AI-Driven Multi-Purpose Platform for Smart Healthcare Systems

by

Md. Milon Islam

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2024

© Md. Milon Islam 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	M. Shamim Hossain Professor, Dept. of Software Engineering, College of Computer and Information Sciences, King Saud University
Supervisor(s):	Fakhri Karray Professor, Dept. of Electrical and Computer Engineering, University of Waterloo
Internal Member:	Sagar Naik Professor, Dept. of Electrical and Computer Engineering, University of Waterloo
Internal Member:	Ramadan El Shatshat Professor, Dept. of Electrical and Computer Engineering, University of Waterloo
Internal-External Member:	Ali Elkamel Professor, Dept. of Chemical Engineering, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The following publications have resulted from the work presented in the thesis:

1. **Md. Milon Islam**, Fakhri Karray, and Ghulam Muhammad, "MSF-Net: Multi-Stage Fusion Network for Emotion Recognition from Multimodal Signals in Scalable Healthcare," Information Fusion, Elsevier, 2024. [Under Review].

2. **Md. Milon Islam**, S M Taslim Uddin Raju, Sheikh Nooruddin, Fakhri Karray, and Ghulam Muhammad, "Internet of Health Things: An Introduction," Blockchain and Digital Twins for the Internet of Medical Things in Smart Hospitals, Elsevier, 2024. [In Press]

3. **Md. Milon Islam**, Sheikh Nooruddin, Fakhri Karray, and Ghulam Muhammad, "Enhanced Multimodal Emotion Recognition in Healthcare Analytics: A Deep Learning based Model-Level Fusion Approach," Biomedical Signal Processing and Control, Elsevier, vol. 94, pp. 106241, Mar. 2024.

4. Sheikh Nooruddin, **Md. Milon Islam**, Fakhri Karray, and Ghulam Muhammad, "A Multi-resolution Fusion Approach for Human Activity Recognition from Video Data in Tiny Edge Devices," Information Fusion, Elsevier, vol. 100, pp. 101953, Jul. 2023.

5. **Md. Milon Islam**, Sheikh Nooruddin, Fakhri Karray and Ghulam Muhammad, "Multi-level Feature Fusion for Multimodal Human Activity Recognition in Internet of Healthcare Things," Information Fusion, Elsevier, vol. 94, pp. 17-31, Jan. 2023.

6. **Md. Milon Islam**, Sheikh Nooruddin, Fakhri Karray and Ghulam Muhammad, "Internet of Things: Device Capabilities, Architectures, Protocols, and Smart Applications in Healthcare Domain," IEEE Internet of Things Journal, IEEE, vol. 10, no. 4, pp. 3611-3641, Dec. 2022.

7. **Md. Milon Islam**, Sheikh Nooruddin, Fakhri Karray, and Ghulam Muhammad, "Human Activity Recognition Using Tools of Convolutional Neural Networks: A State of the Art Review, Data Sets, Challenges, and Future Prospects," Computers in Biology and Medicine, Elsevier, vol. 149, pp. 106060, Oct. 2022.

8. Sheikh Nooruddin, **Md. Milon Islam**, and Fakhri Karray, "TinyHAR: Benchmarking Human Activity Recognition Systems in Resource Constrained Devices," 8^{th} IEEE World Forum on Internet of Things (WF-IoT), IEEE, Yokohama, Japan, pp. 1-8, 26 Oct.–11 Nov., 2022.

9. **Md. Milon Islam**, Sheikh Nooruddin and Fakhri Karray "Multimodal Human Activity Recognition for Smart Healthcare Applications," IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Prague, Czech Republic, pp. 196-203, 9-12 Oct., 2022.

10. Mahmoud Nasr, **Md. Milon Islam**, Shady Shehata, Fakhri Karray, and Yuri Quintana, "Smart Healthcare in the Age of AI: Recent Advances, Challenges, and Future Prospects," IEEE Access, IEEE, vol. 9, pp. 145248-145270, Oct. 2021.

In paper 1, Md. Milon Islam was responsible for the conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, and writing–original draft.

In paper 2, Md. Milon Islam was responsible for the conceptualization, data curation, formal analysis, investigation, methodology, validation, and writing–original draft. S. M. Taslim Uddin Raju and Sheikh Nooruddin were responsible partly for the data curation, formal analysis, investigation, methodology, and writing–original draft.

In papers 3, 5-7, 9, Md. Milon Islam was responsible for the conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, and writing–original draft. Sheikh Nooruddin was responsible partly for the data curation, formal analysis, investigation, methodology, and writing–original draft.

In papers 4, 8, Sheikh Nooruddin was responsible for the conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, and writing–original draft. Md. Milon Islam was responsible partly for the data curation, formal analysis, investigation, methodology, and writing–original draft.

In papers 1-10, Prof. Fakhri Karray was responsible for the supervision, writing–review & editing, guidance, and suggestion. Prof. Ghulam Muhammad aided in the writing–review & editing, guidance, and suggestion for papers 1-7.

In paper 10, both Mahmoud Nasr and Md. Milon Islam were responsible for the conceptualization, data curation, formal analysis, investigation, methodology, validation, and writing–original draft. We contributed equally to this article. Dr. Shady Shehata and Dr. Yuri Quintana aided in the writing–review & editing, guidance, and suggestion.

Chapters 1 to 8 of this thesis contain relevant parts of papers 1 to 10.

Abstract

The significant increase in the number of individuals with chronic ailments has dictated an urgent need for the development of an innovative model for healthcare systems. The design of smart healthcare platforms, a subject of recently growing interest, has become technologically feasible due to the emergence of modern technologies, including smartphones, wearable sensors, 5G communication networks, Internet of Things (IoT), cloud computing, and Artificial Intelligence (AI), particularly machine learning and deep learning. Together, these technologies enable expanded levels of data storage, computation, and secure communication for devices and servers, thus drastically increasing the degree of mobility, security, and available functionality. The thesis is focused on the development of AI-driven data fusion architecture in smart healthcare platform for a scalable, intelligent, and easily deployable remote monitoring system, aimed at providing cost-effective quality healthcare services in assisted living centers.

The contributions of the thesis are two-folds: we mainly focus on two major aspects of human behavior, including Human Activity Recognition (HAR) and emotion recognition to monitor the elderly in smart healthcare. In the first part, we propose two AI-enabled multimodal fusion approaches for HAR in ambient assisted living. We present a deep learning-based fusion approach for multimodal HAR that fuses the different modalities of data to obtain robust outcomes. Initially, Convolutional Neural Networks (CNNs) retrieve the high-level attributes from the image data, and the Convolutional Long Short Term Memory (ConvLSTM) is utilized to capture significant patterns from the multi-sensory data. Afterwards, the extracted features from the modalities are fused through self-attention mechanisms that enhance the relevant activity data and inhibit the superfluous and possibly confusing information by measuring their compatibility. Later, we propose a multi-level feature fusion technique for multimodal HAR using multi-head CNN with Convolution Block Attention Module (CBAM) to process the visual data and ConvLSTM for dealing with the time-sensitive multi-source sensor information. The architecture is developed to be able to analyze and retrieve channel and spatial dimension features through the use of three branches of CNN along with CBAM for visual information. The ConvLSTM network is designed to capture temporal features from the multiple sensors' time-series data for efficient activity recognition.

The second part of this thesis focuses on multimodal emotion recognition in connected healthcare to monitor the patient's health status. To achieve this goal, we introduce two feature fusion approaches through AI tools for emotion recognition. We propose a novel model-level fusion technique based on deep learning for enhanced emotion recognition from multimodal signals to monitor patients in connected healthcare. The representative visual features from the video signals are extracted through the Depthwise Separable Convolution Neural Network, and the optimized temporal attributes are derived from the multiple physiological data utilizing Bi-directional

Long Short-Term Memory. A soft attention method fused the high multimodal features obtained from the two data modalities to retrieve the most significant features by focusing on emotionally salient parts of the features. We exploited two face detection methods, Histogram of Oriented Gradients and CNN-based face detector (ResNet-34), to observe the effects of facial features on emotion recognition. Lastly, we introduce a Multi-Stage Fusion Network (MSF-Net) for emotion recognition capable of extracting multimodal information and achieving significant performances. We propose utilizing the transformer-based structure to extract deep features from facial expressions. We exploited two visual descriptors, local binary pattern and Oriented FAST and Rotated BRIEF, to retrieve the computer vision-based features from the facial videos. A feature-level fusion network integrates the extraction of features from these modules, directing the output into the triplet attention technique. This module employs a three-branch architecture to compute attention weights to capture cross-dimensional interactions efficiently. The temporal dependencies in physiological signals are modeled by a Bi-directional Gated Recurrent Unit (Bi-GRU) in forward and backward directions at each time step. Lastly, the output feature representations from the triplet attention module and the extracted high-level patterns from Bi-GRU are fused and fed into the classification module to recognize emotion.

Finally, we conduct a series of extensive experiments to demonstrate the performance against State-of-the-Art (SOTA) approaches. The findings from the experimental results reveal that the developed multimodal fusion networks surpass the existing SOTA methods in terms of multiple performance metrics. We deployed an IoT system to test the developed feature-fusion networks in real-world scalable smart healthcare application. The developed multimodal predictive analytics frameworks residing in the cloud and trained on large datasets are continually analyzed the nature of the ingested data, making any appropriate notifications to the patients themselves or the healthcare provider through a user-friendly human-machine interface.

Acknowledgements

In the name of Allah, Most Gracious, Most Merciful

First and foremost, I wish to convey my profound appreciation to Allah (SWT), the Almighty, for bestowing upon me the strength, patience, and perseverance required to achieve this endeavor. He has bestowed upon me opportunity, knowledge, and the capacity to tackle challenges, for which I am sincerely grateful. I would like to express my deep gratitude to my PhD supervisor Dr. Fakhri Karray for his continuous help and support throughout my PhD journey. His visionary thinking about smart healthcare research had a great influence on me. I appreciated the freedom he offered me to investigate my research topic and the opportunities he gave me to gain knowledge while working on my thesis.

I would also like to thank my committee members, Dr. M. Shamim Hossain, Dr. Sagar Naik, Dr. Ramadan El-Shatshat, and Dr. Ali Elkamel for their valuable feedback and their time. Moreover, I thank my former committee member Dr. Shady Shehata for his feedback on the background examination to improve my theoretical knowledge. I was extremely fortunate to have such an intellectual committee with me along the path. I would like to express my gratitude to Prof. Ghulam Muhammad for his evaluation, revisions, and valuable suggestions regarding our articles.

Additionally, I want to take a moment to express my gratitude to my dear friends and colleagues, whose inspirational admiration, constructive criticisms, and helpful comments helped me to achieve my goals. Sheikh Nooruddin, I will always remember our extensive conversations and brainstorming sessions, which led to generate ground-breaking concepts and successful publications. Finally, I would like to thank the amazing fellow members, especially, Md. Rezwanul Haque, S M Taslim Uddin Raju, Dr. Hamdi Altaheri, Islam Mohamed Mahmoud Nasr, and Youssef Fathi at the Centre for Pattern Analysis and Machine Intelligence for their help, encouragement, and insightful discussions. I am proud to be a part of such an outstanding community.

At the end, I am so thankful to my mother, father, wife, daughter, son, sister, and brother-in-law who were always next to me, and helped me, during all the ups and downs of starting a new life in a new country. I am very blessed to have you.

Dedication

This is dedicated to the ones I love, especially to my mother and father.

Table of Contents

Examining Committee Membership ii

Author's Declaration iii

Statement of Contributions iv

Abstract vi

Acknowledgements viii

Dedication ix

List of Figures xvi

List of Tables xxii

List of Abbreviations xxv

1 Introduction 1

1.1 Motivation 1

1.2 Problem Statement 3

1.3 Objectives 3

1.4 Contributions 4

1.5 Organization 6

2 Advances and Challenges in AI-Driven Smart Healthcare **8**
2.1 Introduction . . . 8
2.2 Healthcare Monitoring Devices in IoT . . . 11
2.2.1 Wearable Devices for Health Monitoring . . . 12
2.2.2 Smartphone Solutions for Health Monitoring . . . 15
2.3 Artificial Intelligence in IoHT . . . 18
2.3.1 Novel Coronavirus (COVID-19) . . . 18
2.3.2 Heart Disease . . . 21
2.3.3 Diabetes . . . 24
2.4 Ambient Assisted Living . . . 27
2.4.1 User Needs Studies . . . 27
2.4.2 Smart Homes . . . 28
2.4.3 Social Robots . . . 29
2.5 Software Integration Architectures . . . 33
2.6 Discussions, Challenges, and Future Directions . . . 35
2.6.1 Open Discussions . . . 35
2.6.2 Challenges and Further Research Directions . . . 40
2.7 Conclusion . . . 42

3 IoT in Healthcare: Devices, Architectures, Protocols, and Applications **43**
3.1 Introduction . . . 43
3.2 Paper Collection and Selection . . . 47
3.3 IoT Device Capabilities . . . 49
3.3.1 Hardware . . . 51
3.3.2 Software . . . 53
3.4 IoT Architectures . . . 56
3.4.1 Three-Layer Architectures . . . 56
3.4.2 Service-Oriented Architecture . . . 58

3.4.3 Middleware-based Architectures . . . 58
3.5 IoT Protocols . . . 59
3.6 Smart Healthcare Applications . . . 64
3.6.1 Biophysical Parameter Monitoring . . . 65
3.6.2 Chronic Disease Detection . . . 70
3.6.3 Medication Management . . . 72
3.6.4 Telehealth . . . 73
3.6.5 Telesurgery . . . 74
3.6.6 Home and Elderly Care . . . 75
3.7 Discussions, Open Challenges, and Future Trends . . . 78
3.8 Conclusion . . . 86

4 Multimodal Human Activity Recognition through Self-Attention 88
4.1 Introduction . . . 88
4.2 Related Works . . . 89
4.3 Proposed Self-Attention based Fusion Approach . . . 92
4.3.1 Convolutional Neural Networks . . . 92
4.3.2 Convolutional LSTM Network . . . 93
4.3.3 Self-Attention Mechanism . . . 94
4.3.4 Evaluation Protocols . . . 95
4.4 Experimental Results Analysis . . . 96
4.4.1 Dataset Collection and Description . . . 96
4.4.2 Quantitative Results . . . 98
4.4.3 Comparison with the State-of-the-Art . . . 101
4.5 Conclusion . . . 102

5 Hierarchical Architecture for Feature Fusion in Multimodal HAR 103
5.1 Introduction . . . 103
5.2 Related Works . . . 105
5.2.1 Single-modal based HAR . . . 105
5.2.2 Multimodal based HAR . . . 106
5.3 Proposed Hierarchical Architecture for Feature Fusion in Multimodal HAR . . . 108
5.3.1 Multi-head Convolutional Neural Network . . . 110
5.3.2 Convolutional Block Attention Module . . . 111
5.3.3 Convolutional Long Short Term Memory . . . 112
5.3.4 Classification Module . . . 114
5.3.5 IoT Deployment . . . 114
5.3.6 Performance Evaluation . . . 116
5.4 Experiments Results and Discussions . . . 116
5.4.1 Experimental Setup . . . 117
5.4.2 Dataset Description . . . 117
5.4.3 Quantitative Results Analysis . . . 119
5.4.4 Performance in IoT Environment . . . 128
5.4.5 Comparison with the State-of-the-Art . . . 130
5.4.6 Ablation Studies . . . 132
5.5 Conclusion . . . 134

6 Multimodal Emotion Recognition in Healthcare through Model-Level Fusion 135
6.1 Introduction . . . 135
6.2 Related Works . . . 137
6.3 Proposed Model-Level Fusion Approach for Emotion Recognition . . . 140
6.3.1 Depthwise Separable Convolutional Neural Networks . . . 140
6.3.2 Bi-directional Long-Short Term Memory . . . 141
6.3.3 Soft Attention Mechanism . . . 143

6.3.4 Classification Block . . . 143
6.4 Results and Discussions . . . 144
6.4.1 Dataset . . . 144
6.4.2 Results Analysis . . . 146
6.4.3 Comparative Study . . . 154
6.4.4 Ablation Study . . . 156
6.5 Conclusion . . . 158

7 Integrated Design for Multimodal Emotion Recognition in Scalable Healthcare 159
7.1 Introduction . . . 159
7.2 Related Study . . . 161
7.2.1 Emotion Recognition from Physiological Data . . . 161
7.2.2 Emotion Recognition from Facial Cues . . . 162
7.2.3 Emotion Recognition from Multimodal Data . . . 163
7.3 Proposed Multi-Stage Fusion Architecture . . . 165
7.3.1 Vision Transformer . . . 167
7.3.2 Local Binary Pattern . . . 169
7.3.3 Oriented FAST and Rotated BRIEF . . . 170
7.3.4 Triplet Attention Module . . . 170
7.3.5 Bi-directional Gated Recurrent Unit . . . 172
7.3.6 Classification Architecture . . . 173
7.4 Experimental Results and Analysis . . . 173
7.4.1 Experimental Environment . . . 174
7.4.2 Multimodal Datasets . . . 174
7.4.3 Performance Metrics and Baseline Methods . . . 175
7.4.4 Implementation Details . . . 176
7.4.5 Results and Comparison . . . 178
7.4.6 Visual Results . . . 183

7.4.7 Ablation Studies 186
7.4.8 IoT Performance 189
7.5 Conclusion 190

8 Conclusions and Future Works **191**
8.1 Conclusions 191
8.2 Future Works 192

References **195**

List of Figures

2.1 The overall workflow of the reviewed systems of smart healthcare. 10

2.2 A general pipeline of a health monitoring system based on wearable devices. . . . 11

2.3 An overall system architecture of machine learning-based framework for disease diagnosis in IoHT environment. 18

2.4 Generalized architecture for IoMT, comprising of main users, a device layer for data collection, a cloud layer for processing, and user interaction devices. 38

2.5 Generalized architecture for IoMT integrating fog and cloud layers enabling real-time analytics, rapid response, and scalable computing. 39

3.1 Graphical view of the major focuses of this chapter including IoT device capabilities, architectures, protocols, and smart healthcare applications. 46

3.2 Flow diagram showing article selection for final review. The final selection is done based on the duplicate records, abstract review, and full paper review. . . . 48

3.3 IoT device capabilities considering hardware and software platforms. The hardware platform includes several types of tiny processing devices, such as Raspberry Pi, Arduino, and ESP8266. The software module comprises several OSs and built-in tools. 50

3.4 Most common and frequently used IoT architectures that incorporate the three-layer architecture, SOA-based architecture, and middleware-based architecture. . 57

3.5 Most common emerging IoT communication protocols that are extensively used to develop smart IoT applications. 60

3.6 General pipeline of IoT-based smart healthcare systems. The sensors (heart rate, temperature, blood pressure, SpO_2, glucose, and smartphone) collect the data from the real-time healthcare environment. Some of the data are processed in an edge/fog server and the major data analysis occurs in a cloud server where the data analytics engines predict the patient status. The emergencies, families, and physicians receive the patient's condition and take further required actions. 64

3.7 Major healthcare applications focusing on IoT concepts in the area of biophysical parameter monitoring, chronic disease detection, medication management, telehealth, telesurgery, as well as home and elderly care. 66

3.8 Challenges and future research directions of IoT. The major challenges encompass security, reliability, scalability, availability, and cost-efficiency. The potential future directions include the area of blockchain, TinyML, reinforcement learning, and big data processing. 79

4.1 The system architecture of the proposed fusion approach for multimodal human activity recognition. From top to bottom, the two modalities represent input data to the networks, CNN, ConvLSTM, Self-Attention, GAP, fully connected layer with softmax activation function, and activity classification. 91

4.2 t-SNE visualization of multi-sensor human activity recognition dataset on two dimensional space. 97

4.3 Accuracy and loss values of the proposed fusion approach along with two baseline models. (a)-(c) Training (b)-(d) Validation. The first row represents the accuracy values, and the second row depicts the loss vaules. 98

4.4 Confusion matrix for the proposed fusion approach along with two baseline models. The actual activity is shown in the leftmost columns, followed by a prediction in the subsequent columns. (a) CNN (b) ConvLSTM (c) Fusion. 99

4.5 Visualization of the predicted results of the proposed system. The left column represents the original images, and the right column shows the corresponding Grad-CAM. 101

5.1 System architecture of the developed multimodal human activity recognition system in Internet of Healthcare Things. Here, the cameras and multiple sensors collect the human daily activities in the shape of visual and time-series. The deep learning-based fusion network is developed and deployed in the cloud server. The notification is sent to the stakeholders through 5G communication. 108

5.2 Schematic diagram of the developed fusion architecture for multimodal human activity recognition in a smart healthcare framework. The time-series data recorded from multi-sensors are fed into ConvLSTM architecture to get the temporal features (see top part). A three-head CNN architecture with CBAM extracts the high-level channel and spatial dimension features from the visual data and the outputs of each head are get fused (see lower part). Additionally, the output features of two-branch networks are fused and passed to the classification module through the GAP layer. Finally, the Softmax function is utilized to provide output for HAR. 109

5.3 The architecture of ConvLSTM. The new memory c_t and output h_t are formed by updating the internal memory c_{t-1} based on the current input x_t and the prior output h_{t-1}. 112

5.4 Block diagram of user application and cloud server application for interaction for multimodal human activity recognition in IoHT. 115

5.5 Performance measures (accuracy and loss curves) of the developed fusion approach along with two baseline architectures. (a)-(c) Training (b)-(d) Validation. The first row represents the accuracy, and the second row depicts the loss. 118

5.6 Confusion matrices for the developed fusion network along with two baseline architectures. The rows (activity label) represent the actual activities, and the columns (predicted label) show the recognized activities. The deeper color presents the better results. (a) CNN ($C2$) (b) ConvLSTM (TS) (c) Fusion ($C1+TS$) (d) Fusion ($C2+TS$) (e) Fusion ($C1+C2+TS$). 120

5.7 Boxplot for the comparative analysis of the proposed fusion architecture with the baseline models. (a) Accuracy (b) Precision (c) Recall (d) F1-Score. 127

5.8 Grad-CAM of each activity sample on UP-Fall detection dataset for the proposed Fusion ($C2+TS$) architecture. 128

5.9 Performance in IoHT deplyment for multimodal HAR. (a)-(c) Latency (b)-(d) Response time. The first row represents the performance in smart home environment, and the second row depicts the performance in virtual care environment. . . 129

6.1 The proposed model-level fusion technique for multimodal emotion recognition in healthcare analytics. The data from the humans are collected in the form of videos, and physiological signals. Later, the Bi-LSTM network extracts the features from the physiological signals and the DSCNN architecture retrieves the significant patterns from the faces. Finally, the soft attention mechanism extracts the most essential features from the fused attributes that are passed to the classification block over the GAP layer for emotion recognition. 139

6.2 The internal structure of Bi-LSTM network. 142

6.3 Accuracy curves for multimodal emotion recognition of the proposed fusion network trained for 500 epochs. (a)-(b) Training and Validation outcomes for the HOG face feature detector. (c)-(d) Training and Validation outcomes for the ResNet-34 face feature detector. 145

6.4 Loss curves for multimodal emotion recognition of the proposed fusion network trained for 500 epochs. (a)-(b) Training and Validation outcomes for the HOG face feature detector. (c)-(d) Training and Validation outcomes for the ResNet-34 face feature detector. 147

6.5 Confusion matrices of the developed models using HOG feature detector for multimodal emotion recognition. (a) Attention (b) No Attention (c) DSCNN (d) Bi-LSTM. The Bi-LSTM network has not used the HOG features as the inputs of this architecture are the physiological signals. Here, E1: Amusement, E2: Anger, E3: Disgust, E4: Fear, and E5: Sad. 148

6.6 Confusion matrices of the developed models using the ResNet-34 feature detector for multimodal emotion recognition. (a) Attention (b) No Attention (c) DSCNN. Here, E1: Amusement, E2: Anger, E3: Disgust, E4: Fear, and E5: Sad. 149

6.7 ROC curves of the proposed fusion architecture and different baseline methods for multimodal emotion recognition. (a) present the ROC curves for the HOG face feature detector. (b) present the ROC curves for the ResNet-34 face feature detector. 150

6.8 Visual representations of the results obtained by the three models: 'DSCNN', 'No Attention', and 'Attention' using two facial feature descriptors, including HOG and ResNet-34. All measures are in percentage. 151

6.9 t-SNE to visualize the feature distributions of the developed network along with the baseline methods. (a) & (e) Attention (b) & (f) No Attention (c) & (g) DSCNN (d) Bi-LSTM. (a)-(c) utilized the HOG facial feature detector. (e)-(g) exploited the ResNet-34 facial feature detector. 154

7.1 A framework of the proposed architecture for emotion recognition from the multimodal data in scalable healthcare. The visual and physiological signals are collected from the patients in AAL environment through multiple sensors. In the cloud platform, the MSF-Net developed using deep learning tools is deployed for efficient emotion recognition. The patients status is informed to the immediate mediators through proper communication channel. 165

7.2 An overview of the proposed MSF-Net for multimodal emotion recognition. MSF-Net includes several stages: visual feature extraction, triplet attention, temporal feature extraction, and emotion recognition. The visual feature extraction module comprises ViT, LBP and ORB methods that simultaneously extract the deep and computer vision-based features from the facial expression. The retrieved features are fused in the visual feature extraction module and sent to the triplet attention module. The triplet attention exploits three attention branches to capture the cross-interaction among the extracted signals. On the other hand, MSF-Net encodes the physiological signals with Bi-GRU to obtain the temporal features. Afterward, the extracted features from two branches are fused to achieve the feature representation. Finally, we classify the final feature representation in the emotion classifier. 166

7.3 Visualization of confusion matrices of MSF-Net on the testing set of the BioVid Emo DB database. The columns represent the actual labels, while the rows represent the predicted ones. (a) ViT+Attention+Bi-GRU (b) ViT+LBP+Attention+Bi-GRU (c) ViT+ORB+Attention+Bi-GRU (d) ViT+LBP+ORB+Bi-GRU (e) ViT + LBP+ORB+Attention+Bi-GRU. Here, B1: Amusement, B2: Anger, B3: Disgust, B4: Fear, and B5: Sad. 183

7.4 Visualization of confusion matrices of MSF-Net on the testing set of the MGEED database. The columns represent the actual labels, while the rows represent the predicted ones. (a) ViT+Attention+Bi-GRU (b) ViT+LBP+Attention+Bi-GRU (c) ViT+ORB+Attention+Bi-GRU (d) ViT+LBP+ORB+Bi-GRU (e) ViT+LBP + ORB+Attention+Bi-GRU. Here, M1: Neutral, M2: Happy, M3: Sad, M4: Angry, M5: Fear, M6: Disgust, and M7: Surprise. 183

7.5 t-SNE feature visualization of the MSF-Net for multimodal emotion recognition. (a) & (f) ViT+Attention+Bi-GRU (b) & (g) ViT+LBP+Attention+Bi-GRU (c) & (h) ViT+ORB+Attention+Bi-GRU (d) & (i) ViT+LBP+ORB+Bi-GRU (e) & (j) ViT+LBP+ORB+Attention+Bi-GRU. Different categories of emotions are represented with distinct colors and shapes. The first row represents the results from the BioVid Emo DB dataset, and the second row depicts the results from the MGEED database. 184

7.6 Performance measures for IoT deployment in scalable healthcare scenario. (a) Latency (b) Response time. 189

List of Tables

2.1 Summary of the smart health monitoring systems based on wearable devices. . . 14
2.2 Summary of the smart health monitoring systems based on smartphone. 17
2.3 Summary of COVID-19 detection frameworks in IoHT environment. 20
2.4 Summary of heart disease detection systems in IoHT environment. 23
2.5 Summary of diabetes detection frameworks in IoHT environment. 26
2.6 Summary of the social robots implementations for care of older adults. 30

3.1 IoT hardware components with some properties. 52
3.2 A short look at the IoT operating systems. 55
3.3 IoT communication protocols with some fundamental features. 63
3.4 Summary of challenges in large-scale deployment of IoT networks and future research recommendations. 85

4.1 Recognition performance of the proposed fusion approach. 100
4.2 Performance comparison of the proposed approach with the state-of-the-art for UP-Fall detection dataset. 102

5.1 Description of the activities performed by human participants in UP-Fall detection dataset. 117
5.2 Percentile precision (precision % $\pm$ standard deviation) of the proposed fusion architecture for each activity classification. 122
5.3 Percentile recall (recall % $\pm$ standard deviation) of the proposed fusion architecture for each activity classification. 123

5.4 Percentile F1-Score (F1-Score % ± standard deviation) of the proposed fusion architecture for each activity classification. . . . 124
5.5 Activity recognition results of the proposed fusion approach along with baseline models for the test samples. . . . 125
5.6 Performance comparison of the results obtained by the proposed fusion approach and the recent studies on UP-Fall detection dataset. . . . 131
5.7 Experimental findings of different variants of the proposed fusion architecture on UP-Fall detection dataset. . . . 133

6.1 Performance measures of the proposed fusion architecture for multimodal emotion recognition. All measures are in percentage. . . . 151
6.2 Emotion-wise performance measures of the proposed fusion approach for Bio Vid Emo DB dataset. All measures are in percentage. . . . 153
6.3 Accuracy comparison of the proposed fusion approach with the existing literature for the Bio Vid Emo DB dataset. . . . 155
6.4 Effects of attention module placement on the performance for emotion recognition in the proposed model-level fusion approach. All measures are in percentage. 157

7.1 Performance (%) of MSF-Net against SOTA methods on the BioVid Emo DB database. The best performances are represented by bold front, while the second-best outcomes are marked with underlined text. . . . 177
7.2 Performance (%) of MSF-Net against SOTA methods on the MGEED database. The best performances are represented by bold front, while the second-best outcomes are marked with underlined text. . . . 179
7.3 Performance (%) comparison of different approaches on the BioVid Emo DB dataset with multi-modality. We provide each emotion categories classification performances. The best performances are represented by bold front, while the second-best results are marked with underlined text. . . . 181
7.4 Performance (%) comparison of different approaches on the MGEED dataset with multi-modality. We provide each emotion categories classification performances. The best performances are represented by bold front, while the second-best results are marked with underlined text. . . . 182

7.5 Ablation analysis (%) of MSF-Net with different combinations of multimodal signals on the BioVid Emo DB and MGEED datasets. We provide the overall performance on the testing set. The best performances are represented by bold front, while the second-best results are marked with underlined text. 185

7.6 Ablation analysis (%) of MSF-Net with different combinations of multimodal signals on the BioVid Emo DB dataset. We provide each emotion categories classification performances. The best performances are represented by bold front, while the second-best results are marked with underlined text. 187

7.7 Ablation analysis (%) of MSF-Net with different combinations of multimodal signals on the MGEED dataset. We provide each emotion categories classification performances. The best performances are represented by bold front, while the second-best results are marked with underlined text. 188

List of Abbreviations

AAL Ambient Assisted Living 2–4, 6, 7, 9–11, 27–29, 31–37, 41, 42, 75, 88, 103, 107, 194

ADL Activities of Daily Living 4, 34, 37, 96, 97, 106, 108, 117, 126

AI Artificial Intelligence 1, 3, 4, 6, 18, 24, 56, 65, 74, 80, 83, 88, 127, 193, 194

AMQP Advanced Message Queuing Protocol 46, 61

ANFIS Adaptive Neuro-Fuzzy Inference System 22, 36

AUC Area Under the Curve 89, 150

AWS Amazon Web Services 45, 55, 56, 67, 76, 78

BCN Bilinear Convolution Network 139

Bi-GRU Bi-directional Gated Recurrent Unit 105, 167, 172, 173, 175, 176, 190, 192

Bi-LSTM Bi-directional Long Short-Term Memory 5, 90, 136, 139–142, 145–147, 149, 150, 152, 154, 156–158, 161, 192

BLE Bluetooth Low Energy 28, 35, 37, 46, 62, 63, 67, 70, 75, 78

BNN Backpropagation Neural Network 24

BP Blood Pressure 13, 15, 16, 22, 36, 65, 67, 68, 76, 193

BR Breathing Rate 16

C-MHAD Continuous Multimodal Human Action Dataset 107

CAM Channel Attention Module 111, 132, 134, 191

CBAM Convolution Block Attention Module 5, 7, 105, 108, 109, 111, 132, 134, 191

CK+ Cohn Kanade 163

CNN Convolutional Neural Network 4, 5, 7, 19, 20, 32, 36, 74, 76, 89–92, 97–102, 104–110, 112–114, 119, 121, 126, 130, 132, 134, 136–138, 141, 145, 160–164, 175, 176, 191

CoAP Constrained Application Protocol 46, 53, 55, 59, 60

ConvLSTM Convolutional Long Short Term Memory 4, 5, 7, 92–94, 97–100, 102, 105, 108, 109, 112–114, 118, 119, 121, 126, 132, 134, 160, 191

CoRE Constrained RESTful Environments 59

COTS Commercial off-the-Shelf 33, 39, 42

CSI Channel State Information 76, 77

DaaS Database-as-a-Service 84

DBN Deep Belief Network 138, 154, 175

DCCA Deep Canonical Correlation Analysis 163

DDS Data Distribution Service 46, 61

DL Deep Learning 4, 19, 41, 74, 88–90, 92, 102, 104–108, 110, 136–138, 140, 150, 158, 160, 161, 163–165, 174, 175, 189, 191, 192

DLaaS Deep Learning as a Service 115

DLMNN Deep Learning Modified Neural Network 21

DSCNN Depthwise Separable Convolution Neural Network 5, 136, 140, 141, 146, 147, 150, 152, 156–158, 192

DSS Digital Signature Standard 80

ECG Electrocardiogram 12, 13, 20, 22, 24, 32, 65–67, 71, 74, 135, 136, 140, 142, 144, 145, 161, 165, 174–176, 186, 189, 190, 192

EEG Electroencephalogram 108, 135, 136, 139, 160–165, 174–176, 189

EMG Electromyogram 135, 142, 161, 165

FC Fully Connected 90, 92, 93, 95, 98, 99, 106, 109, 113, 114, 156–158, 163, 164, 169, 190

FER Facial Emotion Recognition 162, 163

FMCW Frequency Modulated Continuous Wave 76

FN False Negative 96

FP False Positive 96

GAP Global Average Pooling 90, 92, 95, 99, 109, 114, 140, 143, 146, 157, 158, 192

GPIO General Purpose Input Output 51

GPS Global Positioning System 15, 21, 50, 57, 76, 193

Grad-CAM Gradient-weighted Class Activation Mapping 100, 127

GRU Gated Recurrent Unit 90, 101, 106, 130

GSR Galvanic Skin Response 161, 174, 175, 186, 189

HAR Human Activity Recognition 3–5, 7, 76, 77, 88–90, 93, 96, 99–108, 116, 117, 119, 121, 126, 130, 134, 191–193

HCI Human-Computer Interaction 135, 136, 159

HD Heart Disease 8, 11, 21, 22, 24, 36, 41, 42, 70, 71

HOG Histogram of Oriented Gradients 6, 140, 145–147, 149, 150, 152, 156–158, 192

HR Heart Rate 12, 13, 15, 16, 21, 31–33, 35, 65, 73, 135, 193

HRI Human-Robot Interaction 29, 31, 34, 164

HSI Human System Interface 74

HTTP Hypertext Transfer Protocol 26, 55, 56, 59, 67

HWU-USP Heriot-Watt University/University of Sao Paulo 107

IETF Internet Engineer Task Force 59

IMU Inertial Measurement Unit 32

IoHT Internet of Health Things 5, 6, 8, 18, 21, 22, 24, 26, 34, 36, 41, 90, 103, 104, 108, 116, 134, 191

IoMT Internet of Medical Things 8, 22, 28, 33, 36, 77, 78, 137

IoRT Internet of Robotic Things 31

IoT Internet of Things 2–9, 11–13, 18–22, 24–27, 31, 32, 37, 40, 43–62, 64–78, 80–84, 86–89, 105, 106, 108, 109, 114, 116, 128, 134, 137, 165, 189–193

IP Internet Protocol 43, 52, 53, 62, 81, 115, 116

k-NN k-Nearest Neighbors 19, 26, 89

LBP Local Binary Pattern 6, 167, 169, 170, 176, 178, 190, 192

LCD Liquid Crystal Display 12, 13, 72, 73

LLM Large Language Model 7, 193

LN Layer Normalization 169

LoWPAN Low Power Wireless Personal Area Networks 46, 53, 61, 62

LPG Liquefied Petroleum Gas 28

LPWAN Low-Power Wide-Area Network 62

LSTM Long Short-Term Memory 89, 90, 93, 101, 105–107, 112, 113, 130, 136, 141, 142, 160, 162, 164

LTE Long-Term Evolution 62

M2M Machine-to-Machine 60, 61

MDCNN Modified Deep Convolutional Neural Network 22

MFCC Mel Frequency Cepstral Coefficients 19, 160, 164

MGEED Multimodal Genuine Emotion and Expression Detection 172–176, 178, 180, 183–186, 189, 190

MHEALTH Mobile HEALTH 106

MHSA Multi-Head Self-Attention 167, 169

ML Machine Learning 8, 10, 11, 15, 18, 19, 21, 22, 24–26, 28, 35–37, 41, 42, 44, 56, 67, 71, 74, 76, 83, 88–90, 104–106, 130, 136–138, 160

MLP Multilayer Perceptron 89, 90, 111, 134, 160, 167, 169

MQTT Message Queuing Telemetry Transport 26, 29, 31, 34, 35, 37, 46, 54–56, 60, 67, 78

MSF-Net Multi-Stage Fusion Network 6, 167, 169, 170, 173–176, 178, 180, 183–185, 190, 192

MSSO Modified Salp Swarm Optimization 22, 36

NB-IoT Narrowband Internet of Things 62

NFT Non-Fungible Token 80

NLU Natural Language Understanding 31, 32

NR New Radio 62

OMG Optomyography 174–176, 189

ORB Oriented FAST and Rotated BRIEF 6, 167, 170, 176, 178, 190, 192

OS Operating System 45, 46, 51, 53, 54, 78, 80, 84

P2P Peer-to-Peer 61, 63

PaaS Platform-as-a-Service 84, 130

PCG Phonocardiogram 15

PPG Photoplethysmogram 15, 16, 70, 76

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses 47

QoS Quality of Service 2, 56, 59, 61, 64

RAT Radio Access Technology 62

ReLU Rectified Linear Unit 98, 111, 114, 143, 146

REST Representational State Transfer 31, 37, 54, 59

RF Random Forest 89, 105, 137, 160

RF Radio Frequency 67, 68

RHM Remote Health Monitoring 2, 3, 12, 33, 34, 65

RISC Reduced Instruction Set Computer 52

RNN Recurrent Neural Network 71, 104, 112, 136, 138, 160

ROC Receiver Operating Characteristic 91, 105, 149

ROS Robotic Operating System 29, 31, 32, 34, 37

RSSI Received Signal Strength Indicator 28, 37, 76, 77

SA Self-Attention 4, 7, 89, 90, 92, 94, 95, 98, 99, 102, 167, 169, 191

SaaS Software-as-a-Service 84

SAM Spatial Attention Module 111, 112, 132, 134, 191

SCL Skin Conductance Level 135, 136, 140, 142, 144, 145, 174, 186, 190, 192

SMO Sequential Minimal Optimization 25

SOA Service-Oriented Architecture 35, 45, 46, 58, 82, 86

SOTA State-of-the-Art 4–7, 10, 42, 44, 78, 89, 96, 105, 117, 130, 134, 156, 159, 173, 175, 178, 180, 183, 192

SPI Serial Peripheral Interface 52

SpO_2 Oxygen Saturation 12, 15, 16, 19, 32, 65, 69, 70, 74, 193

SVM Support Vector Machine 19, 21, 25, 26, 71, 74, 77, 89, 90, 101, 130, 136–138, 154, 175, 176

t-SNE t-distributed Stochastic Neighbor Embedding 97, 152

TCP Transmission Control Protocol 52, 59–61, 116

tEMG Trapezius Electromyogram 140, 144, 145, 174, 190, 192

TinyML Tiny Machine Learning 7, 80, 83, 84

TN True Negative 96

TP True Positive 96

UCI University of California Irvine 89, 90, 106

UDP User Datagram Protocol 59, 116

UTD-MHAD UTD Multimodal Human Action Dataset 107

VANET Vehicular Ad-hoc Network 83

VDA Village Doctor App 73

ViT Vision Transformer 6, 167, 169, 176, 178, 192

WISDM Wireless Sensor Data Mining 90, 106

wkPCA Wavelet-based Kernel Principal Component Analysis 24

XMPP Extensible Messaging and Presence Protocol 46, 60, 61

3GPP 3rd Generation Partnership Project 62

5G Fifth Generation 2, 19, 25, 45, 54, 62, 74, 108, 137, 167

6G Sixth Generation 9, 54

Chapter 1

Introduction

1.1 Motivation

With projections of 22% of the population reaching the age 60 or more by 2050 [1], people affected by chronic diseases are growing along with health-related emergencies, therefore resulting in a higher pressure on the healthcare industry [2]. With a decline in the ratio between working-age people, fewer professional healthcare workers for increased demand. Besides, the cost of said health care, medications, and medical devices continuously soar, making it harder to cover such costs for the average citizen as the need for more caregivers and healthcare facilities increases to with-stand the increase in demand [3]. Combined, these conditions call for cheaper, more inclusive, and better health care solutions. For a few years now, around the world, public and private healthcare industry stakeholders have been hard at work devising a smart healthcare infrastructure that will revamp and help reshape the healthcare sector. This goal has never been more pressing than in a time of pandemics such COVID-19, when hospitals have been overwhelmed with patients who require immediate care and treatment. The design and development of remote healthcare systems have become an active and well-funded research area, as evidenced by the recent advances in the field. A great candidate for such a situation is utilizing the recent advancements in smart and miniaturized sensors, communication technologies, and Artificial Intelligence (AI) to provide technological solutions at an affordable price to the broadest range of the population without sacrificing the quality of care. Wearable and non-invasive sensors, actuators, and new communication technologies [4] are providing an accurate and cost-effective approach for the remote monitoring of patients. Blockchain technology ensures security and privacy while data from the end user can be streamed to servers and edge/fog devices over the Internet [5]. AI techniques [6] are becoming more adept at classifying a wide range of ailments based on patient

symptoms and vitals.

A healthcare platform is described as a combination of hardware and software components that work together to give a wide range of healthcare services and applications to people including healthcare professionals and patients, to promote health in an efficient and widespread manner [7], [8]. Nowadays, there are numerous healthcare frameworks available that integrate various technologies to monitor multiple human biophysical symptoms and environmental data utilizing various wireless communications techniques including Zigbee, Fifth Generation (5G), Bluetooth, and Wi-Fi [9], [10]. Internet of Things (IoT) technology supports healthcare systems by allowing people to remain at smart homes and be monitored in real-time rather than being admitted to hospitals or clinics which reduces emergency costs [11], [12]. There are several advantages of IoT technologies to design smart healthcare systems such as ease of access, ensuring patient comfort and safety, and reducing patient's burden in hospitals. In general, IoT enables networks of smart devices, cloud applications, and solutions to make data transfer and storage easier. The most promising applications of IoT in the area of healthcare include remote monitoring, smart medical devices, smart homes, as well as wearable devices. In recent years, several academic and industry researches on IoT interoperability including in the healthcare domain have been done, with an emphasis on the standardization of communication protocols to offer interoperability of diverse devices, networks, and data structures [13], [14]. To ensure Quality of Service (QoS) in healthcare systems, IoT should incorporate some aspects such as standardization of reliable communication protocols, enhanced mobile and wearable devices, and low-cost as well as low-power embedded processors [15], [16]. The IoT communication architecture could be the primary enabling mechanism for decentralized pervasive healthcare applications, considering the prevailing availability of distant wireless healthcare platforms and the rising topic of electronic patientcare datasets [17], [18]. One of the main drivers for smart healthcare adoption is the increasing ratio of older adults in societies worldwide. Consequently, Ambient Assisted Living (AAL) focuses on creating environments for older adults that integrate smart healthcare techniques for better care without human intervention.

While many advances have been made in the area of Remote Health Monitoring (RHM) and commercial products have been created, issues related to the implementation of technology for a wider population range has yet to be addressed. The challenges are due primarily to shortcomings in tackling efficiently the information gathered from the wearable devices, and the nature of the closed architecture of these systems, preventing them from being interoperable among various points of care such as test/imaging labs, pharmacies, hospitals, specialists, long term care, and home care [19].

1.2 Problem Statement

The existing healthcare system is facing difficulties to ensure cost-effective and high-quality health services to the growing global population and the increasing patients burden. In typical healthcare services, patients occasionally consult physicians for health assessments, which are unable to satisfy the increasing expectations for personalized care. The new model will be more personalized and less reliant on traditional brick-and-mortar healthcare institutions such as hospitals, nursing homes, and long-term healthcare centers.

In recent years, wearable sensors and IoT devices have collected an enormous amount of unstructured information from patients; however, the generated raw data in various forms poses major issues. The crucial aspects of healthcare, such as human actions and emotions, present major hurdles while dealing with heterogeneous data captured by IoT devices for real-time recognition. Human Activity Recognition (HAR) and emotion recognition are critical human behaviors for patient condition monitoring, requiring reliable architectures that can handle data from a variety of sources, including physiological data from wearable devices, visual data from cameras, and ambient information from sensors. Most existing systems for HAR and emotion detection fail to fuse and analyze various forms of data, resulting in incorrect health status recognition which decreases the efficiency of healthcare system.

Given these limitations, there is an urgent need to develop an intelligent platform capable of fusing multimodal data in real-time, retrieving significant features and patterns for accurate recognition of both physical behaviors and emotional states, as well as assisting in decision-making. Hence, the major issue addressed in this thesis is designing AI-driven multi-purpose platform to monitor the patients activities and health status through multimodal data fusion in AAL environment. The challenge is to develop an intelligent system capable of integrating data from diverse sources, interpreting them efficiently, and offering relevant insights to healthcare professionals.

1.3 Objectives

The smart healthcare sector, which includes telemedicine and eHealth, is fast becoming a game-changer in the healthcare industry. As a contribution to this expanding sector, the proposed research is aimed at developing a suite of technologies centered around a smart healthcare platform for RHM and AAL. The thesis is carried out to meet the specific goal to develop an AI-driven multi-purpose platform for smart healthcare system that will allow patients to be monitored remotely using multimodal signals collected from their different body parts. This thesis focuses on two human behaviors: HAR and emotion recognition, and addresses difficulties of multimodal data fusion, real-time monitoring, and scalability.

The objectives of this research are given below.

- **Introduce feature fusion architectures for multimodal HAR for activity monitoring**: This objective focuses on fusing heterogeneous data from multiple sources, including wearables, stationary cameras, and ambient sensors to recognize Activities of Daily Living (ADL) through AI tools in AAL.
- **Design multimodal emotion recognition platform for health status monitoring**: This objective aims to develop reliable emotion recognition frameworks leveraging model-level and multi-stage fusion approaches derived from physiological signals and behavioral data within healthcare contexts.
- **Develop a deployable platform for patient's monitoring through IoT**: This objective emphasizes on deploying the proposed fusion architectures in IoT networks for remote monitoring to manage substantial data amounts and ensure adaptability to different healthcare needs without considerable technical overhead.

1.4 Contributions

The thesis makes several contributions in the area of AI-driven smart healthcare to monitor the elderly. The major contributions are described as follows.

Firstly, the proposed research is focused on developing HAR system from multimodal data through Self-Attention (SA) based fusion for smart healthcare applications. The major contributions of this research are mentioned below.

- Design a Deep Learning (DL)-based fusion network to efficiently recognize human activities through the signals from wearable and stationary devices in smart healthcare.
- Exploit Convolutional Neural Network (CNN)s and Convolutional Long Short Term Memory (ConvLSTM) to retrieve features from images and wearable sensors signals.
- Propose to utilize SA technique to fuse the extracted features from both branches to explore the prominent features from the original data.
- Demonstrate the efficiency of the proposed work through extensive experiments against State-of-the-Art (SOTA) and baseline models.

Secondly, we introduce a hierarchical fusion architecture for multimodal HAR in Internet of Health Things (IoHT) environment. This is the first work that applies multi-head CNNs integrated with Convolution Block Attention Module (CBAM) to concurrently perform channel and spatial feature fusion of visual information based on the existing SOTA literature in this area. Additionally, we first used all the data including time-series and images of this dataset for our experiments. The key contributions of this work are given below.

- A fusion technique is introduced for multimodal HAR that can fuse the information of cameras and multiple sensors simultaneously through the fusion of the output of multi-head CNNs with ConvLSTM in IoHT environment.
- An efficient feature fusion approach is developed to retrieve the channel and spatial dimension attributes using CBAM and fusing the output of each head of CNN architecture.
- A ConvLSTM-based architecture is developed with a multi-sensor based data fusion method to handle sequential data and retrieve the high-level temporal features from the raw signals.
- We compare the experimental findings of the proposed fusion architecture with two baseline models along with the SOTA systems on a large multimodal dataset called UP-Fall detection dataset to verify the performance of our developed architecture.
- We compare the model performances on two different IoT deployment scenarios.

Thirdly, we propose a model-level fusion architecture for multimodal emotion recognition. This is the first research, we are exploiting soft attention mechanisms along with Depthwise Separable Convolution Neural Network (DSCNN) and Bi-directional Long Short-Term Memory (Bi-LSTM) networks to detect emotion from multimodal data. This research provides the following significant contributions, to sum up.

- We develop a novel model-level fusion technique for emotion recognition to monitor patients' health conditions in connected healthcare through multimodal signals.
- We exploit the DSCNN architecture to capture the high-level attributes from the facial videos and Bi-LSTM network to retrieve the temporal features from the physiological data.
- We fuse the extracted features through soft attention mechanism that enabled to find out the most relevant features to recognize emotion.

- We employ two distinct facial detection techniques, namely Histogram of Oriented Gradients (HOG) and ResNet-34, to investigate the influence of facial attributes on emotion recognition performance.
- We perform extensive experiments considering a few baseline architectures with the Bio Vid Emo DB dataset to demonstrate the efficiency of the proposed fusion architecture.

Fourthly, we present an integrated fusion architecture for emotion recognition from multimodal signals in scalable healthcare platform. This is the first research of proposing multi-stage feature fusion architecture to recognize emotion from multimodal data to fuse deep features and computer vision-based features by exploiting Vision Transformer (ViT) along with the Local Binary Pattern (LBP) and Oriented FAST and Rotated BRIEF (ORB) feature descriptors. Moreover, this architecture exploits the triplet attention mechanism to exhibit more importance on the fused visual features for the first time. This proposed Multi-Stage Fusion Network (MSF-Net) has the following main strengths.

- We introduce a multi-stage fusion network, called MSF-Net, to effectively leverage the visual features and temporal relationships of the data for recognizing discrete dimensional emotions in a multimodal context.
- We propose to exploit an approach that enables the efficient extraction of both visual features and temporal dependencies from the data, without any interference between them.
- We design an innovative architecture based on a triple attention technique, enabling the extraction of significant emotional feature representations by developing three attention branches to record cross-interactions among the visual features.
- We conduct comprehensive evaluations to investigate the performance of our proposed fusion network on two popular multimodal databases and obtain the new SOTA record. Moreover, the proposed MSF-Net is deployed in IoT network to demonstrate the scalability.

1.5 Organization

The rest of the thesis is organized as follows:

Chapter 2 presents an overview of the AI-driven smart healthcare platform, emphasizing on current advancements and challenges. The key subjects covered in this chapter include healthcare monitoring devices, AI models for IoHT, AAL solutions, and software integration architectures.

Chapter 3 covers various theoretical aspects of IoT, such as IoT device capabilities (hardware and software), IoT architecture, and IoT communication protocols. This chapter also discusses IoT applications in smart healthcare, including biophysical parameter monitoring, chronic disease management, medication adherence, telehealth services, and home care solutions for the elderly. Chapter 4 introduces a SA-based fusion network for multimodal HAR in healthcare applications. This chapter demonstrates the proposed fusion network, that fuses data from wearable devices and stationary cameras in an AAL environment. It also describes thorough experimental assessments performed on the benchmark multimodal UP-Fall detection dataset.

Chapter 5 offers a hierarchical feature fusion network for multimodal HAR based on multi-head CNNs, ConvLSTM, and the CBAM network. This chapter compares the performance to SOTA approaches and explains the way to deploy the trained model in an IoT environment. Chapter 6 demonstrates the model-level feature fusion architecture for emotion recognition considering multiple data modalities, including facial cues and physiological data from the patients. The chapter illustrates the thorough analysis of the experimental outcomes along with SOTA performances. Chapter 7 presents the model-level feature fusion architecture for emotion recognition leveraging various data modalities, including facial cues and physiological data from patients. The chapter provides an in-depth investigation of the experimental results as well as SOTA performance.. Chapter 8 briefly mentions the major contributions of the dissertation focusing on multimodal HAR and emotion recognition in healthcare context. This chapter outlines potential research directions, such as incorporating more human behaviors and data modalities, deploying Large Language Model (LLM)s, introducing novel Tiny Machine Learning (TinyML) solutions, and developing a fully functional healthcare platform.

Chapter 2

Advances and Challenges in AI-Driven Smart Healthcare

2.1 Introduction

The IoT [20], [21] has been steadily increasing in popularity over the past years. Due to the advancements in communication technologies and data transfer speed, the ability to transmit large amounts of data has grown drastically. In addition, more robust and advanced storage and processing capabilities provided by big data analytics [22], and cloud technologies [23] opened the doors for new applications and markets for IoT in real-time analytics and predictive modelling. As a result of the earlier described advancements in combining smart sensors, actuators, and data analytics in an IoT environment for real-time and continuous healthcare brings great promise to the healthcare industry. The field, called the IoHT [24] or the Internet of Medical Things (IoMT) [25] offers the potential of transforming the healthcare paradigm. The method which is pursued in research and practice can be divided into several areas. First, smart sensors are combined in the IoMT environment to continuously monitor health vitals using wearable devices [26], [27] and smartphone solutions such as those proposed by [28], [29]. After the data is collected using smart sensors, Machine Learning (ML) techniques interpret and present predictive analytics such as predicting illnesses [30], [31]. In addition, other algorithms are used to keep track of chronic conditions such as diabetes [32] and Heart Disease (HD)s [33] and detect abnormalities in the patient's health.

Since around 90% of older adults prefer staying in their own homes, many solutions are based on smart home systems proposed in [34], [35]. User studies [36], [37], however, uncover the importance of including robotic agents capable of social interactions with the user to provide both

psychological and physical assistance to older adults. Several studies tackle using robotic agents for the care of older adults [38], [39], while others propose taking a step further by integrating such robots in the AAL environments with other smart sensors [40], [41]. Integrating various sensors, actuators, and user interfaces requires rigorous work on a scalable and personalization to different user needs. Therefore, studies such as [42], [43] attempt to formalize different architecture to tackle this problem and create integrated smart healthcare.

The concept of fog and edge computing plays a vital role in smart healthcare by reducing the computing burden of cloud servers and ensuring real-time healthcare services with the fastest response time. Fog computing is heavily decentralized and serves as a bridge between end devices and the cloud for computing, storage, and networking. The fundamental idea behind fog computing is to move data center tasks to fog nodes located at the network's edge. The fog devices carry out the tasks located at the network's edge and achieve a high data transfer rate and low response time. The data in traditional healthcare systems is analyzed in the cloud server, which results in high latency and high bandwidth requirements for large data. In the fog-enabled healthcare framework, the data from the IoT devices is collected and processed in the fog layer that minimizes the latency and ensures real-time medical care [44]. Edge computing refers to the installation of computing and storage resources at the point of data is being acquired and analyzed. This framework includes sensors for data collection and edge servers for secure real-time data processing. Edge computing enables intelligent and real-time healthcare services that address energy efficiency and latency requirements through IoT devices and Sixth Generation (6G) environments [45], [46]. The edge computing schemes offer potential solutions to increase reliability and responsiveness in decentralized applications, including healthcare, as mapping IoT devices and sensors and managing resources are critical components for a smart healthcare platform [47].

With the rapid increase of IoT devices, huge amounts of electronic health data are generated daily. It is necessary to keep the data more secure. The integration of blockchain and smart healthcare could mitigate the shortcomings of conventional smart healthcare in data sharing, data security, and privacy maintenance [48]. Blockchain technology works by linking secure blocks of data together using an encrypted data record. Here, the data is kept in synchronized database systems recreated with no need for a centralized administrator. It provides effective data collecting while ensuring system security because it is a distributed system disseminated across a network. Although several works have already been done to ensure the security and privacy of the health data, some of the most relevant architectures that are used for ensuring the security of the medical data in smart healthcare using blockchain technology are found from [49], [50]. The authors of [49] proposed a framework in the healthcare environment that can share data securely using the consortium blockchain through cryptographic primitives. The data preservation scheme for electronic health data using blockchain technology is discussed in [50], where cryptographic techniques are implemented to obtain data privacy. An architecture named BMPLS is developed

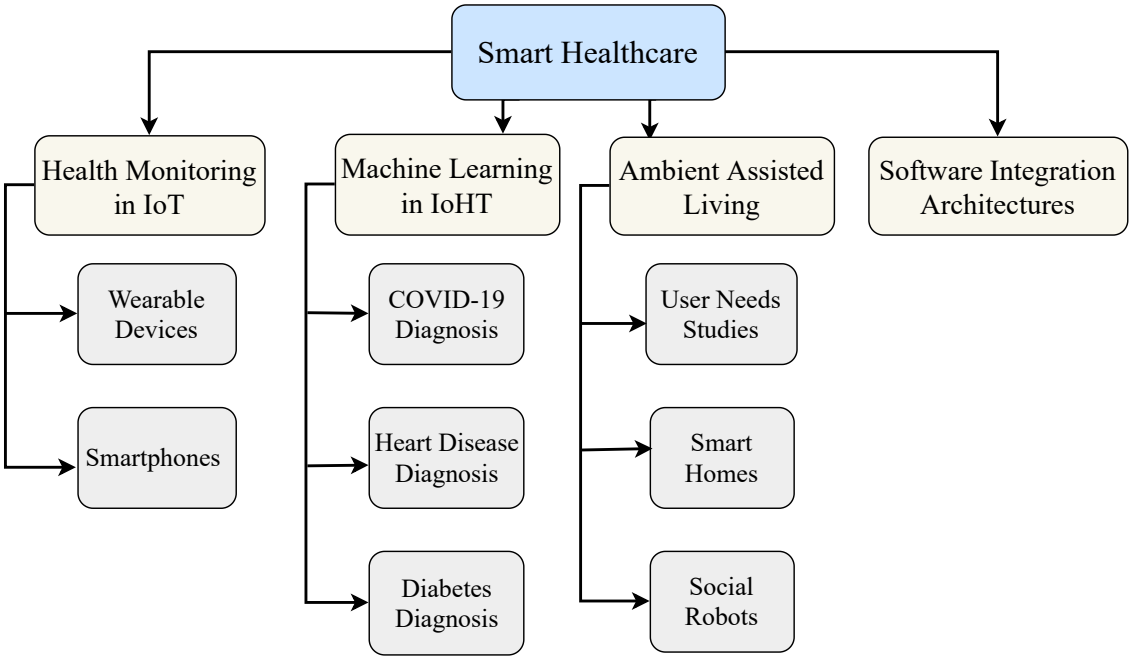


Figure 2.1: The overall workflow of the reviewed systems of smart healthcare.

in [51] for telehealth systems to obtain multi-level location privacy through order-preserving encryption techniques.

The goal of this chapter is to explore the SOTA smart healthcare systems that highlight the significant areas of research, including wearable and smartphone-based health monitoring, ML for predictive analytics, and assistive frameworks developed for assisted living environments, including social robots. The main contributions of the chapter can be summarized as follows:

- Provide a systemic review of SOTA research in smart medical devices, ML for disease prediction, AAL, and software architectures.
- Compare approaches to each problem, highlight their advantages and challenges, and present recommendations for improvement in future studies.
- Present a holistic overview of the smart healthcare field to provide a complete view of how technologies in different areas can be combined to accelerate smart healthcare.

The overall workflow of the reviewed systems is shown in Fig. 2.1. In Fig. 2.1, the comprehensive review of smart healthcare systems is divided into three major areas: health monitoring, disease diagnosis, and supportive devices in AAL. Additionally, the software integration

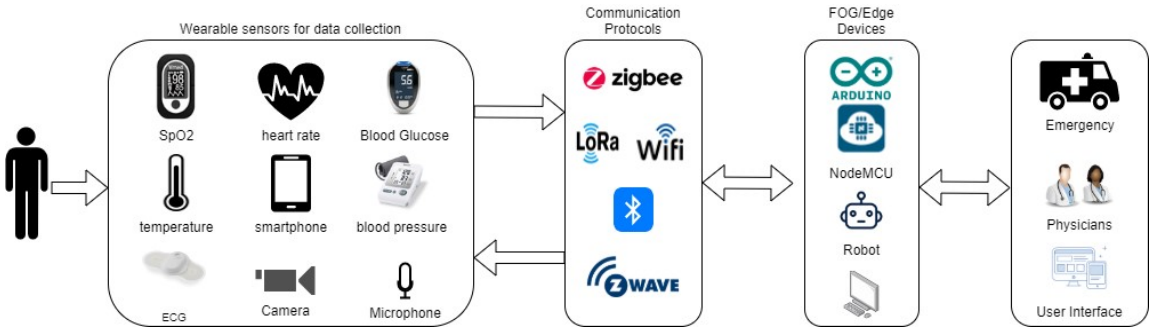


Figure 2.2: A general pipeline of a health monitoring system based on wearable devices.

architectures are described in this review. The health monitoring prototypes are divided according to wearable devices or smartphones, mentioned in section 2.2. Three major diseases, like COVID-19, HD, and diabetes detection frameworks based on ML algorithms, are demonstrated throughout section 2.3. Section 2.4 discusses the assistive prototypes in AAL, the supportive tools in smart homes, and social robots. The major directions in creating software architecture for smart healthcare are discussed in section 2.5. Section 2.6 presents an open discussion of the reviewed studies and guidelines for areas of future work. Finally, section 2.7 concludes the study and summarizes the main key points.

2.2 Healthcare Monitoring Devices in IoT

The IoT is a possible solution to relieve the stress on healthcare infrastructures and has become a prominent research issue in recent times [52], [53]. Health monitoring can be ensured through wearable sensors and smartphone applications. A broad pipeline of a health monitoring system based on wearable devices is shown in Fig. 2.2. Various sensors collect data from patients and transfer it to an edge/fog/cloud server for processing through Wi-Fi or LoRa gateway. The health status of the patients is monitored through physicians and their family members through several user interface tools like web or mobile applications. In some cases, emergency services are deployed to handle critical situations of the patients. The significant developments of health monitoring through IoT are described in this section, digging deeper into the details of implementation and technologies utilized.

2.2.1 Wearable Devices for Health Monitoring

Smart healthcare services using wearable sensors provide an appropriate and cheaper alternative to the costly hospital environment [54], [55]. These systems enable medical professionals to screen the significant symptoms of the patients, evaluate the general health of the users, and detect abnormalities remotely.

Recently, Islam et al. [56] developed a smart healthcare system that monitored patients' health using five sensors: two sensors (Heart Rate (HR) sensor and body temperature sensor (LM35)) for patient condition monitoring and three sensors (room temperature sensor (DHT11), CO sensor (MQ-9), and CO_2 sensor (MQ-135)) for detection of the living environment condition. Here, the processing device is ESP32, and Wi-Fi is used as communication media to transfer data from the patient's side to a web server. However, the developed prototype is not adequately manufactured for application purposes. Afterward, the authors of [57] proposed a wearable RHM framework that can monitor the HR, body temperature and detect falls. The proposed scheme used NodeMCU as a processing device and heartbeat sensor, body temperature sensor (LM35), and accelerometer (MPU 6050) as sensing elements. The ThingSpeak web service is used for data visualization to aid physicians in monitoring the patients from remote locations with a hand-held device as a prototype. In another study, the authors of [58] demonstrated a healthcare monitoring framework utilizing the concept of IoT and cloud computing. The prototype used an HR sensor, Electrocardiogram (ECG), Oxygen Saturation (SpO_2), and body temperature sensors for monitoring the corresponding HR, ECG, SpO_2, and body temperature of the patients, respectively. Two microcontrollers (Arduino and NodeMCU) were utilized to collect data from sensors and transfer data from edge devices to the cloud using Wi-Fi. Blynk cloud service was used to monitor the patients' health parameters from remote locations. However, the system is not able to handle the emergencies of the patients.

Chigozirim et al. [59] introduced a patient monitoring prototype that allows the doctors to monitor the patient's status through the developed tool using IoT. In this system, HR and body temperature are monitored through pulse and body temperature sensors. ATmega328P microcontroller and NodeMCU are used as edge devices, and the collected data from sensors are transmitted to the internet through a Wi-Fi connection. In addition, the doctors can check the patient's condition using Liquid Crystal Display (LCD). However, the developed system has not been entirely fabricated for real-time tests. In another research, Mohapatra et al. [60] demonstrated a smart healthcare management framework in IoT and cloud services to ensure the patient's condition monitoring from remote locations. The proposed scheme utilized HR and temperature sensors to perceive the data from patients, and an Arduino is used as an edge device connected to sensors and used to transmit data to the cloud through Wi-Fi. The doctors can easily connect with the cloud server through internet-connected devices to check the patient status and suggest proper

medications in an emergency. However, no security concerns are mentioned here for the transmitted data. Further, Swaroop et al. [61] developed a framework for basic symptoms monitoring in IoT environments. Multiple channels like Wi-Fi, messaging services, and mobile applications ensure a reliable link between the sensors for data transmission. The hardware components utilized here are sensors to measure Blood Pressure (BP), HR, body temperature, and Raspberry Pi 3. The prototype is developed in hand-held form. However, the latency for the Wi-Fi communication channel is comparatively high (125.95 seconds on average). Afterward, Al-khafajiy et al. [62] introduced a health monitoring framework based on wearable sensors for older people, enabling patients to take healthcare facilities from their home environment. The sensors used in this framework are pulse, temperature, blood oxygen, and blood glucose sensors. All the components are connected to Arduino UNO, and the collected data are sent to the patient's mobile application through a Bluetooth module. Finally, the smartphone application is used as a gateway to send the data to the cloud server. The doctors can easily monitor the sensors data and patient records through the developed monitoring platform. However, no data analytics tools are used for automated decisions.

In another work, Semwal et al. [63] presented a cost-effective and portable healthcare platform to ensure essential health services from remote locations. The sensory elements utilized in this system are ECG, pulse oximeter, body temperature, and BP sensors. ATmega328P microcontroller and Bluetooth module are used for data collection and transmission, respectively. LabVIEW tool is utilized for data visualization in the cloud server. The proposed prototype provides the offline accumulation of data from various sensors in low network connectivity to be updated to cloud accessibility. However, no security measures are taken into consideration in this system. Afterward, Kumar et al. [64] introduced a smart healthcare monitoring system where patients and doctors can interact through a camera. The health parameters from input sensors (HR and temperature sensor) are sent to the processing module (Raspberry Pi) and displayed on LCD. The processed data are transmitted to the web server using an internet connection that assists physicians in monitoring the patient status in real-time. However, the developed system is linked to a limited number of sensors which were not enough to monitor a patient's complete status. Furthermore, Wan et al. [65] presented a wearable health monitoring framework that assists doctors in monitoring the patients in real-time through an IoT network. In this system, BP, heartbeat, and body temperature sensors are used in the sensing node, and Arduino is used as an edge device. All the collected data are transmitted via Wi-Fi to the cloud server using the body area sensor network. The physicians can monitor the patients through their own devices like a laptop. However, the developed prototype did not mention any solutions for emergency cases.

Table 2.1 summarizes the developed health monitoring systems considering some properties like the used sensors and edge devices, the communication channel, the data visualization tools for ensuring the real-time monitoring, and comments of each developed system.

Table 2.1: Summary of the smart health monitoring systems based on wearable devices.

Refs.	Sensors	Edge Devices	Communication	Comments
[56]	Heartbeat and body temperature sensor	NodeMCU	Wi-Fi/ HTTP	The developed prototype is not adequately manufactured for application purposes.
[57]	Heartbeat sensor, body temperature sensor, and accelerometer	NodeMCU	Wi-Fi/HTTP, MQTT	No emergency services are mentioned in the proposed system.
[58]	Heart rate, ECG, SpO_2, and body temperature sensor	Arduino, NodeMCU	Wi-Fi	The system is not able to handle the emergencies of the patients.
[59]	Heartbeat and body temperature sensor	ATmega328P, NodeMCU	Wi-Fi	The developed system has not been fully fabricated for real-time tests.
[60]	Heartbeat and body temperature sensor	Arduino, NodeMCU	Wi-Fi/ HTTP	No security concerns are mentioned here for the transmitted data.
[61]	Heartbeat, blood pressure, and body temperature sensors	Raspberry Pi	Wi-Fi/ MQTT	The latency for the Wi-Fi communication channel is comparatively high.
[62]	Pulse, temperature, blood oxygen, and blood glucose sensors	Arduino	Bluetooth HC-06	No data analytics tools are used for automated decisions.
[63]	ECG, pulse oximeter, body temperature, and blood pressure sensors	ATmega328P	Bluetooth HC-05	No security measures are taken into consideration in this system.
[64]	Heartbeat and temperature sensor	Raspberry Pi	Wi-Fi/ HTTP	A limited number of sensors are not quite enough to monitor the patient's complete status.
[65]	Heartbeat, blood pressure, and body temperature sensors	Arduino	Wi-Fi/ HTTP	The developed prototype did not mention any solutions for emergency cases.

2.2.2 Smartphone Solutions for Health Monitoring

The growing penetration of mobile phones, integrated sensors, and advanced communication technologies makes it an appropriate infrastructure that allows continuous and virtual monitoring of patients' health. The built-in sensors in smartphones for health monitoring are a camera, accelerometer, gyroscope, proximity sensor, microphone, light sensor, and Global Positioning System (GPS) [66]. The major health parameters that can be monitored through smartphone sensors are HR and variability, BP, oxygen levels (SpO_2), and respiratory rate. They are used to identify skin, eye, ear diseases. As almost all people are now using a smartphone, it has become a great choice to research smartphone applications that ensure portability and reduce the additional cost of the developed systems [67], [68]. The systems that are designed for health monitoring using data collected from smartphone sensors are discussed here.

Zhang et al. [69] proposed a framework for blood glucose monitoring using Photoplethysmogram (PPG) data in the form of a video from a smartphone. The blood glucose level is collected from patients through smartphones, and the collected data is processed in the cloud. Lastly, the features are extracted using Gaussian Fitting and classified into normal, warning, and borderline based on blood glucose range using ML algorithms. The accuracy for blood glucose level estimation achieved from the developed system is 81.49%, which can be deemed lower than needed for a reliable glucose monitoring system. In another work, Nemcovaa et al. [70] developed a framework for monitoring the SpO_2, BP, and HR utilizing a mobile phone. The rear camera and microphone of a smartphone are used as sensing elements in this system. The camera data is converted to PPG and used for HR and SpO_2 estimation. At the same time, the BP is estimated from PPG and Phonocardiogram (PCG) recorded by the microphone. A smartphone application is developed to determine the feasible position of the data collection device for BP estimation. However, the synchronization between PPG and PCG signals is not handled, although there are different time bases. Afterward, the authors of [71] introduced a respiratory monitoring system using smartphone sensors based on imaging and the Fourier transform technique. The skin surface video data is captured in the presence of a flashlight using an embedded smartphone camera, and Plethysmography data is collected using developed hardware. The collected data is transferred to a PPG signal, and the processed data is analyzed using the concept of the discrete wavelet transform to estimate respiratory rate. The experimental results depicted that the system obtained an accuracy of 97.8% and an error of 2.2%. However, the system did not consider temperature and skin colour as well as the condition effect.

Recently, Tabei et al. [72] presented a framework for monitoring BP using smartphones' cameras. The data for the proposed scheme is collected from the user's finger index through a smartphone camera. Filtering and peak detection techniques are used to minimize the motion and noise from the collected PPG signals. The estimation of BP is done with the use of a linear regression

algorithm. It is revealed from the experimental results that the system obtained mean absolute error, standard deviation, and correlation parameters of 2.10, 1.96, and 0.90, respectively. However, the derivation of pulse transit time from two sides of the arterial is not mentioned. In another research, Dey et al. [73] proposed a cuff-less BP measurement system using a HR sensor embedded in a smartphone. The sliding window technique is used to convert the collected PPG signals to 15 s epochs. Approximately 233 features are derived from the raw signals from a PPG pulse in the domain of time and frequency. Finally, the BP estimation is conducted using the Lasso regression technique. A smartphone application is developed to monitor psychological signals in real-time. The developed system can calculate the 95% confidence interval of the BP of the patient. However, the results do not satisfy the precision in terms of standard value. Using the built-in accelerometer and camera of a smartphone, Wang et al. [74] introduced a BP monitoring device named Seismo to interpret the vibration generated by heartbeat and finger pulses. The blood ejection time is measured from the seismocardiography signal through the accelerometer, and the fingertip is used to calculate the arrival time from PPG data using a camera module. The embedded speaker synchronized the data from the accelerometer and camera. The developed system obtained a Pearson correlation coefficient between 0.20 and 0.77 for the volunteers. However, the prototype is not able to monitor BP continuously.

Further, a HR monitoring framework based on photoplethysmographic data from the smartphone is proposed in [75]. The data is collected using PPG's visible light reflected mode using a built-in smartphone camera from the user's index fingertip. The data from smartphone storage is transferred to the processing device using Bluetooth communication. Among the three channels of the PPG signal, the red channel is utilized for HR estimation in this system. The proposed scheme appraised an accuracy of 99.7%, and the found absolute error is within the range of 0.04–0.3 beats/min. However, the duration of the video is relatively low. In another study, Lomaliza and Park [76] developed a reliable and accurate system for HR monitoring using the camera images of a mobile phone of the fingertip. In this system, the signal is extracted using the concept of Region of Interest, and the noise from the raw data is eliminated through the adaptive threshold method. The developed scheme is adopted in any level of smartphone. The experimental finding depicted that the developed framework estimated the HR in real-time, with less than a 5% error rate. However, the proposed system ignores the effects of different camera modules of other smartphones. Furthermore, Qayyum et al. [77] demonstrated a vital sign monitoring system using video from a smartphone camera. In this framework, the collected data is pre-processed to reduce the noise from the raw signals using the colour distortion filtering technique. The inter-beat interval is used to detect HR variability, and the Breathing Rate (BR), HR, and SpO_2 are calculated from PPG signals. The developed system obtained a mean absolute percentage error of 2.965 from the experimental findings. However, no real-time prototype is shown in this study.

Table 2.2 summarizes the developed health monitoring systems considering some properties

Table 2.2: Summary of the smart health monitoring systems based on smartphone.

Refs.	Signs	Sensors	Smartphone	Users	Video	Rate	Comments
[69]	Glucose level	Camera	iPhone 6s Plus	80	60 s	30 Hz	The performance of the developed system is relatively low.
[71]	Respiratory rate	Camera	Samsung Galaxy S6	15	30 s	30.30 fps	The system did not consider temperature and skin colour as well as the condition.
[72]	BP	Camera	iPhone X	6	120 s	30 fps	The derivation of pulse transit time from two sides of the arterial is not mentioned.
[73]	BP	HR sensor	Samsung Galaxy S6	205 (90 M, 115 F)	15 min	125 Hz	The results do not satisfy the precision in terms of standard value.
[74]	BP	Rear camera, accelerometer	N/M	9	N/M	N/M	The prototype is not able to monitor blood pressure continuously.
[75]	HR	Camera	Samsung Galaxy	19	N/M	30.30 fps	The duration of the video is relatively low.
[76]	HR	Rear camera	Pantech Vega Racer 3, Samsung Galaxy S2, LG G3, HTC Desire HD	5 (3 M, 2 F)	10 s	10 fps	The proposed system ignores the effects of different camera modules of different smartphones.
[77]	HR, BR, SpO_2	Camera	iPhone 4	20 (10 M, 10 F)	60 s	30 fps	No real-time prototype is shown in this study.

M=Male, F=Female, N/M=Not Properly Mentioned.

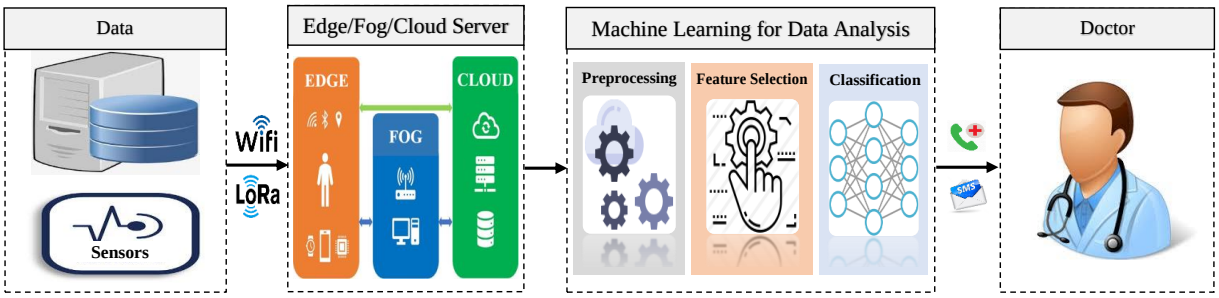


Figure 2.3: An overall system architecture of machine learning-based framework for disease diagnosis in IoHT environment.

like the monitored signs, the used sensors, the smartphone model, the number of subjects, the video length, sampling rate, and comments of each developed system.

2.3 Artificial Intelligence in IoHT

The IoHT comprises various interlinked devices that can share and handle data to enhance patient health. It has become a fast-growing area with numerous investments associated with the development and use of IoT [78]. Statistics from the McKinsey study depicts that IoHT will have a financial impact of 11.1 trillion dollar in a year by 2025 [79]. ML has become a significant tool in the arsenal of AI techniques used in healthcare. It enables IoT devices with outstanding capabilities for information inference, data analytics, and intelligence. ML has become a powerful and effective solution for various IoHT technology contexts, from big-data cloud computing to smart sensors [80], [81]. An overall system architecture for disease diagnosis using ML algorithms in the IoHT environment is shown in Fig. 2.3. The used data in these frameworks are from benchmark datasets or real-time sensor data sent to the fog/edge/cloud for processing. Afterward, the data are preprocessed, and necessary features are extracted to fit in the ML techniques. Finally, the decision is transferred to the concerned person to take proper action. The significant developments of ML-based IoHT solutions are demonstrated in this section. We have described some major disease solutions using ML in the IoHT platform that are becoming significant threats for human-being in recent times.

2.3.1 Novel Coronavirus (COVID-19)

The novel coronavirus 19 has become a public health crisis due to this virus's communicable nature in recent times. This is an ongoing pandemic, and all the sectors of the whole world are fighting to recover from this ailment. The statistic shows that approximately 98 million cases

have already found, and the death cases are about 2 million worldwide [82]. Numerous works were conducted to reduce the severity of this disease using modern technologies [83], [84]. In the current section, we describe the frameworks that used ML algorithms to diagnose COVID-19 in IoT environments.

Very recently, Le et al. [85] developed IoT based system to diagnose COVID-19 using the concept of CNN and Support Vector Machine (SVM). This framework retrieves the data from the patients utilizing IoT sensors and transfers it to cloud storage through 5G networks. In addition, the CXR dataset [86] is utilized to conduct an experiment. To reduce noises from the raw images, Gaussian filtering is utilized. The depth-wise separable CNN extracted the features from the pre-processed samples, and SVM categorized the extracted features to detect COVID-19. The proposed system appraised an accuracy of 98.54% for binary class and 99.06% for multiclass scenarios. However, no monitoring system from the doctor's end is developed here. Afterward, Ramallo-González et al. [87] introduced an IoT platform named CIoTVID for the detection of coronavirus. The scheme includes various levels of sensorization, which can handle and evaluate the data that assists in making a decision. The data collection layer collects various symptoms like voice signals, SpO_2, respiration rate from the patients. Mel Frequency Cepstral Coefficients (MFCC) transferred the raw voice signals to spectrogram as an image format. CNN architecture extracted the features and classified them in this system. The use case analysis found that the system appraised an accuracy of 66.67% in the testing phase. However, the outcome is relatively low for real-time use. In another research, Ahmed et al. [88] proposed a DL-based framework to diagnose COVID-19 in an IoT environment. A combined architecture like Faster-RCNN with ResNet-101 is utilized to diagnose the coronavirus cases from chest X-ray samples. The used data are retrieved from various open-access data sources where the COVID-19 cases are about 4000, and the negative cases are 7000. The data from the medical sensors are directly sent to cloud storage using Wi-Fi communication, where the proposed architecture is trained and finally diagnoses the positive cases. The radiologist can monitor the outcome through the internet. It is evident from the experiments that the developed system obtained an accuracy of 98%. However, no usability study is mentioned in this system.

Otoom et al. [89] introduced a scheme using ML techniques in an IoT environment to detect and monitor coronavirus-infected patients. The real-time symptom data from the patient's end are retrieved utilizing IoT devices and transferred to the cloud server for storage. Benchmark data called COVID-19 Open Research Dataset (CORD-19) [90] is being used for the analysis in addition to real-time data. The collected data are analyzed using ML classifiers to detect coronavirus infections. Among the eight classifiers, neural networks and k-Nearest Neighbors (k-NN) performed the best, and the accuracy was 92.89%. The system notified the medical experts of the suspected cases for further clinical treatment. However, the performance of the developed system is relatively low for practical uses. Further, El-Rashidy et al. [91] demonstrated a DL ar-

chitecture based on end-to-end nature to diagnose and monitor coronavirus-infected patients. In the proposed system, the patients are monitored through wearable sensors and smartphone app. A fog network is used to handle the data storage and transmission issues, and a CNN architecture with transfer learning diagnosed the COVID-19 patients from X-ray samples. The experimental data is collected from two publicly available datasets [86], [92], and wearable sensors. It is found from the experiments that the developed scheme obtained accuracy and specificity of 97.95% and 98.85%, respectively. The physicians monitored the patients in real-time and guided the individuals properly. However, energy consumption and storage issues are still a challenge for this system. In another research, Karmore et al. [93] developed humanoid software for the diagnosis of coronavirus in IoT networks that can identify whether an individual is infected with this ailment or not. The robotic system used IR sensors and a camera module for navigation, and the E-Health sensor kit and chest X-ray scans were utilized for diagnosis. The developed humanoid robot used NodeMCU, Raspberry Pi, temperature sensor, ECG sensor as hardware components. Three pre-trained architectures like InceptionV3, ResNet50, and Inception ResNetV2 diagnosed positive cases from X-ray samples. The average accuracy found from the proposed system is 97.95%. However, the security issues during the data transfer are not depicted here.

Table 2.3: Summary of COVID-19 detection frameworks in IoHT environment.

Refs.	Data	Techniques	Accuracy (%)	Comments
[85]	CXR dataset and real-time sensor data	CNN, SVM	99.06	No monitoring system from the doctor's end is developed.
[87]	Real-time sensor data	CNN	66.67	The outcome is relatively low for real-time use.
[88]	Various public datasets	Faster-RCNN, ResNet-101	98	No usability study is mentioned.
[89]	CORD-19 dataset, and real-time sensor data	Eight ML techniques	92.89	The performance of the developed system is relatively low for practical uses.
[91]	Two open-access datasets and real-time sensor data	ResNet-50	97.95	Energy consumption and storage issues are still a challenge for this system.
[93]	Real-time sensor data	InceptionV3, ResNet50, Inception ResNetV2	97.95	The security issues during the data transfer are not depicted.
[94]	Real-time sensor data	Random Forest, Naïve Bayes, SVM	73	The system achieved comparatively low performance.
[95]	Real-time sensor data	K-means algorithm, SVM, Bayesian classifier	86	The accuracy rate of the system is not up to the mark.

Furthermore, Cacovean et al. [94] introduced an IoT-based framework for COVID-19 detection where ML techniques are utilized for diagnosis. The data from the participants are collected using wearable devices like GPS, temperature, and HR sensors. The retrieved data are sent to the oracle cloud server for processing through the Bluetooth module. Random Forest obtained the best outcome from the experiments among the three classifiers, and the accuracy value is 73%. The prediction outcomes are directly sent to doctors and patients' guardians to take further steps for proper treatment. However, the system achieved comparatively low performance. Afterward, Kumar et al. [95] presented a system to monitor the COVID-19 patients using sensor and IoT technology. The participants retrieve the real-time data using IoT sensors and feed it into the Bayesian network for preprocessing. The IoT devices are configured and accessible using wireless sensors to send the data to the patient's repository. The data are trained with SVM and predict the coronavirus cases from the test samples. The scheme appraised 87.23% and 86% accuracy for recovery and prediction, respectively, using SVM. In addition, the K-means algorithm estimated the spread as well as recovery rate. However, the accuracy rate of the system is not up to the mark.

Table 2.3 briefly discusses COVID-19 detection systems highlighting features such as the datasets and techniques used for detection, the accuracy as a performance metric, and comments of each reviewed system in the IoHT environment.

2.3.2 Heart Disease

Heart disease has become a very crucial and acute ailment for every aged people, especially for adults. An estimation shows that heart disease is responsible for approximately 30% (18 million individual) deaths among all death cases per year [96], [97]. Hence, the researchers focus on the development decision support system in the smart healthcare environment to reduce the severity of heart disease. The significant developments of heart disease diagnosis using ML in the IoT environment are demonstrated here.

Recently, the author of [98] developed a patient monitoring system for heart patients in an IoT environment where the data from the patients are analyzed using a modified Deep Learning Modified Neural Network (DLMNN). The body-worn sensors collected data from the patients and securely sent them to the cloud for further processing. In addition, the proposed system used the Hungarian HD dataset [99]; benchmark data for HD classification to detect the presence of abnormality. An alert message is delivered to the doctors when any abnormality is detected. However, the developed scheme obtained comparatively low performance in the case of a small data size. In another work, Ali et al. [100] proposed a smart healthcare monitoring framework for HD-infected patients using the concept of ensemble learning and feature fusion. The extracted

features from sensor data and patient history are merged through the feature fusion technique in this system. The information gain method eliminated the unnecessary and redundant features, selected the most appropriate features responsible for the disease. A semantic web rule language is introduced that recommends the activities of the infected patients automatically. Lastly, the LogitBoost technique, an ensemble learning classifier, is used to predict HD and obtained an accuracy of 98.5% from the experiments. However, the developed system used traditional techniques for feature selection, reduction, and classification. Afterward, Deperlioglu et al. [101] introduced a framework for HD diagnosis using an autoencoder network in an IoHT environment. The developed system comprises a cloud environment where beacons are used for data sharing and a central system to synchronize the cloud and devices' communication and ML architecture. This system uses two heart sounds datasets named PASCAL B-Training [102] and Physiobank-PhysioNet A-Training [103]. The developed system obtained accuracy, sensitivity, and specificity of 100% for the PASCAL dataset. In addition, an accuracy of 96.03%, 91.91%, and 90.11% are achieved for healthy heart sounds, extrasystole, and murmur, respectively, from 479 real-time participants. However, no voice command facility is available in this study to ensure less physical interaction.

In another research, the authors of [104] presented a ML-based HD diagnosis system in the IoMT cloud environment using Modified Salp Swarm Optimization (MSSO) and an Adaptive Neuro-Fuzzy Inference System (ANFIS). The data from the IoMT sensors and UCI [99] and Framingham database [105] are used to diagnose the presence of HD. The MSSO technique optimized the dataset's attributes to find the best features, and the ANFIS trained the most appropriate features and diagnosed the disease. The experimental results found that the system achieved accuracy, AUC, and precision of 99.45%, 99%, and 96.54%, respectively, using the datasets, yet no results are presented for real-time data. The authors of [106] proposed an IoT-based system using Modified Deep Convolutional Neural Network (MDCNN) to predict HD. The data (BP and ECG) used for this study are collected from smartwatches and heart monitor devices attached to the patient's body. In addition, some open-access databases like UCI [99], Public Health, and Framingham [105] are also used to train the network. In this framework, Long-range (LoRa) communication protocols, LoRa cloud, and servers are used to ensure the real-time monitoring of the patients. The developed framework categorized the sensors' data into two classes (normal and abnormal) and obtained an accuracy of 98.2% from the experiments. However, no wearable prototype is mentioned here. Further, Tuli et al. [107] developed a smart healthcare framework named HealthFog to diagnose HD using ensemble learning in IoT and Fog computing environments. To evaluate the performance of the developed system with respect to energy consumption, accuracy, latency, and execution time, FogBus (Fog-based cloud environment) is utilized. The FogBus is comprised of the worker node, the cloud data center, and the broker node. The Bagging classifier categorized the data collected from sensors and benchmark datasets [99] of HD. The developed prototype achieved an accuracy of 89% from the experiments for the test cases. As the

Table 2.4: Summary of heart disease detection systems in IoHT environment.

Refs.	Data	Techniques	Accuracy (%)	Comments
[98]	Hungarian HD dataset and real-time sensor data	DLMNN	96.8	The developed scheme obtained comparatively low performance in the case of a small data size.
[100]	Cleveland dataset and real-time sensor data	Feature fusion and ensemble learning (LogitBoost)	98.5	The developed system used traditional techniques for feature selection, reduction, and classification.
[101]	PASCAL and PhysioNet dataset, and real-time sensor data	Autoencoder neural network	96.03	No voice command facility is available in this study to ensure less physical interaction.
[104]	Hungarian and Framingham dataset, and real-time sensor data	MSSO-ANFIS	99.45	No real-time study is shown in this system.
[106]	Cleveland database and real-time sensor data	MDCNN	98.02	No wearable prototype is mentioned.
[107]	Cleveland dataset, and real-time sensor data	Ensemble learning (Bagging)	89	The computational time is relatively high.
[108]	Real-time sensor data	wkPCA, BNN	98.03	No notification system to alert physicians is developed.
[109]	Statlog dataset and real-time sensor data	J48, logistic regression, multilayer perceptron, SVM	91.48	No real-time study is illustrated in this system.

ML architecture is trained in each worker node of every fog node, the time consumption becomes comparatively high.

In another study, Nguyen et al. [108] introduced a scheme for diagnosing HD using ML in an IoT environment. ECG devices collect the data from patients and send them to cloud storage through Wi-Fi. Wavelet-based Kernel Principal Component Analysis (wkPCA) technique pre-processed the raw data and extracted the most relevant classification features. The extracted features are fed into a Backpropagation Neural Network (BNN) that diagnoses the HD based on the input data. The developed system achieved an accuracy of 98.03%. However, no notification system to alert physicians has been developed yet. Furthermore, the authors of [109] demonstrated an IoT-based framework to diagnose HD using cloud storage and ML algorithms. The data is collected from the human body using medical IoT sensors, and benchmark datasets [110] from UCI are also used for the experiments. All the data are stored in a cloud database, and ML techniques are applied to the cloud database to predict the presence of HD. Among four classifiers, J48 performed the best and obtained accuracy, precision, recall, and F1-Score of 91.48%, 91.50%, 91.50%, and 91.50%, respectively. However, no real-time study is illustrated in this system.

Table 2.4 summarizes HD detection systems considering some features such as the used datasets, the used algorithms for detection, the accuracy as an evaluation metric, and comments of each developed system in the IoHT environment.

2.3.3 Diabetes

Diabetes is another life-threatening disease for humankind that results in many deaths per year. An estimation shows that almost 463 million individuals had diabetes in 2019, and the numbers are expected to grow to 578 million and 700 million by 2030 and 2045, respectively [111]. As this ailment is rising rapidly, early diagnosis of diabetes is necessary for the sake of people. Various studies are conducted to diagnose diabetes early, utilizing AI, IoT, and Big data [112], [113]. The works that have been developed recently for diabetes detection are illustrated in this section.

In recent times, Rghioui et al. [114] developed a framework to monitor and predict diabetic patients using ML techniques in IoT networks. In this system, the glucometer is connected to NodeMCU to record the data from patients seamlessly. A vast amount of collected data are sent to the cloud database using IoT platform and processed using ML algorithms, and the decisions are sent to doctors for further treatment. The dataset comprises five features with 12612 records. Among the four algorithms, Random Forest achieved an accuracy of 96.05% from the experiments. In another research, Allugunti et al. [115] proposed a diabetes prediction framework using the concept of IoT and a decision tree to monitor the infected patients in real-time. The

data is collected from IoT sensors and contains eight attributes and 15,000 realities. The best features from the dataset are selected using the concept of entropy measurement. The experimental outcomes show that the developed framework obtained accuracy and an error rate of 96.43% and 5.37%, respectively, in prediction. However, the detailed data collection procedure is not mentioned here. Afterward, Efat et al. [116] demonstrated a health monitoring system focusing on diabetic patients that can monitor the level of sugar, sleep time, food intake, and pulse rate. The patient's data is continuously sent to a neural network using wearable sensors through Bluetooth, and the developed architecture categorizes the data based on the severity of diabetes cases. An alert message/call is sent to the patient's guardians and caregivers in an emergency. The developed scheme appraised an accuracy of 84.29% from 25 diabetes patients' data. A web portal monitors the patients' health status continuously. However, the performance is relatively low for practical use.

In another work, a diabetes monitoring and prediction framework is proposed in [117] utilizing IoT and ML techniques. The system used a blood glucose meter, Arduino, and GSM modem as hardware components. The experimental data are retrieved using the glucose meter using an edge device like Arduino and processed in the microcontroller. The decision of the processed data is automatically sent to mediators through a GSM modem. The author found the Random Tree classifier to provide the highest accuracy and lowest training time among the four used algorithms, and the value is 97.87% and 0.03 seconds, respectively. However, the glucose sensor and Arduino could not be operated at the same time. Further, Rghioui et al. [118] introduced an intelligent framework for diabetes-infected patients monitoring using ML architectures in IoT networks to monitor physical activity, glucose level, and body temperature. A glucometer, temperature sensor, and motion sensor are used for data collection purposes at the patient's end. The collected data are transferred to the database station using a smartphone through 5G networks. The patient's records are classified using six classification algorithms, and the Sequential Minimal Optimization (SMO) obtained the best accuracy of 99.66%. Whenever any abnormality is found, a notification text is sent to the doctors to take proper steps for treatment. However, the latency is comparatively high in this system. Afterward, Godi et al. [119] developed a healthcare monitoring framework to diagnose and monitor disease using ML modalities through IoT networks. Various wearable devices are utilized to retrieve data from patients from different scenarios like homes and hospitals. In addition, a diabetes dataset from the Kaggle repository [120] is used for experiments. ML techniques classified the data based on the presence of abnormality. Among four classifiers, SVM achieved an accuracy of 80.51%, precision of 76%, recall of 65%, and F1-Score of 70% for the positive diabetes class. The predicted results are shared with the physicians, mediators, and patients' caregivers. However, the performance is relatively low for real-time implementation.

To predict diabetes mellitus, the authors of [121] introduced a novel framework using ML

in the IoT environment. The glucose sensors collected the blood sugar data from the patients, normalized it into the proper format, and transferred it to the storage device using Hypertext Transfer Protocol (HTTP) and Message Queuing Telemetry Transport (MQTT) protocol. The system also used a benchmark dataset for the experiment. Two ML classifiers named SVM and k-NN are used for diabetes prediction, where SVM achieved accuracy and F1-Score of 90% and 89%, respectively. However, real-time cases are not found there. Furthermore, Kaur et al. [122] introduced a framework named CI-DPF to predict diabetes in a cloud-based IoT environment. The blood glucose level from the patients is collected using smart sensors and sent to the cloud environment for storage and further processing through IoT devices. The proposed system also used benchmark data named Pima Indians Diabetes dataset [120] for the experiment. Ensemble learning is used to diagnose diabetes from the patient's records. The decision tree and neural network ensemble obtained accuracy, sensitivity, and specificity of 94.5%, 79.5%, and 83.12%, respectively. However, real-life clinical tests have not been conducted here.

Table 2.5 briefly discusses diabetes detection frameworks highlighting properties such as the used datasets, the used algorithms for detection, the accuracy as a performance metric, and comments of each reviewed system in the IoHT environment.

Table 2.5: Summary of diabetes detection frameworks in IoHT environment.

Refs.	Data	Techniques	Accuracy (%)	Comments
[114]	Real-time sensor data	Naive Bayes, Random Forest, OneR, SMO	96.05	No user prototype is shown in this system.
[115]	Real-time sensor data	Decision Tree	96.43	The detailed data collection procedure is not mentioned.
[116]	Real-time sensor data	Neural network	84.29	The performance is relatively low for practical use.
[117]	Real-time sensor data	Six ML algorithms	97.87	The glucose sensor and Arduino could not be operated at the same time.
[118]	Real-time sensor data	Six classification algorithms	99.66	The latency is comparatively high in this system.
[119]	Benchmark data from Kaggle and real-time sensor data	Decision tree, Regression, k-NN, SVM	80.51	The performance is relatively low for real-time implementation.
[121]	Real-time sensor data	k-NN, SVM	90	Real-time cases are not found in this study.
[122]	Pima Indians Diabetes dataset and real-time sensor data	Ensemble of decision tree and neural network	94.5	Real-life clinical tests have not been conducted.

2.4 Ambient Assisted Living

AAL involves combining sensors and actuators in an IoT environment to communicate and provide enhanced lifestyle and human-independent care for older adults. With 87% of older adults preferring living in their own homes over senior homes [123], there are two target environments for AAL; senior care homes and older adults' private homes. We first go in-depth into the studies carried out to define the needs of the seniors, then discuss the implementations both with and without a robotic social agent.

2.4.1 User Needs Studies

Several studies are directed towards studying the needs of older adults. Such requirements can be categorized into physical assistance, emotional support, reminders, or social support. Bedaf et al. [124] performed a user study with different stakeholders, including 11 formal caregivers, seven informal caregivers, and ten older adults as part of the ACCOMPANY project. After the users interacted with a robotic system (Care-O-Bot 3) in a fetch-and-carry scenario and a scenario where the robot reminded them to drink water, the authored received several suggestions on the functionalities they needed. As such, the authors concluded that a social robot for elderly care needs to have advanced speech interaction capabilities, fetch and carry various objects, detect dangerous situations, alert the caregivers, and be adaptable to individual user needs. Likewise, the HomeMate project [125] defined five main scenarios that would benefit users the most: fetch-and-carry, infotainment (music and movies), gaming services, video chatting, and reminders for various events.

After implementing a physical prototype and its testing with older adults, the authors emphasized the importance of natural interaction, specifically through speech. Other studies, such as [126], highlight the importance of combining both AAL sensor technologies with a robot for task achievement as well as social companionship. Similar conclusions were made in [127], where the authors highlighted how 50% of older adults requested efficient speech interaction and added that the inclusion of a robotic platform provides multiple benefits. Syed et al. [128] highlight that movement should be monitored for older adults, citing falls as one of the major causes of death in the senior community. Other studies that highlighted the needs of older adults include [129], [130]. The requirements can be summarized as:

- Provide a natural means of interaction that require minimum to no learning by the older adult.
- Remind the users of medications, appointments, and events.

- Provide infotainment services such as music, movies, and cognitive games.
- Real-time monitoring of health vitals and detection of emergencies.
- Include a robotic platform for task achievement as well as social companionship.

In the remaining section, the implementation of various platforms for AAL is discussed. First, we showcase research done with IoMT environments without robotic agents, followed by studies involving social robots as a central system component.

2.4.2 Smart Homes

Komai et al. [131] present a system to monitor the activity of multiple seniors simultaneously based on Bluetooth Low Energy (BLE) with a beacon in the user's name card and the Received Signal Strength Indicator (RSSI). Likewise, [132] uses BLE and RSSI to create a low-cost indoor-localization method to track and estimate users' location. They propose the method as a low-cost system to detect older adults' activity and early signs of frailty using a Random Forest classifier. Although both systems are proposed based on low-cost solutions for indoor localization, they can only achieve room accuracy. Marques et al. [123] present an indoor environmental monitoring system that measures room temperature, relative humidity, CO, CO_2, light detection, and transmitting the messages through XBee through Zigbee networking protocol. The system proves to be a modular and cheap solution for indoor air quality monitoring. The work is extended in [133], where a robotic platform is integrated equipped with a gas sensor to detect levels of Liquefied Petroleum Gas (LPG), isobutane, and propane, which can lead to explosions when they reach specific levels. Based on the famous Turtlebot platform, the robot can notify the user through the Facebook social platform. The system provides a safe way for monitoring gas levels but allows for little to no control by the user. Diraco et al. [134] propose a sensory system based on radars to monitor heart and respiration rates of older adults without contact (from a distance), achieving 95% and 91% accuracy, respectively. The authors also utilized the radar for fall detection, resulting in a sensitivity of 97% and specificity of 90%. Nevertheless, ultra-wideband radio signals can measure a few vital signs and need a more versatile system to integrate more sensors. The authors of [128] propose a framework for monitoring physical activities and utilizing ML algorithms for more accurate and faster predictions and decisions. They build on the mHealth framework, initially proposed in [135] to collect data from multiple sensors and combine them to predict 12 different physical activities using a multinomial Naïve Bayes classifier, achieving an accuracy of 97.1% on the mHealth dataset. Although the framework sounds promising, it is yet to be tested in a real-world environment.

A mobile application called InfoSage is offered by Quintana et al. [136] to connect older adults to their formal and informal caregivers, centred around the older adult as the keystone user. The solution focuses on dementia patients and offers a tracker and reminders for appointments and medications. The system provides capabilities for exchanging messages between family members (informal caregivers), doctors, and older adults and share information. The authors also perform several user studies on the acceptance and usability of their platform. Stavrotheodoros et al. [129] propose the IN LIFE platform, a cloud-based solution that combines various sensors focusing on personalization and easy installation for cognitively impaired older adults. The system is capable of monitoring user activities through unobtrusive sensors and multilayered architecture. The system is comprised of 3 layers, a perception layer with the various sensors for data collection, a gateway layer that combines the data and transmits them to the final layer (cloud layer) using MQTT protocol. The cloud layer stores and analyzes the data. The work is extended in [137] by establishing the system and using motion sensors and door sensors to identify user habits and a panic button for older adults when there is an emergency. The authors of [138] propose an AAL system that utilizes fog computing. A system is put in place that incorporates radar sensors to detect daily activities and implements an algorithm to detect whether the patient is suffering from a neurological disease attack, if they are idle, or if no patient is detected. Utilizing the fog layer in the system leads to minimal response delay and energy consumption coupled with more bandwidth efficiency and overall performance.

Several of the proposed solutions provide promising solutions for AAL, all aiming at better and healthier living. Nevertheless, the area still requires much work, specifically applying the proposed systems in real-world environments for elaborate testing and user feedback.

2.4.3 Social Robots

Previous studies did not harvest the power of robotics to provide more functionalities for older adults. The following discussion tackles studies that involve social robots, a summary of the most prominent works is provided in Table 2.6. Portugal et al. offer SocialRobot in [142], a modular robotic platform with independent layers. SocialRobot can adapt to user's preferences and includes Human-Robot Interaction (HRI), emotion and facial recognition, and speech interaction. The robot is based on Robotic Operating System (ROS) and uses a sequential database (MySQL) to store data for more personalized interactions, which the authors called SoCoNet. The work was extended in [144] and [127]. The authors studied the effect of including environmental context on the decision process and tested SocialRobot in an elderly care home for a week. Besides navigation and storing data on its users, SocialRobot can also recognize faces and users' emotions. After testing SocialRobot in an elderly care home in the Netherlands, the authors concluded that

Table 2.6: Summary of the social robots implementations for care of older adults.

Refs.	Sensors	Functionalities	HRI Methods	Comments
[139]	Fitbit Versa2 Smart Watch	Autonomous navigation	Speech interaction	Needs validation for acceptance and usability through testing it with older adults in a real-world environment.
[133]	Gas sensor: LPG, iso-butane, propane	Autonomous navigation	Notifications on Facebook (no user input)	The system lacks HRI methods and provides no control for users over the system. The system needs a user interface with more functionalities needed for older adults.
[125]	N/M	Autonomous navigation, fetch-and-carry, object handover, telepresence (video chat)	Speech interaction, gesture control	The system could only have simple speech interaction and is missing a link to an IoT environment with smart sensors.
[140]	Custom bracelet: HR, body temperature, angular acceleration, and heading	Tele-operation, obstacle avoidance, facial recognition, telepresence (video chat)	Speech interaction and vision	Needs validation for acceptance and usability through testing it with older adults in a real-world environment. The robot needs autonomous navigation to eliminate privacy concerns.
[141]	Cameras	Autonomous navigation, depth vision, object detection	Alarms	The system needs efficient means for user interaction and sensors/wearables for tracking body vitals.
[127], [142]	N/M	Autonomous navigation, personalized interactions, facial recognition, emotion detection	Touch Screen	The robot has no speech interaction capabilities and is not linked to an IoT environment with smart sensors.
[143]	Smartwatch, IMU, Passive Infrared Sensor, microphones	Autonomous navigation, sound detection	Touch screen UI (mobile phone or tablet)	It needs validation for acceptance and usability through testing it with older adults.

N/M=Not Properly Mentioned.

considering the current context improves the accuracy of predictions. They also noted that although the robot was generally deemed acceptable by the older adults, it should incorporate more effective speech interaction and anthropomorphism (arms) for better functionality, which it is not equipped with. SocialRobot covers several requirements of the older adult community. However, it has been shown that integrating social robots in an IoT environment with sensors that track older adults provides much greater promise. Moreover, the caregivers and older adults requested functionalities such as playing music, movies, games, and memory training activities.

The HomeMate project [125] conducted studies to define older adults' requirements for creating a robotic platform. The HomeMate robot can play movies and music, link older adults with their family members and friends through video chat. Besides that, it was also able to communicate through touch and voice interaction and schedule reminders for different events, thus tackling the missing functionalities from the SocialRobot project. However, the users still requested more natural and intuitive speech interactions, possibly through a more intelligent Natural Language Understanding (NLU) module. Moreover, the robot was also missing a link to smart devices for monitoring environmental conditions and users' health. While both HomeMate and SocialRobot tackled senior home environments, Gross et al. proposed a robot companion for private homes called Sympartner [145]. The companion can provide reminders, health updates, daily routines, greet visitors at the apartment door, detect and identify objects and faces. Sympartner had autonomous navigation but did not have a manipulator to carry objects. HRI was done through a graphical user interface as well as through simple speech commands. The robot was deployed with 20 participants in their private homes for five days each in Germany. Although the overall feedback from the participants was positive, they requested speech understanding capabilities. Moreover, the robot was reported to have failed several times and required remote teleoperation by the researchers, presenting privacy concerns to its users.

The previous studies involved social robot implementations. However, none of them were integrated into smart home environments with other environmental and health monitoring sensors. The idea of including robots in the AAL environment is recommended by [143], [146]. This can be thought of as an Internet of Robotic Things (IoRT) framework. Nasr et al. introduce a solution for AAL environments with robotics in [139]. The platform connects heterogeneous agents such as mobile robots, virtual assistants, and mobile phones with smart sensors and wearables. The system was designed with an emphasis on HRI and intuitive speech interaction with its users. The authors developed the platform with a MySQL database for storing reminders, a mobile phone to utilize Google Assistant's speech-to-text functionality, a Fitbit Versa2 smartwatch for HR monitoring, and a simulated robot in ROS and Gazebo. Two different protocols for data sharing are provided in the system, MQTT and Representational State Transfer (REST) API, to allow all kinds of smart devices and agents. The system is shown to respond to natural ways of giving commands and without the training of users. It provides its users with control over the robotic

system by sending navigation commands and teleoperation through speech. Also, the authors emphasize the idea of offering the same functionality across all devices through utilizing a common NLU and Dialogue Management agent. While the system shows great promise in integrating heterogeneous systems in a modular way, it still needs to be tested in a real-world environment. It requires a functional robot capable of at least autonomous navigation and moving objects. Do et al. propose a similar approach to the problem through the robot-integrated smart home platform (RiSH) [143]. They utilize a Pioneer 3DX platform based on ROS and are equipped with a microphone for acoustic detection. The robot is linked to a network with body sensors, including an Inertial Measurement Unit (IMU), motion sensors, and a smartwatch to obtain ECG, SpO_2, and respiration rate readings and a home sensor network which includes a passive infrared for binary motion detection and microphones for acoustic data. The authors present a system that is extensible and capable of leveraging smart home sensing capabilities. Furthermore, they conduct experiments with 12 older adults to detect human trajectories down to a 0.2-meter accuracy and recognize 37 different human activities and falls with accuracy up to 88% and 80%, respectively.

Another study that follows the integration of robots for AAL is provided by Loza-Matovelle et al. [140]. The system is composed of a heterogeneous network of sensors both on a robotic platform and a bracelet that the user wears. The bracelet can measure HR and body temperature, and angular acceleration and heading (gyroscope). The accelerometer and gyroscope are used to detect falls of the older adult and send warnings to the family members. The robot is capable of localization and obstacle avoidance but needs to be teleoperated. In addition, it is capable of facial recognition to maintain contact with the user during interaction and allows for video conferencing with formal and informal caregivers (telepresence). The user is provided with the ability to interact with the robot and a hologram (called the interactive pyramid) through speech and a chatbot. The interactive pyramid's speech interface updates the user on the weather, time, reminders on medications and visits, as well as health recommendations. In essence, the system is like the works in [139], [143]. However, it uses teleoperating, which creates a privacy risk and lower autonomy of the system, and the system is yet to be tested in a more realistic environment. Gomez-Donoso et al. [141] integrate a robotic system into an AAL environment equipped with cameras to detect dangerous situations. The authors found that the existing system does not detect dangers such as objects on the floor, knives (which change location), and dangers in occluded areas. Therefore, they added a Pepper robot based on ROS with an RGB-D camera to detect such dangers. The robot was capable of ground plane detection and clustering of pixels to find objects on the floor. Also, as the robot moves, it uses an R-CNN to detect smaller sources of danger, such as electrical outlets and knives. Pepper was also capable of autonomous navigation and detection of people who had fallen. The robot was linked to Wi-Fi and would send out an alarm whenever the older adult was in danger. The authors tested the robot in IoT environments covering residential areas, clinical areas (nursing homes), and offices. The robot is shown to

increase the ability of the AAL to detect potentially harmful circumstances in different scenarios. Such a system is powerful but would benefit when integrated with other sensors that monitor the user's health vitals.

2.5 Software Integration Architectures

Global business processes in IoMT demand information to be shared quickly and efficiently across many different software, tasks, and applications. Having discussed the general areas of research in IoMT, we find much common ground in approaching and the architectures utilized. To unify the approach and paving the road from research to implementing the systems in real-time environments, various architectures are provided in literature that attempts an inclusive solution for an IoMT environment that is modular and easily extensible in its functionalities. However, each architecture is tailored to a specific use case. This section gives an overview of the most promising architectures and frameworks proposed in this area and an in-depth analysis of their advantages and shortcomings.

Petrovic et al. [147] popped the following question: Why recreate devices, sensors, and systems. At the same time, we can use off-the-shelf solutions and augment their capabilities to provide a far more functional system. Therefore, the authors use Commercial off-the-Shelf (COTS) smart devices (COTS component) and combine them through an Interactivity component that provides an interface to retrieve data from the COTS devices and performs pre-processing stages. This layer can also detect some patterns and issue warnings in danger to a patient's health. The third and last component of the system is the cloud component which synchronizes the COTS devices and takes care of big data analytics. While this solution offers faster solutions and lower costs, it is mentioned as an idea, and no validation is provided. Both [148] and [128] design a very similar architecture to approach the integration of IoMT. In [148], the authors developed an integrated medical platform for RHM. The platform is multi-layered to enable easy integration and expandability. The first layer, the perception layer, combines all the sensors that collect data about the patient and their surrounding environment that could affect the patient's health. The next layer is the network and gateway layer, which transfers the data from the perception layer and processes it. The last layer stores and integrates the data received from various sensors and makes informed decisions. The authors showcased the architecture using sensors to measure multiple vitals, including HR and body temperature and environmental data like light intensity, humidity, and temperature. They also utilized the system for fall detection. The system is promising but does not present a means of interacting with users. It also presents a great resemblance to the architecture used in [128], which augmented it with the mHealth dataset and a user interface for caregivers (in the application layer), yet not its primary users; patients.

Most architectures in literature highly depend on cloud infrastructure for storage and analytics. Other architectures include Hadoop Map Reduce techniques to process vast amounts of data in parallel. Although utilizing the cloud provides virtually unlimited space and computational power, a side effect is the time delay in transferring data. This led researchers to use faster communication, analysis, and temporary storage within the user's local network to promote a more effective response to dangerous situations with minimum to zero delays. The devices that provide these functionalities are referred to as fog devices. The name comes from the real-life fog to resemble its proximity to the ground (environment) instead of the cloud. One of the first research works in that area was presented by Vora et al. [138]. The system implemented is used for the detection of ADL based on a radar sensor. The authors highlight further advantages in using fog devices for IoHT. It increases bandwidth, lowers data transfer latency, and offers a complementary decrease in energy consumption and data overload compared to cloud computing. Loza-Matovelle et al. [140] propose an architecture that focuses on integrating robotics and HRI methods. The system combines a network of heterogeneous sensors and actuators in a decentralized manner that decouples the functionalities of various agents. It is made of two servers that communicate together. A local (ROS-based) server oversees task achievement while a server for web services integrates with interactions with the users. All communications in this system use the MQTT protocol. The system is implemented with different sensors, a robotic platform, and a hologram for user interaction, thus showcasing its functionality. While the system shows great promise of integrating various kinds of agents and modularity, it will need a ROS network integration for each new agent, which is not useful for non-robotic agents. Therefore, further breakdown of the local server can prove more practical, as shown in other works discussed here. Nasr [139] provides a framework that focuses on modularity and scalability and integrates heterogeneous agents, sensors, robots, and HRI devices for AAL scenarios. In [149], the framework is deployed in two scenarios (RHM and AAL), using the same building blocks and combining both a cloud and a fog layer. The framework is divided into three independent layers, namely:

- Device Layer: Includes smart sensors in the environment or on the user's body, robotic agents, and agents used for user interaction such as mobile phones, virtual assistants, or gesture control devices. This layer covers all kinds of objects that can interact with the real world and users.
- IoT Fog Layer: Decentralized computing and storage device(s) that receive the data from the device layer, preprocess it, and sends required data to the cloud. Fog devices reside in the local network of their users.
- Cloud Layer: Central hub for large data storage and analytics. It allows easily expanding the capacity of the system for processing and prediction models.

The authors argue that the proposed division of layers and clearly defined communication protocols and methods allow the framework to be flexible enough to fit into different use cases and personalize each user according to their needs. Two prototypes are created using the framework with different agents and capabilities, but the prototypes still need user testing and feedback in real-world scenarios. Similarly, a multi-layer architecture made of a device layer, fog computing layer, and a cloud layer is presented in [150]. The device and fog layer house the physical devices in the environment and the fog devices, respectively. The cloud layer separates the data and devices from the application and manages the data, contains the rule and data analytics engine. The architecture is implemented in an AAL environment with real-time monitoring of HR and an indoor positioning system. The communication was done through the MQTT protocol. The data is stored using a Redis database in the fog layer and MongoDB in the cloud. Feria et al. [151] created an architecture of 3 separate layers. The remote portable device layer combines all the sensing and action devices in the physical world and allows data collection and manipulation of the environment. In their implementation, the devices in this layer contain either a sensor or an actuator, microcontroller, and a communication module, BLE. The second layer receives data from the remote portable device layer and coordinates the devices through adding, removing, or applying changes to devices. The collected data is also manipulated and temporarily saved before being sent to the next layer, and time-sensitive reactions are made at this layer. Therefore, it can be thought of as a fog layer like that proposed in [139], [150], [138]. The final layer is called the web service application layer. It reorganizes the data to be presented to the users in different forms of user interfaces. The architecture utilizes BLE for communication with devices and JSON and RESTful API for communication with the web service layer. The system is presented to the user as a Service-Oriented Architecture (SOA), where the users can interact and control the devices and functionalities in the form of services. The system, however, was not tested in a real-world environment or implemented as a prototype.

2.6 Discussions, Challenges, and Future Directions

This section describes the open discussions of the reviewed frameworks and the challenges available in existing systems. In addition, the potential future research directions are highlighted to demonstrate the scope for further study.

2.6.1 Open Discussions

In this chapter, we have described smart healthcare frameworks highlighting areas such as health monitoring systems based on wearable devices and smartphones, disease detection using ML,

utilizing IoMT and social robots for AAL, and software integration architectures used to develop such assistive frameworks.

The summary of the health monitoring systems based on wearable devices and smartphones is illustrated in Tables 2.1 and 2.2. From Table 2.1, it is evident that almost all the systems can measure the heartbeat and body temperature of the patients, which shows the importance of these vitals. Additionally, some of the developed frameworks [61], [62], [63]. Furthermore, [65] measure BP along with the heartbeat and body temperature. The commonly used edge device is NodeMCU in most cases; some of the schemes [58], [59], and [60] used two devices as edge computing device. Almost all the systems used Wi-Fi for data transfer; only the systems introduced in [62] and [63] utilized Bluetooth for data communication. ThingsPeak and web applications are very common for data visualization that assists physicians in monitoring the patients. Further, Adafruit and LabVIEW are used in [60] and [63] for data visualization. A common negative aspect among the surveyed studies is that the frameworks are not adequately manufactured for clinical uses. From Table 2.2, it is found that most of the reviewed systems monitored a single sign for the patients except the frameworks developed in [70] and [77]. The camera module (rear and front) is used in most cases as a sensing element; only the schemes introduced in [70] and [74] utilized a microphone and accelerometer for data perception. The different iPhone and Samsung brand smartphones are used for the experiment, although the prototype presented in [74] did not mention any smartphone model. The highest number of participants (205 people) are found in [73], and the lowest number (5 individuals) is in [76]. The maximum and minimum video duration of the collected data is about 15 min and 10 seconds for [73] and [76], respectively. A data sampling rate of 30 Hz is utilized almost in all cases. No security concerns are handled in most smartphone-based health monitoring systems.

The summary of the ML-based disease diagnosis (COVID-19, HD, and diabetes) in the IoHT environment is shown in Tables 2.3, 2.4, and 2.5. It is observed from Table 2.3 that the reviewed systems utilized both the benchmark and real-time sensor data, as the amount of data for COVID-19 cases is relatively small. Most of the systems used CNN or variants of CNN as a classification algorithm; conventional ML technique is also used in some cases [89], [94], and [95]. The highest and lowest accuracy of 99.06% and 66.67% are found from [85] and [87]. Almost all the systems are not practically used in the target environment. It is shown from Table 2.4 that some common benchmark datasets like Hungarian and Cleveland databases, along with real-time sensor data, are used for the experiments. The highest accuracy of 99.45% is achieved from [104] using the MSSO-ANFIS classification technique, and the minimum value of 89% as an accuracy measure is obtained from [107] using bagging ensemble learning. No wearable prototypes are developed in almost all HD diagnosis systems. In diabetes detection, only a popular benchmark dataset (Pima Indians Diabetes dataset) is used in [122], and the other frameworks utilize real-time data collected from targeted individuals. Traditional ML techniques are applied to develop support

systems for diabetes patients. The accuracy values 99.66% and 80.51% are appraised from [118], and [119] are treated as maximum and minimum. The clinical trials of the developed frameworks are absent from the study.

AAL has taken a huge portion of researchers' interest in healthcare frameworks and IoT. The review presented here divided the efforts in the field into two portions. First, smart home environments utilize sensors and actuators to assist older adults in living longer, healthier lives while reducing the need for specialized healthcare professionals and its associated cost. Researchers focus on methods to track older adults' motion and identify their activities in their home environments. The systems developed in [131], [132] utilized BLE and RSSI to develop a low-cost localization system and deployed ML models for activity recognition [128]. Moreover, unobtrusive sensors have been of increasing importance over recent years. The primary motivation is to present solutions that would require minimal effort from older adults to incorporate into their daily lives. To that end, [129], [134], [137], [138] propose solutions that depend on radar, motion sensors, and door sensors to identify the senior's location, activity and detect dangerous situations such as falls. Finally, approaches that involve more dependency on user interaction, cloud computing, and storage for better predictive analysis are presented in [129], [136], [138], [152].

These studies are all presented as work-alone systems and do not attempt to capture the full capabilities that IoT offers to this sector. For example, several studies propose utilizing robotic systems to provide social companionship, assistance in ADL, and natural interaction with the provided home systems. Table 2.6 summarizes the approaches to develop social robots in recent literature. A main component of the provided robots is their ability to autonomously navigate the indoor environment and assist older adults without human interference. Other capabilities requested by various stakeholders include fetch and carry facial recognition and video chatting for telepresence. Researchers developed robots to act as the sole agent for the care of older adults [125], [142], [145]. Although this brings users personalization, it fails to capture the power of fully integrated smart sensors. To that end, [139], [140], [141], [143] integrate existing commercial robotic agents into IoT environments. The main system behind the robotic systems was ROS, and the main communication protocols utilized were REST and MQTT. The sensors linked to these robotic systems included smartwatches/bracelets and cameras. A fallback of the proposed systems is the lack of real-world validation and extensive testing. Such testing is needed with user feedback to showcase its applicability beyond research and improve the design and functionalities in a user-based approach.

The presented review also highlights the most recent suggestions for software architectures for smart healthcare. The main factor that combines the proposed systems is modularity and multi-layered architecture. The proposed architectures can be summarized to a generic framework as shown in Fig. 2.4, comprised of the main users, a device layer, a cloud layer, and user interaction devices. A perception layer or device layer is utilized to combine the sensory data from multiple

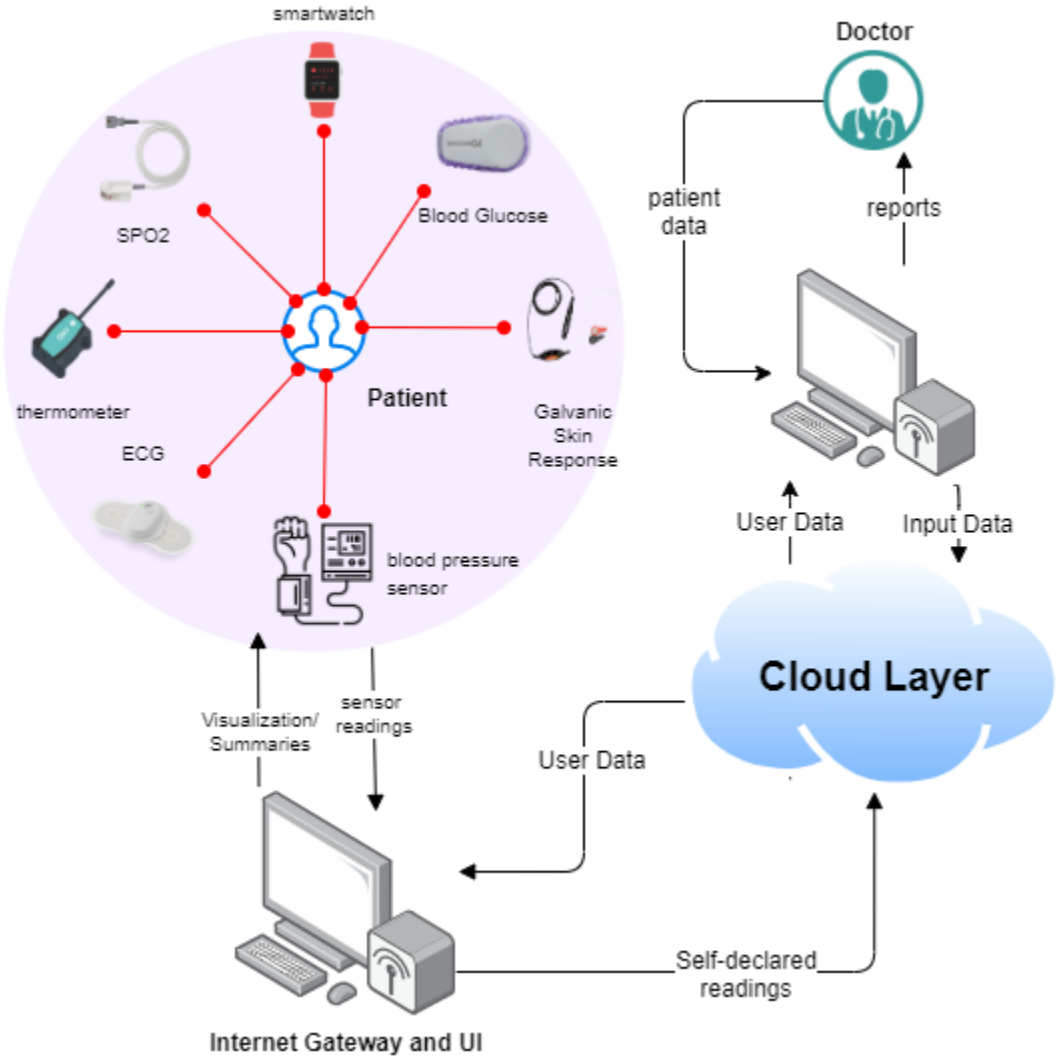


Figure 2.4: Generalized architecture for IoMT, comprising of main users, a device layer for data collection, a cloud layer for processing, and user interaction devices.

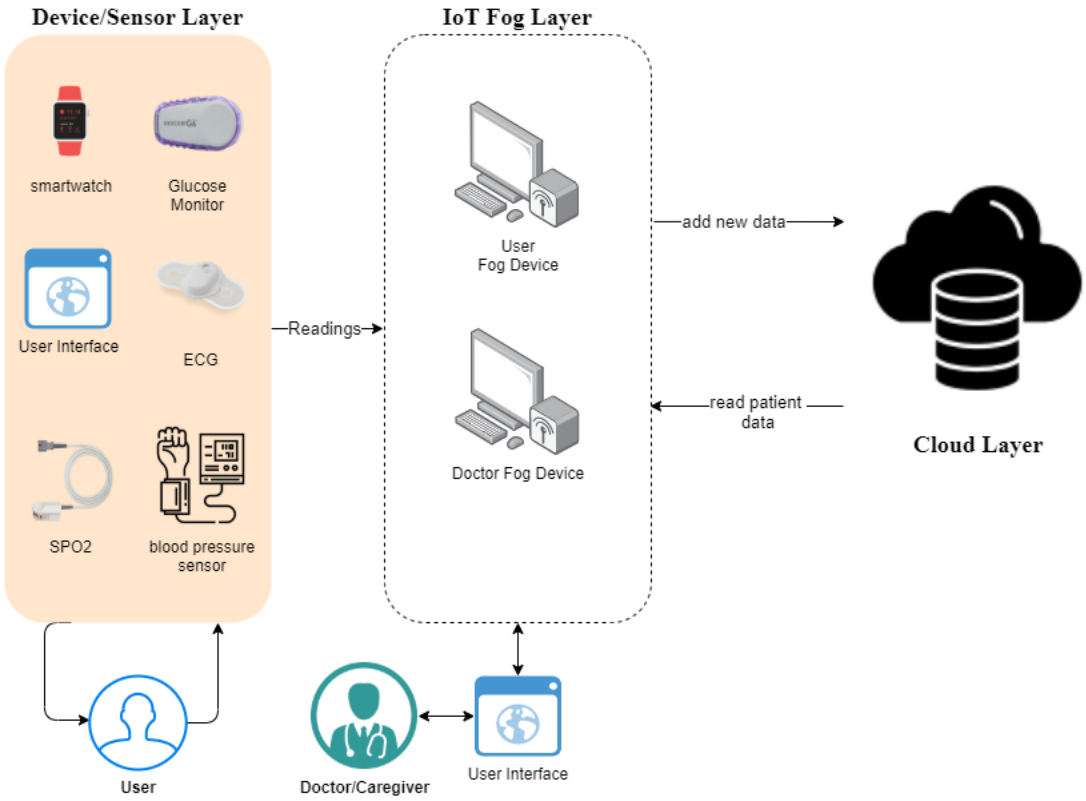


Figure 2.5: Generalized architecture for IoMT integrating fog and cloud layers enabling real-time analytics, rapid response, and scalable computing.

sources and stream it to the rest of the system for processing. Such a layer is utilized in most recent literature. The collected data is then processed in either a local device or in the cloud. Fog computing is proposed to accelerate response to real-time data and serve as temporary storage [138], [139], [150], [151]. Furthermore, researchers propose a link to the cloud for greater computational and storage power and enhanced scalability. This leads to the second approach, which combines both a fog layer and a cloud layer, as shown in Fig. 2.5. Several discussed studies have used the fog layer to provide real-time analytics and faster response to dangerous situations. The modularization of smart healthcare is essential, as shown in several studies explored in this chapter. However, this emphasizes standardizing communication protocols between the layers, especially between the devices and the device or perception layer.

Several studies opted to create their sensors to conform to their proposed architectures, thus decreasing the potential benefit. Conforming to COTS sensors and reaping the benefits of the wide-spread spectrum of sensors for remote healthcare is crucial, which led studies such as [139],

[147] to provide a direct link to incorporate these systems. A major fallback in most software architectures is their abstract nature. Developing market-ready software architectures goes beyond linking the layers and transmitting data. Other aspects such as security and privacy need to be addressed heavily to enable a commercial smart healthcare framework. Such aspects are usually addressed independently from the framework. The authors see a vital need to include them in the software architecture design to enable a complete framework for smart healthcare.

2.6.2 Challenges and Further Research Directions

Although several assistive frameworks have been developed using modern technologies to ensure smart healthcare, some challenges need to be addressed to ensure a scalable, secure, easily accessible, and efficient healthcare system. The main challenges, along with the potential future research directions, are demonstrated here.

The major challenge for implementing smart healthcare using wearable devices, including smartphones, is integrating the data from different sensors. The various sensors generate several data types. It is important to convert the signals from heterogeneous sensors attached to patients to a meaningful format for health monitoring applications. Several data fusion techniques [153], [154], [155] for integrating information derived from multi-sensory devices can be investigated as a means of providing streamlined signals for improving reliability and minimizing the bandwidth required for communication with the cloud layer as future work. Further, a hybrid body-sensor network architecture based on multi-sensor data fusion approaches will be investigated based on the work of [156], [157], [158].

Another key issue relating to the healthcare system based on wearable devices is the security and privacy of patients' responsive health records. The security issue has become a widespread and continuous challenge for wearable devices in IoT environments because of the increasing complexity of the data and the progressive network attacks. In the future, more secure and privacy-preserving frameworks using different security ensuring protocols like Blockchain [48], [159], [160] are recommended that can provide secure data communication among the users (patients and their families, medical experts, and caregivers). Low power consumption and energy efficiency are very significant for smart healthcare systems based on wearable devices and smartphones, especially for long-term patient monitoring. These issues can be handled by using low power equipment [161], long-life batteries [162], and energy harvesting techniques [163], [164] in future research. Another way to increase the battery's lifetime is the sensors 'sleep' and 'wake-up' properties employed to ensure the desired goal.

Along with the previous challenges, smartphone-based health monitoring systems face some noise as the collected data from smartphone cameras are in image/video format. Generally, the

noise in the data delivers misinformation to the users. Some major developments for healthcare applications are needed that are capable of handling noisy environments [165], [166], or some noise-free solutions [167] as future works. The developed prototypes should be maintained like low-cost, easy-to-use, and compatible platforms to increase the acceptance rate. More research and development activities can be ensured for developing assistive devices for smart healthcare considering the user needs. ML-based disease diagnosis systems also suffer from various unique challenges that need to be resolved to develop efficient and accurate frameworks for disease detection in IoHT environments.

Remote patient monitoring raises several real-world challenges, such as what to do with missing or incomplete data. Loss of electric power may cause the loss of some data being collected. In the worst case, a natural catastrophe such as an earthquake or weather-related event may cause data loss before archiving it at a central cloud location. This would be particularly problematic for patients with serious illnesses at home. Also challenging is when multiple patients have severe conditions that require assistance beyond what healthcare teams can respond to quickly. There will need to be ways to have a way to send those requests to another healthcare provider. Wearable devices may also fail, and so there may be incomplete and inconsistent data. We will need ways to deal with missing data, such as those proposed by Kaur and colleagues [168], [169].

Another challenge is to fit ML and DL algorithms with a small amount of data. To resolve the issues related to data shortage, optimized learning algorithms [170], end-to-end architecture [171], and synthetic data generation using generative adversarial networks [172] are highly recommended for future studies. Furthermore, there are several unnecessary features in the dataset for detecting HD responsible for occurring the disease, and these features often degrade the performance of the developed systems. In this scenario, the use of some optimization and feature selection algorithms such as genetic algorithm [173], particle swarm optimization [174], principal component analysis [175], etc., would be a potential solution to improve the performance of the detection procedure in future research. Furthermore, most of the studies depicted that the diabetes detection systems used data from the glucose sensors to achieve their goal. Hence, the proper design of sensors with long life would be an excellent approach for diabetes monitoring and detection. The design of lightweight ML frameworks [176], [177] would be better suited in the embedded devices to ensure smart healthcare systems in the future study. Overall, it is found from the reviewed systems that the developed prototypes are not entirely manufactured for practical uses. In some cases, no clinical tests have been conducted yet. Most of the systems represented their results considering the laboratory environment. Addressing these issues considering human health conditions will lead to potential research directions in the future.

AAL is a crucial application for smart healthcare, owing to its benefits in improved and 24-hour monitoring of health and cost reductions on both the older adults and the healthcare system. Nevertheless, there are some challenges left to tackle. First, a user-based approach is required

with feedback from older adults and redesign of the system accordingly. The aging society has special needs that are tough to identify without real-life tests and formal user studies. Moreover, using robotic agents and their integration into a smart environment is important for physical assistance and social presence. Usability and acceptability of smart healthcare systems by older adults is linked to enabling independent living in their own home [123], [127], [128], [143], personalization [128], [129], and the intuitiveness of user-interface as highlighted in [125], [127], [145], [139]. As a result, it is important to directly tackle these three aspects in designing AAL systems and follow user-based testing and improvement with older adults in future studies.

Software architectures meant for smart healthcare have been improving and following the same direction of modularization and scalability. Recent studies highlighted the need for providing means of integrating COTS and therefore standardizing communication technologies and protocols. Such protocols need to enable multiple user interfaces to accommodate different users and applications. Moreover, both fog devices and cloud integration are needed for real-time response and the power of the cloud for big data analysis, storage, and scalability. Finally, security and privacy concerns are of utmost importance when it comes to health-related data. Consequently, these aspects need to be part of the design of the smart healthcare frameworks at an early stage, utilizing the most recent advancements in Blockchain technologies and allowing access to data appropriate to users in question.

2.7 Conclusion

Smart healthcare provides a secure, effective, and easily deployable health monitoring system that can ensure quality healthcare services at a fraction of the cost currently incurred by hospitals or assisted living centers. In this chapter, we briefly discussed the SOTA wearable devices and smartphones for basic signs monitoring, ML for three significant diseases (COVID-19, HD, and diabetes) diagnosis, and the frameworks developed to aid the adults in AAL. The software integration frameworks that are very substantial to develop smart healthcare are demonstrated in a nutshell in this chapter. We have reviewed the advantages and shortcomings of a wide range of systems. In addition, we discussed the major challenges of recently developed smart healthcare frameworks that are the main obstacles to develop assistive prototypes. Some potential future research directions are recommended for the further improvement of the existing healthcare system. It is quite impossible to replace the whole medical system with technology, but it can reduce the burden of medical experts by introducing some novel architectures. The development of such assistive systems would be feasible while the medical experts and the researchers would work jointly in a platform.

Chapter 3

IoT in Healthcare: Devices, Architectures, Protocols, and Applications

3.1 Introduction

The IoT concept represents a set of Internet-connected devices that are empowered with computing capability, are easily recognizable, and can share data over the Internet without the need for physical contact. The IoT revolution has resulted in unprecedented levels of interconnection among devices, allowing for the development of smart environments. Smart connectivity with the existing network, as well as context-aware information processing employing network resources, are essential components of IoT [178], [179]. The IoT allows physical things to see, hear, think, and execute a variety of tasks by allowing them to "speak" to one another, exchange data, and make choices. The concept of IoT turns these devices from being conventional to smart by leveraging its wide range of technologies including embedded applications, Internet Protocol (IP) and applications, wireless communications, distributed networks, and sensor technologies [180], [181]. An important advancement of the current Internet into a network of interconnected objects is that it not only comes into contact with the physical surroundings via actuation, controlling, and monitoring, nor merely extracts information from the surroundings, but it also uses traditional network requirements to enhance data transfer, predictive analysis, and communication [181]. The IoT covers a huge number of applications including healthcare, automotive, utilities, smart cities, wearables, smart homes, and smart farming [182]. Numerous research projects, enterprises, and businesses are currently engaged in the creation of various IoT features to accommodate the growing evolving needs that emerge with such rapid development.

In general, IoT technology contains several unique features [183], [184] that make this area

of research from traditional ones. The main feature of IoT includes connectivity, which means that there needs to be a link among different levels for everything that happens in IoT hardware and devices, including sensors, connected hardware, control mechanisms, and other electronics. Additionally, connected devices need to share data with other devices in the IoT environment. Another fundamental feature of IoT is scalability, which is very important as the number of devices connected to the IoT environment is increasing. An IoT system needs to be able to handle the massive expansion required to properly process the generated data in order to keep up with this rapid development. Intelligence is another significant feature of an IoT system that enables an intelligent framework to extract knowledge from the generated data. It is significant as it offers a means for a device to gradually become more intelligent. Another key feature of IoT is security which ensures the security of Internet devices and networks. The security feature is useful for protecting, detecting, and tracking threats associated with a variety of devices. Lastly, dynamic nature is the key another element of an IoT system. This feature ensures that IoT devices should dynamically adjust to their area of operations, sensing environment, and contextual information as well as take appropriate actions.

Recent developments in wireless networks and IoT have motivated various research attempts to analyze related literature. Every day, innovations in IoT technology, especially in the field of healthcare are made. In recent times, a number of survey papers that were published reviewed the IoT technology from distinct viewpoints. The survey work done by Alam et al. [185] presented several communication protocols and standards that are highly applicable to healthcare applications. This survey demonstrated the SOTA with five healthcare application scenarios. However, this survey did not highlight any current developments or contributions towards alleviating the problems and difficulties in IoT-based healthcare applications. Baket et al. [186] discussed various parts of IoT-based healthcare applications including the IoT architecture, and components. The chapter highlights the use of wearable sensors, communication technologies, cloud applications, and big data in IoT-based healthcare applications. However, some of the important aspects of IoT-based healthcare applications such as data processing techniques, ML, and network management are excluded from this survey. Baali et al. [187] reviewed wearable sensors with their working principles and operations in IoT-based healthcare applications. This study explored wireless communication technologies, on-sensor pre-processing abilities, and power systems in sensor nodes. However, this survey failed to describe network-layer protocols which are restricted to the legacy protocols and methodologies. Moreover, some significant aspects of healthcare in IoT including IoT device capabilities, and ML approaches are not demonstrated here. In another survey, Sethi and Sarangi [188] discussed several IoT stack taxonomy with a synopsis of the technology and protocols used at each layer. This review analyzed several kinds of IoT sensors with their relevant applications. Moreover, the difficulties in introducing the concept of middleware were highlighted along with their various forms. Silva et al. [189] provided a brief overview of one

kind of IoT architecture and related technologies. This survey also briefly explained some of the common communication standards and protocols used in the IoT ecosystem. Following describing a few IoT applications, certain problems and security challenges were briefly discussed in this survey. However, an inadequate investigation that raised a lot of issues regarding the complexities of the topic misguided the readers. In [190], the authors demonstrated the features and visions of IoT as well as insights into some supporting technologies and communication protocols based on their functionalities. Besides, this review briefly examined the middleware and network domains. Moreover, this survey focused on addressing and exploring the difficulties and open problems that the IoT paradigm faces. In another review, Balaji et al. [191] discussed some of the technologies and protocols used in the IoT domain, followed by providing a synopsis of some of the security challenges this area faces. The authors highlighted some real-time applications as well as well-known IoT-based lifesaving technologies. Lastly, a few of the common IoT problems were discussed, but the future scopes were omitted in this study. Ahad et al. [192] reviewed the architecture and taxonomy of the 5G-based smart healthcare network, including the communication technologies, goals, performance measures, and requirements. Besides, this paper presented a detailed review of several scheduling and routing strategies to meet various goals and needs of smart healthcare. Further, the authors discussed several open problems and difficulties in smart healthcare. The review work done by the authors of [193] presented a thorough analysis of several IoT-based smart healthcare systems that support data transmission and data reception. Moreover, this survey offered a thorough review of power management, resource management, security and privacy, and energy management in relation to IoT-based smart healthcare. Qi et al. [194] reviewed various IoT applications for smart healthcare, focusing on the aspect of basic symptom monitoring. Additionally, this review discussed in detail current IoT technology for smart healthcare applications, covering a variety of topics including networking, data processing, and sensor devices.

This chapter is aimed at highlighting the fundamental issues of IoT device capabilities, architectures, protocols, and healthcare applications (see Fig. 3.1) in detail. The IoT device incorporates all the necessary components for hardware and software platforms. The hardware platform describes the most popular portable processing devices such as Raspberry Pi, Arduino, and ESP8266 highlighting some significant features including power requirement, memory capacity, and processing power. The software platform is described focusing on the Operating System (OS)s for small embedded devices and built-in software for IoT-related tasks. The OSs named Contiki, Mbed, RIOT, Embedded Linux, and Windows 10 IoT are described here with their distinguished features. The popular and widely used built-in tools for IoT smart applications include DeviceHive, Kaa, ThingsBoard, Amazon Web Services (AWS) IoT, and Google Cloud IoT. The IoT architectures such as three-layer, SOA-based, and middleware-based architectures are depicted to provide cutting-edge insights. The three-layer architecture contains the

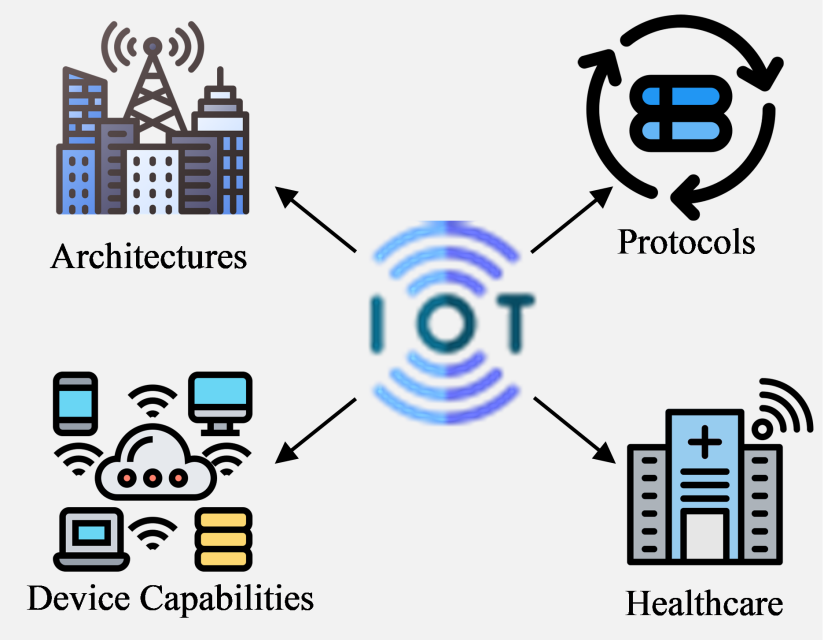


Figure 3.1: Graphical view of the major focuses of this chapter including IoT device capabilities, architectures, protocols, and smart healthcare applications.

layers of perception, network, and application. Additionally, the SOA-based architecture comprises a service layer along with the layers of the three-layer architecture. The middleware and business layers are the additional layer in the middleware-based architecture over the three-layer architecture. The IoT communication protocols are also demonstrated here to provide an overview of communication interfaces. The eight popular protocols that are described in this chapter are Constrained Application Protocol (CoAP), MQTT, Extensible Messaging and Presence Protocol (XMPP), Advanced Message Queuing Protocol (AMQP), Data Distribution Service (DDS), Low Power Wireless Personal Area Networks (LoWPAN), BLE, and Zigbee. The most promising healthcare applications for monitoring, diagnosis, and treatment for individuals using IoT technologies are demonstrated. The major applications include biophysical parameter monitoring, chronic disease detection, medication management, telehealth, as well as home and elderly care. The current challenges for IoT technologies and potential future trends are explored to find out the research gaps and potential solutions to develop smart applications. The key contributions of the chapter are highlighted as follows.

- Providing a comprehensive discussion on IoT device capabilities highlighting the area of hardware and software systems including the processing capabilities, OSs, and built-in

tools.

- Addressing core technologies for IoT architectures that are the fundamental issues in diverse application domains of IoT and a detailed discussion on it.
- Analyzing typical IoT communication protocols that deal with data transmission concerns via networks, as well as providing an overview of each.
- Exploring the in-depth analysis of the most recent and promising IoT-based healthcare applications incorporating different architectures and protocols for monitoring, diagnosis, and treatment of patients in hospitals as well as smart home environments.
- Presenting a study of the major open issues for application approaches in the IoT context, as well as the future trends to provide the scope of further research for smart applications.

The remaining parts of the chapter are organized as follows. The paper collection and selection criteria are demonstrated in section 3.2. Section 3.3 provides the concept of IoT device capabilities with proper illustrations of hardware and software platforms. The architectures that are available in recent times to develop smart systems are described in section 3.4. Section 3.5 demonstrates the IoT protocols that are generally used to transfer data over the Internet. The healthcare applications of IoT technology are briefly described in section 3.6. Section 3.7 demonstrated the open discussions of the IoT technology with the current challenges as well as the potential future works. Section 3.8 concludes the chapter.

3.2 Paper Collection and Selection

To provide a systematic review of the application of IoT in smart healthcare after discussing the components of IoT, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology [195]. PRISMA streamlines the paper selection process for systematic reviews. PRISMA process has the following sequential steps for completing a systematic review: identification of papers related to the research domain, elimination of duplicate papers using scanning, filtering of remaining papers based on eligibility criteria, and creating the final list of papers to be reviewed by including the selected papers. The paper selection process through PRISMA for selecting the papers for a systematic review is illustrated in Fig. 3.2. For this review work, we considered the peer-reviewed research works published in reputed international venues mostly within the years from 2017 to 2022. We only considered the research works that were written in English. We initially collected 739 papers from Google Scholar, 157 papers from

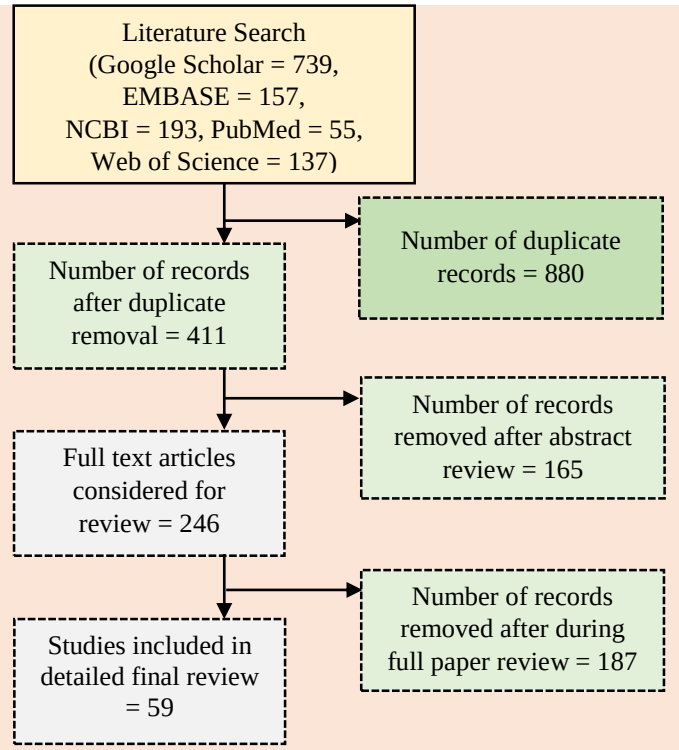


Figure 3.2: Flow diagram showing article selection for final review. The final selection is done based on the duplicate records, abstract review, and full paper review.

EMBASE, 193 papers from NCBI, 55 papers from PubMed, and 137 papers from Web of Science. The majority of the papers were collected from Google Scholar as Google Scholar is a research work aggregator service, and its database is oftentimes comparable to other commercial services. The keyword "Healthcare applications of IoT" resulted in the highest number of relevant papers. Searching on Google Scholar with this keyword and the minimum year set to 2017 resulted in 60,600 entries. From all the collected papers, 880 papers were discarded due to duplicity. We removed 165 papers from the review due to topic mismatch. These papers were removed after we read their abstracts to determine if their topics aligned with the research domain. We fully read a total of 236 papers. The quality of these papers was assessed based on plausibility, representation, methodological mistakes, and the significance of IoT in the work. After careful consideration, a total of 59 works were selected for inclusion in our review work. The systematic review follows the three steps presented in [196]: definition of research questions, paper collection process, and information extraction. We expand on these three steps in the context of our systematic review in the followings.

Research Questions: After discussing the various components of IoT in detail, we aim to provide a systematic review of the application of IoT in the healthcare domain. We provide a comprehensive survey of the various applications of IoT including applications in biophysical parameter monitoring, chronic disease early detection and monitoring, medication and habit man-

agement, elderly monitoring and care, virtual care, telehealth services, telesurgery services, and telemedicine services. We discuss the overall methodology, important components, and results of the selected systems. An in-depth discussion on the various open challenges and recommendations to overcome them is provided. We also discuss the future directions for IoT as a whole. The following research questions guided us in this systematic review: "What are the applications of IoT in current smart healthcare systems?", "How can IoT improve current healthcare systems?", "How do the research works in a specific healthcare application differ from one another?", "What common issues are currently present in smart healthcare systems?", "How can we overcome the current challenges in smart healthcare systems?", and "What are the future research directions for improving the smart healthcare systems?".

Paper Collection: Some notable sources used for searching and collecting relevant papers for this systematic review work are Google Scholar, EMBASE, PubMed, and NCBI. We used the following keywords while searching the web and databases: "applications of IoT in smart healthcare", "IoT based smart healthcare solutions", "IoT based biophysical parameter monitoring systems", "IoT for disease detection and monitoring", "IoT based solutions for medication management", "applications of IoT for elderly monitoring", "applications of IoT in virtual care", "IoT based telehealth systems", "IoT telesurgery systems", and "IoT for telemedicine services". The abstracts and titles of the search results were analyzed to eliminate duplicates and to make sure the selected articles fall within the scope of the systematic review work. The specific topics of interest were identified and quantified after thoroughly analyzing each selected article. Some of the articles were removed from the research work during the quality assessment process due to reasons such as IoT not playing a significant role in the overall system and not showing the IoT portion in detail in the system overview. We used Mendeley as the citation management tool to create a bibliography. The broad requirements for all papers to be selected for the systematic review are as follows: the papers should be written in English, the paper should be related to healthcare, and IoT should play a significant role in the overall system.

Information Extraction: The following information were extracted and discussed from the reviewed papers: overall methodology, used sensors, details of the employed IoT components, details of the physical prototypes, hardware specifications, specific advantages, drawbacks, and quantitative and qualitative results if available. The information was extracted and summarized after reading through the full text of the selected papers.

3.3 IoT Device Capabilities

In general, an IoT framework comprises devices that perform sensing, actuation, control, and surveillance tasks. IoT devices are able to transmit data to other devices and applications over

the Internet, or they can retrieve data from other connected devices and process it locally or on a cloud server after sending it to centralized servers. The tasks are completed on the local server or within the IoT architecture depending on some temporal and spatial restrictions such as memory, communication delays, deadlines, computational power, and speed of operation. An IoT device may have multiple interfaces for communicating with other devices, in both wired and wireless modes. The common interfaces that are equipped with most IoT devices are I/O interfaces for sensors, Internet connectivity interfaces, memory and storage interfaces, and audio/video interfaces. There are various types of IoT objects such as wearable sensors, accelerometers, cameras, smart watches, smart locks, GPS, and LED lights that are available for smart applications. Most IoT devices generate a lot of data in different forms, and when these data are analyzed by data analytics engines, they include vital information that may be utilized to guide nearby or remote activities. The IoT device capabilities highlighting hardware and software platform is shown in Fig. 3.3. In the following section, the device capabilities including hardware and software platforms are demonstrated in brief.

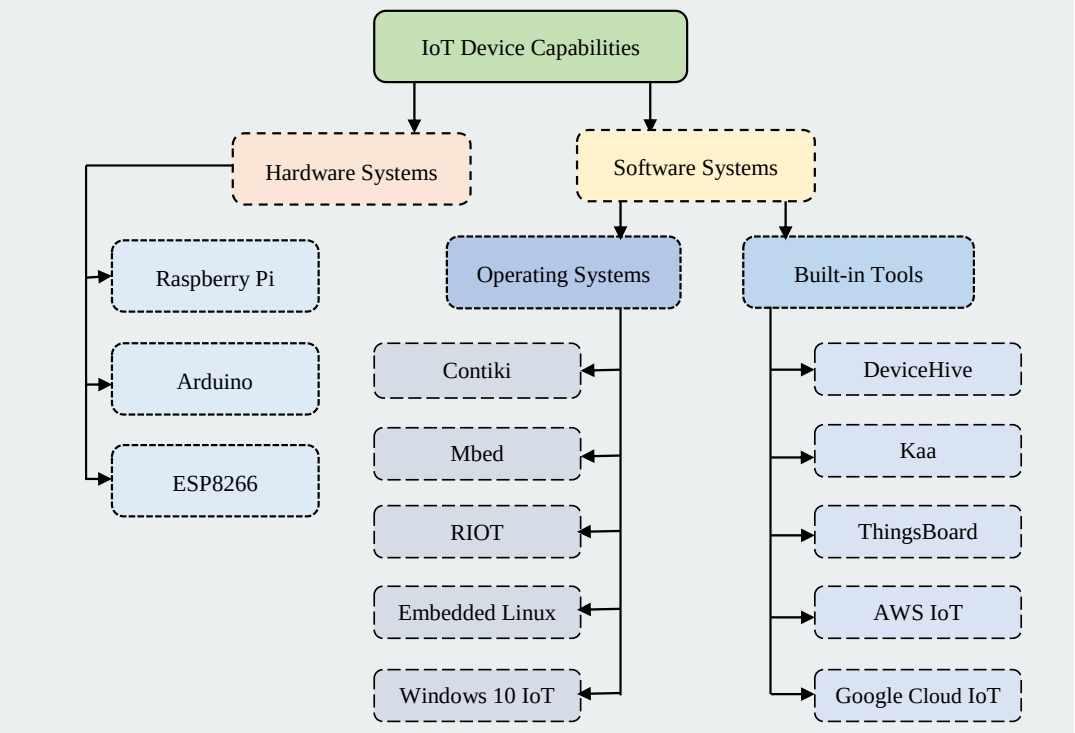


Figure 3.3: IoT device capabilities considering hardware and software platforms. The hardware platform includes several types of tiny processing devices, such as Raspberry Pi, Arduino, and ESP8266. The software module comprises several OSs and built-in tools.

3.3.1 Hardware

The hardware components for the IoT have arrived in a variety of shapes and sizes, depending on the tasks at hand. For instance, the core CPUs are used to operate the phones, the sensors are utilized to perceive data from the physical environment, and the data are processed and analyzed on edge devices. In the case of wired mode, the hardware aspects are the essential components of the IoT platform, and the capabilities of these devices have only grown in significance as IoT has progressed. The main hardware kits that are widely used to develop IoT applications are described here focusing on some properties such as processing power, memory capacity, and power requirement.

Raspberry Pi: Raspberry Pi is a series of single-board computers that compress a lot of processing power into a tiny box and offer desktop-like functionality [197]. The components of Raspberry Pi vary based on its version. In the Raspberry Pi B+ model, there are four USB ports, two ports for the camera, an Ethernet port, an audio jack port, and an HDMI port. To implement real case studies, there are 17 General Purpose Input Output (GPIO) pins that are available to connect with any electronic components such as LEDs, sensors, switches, and motors. The minimum required logic level for the Raspberry Pi is almost 3.3V. In general, it is run over the Raspbian OS. Additionally, it is also compatible with other OSs such as Lakka, RaspBSD, and RetroPie [198]. The programming languages that are supported in Raspberry Pi are Python, Java, C++, and C. The combination of GPIO pins with the Linux OS gives this device a lot of flexibility and gives users a lot of choices. The CPU speed of the standard Raspberry Pi model is 700 MHz to 1.4 GHz, the memory capacity is 256 MB to 4 GB, and the power requirement is 2.5 Amp. These measures are varied based on the version of the Raspberry Pi model. The Raspberry Pi is well-suited for a wide range of applications due to its high connectivity, ease of use, portability, low power consumption, high flexibility, and high transferability [199]. However, there is no power button to keep it in sleep mode while it becomes idle, no built-in ADC, and replicability is quite hard for this device.

Arduino: Arduino is a popular and frequently used open-source electronics framework based on simple hardware and software that can be utilized to develop electronic projects. It offers a wide variety of products, from straightforward 8-bit microcontroller boards to wearable technologies, IoT components, and 3D printing [200]. The set of instructions to complete any tasks is executed through microcontrollers. There are several versions of Arduino such as UNO, micro, pro, mega, and nano available in the market for electronic system development where the use of the Arduino model varies based on the nature of the applications. It is easily compatible with MAC, windows as well as Linux platforms. A type B USB cable is utilized to do programming on the Arduino board with the support of an IDE [201]. The supported language for Arduino is C++ where the fundamental issue of peripherals such as sensors and actuators are controlled through the use of

header files. Arduino Uno has analog pins, digital pins, a USB connection, a reset button, a crystal oscillator, and a power jack. Arduino is highly suited to perceive data from physical environments i.e., analog data [202].

ESP8266: ESP8266 is a mini-sized, Wi-Fi enabled device with a full Transmission Control Protocol (TCP)/IP stack and microcontroller capability which is frequently used for embedded development in IoT environments [203]. This device can host an application that offloads all Wi-Fi network services to another underlying hardware. The Wi-Fi module of ESP8266 is pre-programmed with an ESP-AT command set firmware that enables any Arduino device to get a Wi-Fi connection. It runs at 80 MHz and uses a 32-bit Reduced Instruction Set Computer (RISC) CPU based on the Tensilica Xtensa L106. The boot ROM, instruction RAM, and data RAM size of ESP8266 are 64 KB, 64 KB, and 96 KB respectively. The Serial Peripheral Interface (SPI) accesses the external flash memory. ESP8266 is quite fit for low-cost systems [204] that do not require a high amount of CPU support, huge storage, and some special features such as a camera, USB, and HDMI ports.

Table 3.1 shows the different properties including CPU speed, memory capacity, power requirement, supported programming languages, and storage of the above-mentioned hardware components.

Table 3.1: IoT hardware components with some properties.

Property	Raspberry Pi	Arduino	ESP8266
Programming	Python/Java/C++	C++	C++/Lua
Wi-Fi	USB dongle	Shield or ESP8266	Built-in
Code Distribution	In-Situ	USB/SPI/Serial	Serial/OTA
Storage	SD-card	Built-in	Built-in
I/O	17 GPIO	Arduino UNO: 13 GPIO/6 ADC	10 GPIO/1 ADC
CPU Speed	700 MHz to 1.4 GHz	16 MHz	80 MHz
CPU	ARM Cortex-A7, ARM Cortex A-53, ARM Cortex-A72	Atmel AVR, ARM Cortex-M0+, ARM Cortex-M3, Intel Quark	Tensilica L106 32-bit RISC processor
Memory Capacity	256 MB-4 GB	16 KB-64 MB	128 KB-4 MB
Power Requirement	5V, 2.5 Amp	5V, 250 mA	3.3V, 70.5 mA
Cost	Expensive ($35)	Moderate ($3-$30)	Low cost ($3)

3.3.2 Software

In general, IoT software outlines the key aspects of communication networks and action via numerous channels, integrated devices, partner systems, and development tools. IoT software is the underlying platform that allows a connected device to capture and communicate data in real-time where the data are generally processed by computers and relevant applications. The connected devices can operate effectively owing to the intelligence that IoT software ensures. Smart devices are limited to their computing power and in such scenarios, intelligence is ensured through cloud computing technology where the IoT software is located. The OSs along with built-in tools which are the parts of IoT software are described here.

Operating Systems: An IoT OS is a part of software platforms that ensures how IoT smart devices interact with users and handle all software and hardware resources. IoT OSs are integrated OSs that are intended to work in small IoT devices with limited memory and low processing capacity. IoT devices can interact with cloud platforms and other intelligent devices over a global network thanks to embedded OSs. The OSs for a particular application are chosen based on some performance measures including flexibility, compatibility, reliability, simplicity, and consistency. The popular IoT OSs are demonstrated here to provide a short overview of embedded OSs.

Contiki: Contiki is a freely available OS that is well-known for its way of interacting with extremely small, low-power, and cost-effective embedded systems to the Internet [205]. The programming language that is used in Contiki is C. It supports multithreading via protothread, and the processes are managed through cooperative or preemptive scheduling. Generally, multiple stacks are supported in Contiki including some features such as 6LoWPAN, CoAP, and IPv6 [206]. This OS is appropriate for developing complex wireless systems due to its high memory efficiency.

Mbed: Mbed OS is a publicly available, open-source OS that is quite popular for its integration with ARM processors and has a wide range of connectivity features such as Wi-Fi, Ethernet, 6LoWPAN, Cellular, and Bluetooth [207]. It is particularly tailored for 32-bit ARM processors. This OS has a multilayer security system that provides users with a high level of reliability. Mbed OS is well-suited to IoT applications on ARM cortex M-based devices, as it preserves code clean and portable features [208]. It is a popular IoT OS due to its low system requirements and compatibility with a variety of development kits.

RIOT: RIOT is another widely used and popular OS specifically designed for embedded devices [209]. This OS targets wireless sensor networks as it is developed based on a microkernel architecture named FireKernel. It ensures the usage of multithreading and SSL/TSL libraries, as well as provides the facilities for the use of 8, 16, and 32-bit processors. The supported language for RIOT is C and C++. RIOT can run as a Linux or macOS process as there is a dedicated port

for this OS. RIOT enables modular design and provides access to separate hardware abstraction for minimal memory and energy consumption [210].

Embedded Linux: Embedded Linux is a significantly modified version of the Linux kernel, designed specifically for embedded devices. The compact size and low power consumption of this OS make it well-suited for fulfilling all the requirements of IoT devices such as tablets, navigational devices, and wireless routers [211]. It requires a small amount of storage like 100 KB of memory which makes it quick and dynamic. In the IoT OS world, this OS also offers an unparallel level of customization. Embedded Linux OS is easily compatible with any single-board computer such as Raspberry Pi (Raspbian, Ubuntu Core, OSMC, and Gentoo. Linux distros are popular options for IoT applications due to some performance factors such as flexibility, consistency, architectural layers, cloud support, and visualization [212].

Windows 10 IoT: Windows 10 IoT is a product of Microsoft Windows 10 OS, but it is dedicatedly designed for IoT devices [213]. Windows IoT is divided into two parts: Windows 10 IoT core, which is considered to interact with tiny embedded devices, and Windows 10 IoT Enterprise, which is developed for industrial use. Windows 10 IoT includes a user-friendly interface, more user control than some other OSs, and is universally recognized as a capable IoT OS among IoT experts. Windows OS is a quite better choice over free or open-source ones whenever reliability and security become critical issues [214].

Table 3.2 provides an overview of the discussed IoT OSs including some dominant features such as accessibility, provider, license, processor, and use cases.

Built-in Tools: Generally, IoT software has two main attributes such as (i) Connectivity with wired or wireless devices via Bluetooth, Zigbee, Ethernet, 5G/6G, or satellite communications. (ii) Automated communication between the hardware, devices, or machinery without human involvement to initiate an operation or carry out actions. The available built-in tools are developed based on these two characteristics to carry out automated data analysis and transfer, as well as autonomous response. The most popular IoT built-in tools are described here to provide in-depth knowledge to the researchers.

DeviceHive: DeviceHive is a publicly accessible platform that offers communication and management tools for connected devices. There are several things included in DeviceHive such as a communication layer, control software, multi-platform modules and users to launch smart home development, remote control and monitoring tool, and telemetry [215]. DeviceHive is compatible with Python, Node.js, and Java client modules. It uses the REST, WebSocket, and MQTT protocols to connect with other devices. This platform is a modular, hardware, and cloud-independent micro service-oriented solution with access control APIs in several protocols that enable users to monitor, control, and analyze the behavior of their devices.

Kaa: Kaa is developed for enabling smart devices to communicate with each other across a

Table 3.2: A short look at the IoT operating systems.

Operating Systems	Features	Provider	License	Processor	Use Cases
Contiki	Open-source, free	ThingSquare	3-clause BSD	ARM Cortex-M, MSP430, AVR, and x86	Networked memory-constrained systems
Mbed	ARM-based, high-grade security	ARM	Apache 2.0	ARM Cortex-M	For portable code
RIOT	Open-source, full multithreading	FU Berlin	LGPLv2.1	ARM Cortex-M, MSP430, ARM7, AVR, x86, and Cortex-M23	Can be run as a MacOS process
Embedded Linux	Linux kernel, free	Multiple providers	GPL, and GPLv2	ARM Cortex-A8, ARM Cortex A-53, and Intel Core i3, i5, i7	Versatile, can be used for various use cases
Windows 10 IoT	Proprietary, high-grade security	Microsoft Corporation	Commercial	ARM Cortex-A7, Snapdragon 400, ARM Cortex-A53	Ideal for heavy-duty industrial use cases

big cloud. Kaa is an open-source development tool, and it is hardware-oriented i.e., it is well-suited for any hardware components such as sensors, gateways, and devices. Kaa's multi-purpose middleware enables users to develop IoT solutions, application support, and a variety of intelligent devices [216]. The setup of the Kaa tool is quite easy and it provides a large number of features that can be integrated into the platform easily. Kaa is a scalable IoT platform that uses a microservice design to provide obvious layers of abstraction, flexibility, and scalability [217].

ThingsBoard: ThingsBoard is a free server-side built-in tool for data collection, processing, monitoring, and device management [218]. It is deployable in any smart application including personal and commercial purposes. ThingsBoard supports several protocols such as CoAP, MQTT, and HTTP. The IoT devices are easily connected to the platform via the ThingsBoard gateway. Although the back end of ThingsBoard is written in Java, it supports several micro services based on Node.js. Some of the major reliant features of ThingsBoard are scalable, fault-tolerant, robust and efficient, adaptable, and resilient [219]. In practical scenarios, it is well-suited for predictive analytics tasks such as smart farming, smart metering, and smart energy.

AWS IoT: The AWS IoT tool is a reliable and secure development platform for industrial, business, and retail solutions. This platform can be used to analyze and manage data from Internet-connected devices and sensors in residences, businesses, healthcare, automobiles, and other loca-

tions [220]. It supports HTTP, LoRaWAN, the lightweight communication protocol, and MQTT to publish and subscribe to messages. This tool incorporates layers of security like encryption, secure access, constant monitoring, and audits. AWS IoT can handle smart objects and trillions of signals, and it can securely and efficiently interpret and transmit those signals to AWS endpoints and other devices [221].

Google Cloud IoT: Google Cloud provides scalable services that are quite managed and incorporated with IoT devices. The functionalities of this tool include connecting, preserving, and investigating data at the edge and in the cloud. The common aspects of Google Cloud IoT are data warehousing and quick querying, streaming and batch analysis, device connectivity and management, as well as AI and ML models [222]. This platform is capable of predicting equipment maintenance and optimizing operations, as well as monitoring resources in real-time. Google Cloud IoT covers a wide range of applications including predictive analytics, logistics and supply chain management, smart cities and infrastructures, as well as real-time asset surveillance [223].

3.4 IoT Architectures

The IoT should be competent at networking billions or trillions of diverse things over the Internet, therefore, a flexible architectural pattern is crucial. Some of the major factors like reliability, scalability, privacy and security, QoS, and data storage should be considered at the time of developing the architecture of an IoT network [224], [225]. To fulfill the demand for the defined key features, several research groups have proposed several IoT architectures. The ever-growing number of potential architectures has been unable to settle on a standard model [226]. There is no specific base architectural pattern till now despite numerous standardization initiatives. This is because of its wide area of applications having different design patterns and variables/factors to measure the performance of the existing systems. Fig. 3.4 depicts the most common and widely used IoT architectures, which are explained as follows.

3.4.1 Three-Layer Architectures

The most fundamental architecture is the three-layer architecture which was developed in the beginning periods of research in the IoT area [227], [228]. This architecture is divided into three base architectures such as perception layer, network layer, and application layer which are demonstrated below.

Perception Layer: In the IoT ecosystem, the perception layer is sometimes referred to as the sensor layer, and it is deployed in the bottom layer. The perception layer interferes with the en-

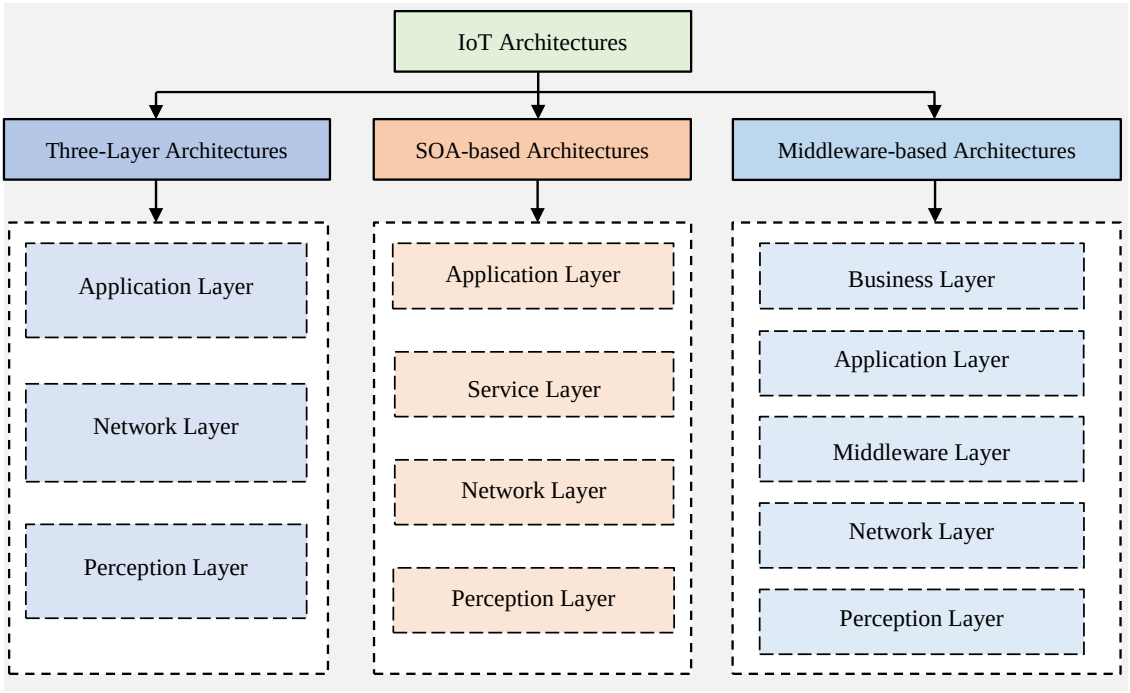


Figure 3.4: Most common and frequently used IoT architectures that incorporate the three-layer architecture, SOA-based architecture, and middleware-based architecture.

vironment via smart devices like sensors, actuators, cameras, and GPS terminals to collect and process data. This layer's primary functions include connecting devices to the IoT framework, measuring, and processing real-time physical parameters (data) from connected things via intelligent devices, and transmitting data to the higher layer via its interfaces [229], [230]. In the perception layer, heterogeneous devices are configured using standardized plug-and-play techniques [231]. The major success of IoT, known as big data, begins in this layer.

Network Layer: The network layer, also known as the communication layer and incorporated in the middle layer of the IoT framework, is referred to as the core of the IoT ecosystem [232]. The communication layer [233] is responsible for providing data with routing channels so that it can be transmitted in packets through the network. This layer is in charge of network connectivity, message passing among various servers, and routing between the cloud and devices. The network layer serves as a link between the perception layer and the application layer, allowing data to be transferred to IoT servers, applications, and devices through interconnected networks. This layer combines the Internet and communication networks. The network layer consists of different communication standards like Wi-Fi, Bluetooth, and Zigbee [234] and IoT objects like gateway, hub, and switching [235]. The use of wireless protocols is very important in this layer

because wireless devices can be deployed in critical environments with less material and human effort [236]. This layer generally transmits/receives data to/from various applications via interfaces/gateways using a variety of communication standards and protocols. Furthermore, unique addressing and routing capabilities guarantee the easy integration of several devices into a single collaborative framework, achieving the universality of the IoT concept.

Application Layer: The application layer is located at the top layer of the IoT network, and it serves as the interface between users and applications. This layer is in charge of retrieving data sent from the network layer and utilizing the collected data for specific operations [237]. As a task of the application layer, it provides a storage facility that can be utilized to store a backup of the acquired data into the database. Additionally, this layer predicts the future condition of physical things by analyzing the nature of the received information. From an application perspective, this layer incorporates the functionalities of the IoT framework to develop real-time applications including the healthcare system [238], smart cities [239], smart transportation [240], and smart agriculture [241].

3.4.2 Service-Oriented Architecture

Generally, SOA is a component model-based framework that connects various functional blocks of applications via interfaces as well as standard protocols [242]. SOA provides viability by allowing software and hardware components to be reused and coordinating the intended services. A new layer named the service layer is added between the network layer and application layer that ensures services are sufficient to assist the application layer in extending the three-layer architecture in the SOA framework [243], [244]. As a result, the SOA architecture differs from the conventional three-layer framework by including service layer functionalities. The service layer consists of several components such as service management, service discovery, service composition, and service interfaces [245]. The service management maintains and measures the trust processes in order to fulfill the service's requests. The service discovery unit finds the service's requests. The purpose of service composition is to communicate with the associated things and incorporate services to obtain the requests efficiently. Additionally, all the services are interfaced with one another through the service interface unit. The application layer, in general, complies with the user's request for services.

3.4.3 Middleware-based Architectures

Middleware-based architecture is also popular and widely used in the IoT ecosystem nowadays; it is also known as a five-layer architecture. To ensure some of the performance factors like

reliability, scalability, and QoS in today's applications, the middleware-based architecture plays a vital role by creating applications efficiently and providing links among the applications, users, and data [246]. The five-layer architecture has been introduced due to the recent developments of IoT and its huge applications in business models. As shown in Fig. 3.4, the five layers include the middleware layer in between the network and application layer as well as the business layer at the top of the layer of the three-layer architecture [247]. The middleware layer aggregates, and filters data retrieved from the hardware, conducts data exploration, and offers access control to the objects for a variety of applications [248]. This layer has grown in importance in recent years since it allows for the growth of streamlined novel services and the incorporation of existing methods into new designs. The other layer of middleware-based architecture named the business layer organizes the entire services and activities of the IoT infrastructure. This layer takes data from the application layer and creates a business model, graphs, and other representations based on it [249]. This layer is in charge of designing, implementing, and monitoring all IoT components. In addition, the output of the previous four layers is compared with the desired output in this layer to enable the services and user's privacy [250].

3.5 IoT Protocols

Various standards have been introduced to assess the services and relevance that are utilized for IoT solutions to link several things to the Internet in IoT common standards [251], [252]. Although multiple protocols have been developed, they are not all required for a single IoT application at the same time. The IoT protocols for a given application are chosen considering the nature of the application. The most common IoT protocols (depicted in Fig. 3.5), which are utilized in a variety of applications, are listed below.

Constrained Application Protocol: The Internet Engineer Task Force (IETF), Constrained RESTful Environments (CoRE) research team developed the CoAP, which is a HTTP functional and lightweight application layer protocol [253]. As most IoT devices have limited power and storage, the CoAP protocol extends the functionalities of HTTP (which has a relatively high complexity) by fulfilling the needs of IoT devices [254]. This protocol shows how to build a web transfer protocol called REST on the upper level of HTTP. The CoAP uses the User Datagram Protocol (UDP) because it is simple in nature and has a small message size and layout, which helps to decrease the need for bandwidth, utilize resources, and decrease the overhead of TCP handshaking before data transfer [255]. This protocol has two sub-layers such as the messaging sub-layer and the request/response sub-layer. The first sub-layer (messaging) determines the replications and ensures efficient data transmission over the UDP through exponential backoff as UDP is constrained to the error recovery technique. On the contrary, the REST communica-

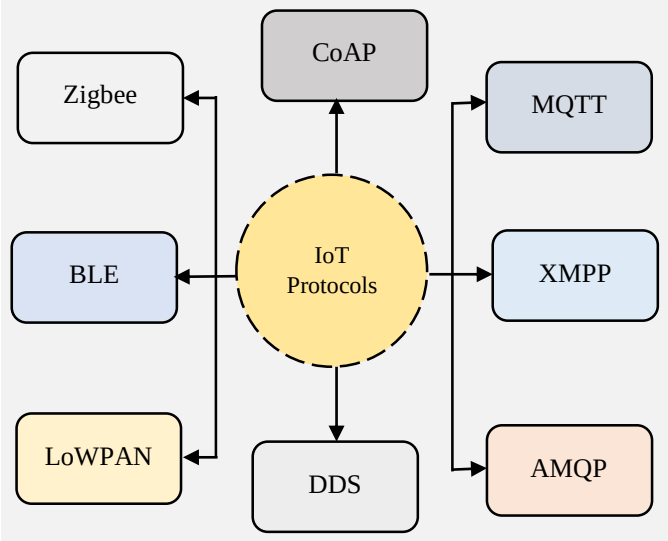


Figure 3.5: Most common emerging IoT communication protocols that are extensively used to develop smart IoT applications.

tions are handled by the request/response sub-layer. There are four kinds of messages in CoAP such as confirmable, non-confirmable, reset, and acknowledgment. The CoAP enables efficient delivery, congestion control, and flow control for IoT applications in resource-constrained and unsynchronized objects [256]. CoAP has several drawbacks, including increased communication delay, packet delivery instability, and the inability to transfer complicated data [257].

Message Queuing Telemetry Transport: MQTT is a messaging transport protocol that performs data aggregation of the environmental data and sends it to a web server [258]. This protocol is based on the TCP subscribe and publish messaging model and is intended for lightweight Machine-to-Machine (M2M), server-to-server, and machine-to-server interactions [259]. Here, the clients act as publishers/subscribers and the server is as a broker where the clients are connected to the server through TCP. Generally, the subscriber registers for a particular task on a device, and the data are generated and transferred to subscribers by the publishers through brokers [260]. MQTT is appropriate for utilization in things with limited resources, like those with low power and computing capabilities, and are connected to low bandwidth or unstable networks. However, the MQTT protocol is not suited for usage with all forms of IoT applications because it operates over TCP, and the overhead is raised because it uses topic names as texts [261].

Extensible Messaging and Presence Protocol: XMPP is a protocol that ensures low bandwidth communication and short message transfer, making it ideal for video conferencing, publish-

subscribe systems, telepresence, multi-party chatting, and talking in IoT [262]. For instant messaging applications, XMPP is appropriate for authentication, security measures, access control, hop-by-hop and end-to-end encryption, and interoperability with various protocols. This protocol serves three functions: client, server, and gateway, and it facilitates two-way communications between any two of these roles [263]. In this scenario, the client connects to the server through TCP protocol and transfers data using the XML streaming standard; the server is in charge of the connection management and routing of the message, and the gateway ensures reliable connectivity among distributed systems. This protocol allows communication among a variety of applications as it is flexible and simple. However, XMPP requires high computing capabilities devices, consumes bandwidth of the network, transfers simple types of data, and is unable to provide QoS [264].

Advanced Message Queuing Protocol: AMQP is an open platform messaging standard which is utilized at the application level to provide message services such as privacy, queuing, durability, and routing [265]. AMQP ensures reliable and consistent information exchange by using message passing primitives such as one-to-one, one-to-many, and exactly-once delivery. This protocol necessarily needs a stable transport protocol architecture and middleware serves as a gateway between applications and available resources, connecting institutions and mechanisms throughout time and space. The message queue and the exchange queue are the two main steps in the AMQP data transmission process. In a message queue paradigm, messages are kept until they are delivered to the recipient. The messages are transmitted in an appropriate sequence in another scenario (exchange queue model) [266]. AMQP also enables the publish/subscribe communication architecture in addition to point-to-point data transfer. There are two kinds of messages found in AMQP, bare messages provided by the sender and annotated messages available to the recipient. However, AMQP requires comparatively higher bandwidth, and it does not guarantee resource discovery [267].

Data Distribution Service: DDS is a data distribution service protocol based on publish / subscribe model that allows highly dependable M2M, real-time, flexible, and accessible connections [268]. DDS retrieves the edge outliers and gives notifications as well as pushes them to the data analytics engine. The distributed relational information model, which uses a distributed server to keep the application's specific data, is generated/consumed by the publishers/subscribers. The powerful QoS architecture of DDS is worked based on a series of two principles that provide end-to-end QoS control [269]. DDS allows asynchronous data transfer via a data bus through Peer-to-Peer (P2P) and distribution communication, making it a good fit for IoT applications [270]. As this protocol supports multicasting and broker-less models, it is possible to obtain high reliability and QoS through the use of this protocol.

Low Power Wireless Personal Area Networks: LoWPAN are made up of variety of cost-effective devices that are linked via wireless communication. This protocol has a huge number

of applications in IoT architectures due to its small packet sizes, low computing power, low data throughput, and low latency [271]. Additionally, the 6LoWPAN protocol was introduced by incorporating the most recent release of the IPv6 and LoWPAN. 6LoWPAN makes it easier to maintain the administration process by allowing each constrained object to be accessed independently within the network. Additionally, it is in charge of segmenting and reorganizing IPv6 traffic, guaranteeing unitary routing, reducing protocol stack headers, and providing compliance with the higher levels [272]. This protocol eliminates overall packet overhead as it does not include the extra header information during routing. Besides, 6LoWPAN contains a mesh address header to enable packet routing in a mesh architecture, but it is unable to provide detailed information on routing to the link layer. 6LoWPAN has various advantages, including ad-hoc self-organization, robust connectivity and standard compatibility, and low power consumption.

Narrowband Internet of Things (NB-IoT) [273], [274] is a Low-Power Wide-Area Network (LPWAN) radio technology standard introduced by the 3rd Generation Partnership Project (3GPP) for mobile devices and services. It can be considered as a novel air interface despite being incorporated into the Long-Term Evolution (LTE) standard. This protocol has been made as straightforward as possible to reduce device costs and battery usage. It eliminates the numerous functionalities of LTE including carrier aggregation, handover, measures to check the channel quality, and dual connection [275]. The technical details of NB-IoT [276] are as follows. The NB-IoT protocol has a 200 kbps downlink and 20 kbps uplink data rate capabilities. In this protocol, each message may have a payload up to 1600 bytes in size. With NB-IoT, each cell may link up to 100 K endpoints. It can cover the smallest areas and has a range of only about 10 kilometers. The devices with an ultra-narrow bandwidth of 200 kHz are supported by NB-IoT.

The 5G mobile network uses the new Radio Access Technology (RAT) known as 5G New Radio (NR) [277], which was developed by the 3GPP. The ultra-lean architecture [278] of 5G NR reduces always-on transmissions, which is a key component to extend the lifespan of IoT devices. This reduces interference and increases energy efficiency. 5G NR is beam-centric by broadening beamforming and multi-antenna methods from data transfer to control-plane procedures and early entry [279]. As 5G NR is ready for its future expansion in application use cases and technology, it also guarantees forward compatibility. Moreover, it can ensure low-latency needs due to the addition of mini-slots. In general, 5G NR employs frequency bands in two frequency ranges such as frequency range 1 (410 MHz–7125 MHz), and frequency range 2 (24250 MHz–71000 MHz) [280].

Bluetooth Low Energy: BLE, an extended version of Bluetooth offers a small radio with reduced power consumption to operate for a longer period for controlling and monitoring applications [281]. The protocol stack that is utilized in BLE is almost identical to the standard of conventional Bluetooth technology, but it has a larger coverage; approximately 100 meters with low latency [282]. The devices that employ the BLE standard are categorized into two groups:

master and slave. The master devices are the ones that play the most important roles and link to slaves. Additionally, the slaves can access and subscribe to several master devices. BLE enables devices to investigate as masters or slaves channels in star topology [283]. This technology turns off the radio while idle time and only turns it on to broadcast or receive minimal data packets, resulting in minimal energy consumption. There is a gateway required (another BLE device with network connectivity) for BLE devices while transferring data over the Internet.

Zigbee: Zigbee is a communication standard that ensures reliable, low power, and cost-effective data transfer, but it covers a small range of communication [284]. Zigbee supports star, cluster-tree, and P2P network topologies. A controller, in general, is in charge of the structure and can be found in the middle of a star network, at the root of a tree or cluster architecture, or anywhere else in a P2P topology [285]. There are two stacks in Zigbee standard such as Zigbee and Zigbee Pro; these stacks allow mesh network architectures to operate with a variety of applications, allowing for low storage and processing power implementations. Zigbee Pro adds new functionalities like a symmetric key to ensure security, effective many-to-one relaying for enhanced performance, and probabilistic address allocation for flexibility. This protocol offers

Table 3.3: IoT communication protocols with some fundamental features.

Protocol Name	Network Standard	Transport Protocol	Messaging Model	Network	Security and QoS	Applications
CoAP	IETF, Eclipse Foundation	UDP	Request/response	Tree	Both	Utility field
MQTT	OASIS, Eclipse Foundations	TCP	Publish/subscribe, request/response	Tree	Both	IoT messaging
XMPP	(RFC 3920-RFC 3923) RFC 4622, RFC 4854, RFC 4979, RFC 6122	TCP	Publish/subscribe, request/response	Client-server	Security	Remote management
AMQP	OASIS, ISO/IEC	TCP	Publish/subscribe	P2P	Both	Enterprise integration
DDS	OMG	TCP/UDP	Publish/subscribe, request/response	Bus	QoS	Military
LoWPAN	IEEE 802.15.4	TCP	Publish/subscribe	Star, mesh	Security	Structural monitoring
BLE	802.15.1	TCP/UDP	Publish/subscribe, request/response	Star	Security	Wearable devices
Zigbee	IEEE 802.15.4	UDP	Publish/subscribe, request/response	Star, mesh, hybrid	Security	Consumer electronics

stringent security and listening mechanisms, making it quite appropriate compared to other wireless communication standards [286]. The Zigbee protocol has a wide range of control and monitoring applications, including smart healthcare, smart homes, industry automation, lighting, and commercial control.

Table 3.3 describes the IoT communication protocols with some basic features such as network standard, latest version, used transport protocol, used messaging model, utilized architecture, the security and QoS measures, as well as potential application levels for the protocols that are necessary to compare each of the protocols.

3.6 Smart Healthcare Applications

Considering the state of the world today and the disperse of outbreaks and communicable diseases including COVID-19, as well as other things relevant to this global epidemic like massive cost, long distance, and the need for isolation during this crucial timeframe, heading to healthcare institutions is challenging, even sometimes become impossible particularly for the disabled and

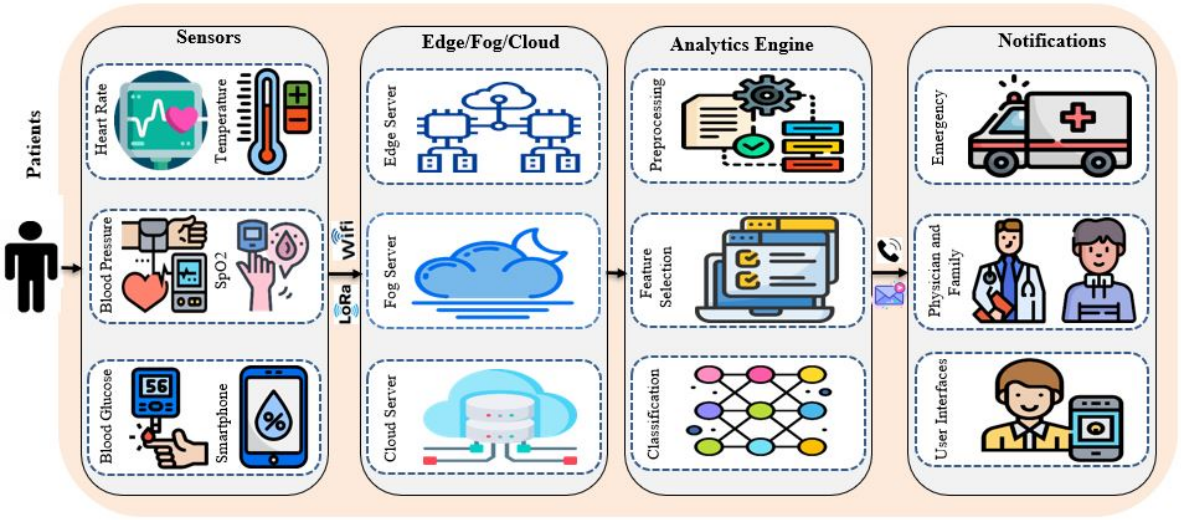


Figure 3.6: General pipeline of IoT-based smart healthcare systems. The sensors (heart rate, temperature, blood pressure, SpO_2, glucose, and smartphone) collect the data from the real-time healthcare environment. Some of the data are processed in an edge/fog server and the major data analysis occurs in a cloud server where the data analytics engines predict the patient status. The emergencies, families, and physicians receive the patient's condition and take further required actions.

elderly, who experience severe from at least one chronic illness [287], [288]. As a result, efficient, substantial, and computer-aided technology is very important to address the necessities for long-term care and RHM to ensure patients with an adequate quality of life while also reducing the economic strain [289], [290]. IoT has become a revolution for healthcare applications, allowing for the development of predictive and smart systems by linking various IoT devices to perceive real-time biophysical information from patients including HR, BP, glucose level, body temperature, and SpO_2 level [291], [292]. The most significant contributions of IoT in the healthcare industry are remote patient monitoring, disease management, preventative care, and assisted living [293], [294]. Additionally, the application environment of IoT for smart healthcare includes hospital management, home healthcare, and e-healthcare/mobile healthcare [295]. A general pipeline of IoT-based smart healthcare systems with the basic components is shown in Fig. 3.6. Various sensors/medical devices are utilized to collect data from the patient end through the help of healthcare professionals or the patients. The data collection step for healthcare is done in the perception layer and the retrieved data are sent through the network layer using Wi-Fi or LoRa communication technologies. Some portions of the collected data are processed in the edge/fog server to decrease the burden of the cloud server. The main data analytics including preprocessing, feature selection, and classification are done through predictive analytics engines with the use of AI techniques; resided in the cloud server. All the processes for the data analysis are performed in the middleware layer. The outcomes of the analytics engines are sent as notifications through call/short messages to emergencies, physicians and families, as well as interactive user interfaces. The application layer and business layer performed the information-consuming step. Although IoT has diverse applications in the healthcare domain, some of the most recent and promising works are described here. The major applications that IoT includes in healthcare domain are depicted Fig. 3.7.

3.6.1 Biophysical Parameter Monitoring

The monitoring of biophysical parameters is mainly used to observe the changes or absence of changes in the health findings of a patient. The goal of health monitoring is to provide early notification of threatening deterioration by achieving an optimal compromise that included many medical, engineering, and financial design factors. IoT plays a significant role to monitor patients remotely and condensing the burden of patients in the hospital as well as decreasing the number of cases. The major biophysical symptoms of a particular patient involve ECG, BP, temperature, glucose level, and SpO_2. The frameworks that are developed to monitor the above-mentioned parameters of the patients in the IoT environment are described below.

ECG Monitoring: Wu et al. [296] introduced a tiny-sized and low-powered ECG monitoring framework that was incorporated into a t-shirt. To collect the ECG signal, the system

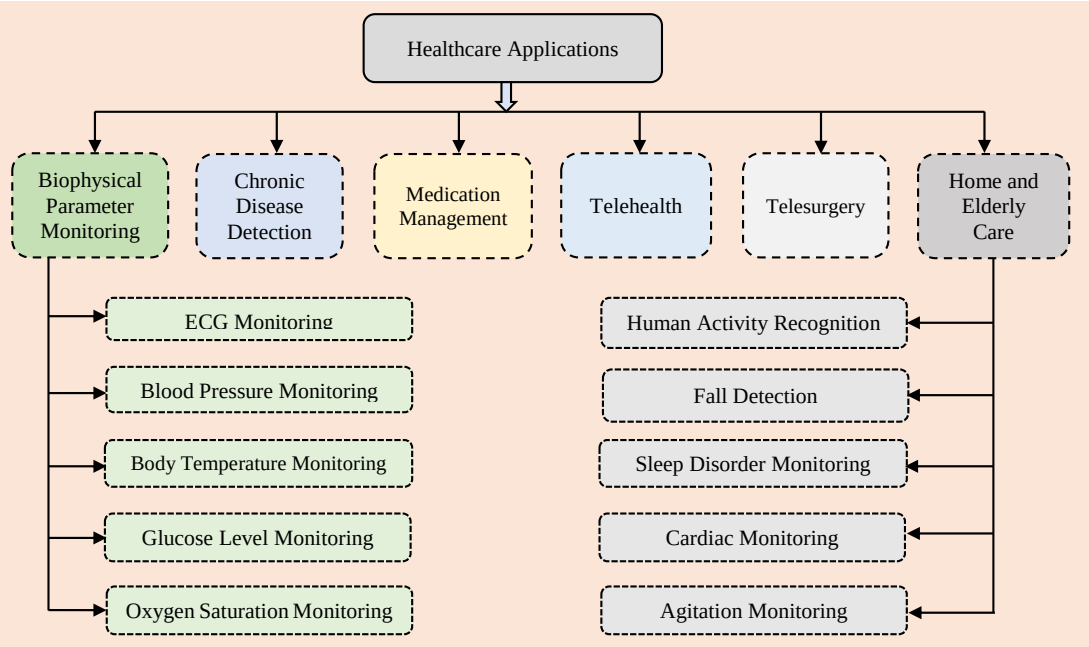


Figure 3.7: Major healthcare applications focusing on IoT concepts in the area of biophysical parameter monitoring, chronic disease detection, medication management, telehealth, telesurgery, as well as home and elderly care.

used biopotential chips. The retrieved data from the patients are transferred to end-users (physicians/families) using Bluetooth. A smartphone application has been developed in this system to visualize the collected data. The minimum power required for the developed system to operate is 5.2 mW. The system is no longer in real-time mode; it is possible to monitor the ECG in real-time by utilizing big data analytics to accommodate larger data capacity. Shaown et al. [297] developed an ECG monitoring system in IoT for ensuring smart health. The developed system used Raspberry Pi, Arduino, ECG sensor (AD8232), and bio pad as hardware components. The collected data are sent to a cloud server for processing through a Wi-Fi network. The emergency cases of the individuals are automatically sent to doctors to take fast initiatives in a form of an email. The overall accuracy found from the developed system is approximately 80%.

In another study, Djelouat et al. [298] presented a wearable ECG monitoring platform that resolved the problems related to power consumption. The device featured a novel design called compressive sensing, which can reduce power consumption and improve ECG monitoring efficiency. It is shown from the experimental results that the multicore processors are quite appropriate for a gateway-centric solution while maintaining processing capacity and power efficiency.

Sahu et al. [299] proposed a cloud-based remote ECG monitoring scheme for cardiovascular patients. The ECG data are transmitted to AWS through a mobile gateway service. BLE is used as a communication medium between the users and end devices. For data visualization, fast response, and long-live connection, AWS used HTTP and MQTT protocols in this system. Further, Basu et al. [300] introduced a platform for remote patients to monitor their cardiac conditions using the concept of IoT technology. The system used an ECG sensor (AD8232); interfaced with a notch filter (removes the power-line frequencies), and Arduino to process the data initially. The final processing is done using the MATLAB tool to remove baseline wander and the processed data are transferred to the cloud as a decision via Wi-Fi that can be accessed by the registered users.

Blood Pressure Monitoring: The measurement of BP is one of the most important aspects of diagnosing any disorders in the human body. In general, trained professionals measure BP while recording the data of humans. Nowadays, IoT and several sensing modalities are used to measure BP automatically and efficiently. Numerous works [301], [302] have been conducted to make the BP measurement process automated in recent years.

The authors of [303] developed a wearable device to measure systolic and diastolic pressure accurately. The developed system used light-based sensors to collect data and the recorded data are stored in the cloud server to make further analysis. The system was tested with targeted users (more than 70 individuals) and the performance was validated. Singh et al. [304] demonstrated a BP monitoring framework using the concept of IoT, Arduino, and BP sensors. Communication technologies including Bluetooth and Wi-Fi are utilized to transmit data from the patients to the cloud database. The developed system is flexible, cost-effective, compact, and provides the facilities to monitor BP from the home environment. In another study, Adi and Kitagawa [305] developed an IoT platform for BP monitoring using a tiny embedded device. The used components in this prototype are the pressure sensor, Zigbee, Radio Frequency (RF) module, and Raspberry Pi 3. The purpose of the Zigbee module is to send BP data to a database, which is then transferred to an IoT network via Zigbee end-device connection to the Zigbee coordinator. However, the developed system is still working in localhost; not in real-time. Hashim et al. [306] introduced a wearable system for BP monitoring in an IoT environment. A BP sensor is connected to Raspberry Pi to collect data from the patients. The system is tested with various positions of the patients such as sitting, standing, and lying where the BP value of the sitting position is more correct. The readings of the BP sensor are sent to registered Telegram as well as Gmail applications for patient monitoring. Furthermore, the authors of [307] presented the use of ML techniques to classify BP in wireless sensor networks environment that can provide BP status and abnormalities in real-time. In this prototype, a pressure sensor (MPS20N0040D-D) is used for data collection, a microcontroller (ATmega328) is utilized for data processing, and the data transfer is done through a transceiver (NRF24L01). Among the used ML techniques, the decision tree algorithm achieved the highest accuracy (97.9%) for BP status detection. Finally, all the data

of BP status are visualized in IBM Watson Cloud.

Body Temperature Monitoring: The body temperature of individuals is a leading determinant of homeostasis maintenance and is used in a variety of diagnostic procedures. In many disorders, maintaining records of temperature changes might help doctors to draw assumptions about a patient's health. The traditional way of temperature measurement is the use of a thermometer that takes the records from the different positions of the patient's body such as the ear, mouth, or rectum. However, the traditional techniques for body temperature measurement are low comfortable, less flexible, and less accurate. In recent days, several frameworks [292], [308] have been developed thanks to IoT technology for human body temperature measurement.

Ota et al. [309] developed a wearable prototype for body temperature measurement that is printed using a 3D printer and is applicable to wear on the ear. The developed device records data from the tympanic membrane utilizing an infrared sensor. In this system, a wireless sensor network and data processing unit are utilized where the prototype is integrated. The environmental conditions and other physical activities could not interfere with the readings of the body temperature. Zakaria et al. [310] proposed a lightweight device for real-time temperature measurement targeting infant patients. The major components of this system include a wearable sensor (LM-35), microcontroller (Arduino), ESP8266, and power module. The collected body temperature is visualized in the cloud server through ThingSpeak and Blynk applications. The developed system can provide alert messages to the mediators to take immediate actions when there is a high change of temperature. In another work, Rahimoon et al. [311] introduced a contactless body temperature measurement system to monitor patients remotely. Two temperature sensors named LM-35 and MLX-90614, Arduino, ESP Wi-Fi shield are used in this prototype. The collected data are transferred to the cloud server for remote access utilizing a Wi-Fi connection. However, the developed prototype is quite accurate, real-time, as well as reliable. Boonsong et al. [312] demonstrated a contactless body temperature monitoring platform in an IoT environment. The infrared temperature sensor MLX90614 DCI is used here to perceive the data from the environment and send it to the host computer as well as the cloud server using a 2.4 GHz microwave frequency. The experimental findings revealed that the developed framework obtained a reliability calibration of 74.7%. Further, Fang et al. [313] used a wearable smart bracelet to monitor the temperature for contact tracing applications during the time of the pandemic. The bracelet used RF technology for data collection and the retrieved data are sent to the cloud database via wireless technology. However, the system is cost-effective, easy to operate, and provides better data privacy.

Glucose Level Monitoring: Glucose level monitoring is one of the most significant aspects of managing type 1 or type 2 diabetes. Diabetes is a chronic disease wherein the glucose levels in the body keep going up for an extended time. Fingerpicking is the most common method for detecting diabetes, which is validated by blood glucose levels. The development of numerous wearable

devices [314], [315] for blood glucose monitoring using modern IoT technologies ensures some features such as comfort, safety, non-invasiveness, and convenience.

Gia et al. [316] introduced an architecture for real-time blood glucose level monitoring in an IoT environment. The proposed scheme was designed as an energy-efficient sensor module using nRF communication protocol and developed an energy harvesting unit to extend the battery life of the sensor device. There is a push-button service in the developed prototype that notifies the mediators in the case of abnormal behaviors of glucose like too high or too low. Rahmat et al. [317] proposed a prototype named GluQo to monitor the glucose level efficiently. The data from the fingertip are collected using near-infrared LED to determine blood glucose optically and the intensity of the retrieved light calculated the concentration of the glucose in the blood. The glucose level was sent to the mobile application through a Wi-Fi signal from the microcontroller. It is found from the experimental findings that the percentage error of the developed prototype is 7.20%. In another research, Alarcón-Paredes et al. [318] developed a prototype in the shape of a glove that can measure a patient's glucose level and is equipped with a Raspberry Pi camera and visible laser beam. The developed framework used a set of images collected from the patients' fingertips to detect their diabetic status. For data analysis, an artificial neural network is used, and the obtained mean absolute error and Clarke grid error of this system are 10.37% and 90.32% respectively. However, the developed prototype is a little bit bulky to use. Valenzuela et al. [319] used a double moving average technique enabled by an IoT network to offer a framework for blood glucose monitoring to minimize potential problems in elderly adults. The data for the developed system are collected using a glucometer and this sensor is connected with NodeMCU to make IoT environments. The system can notify doctors and families about a potentially critical state in patients at an early stage. Further, the authors of [320] used near-infrared sensors to demonstrate a portable non-invasive glucose monitoring system. The data for the experiments are collected from the fingertips of the patients and the analyzed data are displayed on the Android application. The obtained percentage difference of the reliability test of the developed prototype is less than 25% and the approximate cost is $15.

Oxygen Saturation Monitoring: The measurement of SpO_2 is a crucial thing in the management and interpretation of the healthcare system. Pulse oximetry is the process of measuring SpO_2 in a non-invasive way, and it is used as a significant factor in the patient care system. The non-invasive techniques for SpO_2 monitoring solve the problems associated with traditional approaches and ensure real-time monitoring. The integration of IoT-based technologies [321], [322] with pulse oximetry has shown significant potential in the medical field.

Son et al. [323] proposed a SpO_2 monitoring wearable tool that uses IoT technology to give real-time monitoring and prediction of patient health states. The data from the sensor are transferred to a cloud database through the Internet and this data can be accessed via a web server for the registered users. The experimental findings found that the highest deviation of SpO_2

measurement for this device is approximately 4.4%. It is evident that the prototype is complete, cost-effective, and has low-power consumption. Adiputra et al. [324] presented a small, wearable, and cost-effective SpO_2 device in an IoT environment for blood oxygen monitoring. To create an IoT environment, the developed prototype is connected to NodeMCU which functions as a data processor and Internet network gateway. A dedicated website is developed to ensure online and real-time data accessibility from the cloud. The error rate for SpO_2 monitoring is quite low (1.5% only). In another research, Utomo et al. [325] demonstrated a scheme for SpO_2 monitoring using oximeter sensor data through the concept of IoT technology. The data are collected from the fingers of the patients in different body positions like standing, sitting, and lying. The sensor is connected to the ESP8266 microcontroller, and the collected data are transferred via the built-in Wi-Fi module to the cloud server where the data visualization is done using the ThingSpeak application. The experimental findings revealed that the uncertain measure of the SpO_2 is 0.04. The authors of [326] demonstrated a wearable prototype for blood SpO_2 monitoring from PPG signal using the concept of BLE. Here, the system introduced a novel adaptive cancellation algorithm using adaptive filtering to remove motion-induced noise. The proposed architecture captured the blood oxygen signals as soon as it received the instruction and sent the data to the mobile phone gateway via Wi-Fi, where it displayed the mathematical waveform. Furthermore, Zhang et al. [327] proposed a new wearable tool for blood SpO_2 monitoring that sensed two-channel PPG signals and sent the signals to the smartphone wirelessly. The dataset for this work is retrieved from the users (eleven individuals) to measure the performance of the proposed framework. It is evident from the experiments that the root mean square error for SpO_2 estimation is 1.8%. However, the performance of the proposed framework is calculated using the data from healthy patients.

3.6.2 Chronic Disease Detection

Chronic disease is a patient health issue whose consequences are continuous or otherwise lengthy, or an illness that develops over time. Many individuals around the world suffer from chronic illnesses including HD, diabetes, asthma, and cancer. In most cases, these illnesses lead to mental health problems, which is one of the most common adverse events of chronic disease. The recent revolution of IoT contributes a lot to develop healthcare frameworks [328], [329] for people infected with chronic diseases.

Shah et al. [330] proposed a health parameter monitoring scheme for asthma patients using the concept of IoT and cloud technology. The system utilized a smart sensor to record the respiratory rate of the patients. A watermarking technique ensures data security and authenticity whenever transferred from the patient's side to the cloud as well as the cloud to physicians. A dedicated cloud server is used here to keep the record of the patients and it provides access to mediators for

diagnostic and monitoring purposes. Rehman et al. [331] developed an IoT cancer monitoring framework based on a multi-layered architecture that can diagnose and monitor cancer patients remotely. Here, three-layered architectures are utilized which are the patient layer, connectivity layer, and medical layer. The prototype is in the shape of a wristband and can monitor the basic health issues of cancer patients. The major advantage of this architecture includes that it gives continuous monitoring of the patients that is helpful for the physicians to generate a complete history of the patients which in turn would be helpful for treatment in the later stages. Shihab et al. [332] developed a system for HD detection using the concept of Recurrent Neural Network (RNN) in an IoT environment. In this system, an ECG sensor, Arduino, and biomedical patch are used for data collection and to perform an experimental study. The parameters from ECG recordings are retrieved and fed into RNNs with recurrent layers as input. The major advantage of this developed system is that it is quite beneficial for doctors to be able to prescribe medication to patients without having to be physically there. The authors of [333] proposed an IoT-based novel HD diagnosis system using ML techniques. In the preprocessing step, the noises from the raw ECG signals are removed efficiently ensuring a more accurate diagnosis. As hardware components for ensuring the IoT platform, the system is comprised of Arduino, ECG sensor, pulse sensor, and temperature sensor. The decision tree algorithm obtained comparatively better performances for HD diagnosis using the benchmark dataset.

Hebbale et al. [334] presented a non-invasive self-care framework that monitors the blood sugar of diabetes patients using ML techniques and IoT. In this system, predictive analytics engines reside on the cloud, and they are used for data analysis for diabetes detection. A smartphone application facilitated physicians as well as patients to monitor their blood sugar and related risk easily. The decisions from the ML approaches are forwarded to the doctors for further analysis and the necessary recommendations from the doctors are sent to the parents through the developed mobile application. The SVM obtained the highest accuracy of 82.41% from the experiment. The authors of [335] described an IoT-based diabetes prediction scheme using ML approaches. The data for this system are collected using an IoT diabetes sensor. The dataset for the experiment is divided into training and testing sets in the percentage of 75% (11,250 records) and 25% (3,750 records) respectively. It is found that the achieved accuracy and error rate of the developed system are 94.23% and 5.77%, respectively. The author of [336] demonstrated an architecture to predict asthma using a decision tree algorithm in an IoT environment. The major components of the developed system are Arduino, ESP8266, GP2Y1010AU0F Optical Dust sensor, LM-35 sensor, and 'NO' Gas sensor. The prototype can notify patients when they need to take pain relievers. It is observed from the experiments that the decision tree algorithm achieved an accuracy of 97.7% for asthma prediction. However, the developed prototype is not fully integrated and bulky. In [337], the authors demonstrated various types of architectures and frameworks to develop IoT-based healthcare platforms targeting cancer patients. For communication purposes, a wireless sensor

network is developed that can accommodate a high number of sensors and the data analysis is done on the cloud server. The security measures and operational challenges regarding the IoT healthcare system for cancer patients are also described here.

3.6.3 Medication Management

Smart medication is one of the most significant concerns in the IoT-based smart healthcare application. It is very critical as it is intrinsically related to human health and missing a dose of medication can have serious effects. Medication nonadherence is particularly common in older individuals as they age and eventually increases the chance of diseases including dementia, and cognitive decline. Hence, it is badly needed for them to use assistive technologies to keep track of medication timing and dose. In recent times, several studies [338], [339] have been performed emphasizing on employing IoT technology to track patients' medication adherence.

Bharadwaj et al. [340] developed an IoT device in the shape of a medical box to remind patients of the time of their medication. In this prototype, there are three trays to keep the medicine for three different periods such as morning, noon, and evening. The data are stored on the cloud server and the smartphone application retrieved the data from the cloud and used it to remind the patients about medication. Tsai et al. [341] demonstrated an intelligent pillbox for the older individual and nursing home needs using the concept of electronic technology and network functionality. Using embedded sensors, the developed device detects the signal when patients take their medications and displays the schedules and deadlines as text, pattern, or voice on the LCD screen. The hardware unit is made up of webduino for data transmission and can interact with a web server via the IoT architecture. The sensed data are transferred to the Arduino module, which is then sent to the developed pillbox, allowing the elderly to maintain their medications. Srinivaset al. [342] introduced a home-based medicine box named iMedBox to manage medicine for patients in an IoT environment. The used sensors in this prototype notified the patients via a wireless connection and an Android application to take their medications on time and to keep in closer contact with their doctors. In another research, Latif et al. [343] presented a wearable IoT device called I-CARES to diagnose health issues and medication for elder patients. The system described an algorithm for data analysis and aided medical professionals to make decisions for early-stage treatment of predicted diseases or illnesses. Additionally, the developed prototype can predict emergencies and call emergency personnel when necessary. Vardhini et al. [344] proposed an assistive system to help elderly people with memory impairment for keeping track of their medication schedules and informing them to consider taking their medicines at the appropriate times. The main hardware components for the prototype include NodeMCU, LDR sensor, motor, switch, power supply (battery), LCD monitor, and buzzer. The developed system is low-cost (approximately $15), and easy to use. However, the system has not been tested

in any hospital environment. The authors of [345] designed a smart medication box for ensuring medication monitoring in the IoT platform. The prototype includes six distinct parts to keep track of six different pills, and it can send timely reminders to the caregivers/patients through the smartphone. A simple authentication process is developed to avoid overdosage and improper intake of medicines. In addition, the system can provide monitoring facilities for temperature and heartbeat through the use of biosensors.

3.6.4 Telehealth

Telehealth is the use of technologies such as digital devices, smartphone apps, and webpages to instruct, detect, and even prescribe a remedy to individuals suffering from health problems. It primarily allows individuals of all ages to be monitored remotely. The telehealth system monitors significant clinical data and takes oversight of the individual's well-being. Numerous frameworks [346], [347] have been done to ensure telehealth services using IoT technology in recent times.

Ganesh et al. [348] introduced a framework named 'AutoImpilo' to improve telemedicine and telehealth services in IoT networks. The designed approach ensured that all online medical checks were conducted in collaboration with consulting doctors or specialists. This system is quite beneficial for individuals who are unable to go outside due to age or quarantine issues. The system used a smart card in the authentication step to ensure privacy, but it is not linked to the cloud. In another work, Tsiouris et al. [349] integrated many online technologies and compatible components into a unified framework for digital coaching, encouragement, and empowerment of the elderly with balance problems. This approach has been proven in a telerehabilitation system for individuals with balance issues that uses augmented reality to provide a surrogate physiotherapist as a superimposed hologram, along with easy-to-use wearable sensors. Hewa et al. [350] employed many decentralized applications to transmit cryptocurrency to patients' wallets effectively and reliably as an incentive for providing their health information utilizing edge computing and blockchain technology. To provide scalability with a large number of connected biomedical sensors to the cloud, the system included memory offloading functionality into the blockchain. The proposed platform is tested in a near real-world setting employing the Hyperledger Fabric blockchain infrastructure and Raspberry Pi modules to imitate smart sensor activities. Rokonuzzaman et al. [351] developed a telehealth platform based on the IoT concept to create warnings automatically for providing individuals with desirable healthcare under constant health contexts. A smartphone application called 'Village Doctor App (VDA)' is deployed for real-time monitoring. This scheme provides monitoring facilities for HR and temperature. The hardware components of the developed prototype are HR sensor, temperature sensor, Arduino, NodeMCU, battery, and LCD display. The physicians can access the data from the cloud server through the VDA application. Further, Hamil et al. [352] described a telehealth platform for bio-signal diagnosis

and classification using e-health sensors data considering the IoT environment. The Arduino and Raspberry Pi with sensors are used for data acquisition and processing. The used sensors here are ECG sensors, body temperature sensors, SpO_2, position sensors, airflow sensors, and galvanic skin sensors. Various ML algorithms are tested with PhysioNet databases where the best accuracy of 99.56% is found from a CNN-SVM for atrial fibrillation classification.

3.6.5 Telesurgery

In the age of 5G communication technology, telesurgery offers enormous promise for bringing surgical healthcare services to distant locations via high-speed data transfer using a wireless communication channel. The ability to diagnose patients more accurately and precisely, even from remote locations, is a major benefit of telesurgery services to society [353]. Telesurgery enables the connection of patients and surgeons utilizing a wireless network and robotic device. It functions as a master-slave system, where the master domain employs a Human System Interface (HSI) to command the surgical robot at the operating site utilizing a variety of control instructions [354]. The HSI includes audio-visual headphones, a video console, and haptic devices. Telesurgery can be carried out at the patient's location by a doctor at the master site using a surgical robot. The major components of the surgical robot are a 3D camera with good resolution, a top-notch microphone, and trustworthy touch sensors. In a telesurgery operation, a surgeon can perform surgery without directly touching the patient by controlling a robotic arm at the patient's site via HSI, and monitoring the progress in real-time on a master domain console display [355].

Shabana and Velmathi [356] developed an IoT platform for telesurgery that used vision-oriented VR goggles and hand motion acquisition devices accessed by the surgeon. In this system, the sensor data is put into a microcontroller, which transmits it in real-time, without interruption or data loss, to two connected Zigbee systems. The IoT network collected the data from MEMS sensors and transferred the processed data to the surgeon's site. However, the quality of live video streaming of this system is not quite good. Miao et al. [357] demonstrated a telesurgery framework employing 5G tactile Internet where the surgeon and the master control site are placed on one side, and the patient and his controllable desk are located at the edge of the network. The fog computing terminal enabled a portion of the network to support 5G customers' communication demands based on software-defined network and network functions virtualization approaches. Moreover, the system employed AI algorithms to predict the physician's following actions specific to the patient. Fekri et al. [358] presented a novel technique to improve security in telesurgery frameworks by exploiting DL algorithms. The DL algorithms are trained by the instructions from the surgeons, which are then utilized to classify the commands given by a surgeon or other users. The developed system made it possible to secure telesurgery procedures against portable security intrusions. Gupta et al. [359] introduced a telesurgery framework called HaBiTs that exploited a

private blockchain network. The personal health records of the patients are secured through the blockchain architecture as it allowed only authenticated users. By allocating one block for each surgeon, the problem with the existing system is resolved and all information about the surgeries performed by the surgeons is saved in this approach. However, the end-to-end delay of this platform is relatively high which is not appropriate for complicated surgeries. In another work, Iqbal et al. [360] proposed a scheme called SecureSurgiNET to provide security and reliability for telesurgery systems. The system used several full-proof trustworthy communication protocols along with telesurgical authority, database server, and identity management server to ensure security. However, the developed platform is insufficient due to the lack of security, privacy, and reliability in some critical patient information.

3.6.6 Home and Elderly Care

Elderly people can continue living at home freely with the assistance of smart home care services, as IoT devices provide them with the security of being observed and the capability to instantly call for help in the case of an emergency. Modern IoT technologies, as well as relevant wearables devices, are frequently utilized in AAL environments to monitor older patients who are unable to move rapidly and need a longer time to reach clinics for routine or urgent medical facilities [361], [362].

Abdelgawad et al. [363] proposed a system for health monitoring of older people in an AAL environment where the data are retrieved from several sensors and transferred to the cloud server for further investigation. The prototype incorporates six different sensors, Raspberry Pi, a Wi-Fi module, a cloud server, sensor interface circuits, and an indoor positioning module enabled by BLE. A feedback signal is sent to the users based on the behavior of the data so that they can take necessary actions in the case of emergencies. Debauchea et al. [364] demonstrated a novel architecture for elderly psychological signals monitoring utilizing the concept of IoT and fog computing. The framework includes a wireless sensor network consisting of various psychological and environmental signals, a local gateway for local and fast storage, and Lamda cloud server for data processing as well as storage. The main advantage of this architecture is that it can provide local gateway monitoring of new and current patient data. The latency and CPU usage of the fog architecture is 7 ms and 7.86% respectively. In another work, the authors of [365] presented a novel and scalable platform for smart home and elderly care based on the IoT concept. The system employed various sensors' data such as heartbeat, pressure, and gas sensor for patients as well as smart home monitoring through the use of Raspberry Pi and a cloud server. The medications are suggested by the registered doctors based on the health conditions of the patients which are delivered through wireless sensor networks in the form of texts or email. The experimental findings revealed that the mobility absorption of the designed network is approximately

90%. However, a few numbers of signals are considered here for health and home monitoring. Srinivasan et al. [366] used the concept of IoT and ML to develop an elder care system. The prototype of this framework contains sensors connected to an ESP32 and a home module with an IoT camera attached to a Raspberry Pi. The biophysical parameters are retrieved through sensors and processed in the AWS cloud utilizing ML approaches. The warnings are directed to caregivers automatically in the case of abnormalities. The locations are tracked through a GPS module and transferred to the mediators to let them know the location of the individual patient. All the data for the experiments are visualized through the developed smartphone application. Furthermore, the authors of [367] developed a wearable framework for health data monitoring (stress, BP, and position of the users) for the elderly in smart home environments. The stress of the elders is detected using electrodermal activity, PPG, and skin temperature sensors. The BP is estimated utilizing the data from the PPG sensor. The location of the individuals is detected using a voice-assisted indoor location module. The highest accuracy of 94% is obtained from the combination of multiple signals for stress detection. Finally, a prototype is developed to test the location detection system in a smart home environment.

From another perspective, detection and localization systems are crucial parts of various application areas including elderly monitoring, intrusion detection, gesture recognition, and HAR. Vision systems are ideal for various detection and localization tasks. However, IoT-based vision systems can be a major privacy and security risk. Wireless sensing technologies can be a major replacement for vision systems as they do not collect any direct views of the environment-instead using the various properties of wireless signals for detection and localization [359]. Common wireless sensing techniques include RSSI based detection-which detects changes in wireless signal strength due to motions of physical objects, Channel State Information (CSI) based detection-which detects changes such as scattering, fading, and power decay of wireless signals due to motions of physical objects, Frequency Modulated Continuous Wave (FMCW) based detection-which uses a specialized radar to detect motion via frequency difference of transmitted and received echo signal from moving bodies, and Doppler Shift based detection-which uses a specialized radar to detect movement using the frequency changes of wireless signals due to relative motion [360].

Chuma et al. [361] proposed an IoT system that utilized a continuous-wave Doppler radar sensor. The sensor captured the data associated with the unique time-varying characteristics associated with different body motions. The captured data is then classified into appropriate classes using a pre-trained CNN model. The system achieved high accuracy for five action classes: falling, jumping, stopping, crouching, and walking. Abdu et al. [362] proposed an FMCW radar-based elderly monitoring system. The system extracts features from the radar data using pre-trained AlexNet and VGG models. Useful features are extracted from the outputs using a channel attention module and a canonical correlation analysis algorithm is used for further feature fusion. The

fused features are finally classified using a SVM classifier into the following classes: walking, sitting, standing, picking up objects, drinking water, and falling. The system achieved 99.77% test accuracy. Chi et al. [363] proposed a HAR system that utilized the coarse-grained RSSI measurements from radios of IoT devices. The coarse-grained measures are pervasively available. The authors designed and developed the system as a self-contained middleware. The system performed similar to more fine-grained CSI-based approaches and was able to distinguish among nine activities. The system achieved 90% accuracy when RSSI readings from three IoT devices were used.

In most modern systems, sensors perceive the environment around them and send the information to another location for processing and assessment. Modern systems often have multiple sensors. Collaborative sensing requires the individual sensors of a modern system to operate and collaborate as a collective [368]. Collaborative sensing has become possible due to the recent proliferation of small-size and affordable sensors in all appliances around us. Network-wide collaborative sensing optimizes the sensing elements to gather the required information from regions of interest. This enables real-time monitoring as well as reliable control of the physical systems. Based on the dimension of the sensing targets, collaborative sensing can be divided into three general categories: point coverage, barrier coverage, and area coverage. There are two main strategies for applying collaborative sensing in an application: centralized and distributed. In centralized settings, a dominant sink node collects and processes global information. In distributed settings, all nodes are given the same priority and responsibilities [369]. Collaborative sensing has numerous applications such as improving smart cities using collaborative drones [370], surveillance of large areas for monitoring accidents or disasters [371], smart transportation [372], and smart building [373]. Collaborative systems employ distributed intelligence. The sensors are required to have the logic to opportunistically collaborate.

Belkacem et al. [374] developed a collaborative IoT system that collects patients' medical information through several distributed websites. The collected medical information includes sensing data from assisted living sensors and localization data. Patient profiles are then created based on the collected data. The Pearson correlation coefficient is utilized over a long period of time to compare the patient profiles with new patient data. This comparison helps in detecting emerging health risks and chronic diseases. This can also help in providing personalized recommendations. Jaleel et al. [375] proposed MeDIC-a framework that provides probing and translating agents at the network edge to facilitate collaborative sensing. The sensing devices connected via the MeDIC network can request data conversion services from one another when some nodes of the network are overloaded. This framework can decrease the overall number of cloud accesses of IoMT networks while improving response time-thus making the framework suitable for real-time medical applications. In another research, Nguyen et al. [376] proposed a way to significantly reduce power requirements for sensor movements and communication sys-

tems using a distributed compressive and collaborative sensing solution. The proposed solution builds a scalar field map for mobile sensor networks. Some sensors deployed in the mobile sensing area build the maps by themselves and collect sensory data from neighboring sensors based on compressive sensing algorithms. As the sensors recover all data based on a small number of compressive sensing measurements instead of collecting all readings from the networks, the overall power consumption of the network is reduced. Further, Djenouri et al. [377] proposed a collaborative sensing-based disease detection system. The detection system collects medical data about the patients from various IoMT and image sources. Intelligent agents trained using reinforcement learning then explore the interactions among the data sources to reach a reliable conclusion regarding the target disease. This collaborative sensing approach drastically improves the overall disease detection accuracy.

3.7 Discussions, Open Challenges, and Future Trends

In the reviewed IoT-based healthcare systems, we have described each of the systems by combining the information of IoT device capabilities, architectures, and protocols. Among the used processing devices, Arduino is used in most of the reviewed systems [300], [303], [332], [333], and [341] due to its lightweight and portability. Raspberry Pi is also exploited in some use cases [305], [306], [318], and [350] where the developed systems require more processing power, especially for image data processing. Only the system developed in [325] used ESP8266 as a processing device. The Raspberry Pi and ESP8266 are capable of providing real-time operation as they have a built-in Wi-Fi module. Some of the developed healthcare systems used both Arduino and ESP8266 [310], [311], [336], or Arduino and Raspberry Pi [297], [352] to carry out the real-time operation. Approximately, all the reviewed systems did not mention which OSs they used to develop the healthcare framework. In the case of built-in tools, the Google cloud server is used in most of the reviewed systems. The healthcare frameworks developed in [299], and [366] utilized AWS cloud servers. Besides, ThinkSpeak is exploited in [310], and [325] for patient status monitoring. In the case of IoT architectures, most of the reviewed systems did not mention which architectures they have utilized for developing the healthcare platform. Only the system introduced in [331] mentioned that they used three-layer architectures. Regarding the IoT protocols, a few of the protocols are frequently used in the reviewed healthcare systems. The framework presented in [299] exploited both MQTT and BLE protocols. In some other works [326], and [363], BLE has been utilized for short-range data communication. Moreover, the works introduced in [305], and [356] utilized Zigbee protocols for their intended purpose of application.

IoT and the technology and research state related to it are in their nascent state compared to other networking research topics. The SOTA connection models and architectures are recently

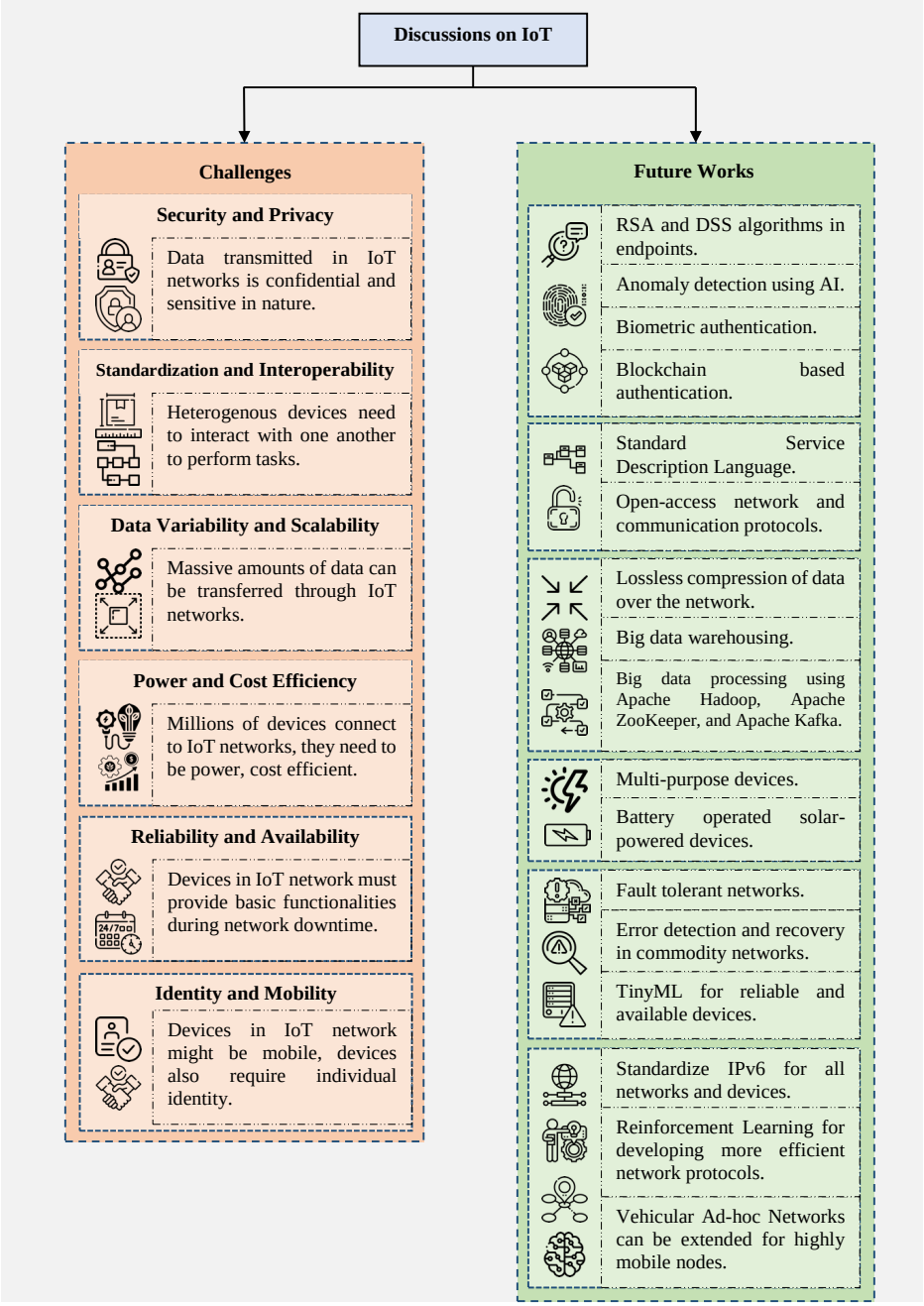


Figure 3.8: Challenges and future research directions of IoT. The major challenges encompass security, reliability, scalability, availability, and cost-efficiency. The potential future directions include the area of blockchain, TinyML, reinforcement learning, and big data processing.

being developed. IoT in general and IoT in healthcare in specific still face many challenges. Understanding and exploring these challenges such as the security and privacy risks, lack of standardization, big data management, and costs are imperative to realize the full potential of IoT in industrial, healthcare, and personal use cases among other promising domains. The major challenges and the potentials future directions for IoT research are illustrated in Fig. 3.8 in brief. The open issues, as well as future research trends related to IoT, are discussed as follows.

Security and Privacy in IoT: Security is a huge concern in any networked system. IoT is a collection of devices connected via a single network or multiple networks. Thus, the security issues concerning networked systems are also relevant to IoT systems [378], [379]. Devices in IoT are also heterogeneous in nature–which adds to the complexity of maintaining the security of the transmitted data. To provide complete security, all of the types of devices connected to IoT networks, would need to be completely secure–which is a near impossible task. Security can be breached on both hardware and software levels. Modern smartphones often have exotic zero-click exploits, meaning the devices can be hacked without any action by the user of the device. Modern OSs – be it for desktops or smartphones often have zero-day exploits, meaning the devices running these updated OSs can be exploited by hackers even on the first day of the update. Exploited devices can act as Byzantine nodes in IoT networks, making the entire network vulnerable and stealing data from other devices. Especially in healthcare systems, the underlying data is extremely sensitive and confidential. Healthcare data not only contain health-related information but also contains identifying information of the patients including social security numbers, social insurance number, and insurance information. Any leak or compromise in such networks might result in catastrophic ramifications for the patient such as identity theft and customized scamming. Improving the security and privacy of data in healthcare systems is a research challenge.

In the future, cryptosecurity algorithms such as the RSA encryption algorithm, Digital Signature Standard (DSS), and ElGamal can be utilized in all endpoints in IoT networks to provide security. Secure hash algorithms can be utilized to ensure data integrity. Blockchain technology can be leveraged to reduce man-in-the-middle and network eavesdropping attacks [380], [381]. Non-Fungible Token (NFT) technology can save patients from counterfeit products or services. NFTs can also help patients to decentralize their sensitive data and to share healthcare data while keeping their identities anonymous. Biometric authentication techniques such as fingerprints, voice recognition, face recognition, and iris recognition as well as software-based authentication such as two-factor authentication, access cards, and seed phrases should also be standardized in the future for systems connected to IoT networks. AI-based approaches for anomaly detection in networks should be explored in the future for improving security. Due to advances in TinyML [382], it might be possible to run anomaly detection models in the future in resource-constrained devices [383]. These would greatly improve the security and privacy of the users and the transmitted data.

Standardization in IoT: Heterogenous types of devices are interconnected via complicated networks in an IoT environment. Each of these devices might use distinct architectures or protocols for generating or transmitting data. The devices are also different from software and hardware perspectives. The management of such devices including the collaboration of the devices at a software or hardware level, proper addressing, marshalling, de-marshalling, optimization at a protocol or architectural level, and identification is a huge research challenge. Measures should be taken to draft and standardize IoT architecture and network protocols in the future [384].

Data Variability in IoT: Data in IoT can vary in type, size, and formation. Due to the sheer number of connected devices in IoT networks, storing the data in solutions such as in conventional RAID systems or conventional databases is very inefficient. Huge volumes of data are generated in a very short period in IoT networks. Conventional database systems are not capable of storing or efficiently retrieving them. Management of a huge volume of data such as handling spatial or temporal information, handling differing data sizes, and handling different types are massive research challenges. Devices in healthcare systems transmit data in various formats such as 2D image formats, 3D imaging formats, and numerical sensor readings in time-series format.

Big data warehousing and processing frameworks such as Apache Hadoop, Apache ZooKeeper, and Apache Kafka can be utilized to efficiently deal with the huge volume of data in the future [385], [386]. Some of these frameworks also provide error detection and error recovery of the stored data, which will help in persisting historical data. Some of these frameworks are also capable of dealing with multiple types of data at the same time. These frameworks are also highly cost and space-efficient. Lossless compression models can also be researched to transmit huge volumes of data over IoT networks [387].

Identity in IoT Networks: IoT networks generally consist of a huge number of devices. Every device needs a unique identity in the network. The most used core protocol in standard-based interconnecting networks is IPv4. The IPv4 protocol uses 32 bits to represent unique addresses which provide a maximum of 232 address spaces. IPv4 addresses are represented in 4 numbers separated by 4 dots. Due to the huge number of devices connected to IoT networks, it is possible to run out of address space. Healthcare systems have multiple monitoring systems per person. For a sufficiently large elderly population in a single network, the address space for all devices might be insufficient. While a lot of devices support IPv6 nowadays, developing more lightweight and robust solutions for addressing the identified problem is a research challenge.

In the future, IPv6 protocol can be standardized for all IoT networks [388], [389]. IPv6 addresses are represented as eight groups of numbers. Each group contains four hexadecimal digits. The address space of IPv6 is 2128 which is suitable for addressing the huge number of devices in IoT networks. IPv6 also supports shorthand representation and compression of network addresses, which reduces the total used bandwidth over the network.

Lack of Service Description Language: A service description language is a special form of Interface Description Language that is used for describing a particular service. Interface description languages enable programs to interact with one another regardless of origin or language. Similarly, service description language enables services from varying devices and software/hardware combinations to interact with one another. Due to a lack of standardization, devices in IoT networks use vastly different protocols, architectures, and data types. The underlying service routines dealing with these differences are also vastly different from one another. A unified service description language would greatly benefit IoT networks by enabling seamless collaboration and interoperability among devices. Creating a unified service description language is a massive research challenge. In the future, service description languages such as XML, and UML. can be extended to create a standardized service description language for IoT networks. Research should also be focused on developing specialized service description language for IoT networks [390], [391].

Design of Service-Oriented Architecture: SOA is a style of software architecture where components of an application interact with other components of that application or other applications via services through utilizing a communication protocol over a network. The connection is independent of vendor or other limiting hardware or software technologies. The design of SOA for IoT systems is a huge challenge. On one hand, SOA simplifies the management of heterogeneous types of connected devices. On the other hand, SOA introduces additional overhead to devices. Most IoT devices are resource-constrained. Adding overhead to these resource-constrained devices slows down the entire network making data processing, collection, and transferring more difficult. Thus, researching lightweight SOA with minimum overhead that facilitates big data warehousing and processing is a huge research challenge [392], [393]. In the future, novel architectures for IoT networks should be developed to be modular and service-oriented in nature.

Hardware Customization: IoT networks currently degenerate into traditional ICT-oriented networks. Similarly, the hardware of various intelligent devices is not customized to specific protocols. IoT devices generate huge amounts of data within short amounts of time. Higher bandwidth networks are necessary alongside big data measures to store and process the data in real-time. Developing novel methods to transport huge amounts of data in commodity networks is a huge research challenge, as it is impossible to deploy high bandwidth high-frequency networks all around us. Similarly, optimizing the hardware of general intelligent things to efficiently use certain universal IoT protocols is a research challenge. In healthcare systems, various sensors are used to monitor patients. Developing novel hardware design solutions to contain various sensors in small housing would enable a single device to monitor various aspects of a patient. In the future, various bio-inspired optimization techniques such as swarm optimization, and genetic algorithms should be utilized to automatically develop optimized hardware design for various use cases.

Mobility of IoT Devices: In an interconnected world from a network perspective, intelligent

devices are expected to disconnect from one network and connect to other networks. In traditional networks, mobility is a huge issue. For mobile nodes, routing tables have to be re-calculated each time nodes move out of range from one another and come into the range of other nodes. This calculation puts a huge overhead on the entire network when numerous nodes with high mobility are present [394]. Modeling and handling highly mobile nodes in IoT networks are a huge research challenge. In healthcare systems, some monitoring devices might be mobile, while others might be stationary. For example, wearable monitoring devices might be highly mobile, while surveillance monitoring devices would be stationary. Patients would wear and bring highly mobile monitoring devices to their activities of daily life. Thus, optimizing the various aspects of these mobile devices, such as route discovery, routing table generation, efficient data transfer, and reducing data costs is a research challenge.

In the future, AI-based approaches such as reinforcement learning-based techniques can be used to develop networks facilitating highly mobile devices [395], [396]. Vehicular Ad-hoc Network (VANET) architectures can be extended in the future to IoT networks involving highly mobile nodes [397].

Reliability in IoT Networks: Reliability is a huge issue in commodity networks. Regardless of the network and type of protocol, network packets get lost with the data. While some data warehousing solutions provide checkpointing and state management for error detection and recovery, it is not the norm and requires extensive installation and support. Developing lightweight reliable networks with error detection and error recovery is a huge challenge. Devices connected to IoT networks themselves need to be reliable [398], [399]. Erroneous reporting might result in faulty directions from central controller servers. Healthcare systems in general require reliable and fast networks. Any drop in data, as well as monitoring results, might be irreversibly harmful to the users. For example, in stroke, diabetes, Alzheimer's, or fall monitoring systems, it is imperative to monitor the patients without interruptions, and any kind of delay in transmission or monitoring might prove fatal for the patients. Developing specialty fault-tolerant and state-recovery-capable networks for healthcare systems is a challenge. Similarly, unreliable devices in healthcare IoT networks can result in the loss of life of patients in extreme cases.

In the future, network traffic forecasting models should be developed to predict traffic load and allocate network resources accordingly. Various checkpointing algorithms can be used for error detection and correction in networks. Network designs should also be fault tolerant. Advances in TinyML enable us to deploy and run ML models in resource-constrained environments. In the future, devices connected in IoT networks should use TinyML to run basic models to provide basic service functionalities during network downtime or other failures. This would increase the reliability factor of the devices. To reduce the downtime of the devices, advanced battery solutions should be used in the future.

Scalability in IoT Networks: IoT networks might have millions or billions of interconnected nodes. General networks and services are not scalable in nature. Scalability is a huge challenge for such IoT networks [400], [401]. In healthcare systems, each patient might have multiple types of monitoring going on at the same time. For example, elderly patients might be monitored for heart diseases, glucose levels, diabetes, Parkinson's, Alzheimer's, and falls at the same time. For a sufficiently large elderly population, the number of these systems would reach millions, requiring specialty scalable services and applications.

In the future, the potential of using existing Database-as-a-Service (DaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) to increase the scalability of IoT services and networks should be explored. The research focus should be on developing modular IoT solutions that can scale on need.

Availability in IoT Networks: IoT has applications in industrial and healthcare domains among other potential application areas. Industrial devices are placed for continuous monitoring of various parameters such as temperature, humidity levels, and pressure. Even during network failures, the devices should be able to provide basic monitoring services to avoid catastrophic failures in industrial cases. In the healthcare space, monitoring devices need to be available at all times similar to industrial use cases. Lack of availability might result in serious bodily harm or even death. Devices in IoT networks must provide basic monitoring and action services even when they are disconnected from networks or when network failures happen. The availability should be implemented in both hardware and software levels.

Similar to the reliability, research should be focused on deploying basic models into devices using TinyML, which would increase overall availability. Novel error detection and error correction in fault-tolerant networks should be explored to increase the availability of the network.

Interoperability in IoT Networks: Interoperable applications and services can communicate and transmit data among themselves regardless of hardware or software differences. However, interoperability is very hard to achieve. New protocols or architectures that are constantly being developed do not always support all legacy protocols and architectures, making data transmission and interaction between devices very complex and problematic. Similarly, OSs and various applications also have different implementations, making data sharing very difficult. Vendors often have patented and exclusive implementations that make their devices only compatible with other devices of the same vendor. Thus, interoperability is a huge research challenge for IoT [402], [403]. Interoperability is very important in healthcare systems as various medical equipment have differing hardware, protocol implementations, and data representation. For a smooth deployment and operation of IoT networks in the healthcare sector, the various devices and services need to be interoperable.

In the future, the research focus should be on developing open-access network architecture

Table 3.4: Summary of challenges in large-scale deployment of IoT networks and future research recommendations.

Type	Concerns	Recommendations
Security and Privacy	**Healthcare data** is highly sensitive and contains personal identifying information such as social security information, and insurance information.	**More secure overlay networks** such as The Onion Router (TOR) network might be used to transfer confidential data. **Technologies** such as blockchain and NFT tickets can be used to provide authentic and customized services. Existing security solutions such as RSA, seed phrases, and DSS might be employed in all connection endpoints. **Authentication and identity verification methods** such as password, fingerprint scanning, signature, voice pattern, and smart card can be employed in application protocols. **AI-based approaches** can be used to detect anomalies in IoT networks.
Standardization and Interoperability	**Different devices** in IoT networks use differing application protocols, data representation methods, network protocols, marshalling, de-marshalling, and identification. **No standard** service description language. **Limited interoperability** among devices.	**Developing a standardized service description language** would benefit the standardization and interoperability of IoT devices. **Open-source implementations** of vendor-locked applications and networking protocols would enable interoperability. **Standardizing application and networking protocols** would simplify interoperation among devices in IoT networks.
Data Variability and Scalability	**Data transmitted in IoT networks** vary by size, type, and representation over the network. **Massive amounts of data** can be transmitted over an IoT network in a very short amount of time. **General applications and networks** are not scalable in nature.	**Utilizing big data warehousing techniques and open-access application frameworks** such as Apache Hadoop, and Apache Zookeeper will allow the storing and processing of huge amounts of data. **Modularity** can be prioritized to make devices in IoT networks more scalable. **Lightweight SOA of devices** would enable scalable big data warehousing and processing. Utilizing available SaaS, PaaS, and DaaS platforms such as AWS, Google App Engine would increase the scalability of IoT-orientated solutions. **Lossless compression algorithms** might be utilized to transmit huge volumes of data over standard IoT networks.
Power and Cost Efficiency	**Increasing number of intelligent devices** connected in IoT networks has harmful effects on nature. **Devices** are not as power-efficient as they could be. **Setting up and maintaining** all-day monitoring via multiple devices is not cost-efficient.	**Solar-powered hardware solutions** would increase the sustainability and eco-friendliness of the huge number of devices used in IoT networks. **Creating multi-purpose generalized devices** would reduce the number of separate devices necessary to monitor or surveil patients or environment. This reduction in number would also reduce the overall cost. **Specialized hardware** for task-specific purposes might reduce overall power consumption. **Optimized software** would also reduce the amount of necessary processing, which might optimize power consumption.
Reliability and Availability	**Devices** used in monitoring purposes in both industrial and healthcare use cases need to be available all the time regardless of network or other failures for avoiding dangerous or life-threatening situations. **Intelligent devices** used for monitoring and surveillance purposes need to be reliable at all times. Unreliable monitoring can result in erroneous diagnoses which might even lead to death of patients. **IoT networks** need to maximize availability and reliability. Downtime of networks might lead to loss of valuable data.	**Fault-tolerant development principles** should be followed while developing specialized IoT networks. **Error detection and recovery techniques** in networking such as distributed checkpointing and coordinated checkpointing can be implemented to increase network reliability. **Network traffic analysis and traffic forecasting** might be used to anticipate network load, which will help in allocating appropriate resources to increase network availability. **Simplified versions of applications** must be available in devices to facilitate basic services even during network failure or downtime. **TinyML deployment techniques** can be leveraged to ensure minimal functionality during network downtime. **Sensing modules** in devices must be within standardized error tolerance limits and must undergo scheduled testing to guarantee utmost reliability. **Using state-of-the-art battery technologies** might improve the reliability and availability of devices.
Identity and Mobility	**IPv4 network protocol** might not be sufficient to address millions of devices connected in IoT networks. **Networks and routing protocols** are not efficient in dealing with numerous highly mobile devices in IoT networks.	**IPv6** can be adopted as the standard IP for IoT networks. **AI approaches** such as reinforcement learning techniques can be utilized to develop and test new routing protocols capable of efficiently dealing with highly mobile nodes. **VANETs** already provide networking solutions for mobile and stationary nodes in networks. Relevant components from VANETs can be evolved for use in IoT networks.

and protocols to facilitate interoperability among devices. Open-access implementations of specialty architecture and protocols would also facilitate this issue.

Power and Cost Efficiency in IoT Networks: IoT networks have millions of nodes. Power and cost are huge concerns for such a huge number of devices. The entire world is also experiencing various global natural catastrophes due to global warming and climate change caused by extracting and depending on fossil fuels, trashing the oceans with electronic waste, and chemical waste. With an increased understanding of the causes behind the sudden climate change and global warming, researchers are increasingly moving towards sustainable, eco-friendly, and biodegradable solutions. Making IoT devices more eco-friendly, cost-efficient, and power-efficient is a massive research challenge [404]. Similarly, converting intelligent medical monitoring devices in healthcare IoT networks into more multi-purpose, power and cost-efficient, long-lasting, eco-friendly versions is a research challenge.

In the future, research should be focused on simplifying and combining multiple sensing modalities in single devices to reduce the number of total sensors necessary to monitor and surveil users. The focus should also be on developing more area-surveillance solutions. Area-surveillance technologies would reduce the number of devices per individual. Reducing the total number of devices needed per user would also drastically reduce the cost of effectively deploying an IoT solution. In the future, research should be focused on transforming commodity devices into solar-powered versions. Research on hardware and software optimization would also increase the power the efficiency of devices. Making the devices solar-powered would reduce the overall pressure on fossil fuels, which has beneficial effects on nature.

The summary of challenges in large-scale deployment of IoT networks and future research recommendations are shown in Table 3.4 to provide a broad idea to the readers in a short look.

3.8 Conclusion

The Internet has significantly shaped the way people live, allowing individuals to interact with one another on a virtual level in a variety of situations ranging from professional to social. The IoT could contribute meaningfully to this system by enabling communication among connected devices, which tends to result in the dream of "anymedia, anytime, anything, anywhere" connectivity. This chapter focuses on the cutting-edge IoT device capabilities, architectures, and protocols that describe the IoT conceptual framework. The IoT device capabilities for hardware and software tools are thoroughly demonstrated in this chapter. There have been discussions about possible IoT architectures, such as the conventional three-layer architecture, the SOA-based architecture, and the middleware-based architecture. Additionally, the IoT communications protocols

with their distinguished features are incorporated briefly in this chapter. We have highlighted various IoT applications for smart healthcare that have been developed in recent decades to ensure some facilities for medical practitioners such as monitoring and identifying a variety of health problems, measuring various health factors, and offering diagnosis and treatment at remote locations. Lastly, the current challenges and gaps related to the development of IoT-based healthcare applications, along with future improvement and research focus in the coming years have been explored here. Hopefully, this chapter will provide a stable platform for scholars and scientists willing to engage in learning more about IoT technologies in the context to fully grasp the general architecture and roles of the multiple elements and processes that comprise the concept of IoT.

Chapter 4

Multimodal Human Activity Recognition through Self-Attention

4.1 Introduction

HAR is the task of detecting the human activities being conducted by individuals based on the inputs from numerous wearable or stationary devices. HAR has a wide range of applications in various domains thanks to the rapid development of modern innovative techniques including sensor technology, AI specifically DL as well as ML, IoT, mobile computing, and cloud computing [405]. Applications of HAR include real-time always-on healthcare monitoring systems, gaming systems, surveillance systems, as well as crowd control and management systems [406]. The most recent works in HAR and AAL have found a significant correlation between physical activity levels and preserving a healthier life. Smart healthcare has become a crucial part of our daily life due to its capability of reducing the growing strain on public healthcare systems [407]. The monitoring of real-time human activities of the older individuals is an important concern in smart healthcare that can obviously improve medical rehabilitation and care for the elderly utilizing portable and mobile sensing devices [408].

Different kinds of wearable sensors [409] are placed on different body positions to collect the activity data and transfer the real-time collected data using IoT technology. Typically, accelerometer trends to measure the acceleration along multiple axes during human motion, while gyroscope trends to measure rotation information. In some cases, cameras are also used to capture human activities in AAL environments. Camera systems are mostly stationary in nature. Considering the privacy issues, sensors data are more convenient than the camera data for HAR.

Most of the existing frameworks used traditional ML techniques [410] that take fixed-length windows from time-series sensor data as input and classify each window to one activity. Over the last decades, conventional ML approaches such as Random Forest (RF), Decision Tree, k-NN, Multilayer Perceptron (MLP), and SVM have been widely utilized to address the HAR problem. However, these approaches need pre-processing steps which are time-consuming to generate handcrafted or domain-specific patterns. Nowadays, DL [411] has emerged as a dominant strategy in ML techniques for automatically extracting discriminant information, reducing the load of handcrafted attributes. Especially, CNNs and Long Short-Term Memory (LSTM) have received much attention for HAR due to their exceptional performance. CNN approaches [412] utilize the local dependency and translation invariance of information making it ideal for signal inference of the collected data. Additionally, the LSTM [413] is proven to be fit for handling long-term dependencies in information that ensure the data diversity in the architecture. The recent trends for HAR are focused on the attention mechanisms [414] that consider the important parts of the data, and compute the outputs with minimum noise information.

The organization of the remaining part of the chapter is given below. Section 4.2 demonstrates the most recent intelligent frameworks for DL-based HAR. Section 4.3 presents the proposed fusion architecture with a brief description of the used evaluation protocols. The experimental findings, the description of the dataset, and comparisons with the existing SOTA are briefly described in section 4.4. Section 4.5 concludes the research with potential future works.

4.2 Related Works

Several efficient and accurate frameworks have been developed for HAR in smart healthcare environment with the evolution of modern technologies including DL and IoT. The most recent and promising frameworks for HAR based on DL architecture are demonstrated as follows.

Betancourt et al. [415] developed a HAR platform using SA networks that used data from smartphones for recognition purposes. The proposed approach extracted the necessary features from time-series data by precisely distributing their emphasis among the important input attributes. The system is tested with two public datasets and obtained an average accuracy of 97.1% for University of California Irvine (UCI)-HAR dataset. The proposed architecture has a drawback that it is only applicable to data from wearable devices, and the learning period is not appropriately considered. Nazari et al. [416] compared various ML approaches for HAR utilizing the data from the knee angle. The ML techniques are trained with raw data and handcrafted extracted features. The experimental result shows that Gradient Boosting algorithm found an Area Under the Curve (AUC) of 0.94 for raw data using 11-fold cross-validation. However, the use of manually extracted features decreased the recognition performance. Tan et al. [417] introduced a fusion

architecture for HAR using faster region-based CNN and attention-based LSTM. The features are extracted from raw RGB images using CNN and the extracted patterns are fed into Bi-LSTM architecture for activity recognition. The attention layer is put between the Bi-LSTMs layers. The experimental findings reveal that the developed system obtained a precision of 97.02% and a recall of 96.83% for classifying twelve activities.

Further, Mekruksavanich et al. [418] proposed a lightweight and efficient framework for HAR from smartphone data using CNN with squeeze-and-excitation modules that considered entirely swapping in the same layer. The developed scheme enabled lower-level superficial CNNs to collect more important data. The proposed approach achieved an accuracy of 99.27% for identifying six human activities for a publicly available dataset (UCI-HAR). Singh et al. [419] described a DL architecture that used SA mechanisms to select significant information of time-series data for HAR. The SA layer was utilized to learn weights that represent the latent relationship between input time points of raw sensory information to decode activities effectively. The experiments have been continued for six different public datasets and the best accuracy, recall, and F1-Score of 94.86%, 94.35%, and 94.74% individually are obtained for MHEALTH dataset. Wan et al. [420] compared five ML algorithms namely CNN, LSTM, Bi-LSTM, SVM, and MLP for HAR utilizing smartphone-inertial-sensor data. In this system, CNN is exploited to retrieve the high-level patterns from the original data as well as classify the extracted features whereas the other techniques are used for activity recognition. The best performance measures such as accuracy of 92.71%, precision of 93.21%, recall of 92.82%, F1-Score of 92.93% are found from the experimental outcomes using CNN for UCI-HAR dataset.

Furthermore, Dua et al. [421] introduced an end-to-end DL platform that used CNN and Gated Recurrent Unit (GRU) to extract features automatically to recognize human activities. In this architecture, the CNN extracts local features, whereas the GRU layers effectively capture long-term data dependencies. The proposed model is validated through the experimentation of three benchmark datasets. The accuracy measures of the developed system are 96.20%, 95.27%, and 97.21% for UCI-HAR, PAMAP2, and Wireless Sensor Data Mining (WISDM) datasets, individually. Challa et al. [422] demonstrated a multibranch network using CNN and Bi-LSTM for HAR from the wearable sensors data. The proposed scheme utilized several filters in varied sizes to learn temporal local dependencies resulting in the improvement of the feature extraction stage. Additionally, a Global Average Pooling (GAP) replaced the Fully Connected (FC) layer after convolution that decreased the parameters of the developed model. Although the developed architecture has experimented with three benchmark datasets, the best accuracy, and F1-Score of 96.37%, and 96.31% are achieved, respectively, for UCI-HAR dataset. Zhou et al. [423] used LSTM technique for HAR targeting the IoHT application. To address the shortage of labeled data, the developed architecture exploited a Q-network to label data automatically with a reward based on the distance. A multi-sensor data fusion approach has been introduced to handle the

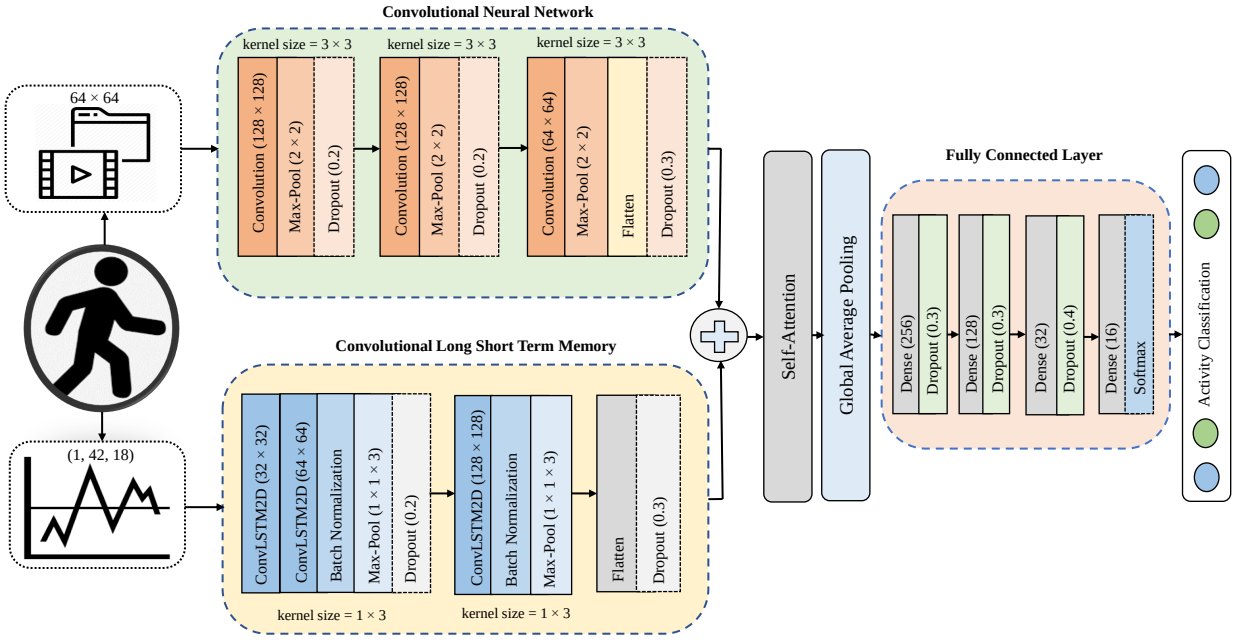


Figure 4.1: The system architecture of the proposed fusion approach for multimodal human activity recognition. From top to bottom, the two modalities represent input data to the networks, CNN, ConvLSTM, Self-Attention, GAP, fully connected layer with softmax activation function, and activity classification.

sequential motion data, and environmental data to determine fine-grained features according to the extracted high-level patterns. The experimental findings revealed that this scheme outperformed the baseline techniques with a Receiver Operating Characteristic (ROC) curve up to 0.95. Espinosa et al. [424] presented CNN inference architecture for human fall detection utilizing the UP-Fall detection dataset. This method used an optical flow approach to collect data on the relative motion between two successive images to evaluate images in specified time frames and extract features. Although the dataset contained the features of several types of human activities and fall, this system considered two-class classification problem whether the performed activity is fall or not. From the experiments, it is evident that the proposed multi-vision-based approach appraised an accuracy of 95.64% considering the image data.

4.3 Proposed Self-Attention based Fusion Approach

The main purpose of our proposed approach is to recognize human activities from different modalities of data to ensure robustness. The proposed DL-based fusion architecture is shown in Fig. 4.1 which consists of CNN, ConvLSTM, and SA modules followed by GAP and a FC layer. In this system, the data is collected from the subjects through various sensors and cameras that are sent to the ConvLSTM and CNN branches as input. The sensor modalities are wearable while the camera modality is stationary. The high-level features are extracted using ConvLSTM and CNN from time-series and image data, respectively. They are then fused through the SA mechanism. The attention mechanism can recognize the most important activity data regions and boost their relevance by limiting superfluous and potentially misleading data in other activity data regions. The output of the GAP layer is fed into a FC layer followed by a softmax activation function for the final activity classification. The details of the proposed fusion architecture are described as follows.

4.3.1 Convolutional Neural Networks

CNNs are very popular and widely used deep neural networks that learn kernels and use convolution operations to retrieve relevant features from raw input samples [425]. The convolution filters of CNN can highlight the most relevant features while minimizing domain-specific issues efficiently. Additionally, each layer's filter is able to retrieve some significant features for the same receptive field leading to a variety of outputs known as feature maps. Another important property of CNN architecture is the pooling mechanism that is responsible for receiving the information of feature maps and sending them to other layers in the network [426]. The functionalities of the CNN architecture are presented below.

For a feature f, the output (O) of a neuron (xy) of the i^{th} convolutional layer is expressed as in (4.1) considering the activation function as ReLU, the number of conventional cores as C, and the feature map has a bias term called b.

$$O_{xy}^{(i,f)} = \text{ReLU}\left(\sum_{i=0}^{C-1} \sum_{j=0}^{f_n} \sum_{k=0}^{f_w} W_{(j,k)}^{(f,i)} O_{(x+j),(x+k)}^{(x-1),i} + b \right) \tag{4.1}$$

Similarly, the output of neuron (xy) for sub-sampling layer s considering the feature f can be calculated using (4.2).

$$O_{xy}^{(i,f)} = \text{ReLU}\left(W^f \sum_{j=0}^{s_n} \sum_{k=0}^{s_w} O_{(x*s_n+j),(y*s_{w+k})}^{(i-1),f} + b \right) \tag{4.2}$$

Let, the number of features in a particular layer is m and the hidden layer is h. Hence, the output of a neuron n is given by (4.3).

$$O_{(i,n)}^{(i,f)} = \text{ReLU}\left(\sum_{j=0}^{m-1} \sum_{x=0}^{m_n} \sum_{y=0}^{m_w} O_{(x,y)}^{(i-1),j} + b \right) \tag{4.3}$$

For multi-class HAR, the output is taken from the softmax layer. The output layer is expressed using (4.4) for c^{th} neuron and feature map F.

$$O(j,c) = \text{Softmax}\left(\sum_{n=0}^{F} O_{(j-1),n} W_{(c,n)}^{j} + b \right) \tag{4.4}$$

In the proposed architecture, we considered the input image size of 64×64. The convolution layer with filter sizes 128, 128, and 64 is utilized to retrieve relevant attributes. The size of kernel is 3×3 for all convolution operations. The max-pooling layer with size 2×2 is utilized for all pooling operations. We used a dropout layer with the percentage of 0.2, 0.2, and 0.3 to prevent overfitting.

4.3.2 Convolutional LSTM Network

The conventional LSTM network is particularly designed to learn long-term dependencies. The performance of such a network is decreased while handling sequential data having temporal and spatial information [427]. To cope with spatial data, the LSTM network uses a pooling/FC layer to convert spatio-temporal information to temporal information, discarding the spatial features. The ConvLSTM takes the inputs and previous states of its local neighbors to determine the future state of a cell in the matrix [428]. Thus, this architecture solves the above issue by the use of convolution operators in the state-to-state and input-to-state transitions for multiplication. The major operations of ConvLSTM are demonstrated below mathematically.

Let, the input features are denoted as $x_{(1)}, \ldots, x_{(t)}$, the cell outputs are $c_{(1)}, \ldots c_{(t)}$, the hidden states are $h_{(1)}, \ldots h_{(t)}$, as well as all the gates are marked as $i_{(t)}$, $f_{(t)}$, and $o_{(t)}$ for input, forget, and output, respectively. Here, the cells, the hidden states, and all the gates are 3D tensors. The input gates are formulated using (4.5).

$$i_{(t)} = \sigma(W_{xi} * x_{(t)} + W_{hi} * h_{(t-1)} + W_{ci} \circ c_{(t-1)} + b_i) \tag{4.5}$$

The forget gates are represented as in (4.6).

$$f_{(t)} = \sigma(W_{xf} * x_{(t)} + W_{hf} * h_{(t-1)} + W_{cf} \circ c_{(t-1)} + b_f) \tag{4.6}$$

The cell states are calculated using the output of input gates, and forget gates which is calculated in (4.7).

$$c_{(t)} = f_{(t)} \circ c_{(t-1)} + i_{(t)} \circ \tanh\left(W_{xc} * x_{(t)} + W_{hc} * h_{(t-1)} + b_c\right) \tag{4.7}$$

Using the output of the cell states, the representation of the output gates is given by (4.8).

$$o_{(t)} = \sigma(W_{xo} * x_{(t)} + W_{ho} * h_{(t-1)} + W_{co} \circ c_{(t)} + b_o) \tag{4.8}$$

Finally, the hidden states are determined through the use of output gates and cell states as in (4.9).

$$h_{(t)} = o_{(t)} \circ \tanh(c_{(t)}) \tag{4.9}$$

where, the convolution operation is denoted by $*$, the Hadamard product is marked as $\circ$, and σ is the sigmoid activation function. Additionally, the 2D convolutional kernels are represented as $W_{x\sim}$ and $W_{h\sim}$.

Moreover, the ConvLSTM can take the feature maps of each sequence while keeping the spatial dimensions fixed. In our proposed system, the input sequence from the time-series multi-sensory data is (1, 42, 18). The size of the ConvLSTM operation filters are 32, 64, and 128. The same kernel size of 1×3 is utilized for each ConvLSTM operation. Batch normalization layer is added after activation layers which continuously receives the output from the previous layer and normalizes it before passing it to the next layer to stabilize the network. The max-pool operation with kernel size of $1 \times 1 \times 3$ is adopted. A dropout rate of 0.2 and 0.3 is selected to reduce overfitting during training the network.

4.3.3 Self-Attention Mechanism

The SA network is a widely used approach for calculating correlation and the weighted combination of all timestamps in the input sequence. The attention technique attempts to extract the most significant timestamps from the sequence feature maps, enabling more efficient recognition [429]. Additionally, the SA method computes the weighting factors for each timestamp in the sequence feature map by comparing it to all other timestamps in the sequence. The attention mechanism transforms a query and a set of key-value pairs into an output, where each part of the input is calculated as a weight depending on its relative importance to the output [430]. The major components of the attention mechanism are the query, keys, and values that are denoted as Q, K, and V respectively. Additionally, the same dimension of d_k is used for Q, and K, as well as the outputs and values have the same dimension of d_v. The attention score is calculated by implementing the softmax function to the scaled dot products of queries and keys and then using

the weighted form of the values. The overall implementation is carried out in mathematical form, as specified in (4.10).

$$f^{sa}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4.10}$$

Here, $Q = IW_Q$; $K = IW_K$; $V = IW_V$ where W_Q, W_K, W_V are weight matrices; I is the input and $Q \in \mathbb{R}^{t \times d_k}$; $K \in \mathbb{R}^{t \times d_k}$; and $V \in \mathbb{R}^{t \times d_v}$.

The output of the SA block is flattened to one dimension using the GAP layer and is forwarded to the FC layer. This layer comprises several dense layers with filter size of 256, 128, 32, and 16. The dropout rate in this FC layer is 0.3, 0.3, and 0.4. Lastly, the generated vectors are sent to the softmax function to calculate the probabilities of each class. The loss function is utilized to measure the performance of the model. In general, the softmax-loss measures the final output based on the probability of each activity belonging to a class. The loss function for data samples D with class labels (y_i), and the predicted labels ($\hat{y}_i$) is given by (4.11).

$$Loss = -\frac{1}{D}\sum_{i=1}^{D}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right] \tag{4.11}$$

4.3.4 Evaluation Protocols

Several statistical measures such as accuracy, precision, recall, and F1-Score are adopted to evaluate the performance of the proposed fusion architecture. The performance measures are mathematically represented as follows.

- **Accuracy**: Accuracy is the ratio of correctly recognized samples in all data samples that represent the rate of recognition of human activities.

$$Accuracy = \frac{1}{D}\sum_{k=1}^{D}\frac{TP_k + TN_k}{TP_k + TN_k + FP_k + FN_k} \tag{4.12}$$

- **Precision**: Precision is the ratio of real positive samples to the total number of recognized positive data instances.

$$Precision = \frac{1}{D}\sum_{k=1}^{D}\frac{TP_k}{TP_k + FP_k} \tag{4.13}$$

- **Recall**: Recall represents the proportion of correctly recognized positive samples to the total number of real positive data instances.

$$Recall = \frac{1}{D}\sum_{k=1}^{D}\frac{TP_k}{TP_k + FN_k} \tag{4.14}$$

- **F1-Score**: F1-Score is a measure to evaluate the performance of multi-classification architectures that is represented by the weighted average of recall and precision, which facilitates to give more insights into HAR tasks.

$$F1 - Score = \frac{1}{D}\sum_{k=1}^{D}\frac{2 * TP_k}{2 * TP_k + FP_k + FN_k} \tag{4.15}$$

Here, the correctly recognized samples are denoted as True Positive (TP), and True Negative (TN). Additionally, the incorrectly classified samples are known as False Positive (FP), and False Negative (FN).

4.4 Experimental Results Analysis

In this section, we are describing the experimental study including the dataset collection, quantitative results, and comparative study with the existing SOTA. We performed our experiments on a high-computing machine (GPU: GTX 1070 with 8 GB GDDR5 memory, CPU: Intel® Core™ i7-6700K CPU @ 4.00GHz × 8, RAM: 32 GB DDR4). All the experiments were conducted using Python(v2.10) and TensorFlow(v2.8.0).

4.4.1 Dataset Collection and Description

The dataset used to train and evaulate the proposed fusion model is retrieved from a large multimodal benchmark namely UP-Fall detection dataset [431]. The dataset is collected in a controlled environment through the use of accelerometers, gyroscopes, RGB cameras, ambient luminosity sensors, electroencephalograph headsets, and context-aware infrared sensors. The wearable sensors are placed in the waist, chest, foot, left wrist, and thigh pocket of the test subjects. The stationary cameras captured the particular activities at two different viewpoints: lateral and front views. The volunteers (17 healthy subjects) performed 11 different activities: 6 human ADL and

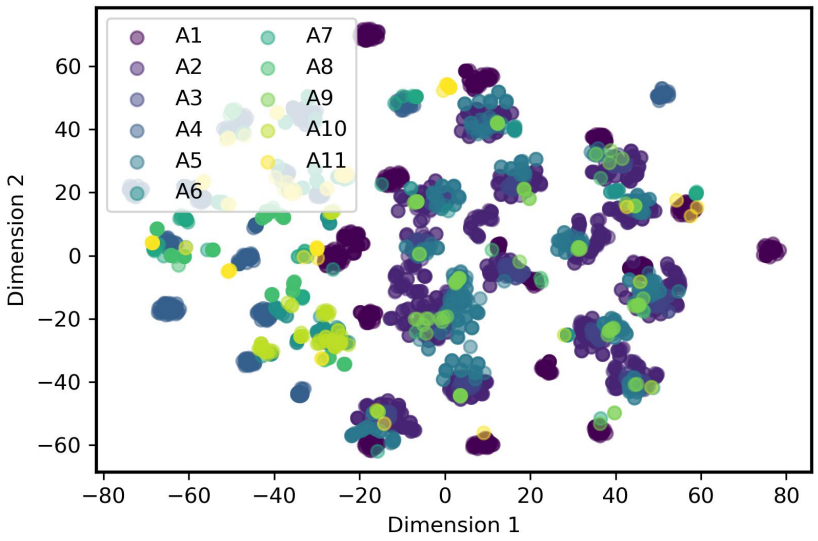


Figure 4.2: t-SNE visualization of multi-sensor human activity recognition dataset on two dimensional space.

5 human falls. The fall scenario data were captured for 10 seconds duration where the other activities had variable durations. The activities are demonstrated as follows. "Falling forward using hands (A1)", "Falling forward using knees (A2)", "Falling backwards (A3)", "Falling sideward (A4)", "Falling sitting in empty chair (A5)", "Walking (A6)", "Standing (A7)", "Sitting (A8)", "Picking up an object (A9)", "Jumping (A10)", and "Laying (A11)". Each activity was performed for three trials by all the subjects. The t-distributed Stochastic Neighbor Embedding (t-SNE) visualization for the representative multi-sensor dataset is shown in Fig. 4.2 to observe the data distribution on two dimensional space. The distribution shows clear inter-class distinction between fall activity clusters (A1-A5) and ADL clusters (A6-A11). However, very little intra-class distinction is visible as multiple separate activities seems to be clustered together.

In our proposed architecture, we considered trials 1 and 2 of each individual as the training set for each activity of the UP-Fall dataset, and trial 3 of each individual as the testing set for each activity. During preprocessing stage, we resized the image data to 64×64 from 640×480 to feed into the CNN. The image data were also rescaled to $[0, 1]$ range. There are 42 attributes in the time-series data; collected by different sensors for different activities. To train the ConvLSTM architecture, we at first standardized the data and grouped the data points into sequences. We considered the sequence size of 18 based on the smallest number of data points in a trial file.

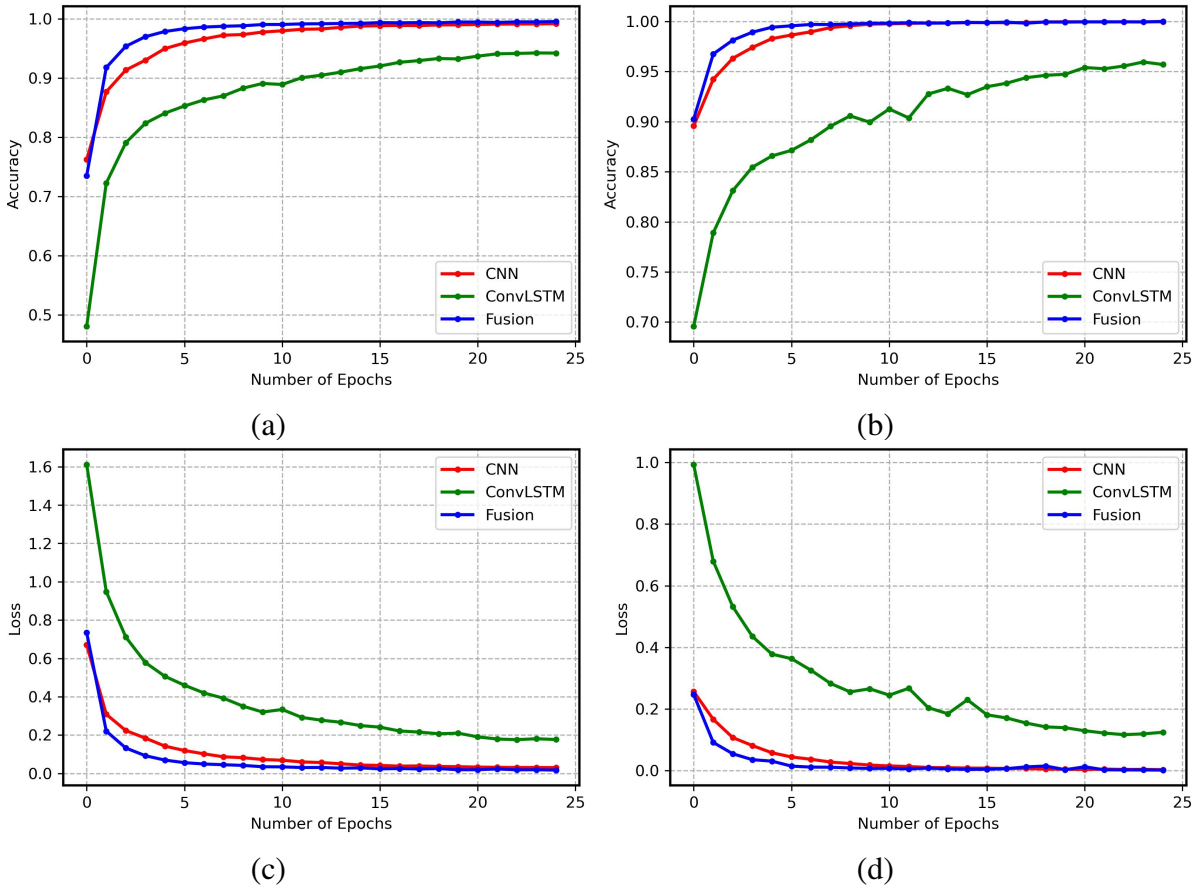


Figure 4.3: Accuracy and loss values of the proposed fusion approach along with two baseline models. (a)-(c) Training (b)-(d) Validation. The first row represents the accuracy values, and the second row depicts the loss vaules.

4.4.2 Quantitative Results

We performed the experiments for 25 epochs to train the proposed architecture. The following hyperparameters were considered during experimentation: batch size was set to 128, learning rate was set to 0.001, Adam was used as optimizer, and sparse categorical cross-entropy was used as the loss function. The Rectified Linear Unit (ReLU) was used as the activation function for all layers except the last layer. We trained the two baseline architectures: CNN and ConvLSTM to obtain their performance individually. In our proposed fusion approach, these two baseline architectures are used as a feature extractor by omitting the FC layers. The extracted high-level features are fused through SA mechanism to extract the most important features. The output

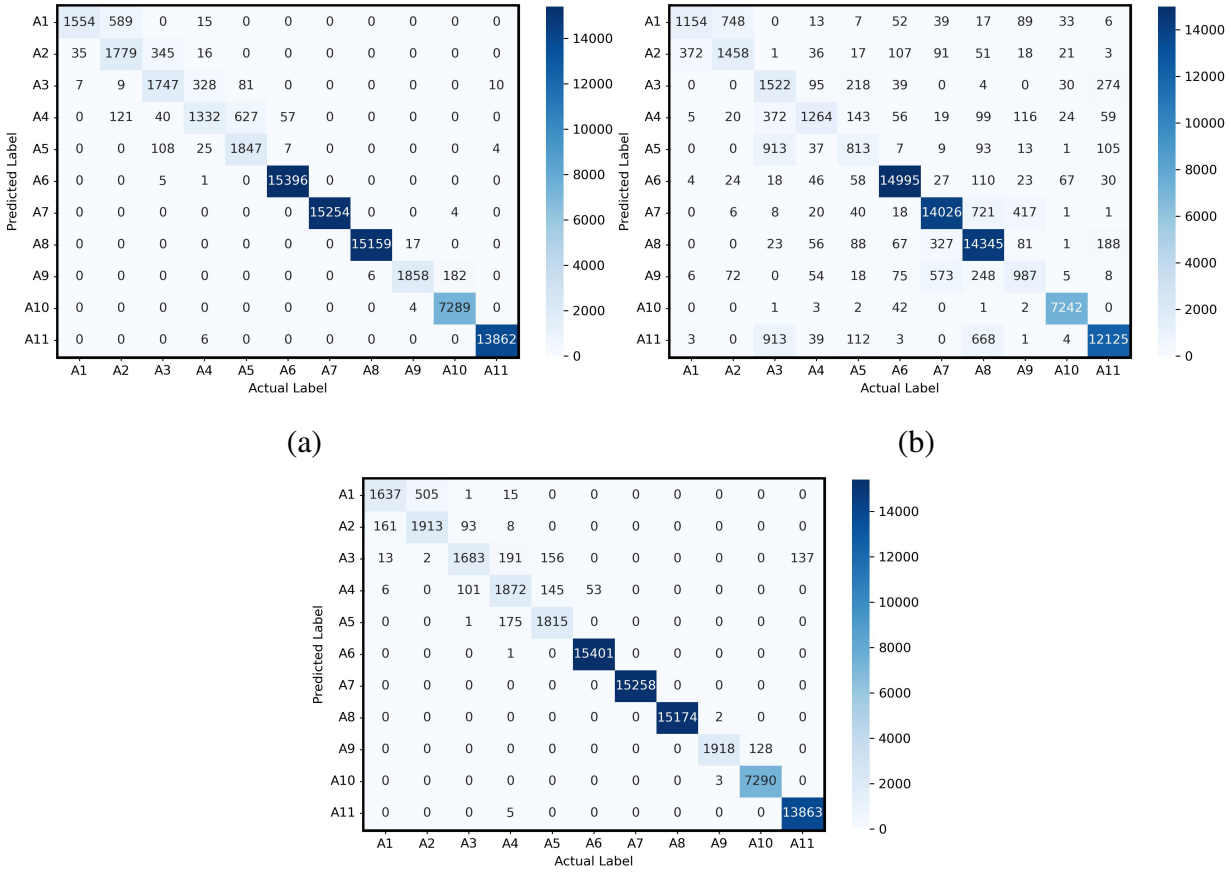


Figure 4.4: Confusion matrix for the proposed fusion approach along with two baseline models. The actual activity is shown in the leftmost columns, followed by a prediction in the subsequent columns. (a) CNN (b) ConvLSTM (c) Fusion.

of the SA is passed to the GAP layer. Finally, the activities are recognized by connecting the FC layer with the output of the GAP layer. We used 20% of the training data as a validation set to validate the developed framework. Fig. 4.3 (a)-(b) depicts the accuracy per number of epoch for multimodal HAR of two baseline approaches along with our fusion model for training and validation stages. In both training and validation stages, the fusion approach reaches higher accuracy levels in less number of epochs than the other baseline approaches. The ConvLSTM baseline approach reaches significantly lower accuracy in general than the other two approaches. During the training stage, the fusion approach reached slightly higher accuracy than the CNN

baseline approach.

The loss values for the training and validation process per number of epoch are shown in Fig. 4.3 (c)-(d). Similar to Fig. 4.3 (a)-(b), the fusion approach reaches lower levels of loss in less number of epochs than the other two baseline approaches. During the training stage, the fusion approach reached slightly lower loss value than the CNN approach. The ConvLSTM baseline approach had significantly higher loss value than the other two approaches at the end of both training and validation stage.

Fig. 4.4 illustrates the confusion matrix for multimodal HAR framework for (a) CNN, (b) ConvLSTM, and (c) fusion approach. It is evident that the proposed fusion approach recognizes the activities more perfectly than the baseline approaches. The recognition rate for activity seven (A7) is high for the fusion and CNN approach and they recognized almost all standing activities. For activities A6, A7, and A8, the fusion approach classified more samples correctly than the other baseline approaches. On average, the fusion approach misclassified lesser number of samples than the baseline approaches.

The obtained performance for the experimental study from the testing dataset is shown in Table 4.1. The proposed fusion approach achieved an accuracy of 97.61% whereas the CNN and ConvLSTM models obtained an accuracy of 96.67% and 87.71%, respectively. The precision, recall, and F1-Score of the developed fusion architecture are 97.61%, 97.64%, and 97.58%, respectively. The proposed network improved results of multimodal HAR by ∼1% up to ∼10% compared with two baseline architectures (CNN, and ConvLSTM).

The Gradient-weighted Class Activation Mapping (Grad-CAM) visualization of the predicted results is depicted in Fig. 4.5 for two particular activities: sitting and falling backwards. Both of these images were randomly selected from the test set. The generated map highlights the relative importance of each subregion's contribution to the recognition. In both cases, the model is tracking the person to detect the relevant activity.

Table 4.1: Recognition performance of the proposed fusion approach.

	Accuracy	Precision	Recall	F1-Score
CNN	96.67%	96.67%	96.86%	96.64%
ConvLSTM	87.71%	87.71%	88.29%	87.73%
Fusion	**97.61%**	**97.61%**	**97.64%**	**97.58%**

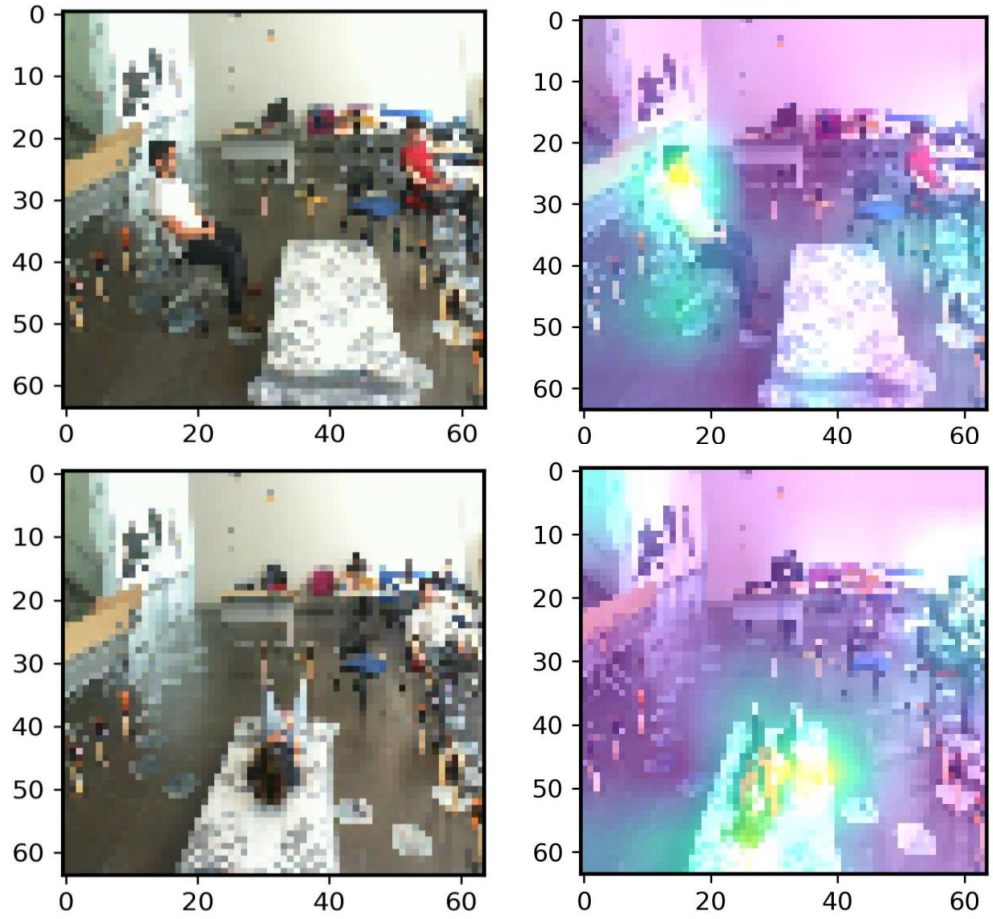


Figure 4.5: Visualization of the predicted results of the proposed system. The left column represents the original images, and the right column shows the corresponding Grad-CAM.

4.4.3 Comparison with the State-of-the-Art

A comparative study with the existing literature for multimodal HAR using the UP-Fall detection dataset is demonstrated in Table 4.2. The system developed in [432] used SVM to recognize human activities and obtained relatively low performance. The authors from the same group used only image data in [431], and images as well as time-series data in [433]. The highest accuracy of 96.4% is achieved in [433] which used CNN and LSTM architecture to recognize the activities. In the literature, the highest accuracy of 96.7% is obtained in [434] utilizing CNN and GRU for activity recognition. They only used the image modality of the UP-Fall dataset. In this study,

Table 4.2: Performance comparison of the proposed approach with the state-of-the-art for UP-Fall detection dataset.

Refs.	Year	Data	Methods	Accuracy	Precision	Recall	F1-Score
[432]	2018	Time-series	SVM	–	66%	69%	67%
[431]	2019	Image	CNN	95.10%	71.80%	71.30%	71.20%
[433]	2019	Image + Time-series	CNN + LSTM	96.40%	84.20%	81.50%	82.30%
[424]	2019	Image	CNN	95.64%	96.91%	**97.95%**	97.43%
[434]	2022	Image	CNN + GRU	96.70%	96.90%	96.70%	96.60%
Ours	–	Image + Time-series	Fusion approach	**97.61%**	**97.61%**	97.64%	**97.58%**

we proposed a multimodal fusion architecture that utilizes both stationary RGB camera data and wearable sensor data to identify human activities. The accuracy, precision, recall, and F1-Score of the proposed fusion approach are 97.61%, 97.61%, 97.64%, and 97.58%, respectively. Our proposed approach outperformed the current literature with respect to accuracy, precision, and F1-Score. However, the best recall rate of 97.95% was found in [424].

4.5 Conclusion

HAR plays a key role in monitoring human health status using data from wearable and stationary devices in smart healthcare environment. In this research, we develop a DL-based fusion architecture for multimodal HAR that is able to process multi modality signals from smart devices. In our proposed architecture, visual patterns are retrieved from sequence of images using the CNN model and multi-sensory features from time-series are retrieved using ConvLSTM model. The features are then fused using a SA mechanism. The proposed architecture along with two baseline frameworks was evaluated using a large benchmark dataset called UP-Fall detection dataset to depict the robustness and effectiveness of this architecture. The experimental results revealed that the best accuracy of 97.61% is achieved from the developed fusion system for classifying eleven activities. The experimental results demonstrate that the fusion scheme contributes to a significant improvement in classification accuracy. The practical applications of the proposed system include health monitoring, sleep disorder monitoring, eating disorder monitoring, medication intake monitoring, and exercise aid systems.

Chapter 5

Hierarchical Architecture for Feature Fusion in Multimodal HAR

5.1 Introduction

HAR refers to the recognition of movement behavioral patterns by analyzing human activity signals, which has a significant economic value and scientific research importance in the area of smart healthcare [435]. Real-time monitoring of the activities of the elderly is a key task in the development of smart health monitoring that employs signals from wearable and mobile devices to enhance medical decision support systems [436], [437]. Hence, HAR has evolved into a dynamic and peremptory research area for the effective detection of human activities and interactions in AAL environments, which is being widely investigated for the aptly called IoHT [438], [439].

Nowadays, there are numerous wearable sensors integrated with intelligent devices including smartphones and smartwatches, as well as numerous wearable components that can be attached to multiple on-body positions to collect and transfer real-time activity data through wireless sensor networks [440], [441]. These sensing paradigms offer effective data fusion for efficient HAR that supports the development of services and applications for the elderly in smart healthcare frameworks based on the generation of real-time fused information [442], [443]. Over the years, several HAR frameworks have been introduced based on various approaches including sensors [444], cameras [445], Wi-Fi signals [446], and others. The sensors are more useful for HAR because they are more reliable in testing situations, do not need proper lighting, and preserve human privacy. However, the sensor-based HAR systems are expensive and burdensome for their users. The major advantages of vision-based HAR systems are that they spare regular users from having to wear a number of cumbersome devices on various parts of their bodies. Nevertheless, these

systems have privacy concerns, and their efficiency is influenced by the level of ambient lighting. On the contrary, Wi-Fi based HAR techniques use signal descriptors to identify the human activities that are reflected in the mobility of Wi-Fi signals. However, these techniques have a limited range, require significant investment, and entail unified protocols as well as benchmarks.

Conventional ML techniques [410] are widely used for HAR in smart healthcare applications. The major task of ML approaches includes the extraction of the most significant features from the raw data. In general, time-domain and frequency-domain attributes are exploited for ML-based HAR. The derivative features are fed into ML algorithms for HAR purposes. Several ML-based HAR frameworks [447], [448] have been developed to address the challenges associated with real-time HAR in IoHT. However, the hand-crafted features that need subject-matter expertise and human experience present some serious challenges for ML algorithms. DL algorithms [449] have recently been introduced for HAR purposes as they have demonstrated their exceptional efficiency in automatic feature extraction as well as recognition. DL technique [450] extracts the necessary features from the multimodal raw data that are quite effective to enhance the performance of HAR compared with conventional techniques. In literature, various HAR systems have been proposed based on DL approaches. The most popular and extensively exploited approaches for HAR are CNN [412], which use several convolutional kernels to extract hidden patterns from raw visual data. To handle sequential data, RNNs [451], [452] have shown promising performance for long-term dependency on multi-sensor information. Besides, the time-series data collected from wearable sensors are encoded to image representation including recurrence plot, Gramian angular field, and Markov transition field in recent studies [453]. Moreover, the current researches are highly focused on attention mechanism with DL architectures for HAR applications in IoHT environment. For many time-series and visual data classification applications [454], the attention mechanism has become essential and performs better than traditional methods. Attention mechanisms focus on the most significant extracted attributes and remove the unnecessary noisy attributes during the recognition - resulting in overall performance improvement.

As multiple wearable devices are utilized to develop HAR frameworks, it is required to synchronize and combine the various sensor information using a uniform data fusion approach to record more complicated activities from and multimodal and multi-positional perspective. Keeping this in mind, the researchers are now introducing various fusion strategies to develop multimodal HAR. In most of the works, single nature data (time-series) collected from accelerometer, and gyroscope are utilized for developing HAR based on fusion techniques [455]. The research gaps exist in developing fusion approaches for multi-nature data such as fusing the information of multi-sensory (time-series) and camera (visual) for multimodal HAR.

The remaining parts of the chapter are structured as follows. Section 5.2 presents the relevant literature of the most recent HAR systems in the view of single-modal and multimodal data using DL techniques for smart healthcare applications. Section 5.3 demonstrates the proposed fusion

architecture incorporating the multi-head CNN with CBAM, ConvLSTM, IoT deployment, and evaluation measures. Section 5.4 demonstrates the experimental specifications, the details of the dataset, the experimental findings, the comparison with the SOTA, and ablation studies. Lastly, the conclusion and potential future works are discussed in section 5.5.

5.2 Related Works

In recent years, wearable technologies are in ongoing development stage incorporating complicated and sensitive embedded sensors. In general, DL and ML algorithms are utilized to improve the efficiency of HAR using data retrieved from these sensors. This section describes the recent developments of HAR platforms in the area of DL, ML, and IoT. Here, we discussed a few most recent HAR approaches based on the single-modal and multimodal data.

5.2.1 Single-modal based HAR

Single-modal based HAR systems use only single source data for recognition purposes. Single-modal based HAR utilizes any sensor data such as accelerometer, gyroscope, or RGB camera [456]. The development of the single-modal based HAR platform is quite cost-effective but it lacks reliability and robustness.

Al-qaness et al. [457] proposed a multilevel residual network with attention called Multi-ResAtt for HAR from wearable sensors data. The system consists of a variety of parallel-oriented residual modules and initial blocks. A Bi-directional Gated Recurrent Unit (Bi-GRU) with attention blocks are preceded by two dense layers for HAR in this proposed system. The developed scheme is tested by three publicly available datasets and the best accuracy of 90.08% is found from the PAMAP2 dataset. Lu et al. [458] developed a wearable prototype to capture human activities from the environment through the use of one tri-axial accelerometer for HAR. The system extracted global and local features from the data to find out the effect of different perspectives of human activities. The extracted features are fed into several ML techniques to develop classification models and an average accuracy of 96% is obtained from RF leveraging both global and local features. Zhou et al. [423] introduced a semi-supervised learning architecture exploiting the concept of LSTM with deep Q-network to enrich the performance of HAR on weakly labeled sensor information. Deep Q-network automatically classifies the data with a reward based on distance to address the issue of lack of labeled data. The LSTM network recognized the fine-grained features contextually retrieved from sequential motion information. The experimental findings reveal that the presented scheme achieved a ROC curve up to 0.95.

In another research, Abdel-Basset et al. [459] presented a lightweight DL architecture for HAR utilizing the data from the wearable heterogeneous sensors. The collected sensor data is encoded into RGB images and the features are extracted using a hierarchical multi-scale extraction module. The evaluations are performed with two open-access datasets and the accuracy of 98% and 99% are obtained from UCI-HAR and Mobile HEALTH (MHEALTH) datasets respectively. However, the proposed framework did not mention about power consumption and inference latency in IoT environment. Further, Abdel-Basset et al. [460] used CNN, LSTM, and attention techniques to develop a dual-channel network called ST-DeepHAR for HAR. In this system, an adaptive channel-squeezing technique is introduced to fine-tune the feature extraction capacity of CNN by using multichannel dependency. The proposed framework is evaluated with two publicly available datasets and obtained the best accuracy of 98.9% from the WISDM dataset. Zhang et al. [461] developed a HAR platform based on multi-head techniques along with attention mechanisms. The high-level features are retrieved from the raw data exploiting multi-head CNN and fused to generate a single feature vector. Thirty parallel attention heads are exploited to select the most significant attributes for HAR. Besides, the system appraised an F-measure of 0.954 using the WISDM dataset with the parameters of 2.77 million. However, the complexity of the proposed network is relatively high.

5.2.2 Multimodal based HAR

Multimodal based HAR frameworks use multiple sensors to record real-world environmental data. Multimodal based HAR systems are merely deployed to classify ADL from a variety of data sources including sensors, and visuals [406]. The multimodal based frameworks depend on a fusion of sensors including accelerometers, gyroscopes, or depth cameras for data collection purposes.

Yadav et al. [434] proposed an activity recognition framework called ARFDNet utilizing pose estimation based classification architecture. The skeleton features of the individuals are extracted from the RGB videos using the pose estimation model. The skeleton features are preprocessed and fed into CNNs preceded by GRUs to learn the spatiotemporal dynamics. Finally, the classification is done in a FC layer using the outcome of the GRUs. The experimental findings revealed that the developed network obtained the best accuracy of 96.7% from the UP-Fall detection dataset. However, the system used only data from camera 2 (frontal view). Ramirez et al. [462] demonstrated a HAR system that used human skeleton features from raw images collected from a standard video camera. The scheme is able to detect the human falls as well as different types of activities for several individuals in the same scene through the use of human skeleton features. The system applied four different ML algorithms and the average accuracy of 98.59% is found from Camera 1 (lateral view) of the UP-Fall detection dataset. Inturi et al. [463] introduced a DL architecture

that is able to detect different kinds of human activities and falls. The system used human joint points that are captured by applying the AlphaPose pre-trained architecture. The retrieved key points are processed using CNNs to analyze the spatial correlation of the key points. The LSTM network preserved the long-term dependencies in this system. UP-Fall detection dataset is utilized to prove the efficiency of the developed architecture. An accuracy of 98.59% is achieved from the experimental results with the camera data only.

In another study, Lin et al. [464] proposed an adaptive multimodal fusion architecture for HAR using videos and inertial data. The skeleton sequence patterns are retrieved through the proposed Spatio-temporal graph CNN with adaptive loss function and the inertial data features are extracted using LSTM with a fully convolutional network. The contribution of two modalities at the decision level are learnt using the proposed adaptive learning technique. The architecture is tested with the data called H-MHAD collected from a laboratory environment along with two open-access datasets UTD Multimodal Human Action Dataset (UTD-MHAD) and Continuous Multimodal Human Action Dataset (C-MHAD). Moreover, the best accuracy of 91.18% is found from the C-MHAD dataset. Ranieri et al. [465] used a two-stream ConvNet for HAR improved with LSTM and a temporal CNN to analyze the temporal data on videos and inertial sensing devices. Here, the authors deployed two fusion strategies such as feature-level and late fusion to enrich the recognition performances. The experiments have been carried out with two publicly available popular datasets including egocentric multimodal and UTD-MHAD. The proposed network obtained the best accuracy of 85.47% from the UTD-MHAD dataset. Further, Ranieri et al. [466] presented a HAR platform utilizing different types of data such as videos, ambient sensors, and inertial units in an AAL environment. The main focus of this research is to generate a multimodal dataset called Heriot-Watt University/University of Sao Paulo (HWU-USP) activities dataset in a small kitchen. As a DL framework, the system used CNN and LSTM for recognition purposes and achieved an accuracy of 98.61% from the experiments. The framework is also evaluated with the UTD-MHAD dataset and found an accuracy of 92.33%. Furthermore, Gao et al. [467] developed a framework for recognizing human action that incorporated latent information from various viewpoints. The system proposed a category-level dictionary learning architecture based on adaptive fusion. The design of query sets and the development of the regularization scheme for the assignment of the adaptive weight were performed in order to incorporate dictionary learning. This method found excellent results, with an increase in accuracy of between 3% to 10%.

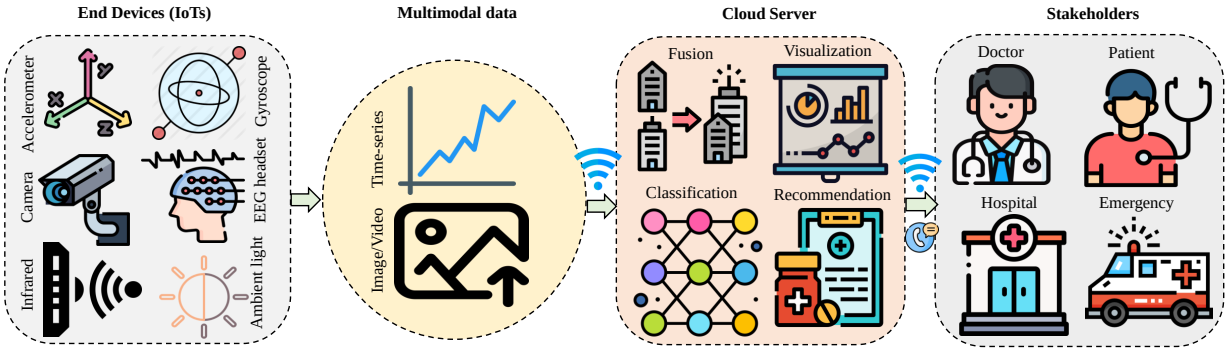


Figure 5.1: System architecture of the developed multimodal human activity recognition system in Internet of Healthcare Things. Here, the cameras and multiple sensors collect the human daily activities in the shape of visual and time-series. The deep learning-based fusion network is developed and deployed in the cloud server. The notification is sent to the stakeholders through 5G communication.

5.3 Proposed Hierarchical Architecture for Feature Fusion in Multimodal HAR

In this chapter, we propose a multi-level feature fusion architecture for multimodal human recognition in smart healthcare applications. The proposed system consists of several components such as the end IoT devices, DL-based fusion architecture, cloud server, 5G communication, and stakeholders. Fig. 5.1 shows the overall system architecture for multimodal HAR in IoHT. In our system, the end IoT devices include accelerometer, gyroscope, camera, Electroencephalogram (EEG) headset, infrared sensor, and ambient light sensor to collect the different ADLs in visual and time-series modalities. The fusion architecture is developed to be able to handle multimodal signals (visual and time-series) using multi-head CNN along with CBAM as well as ConvLSTM. The proposed DL model is deployed in the cloud server using large training data samples. The parameters of the fusion network are stored in the cloud to facilitate smart healthcare applications. All the data transfers from end IoT devices to the cloud server as well as the cloud server to stakeholders are done through Wi-Fi signals. The decision of the proposed fusion architecture is conveyed to the stakeholders through 5G communication so that they can take further actions.

The proposed multi-level feature fusion architecture contains several blocks to recognize human actives from the multimodal data. Fig. 5.2 illustrates the several blocks of the proposed architecture for multimodal HAR for smart healthcare applications. The data collected from multiple wearable sensors in the form of time-series are passed to ConvLSTM architecture to obtain the high-level temporal features. The visual information from cameras are sent to three-head

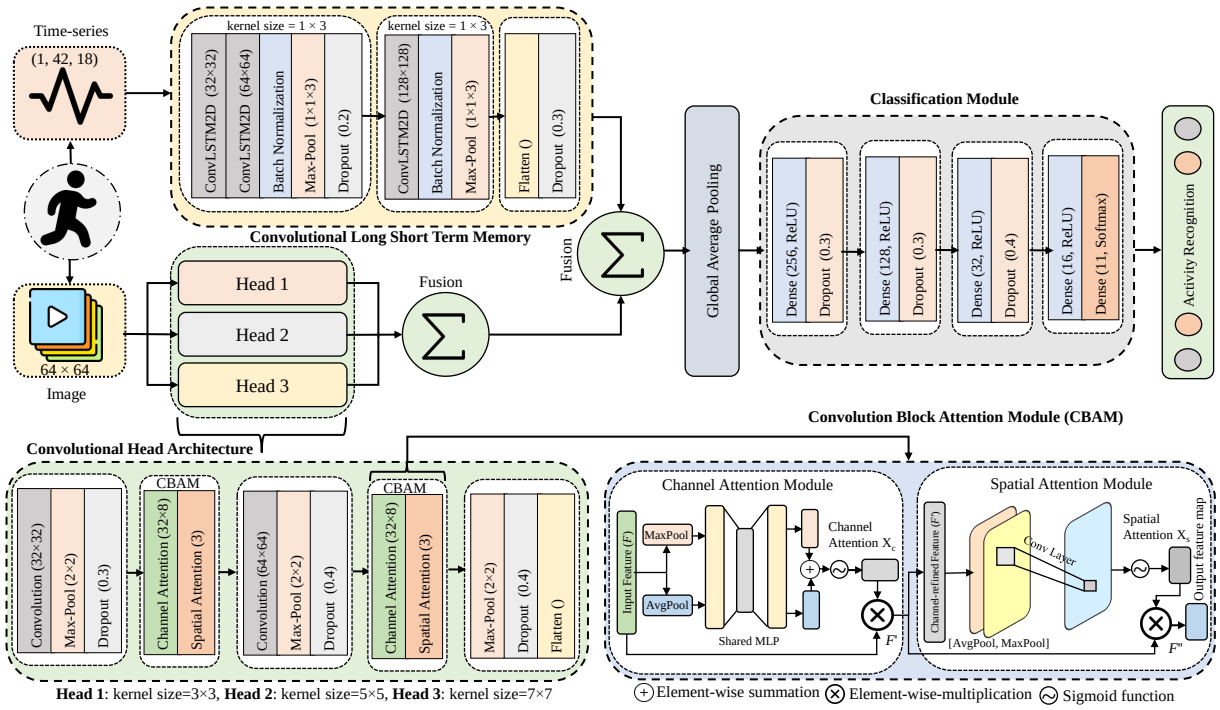


Figure 5.2: Schematic diagram of the developed fusion architecture for multimodal human activity recognition in a smart healthcare framework. The time-series data recorded from multi-sensors are fed into ConvLSTM architecture to get the temporal features (see top part). A three-head CNN architecture with CBAM extracts the high-level channel and spatial dimension features from the visual data and the outputs of each head are get fused (see lower part). Additionally, the output features of two-branch networks are fused and passed to the classification module through the GAP layer. Finally, the Softmax function is utilized to provide output for HAR.

CNN networks where each contains the CBAM network to extract the patterns from the channel and spatial dimensions. In this level, the outputs (homogeneous features) of each head are fused to get the high-level features from the raw images. In the next level, the extracted features from visual data from multi-head network and multi-sensor information from ConvLSTM architecture are fused to get the final features and are sent to FC layer through the GAP layer. The GAP layer generates one feature map for each class of human activities. Finally, the classification module containing FC layer with Softmax function recognizes the human activities. The following sections demonstrate the proposed fusion architecture with IoT deployment and its blocks in detail.

5.3.1 Multi-head Convolutional Neural Network

CNN [426] has become one of the most representative neural networks in the area of DL for image recognition. In this chapter, a three-head CNN (as illustrated in Fig. 2) is designed to retrieve the relevant patterns from the input images. The most pivotal component in CNN is the convolutional layer [468] that contains several convolutional filters which produce the output feature map from the input images through convolution operation. In the convolution layers, the output feature maps from the previous layer are convolved using a number of convolutional kernels. Moreover, a bias is used to enhance the output of the convolution operation that is passed through the activation function to generate the feature map for the following layer. Mathematically, the q^{th} feature map at the p^{th} layer of h^{th} head of multi-head CNN is a matrix, and the value at the x^{th} row is represented as $z_{pq}^{x,h}$. The value is calculated as in (5.1).

$$z_{pq}^{x,h} = f_{ReLU}\left(f_{conv2d}^{h}\left(z_{p-1}^{x+l}\right)\right) \qquad \forall_h = 1, 2, 3 \tag{5.1}$$

Here, f_{ReLU} is the activation function that substitutes all negative values with zero in the feature map, and f_{conv2d}^{h} represents convolution function of the h^{th} head in our multi-head CNN, as expressed in (5.2).

$$f_{conv2d}^{h}\left(z_{p-1}^{x+l}\right) = b_{pq} + \sum_{r} \sum_{l=0}^{n_p^h - 1} w_{pqr}^{l,h} z^{x+l,h}_{(p-1)m} \tag{5.2}$$

Here, the bias for a specific feature map is b_{pq}, r is the index of the feature maps at the $(p-1)^{th}$ layer, $w_{pqr}^{l,h}$ is the weight matrix at the position l of the convolution kernels, and n_p^h is the length of the kernel of the h^{th} head in our multi-head CNN.

Another important component in our proposed multi-head CNN architecture is the pooling layer [469] that minimizes the number of parameters and calculations by scaling down the spatial size of the feature description. The most popular and widely used pooling technique is called max pooling, which selects the largest component within each receptive area as shown in (5.3).

$$\mathcal{Y}_{kpq} = \max_{(a,b)\in\mathbb{R}_{p,q}} \mathcal{V}_{kab} \tag{5.3}$$

Here, $\mathcal{Y}_{kpq}$ denotes the pooling operation of k^{th} feature maps, and $\mathcal{V}_{kab}$ stands for the component at position (a, b) enclosed by pooling region $\mathbb{R}_{p,q}$ that represents a receptive field around the location (p, q).

In the proposed multi-level feature fusion architecture, two convolution operations with filter sizes of 32, and 64 are utilized to retrieve the high-level patterns from the input images. The size of the input data is 64×64 in this network. We consider the kernel size of 3×3, 5×5, and 7×7

in head 1, head 2, and head 3, respectively. Three max-pooling operations with pool size 2×2 are exploited to minimize the number of parameters and calculations. A dropout rate of 0.3, 0.3, and 0.4 are put after each max-pooling operation to avoid over-fitting during the training process.

5.3.2 Convolutional Block Attention Module

In the proposed architecture, each head contains two CBAMs (as shown in Fig. 2) to enhance the training performance by highlighting the channel and spatial features of the activity images. The Channel Attention Module (CAM) network enables the proposed architecture to focus on important channel features and ignores the other features. To measure the significance of each channel, various weight information is applied to diverse feature dimensions and feature channels of the visual data. The Spatial Attention Module (SAM) network allows the proposed architecture to give more attention to the spatial dimension information on the feature map. For feature extraction, CBAM [470] sequentially extracts a 1D channel attention map $X_c \in \mathbb{R}^{C\times 1\times 1}$ and a 2D spatial attention map $X_s \in \mathbb{R}^{1\times H\times W}$ from the given intermediate feature map $\mathcal{F} \in \mathbb{R}^{C\times H\times W}$ of the activity visual data. The overall attention mechanism is given in (5.4).

$$\begin{aligned} \mathcal{F}' &= X_c(\mathcal{F}) \otimes \mathcal{F} \\ \mathcal{F}'' &= X_s(\mathcal{F}') \otimes \mathcal{F}' \end{aligned} \tag{5.4}$$

Here, $\otimes$ presents the element-wise multiplication, and $\mathcal{F}''$ is the final refined feature. During the multiplication, the channel attention features are compressed along the spatial dimension, and vice-versa.

The CAM network (as shown in Fig. 2) enhances the weight of the relevant information and suppresses the weight of the unnecessary information in the feature channel. Thus, the proposed architecture highlights more on the discriminative channels in the activity images. In the CAM network, the average-pooled patterns, and the max-pooled attributes are extracted from the aggregated feature map of spatial information by exploiting both average-pooling and max-pooling operations. The extracted high-level patterns are sent to a shared MLP model that contains one hidden layer. Moreover, the output of the shared network is forwarded through a pipeline containing further max-pooling and average-pooling operations, and a non-linear activation function ReLU to generate channel attention map $X_c \in \mathbb{R}^{C\times 1\times 1}$. The use of two pooling operations facilitates to extract high-level features efficiently. The channel attention is mathematically calculated as in (5.5).

$$X_c(\mathcal{F}) = \sigma(\text{MLP}(\text{AvgPool}(\mathcal{F})) + (\text{MLP}(\text{MaxPool}(\mathcal{F}))) \tag{5.5}$$

Here, σ represents the sigmoid function.

The SAM network (as depicted in Fig. 2) improves the spatial dimension features in the feature map by performing feature filtering on the pixels at various positions in the same spatial dimension and weighting the significant features. The SAM performs the max-pooling and average-pooling operations on the given feature map F' along the channel dimension to get two feature maps. The retrieved features are fused and convolved by a convolution layer with a kernel size of 7×7 to generate the final spatial attention map $X_s \in \mathbb{R}^{1 \times H \times W}$. The mathematical formulation of the SAM is given in (5.6).

$$X_s(\mathcal{F}') = \sigma(f^{7\times7}([\text{AvgPool}(\mathcal{F}'); \text{MaxPool}(\mathcal{F}')])) \tag{5.6}$$

Here, $f^{7\times7}$ is the convolution operation with a filter size of 7×7.

5.3.3 Convolutional Long Short Term Memory

ConvLSTM architectures [428] are an improvement of the CNN, RNN, and LSTM networks. RNN networks transmit historical information through chain network structures. However, as the distances between the chain network structures become larger, the long-term information becomes harder for the RNN networks to learn. LSTM networks improve this condition of RNN via adding three different gate functions that control memory information. However, LSTM's

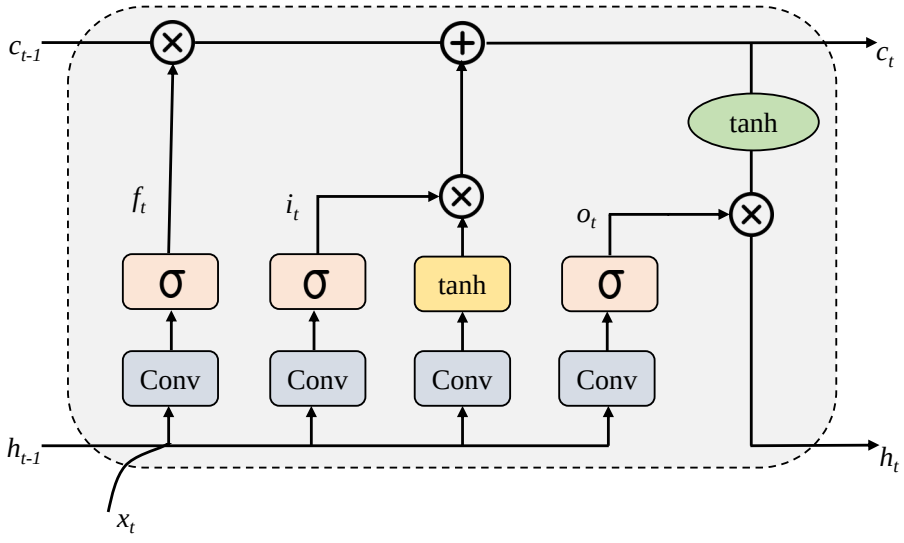


Figure 5.3: The architecture of ConvLSTM. The new memory c_t and output h_t are formed by updating the internal memory c_{t-1} based on the current input x_t and the prior output h_{t-1}.

use FC networks which leads to information redundancy. LSTM networks only keep temporal dependencies but do not take spatial dependencies into account.

CNNs are generally a combination of convolutional layers and pooling layers that can extract spatial information from input data modalities but cannot extract temporal dependencies. ConvLSTM networks embed convolution operations inside LSTM cells, thus alleviating the redundancy of FC layers and resulting in fewer parameters compared to LSTM networks due to weight sharing features of convolutional kernels. ConvLSTM networks can extract both the spatial and temporal dependencies from input data modalities as they combine both CNN and LSTM's. ConvLSTM networks have better spatiotemporal comprehension ability compared to fully connected LSTM networks. The architecture of ConvLSTM is shown in Fig. 5.3. ConvLSTM networks can be expressed using the following major operations.

ConvLSTM networks have three kinds of gates: input, forget, and output. The input gates can be formulated using (5.7).

$$i_t = \sigma(W_{xi} * x_{(i)} + W_{hi} * h_{(t-1)} + W_{ci} \odot c_{(t-1)} + b_i) \tag{5.7}$$

Here, the input features are denoted by $x_{(1)}, x_{(2)}, ..., x_{(n)}$. The cell outputs are represented as $c_{(1)}, c_{(2)}, ..., c_{(n)}$. The hidden cells are denoted as $h_{(1)}, h_{(2)}, ..., h_{(n)}$. The input, forget, and output gates are expressed by $i_{(t)}$, $f_{(t)}$, and $o_{(t)}$. $W_{x\sim}$ and $W_{h\sim}$ represents the 2D convolutional kernels. The convolution operation in all subsequent representations is expressed by $*$. The Hamdard product is expressed by $\odot$.

Equation (5.8) represents the formulation of the forget gates.

$$f_{(t)} = \sigma(W_{xf} * x_{(t)} + W_{hf} * h_{(t-1)} + W_{cf} \odot c_{(t-1)} + b_f) \tag{5.8}$$

The cell states are calculated using the input and forget gates. Equation (5.9) presents the formulation for calculating the cell states.

$$c_{(t)} = f_{(t)} \odot c_{(t-1)} + i_{(t)} \odot \tanh(W_{xc} * x_{(t)} + W_{hc} * h_{(t-1)} + b_c) \tag{5.9}$$

The mathematical expression of the output gates require the cell states. The formulation for calculating the output gates is shown in (5.10)

$$o_{(t)} = \sigma(W_{xo} * x_{(t)} + W_{ho} * h_{(t-1)} + W_{co} \odot c_{(t)} + b_o) \tag{5.10}$$

Equation (5.11) formulates the calculation of the hidden states. The cells, hidden states, and three gates are considered as 3D tensors.

$$h_{(t)} = o_{(t)} \odot \tanh(c_{(t)}) \tag{5.11}$$

In our proposed fusion architecture, the input shape for the ConvLSTM stream is (1,42,18). The input contains data collected from multiple types of sensors in the shape of time-series. Three layers of ConvLSTM were used inside the stream with filter sizes of 32, 64, and 128. In all the ConvLSTM layers, the kernel size was 1×3 and ReLU was used as the activation function. We used Batch normalization, max-pooling layers, and dropout layers to ensure the model does not overfit or underfit the input data. The max-pooling layer has a kernel size of $1 \times 1 \times 3$. We used 20% and 30% dropout in the dropout layers.

5.3.4 Classification Module

The output vectors from ConvLSTM network and multi-head CNN architecture are fused and sent into the classification module to predict the target classes through the GAP layer. The GAP layer creates one feature map of each class of human activities and minimizes the number of trainable parameters. The FC layers contain the neurons of 256, 128, 32, and 16 with ReLU activation function, and the FC layers are followed by dropout layers to prevent over-fitting. Finally, the Softmax activation function calculates the class score of each activity and outputs the correct activity class that obtained the higher probability score. The prediction representation of the Softmax function is formulated as in (5.12) and (5.13), where, $\mathcal{F}$ presents the output features from the previous FC layer and a is the number of activities in the data samples (N).

$$\mathcal{P} = \text{Softmax}(\mathcal{F}) = \frac{exp(\mathcal{F})}{\int_1^a exp(\mathcal{F})} \tag{5.12}$$

$$\hat{y} = \text{argmax}(\mathcal{P}) \tag{5.13}$$

Moreover, the cross-entropy loss was used to reduce the loss value during training. The loss function ($\mathcal{L}$) is demonstrated as in (5.14), where, y_i is the actual activities, and $\hat{y}_i$ is the predicted activities obtained by the proposed architecture.

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right] \tag{5.14}$$

5.3.5 IoT Deployment

Once the fusion architecture is trained and evaluated, it is ready to be deployed as an IoT application. In this scenario, the proposed system would be deployed in the cloud. Individual

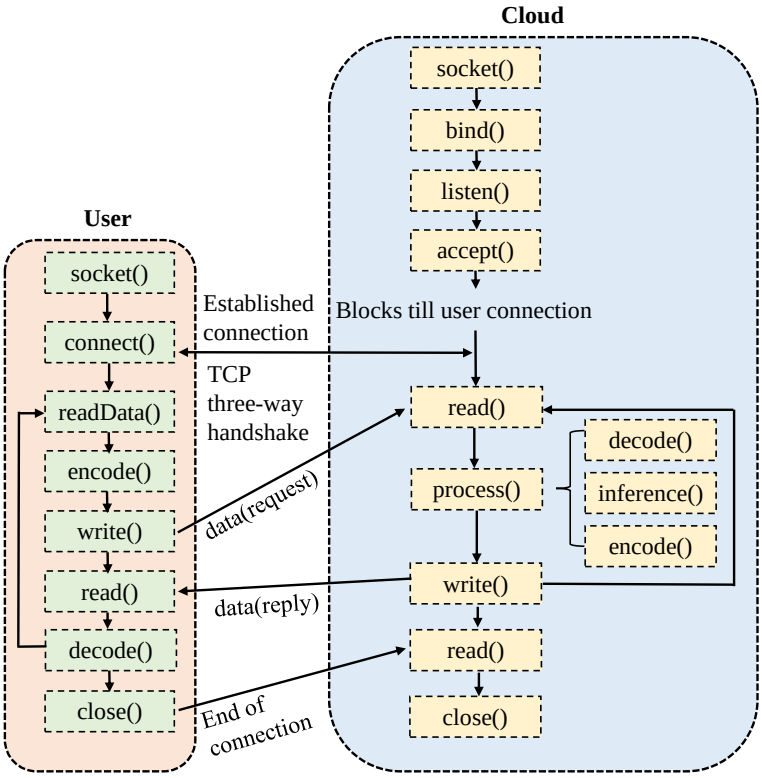


Figure 5.4: Block diagram of user application and cloud server application for interaction for multimodal human activity recognition in IoHT.

user applications can then connect to the cloud server that is serving the model, send data to the server, and receive inference results. The first possible deployment network scenario is where the cloud server and users are in the same network. In this case, the latency between the users and server would be minimal. This scenario is feasible for large nursing homes, hospitals, and care facilities that want to provide localized monitoring services to their customers. The second possible deployment network scenario is where the cloud server and users are in different networks. This scenario is more feasible for corporations that want to provide Deep Learning as a Service (DLaaS). Although the latency between the cloud server and user application would be much larger than in the same network scenario, the different network configuration is more scalable. The cloud server would be more sophisticated, have specialized system architecture, have load balancing, and redundancy to ensure minimal downtime. Fig. 5.4 provides a block diagram of the processes involved in the interaction between the user applications and cloud service. The connection and data transfer requires sockets. Generally, the cloud service has a fixed IP and port

number. The user applications must know the fixed IP address and port number to connect to the cloud service. The internet protocol used in the network connection and data transfer is TCP. TCP provides three-way handshake, retransmission in case of transmission failure, error detection, network congestion avoidance, and improved security over other internet protocols such as UDP. TCP protocol has slightly higher latency than UDP. However, TCP provides much more security-oriented features than UDP. After a connection is established, the client reads data continuously from the sensors, encodes them, and sends them via the network to the cloud server. The server receives the data, decodes it, and performs inference. The server then encodes the inference results, sends it to client, and waits for further data. The client decodes the inference results, takes action based on the predicted activity, and starts sending the next batch of data. The network connection can be closed at any time from the client or server. The diagram shown in Fig. 5.4 depicts a single-thread network. When the number of clients and servers will be large, multithreading and threadpool executor blocks would be required to efficiently complete the tasks.

5.3.6 Performance Evaluation

In this chapter, we measure the efficiency of the developed fusion architecture in terms of four widely used statistical metrics: accuracy, precision, recall, and F1-Score. Additionally, to evaluate the performance of the fusion architecture for multimodal HAR in IoHT environment, we measure two metrics including latency of the deployed IoT network and response time of the proposed fusion network. The lower value of these parameters represents the better performance of the system.

- **Latency**: Latency refers to the time taken for a packet or data message to reach the destination from its point of origin. The low latency network ensures real-time applications with minimal delay times.
- **Response time**: Response time includes latency as well as the time taken for the server to perform inference on the received data. It is mathematically demonstrated as follows.

$$Response\ time = 2 \times latency + processing\ time \tag{5.15}$$

5.4 Experiments Results and Discussions

In this section, we describe the experimental findings of our developed fusion technique for multimodal HAR to show the efficacy of the proposed system. This section includes a brief description

of the experimental setup, the multimodal dataset for the experiments, the quantitative results analysis, and the ablation study. Moreover, a comparative study is conducted to compare our findings with two baseline models as well as recent SOTA methods.

5.4.1 Experimental Setup

The developed multi-level feature fusion architecture for multimodal HAR was written in Python (v2.10) and implemented with the Keras API on top of the TensorFlow (v2.8.0) backend. The training was conducted in a desktop computer equipped with GPU: GTX 1070 with 8 GB GDDR5 memory, CPU: Intel® Core™ i7-6700K CPU @ 4.00GHz × 8, and RAM: 32 GB DDR4. The computer was running in an Ubuntu operating system with a version of 18.04 LTS.

5.4.2 Dataset Description

We used a large and publicly available multimodal dataset called UP-Fall detection dataset [431] for our experimentation. To collect the ADLs data, 17 healthy participants (9 males and 8 females) were considered between 18 to 24 years, with an average height and weight of 1.66 m and 66.8 kg respectively. The dataset comprises 11 human activities including 6 basic ADLs and 5 different kinds of fall activities. The human activities and falls along with the duration of each activity

Table 5.1: Description of the activities performed by human participants in UP-Fall detection dataset.

Activity Number	Description	Duration (s)
$A1$	Falling forward using hands	10
$A2$	Falling forward using knees	10
$A3$	Falling backwards	10
$A4$	Falling sideward	10
$A5$	Falling sitting in empty chair	10
$A6$	Walking	60
$A7$	Standing	60
$A8$	Sitting	60
$A9$	Picking up an object	10
$A10$	Jumping	30
$A11$	Laying	60

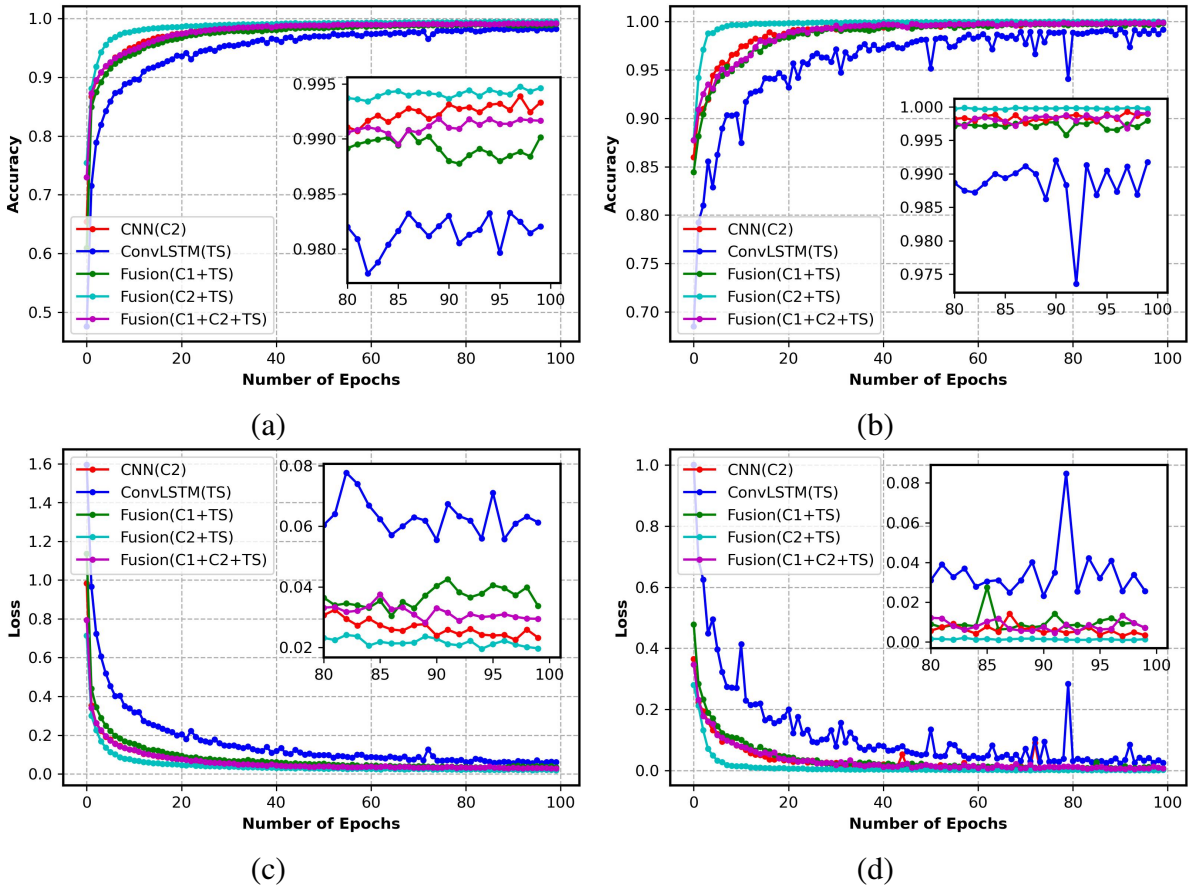


Figure 5.5: Performance measures (accuracy and loss curves) of the developed fusion approach along with two baseline architectures. (a)-(c) Training (b)-(d) Validation. The first row represents the accuracy, and the second row depicts the loss.

stored in this dataset are given in Table 5.1. The duration of all fall events is the same (60 s) and the duration of daily activities is different. Each activity in the dataset was recorded 3 trials by each healthy participant. The data collected from the wearable sensors and vision devices are in the shape of time-series and visual respectively.

In our research, we used all modalities of data to conduct the experiments. We considered trial 1 as well as trial 2 of each activity as the training set and trial 3 of each activity as the testing set. The image data is resized to 64×64 from 640×480 to feed into the network. The data collected from wearable sensors are in time-series format and the total number of attributes is 42. We generated the sequences from the raw time-series data to process them using ConvLSTM. In

the experiment, the sequence size is 18 depending on the smallest number of data points.

5.4.3 Quantitative Results Analysis

We conducted the experiments for 100 epochs to train and evaluate the proposed multi-level feature fusion architecture. We performed Grid Search analysis to select the best hyperparameters for our architecture. Based on Grid Search results, the learning rate is set to 0.001, the optimizer is Adam, the batch size is 128, and the loss function is sparse categorical cross-entropy. In the experiments, two baseline architectures such as CNN and ConvLSTM are considered to get the results from the visual and time-series data individually. Moreover, we run three variants of our proposed fusion architecture based on the nature of the dataset. The CNN ($C2$) is run only with the visual data from camera 2 as the previous studies [434] depicted that the existing architectures showed robust performance for the frontal view of the camera data. The time-series data is fed into ConvLSTM (TS) to achieve the relevant attributes from the raw multi-sensor information as well as the performance of this architecture for time-series data. Three variants of fusion architecture are Fusion ($C1 + TS$), Fusion ($C2 + TS$), and Fusion ($C1 + C2 + TS$) where the $C1$, $C2$, and TS represent the data from camera1 (lateral view), camera 2 (frontal view), and time-series data from multi-sensor, respectively. Besides, we utilized 20% of the training dataset as a validation set to validate the proposed fusion architecture. In the experimental studies, we presented the average performance of ten runs while the weights of the architectures are randomly initialized during each run. We also depicted the best performance from the output of the run that scores the highest performance.

Fig. 5.5 (a)-(b) depicts the relationship between the number of epochs and training as well as validation accuracies of our proposed multimodal HAR architecture. From Fig. 5.5 (a)-(b), it is evident that the training processes stop at nearly 100 epochs. It is observed from the experimental findings that the baseline model ConvLSTM takes more time to generalize as it contains the time-series data only. Another baseline model CNN achieves comparatively better performance compared to ConvLSTM as it has experimented with only visual data. Our fusion architecture containing the visual data from camera 2 (frontal view) along with all time-series data (42 attributes) is generalized well in both training and validation phases compared to two other use cases of our fusion architectures such as fusion for camera 1 data with time-series information and fusion for all data. More time is required for fusing all data as it comprises a relatively large amount of training data samples.

Besides, the loss curves are illustrated in Fig. 5.5 (c)-(d) to show the relationship between epochs and losses (both training and validation) for five use cases. Similar to Fig. 5.5 (a)-(b), the fusion approach containing data from camera 2 and time-series achieves fast convergence with a

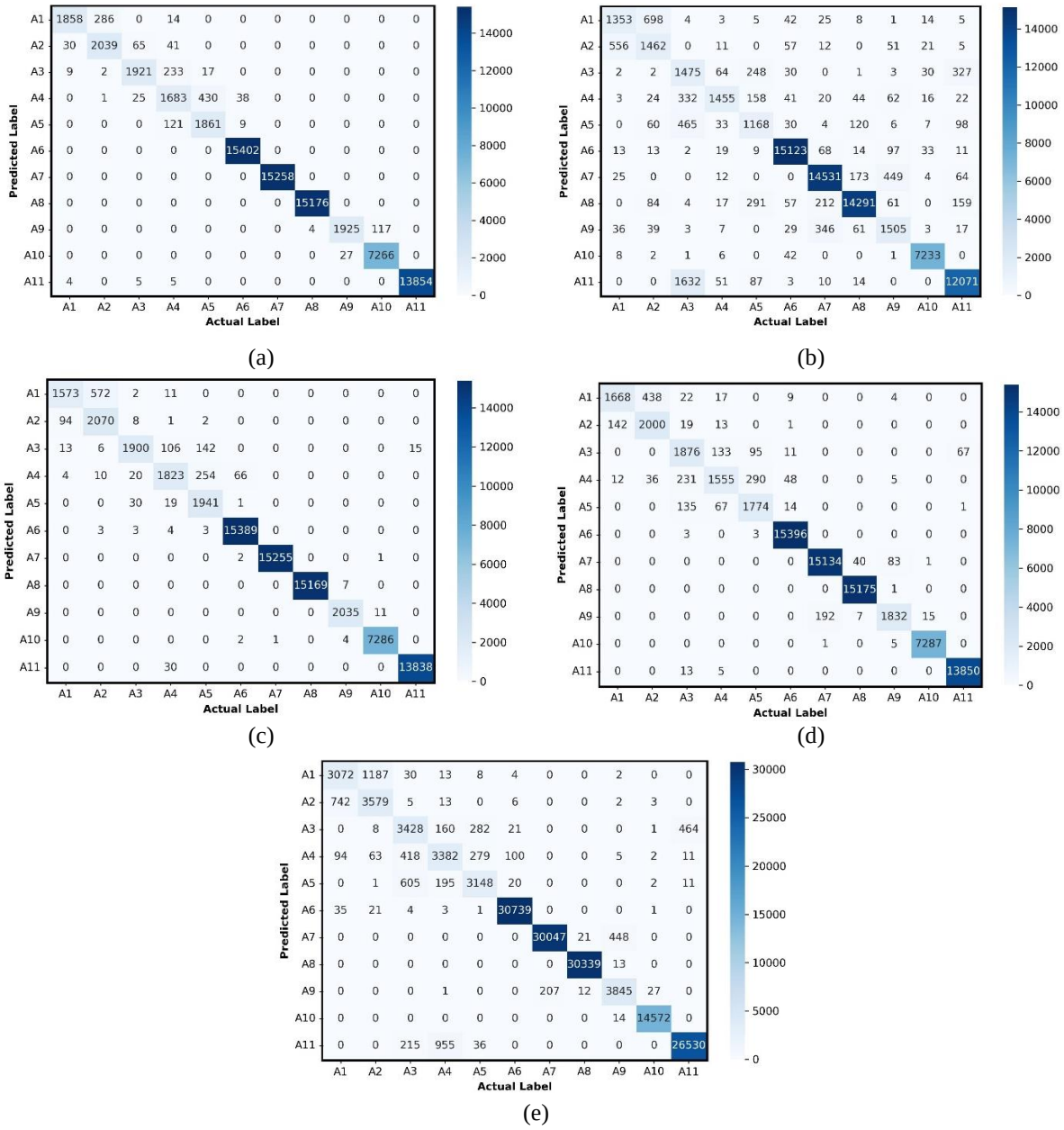


Figure 5.6: Confusion matrices for the developed fusion network along with two baseline architectures. The rows (activity label) represent the actual activities, and the columns (predicted label) show the recognized activities. The deeper color presents the better results. (a) CNN ($C2$) (b) ConvLSTM (TS) (c) Fusion ($C1+TS$) (d) Fusion ($C2+TS$) (e) Fusion ($C1+C2+TS$).

smaller number of epochs. The loss values of the other two use cases of fusion architecture such as Fusion ($C1 + TS$) and Fusion ($C1 + C2 + TS$) is relatively low but it is slightly higher than the Fusion ($C2 + TS$) architecture. The baseline model CNN obtains relatively low loss values compared to another baseline model ConvLSTM both in the training and validation set.

Fig. 5.6 represents the confusion matrices obtained by our proposed fusion architecture for multimodal HAR. It is found from Fig. 5.6 that the fusion architecture with camera 2 data and time-series information recognizes the activities more efficiently than the other two fusion architecture variants. The baseline models misclassified a large number of activities compared to fusion architectures. It is observed from the fusion architecture, the activities $A1$ to $A5$ are achieving comparatively less scores than other activities. The highest recognition rate is obtained for the activity standing. The falling activities ($A1$ to $A5$) and the activity picking up an object ($A9$) which is just like a fall activity receive a relatively low score for activity recognition. In general, falling events consist of transitioning from standing to laying activities. In our dataset, a few of the first frames depict standing while a few of the final frames show laying. Hence, the proposed fusion architecture shows poor generalization for recognizing the fall activities. Moreover, more than 99% (0.99) of walking, standing, sitting, jumping, and laying are recognized while the lowest recognition rate of 73% (0.73) is obtained from the falling activity: falling forward using hands. In this architecture, falling forward using hands is misidentified as falling forward using knees as these behaviours are intimately associated and intertwined. Our fusion network learned that both activities are almost similar which demonstrates the higher error on the activity falling forward using hands.

Table 5.2 demonstrates the results of the developed fusion architecture along with baseline networks for UP-Fall detection dataset for each activity with respect to precision. Notably, the developed architecture discriminates the activities with more than 98% of precision for fusing the camera 2 data and time-series information. The lowest precision rate of 80.79% is obtained for activity $A2$ from CNN ($C2$), 48.20% for the activity $A3$ from ConvLSTM (TS), 74.70% for the activity $A3$ from Fusion ($C1 + TS$), 80.89% for the activity $A2$ from Fusion ($C2 + TS$), and 60.40% is achieved for the activity $A3$ from Fusion ($C1 + C2 + TS$). The lowest precision rate contains between activities $A2$ to $A3$. In some cases, the highest precision of 100% is found from the experiments for the activities $A6 - A8$, and $A10 - A11$ considering the baseline models as well as fusion architectures.

Similar to Table 5.2, the class-specific recall of the proposed fusion architecture along with baseline models are illustrated in Table 5.3. Here, the average recall rate of more than 98% is found from the experiments for Fusion ($C2 + TS$). The highest recall rate of 100% is achieved from the activities $A6 - A8$, and $A11$ using CNN ($C2$), ConvLSTM (TS) architecture receives the best recall rate of 99.20%, Fusion ($C1 + TS$), and fusion of all data achieve 100% recall rate for only two activities ($A8$, and $A10$), and the best recall rate is found from the Fusion ($C2 + TS$)

Table 5.2: Percentile precision (precision % $\pm$ standard deviation) of the proposed fusion architecture for each activity classification.

Activity	CNN ($C2$)	ConvLSTM (TS)	Fusion ($C1 + TS$)	Fusion ($C2 + TS$)	Fusion ($C1 + C2 + TS$)
$A1$	96.300 ± 0.037	72.800 ± 0.051	86.000 ± 0.062	87.700 ± 0.060	80.299 ± 0.064
$A2$	80.799 ± 0.066	64.700 ± 0.027	74.800 ± 0.058	80.899 ± 0.037	70.100 ± 0.065
$A3$	87.300 ± 0.054	48.200 ± 0.079	74.700 ± 0.042	93.599 ± 0.031	60.400 ± 0.078
$A4$	83.500 ± 0.031	76.800 ± 0.069	86.299 ± 0.029	90.399 ± 0.032	78.600 ± 0.086
$A5$	82.900 ± 0.025	56.700 ± 0.077	82.300 ± 0.027	83.500 ± 0.032	80.500 ± 0.042
$A6$	100.00 ± 0.000	97.100 ± 0.007	99.900 ± 0.003	99.900 ± 0.003	99.600 ± 0.004
$A7$	100.00 ± 0.000	91.799 ± 0.022	98.900 ± 0.003	100.00 ± 0.000	99.100 ± 0.003
$A8$	100.00 ± 0.000	93.499 ± 0.035	99.900 ± 0.003	100.00 ± 0.000	99.800 ± 0.004
$A9$	98.600 ± 0.012	63.199 ± 0.095	90.800 ± 0.045	99.100 ± 0.008	86.300 ± 0.061
$A10$	98.000 ± 0.000	98.400 ± 0.006	100.00 ± 0.000	100.00 ± 0.000	100.00 ± 0.000
$A11$	100.00 ± 0.000	95.199 ± 0.009	99.000 ± 0.007	98.000 ± 0.000	98.200 ± 0.008

Table 5.3: Percentile recall (recall % $\pm$ standard deviation) of the proposed fusion architecture for each activity classification.

Activity	CNN ($C2$)	ConvLSTM (TS)	Fusion ($C1+TS$)	Fusion ($C2+TS$)	Fusion ($C1+C2+TS$)
$A1$	78.200 ± 0.091	63.400 ± 0.072	67.699 ± 0.099	79.099 ± 0.063	67.000 ± 0.085
$A2$	85.900 ± 0.082	74.000 ± 0.064	87.500 ± 0.064	89.100 ± 0.054	84.000 ± 0.067
$A3$	88.599 ± 0.025	63.100 ± 0.050	82.999 ± 0.024	80.099 ± 0.043	78.400 ± 0.027
$A4$	78.900 ± 0.044	59.100 ± 0.059	71.600 ± 0.037	83.200 ± 0.033	70.300 ± 0.065
$A5$	94.100 ± 0.025	59.700 ± 0.046	83.700 ± 0.047	96.300 ± 0.022	78.000 ± 0.015
$A6$	100.00 ± 0.000	97.900 ± 0.005	99.900 ± 0.003	100.00 ± 0.000	99.800 ± 0.004
$A7$	100.00 ± 0.000	92.700 ± 0.024	99.869 ± 0.006	100.00 ± 0.000	97.800 ± 0.013
$A8$	100.00 ± 0.000	93.200 ± 0.016	100.00 ± 0.000	100.00 ± 0.000	100.00 ± 0.000
$A9$	93.200 ± 0.007	70.500 ± 0.040	90.100 ± 0.017	98.200 ± 0.011	93.100 ± 0.024
$A10$	99.800 ± 0.004	99.200 ± 0.004	100.00 ± 0.000	100.00 ± 0.000	100.00 ± 0.000
$A11$	100.00 ± 0.000	87.600 ± 0.010	99.800 ± 0.004	100.00 ± 0.000	93.899 ± 0.011

Table 5.4: Percentile F1-Score (F1-Score % ± standard deviation) of the proposed fusion architecture for each activity classification.

Activity	CNN ($C2$)	ConvLSTM (TS)	Fusion ($C1 + TS$)	Fusion ($C2 + TS$)	Fusion ($C1 + C2 + TS$)
$A1$	85.900 ± 0.046	67.300 ± 0.037	75.000 ± 0.061	82.900 ± 0.021	72.400 ± 0.036
$A2$	82.800 ± 0.043	69.000 ± 0.031	80.399 ± 0.036	84.700 ± 0.023	76.100 ± 0.037
$A3$	87.900 ± 0.033	54.000 ± 0.054	78.499 ± 0.027	86.199 ± 0.024	66.799 ± 0.048
$A4$	80.899 ± 0.024	66.300 ± 0.038	78.100 ± 0.020	86.299 ± 0.016	73.799 ± 0.054
$A5$	88.200 ± 0.010	58.100 ± 0.058	83.000 ± 0.026	89.400 ± 0.018	79.100 ± 0.020
$A6$	100.00 ± 0.000	97.300 ± 0.004	99.900 ± 0.003	100.00 ± 0.000	99.900 ± 0.003
$A7$	100.00 ± 0.000	92.199 ± 0.017	98.700 ± 0.004	100.00 ± 0.000	98.500 ± 0.008
$A8$	100.00 ± 0.000	93.399 ± 0.019	100.00 ± 0.000	100.00 ± 0.000	99.800 ± 0.004
$A9$	95.700 ± 0.009	66.199 ± 0.062	90.299 ± 0.023	98.600 ± 0.004	89.399 ± 0.035
$A10$	98.900 ± 0.003	98.800 ± 0.004	100.00 ± 0.000	100.00 ± 0.000	100.00 ± 0.000
$A11$	100.00 ± 0.000	91.200 ± 0.007	99.300 ± 0.004	99.200 ± 0.004	96.000 ± 0.006

Table 5.5: Activity recognition results of the proposed fusion approach along with baseline models for the test samples.

Architectures	Average				Best			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
CNN ($C2$)	97.766 ± 0.002	97.878 ± 0.001	97.766 ± 0.002	97.749 ± 0.002	98.13	98.20	98.133	98.14
ConvLSTM (TS)	89.081 ± 0.005	89.834 ± 0.006	89.081 ± 0.005	89.278 ± 0.005	89.89	91.17	89.89	90.34
Fusion ($C1 + TS$)	96.491 ± 0.004	96.625 ± 0.003	96.491 ± 0.004	96.451 ± 0.004	97.26	97.29	97.26	97.22
Fusion ($C2 + TS$)	97.902 ± 0.001	97.986 ± 0.002	97.902 ± 0.001	97.880 ± 0.001	98.26	98.30	98.18	98.25
Fusion ($C1 + C2 + TS$)	94.980 ± 0.006	95.492 ± 0.005	94.980 ± 0.006	95.096 ± 0.006	95.75	96.15	95.75	95.80

that obtains 100% recall for five activities ($A6-A8$, and $A10-A11$). Considering the fusion and baseline models, the lowest recall of 59.10% is appraised in ConvLSTM architecture.

Moreover, Table 5.4 illustrates the F1-Score of the proposed fusion approach along with the baseline models for each activity recognition. Here, the fusion architecture with camera 2 data along with time-series information achieves the best F1-Score (100%) for four activities ($A6-A8$, and $A10$) whereas the CNN ($C2$) model obtains 100% F1-Score for four activates ($A6-A8$, and $A11$) but it differs by only one activity. The experimental results reveal that the average F1-Score of 97.90%, 89.10%, 96.50%, 98.00%, and 94.89% are achieved from CNN ($C2$), ConvLSTM (TS), Fusion ($C1+TS$), Fusion ($C2+TS$), and Fusion ($C1+C2+TS$) architectures, respectively.

Considering the class-wise performances (precision, recall, and F1-Score) of multimodal HAR, it is evident that the performances of human falls are relatively low compared to other ADLs. It could happen due to the nature of the fall events as it consists of standing and laying activities in UP-Fall detection dataset.

Further, the activity recognition results of the test samples for the proposed fusion approach and two baseline models (CNN and ConvLSTM) are shown in Table 5.5. Two baseline models: CNNs and ConvLSTM are experimented with the visual data from camera 2 and time-series data respectively. The baseline ConvLSTM model achieved a classification accuracy of 89.08%, while the normal CNN obtained relatively better results with a recognition accuracy of 97.76% which almost reaches the recognition level of the fusion architecture. Among the three variants of fusion architecture, the best performances such as accuracy of 96.90%, precision of 97.98%, recall of 96.90%, and F1-Score of 97.88% are obtained from Fusion ($C2+TS$). While fusing all data, the performances have been somewhat decreased as there are some scenarios with almost the same information. For example, the data from camera 1, and camera 2 contain almost identical activity scenarios but they only differ by view perspectives. The proposed fusion architecture considering the frontal view of the visual data along with time-series information outperforms its corresponding counterparts such as Fusion ($C1+TS$), and Fusion ($C1+C2+TS$) by an accuracy improvement of $\sim$1.40%, and $\sim$2.90%, respectively. Moreover, this model suppresses the performance of the baseline architectures (CNN and ConvLSTM) with an accuracy enhancement of $\sim$0.15%, and $\sim$8.80%, respectively.

Fig. 5.7 illustrates the box plots of the performance measures obtained from ten runs of the proposed fusion approach and baseline models for deeper analysis. In all the cases, the CNN ($C2$) and Fusion ($C2+TS$) models outperformed all the other models. In all of the performance metrics, the median values for the Fusion ($C2+TS$) model were better than the CNN ($C2$) model. In all cases, the ConvLSTM (TS) model performed the worst. The Fusion ($C1+TS$) and Fusion ($C1+C2+TS$) models had some outlier values in terms of precision. All the performance measures indicate that the accuracy and reliability are better in the proposed fusion approach

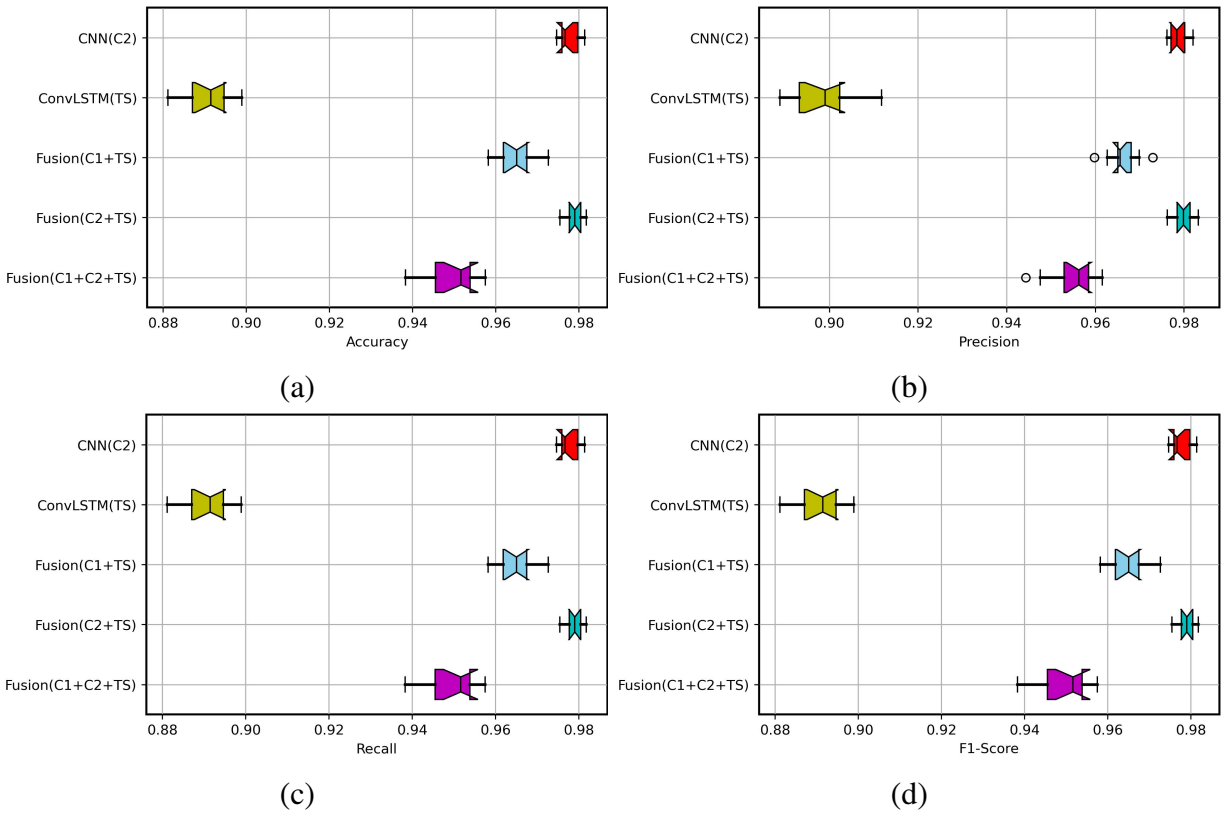


Figure 5.7: Boxplot for the comparative analysis of the proposed fusion architecture with the baseline models. (a) Accuracy (b) Precision (c) Recall (d) F1-Score.

with camera 2 and time-series data compared to others.

Fig. 5.8 presents the Grad-CAM for each activity of our proposed Fusion ($C2 + TS$) model. Grad-CAM is an AI explainability method that visualizes the regions of the input image that the model finds interesting while performing inference. Grad-CAM produces a coarse localization map of the regions that the model deems "important". On all the Grad-CAM of the randomly chosen activity images, the model highlighted the persons. By focusing on the person, their activity can be determined. Thus, the Grad-CAM images indicate that the proposed model is correctly focusing on the correct subject in the images and inferring their activities.

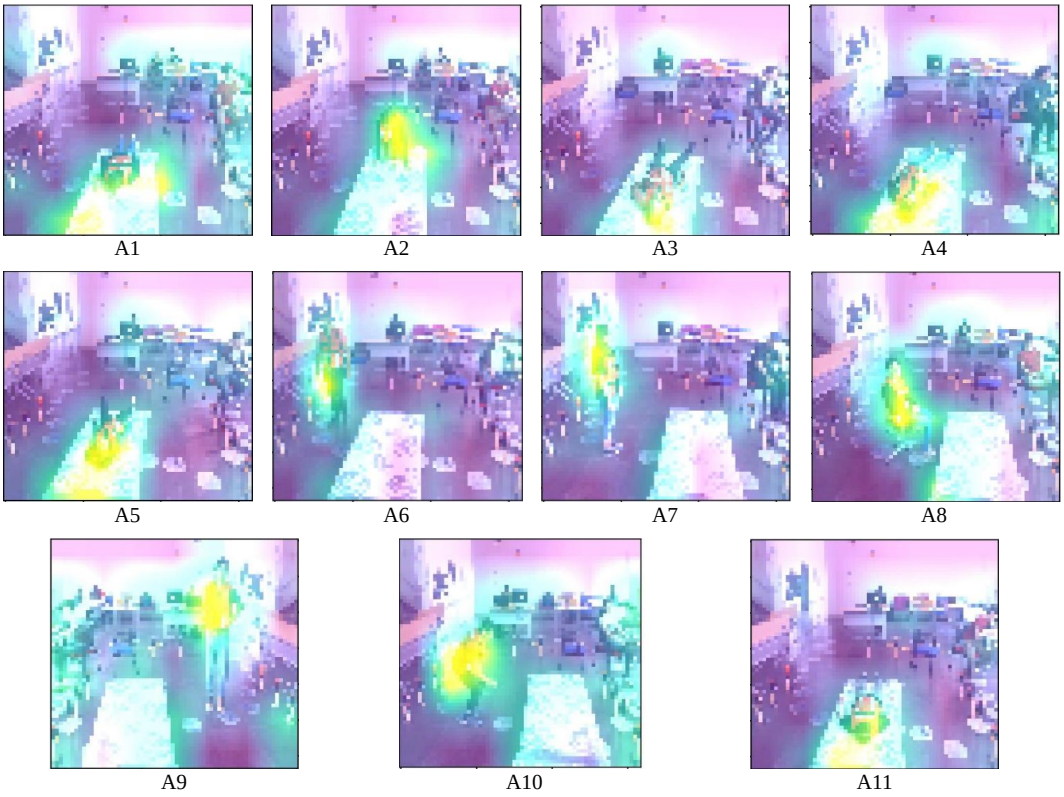


Figure 5.8: Grad-CAM of each activity sample on UP-Fall detection dataset for the proposed Fusion ($C2+TS$) architecture.

5.4.4 Performance in IoT Environment

We evaluated the performance of the models in two IoT deployment scenarios: a smart home environment which represents the same network use case and a virtual care environment which represents the different network use case. The performance in IoT environment was evaluated on the basis of two metrics: latency and response time. Latency and response time depend on various factors including overall network use by other applications, client and server hardware capabilities, as well as the hardware and software configurations of the network. While testing the model performances, we closed all background applications and services.

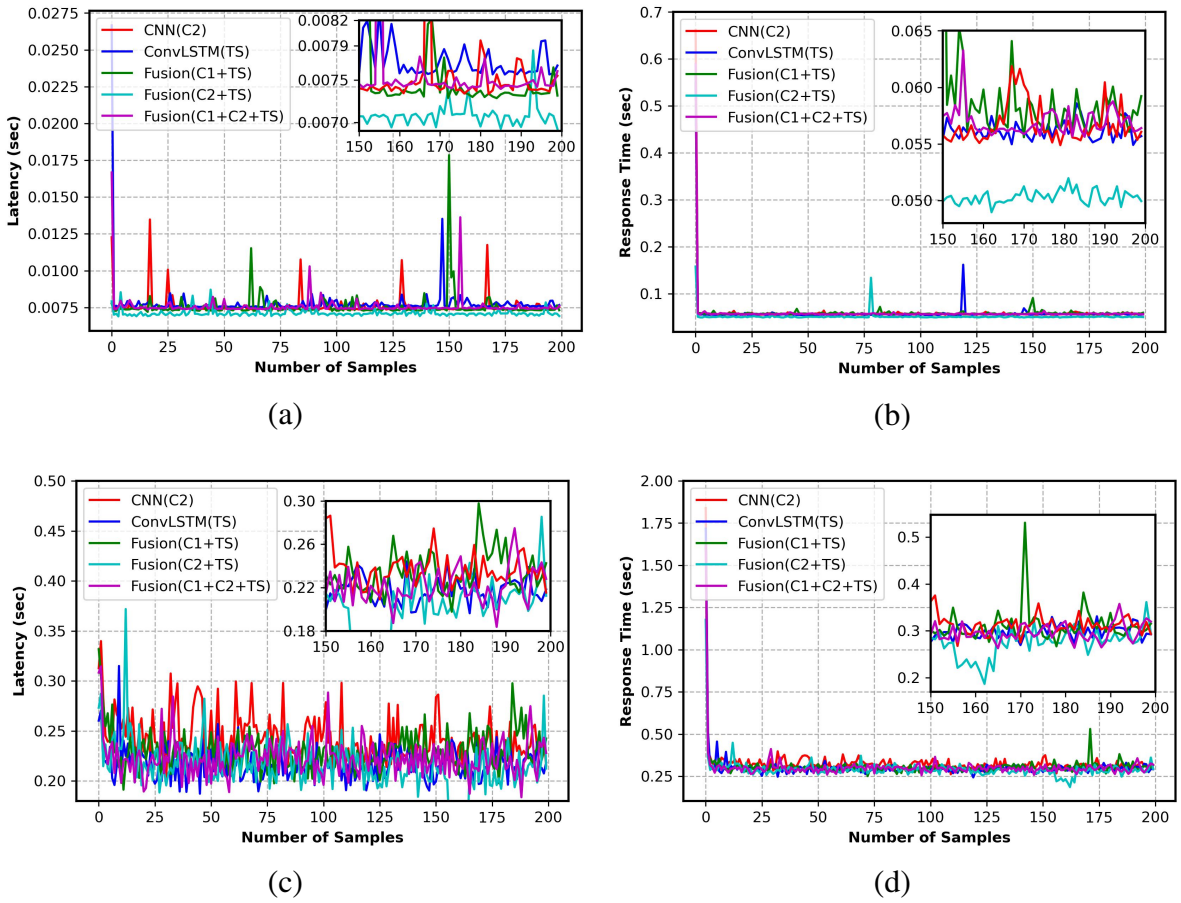


Figure 5.9: Performance in IoHT deplyment for multimodal HAR. (a)-(c) Latency (b)-(d) Response time. The first row represents the performance in smart home environment, and the second row depicts the performance in virtual care environment.

Smart Home Environment

Fig. 5.9 (a)-(b) presents the performance of the models in a smart home environment. In this case, the server and the users are on the same network. On average, the latency of the models is 7.5 ms. The response time of the models is 5.5 ms on average. The fusion model trained on image data from camera 2 and time-series data had marginally lower latency and response time compared to the other models. On average, the Fusion ($C2 + TS$) model had 0.50 ms lower latency and 5 ms lower response time than the other models.

Virtual Care Environment

Fig. 5.9 (c)-(d) presents the performance of the models in a virtual care environment. In this scenario, the users and the server are in a different network. The latency of the models is 220 ms on average. The response time of the models is 300 ms on average. All the models, in this case, had similar latency and response time. The latency and response time for the models are higher than the smart home environment results. Latency and response time in different networks are almost always higher than in the same network. This is due to multiple factors such as the distance of the server from the clients, network firewalls, and network configurations. Modern PaaS providers provide multiple services such as servers in multiple global locations, rapid prototyping and deployment, and workload scaling to mitigate some of these issues.

5.4.5 Comparison with the State-of-the-Art

We compare our findings obtained by the fusion architecture with the SOTA to highlight the competitive advantages of our developed approach for multimodal HAR. Table 5.6 illustrates the performance comparisons of our proposed architecture with the recent studies on UP-Fall detection dataset. We consider the recently published methods possibly 2018 or above as the SOTA techniques. Similar to comparative studies conducted in [424], [434], all the performance measures of these SOTA are retrieved from the corresponding articles. The authors in [432] used only the data from multiple wearable sensors and deployed several ML algorithms. However, the best recall rate of 69% is obtained using SVM. Only the visual data is utilized in [424], [431], [434], where the system developed in [431] used the data from camera 1, the camera 2 data is utilized in [434], and the architecture presented in [424] exploited both the lateral and frontal views. It is found that the lowest F1-Score of 71.20% is obtained from [431] using CNN, the best recall rate of 97.95% is achieved in [424] utilizing CNN, and the highest accuracy of 96.70% is appraised in [434] with the use of CNN + GRU. In the SOTA study, only the system introduced in [433] deployed fusion architecture (CNN + LSTM) and used the data from camera 1 as well as wearable sensors. Although the accuracy measure (96.40%) of this framework is good, the other performance measures are not quite remarkable. In our proposed fusion architecture, we used the full set of data of UP-Fall detection dataset. We obtained the best accuracy, precision, and F1-Score of 97.90%, 97.98%, and 97.88%, respectively from the experiments using the frontal view of the camera data, and the full set of wearable sensors information. The recall rate of our developed system is little bit lower than the framework developed in [424].

Table 5.6: Performance comparison of the results obtained by the proposed fusion approach and the recent studies on UP-Fall detection dataset.

Refs.	Data	Data Sources	Architectures	Accuracy	Precision	Recall	F1-Score
[432]	Time-series	Wearable sensors	SVM	–	66%	69%	67%
[431]	Image	Camera 1	CNN	95.10%	71.80%	71.30%	71.20%
[433]	Image + Time-series	Camera 1 and Wearable sensors	CNN + LSTM	96.40%	84.20%	81.50%	82.30%
[424]	Image	Camera 1, 2	CNN	95.64%	96.91%	**97.95%**	97.43%
[434]	Image	Camera 2	CNN + GRU	96.70%	96.90%	96.70%	96.60%
Ours	Image + Time-series	Camera 1 and Wearable sensors	Fusion ($C1 + TS$)	96.49%	96.62%	96.49%	96.45%
		Camera 2 and Wearable sensors	Fusion ($C2 + TS$)	**97.90%**	**97.98%**	97.90%	**97.88%**
		Camera 1, 2 and Wearable sensors	Fusion ($C1 + C2 + TS$)	94.98%	95.49%	94.98%	95.09%

5.4.6 Ablation Studies

To thoroughly measure the efficiency of our proposed multi-level feature fusion architecture, we conducted several extensive ablation studies considering the different variants of the developed framework. The details description of the models considered for ablation experiments is as follows.

Single-Head

In this experiment, we just considered the single head CNN model instead of three-head from our proposed fusion architecture to see the effects of head architectures for feature extraction. The kernel size of this head architecture is 3×3.

Two-Head

In this model, two-head CNN architecture is considered to extract features from the visual data. The kernel size of 3×3, and 5×5 are considered while omitting the other head architecture with kernel size 7×7.

No Attention

The proposed fusion architecture is modified in such a way that the CBAM network is omitted from the developed framework. In this ablation experiment, the extracted features from ConvLSTM and three-head CNN architecture are directly fused rather than passing through the CBAM architecture.

Only Channel Attention

In the proposed fusion architecture, the SAM network is omitted to perform this experiment. The impacts of channel dimension features are evaluated through this experiment.

Only Spatial Attention

The CAM network of the CBAM architecture is deleted from the proposed fusion approach to perform this study. With this experiment, it is realized how the proposed architecture performs without the channel dimension features.

Table 5.7: Experimental findings of different variants of the proposed fusion architecture on UP-Fall detection dataset.

Architectures	Average				Best			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Single-head	74.566 ± 0.046	82.844 ± 0.029	74.566 ± 0.046	75.689 ± 0.0456	81.23	89.03	81.23	83.15
Two-head	97.444 ± 0.003	97.556 ± 0.002	97.444 ± 0.003	97.418 ± 0.003	97.83	97.93	97.83	97.81
No attention	87.841 ± 0.007	89.651 ± 0.005	87.841 ± 0.007	88.346 ± 0.006	88.81	90.55	88.81	89.09
Only channel attention	96.491 ± 0.004	96.625 ± 0.003	96.491 ± 0.004	96.451 ± 0.004	98.18	98.32	98.18	98.17
Only spatial attention	97.358 ± 0.009	97.612 ± 0.006	97.358 ± 0.009	97.358 ± 0.008	98.02	98.08	98.02	97.99

The results obtained by the different variants of the proposed architecture in ablation studies are shown in Table 5.7. All the experiments in the alation studies are conducted using the frontal view of the visual data and the time-series data from multiple sensors. Considering the single-head and two-head architecture, the lowest performance (accuracy: 74.56%) is achieved by the single-head architecture compared to the performance (accuracy: 97.44%) of two-head architecture. This has happened as the two-head CNN architecture extracts the relevant features using different kernels, compared to the single-head architecture. The next ablation study is conducted to show the impacts of the attention network on activity recognition performance. The model with no attention block obtained quite low performance approximately ∼10% less accurate than our proposed fusion architecture. The reason for the low performance is that the attention module put more attention on the significant features and removes the unnecessary attributes. Without the use of the attention module, all the features are sent to the classification block through the fusion network; thus decreasing the performance. Lastly, we conducted two other ablation experiments considering the CAM architecture and SAM architecture separately. The fusion architecture with SAM only works slightly better compared to CAM on average. The possible reason that could be mentioned in this scenario is that the SAM network contains several convolution operations while the CAM module consists of a shared MLP network. Consequently, the SAM network retrieved more efficient and high-level features than the CAM architecture.

5.5 Conclusion

HAR has become a prominent research area due to the fast development of various emerging computing paradigms and technologies as well as plays a crucial role in IoHT applications by monitoring the physical activities of the individuals. In this chapter, we proposed a multi-level feature fusion network incorporating the concept of multi-head CNN along with an attention mechanism called CBAM as well as ConvLSTM network for multimodal HAR. This chapter enhances the efficiency of multimodal HAR for smart healthcare applications by exploiting ConvLSTM capabilities to model long-term temporal representation from raw multi-sensory information and using multi-head CNN with CBAM to effectively retrieve channel and spatial dimension features from visual information. The extensive experimental findings demonstrated the efficiency and practicability of our proposed fusion architecture compared with two baseline models and SOTA frameworks. The efficiency of the developed framework in IoT platform indicated that this system is quite effective for real-world IoHT applications.

Chapter 6

Multimodal Emotion Recognition in Healthcare through Model-Level Fusion

6.1 Introduction

Emotion is a complicated condition that integrates feelings, ideas, and behaviour, and it is the physiological responses of people to either internal or external stimuli. Human emotions have a significant impact on human daily lives through their simultaneous effects on a person's psycho-physiological and physiological state [471], [472]. It is essential for communication, perception, and decision-making. The task of identifying emotions in multimodal signals is crucial and complicated in the context of applications involving Human-Computer Interaction (HCI). There are many potential uses for emotion recognition systems across a range of areas, such as e-health, e-learning, recommender systems, entertainment, and security [473], [474]. In the healthcare system, emotion recognition has been getting more attention recently. The researchers are currently working to develop intervention strategies that can accurately identify emotion to monitor patients' health status [475], [476]. By monitoring the physical states of the patients through emotion recognition, it is possible to take immediate actions in a connected healthcare framework.

The task of emotion recognition is carried out using various data modalities, such as EEG, speech, gestures, texts, and facial expressions [477], [478]. Moreover, a few of the physiological data [479], [480] including HR, Electromyogram (EMG), ECG, and Skin Conductance Level (SCL) represent the autonomous nervous system that originates from the human physiological system for different emotion categories. Among the data modalities, facial expressions [481], [482] are commonly utilized to recognize human emotions. However, emotion recognition through facial images suffers from different factors such as light intensity, background changes,

and the distance between the user and the sensing devices that make it difficult to identify hidden or unexpressed emotions. Additionally, biosensors [483] are another choice to recognize the emotion that captures and analyzes the users' brain electrical activities. The primary challenge with biosensors for emotion recognition is that the signals captured by various sensors are often contaminated with noise. Moreover, the speech data modality [484], [485] is gaining more attention for emotion recognition nowadays. However, the available datasets for speech signals are small, and most of them contain a single or local language. To solve the issues related to the single data modalities, the current research focuses on multimodal signals [486], [487] to develop more robust and reliable architectures for real-time applications. Although various data combinations are possible, facial expressions fused with physiological signals (EEG, ECG, SCL, etc.) are widely utilized in HCI applications to monitor patients in connected healthcare. In general, emotion recognition from multimodal signals yields better results when robust architectures are developed and used. However, recognizing emotions using single-modal approaches requires less processing time [488].

Over the decades, several frameworks have been introduced for emotion recognition using ML and DL approaches. A few of the systems [489], [490] exploited traditional ML algorithms including SVM, Hidden Markov Model, and Gaussian Mixture Model for emotion recognition from different data modalities. Due to the limited feature extraction capabilities of ML approaches, DL techniques [491], [492] are growing trends to enhance the efficiency to recognize emotion nowadays. The majority of the researchers are focusing on DL approaches for emotion classification as they can extract the most significant features from the raw input signals, which assist in enhancing recognition performance. CNNs are mostly used DL architectures in image classification-related tasks due to their robustness and effectiveness in retrieving high-level features [493], [494]. Hence, the vision-related emotion recognition tasks involve CNNs to efficiently identify emotion from facial expressions. To process the sequential data, including the physiological signals along with speech, the RNNs, as well as LSTM architectures, show significant contributions [495], [496]. Moreover, attention mechanisms along with the DL models are the recent trends for emotion recognition that focus on task-relevant instances [497], [498]. The attention mechanism ensures the different levels of weights to various temporal and spatial features in the data samples. The extraction of representative multi-sensory attributes and the definition of dynamic interactions among different data modalities are focused on emotion recognition from multimodal signals. In this scenario, DL techniques along with attention mechanisms are the best fit to extract representative modality features.

The remainder of the chapter is structured as follows: Section 6.2 presents the most recent research in recognizing emotions from multimodal signals using ML and DL techniques. Section 6.3 describes the proposed DL-enhanced fusion technique to identify the emotion with a detailed description of DSCNN, Bi-LSTM, and soft attention mechanisms. The experimental findings are

elaborated in section 6.4, including the experimental setup, the dataset collection and preparation, the results analysis, and the comparison with the recent literature. The chapter is concluded in section 6.5, which outlines possible directions for future research.

6.2 Related Works

Human emotion recognition systems have become a more popular research topic using multimodal signals, including physiological, speech, video/image, and texts, due to the rapid growth of modern technologies such as DL, ML, computer vision, and the IoT [499], [500]. To develop frameworks for emotion recognition, a few systems exploited single modal data such as physiological/speech/video/texts, and some of the systems utilized a different combination of multimodal signals. In this section, we presented a few current works that are pertinent to the field of emotion recognition.

Ayata et al. [501] developed an emotion-aware healthcare platform to recognize emotion from multimodal physiological signals. The framework developed a decision-level fusion technique utilizing several ML methods such as SVM, RF, and Logistic Regression. The statistical features from the physiological data, including fingertip temperature, photoplethysmography, and respiratory belt of the DEAP dataset are extracted, and the minimum redundancy maximum correlation method is exploited to select the significant patterns. The best accuracies of 73.08% and 72.18% are obtained from the experiments considering arousal and valence emotions. Zhang et al. [502] proposed a framework to monitor the patients in the IoMT platform through the emotional information of the patients. The multimodal data of the patients are collected through several IoMT devices in a surveillance environment. In this research, a modified Weber local descriptor is presented to extract the significant patterns from the retrieved signals and the emotions are identified through the proposed hierarchical deep cognitive architecture. The patients in the IoT environment are monitored through Bluetooth and 5G technology. From the experimental evaluations, the authors achieved an accuracy of 88.1% for their constructed dataset that contains the audio and video data of three types of emotions (happy, pain, and normal). The authors of [503] presented an automated system in smart healthcare to identify emotion from audio and visual data. The system exploited 2D CNN to process the speech data and the video signals by a 3D CNN architecture. The pitch-synchronous peak-amplitude feature vector extracted the features from the speech data in the preprocessing stage. The obtained attributes from the two data streams are combined through extreme learning machine architectures. The proposed architecture is tested with two popular benchmarks (eNTERFACE05 and BAUM-2) and one private dataset. The accuracy of 87.02% and 56.12% are found from the eNTERFACE05 and BAUM-2 datasets through

experiments. Moreover, the system developed an edge computing framework to facilitate real-time evaluations.

Xie et al. [504] introduced an emotion recognition system using wavelet transform features and ML techniques. In this work, the authors exploited three types of physiological signals of Bio Vid Emo dataset. Sequential forward selection and sequential backward selection techniques are utilized to select the relevant features captured from the raw signals through the wavelet transform approach. Moreover, the system investigated how different fusion strategies would affect the outcome of the proposed framework. From the experimental outcomes, the developed method with decision fusion achieved an accuracy of 94.81% by combining three physiological signals using SVM. However, the framework did not use DL algorithms to extract more significant features. Mostafa et al. [505] presented a framework to recognize emotion from facial expression features using several ML methods and RNNs. The developed system extracted 68 facial feature points from local attributes by initially detecting the faces from the videos. Moreover, the geometrically based features are retrieved through facial expressions. The best accuracy of 82% is achieved using RNN architectures to distinguish between anger and disgust emotions from the Bio Vid Emo dataset. Barkur et al. [506] introduced an emotion recognition framework called EnsembleWave based on an end-to-end ensembled approach. The developed architecture used attention wavenet to retrieve global contextual patterns and manually extracted features to integrate a more extensive field of view and seize dataset-independent features. The extracted features from two parallel networks are combined to recognize emotion from the raw data. The presented scheme appraised an accuracy of 58%, 44%, and 68% on three widely used datasets such as EMO-DB, CREMA-D, and SAVEE. However, this work did not examine the covariance between the selection of features and emotion categories.

Recently, a few of the works have been performed based on multimodal data. Zhongmin et al. [507] presented a multimodal emotion recognition based on the Deep Belief Network (DBN) concept to fuse and optimize the retrieved features from the physiological and video data. In this research, the discriminative features of videos are represented by a bimodal DBN, and the significant features are retrieved from video and physiological signals through another BDBN network. For multimodal emotion recognition, a linear SVM technique is exploited lastly. The experimental outcome outperformed the existing literature with an accuracy of 80.89% for the BioVid Emo DB dataset. Kumar et al. [508] proposed a framework called MEmoR for intelligent health prediction in the industry by analyzing people's emotional status. In this work, the frames are generated from the videos, and a pre-trained architecture called ResNet50 is fined-tuned for face recognition to determine emotion. Moreover, the physiological modalities are trained utilizing CNN architecture. Lastly, decision-level weighted fusion is exploited to combine the data obtained from the two modalities in order to recognize the emotion. The proposed architecture appraised an accuracy and F1-Score of 81.54% and 79% from the experimental findings. How-

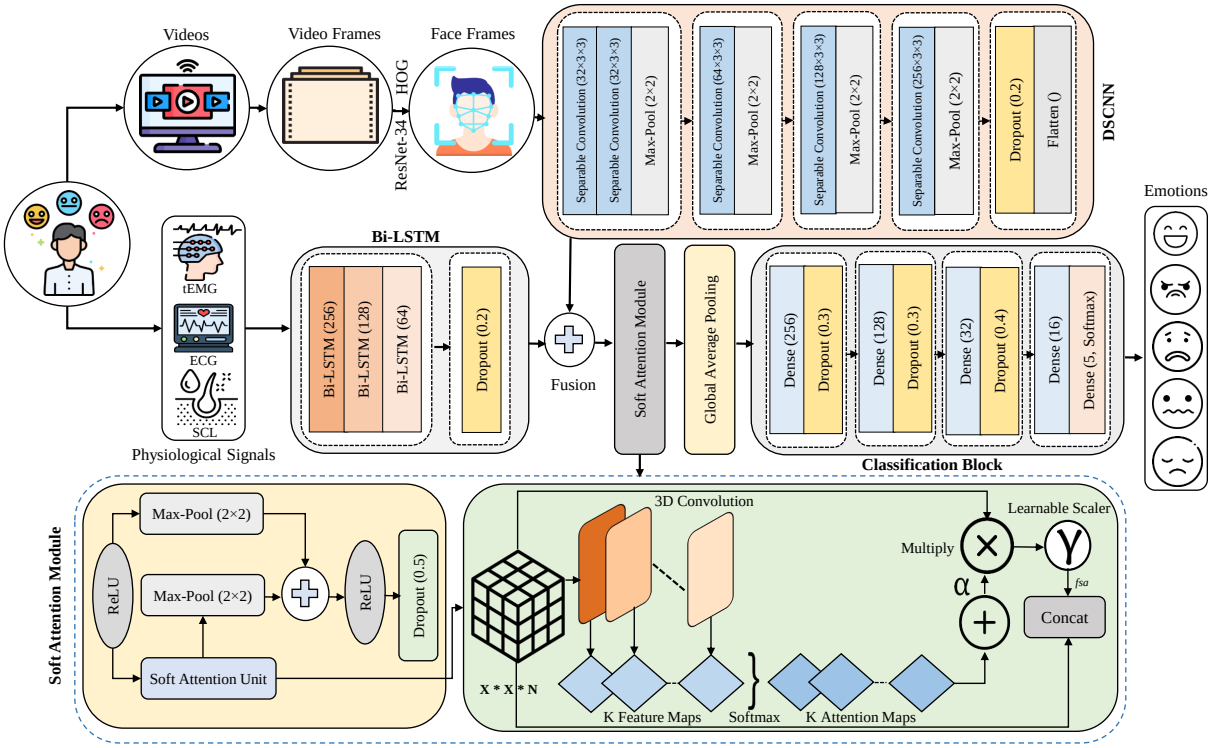


Figure 6.1: The proposed model-level fusion technique for multimodal emotion recognition in healthcare analytics. The data from the humans are collected in the form of videos, and physiological signals. Later, the Bi-LSTM network extracts the features from the physiological signals and the DSCNN architecture retrieves the significant patterns from the faces. Finally, the soft attention mechanism extracts the most essential features from the fused attributes that are passed to the classification block over the GAP layer for emotion recognition.

ever, this study did not consider the data of the dynamic environment. Zhao and Chen [509] introduced a fusion architecture to detect emotions from facial cues and EEG data. The significant features from the facial data are extracted by a Bilinear Convolution Network (BCN), and the EEG were split into three frequency bands to be input into the BCN architecture. Finally, a fusion technique is proposed to fuse the obtained attributes from the two modalities through a three-layer Bi-LSTM with an attention mechanism. The developed system is evaluated with popular benchmark datasets, and the highest accuracies of 86.8% and 86.2% are achieved for arousal and valence emotions from the DEAP dataset.

6.3 Proposed Model-Level Fusion Approach for Emotion Recognition

We propose a novel model-level fusion approach based on DL for enhanced emotion recognition from multimodal data to ensure reliable outcomes for smart healthcare applications. Figure 6.1 illustrates the system architecture of the proposed fusion approach. The developed framework is composed of numerous modules, including face detection, DSCNN architecture, Bi-LSTM technique, soft attention technique, GAP layer, and emotion classification block. The data are collected in two forms: videos captured using cameras and physiological data recorded through various sensors. From the videos, the frames are generated. Later, we used HOG and RestNet-34 techniques to identify the face from the data frames as the face data contains the emotion. Then, the DSCNN architecture receives the face data and generates significant features maps. On the other hand, the Bi-LSTM model receives physiological inputs, including Trapezius Electromyogram (tEMG), ECG, and SCL to extract temporal attributes. The networks receive the collected data from the sensory devices in a synchronized way. Lastly, the soft attention module receives the fused feature maps generated by the two DL algorithms. The soft attention technique ignores the comparatively less significant attributes, which prioritizes the most crucial features. The GAP layer converts the feature vectors into confidence maps to identify emotion. Finally, emotion recognition is carried out in the classification block, which includes the softmax function for multi-class emotion recognition. The proposed fusion technique is described in detail below.

6.3.1 Depthwise Separable Convolutional Neural Networks

DSCNNs [510] are popular DL architectures for image classification problems in real-time mobile and embedded systems due to low memory and lower computing power requirements. The depthwise separable convolution operation factorizes a standard convolution into a depthwise and point-wise convolution.

The depthwise separable convolution [511] accomplishes minimal filtering by applying only one filter to each input channel. Considering that only the input channel for each filter is computed, the total number of computations is thus decreased. The operation of depthwise separable convolution is expressed in (6.1).

$$\mathcal{G}(y,x,j) = \sum_{\upsilon=1}^{\mathcal{K}} \sum_{\upsilon=1}^{k} \mathcal{K}(\upsilon,\upsilon,j) \times I(y+\upsilon-1, x+\upsilon-1, j) \tag{6.1}$$

Here, $\mathcal{K}$ represents the depthwise convolution kernels having the size of $k \times k \times c_{in}$ where c_{in} is the number of input channels. The n_{th} filter in $\mathcal{K}$ is enforced to the n_{th} channel in I to generate the n_{th} channel of the filtered result of convolutional feature $\mathcal{G}$.

Moreover, the point-wise convolution [512] aggregates the outputs of depthwise convolution for generating a new feature map by applying 1×1 convolution. The mathematical calculation of the point-wise convolution is performed in (6.2).

$$O(y, x, l) = \sum_{j=1}^{c_{in}} \mathcal{G}(y, x, j) \times P(j, l) \tag{6.2}$$

Here, the size of the 1×1 convolutional kernel is $1 \times 1 \times c_{in} \times c_{out}$, where c_{out} represents the number of output streams.

Comparing DSCNN to standard CNN, the number of parameters is drastically reduced, as shown in (6.3) and (6.4).

$$C_{DSCNN} = k^2 \cdot c_{in} \cdot h \cdot w + c_{in} \cdot c_{out} \cdot h \cdot w \tag{6.3}$$

$$C_{CNN} = k^2 \cdot c_{in} \cdot c_{out} \cdot h \cdot w \tag{6.4}$$

Here, C_{DSCNN} represents the cost of computation for the depthwise convolution and 1×1 point-wise convolution. Moreover, h represents the height, and w represents the width of the input feature vector I.

In the proposed fusion technique, we consider the image size of 128×128 as input of the DSCNN architecture. Initially, we performed two subsequent separable convolutions with a filter size of 32 followed by a 2×2 max-pooling operation. Later, the filter size of 64, 128, and 256 are applied for every convolution following a max-pooling operation with size 2×2. In each convolution, we select a 3×3 sized kernel. Lastly, a dropout with the percentage of 0.2 is added to avoid the chance of overfitting of the network.

6.3.2 Bi-directional Long-Short Term Memory

The Bi-LSTM architecture [513] is frequently employed to handle sequential data of time instances and resolves the gradient explosion and information morphing problems. This network has two different LSTMs that can encapsulate data from both the forward and backward ways and then combine the information emerging from both directions [514]. The Bi-LSTM architecture can simultaneously consider the data samples' past and future information, considering each sequence's forwarding and reversing state. The final output is then generated by integrating the two hidden states. The general architecture of the Bi-LSTM network is depicted in Fig. 6.2.

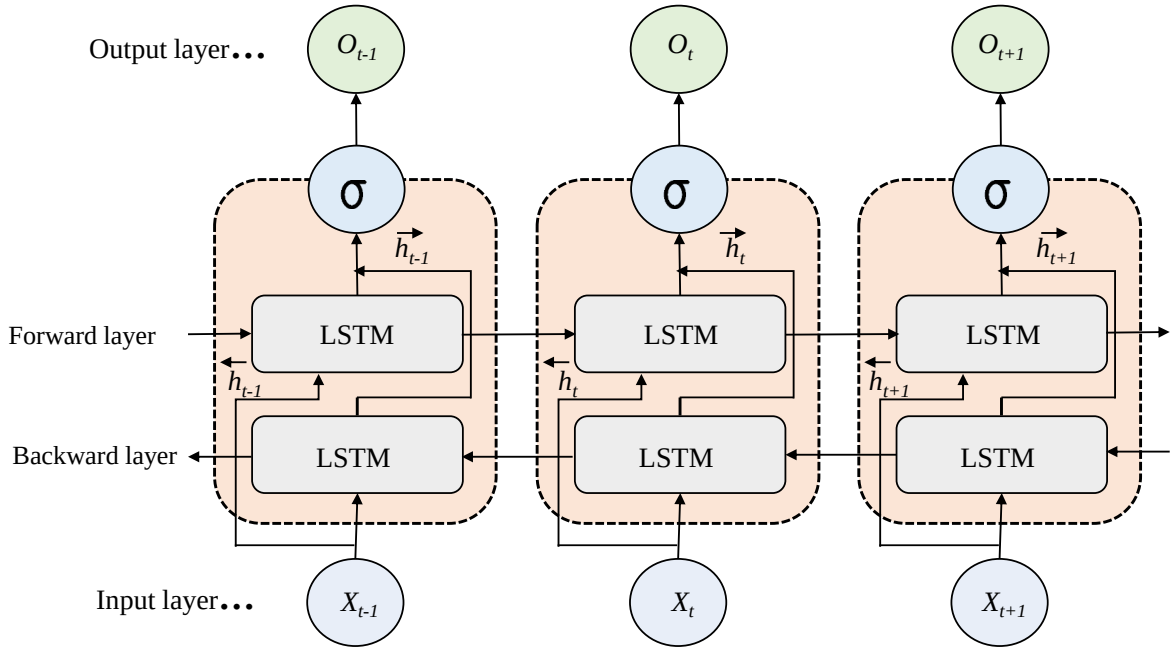


Figure 6.2: The internal structure of Bi-LSTM network.

In each time stamp t, the input samples are encoded as x_t, and the forward LSTM creates the hidden vector $\overrightarrow{h_t}$ using the past hidden layer $\overrightarrow{h_{t-1}}$. Moreover, the backward LSTM generates the hidden vector $\overleftarrow{h_t}$ using the reverse past hidden layer $\overleftarrow{h_{t+1}}$. Finally, the Bi-LSTM architecture result is generated by combining the forward hidden vector and backward hidden vector $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$. The state of the hidden layer H_t of Bi-LSTM network at time sequence t with the cell state c_t comprises forward and backward vector $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$.

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(h_{t-1}, x_t, c_{t-1}), t \in [1, T] \tag{6.5}$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(h_{t+1}, x_t, c_{t+1}), t \in [T, 1] \tag{6.6}$$

$$H_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{6.7}$$

The Bi-LSTM network is exploited in the developed framework to extract the most relevant temporal features from the physiological data to recognize emotion. As the experimental dataset contains three physiological signals, such as EMG, ECG, and SCL, we select the number of features as 3. Moreover, we choose the 20 timestamp data samples at a time depending on the sampling frequency of sensor devices. We consider three consecutive Bi-LSTM layers of filter sizes 256, 128, and 64 to retrieve the significant patterns from the raw input signals. We set a dropout with a percentage of 0.2 after the last Bi-LSTM layer to prevent the network from overfitting.

6.3.3 Soft Attention Mechanism

The soft attention mechanism [515], [516] increases the value of the most significant patterns and reduces the noise-inducing features during the network training. This attention technique multiplies the corresponding feature vectors with low weights to dismiss the insignificant regions of the data samples. The soft attention mechanism is expressed using (6.8).

$$f_{sa} = \gamma\tau((\sum_{k=1}^{K} \text{Softmax}(W_k * \tau))) \tag{6.8}$$

Here, τ is the original features that exist in the data samples. The feature vector $\tau \in \mathbb{R}^{p\times q\times r}$ is sent to a 3D convolution layer as input having weights $W_k \in \mathbb{R}^{p\times q\times r\times K}$,where K is considered as the total number of 3D weights. The tensor shape generated by fused architecture is represented by p, q, and r. Moreover, the attention maps $K = 16$ are generated by normalizing the outputs of the convolution layers through the softmax function. Then, the generated attention maps are combined to develop a weight function called α. This weight function α is multiplied by the feature vector τ and scaled by a learnable scaler called γ. Lastly, the outcome of the attention function is generated by fusing the obtained scaled attention maps with the original feature τ.

In detail, the heat feature maps were created by the soft attention module in two paths after being fed into the ReLU activation function. Two-dimensional max pooling is one direction, and the other path inputs the feature vectors into the soft attention operation before the same pooling operation. Lastly, the two paths' generated feature vectors are fused and fed into the ReLU activation with a dropout rate of 0.5.

6.3.4 Classification Block

The generated feature maps from the soft attention module considering the most significant features for emotion recognition are sent to the classification module over the GAP operation. The GAP layer interprets the feature vectors as emotion categories confidence maps and reduce the complexity as there is no parameter to optimize in this operation. We select multiple dense layers considering the filter size of 256, 128, 32, and 16 followed by a dropout rate of 0.3, 0.3, and 0.4. Lastly, a softmax function is exploited considering the multi-class classification scenario to determine the confidence level of each type of emotion. To fit the training samples in the network, we calculated the loss the function. Suppose, the number of data samples N, the number of emotion categories are C, and $X = x_1, x_2, \ldots, x_n$ as the input feature maps $\theta = \theta_1, \theta_2, \ldots, \theta_n$. The loss

function is expressed as in (6.9).

$$L(\theta, x_n) = -\frac{1}{N}\sum_{c=1}^{C}\sum_{n=1}^{N} W_c \times y_n^c \times log(\hat{y}_n^c) \tag{6.9}$$

Here, W_c is the weight of each emotion type c, $\hat{y}_i^c$ is the output of the trained network, and y_i^c is the actual target label presented in the data samples.

6.4 Results and Discussions

In this section, the experimental findings are described, including a detailed description of the used benchmark dataset to measure the efficiency of the proposed fusion technique, a thorough analysis of the obtained results from the experiments, and a comparison with the recent existing literature to show the robustness and reliability of the developed network. We exploited the following performance measure parameters: Accuracy (A), Precision (P), Recall (R), and F1-Score (F1) to evaluate the efficiency of the proposed system. A high-computing machine is used to perform the experiments. The configuration of the machine is as follows: (GPU: GTX 1070 with 8 GB GDDR5 memory, CPU: Intel® Core™ i7-6700K CPU @ 4.00GHz × 8, RAM: 32 GB DDR4). We utilized the Python language Python(v2.10) to develop the architectures in TensorFlow(v2.8.0) platform.

6.4.1 Dataset

We used a multimodal dataset called Bio Vid Emo DB [517], which is publicly available and widely used for research purposes. This dataset is collected from a laboratory environment to analyze the relationship between human emotions and physiological signals along with facial expressions. This dataset contains three physiological data, including tEMG, ECG, and SCL, as well as the facial videos of the participants. The videos and physiological signals were collected in a synchronous manner. The sensor devices were operated at the data sampling frequency of 512 Hz to generate the data. In this dataset, 94 volunteers of different ages performed five discrete emotions: amusement, anger, disgust, fear, and sad. Among the participants, the data from the 86 volunteers were considered as the data from the other users were incomplete or defiled entries. To collect the emotional level, 15 film clips were chosen, and the participants performed the above-mentioned emotions while watching the film clips. The film clips were in different lengths, varying the duration from 32 s to 245 s with an average length of 68 s.

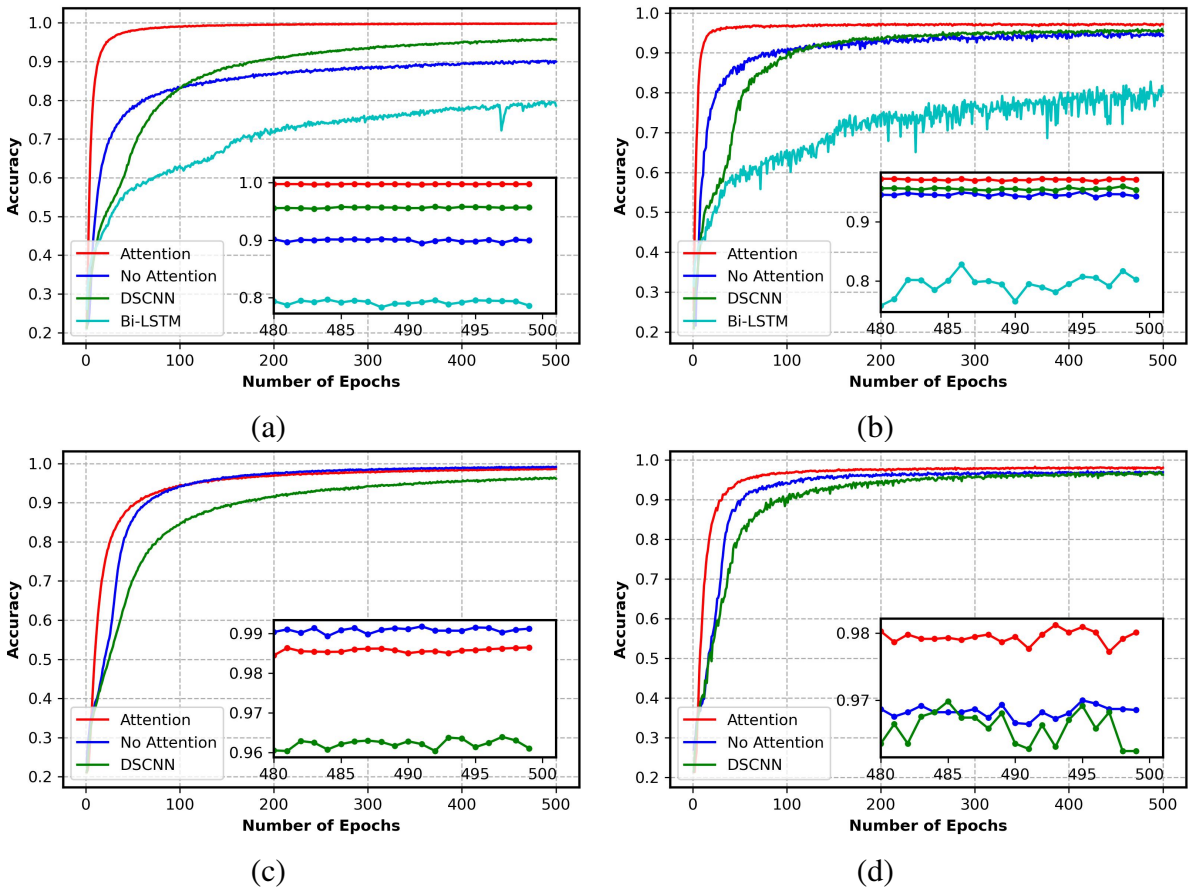


Figure 6.3: Accuracy curves for multimodal emotion recognition of the proposed fusion network trained for 500 epochs. (a)-(b) Training and Validation outcomes for the HOG face feature detector. (c)-(d) Training and Validation outcomes for the ResNet-34 face feature detector.

In the proposed fusion architecture, we considered the physiological signals and facial videos of the Bio Vid Emo DB dataset. We created two sets of data: one with 70% of the data samples utilized as a training set and the other with 30% of the data samples employed as a testing set. At the data preprocessing stage, we generated the video frames from the videos of each participant. We sampled one frame per 1 second of the video clips of the dataset. Later, we detected the facial expressions from the extracted frames using the HOG face detector and CNN-based face detector (ResNet-34). We implemented Dlib HOG and ResNet-34 methods for fast and accurate face recognition. We consider the frame size 128×128 for the experiments. At the same time, we consider the three physiological data, including tEMG, ECG, and SCL, as attributes of the Bi-LSTM architecture. We created the groups that contain the sequences of the data samples.

The sequence size that we consider here is 20 based on the data sampling frequency.

6.4.2 Results Analysis

We run the proposed fusion technique and the baseline methods for 500 epochs to train them properly. We set the following hyperparameters selected by the grid-search strategy to train the models. The hyperparameters are: learning rate–0.0001, batch size–64, optimizer–Adam, and loss function–sparse categorical cross-entropy. During the experimentation, two activation functions, ReLU and softmax, are exploited. The softmax is used at the last to recognize the emotion categories. We considered four models, including our proposed fusion architecture (Attention). The three baseline methods are 'No Attention', 'DSCNN', and 'Bi-LSTM'. In the 'Attention' network, the other two baseline methods are utilized as feature extractors by avoiding the classification block. The retrieved features from the two baseline methods are combined through the soft attention technique and sent to the classification block over the GAP layer in the fusion architecture. The 'DSCNN' network only takes the facial images as input, whereas the 'Bi-LSTM' architecture considers the physiological signals in the input pipeline. It is noted that the three architectures, including 'DSCNN', 'No Attention', and 'Attention', used the facial features from the HOG and ResNet-34 feature detectors. Among the total data instances, 10% of the data from the training samples are utilized to validate the fusion approach as well as the baseline methods.

Fig. 6.3 depicts the accuracy of the developed fusion architecture along with the baseline methods considering both the training and validation phases. Figures 6.3 (a)−6.3 (b) represent the accuracy results for the training and validation stages for the models trained for 500 epochs using the HOG facial feature detector, except for the 'Bi-LSTM' network. Moreover, Figs. 6.3 (c)−6.3 (d) represent the accuracy values for the training and validation cases for the models trained for 500 epochs using the ResNet-34 facial feature detector. It is depicted in Fig. 6.3 that the accuracy of multimodal emotion recognition is enhanced with the increasing number of iterations. The proposed network scored well in both training and validation cases. Initially, the accuracy of the 'No Attention' network was a little bit high in the case of the HOG feature detector, but it converged gradually with the increasing number of iterations. The accuracy results in both training and validation sets in the case of ResNet-34 are smoother compared to the HOG feature detector, meaning that the proposed networks trained well utilizing the features from the ResNet-34 feature detector. The 'Bi-LSTM' network shows relatively lower accuracy with the same number of epochs. Fig. 6.4 demonstrates the training and validation losses of the fusion technique along with the baseline methods. Figures 6.4 (a)−6.4 (b) depict the losses for the training and validation steps achieved by models trained over 500 epochs utilizing the HOG facial feature detector, excluding the 'Bi-LSTM' network. Furthermore, Figs. 6.4 (c)−6.4 (d) show the loss value for the training and validation sets for models trained over 500 epochs employing the ResNet-34

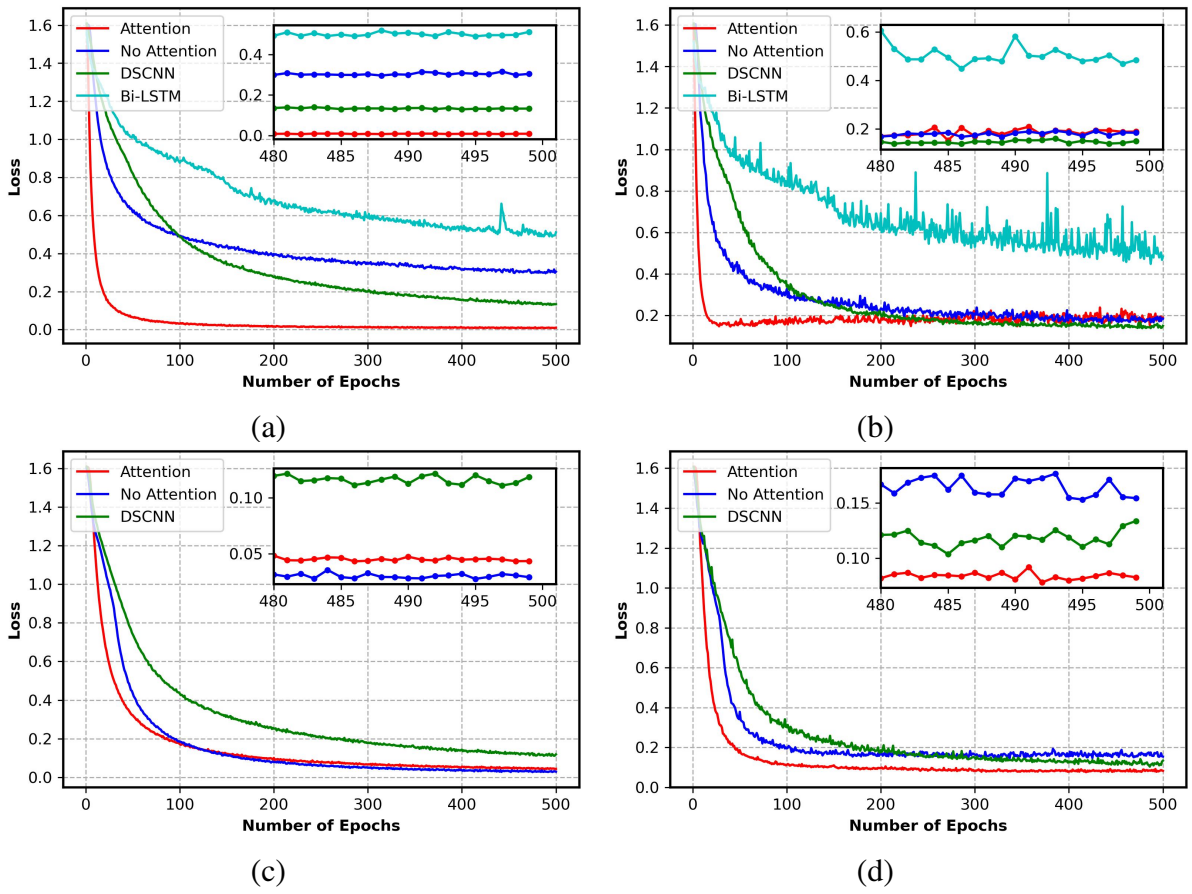


Figure 6.4: Loss curves for multimodal emotion recognition of the proposed fusion network trained for 500 epochs. (a)-(b) Training and Validation outcomes for the HOG face feature detector. (c)-(d) Training and Validation outcomes for the ResNet-34 face feature detector.

facial feature detector. It is found from Fig. 6.4 that the loss values for the 'Attention' network are comparatively lower than the other models in the training phase using HOG features. In the validation phase, the three methods such as 'Attention', 'No Attention', and 'DSCNN', perform almost similarly using the HOG feature descriptor. The loss curves during both the training and validation phases for ResNet-34 exhibit a noticeably smoother trend compared to those generated by the HOG feature detector. This observation suggests that the proposed networks effectively converged when utilizing features extracted from the ResNet-34 feature detector. However, the 'Bi-LSTM' network with physiological signals obtained comparatively high losses in both cases (training and validation).

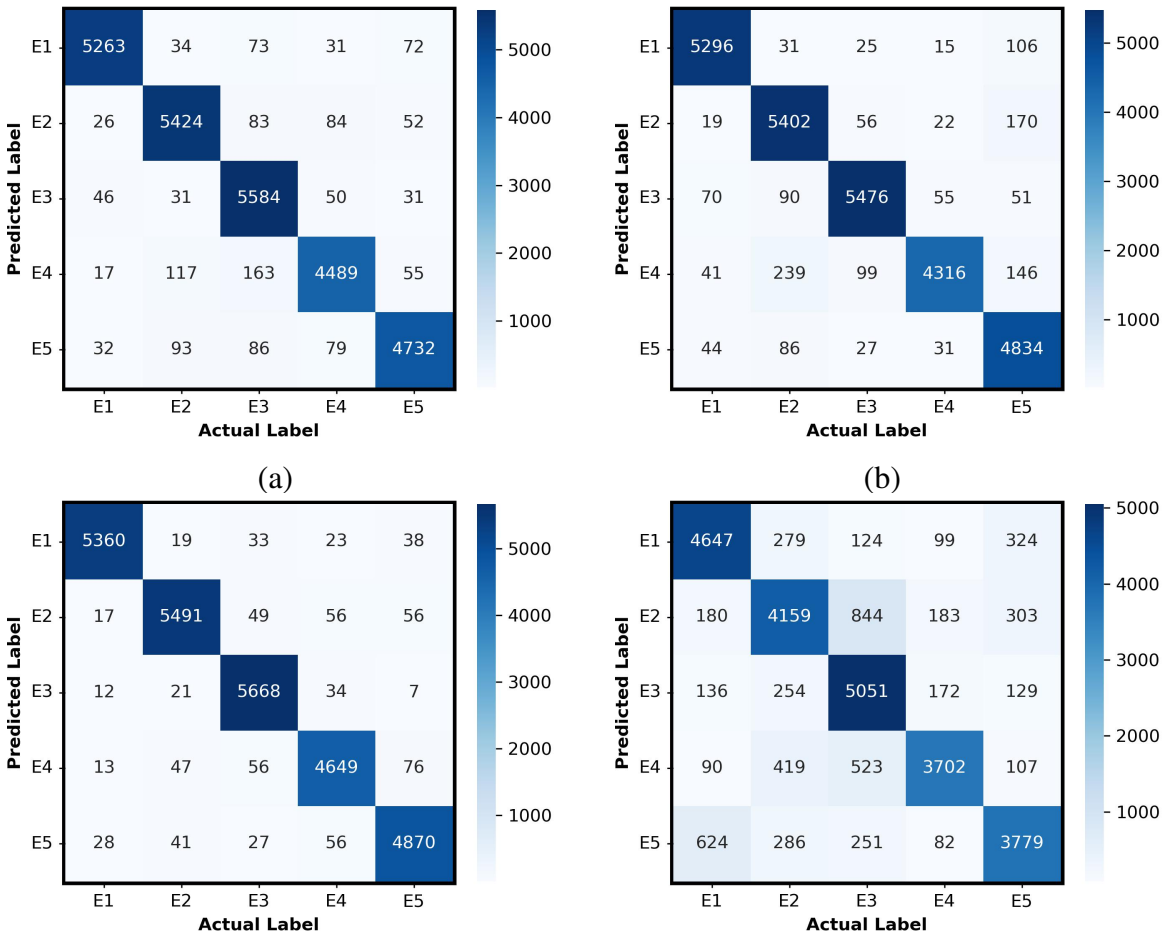


Figure 6.5: Confusion matrices of the developed models using HOG feature detector for multimodal emotion recognition. (a) Attention (b) No Attention (c) DSCNN (d) Bi-LSTM. The Bi-LSTM network has not used the HOG features as the inputs of this architecture are the physiological signals. Here, E1: Amusement, E2: Anger, E3: Disgust, E4: Fear, and E5: Sad.

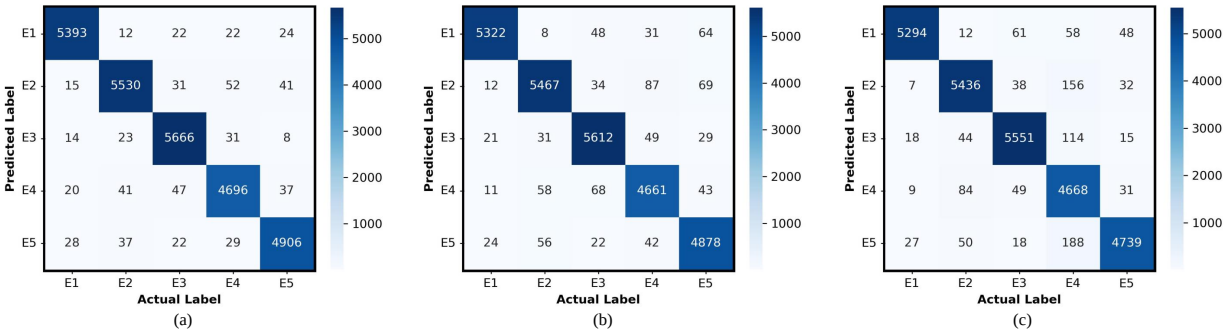
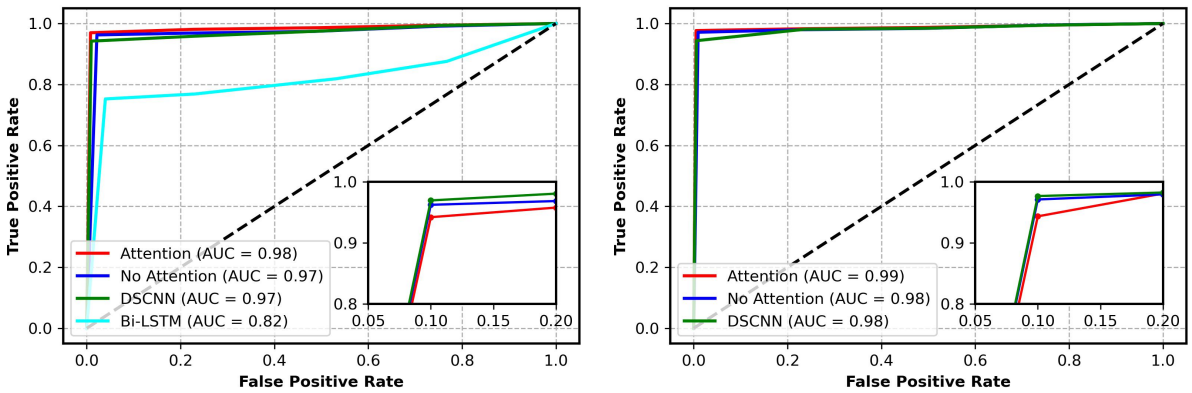


Figure 6.6: Confusion matrices of the developed models using the ResNet-34 feature detector for multimodal emotion recognition. (a) Attention (b) No Attention (c) DSCNN. Here, E1: Amusement, E2: Anger, E3: Disgust, E4: Fear, and E5: Sad.

Figure 6.5 demonstrates the confusion matrices for developed multimodal discrete emotion recognition architectures using the HOG facial features excluding the 'Bi-LSTM' network. Evidently, the proposed 'Attention' network correctly classified all the emotions more than the other baseline models. The emotions, including amusement, anger, and disgust, are correctly identified in all cases, while the other two, fear and sad, got relatively low performances.The recognition of fear emotions is quite complex compared to the other four emotions. Additionally, it shows that the fear emotions are often incorrectly classified as either disgust or anger. Further, fear emotions are highly misclassified as anger, disgust, and sad emotions. In the 'Bi-LSTM' network, anger is incorrectly classified as disgust, which lowers the performance of anger emotion recognition. Moreover, Fig. 6.6 illustrates a visual representation of the confusion matrices corresponding to the developed architectures for multimodal discrete emotion recognition, focusing on utilizing ResNet-34 facial features. From Fig. 6.6, it is evident that the proposed 'Attention' model achieved the highest accuracies for recognizing disgust and anger emotions while showing comparatively lower accuracies for detecting fear and sad emotions. The visual inspection promptly confirms that the number of incorrectly classified samples is notably minimal in both cases (HOG and ResNet-34 feature detectors). Overall, the correct emotion recognition rate for ResNet-34-based feature detector is better than the HOG-based feature detector, particularly for classifying disgust emotion.

The ROC curve for the developed fusion technique and the baseline methods are shown in Fig. 6.7, considering the Bio Vid Emo DB dataset to recognize emotion. Figure 6.7 (a) presents the ROC curves obtained from the models using the HOG features except the 'Bi-LSTM' network. Additionally, Fig. 6.7 (b) shows the ROC curves achieved from the developed architectures exploiting the ResNet-34 facial feature detector. In ROC curves, the best architecture for emotion



(a) (b)

Figure 6.7: ROC curves of the proposed fusion architecture and different baseline methods for multimodal emotion recognition. (a) present the ROC curves for the HOG face feature detector. (b) present the ROC curves for the ResNet-34 face feature detector.

recognition is demonstrated by the highest AUC values. Comparing the proposed fusion network with the baseline methods, the developed DL-enabled fusion technique using HOG features depicted the highest AUC of 0.98, followed by an AUC of 0.97 for 'No Attention' and 'DSCNN' networks. Compared to the baseline methods, the proposed 'Attention' network, when integrated with ResNet-34 features, exhibited the best AUC of 0.99. The 'No Attention' and 'DSCNN' networks followed closely, achieving an AUC of 0.98. The lowest AUC of 0.82 is found by the 'Bi-LSTM' network from the experimental findings.

The experimental results for multimodal emotion recognition considering the HOG and ResNet-34 feature detectors utilizing test set instances are shown in Table 6.1. It is noted that the developed attention network outperforms the baseline methods in terms of recognition performances. The attention mechanism, combined with physiological data and facial cues, improves the recognition outcomes. The proposed architecture using HOG features obtained an accuracy, precision, recall, and F1-Score of 97.34%, 97.35%, 97.39%, and 97.38%, respectively. The proposed 'Attention' network, leveraging ResNet-34 features, demonstrated exceptional outcome with an accuracy of 97.92%, precision of 97.95%, recall of 97.93%, and F1-Score of 97.94%. The Bi-LSTM architecture achieved a lower accuracy of 79.77%, while the other two networks, including 'No Attention' and 'DSCNN' networks, exhibited almost similar outcomes with an accuracy of 94.67% and 95.30% for HOG features, as well as 96.98% and 96.04% for the facial features of ResNet-34.

Figure 6.8 illustrates the overall performance of the three models, including 'DSCNN', 'No Attention', and 'Attention', obtained by using two facial feature detectors: HOG and ResNet-34, to compare the results for these feature detectors. Using the HOG features, the highest measure

Table 6.1: Performance measures of the proposed fusion architecture for multimodal emotion recognition. All measures are in percentage.

Methods	Feature Detectors	A	P	R	F1
Bi-LSTM	–	79.77	80.14	79.78	79.73
DSCNN	HOG	95.30	95.33	95.37	96.34
	ResNet-34	96.04	96.13	96.02	96.06
No Attention	HOG	94.67	94.77	94.69	94.63
	ResNet-34	96.98	96.99	96.95	96.97
Attention (Fusion)	HOG	97.34	97.35	97.39	97.38
	ResNet-34	97.92	97.95	97.93	97.94

* A= Accuracy, P = Precision, R = Recall, F1 = F1-Score

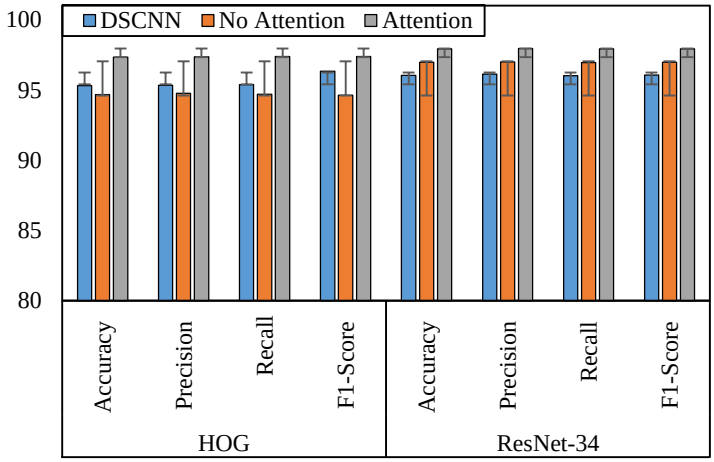


Figure 6.8: Visual representations of the results obtained by the three models: 'DSCNN', 'No Attention', and 'Attention' using two facial feature descriptors, including HOG and ResNet-34. All measures are in percentage.

(recall) of 97.39% is obtained by the 'Attention' network, and the 'No Attention' architecture achieves the lowest measure (F1-Score) of 94.63%. In the context of ResNet-34 features, the 'Attention' network achieved an impressive precision score of 97.95%, making it the top-performing model. At the same time, the 'DSCNN' architecture records the lowest measure with a recall rate of 96.02%. It is noted that the 'DSCNN' architecture performed well compared to the 'No Attention' network in the case of deployment of HOG features, whereas the 'No Attention' architecture outperformed the 'DSCNN' network significantly while using ResNet-34 features, showcasing superior performance for emotion recognition. This happened because the ResNet-34 feature detector extracted the facial landmark more efficiently compared to the HOG feature detector, which enables to develop more accurate and robust models.

Further, the emotion-wise performances of the developed architectures considering the HOG and ResNet-34 feature detectors are illustrated in Table 6.2 using the Bio Vid Emo DB dataset. The best precision of 98.71%, recall of 98.71%, and F1-Score of 98.32% are achieved from the emotion categories amusement, disgust, and amusement, respectively, in the proposed fusion network for HOG facial features. In the proposed model-level fusion network for ResNet-34 facial features, the outstanding performance metrics were attained with a remarkable precision of 98.59%, an impressive recall of 98.67%, and an exceptional F1-Score of 98.56% across the emotion categories of amusement, disgust, and amusement, respectively. In both cases, the fear and sad emotions obtained relatively lower performances than the other discrete emotions. The proposed methods failed to correctly discriminate between fear and sad emotions as they provide almost identical expressions. However, the highest precision of 97.22% is appraised in the 'No Attention' network using HOG features for fear emotion. In contrast, the 'DSCNN' network, leveraging ResNet-34, achieved the highest precision score of 98.86%, specifically for the amusement emotion category. Further, considering all the models and the emotions, the least F1-Score of 75.16% is found for the anger emotion.

Moreover, Fig. 6.9 depicts the qualitative results of the developed architecture and the baseline models considering the HOG and ResNet-34 feature detectors (except the 'Bi-LSTM' network) that demonstrate the emotion recognition capacity of the developed network using the t-SNE method. We selected 4000 samples from the test set to generate the t-SNE. The representations demonstrate that the developed fusion approach has a more dense distribution than the baseline methods. Furthermore, the fusion network maps emotional characteristics to a hyperspace in which the emotions are better presented, making it easier to distinguish among different types of emotions. From Fig. 6.9 (a), it is shown that a few magenta dots are visible in the cyan, red, and purple clusters, meaning the 'Attention' architecture using HOG features is misclassifying some of the sad emotions into fear, anger, and amusement. On the other hand, it is observed from Fig. 6.9(e) that some magenta dots are inaccurately placed within the cyan cluster, a few green samples are erroneously positioned in the red and purple clusters, and a few purple dots appear within

Table 6.2: Emotion-wise performance measures of the proposed fusion approach for Bio Vid Emo DB dataset. All measures are in percentage.

Emotions	Feature Detectors	Attention			No Attention			DSCNN			Bi-LSTM		
		P	*R*	*F1*	*P*	*R*	*F1*	*P*	*R*	*F1*	*P*	*R*	*F1*
Amusement	HOG	98.71	97.93	98.32	96.81	96.76	96.79	97.75	96.16	96.95	81.85	84.90	83.35
	ResNet-34	98.59	98.53	98.56	98.73	97.24	97.98	98.86	96.72	97.78			
Anger	HOG	97.72	96.86	97.28	92.37	95.29	93.80	95.17	95.67	95.42	77.06	73.36	75.16
	ResNet-34	97.99	97.54	97.77	97.27	96.43	96.85	96.62	95.88	96.25			
Disgust	HOG	97.17	98.71	97.93	96.35	95.36	95.85	93.23	97.24	95.20	74.35	87.96	80.59
	ResNet-34	97.89	98.67	98.28	97.02	97.73	97.37	97.09	96.67	96.88			
Fear	HOG	96.49	96.03	96.26	97.22	89.15	93.01	94.84	92.72	93.77	87.35	76.47	81.55
	ResNet-34	97.22	97.04	97.11	95.70	96.28	95.99	90.04	96.42	93.12			
Sad	HOG	96.49	96.97	96.73	91.08	96.25	93.60	95.75	94.22	94.98	81.40	75.24	78.20
	ResNet-34	97.80	97.69	97.74	95.96	97.13	96.54	97.41	94.36	95.86			

* P = Precision, R = Recall, F1 = F1-Score

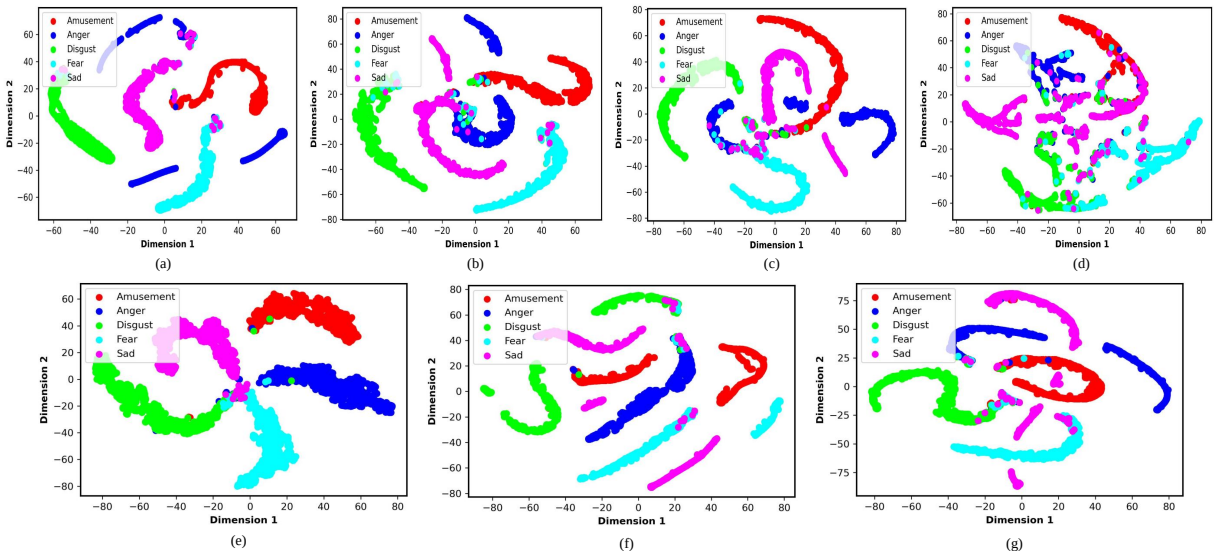


Figure 6.9: t-SNE to visualize the feature distributions of the developed network along with the baseline methods. (a) & (e) Attention (b) & (f) No Attention (c) & (g) DSCNN (d) Bi-LSTM. (a)-(c) utilized the HOG facial feature detector. (e)-(g) exploited the ResNet-34 facial feature detector.

the green cluster. This suggests that the 'Attention' architecture, utilizing ResNet-34 features, is misclassifying a small number of emotions related to sad, disgust, and anger. When visualizing the results of the 'Bi-LSTM' architecture, it is evident that many of the dots are scattered across various clusters, signifying lower performance compared to the other models.

6.4.3 Comparative Study

We compared our proposed architecture to recognize emotion from multimodal signals with the recent literature in this domain to demonstrate the robustness of this network. The accuracy comparison of the developed fusion technique with the recent literature is shown in Table 6.3. The single mode of data was used in the works developed in [504], [505], and [506] to identify emotion. The physiological data were utilized by the systems introduced in [504] and [505], whereas [506] used facial cues. Among the single modal techniques, the highest accuracy of 94.81% is exhibited from [504] while the other systems [505] and [506] obtained an accuracy of 82% and 58%, respectively. The frameworks developed in [507] and [508] used both the physiological signals and the facial cues to derive an accuracy of 80.89% and 81.54% from the outcomes of the experiment. The fusion architecture, which combined DBN and SVM algorithms, was developed in [507].

Table 6.3: Accuracy comparison of the proposed fusion approach with the existing literature for the Bio Vid Emo DB dataset.

Refs.	Year	Data Modalities	Approaches	Accuracy
[504]	2018	Physiological signals	SVM	94.81%
[505]	2018	Physiological signals	RNN	82%
[507]	2020	Physiological signals + Facial expressions	DBN + SVM	80.89%
[506]	2022	Facial expressions	EnsembleWave	58%
[508]	2022	Physiological signals + Facial expressions	Weighted fusion	81.54%
Ours with HOG	–	Physiological signals + Facial expressions	Model-level Fusion	**97.34%**
Ours with ResNet-34				**97.92%**

Moreover, [508] made the use of the weighted fusion. We developed a model-level fusion architecture in the proposed fusion system that recognized human emotions by utilizing both physiological data and facial cues. Our developed fusion network achieved an accuracy of 97.34% and 97.92% for the HOG and ResNet-34 facial feature detectors, respectively, that demonstrated superior performance in identifying emotion from multimodal data compared to both the baseline techniques and the recent literature. In the SOTA, only the work developed in [504] achieved a little bit good outcome, but it exploited only the physiological data, limiting its robustness and reliability.

6.4.4 Ablation Study

The proposed model-level fusion approach is further examined and described using ablation experiments, which generate three variants by altering the position of the soft attention module. The three scenarios are as follows:

Attention with Bi-LSTM

This model is created by fusing the soft attention module with the output of the Bi-LSTM network. In this scenario, the extracted features from the physiological signals are inputted into the attention technique before passing to the FC layer. Here, we observed the effects of soft attention to retrieve the most significant physiological signal features to recognize the emotion.

Attention with DSCNN

In this specific configuration, the soft attention module has been seamlessly integrated with the outcome of the DSCNN architecture. The facial video features are first processed through the attention mechanism, which selects the most relevant facial attributes before being fed into the FC layer. This incorporation of soft attention enables the recognition of highly pertinent facial features, thereby enhancing emotion recognition accuracy.

Attention with Bi-LSTM and DSCNN

This variant of the proposed fusion network is created by integrating the soft attention module with the output of the Bi-LSTM and DSCNN networks separately. The extracted features from the Bi-LSTM are sent to the attention technique, and the retrieved features from the DSCNN

Table 6.4: Effects of attention module placement on the performance for emotion recognition in the proposed model-level fusion approach. All measures are in percentage.

Methods	Feature Detectors	A	P	R	F1
Attention with Bi-LSTM	HOG	96.36	96.37	96.37	96.36
	ResNet-34	96.44	96.46	96.44	96.45
Attention with DSCNN	HOG	97.57	97.58	97.58	97.57
	ResNet-34	97.19	97.20	97.19	97.21
Attention with Bi-LSTM and DSCNN	HOG	97.30	97.31	97.30	97.32
	ResNet-34	96.95	96.99	96.95	96.96

* A= Accuracy, P = Precision, R = Recall, F1 = F1-Score

architecture are also passed to the attention module. Finally, the output patterns from the two-branch attention networks are combined and sent to the FC layer through the GAP layer. We investigated the influence of the attention module on physiological signals and facial videos before fusing the features.

The obtained performances from the ablation studies considering three variants of the proposed fusion technique for multimodal emotion recognition are shown in Table 6.4. The first experiment was carried out to consider the integration of the attention module with the Bi-LSTM network. In contrast, the fusion happened with the features obtained from the DSCNN network considering the HOG and ResNet-34 feature detectors and the output of the attention network using the physiological signals. In this scenario, the best accuracy of 96.44% is found from the experiments, which is higher than the above-mentioned two baseline architectures such as 'Bi-LSTM' and 'DSCNN'. The following evaluation has been done by integrating the soft attention module with the output of the DSCNN network using the HOG and ResNet-34 feature detectors, whereas the physiological signals are fed into the Bi-LSTM architecture only. Lastly, the fusion has happened using the output features of the attention module that takes the input from the DSCNN networks and the features from the physiological signals utilizing the Bi-LSTM network. It is clear from the experimental findings that the best accuracy of 97.57% is achieved from the 'Attention with DSCNN' architecture using the HOG facial features. The performance measures for this model are relatively better than the baseline methods. Finally, we conducted experiments by considering the attention module in two-branch parallelly, meaning that the soft attention network is integrated with Bi-LSTM and DSCNN architecture individually. Then, we concatenate the output features from the two-branch network before passing them to the FC layer. The ex-

perimental outcomes stated that the impressive accuracy of 97.30% and 96.95% are appraised in the 'Attention with Bi-LSTM and DSCNN' model considering the HOG and ResNet-34 feature detectors, respectively. Overall, the performances have been improved with the use of attention module in different positions of the proposed model-level fusion approach than the baseline models. Finally, the best performances have been found from our proposed fusion network that fused the heterogeneous signals first, then deployed the soft attention mechanism to extract the most relevant patterns from the fused data, and afterwards passed the most significant features to the FC layer through the GAP layer for emotion recognition.

6.5 Conclusion

Multimodal emotion recognition plays a crucial role in smart healthcare to monitor the patient's behaviour by analyzing the data from multiple sources. This chapter developed a novel DL enhanced automatic emotion recognition system from physiological and video data in healthcare analytics. The proposed system exploited the features from HOG and ResNet-34 to recognize facial features from the videos. The significant patterns from the face data are retrieved through DSCNN architecture for different emotions. Moreover, the physiological signals are inputted into Bi-LSTM to attain the relevant features for various emotional states. The features from the two models are combined through the soft attention technique to extract the most significant patterns to recognize emotion. From the extensive experiments, it is found that the developed fusion technique surpassed the existing methods and obtained an accuracy of 97.34% and 97.92% for the HOG and ResNet-34 features, respectively, from the popular Bio Vid Emo DB dataset. However, the experimental outcomes imply that emotion recognition from multimodal data is still difficult. Moreover, the proposed fusion architecture has not been tested in real-time environments.

Chapter 7

Integrated Design for Multimodal Emotion Recognition in Scalable Healthcare

7.1 Introduction

Emotion recognition allows machines to perceive and interpret human emotions based on the data provided from the individuals through the use of various smart technologies. Nowadays, the research for emotion recognition draws a lot of interest in analyzing affective data [518]. Emotion recognition encompasses various applications, including HCI, surveillance, security monitoring, psychological analysis, and healthcare monitoring [519]. Automated emotion recognition is essential for the advancement of smart healthcare facilities because it allows patients to be monitored with conditions such as stress and depression that allowed timely treatment and prevention [520].

There are several ways to recognize emotions from the individuals depending on the data modalities. The available data modalities for emotion recognition are video, audio, texts, and physiological signals [521]. In the SOTA, a few of the systems developed based on only the use of single modal data such as facial cues, texts, acoustic signals, posture, or physiological data. Moreover, some of the frameworks exploited different combinations of above-mentioned data modalities for achieving robust outcomes. The unimodal gives useful information but offers only partial information to the systems, making them inadequate, less reliable, and inefficient [522]. Besides, the integration of data from numerous sources through multimodal emotion recognition provides a more efficient depiction of emotions. The multimodal representation of one subject from multiple perspectives ensures a more accurate view of human emotions, allowing a greater understanding of human behavior [523].

Over the decades, several emotion recognition systems have been introduced based on ML methods using several data modalities [502]. In these systems, the features from the raw signals are extracted and input into the ML algorithms for classification. The ML-based systems performed well on comparatively small dataset and are able to demonstrate the interpretability by making relation with the retrieved features to emotional expressions. Domínguez-Jiménez et al. [524] exploited RF method to detect emotion from physiological signals by analyzing the frequency and time domain features. Multiple ML algorithms are employed in [525] to recognize speech emotion by extracting MFCC and mel spectrogram from the corpus. Azam et al. [526] utilized six supervised ML techniques to detect patient's emotions in healthcare where MLP performed the best with 87% accuracy. However, the major issues involved in the ML-based emotion detection systems are the manual feature extraction process, and they are not generalized efficiently. The recent trend for emotion recognition is the exploitation of DL methods to detect and interpret emotions from multiple input modalities, including facial cues, speech features, and physiological data [527]. The capability of automatic complex and hierarchical feature extraction from the raw signals, making the DL-based systems more accurate for emotion recognition. Among the DL algorithms, the CNNs and RNNs are widely adopted to detect emotion due to the capacity of learning hierarchical depiction of image-based features and temporal dependencies. Dwijayanti et al. [528] introduced face and emotion recognition simultaneously ultizing CNN architectures for humanoid robot application. A two-level ensemble classifier is developed based on the pyramidal one-dimensional CNN in [529] to classify emotion states. Kansizoglou et al. [530] trained RNN architecture through facial landmarks to identify emotions for long-term behavior modeling. A LSTM is presented in [495] to recognize emotion from multi-channel EEG data. A few of the difficulties in the DL-based emotion recognition systems are the lack of labeled datasets for efficient model training, and resource intensive for implementation in practical scenarios.

In recent years, a few research groups have focused on developing hybrid networks for emotion recognition [531]. These hybrid networks comprise two parts: CNN to capture spatial hierarchies in visual data and RNN or LSTM to model temporal dependencies in sequential data. The visual network (CNN) extracted the local feature which are sent to the LSTM network for emotion classification. Singh et al. [532] proposed a hybrid architecture consisting of 3D-CNN and ConvLSTM to recognize emotions from facial videos. Similarly, CNN-RNN based hybrid architecture is developed in [533] to detect human emotions using facial expressions derived from visual data. In these scenarios, some part of the features retrieved by CNN are ignored whenever concatenating with the RNN network due to the gradient vanishing problem. To address the challenges associated with feature concatenation, recent research has primarily focused on holistic feature fusion, bringing forward some effective fusion approaches that leverage DL and attention mechanisms. The attention-based fusion techniques focus on relevant feature patterns within multimodal data sources to detect emotions. These types of frameworks provide weights

to various data modalities, enabling the network to prioritize the cues that are most significant for effective emotion recognition. Zhao et al. [534] proposed a sensor fusion module using attention mechanism to retrieve the significant scores of each sensor to create fused feature representation. An attention-based architecture in conjunction with a pre-trained CNN is developed in [535] for emotion recognition from the acoustic signals. However, the most of existing fusion techniques are one-stage fusion, which cannot clearly model the interaction between model pair of the multimodal signals.

The rest of the chapter is structured as follows. We provide an overview of previous related studies, including recognition of emotions from physiological signals, facial cues, and multimodal data in section 7.2. Section 7.3 demonstrates the proposed multi-stage fusion architecture for emotion detection in detail. In section 7.4, we describe two multimodal experimental datasets along with the comprehensive evaluations and experimental results. At the end, section 7.5 provides the conclusions of the chapter.

7.2 Related Study

Multimodal emotion recognition has captured the interest of researchers throughout the globe because of its broad spectrum of applications that contributes a substantial volume of research efforts to this domain. In this section, we describe the frameworks that have been developed based on DL methods for emotion recognition recently. Considering the data modalities used to express emotions, emotions can be classified using physiological data, facial expressions, and multimodal signals.

7.2.1 Emotion Recognition from Physiological Data

The research on emotion recognition through physiological signals has drawn a lot of interest recently since it offers a fair assessment of the true emotional states of people. The most recent research has focused on EEG signals together with a few other physiological signals, including EMG, ECG, and Galvanic Skin Response (GSR), creating notably promising outcomes [536].

Algarni et al. [514] proposed an emotion recognition system aimed at generating more robust outcomes to facilitate accurate medical decision making. This framework retrieved the wavelet features, the Hurst exponent, and statistical patterns from the EEG signals. The Binary Gray Wolf Optimizer was exploited to select the features that enables to find out the most significant patterns. Lastly, the stacked Bi-LSTM network recognized the emotions from the extracted features. The experimental study was carried out with a benchmark dataset called DEAP to evaluate

the proposed architecture. The authors of [537] developed an EEG-based emotion recognition framework that exploited CNN to learn discriminative patterns from raw data. The collected signals are converted to spectrogram-like representations and then input into the 2D CNN to classify emotions. It is found from the experiments that the proposed architecture obtained an accuracy of 96.32% from DEAP database, and 92.54% from SEED dataset. In another study, Rahman et al. [538] presented a system to recognize emotions that converted EEG data to topographic samples that included frequency and spatial information. The two-dimensional topographic images are generated using the relative power spectral density by considering the spatial location of the EEG signals on the brain. A CNN architecture extracted features from the topographic samples and recognized the emotions. The findings revealed that the developed system appraised an accuracy of 94.63% from the SEED database. Ozdemir et al. [539] proposed an approach to recognize emotions that involved taking a series of multi-channel EEG data and transforming them into a set of multi-spectral topology images. The EEG signals are transformed into topologies based on the electrode locations. A 2D CNN extracted the features from the topological images and these high-level features are input into the LSTM to recognize the emotions. The system demonstrated an accuracy of 86.13%, 90.62%, 88.48%, and 86.23% for arousal, valence, dominance, and like–unlike emotions, respectively, from the DEAP database. Islam et al. [540] introduced a simplified system to recognize emotions that required less processing power, less processing time, and less memory. In this system, Pearson's Correlation Coefficient was exploited to transform one-dimensional EEG data into images illustrating the channel correlations of EEG sub-bands. The emotional states of the individuals are recognized through a CNN network by feeding the images into the architecture. With the DEAP dataset, the highest accuracy was achieved, measuring 78.22% for valence and 74.92% for arousal. However, the obtained accuracy falls short of the requirements for an effective real-time emotion recognition system.

7.2.2 Emotion Recognition from Facial Cues

Emotion recognition through facial cues is a natural depiction of an individual's mental state that contains an extensive amount of emotional information. Facial Emotion Recognition (FER) analyzes facial expressions from several sources, such as images and videos, to reveal emotional feelings. Due to the advancements of multimedia technologies, FER has drawn a lot of attention from the researchers for its potential use cases [477].

Sarvakar et al. [541] proposed a method to recognize five different types of emotion from facial expressions. During the initial processing phase, the background of the image is eliminated, followed by extraction of the facial vector. The system proposed a double-layer CNN architecture to detect emotions from expressional vectors. From the experiments, the proposed scheme found relatively low performances with an accuracy of 54% from FER-2013 dataset. Minaee et al. [542]

presented a DL system that exploited an attentional convolutional network for end-to-end emotion recognition in facial expressions. The attention mechanism used in this research focused on the significant parts of the face area for accurate emotion recognition. The proposed approach developed a DL model with fewer than ten layers that achieved an accuracy of 70.02% on the FER-2013 database. In another study, Chowdary et al. [543] demonstrated a transfer learning-based method for FER utilizing four CNN pre-trained architectures: MobileNet, ResNet50, Inception V3, and VGG19. The FC layers of the pre-trained CNNs were removed and new FC layers adapted to the particular requirements of the task were added. The performance was evaluated on 918 images from the Cohn Kanade (CK+) database and obtained the average accuracy of 96% to recognize emotion. Huang et al. [544] introduced a FER framework utilizing CNN architecture that combined squeeze-and-excitation network and the residual neural model. The feature maps are generated from the residual blocks that shows the significant facial landmarks which are nose and mouth. Two popular FER databases, including AffectNet and RAF-DB are used to evaluate the proposed architecture. The developed system was found to obtain an accuracy of 56.54% from the AffectNet database and 65. 67% from the RAF-DB database for emotion detection. Meena et al. [545] developed a solution for FER through CNN that can handle a large amount of signals. For better results, this research suggested to use a large batch size, adding more convolutional layers, including a robust dropout mechanism, and running a large number of epochs. The proposed system is evaluated with two benchmark databases: CK+ and FER-2013, and achieved an accuracy of 79% and 95% for the FER-2013 and CK+ datasets, respectively.

7.2.3 Emotion Recognition from Multimodal Data

Nowadays, multimodal emotion recognition has received a great deal of attention, demonstrating that it is a key approach for understanding human emotions. A single modality is not sufficient to meet the necessary requirements due to the complexity of emotion recognition. There has been substantial advancement in this area for practical applications as the researchers have proposed a number of approaches to achieve great performances in the recognition of emotion from the multimodal signals [487].

Muhammad et al. [546] described an emotion recognition framework based on Deep Canonical Correlation Analysis (DCCA) using the combination of EEG and facial sequences. In this study, CNN-based ResNet50 extracted the features from facial data and 1D-CNN was utilized to retrieve high-level patterns from the EEG signals. The highly correlated features are fused using the DCCA technique, and sent to the SoftMax classifier for emotion recognition. In the MAHNOB-HCI dataset, the accuracy was 93. 86%, and in the DEAP dataset, it was 91.54% from the experiments. Similarly, Aguiñaga et al. [547] introduced a two-stage DL architecture to recognize emotion through the correlation of facial expressions and brain data. The method

leveraged facial expressions as identifiers to derive attributes from EEG signals, then combined the two modalities to determine emotions. CNN architecture is used as the backbone for the experiments, while the 2D CNN is used to recognize facial expression, and the 1D-CNN is used for EEG analysis. The recognition rate for facial expression detection is 88.2%, and 89.6% for the brain signals analysis. Tan et al. [548] fused two data modalities, including EEG and facial cues to detect emotions for the application of HRI. All features of both EEG and facial images are extracted from CNN network. Moreover, the Monte Carlo technique was utilized to combine the outcomes and tackle the issue presented by the small dataset size. In this framework, the recognition rate was appraised near 83.33% while experimenting with HRI application. In another study, Mocanu et al. [549] developed an end-to-end DL technique to recognize emotion, integrating speech and video fusion to exploit the mutually complementary aspects of high-level features. The proposed model combined 2D and 3D CNNs in a single space to record intra-modal features and efficiently detected salient correlations among different data modes. Moreover, the inter-modal features are captured through audio-video cross-attention fusion approach. The experimental study revealed that the proposed fusion technique achieved accuracy scores of 89. 35% from the RAVDESS dataset and 84. 57% from the CREMA-D dataset. Mamieva et al. [550] used facial and speech features to detect emotions through attention-based fusion approach. The facial features from the videos are extracted using a CNN architecture, and the speech encoder utilized MFCC to retrieve spectral features via CNN. The output from two data modalities is fused to generate a multimodal feature vector through the attention network. The weighted accuracy and F1-Score of 74.6% and 66.1% are found from the IEMOCAP dataset, while the CMU-MOSEI database achieved a weighted accuracy and F1-Score of 80.7% and 73.7%.

In another work, Middya et al. [551] studied model-level fusion techniques to explore the best model to recognize emotion, integrating speech and video signals. The fusion of flattened audio and video features is performed at the model level. The combined feature vectors are sent to the FC layer of a CNN architecture for possible emotion recognition. The model-level and decision-level fusion method obtained an accuracy of 86% and 79% from the RAVDESS database. Hu et al. [552] developed a graph-based dynamic fusion technique to incorporate data from audio, video, and textual modalities. This approach analyzed the single-modal and cross-modal relations in a graph architecture to improve the integration of information. The module captured contextual information dynamics across different semantic spaces, therefore minimizing redundancy and enhancing complementarity among different data modes. The extensive experiments on IEMOCAP and MELD datasets found an accuracy of 68.21%, and 62.49%, respectively. Tzirakis et al. [553] fused three data modalities including raw texts, audio, and video to recognize multimodal emotions. The proposed architecture extracted the sentence semantics through a transformer network. Besides, a CNN architecture combining with the attention mechanism retrieved visual descriptors. At the same time, the LSTM network estimated the temporal dynamics of the speech and

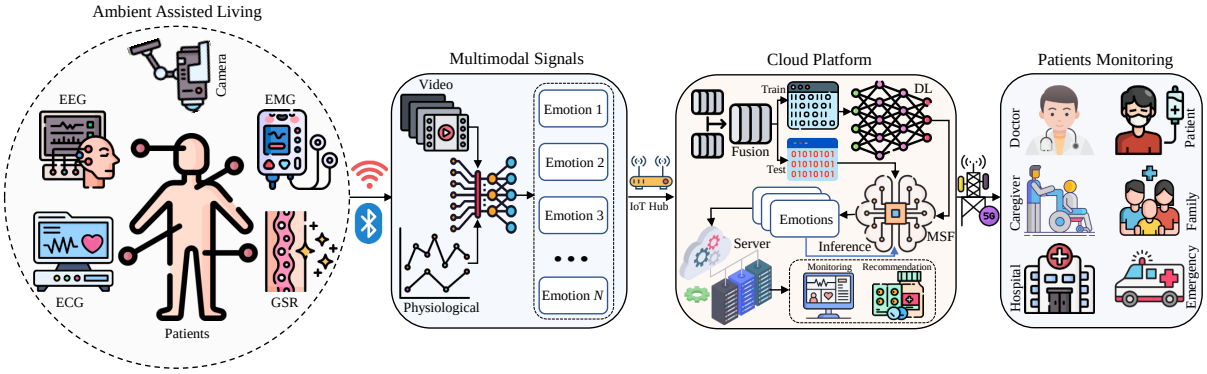


Figure 7.1: A framework of the proposed architecture for emotion recognition from the multimodal data in scalable healthcare. The visual and physiological signals are collected from the patients in AAL environment through multiple sensors. In the cloud platform, the MSF-Net developed using deep learning tools is deployed for efficient emotion recognition. The patients status is informed to the immediate mediators through proper communication channel.

video data. The recognition rates for arousal and valence are 69% and 78.3% from the SEWA dataset, as evident from the evaluation studies. A transformer-based fusion method is presented in [554] to create a joint representation to recognize multi-label emotion from audio, video, and texts. In this system, the label-level representation technique is exploited to enhance the recognition performances and deal with the multi-label classification. The average accuracy scores are 85.9%, and 67.8% are assessed for the IEMOCAP and CMU-MOSEI databases, respectively, for emotion recognition.

7.3 Proposed Multi-Stage Fusion Architecture

The proposed multi-stage fusion architecture aims to recognize emotion from the multimodal signals in the scalable healthcare environment. The developed emotion monitoring system comprises of several requisites as illustrated in Fig. 7.1. The major components are the following: IoT devices, DL architectures to develop fusion model, cloud server, communication technologies, and user interfaces both for physicians and patients. The end-user devices collect the data using several sensors, including camera, EEG, ECG, EMG, and skin sensor. The multimodal data are recorded in the shape of time series (physiological data) and facial videos. The collected signals are directed to the cloud server using short-range communication (Bluetooth) through the IoT hub. In the cloud server, the proposed multi-stage fusion architecture is deployed that is able to receive both physiological signals and visual data. The weights of the trained model are stored

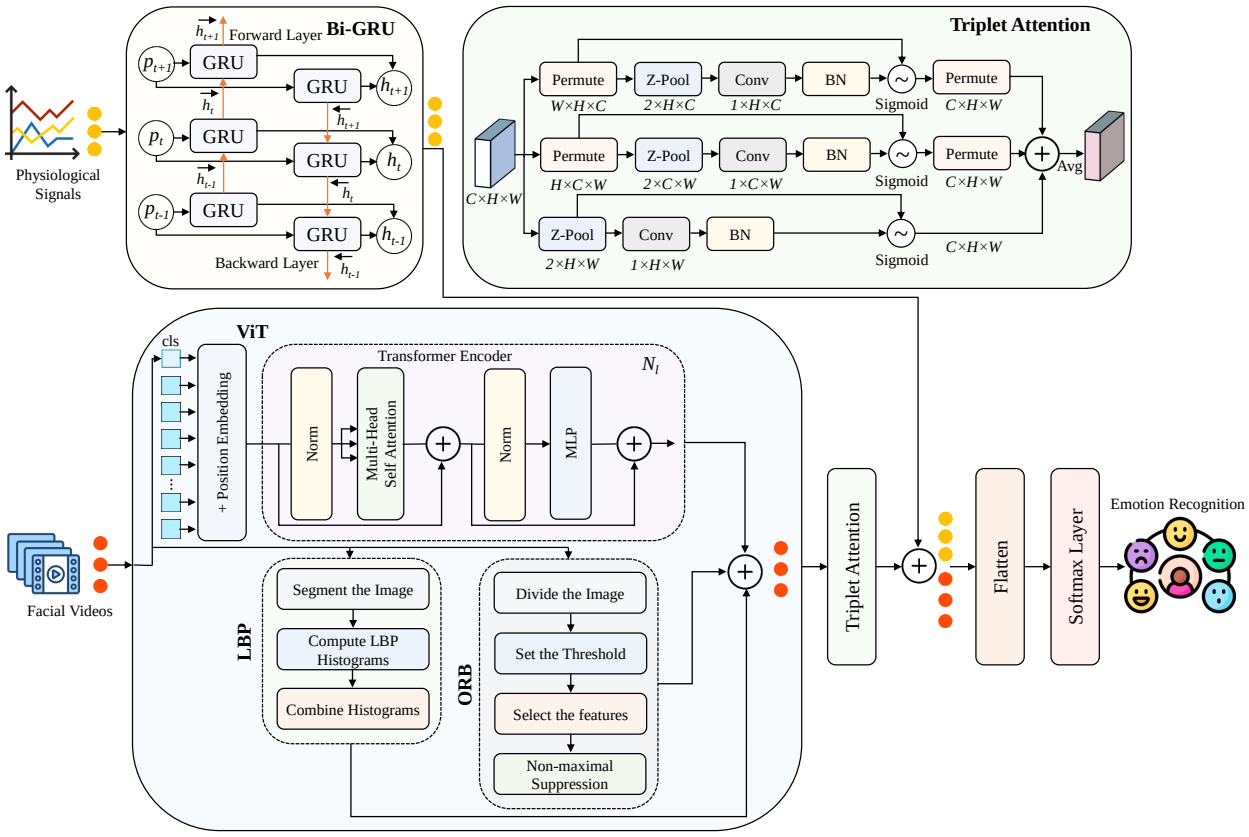


Figure 7.2: An overview of the proposed MSF-Net for multimodal emotion recognition. MSF-Net includes several stages: visual feature extraction, triplet attention, temporal feature extraction, and emotion recognition. The visual feature extraction module comprises ViT, LBP and ORB methods that simultaneously extract the deep and computer vision-based features from the facial expression. The retrieved features are fused in the visual feature extraction module and sent to the triplet attention module. The triplet attention exploits three attention branches to capture the cross-interaction among the extracted signals. On the other hand, MSF-Net encodes the physiological signals with Bi-GRU to obtain the temporal features. Afterward, the extracted features from two branches are fused to achieve the feature representation. Finally, we classify the final feature representation in the emotion classifier.

in the cloud to recognize patient's emotions depending on the input data. The service managers utilize the inference results to monitor the patient's emotional status to provide suitable services that are tailored for the patient's current conditions. At the same time, the decisions from the fu-

sion architecture are forwarded to the patient monitoring interfaces through 5G communication networks.

An overview of the proposed multi-stage fusion architecture that is deployed in the cloud server is shown in Fig. 7.2. The fusion architecture contains several modules, including Bi-GRU network, ViT, LBP, ORB, triplet attention module, and classification module. The visual data are fed into the ViT, LBP, and ORB architectures simultaneously to retrieve the significant deep and computer vision-based features. The retrieved features are fused and input into the triplet attention to refine the most important visual features. In Bi-GRU network, the temporal features are captured from the physiological data using the two-way connection of Bi-GRU's. Finally, the extracted features of two different branches are fused to generate the final feature set and fed into the classification module to recognize emotion. The mathematical flow of the proposed MSF-Net for emotion recognition is shown in Algorithm 1.

7.3.1 Vision Transformer

The ViT [555] has demonstrated impressive image recognition capability by using a transformer architecture (as shown in Fig. 7.2) that receives a series of image patches as input. For 2D image input, the ViT architecture splits the original image $\mathcal{X} \in \mathbb{R}^{C\times\mathcal{H}\times\mathcal{W}}$ into a sequence of flattened 2D $\mathcal{M}$ fixed-size image patches $\mathcal{X}' \in \mathbb{R}^{\mathcal{M}\times\mathcal{P}^2\mathcal{C}}$, where $\mathcal{M} = \mathcal{H}\mathcal{W}/\mathcal{P}^2$. Here, $\mathcal{P}$ = patch size, $\mathcal{H}$ = height, $\mathcal{W}$ = width, and C = number of channels in the original image. The obtained sequence of vectors $\mathcal{X}'$ is sent across a trainable projection layer to form a 1-dimensional facial feature embedding $\mathcal{Z}_0$ as shown in (7.1). A learnable positional embedding is integrated to each feature embedding to represent the positional information of the image samples.

$$\begin{aligned} \mathcal{Z}_0 &= \left[\mathcal{X}_{\text{cls}}; \mathcal{X}_p^1\mathcal{E}; \mathcal{X}_p^2\mathcal{E}; \ldots; \mathcal{X}_p^n\mathcal{E}\right] + \mathcal{E}_{\text{pos}}, \\ \mathcal{E} &\in \mathbb{R}^{D\times(\mathcal{P}^2\cdot C)}, \quad \mathcal{E}_{\text{pos}} \in \mathbb{R}^{D\times(\mathcal{M}+1)} \end{aligned} \tag{7.1}$$

Here, $\mathcal{E}_{\text{pos}}$ represents the position embedding, $\mathcal{X}_{\text{cls}}$ is the learnable embedding, $\mathcal{E}$ represents the linear projection, and $\mathcal{Z}_0$ is the generated position-aware feature sequence.

The next component of the ViT architecture is the transformer encoder, which includes the normalization layer, the Multi-Head Self-Attention (MHSA) mechanism, and the MLP with the residual connection. We input the facial feature embeddings $\mathcal{Z}_0$ into the encoder to design the complex interactions among the features. The MHSA moudule includes multiple SAs which helps the encoder to calculate the weights of embeddings $\mathcal{Z}_0$. The SA transforms queries and key-value pairs into arrays of query ($\mathcal{Q}$), key ($\mathcal{K}$), and value ($\mathcal{V}$) matrices of dimension d_k, and computes attention. The computing process is stated in (7.2).

Algorithm 1: Multi-Stage Fusion Network for Multimodal Emotion Recognition.

Input: Physiological Data and Facial Videos
Output: Classifying the Human Emotions
Pre-Initialization:
$D_{ps} \leftarrow$ Load Physiological Signals;
$D_{fv} \leftarrow$ Load Facial Videos;
$ViT \leftarrow$ Load Vision Transformer;
$LBP \leftarrow$ Load Local Binary Pattern;
$ORB \leftarrow$ Load Oriented FAST and Rotated BRIEF;
$TA \leftarrow$ Load Triplet Attention;
$Bi\text{-}GRU \leftarrow$ Load Bi-GRU;
while $D_{ps} \neq 0$ ***and*** $D_{fv} \neq 0$ **do**
  for D_{fv} **do**
    $F \leftarrow$ Generate Frame from D_{fv};
    if $F^{(t)} == D_{ps}^{(t)}$ **then**
      $L_{fs} \leftarrow F^{(t)}$;
      $L_{ps} \leftarrow D_{ps}^{(t)}$;
    end
  end
  $F_{ViT} \leftarrow ViT(L_{fs})$;
  $F_{LBP} \leftarrow LBP(L_{fs})$;
  $F_{ORB} \leftarrow ORB(L_{fs})$;
  $V_{fuse} \leftarrow F_{ViT} \oplus F_{LBP} \oplus F_{ORB}$;
  $A_{fuse} \leftarrow TA(V_{fuse})$;
  $F_{Bi\text{-}GRU} \leftarrow Bi\text{-}GRU(L_{ps})$;
  $F_{fuse} \leftarrow A_{fuse} \oplus F_{Bi\text{-}GRU}$;
  $Softmax \leftarrow FC(F_{fuse})$;
end

$$\mathrm{SA}(\mathcal{Q}, \mathcal{K}, \mathcal{V}) = \mathrm{Softmax}\left(\frac{\mathcal{Q}\mathcal{K}^{\mathcal{T}}}{\sqrt{d_k}}\right)\mathcal{V} \tag{7.2}$$

Here, $\mathcal{Q} = \mathcal{X}_i'\mathcal{W}^{\mathcal{Q}}$, is the query, $\mathcal{K} = \mathcal{X}_i'\mathcal{W}^{\mathcal{K}}$ represents the key, and $\mathcal{V} = \mathcal{X}_i'\mathcal{W}^{\mathcal{V}}$ denotes the value. All the vectors are generated through linear transformations of $\mathcal{X}_i'$. Moreover, the

MHSA enables transformer architecture to learn several dependencies from different subspaces. The MHSA network divides the input into many embedding subspaces, evaluate them separately in parallel utilizing various SAs, and then combines the outcomes for the final result. The computation procedure for the MHSA is described in (7.3).

$$\begin{aligned} \text{MHSA}(\mathcal{Q}, \mathcal{K}, \mathcal{V}) &= \text{Concat}(h_1, h_2, ..., h_{\mathcal{H}})\mathcal{W}^{\mathcal{O}} \\ h_i &= \text{SA}(\mathcal{Q}_i, \mathcal{K}_i, \mathcal{V}_i),\ i = 1...\mathcal{H} \\ &= \text{SA}(\mathcal{X}_i^t \mathcal{W}^{\mathcal{Q}_i}, \mathcal{X}_i^t \mathcal{W}^{\mathcal{K}_i}, \mathcal{X}_i^t \mathcal{W}^{\mathcal{V}_i}),\ i = 1...\mathcal{H} \end{aligned} \tag{7.3}$$

Here, the projections $\mathcal{W}^{\mathcal{O}}$, $\mathcal{W}^{\mathcal{Q}_i}$, $\mathcal{W}^{\mathcal{K}_i}$, and $\mathcal{W}^{\mathcal{V}_i}$ are parameters' matrices.

In the transformer structure, Layer Normalization (LN) is employed before MHSA and MLP, with residual connections added to each SA and MLP layer to enhance trainability. Each transformer encoder consists of $\mathcal{N}_l$ layers of MHSA blocks, and the output of each encoder is calculated using (7.4) and (7.5). Here, $\mathcal{Z}_l'$ is the intermediate output and $\mathcal{Z}_l$ represents the final output at layer l. The MLP architecture comprises two FC layers and GELU non-linearity activation function.

$$\mathcal{Z}_l' = \text{MHSA}(\text{LN}(\mathcal{Z}_{l-1})) + \mathcal{Z}_{l-1}, \quad l = 1, 2, \ldots, \mathcal{L} \tag{7.4}$$

$$\mathcal{Z}_l = \text{MLP}(\text{LN}(\mathcal{Z}_l')) + \mathcal{Z}_l', \quad l = 1, 2, \ldots, \mathcal{L} \tag{7.5}$$

In our proposed MSF-Net, the original image size is set to 128×128. To deploy the ViT architecture, we set the patch size of 16×16 to divide the original image into patches, which are then embedded into a 768-dimensional space. The proposed architecture contains 12 transformer encoders, each with 12 attention heads to learn the important features from the given signals. Lastly, the hidden dimension of the MLP is set to 4,608 in this study.

7.3.2 Local Binary Pattern

LBP [556] is a textual feature descriptor (as illustrated in Fig. 7.2) that represents the spatial structure of images at the local level. The texture features of the images are obtained by varying the circle's radius and the pixel number. The binary operator assigns labels to the image pixels by evaluating the center pixel value with the 3×3 adjacent pixels, generating an 8-bit binary number that is then transformed into a decimal value. The formulas for calculating the LBP code for an input pixel are shown in (7.6) and (7.7).

$$\mathrm{LBP}_{\mathcal{P},\mathcal{R}}(x,y) = \sum_{n=0}^{\mathcal{P}-1} 2^n \mathcal{S}(i_n - i_{x,y}) \tag{7.6}$$

$$\mathcal{S}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{7.7}$$

Here, $\mathrm{LBP}_{\mathcal{P},\mathcal{R}}(x, y)$ denotes the LBP texture feature centered at pixel (x, y), $\mathcal{R}$ represents the radius, $\mathcal{P}$ indicates the number of neighboring pixels, i_n signifies the gray value of n^{th} neighboring pixel, and $i_{x,y}$ denotes the gray value of the center pixel.

The output is a binary number that represents the local texture adjacent to the center pixel. The LBP values are utilized to generate histograms that depict the textural patterns in the images.

7.3.3 Oriented FAST and Rotated BRIEF

In this chapter, we exploit the region-based ORB [557] feature descriptor (as shown in Fig. 7.2) that divides the region of the images in a unique way compared to conventional ORB. In the region-based ORB algorithm, the number of feature points per region is calculated by dividing the total feature values and the number of regions. In the calculation procedure of this algorithm, the input images are equally divided into $\mathcal{A} \times \mathcal{B}$ regions, where $\mathcal{A}$ indicates the number of rows and $\mathcal{B}$ denotes the number of columns. The feature points are randomly distributed in the created regions, and the regions are labeled as $\{h_1, h_2, \ldots, h_{\mathcal{A}\times\mathcal{B}}\}$. In the next operation, the threshold is computed using (7.8), where the number of feature points is represented by n.

$$\mathcal{T} = \frac{n}{\mathcal{A}\mathcal{B}} \tag{7.8}$$

We detect the feature points in each region depending on the threshold value $\mathcal{T}$. Finally, the best feature values are calculated using non-maximal suppression technique while the generated feature points are greater than n.

7.3.4 Triplet Attention Module

In the proposed MSF-Net, we exploit the triplet attention [558] (as depicted in Fig. 7.2) to extract the most significant and relevant features from visual expressions. The triplet attention module contains three parallel networks. The first network represents cross-dimensional relationships

between the dimensions of the channel ($\mathcal{C}$) and height ($\mathcal{W}$). The second network addresses the interactions involving the dimensions of the channel ($\mathcal{C}$) and width ($\mathcal{H}$). The third network emphasizes on the height ($\mathcal{H}$) and width ($\mathcal{W}$) interactions, serving as spatial attention.

The first network exploits spatial attention using channel pooling (Z-Pool) on the input features $\mathcal{X} \in \mathbb{R}^{\mathcal{C}\times\mathcal{H}\times\mathcal{W}}$, therefore decreasing the zeroth dimension of $\mathcal{X} \in \mathbb{R}^{\mathcal{C}\times\mathcal{H}\times\mathcal{W}}$ to 2. The calculation method is formulated as in (7.9).

$$\text{Z-Pool}(\mathcal{X}) = \left[\text{MaxPool}_{0d}(\mathcal{X}),\ \text{AvgPool}_{0d}(\mathcal{X})\right] \tag{7.9}$$

Here, $0d$ represents the 0^{th} dimension where both the maximum and average pooling are carried out.

Initially, the given feature $\mathcal{X} \in \mathbb{R}^{\mathcal{C}\times\mathcal{H}\times\mathcal{W}}$ is rotated 90° anti-clockwise along the $\mathcal{H}$ axis that generates the feature maps in the form ($\mathcal{W}\times\mathcal{H}\times\mathcal{C}$). The generated tensor ($2\times\mathcal{H}\times\mathcal{C}$) after Z-Pool passes through a convolution layer followed by a batch normalization layer. In this chapter, we set the kernel size of 7×7 in the convolution layer. Now, the compressed tensor size is ($1\times\mathcal{H}\times\mathcal{C}$) which is fed into the Sigmoid activation to generate the optimized attention values. Finally, the original input shape $\mathcal{X} \in \mathbb{R}^{\mathcal{C}\times\mathcal{H}\times\mathcal{W}}$ is retained by the rotation of 90° clockwise along the $\mathcal{H}$ axis. The operation of the first network is computed as in (7.10).

$$\mathcal{X}_{1st} = \mathcal{X}_{\mathcal{CH}} \odot \text{Sig}(\text{BN}(\text{Conv}(z^{th}_{pool}(\mathcal{X}_{\mathcal{CH}})))) \tag{7.10}$$

Here, $\odot$ is a dot product, BN is the batch normalization, and Sig is the Sigmoid activation function.

Similar to first network, the second network converts the given features $\mathcal{X} \in \mathbb{R}^{\mathcal{C}\times\mathcal{H}\times\mathcal{W}}$ to ($\mathcal{H}\times\mathcal{C}\times\mathcal{W}$) dimensions by rotating the input 90° anti-clockwise along the $\mathcal{W}$ axis. This network then performs pooling operation, followed by a 7×7 convolution along with batch normalization. The generated features then passes through the Sigmoid activation and finally converts them into a shape $\mathcal{X} \in \mathbb{R}^{\mathcal{C}\times\mathcal{H}\times\mathcal{W}}$ by rotating 90° clockwise along the $\mathcal{W}$ axis. The overall procedure is formulated as in (7.11).

$$\mathcal{X}_{2nd} = \mathcal{X}_{\mathcal{CW}} \odot \text{Sig}(\text{BN}(\text{Conv}(z^{th}_{pool}(\mathcal{X}_{\mathcal{CW}})))) \tag{7.11}$$

The third network generates the tensor shape ($2 \times \mathcal{H} \times \mathcal{W}$) after the Z-Pool operation which is then passed through the convolution and batch normalization, followed by Sigmoid function to create attention weights of ($1 \times \mathcal{H} \times \mathcal{W}$). The mathematical calculation is followed by (7.12).

$$\mathcal{X}_{3rd} = \mathcal{X}_{\mathcal{HW}} \odot \text{Sig}(\text{BN}(\text{Conv}(z^{th}_{pool}(\mathcal{X}_{\mathcal{HW}})))) \tag{7.12}$$

Finally, the refined attention-based features y are derived by combining the features generated from the three networks for an input tensor $\mathcal{X} \in \mathbb{R}^{C\times\mathcal{H}\times\mathcal{W}}$ through average pooling. The final output representation of the triplet attention module follows (7.13).

$$y = \frac{1}{3}(\mathcal{X}_{1\text{st}} + \mathcal{X}_{2\text{nd}} + \mathcal{X}_{3\text{rd}}) \tag{7.13}$$

7.3.5 Bi-directional Gated Recurrent Unit

In this chapter, the Bi-GRU [559] is exploited to model physiological signals to extract temporal features. The Bi-GRU network (as demonstrated in Fig. 7.2) processes input signals in both directions, including forward and backward simultaneously, to improve sequence modeling. The Bi-GRU architecture has two GRUs: one unit analyzes the input signals from the beginning to the end, called forward direction, and the other unit processes it from the end to the beginning, called backward direction.

For time stamp t, physiological signals $p = \{p_1, p_2, p_3, \ldots, p_{T_p};\ p_t \in \mathbb{R}\}$ are input into the Bi-GRU architecture to extract temporal features. In the Bi-GRU network, the formulas for the forward and backward calculation functions are denoted as $\overrightarrow{GRU}(\cdot)$ and $\overleftarrow{GRU}(\cdot)$. The calculation procedure for the Bi-GRU network is shown in (7.14) and (7.15).

$$\vec{h}_t = \overrightarrow{GRU}(p_t, \vec{h}_{t-1}) \tag{7.14}$$

$$\overleftarrow{h}_t = \overleftarrow{GRU}(p_t, \overleftarrow{h}_{t+1}) \tag{7.15}$$

where, $\vec{h}_t$ indicates the output of the forward direction, and $\overleftarrow{h}_t$ signifies the output of the backward direction on the time stamp t. Finally, the output (h) of the state information of the Bi-GRU architecture is computed through (7.16) considering each time stamp t.

$$h = [\overleftarrow{h_1}, \overrightarrow{h_1}][\overleftarrow{h_2}, \overrightarrow{h_2}], \ldots, [\overleftarrow{h_{T_p}}, \overrightarrow{h_{T_p}}] \tag{7.16}$$

In the proposed fusion network, we exploited a Bi-GRU network with 2 layers, considering the input size of 3 and 92 for Biovid Emo DB and Multimodal Genuine Emotion and Expression Detection (MGEED) datasets. The hidden state dimension is defined to 16. After the operations, the total number of generated hidden state size is 32 due to the bi-directional property.

7.3.6 Classification Architecture

The multimodal features extracted from the Bi-GRU and the triplet attention module are fused and passed to the classification module for emotion recognition. The fused features $\mathcal{F}_{fuse}$ are fed into a linear layer with residual connections and then pass towards a Softmax layer to achieve the probability distribution $\mathcal{P}$ of emotion categories ($\mathcal{C}$).

$$\mathcal{S} = \mathcal{F}_{fuse} + \text{ReLU}(\mathcal{F}_{fuse}\mathcal{W}_{\mathcal{F}_{fuse}} + b_{\mathcal{F}_{fuse}}) \tag{7.17}$$

$$\mathcal{P} = \text{Softmax}(\mathcal{S}\mathcal{W}_s + b_s) \tag{7.18}$$

where $\mathcal{W}_{\mathcal{F}_{fuse}} \in \mathbb{R}^{d_{\mathcal{F}_{fuse}} \times d_{\mathcal{F}_{fuse}}}$, $b_{\mathcal{F}_{fuse}} \in \mathbb{R}^{d_{\mathcal{F}_{fuse}}}$, $W_{\mathcal{S}} \in \mathbb{R}^{d_{\mathcal{F}_{fuse}} \times d_{\mathcal{C}}}$, and $b_{\mathcal{S}} \in \mathbb{R}^{\mathcal{C}}$ are training parameters. Here, $\mathcal{S} \in \mathbb{R}^{\mathcal{N} \times d_{\mathcal{F}_{fuse}}}$ and $\mathcal{P} \in \mathbb{R}^{\mathcal{N} \times \mathcal{C}}$.

We calculate the emotion label with the maximum probability using the argmax function as shown in (7.19).

$$\hat{y}^{(n)} = \arg\max(\mathcal{P}^{(n)}) \tag{7.19}$$

where $\mathcal{P}^{(n)} \in \mathbb{R}^{\mathcal{C}}$ is the probability of emotion categories, and $\hat{y}^{(n)}$ denotes the emotion label predicted by the proposed architecture.

We exploit the cross-entropy loss for training. The formula for loss calculation is given below (7.20).

$$\mathcal{L} = -\frac{1}{\sum_{(m,n)=1}^{(\mathcal{V},\mathcal{U})}} \sum_{m=1}^{\mathcal{V}} \sum_{n=1}^{\mathcal{U}} \sum_{c=1}^{C} y_{m,c}^{(n)} \log_2(\hat{y}_{m,c}^{(n)}) \tag{7.20}$$

In this context, $\mathcal{V}$ and $\mathcal{U}$ are the total number of frames generated from the videos and physiological signals in the dataset. Moreover, $y_{m,c}^{(n)}$ and $\hat{y}_{m,c}^{(n)}$ represent the actual label and the predicted label of class c.

7.4 Experimental Results and Analysis

In this section, we assess the performance of MSF-Net on two multimodal emotion recognition databases, BioVid Emo DB and MGEED by comparing several SOTA emotion recognition

frameworks. We performed comprehensive evaluations to validate the efficiency of our proposed approach and provide a thorough evaluation based on the experimental findings.

7.4.1 Experimental Environment

All experiments for MSF-Net for emotion recognition were performed on a server computer with 8-core 4.00 GHz CPU of a Core i7 processor with 32 GB RAM (DDR4). The experimental setup includes two GPU cards with the specification of GTX 1080 with 8 GB of GDDR5 memory. Pytorch, a DL library, is exploited along with Python for all evaluation processes. The server was running an Ubuntu operating system, version 22.04 LTS.

7.4.2 Multimodal Datasets

The experiments have been performed on two publicly available multimodal emotion recognition datasets: BioVid Emo DB and MGEED. The detailed description of these databases are mentioned below.

BioVid Emo DB

The BioVid Emo DB [517] dataset includes physiological data including SCL, ECG, and tEMG, as well as synchronized videos. The major emotion categories in this dataset are *amusement, sad, anger, disgust,* and *fear*. The raw signals were collected from 94 participants (50 women and 44 men) by showing them 15 video clips. The data from 86 volunteers are available in the dataset due to incomplete and corrupted information. The video frame rate is 25 Hz and the resolution is 1388×1038 pixels. The sampling frequency of the physiological signals is 512 Hz.

MGEED

MGEED [560] dataset is made up of several data modalities including Optomyography (OMG), EEG, ECG, GSR, facial expressions, and depth maps. This is the first open-access dataset that contains OMG signals. The data collection framework is developed to collect multimodal data simultaneously from the participants. In the data collection system, 17 volunteers saw 9 emotional movies and provided their frame-by-frame emotional reactions mentioning seven basic emotions. The emotion classes are *neutral, happy, sadness, angry, fear, disgust,* and *surprise*. The extensive MGEED dataset contains 17 participants and 150,497 facial images, 147,539 depth map frames,

and four physiological signal categories: 70-channel EEG, 20-channel OMG, as well as one-channel ECG and GSR data.

In the proposed MSF-Net, we conducted the experiments using both the BioVid Emo DB and MGEED dataset to evaluate the robustness of the fusion network across multiple data sources. We utilized all physiological data and facial expressions from the BioVid Emo DB database. For the MGEED dataset, we excluded the depth information to maintain the same information in both datasets. We partitioned the data in training and test sets following the ratio of 70:30. Initially, we generated frames from the facial videos at the data preprocessing stage. Afterwards, the face region is detected from the frames using opencv's Caffe model. Later, facial frames and physiological data are synchronously input into the DL network. The frame size is set to 128×128. The attributes for physiological data for BioVid Emo DB and MGEED datasets are considered to value 3 and 92 considering different signals combination, which are fed into the Bi-GRU architecture.

7.4.3 Performance Metrics and Baseline Methods

Performance Metrics

This section presents a summary of the evaluation methods utilized to evaluate the performance of the architecture described in this chapter. This chapter exploits four performance measures, including Accuracy (A), Precision (P), Recall (R), and F1-Score (F1) that are frequently used to evaluate the effectiveness of classification methods.

Baseline Methods

To comprehensively assess our proposed MSF-Net, we exploit the following baselines and SOTA techniques for comparison. We also chose datasets that are extensively used in recent studies to allow for a direct comparison of our findings with them. The works presented in [504], [507], [508], [561], and [562] used BioVid Emo DB dataset. The MGEED dataset is introduced in [560].

WT-SVM [504] retrieves the high-level features from the physiological data through wavelet transform and recognizes the emotions using SVM.

MDBN [507] optimizes and fuses the features from sensor and video data through DBNs. The features are retrieved via bimodal s, and a linear SVM is exploited for emotion recognition.

MEmoR [508] exploits a fine-tuned ResNet50 for face recognition and CNN for physiological signals to analyze emotions. The data from these modalities are fused through decision-level weighted fusion to recognize emotions.

FedCMD [561] proposes the knowledge transfer scheme from teacher model to student model using physiological data and videos i.e., sensor to vision cross-modal distillation. The baseline model in this work is two-dimensional CNN.

TF-SVM [562] retrieves the time-frequency domain features from the sensor data and classifies the signals through SVM.

Fus-CNN [560] employs pre-trained ResNet101 model to retrieve and fuse the video frame and depth features. The physiological signals (EEG, ECG, and OMG) are synchronised, and normalized to their corresponding compact attributes. Lastly, all the extracted attributes are fused and input into classification block for recognition.

7.4.4 Implementation Details

We exploited the validation data for hyper-parameter tuning. We utilized Adam optimizer with a learning rate of 0.001 to fine-tune the parameters. In the evaluations, we trained the models for 25 iterations. The batch size was set to 64. We implemented L_2 regularization technique considering the following parameters: momentum=0.9, decay=0.0005, step size=3, and γ=0.1 to alleviate over-fitting problems. Our proposed model was trained through cross-entropy loss as our objective. For the validation set, we chose 20% of the training data to validate the MSF-Net. For evaluation, we used all physiological signals along with facial cues of the BioVid Emo DB database. From the MGEED dataset, we utilized all physiological data and facial expressions except the depth maps to keep the same information in both datasets. In the experiments, we considered five variants of the developed MSF-Net to recognize emotion. The models vary depending on the modules used in the architecture. In all variants, physiological data are input into the Bi-GRU network to retrieve temporal dependencies. The image frames are input into the ViT, LBP, and ORB networks (depending on the modules used) and then passed directly to the next fusion stage or through the triplet attention module to retrieve the relevant patterns. Finally, the extracted attributes from both branches are fused and passed to the classification module for emotion recognition. The proposed variants of the MSF-Net are given below.

$ViT + Attention + Bi\text{-}GRU$: In this architecture, the image data are fed into the ViT module and passed through the triplet attention block.

$ViT + LBP + Attention + Bi\text{-}GRU$: In this network, the image data are sent to the ViT and LBP modules to extract the features. Afterwards, the extracted attributes from two modules are fused and passed to the triplet attention module.

$ViT + ORB + Attention + Bi\text{-}GRU$: In this variation, the ViT and ORB modules retrieve the image features and the extracted features from both branches are combined and fed into the triplet attention network.

Table 7.1: Performance (%) of MSF-Net against SOTA methods on the BioVid Emo DB database. The best performances are represented by bold front, while the second-best outcomes are marked with underlined text.

Methods	Modalities	Fusion	A	P	R	F1
WT-SVM [504]	SCL	–	70.10	–	–	–
	ECG	–	88.76	–	–	–
	tEMG	–	80.07	–	–	–
	tEMG+SCL	Feature Fusion	63.14	–	–	–
	ECG+SCL		73.90	–	–	–
	tEMG+ECG		90.58	–	–	–
	tEMG+ECG+SCL		63.14	–	–	–
	tEMG+SCL	Decision Fusion	84.94	–	–	–
	ECG+SCL		87.76	–	–	–
	tEMG+ECG		94.81	–	–	–
	tEMG+ECG+SCL		85.19	–	–	–
MDBN [507]	Image+tEMG+ECG+SCL	Feature Fusion	80.89	–	–	–
MEmoR [508]	SCL	–	63.78	–	–	68
	ECG	–	69.56	–	–	64
	tEMG	–	70.93	–	–	61
	tEMG+ECG+tEMG	Feature Fusion	73.29	–	–	76
	Image	–	68.51	–	–	65
	Image+tEMG+ECG+SCL	Early Fusion	76.32	–	–	74
		Decision Fusion	81.54	–	–	79
FedCMD [561]	Image+tEMG+ECG+SCL	–	90.83	–	–	93.82
TF-SVM [562]	ECG	Feature Fusion	79.51	–	–	–
ViT+Attention+Bi-GRU	Image+tEMG+ECG+SCL	Feature Fusion	<u>96.46</u>	<u>96.42</u>	96.39	<u>96.40</u>
ViT+LBP+Attention+Bi-GRU			96.37	96.31	96.28	96.30
ViT+ORB+Attention+Bi-GRU			96.39	96.33	96.33	96.33
ViT+LBP+ORB+Bi-GRU			<u>96.46</u>	96.40	<u>96.40</u>	<u>96.40</u>
ViT+LBP+ORB+Attention+Bi-GRU			**96.66**	**96.63**	**96.56**	**96.60**

$ViT + LBP + ORB + Bi\text{-}GRU$: In this variation, all the three modules including, ViT, LBP, and ORB are exploited to capture the features from the image frame. The three branches extracted patterns are fused and sent to the next fusion stage excluding the triplet attention block.

$ViT + LBP + ORB + Attention + Bi\text{-}GRU$: In this architecture, the ViT, LBP, and ORB blocks extract the image features and fuses them. Then the fused features are passed to the triplet attention module. All the modules proposed in our system are utilized in this architecture.

7.4.5 Results and Comparison

To validate the robustness of the developed MSF-Net, we have conducted extensive evaluations to compare it with the SOTA algorithms on two popular databases, BioVid Emo DB and MGEED. In this part, we evaluate our proposed architecture to SOTA works on overall performance using above-mentioned evaluation metrics. Moreover, we conduct comparisons on recognition accuracy, precision, recall, and F1-Score for each emotion category.

Comparison on Overall Performance

Tables 7.1-7.2 present the comparative analysis of the developed MSF-Net and other SOTA networks for emotion recognition on BioVid Emo DB and MGEED datasets, respectively. For the BioVid Emo DB dataset, our proposed MSF-Net ($ViT + LBP + ORB + Attention + Bi\text{-}GRU$) obtained superior performances: accuracy of 96.66%, precision of 96.63%, recall of 96.56%, and F1-Score of 96.60%. The findings demonstrate an improvement of 1.85% on accuracy and 2.78% on F1-Score over existing multimodal emotion recognition methods, as evident in Table 7.1. For MGEED dataset, the proposed $ViT + LBP + ORB + Attention + Bi\text{-}GRU$ architecture achieved the highest accuracy, precision, and F1-Score of 94.26%, 92.81%, and 92.24%, respectively. Table 7.4 illustrates the increase in accuracy of 65.02% over the work presented in the original MGEED dataset. The best recall of 92.03% is revealed from the $ViT + ORB + Attention + Bi\text{-}GRU$ model from the MGEED dataset. The second-best accuracy: 96.46% and F1-Score: 96.40% are evaluated in the variants of $ViT + Attention + Bi\text{-}GRU$ and $ViT + LBP + ORB + Bi\text{-}GRU$ for the BioVid Emo DB database. On the same dataset, the second-best precision of 96.42% and recall of 96.40% are found from the $ViT + Attention + Bi\text{-}GRU$ and $ViT + LBP + ORB + Bi\text{-}GRU$ models. Considering the MGEED dataset, the variants $ViT + LBP + Attention + Bi\text{-}GRU$, $ViT + LBP + ORB + Bi\text{-}GRU$, $ViT + LBP + ORB + Attention + Bi\text{-}GRU$, and $ViT + ORB + Attention + Bi\text{-}GRU$ revealed the second-best accuracy of 94.02%, precision of 92.54%, recall of 91.86%, and F1-Score of 92.10%, respectively. It is found from Table 7.1 that the works presented in MDBN [507], MEmoR [508], and FedCMD [561] used full set of data from the BioVid Emo

Table 7.2: Performance (%) of MSF-Net against SOTA methods on the MGEED database. The best performances are represented by bold front, while the second-best outcomes are marked with underlined text.

Methods	Modalities	Fusion	A	P	R	F1
Fus-CNN [560]	RGB (weighted loss)	Feature Fusion	20.59	–	–	–
	RGB (over-sampling)		22.00	–	–	–
	RGB+Depth		20.69	–	–	–
	RGB+ECG		19.58	–	–	–
	OMG+EEG		26.77	–	–	–
	RGB+EEG+OMG		29.24	–	–	–
ViT+Attention+Bi-GRU	RGB+OMG+EEG+ECG+GSR	Feature Fusion	93.87	91.90	91.60	91.70
ViT+LBP+Attention+Bi-GRU			<u>94.02</u>	92.12	91.45	91.72
ViT+ORB+Attention+Bi-GRU			93.99	92.29	**92.03**	<u>92.10</u>
ViT+LBP+ORB+Bi-GRU			93.99	<u>92.54</u>	91.70	92.05
ViT+LBP+ORB+Attention+Bi-GRU			**94.26**	**92.81**	<u>91.86</u>	**92.24**

DB dataset. The accuracy obtained from these SOTA methods is 90.83%, 81.54%, and 80.89%, respectively. The other SOTA approaches, WT-SVM [504] and TF-SVM [562] used only the physiological signals available in the BioVid Emo DB dataset. The highest accuracy of 94.81% is obtained in WT-SVM [504] through decision-level fusion using $tEMG + ECG$ signals. In Table 7.2, the best accuracy of 29.24% is achieved in Fus-CNN [560] using $RGB + EEG + OMG$ signals through the feature fusion approach. Our proposed multi-stage fusion architecture obtained superior performances over the SOTA techniques. The reason behind that we fused the deep features captured from the transformer and the significant patterns from the computer vision-based features. Moreover, we extract the most significant visual features through triplet attention along with the deployment of multi-stage feature fusion to exploit considerably fine-grained multimodal interactions.

Comparison on Performance for Each Class

Tables 7.3-7.4 demonstrate class-wise emotion recognition accuracy, precision, recall, and F1-Score for the BioVid Emo DB and MGEED dataset. From Table 7.3, it is shown that the best accuracy, precision, recall, and F1-Score of 97.65%, 97.26%, 97.65%, and 97.45%, respectively, are found for *disgust* emotion from the $ViT+LBP+ORB+Attention+Bi\text{-}GRU$ model. The second-best accuracy of 97.34% and recall of 97.34% are also revealed in the *disgust* emotion for the $ViT+ORB+Attention+Bi\text{-}GRU$ architecture. For the same emotion category, the second-best precision and recall are found considering the values 97.10% and 97.17%. For some variations of the models, a few of the emotions got relatively lower performances. The least accuracy of 95.67% for *anger*, recall of 95.53% for *amusement*, and F1-Score of 95.66% for *amusement* are obtained from the $ViT + LBP + Attention + Bi\text{-}GRU$ network. The $ViT + ORB + Attention + Bi\text{-}GRU$ model found the lowest precision of 95.59% for *amusement* emotion. The findings from Table 7.4 demonstrated that the *sad* emotion scored the highest precision: 96.60% and F1-Score: 96.22% from the $ViT + LBP + ORB + Attention + Bi\text{-}GRU$ network. Moreover, the top accuracy of 95.90% and recall of 96.23% are achieved for *happy* and *sad* emotions from the $ViT + LBP + ORB + Attention + Bi\text{-}GRU$ and $ViT + LBP + ORB + Bi\text{-}GRU$ architectures. Similarly, the second-highest performances are appraised for *happy* and *sad* emotions considering the different variations of the MSF-Net. In the case of lower performances, the $ViT + LBP + Attention + Bi\text{-}GRU$ network recorded the least accuracy of 76.31%, recall of 76.31%, and F1-Score of 81.46% for *disgust* emotion. The *angry* emotion scored the lowest precision of 85.43% from the $ViT + Attention + Bi\text{-}GRU$ model.

Table 7.3: Performance (%) comparison of different approaches on the BioVid Emo DB dataset with multi-modality. We provide each emotion categories classification performances. The best performances are represented by bold front, while the second-best results are marked with underlined text.

Methods	Amusement				Anger				Disgust				Fear				Sad			
	A	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*
MEmoR [508]	91.02	–	–	–	75.31	–	–	–	77.64	–	–	–	80.38	–	–	–	79.92	–	–	–
MDBN[a]	56.19	–	–	–	87.68	–	–	–	27.60	–	–	–	71.93	–	–	–	42.15	–	–	–
MDBN[b]	64.70	–	–	–	57.32	–	–	–	12.21	–	–	–	43.75	–	–	–	93.07	–	–	–
MDBN[c]	80.10	–	–	–	53.65	–	–	–	14.08	–	–	–	67.33	–	–	–	57.64	–	–	–
MDBN[d]	81.86	–	–	–	80.03	–	–	–	22.73	–	–	–	76.95	–	–	–	79.55	–	–	–
MDBN[e]	86.35	–	–	–	87.21	–	–	–	33.57	–	–	–	70.20	–	–	–	85.82	–	–	–
MDBN[f]	89.25	–	–	–	90.74	–	–	–	52.60	–	–	–	86.45	–	–	–	85.42	–	–	–
ViT+Attention+Bi-GRU	95.68	96.05	95.68	95.86	96.04	96.04	96.04	96.04	97.20	96.92	97.20	97.06	96.30	96.57	96.30	96.43	96.72	96.49	96.72	96.61
ViT+LBP+Attention+Bi-GRU	95.53	95.78	95.53	95.66	95.67	96.01	95.67	95.84	97.24	97.10	97.24	97.17	96.17	96.33	96.17	96.25	96.77	96.33	96.77	96.55
ViT+ORB+Attention+Bi-GRU	95.81	95.59	95.81	95.70	95.83	96.48	95.83	96.15	97.34	96.88	97.34	97.11	96.04	96.21	96.04	96.12	96.61	96.48	96.61	96.55
ViT+LBP+ORB+Bi-GRU	95.79	95.85	95.79	95.82	95.93	95.93	95.93	95.93	97.30	97.16	97.30	97.23	96.25	96.54	96.25	96.39	96.70	96.52	96.70	96.61
ViT+LBP+ORB+Attention+Bi-GRU	95.61	96.23	95.61	95.92	96.05	96.50	96.05	96.28	**97.65**	**97.26**	**97.65**	**97.45**	96.49	96.59	96.49	96.54	96.99	96.52	96.99	96.75

[*] a=tEMG, b=ECG, c=SCL, d=combine feature, e=DBN feature, f=MDBN feature. [a−f] results come from [507]

Table 7.4: Performance (%) comparison of different approaches on the MGEED dataset with multi-modality. We provide each emotion categories classification performances. The best performances are represented by bold front, while the second-best results are marked with underlined text.

Methods	Neutral				Happy				Sad				Angry				Fear				Disgust				Surprise			
	A	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*
Fus-CNN[a]	25.73	–	–	–	19.61	–	–	–	7.32	–	–	–	0	–	–	–	30.15	–	–	–	5.35	–	–	–	26.32	–	–	–
Fus-CNN[b]	27.13	–	–	–	24.49	–	–	–	7.39	–	–	–	1.60	–	–	–	27.98	–	–	–	15.30	–	–	–	25.21	–	–	–
Fus-CNN[c]	25.34	–	–	–	23.07	–	–	–	7.27	–	–	–	1.60	–	–	–	28.59	–	–	–	8.89	–	–	–	25.21	–	–	–
Fus-CNN[d]	22.18	–	–	–	14.48	–	–	–	20.55	–	–	–	13.44	–	–	–	32.70	–	–	–	8.84	–	–	–	25.08	–	–	–
Fus-CNN[e]	25.73	–	–	–	20.56	–	–	–	22.14	–	–	–	20.72	–	–	–	55.96	–	–	–	30.17	–	–	–	27.13	–	–	–
Fus-CNN[f]	31.04	–	–	–	37.56	–	–	–	20.36	–	–	–	25.29	–	–	–	28.52	–	–	–	22.34	–	–	–	26.82	–	–	–
ViT+Attention+Bi-GRU	91.72	91.14	91.72	91.43	95.35	94.75	95.35	95.05	95.87	96.11	95.87	95.99	91.39	85.43	91.39	88.31	95.53	95.32	95.53	95.42	77.49	86.37	77.49	81.69	93.82	94.12	93.82	93.97
ViT+LBP+Attention+Bi-GRU	92.32	91.19	92.32	91.75	95.57	95.07	95.57	95.32	95.81	96.38	95.81	96.10	90.23	85.65	90.23	87.88	95.34	95.34	95.34	95.34	76.31	87.36	76.31	81.46	94.54	93.79	94.54	94.17
ViT+ORB+Attention+Bi-GRU	92.13	91.83	92.12	91.98	95.20	94.73	95.20	94.96	96.17	96.07	96.16	96.12	93.26	86.60	93.25	89.80	95.21	95.11	95.21	95.16	78.56	88.31	78.56	83.15	93.67	93.36	93.66	93.51
ViT+LBP+ORB+Bi-GRU	91.82	91.65	91.82	91.74	95.29	94.71	95.29	95.01	96.23	96.17	**96.23**	96.20	91.39	88.11	91.39	89.72	95.28	95.13	95.28	95.20	77.06	89.09	77.06	82.64	94.79	92.89	94.78	93.83
ViT+LBP+ORB+Attention+Bi-GRU	92.68	91.61	92.68	92.14	**95.90**	95.26	95.90	95.58	95.84	**96.60**	95.84	**96.22**	91.86	87.38	91.86	89.56	95.65	95.23	95.65	95.44	76.42	90.02	76.42	82.66	94.62	93.50	94.62	94.06

[*] a=RGB (weighted loss), b=RGB (over-sampling), c=RGB + Depth, d=RGB + ECG, e=OMG + EEG, f=RGB + EEG + OMG. [a−f] results come from [560]

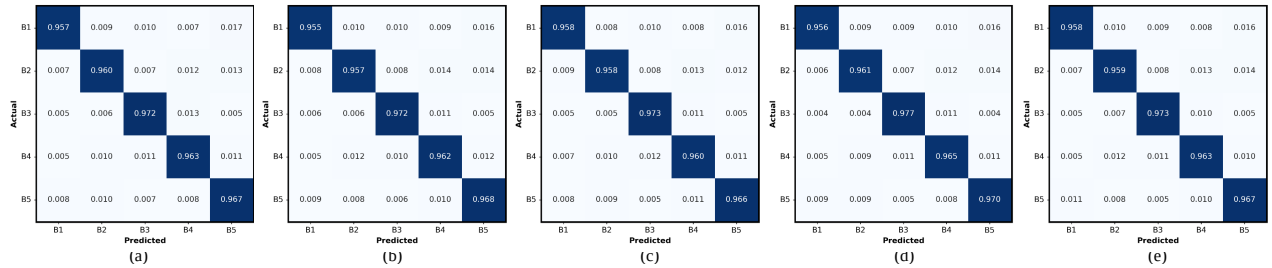


Figure 7.3: Visualization of confusion matrices of MSF-Net on the testing set of the BioVid Emo DB database. The columns represent the actual labels, while the rows represent the predicted ones. (a) ViT+Attention+Bi-GRU (b) ViT+LBP+Attention+Bi-GRU (c) ViT+ORB+Attention+Bi-GRU (d) ViT+LBP+ORB+Bi-GRU (e) ViT + LBP+ORB+Attention+Bi-GRU. Here, B1: Amusement, B2: Anger, B3: Disgust, B4: Fear, and B5: Sad.

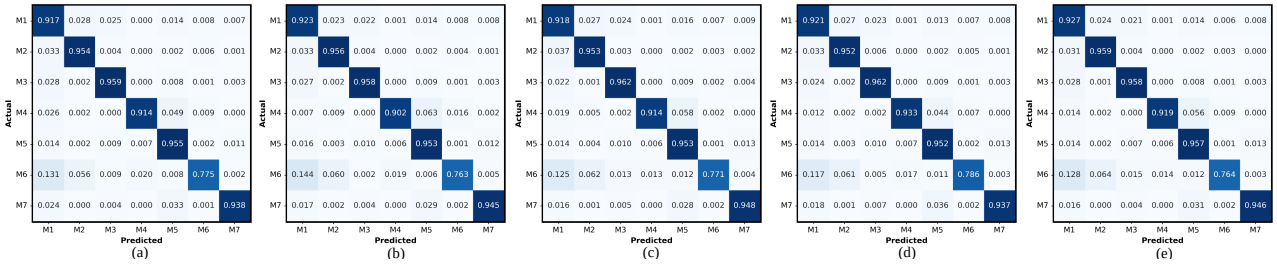


Figure 7.4: Visualization of confusion matrices of MSF-Net on the testing set of the MGEED database. The columns represent the actual labels, while the rows represent the predicted ones. (a) ViT+Attention+Bi-GRU (b) ViT+LBP+Attention+Bi-GRU (c) ViT+ORB+Attention+Bi-GRU (d) ViT+LBP+ORB+Bi-GRU (e) ViT+LBP + ORB+Attention+Bi-GRU. Here, M1: Neutral, M2: Happy, M3: Sad, M4: Angry, M5: Fear, M6: Disgust, and M7: Surprise.

7.4.6 Visual Results

In this part, we provide several examples of the sub-views to demonstrate the intuitive outcomes. Fig. 7.3 and Fig. 7.4 show the confusion matrices to represent the distribution of several emotion categories on two different datasets. Overall, the proposed MSF-Net demonstrates superior predictive performance compared to the SOTA methods on different emotions on the BioVid Emo DB and MGEED datasets. For instance, the *disgust* emotion from the BioVid Emo DB dataset and the *happy* emotion from the MGEED dataset are correctly classified using the proposed MSF-Net.

On the BioVid Emo DB dataset (Fig. 7.3), it is observed that the *amusement* emotion are mostly misclassified by the *sad* emotion. In this case, the developed architecture is failed to

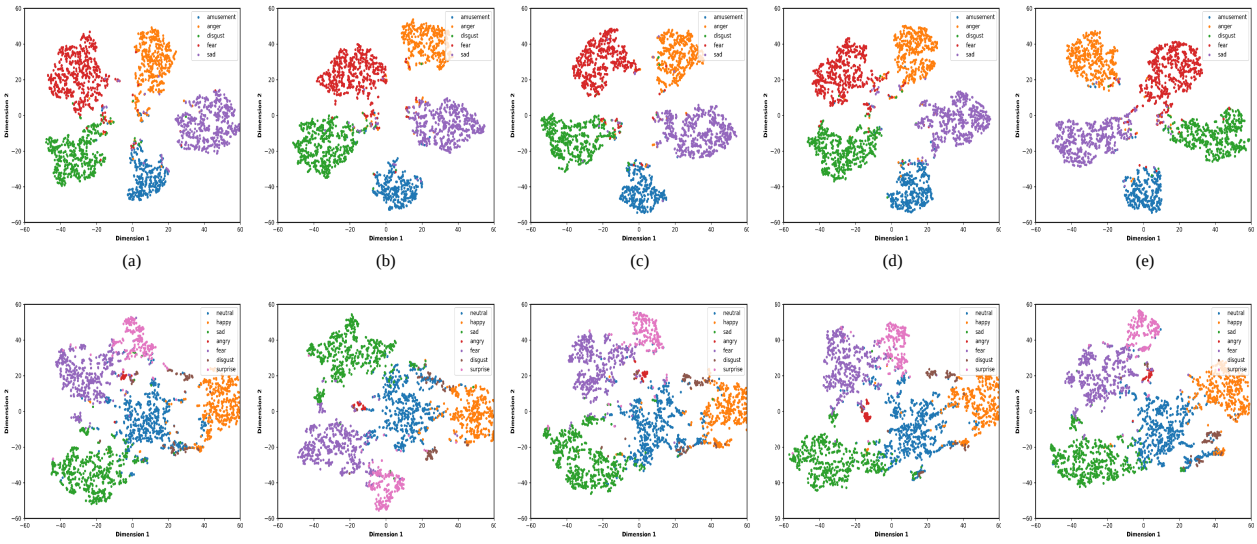


Figure 7.5: t-SNE feature visualization of the MSF-Net for multimodal emotion recognition. (a) & (f) ViT+Attention+Bi-GRU (b) & (g) ViT+LBP+Attention+Bi-GRU (c) & (h) ViT+ORB+Attention+Bi-GRU (d) & (i) ViT+LBP+ORB+Bi-GRU (e) & (j) ViT+LBP+ORB+Attention+Bi-GRU. Different categories of emotions are represented with distinct colors and shapes. The first row represents the results from the BioVid Emo DB dataset, and the second row depicts the results from the MGEED database.

differentiate between *amusement* and *sad* emotions. The *fear* emotion are wrongly identified as *disgust* and *sad* emotions. The reason behind it there are semantic correlations between *fear* emotion and *disgust* or *sad* emotion. At the same time, it is found that the *disgust* emotion has the tend to misclassify as *fear*. Moreover, there is shared semantic information between *anger* and *sad* emotions, which poses a challenge for the MSF-Net to differentiate between these emotions. On the MGEED dataset (Fig. 7.4), it is revealed that the *disgust*, *happy*, and *sad* emotions are confused with the *neutral* emotion in some cases. The reason for this is that the MGEED dataset contains an excess number of neutral emotions. Thus, MSF-Net has a tendency to acquire a greater understanding of the neutral emotion. The *sad* emotion is recognized as *fear* emotion and vice versa. This result is found because the *sad* and *fear* are highly close in the activation and arousal domains, meaning that the speakers have similar perception and interpretation of these emotions. We did not observe misclassification rates for several emotion categories, which are: *happy* and *surprise*, *happy* and *angry*, *sad* and *angry*, as well as *angry* and *surprise*.

Figure 7.5 depicts the feature distribution of the generated fused representations on the BioVid Emo DB and MGEED datasets. We randomly choose 2500 samples from each emotion category and show their combined representations in the shared feature space. It is shown from Fig. 9 that

the data distribution for MGEED dataset is slightly scattered than the BioVid Emo DB dataset meaning the proposed MSF-Net learnt the features more efficiently from the BioVid Emo DB dataset. Considering the BioVid Emo DB dataset, the *amusement* and *anger* emotion categories are more critical to differentiate from other emotions. They are found in nearly all of the other emotion clusters. For the MGEED dataset, it is quite hard to differentiate between *neutral* and *disgust* emotions. The cluster of *disgust* emotion are quite small and is mostly represented in other emotion groups. Our observation revealed that *disgust* emotional data is frequently misclassified into different emotion categories, while the misclassification of other emotions as *disgust* is very rare.

Table 7.5: Ablation analysis (%) of MSF-Net with different combinations of multimodal signals on the BioVid Emo DB and MGEED datasets. We provide the overall performance on the testing set. The best performances are represented by bold front, while the second-best results are marked with underlined text.

Datasets	Signals	A	P	R	F1
BioVid Emo DB	Image+tEMG+ECG	96.42	96.41	96.34	96.38
	Image+tEMG+SCL	96.47	96.41	96.39	96.40
	Image+ECG+SCL	96.41	96.37	96.33	96.35
	Image+tEMG	96.44	96.39	96.36	96.38
	Image+ECG	**96.52**	**96.49**	**96.43**	**96.46**
	Image+SCL	96.39	96.36	96.34	96.35
MGEED	RGB+OMG+EEG+ECG	93.59	91.82	90.61	91.11
	RGB+OMG+EEG+GSR	93.85	91.74	91.18	91.40
	RGB+EEG+ECG+GSR	93.48	91.50	90.44	90.86
	RGB+OMG+EEG	93.88	91.87	91.33	91.55
	RGB+OMG+GSR	93.63	91.73	90.83	91.20
	RGB+EEG+ECG	93.53	91.30	90.99	91.08
	RGB+EEG+GSR	93.58	91.29	90.92	91.05
	RGB+ECG+GSR	93.87	92.14	91.50	**91.73**
	RGB+OMG	93.60	91.72	**91.67**	91.63
	RGB+EEG	93.61	91.86	90.60	91.12
	RGB+ECG	**93.90**	**92.27**	91.14	91.62
	RGB+GSR	93.70	92.23	90.93	91.49

7.4.7 Ablation Studies

This section performs ablation studies on different physiological signals to see the effects of different signals on the model efficiency on the BioVid Emo DB and MGEED datasets. The experimental outcomes are demonstrated in Table 7.5, Table 7.6, and Table 7.7 considering the architecture of $ViT + LBP + ORB + Attention + Bi\text{-}GRU$. Table 7.5 illustrates the overall performance on both datasets. Table 7.6, and Table 7.7 depict the class-wise emotion recognition performance on BioVid Emo DB and MGEED datasets by varying the physiological signals. On the BioVid Emo DB database, the highest performance metric is found for the signal combination of $Image + ECG$. The best scores are: accuracy (96.52%), precision (96.49%), recall (96.43%), and F1-Score (96.46%). The second-best performance is obtained with the combination of $Image + tEMG + SCL$. Similar to BioVid Emo DB, $RGB + ECG$ obtained the best accuracy (93.90%) and precision (92.27%) from MGEED dataset. However, it should be noted that the $RGB+ECG+GSR$ combination has the highest recall (91.50%) and F1-Score (91.73%), showing an excellent balance of precision and recall for the MGEED dataset. The other combination, including $RGB + OMG$ performed the best for recall (91.67%). Multimodal signals are found to improve performance for both datasets. The choice of modalities, including *RGB*, *ECG*, and *GSR*, has an influence on various aspects of performance. Moreover, the efficiency of various signal combinations may be dependent upon the nature of the dataset and its particular features, as evidenced by the variation in performance across datasets.

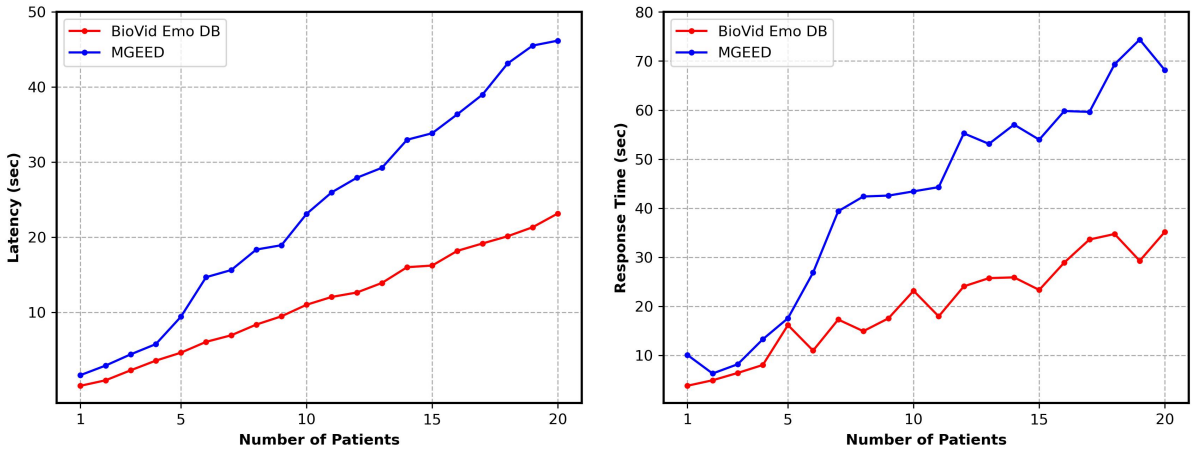
From Table 7.6, it is noticed that the highest metrics are achieved for *disgust* emotion considering the combination of $Image+ECG$ signals. The recorded scores are accuracy (97.57%), recall (97.57%), precision (96.90%), and F1-Score (97.24%). For the *fear* emotion, the signal combination $Image + tEMG + SCL$ received the best precision (96.53%) and F1-score (96.39%), highlighting that the SCL signal can improve the recognition of this emotion. However, the F1-Scores for *amusement* and *anger* have comparatively lower values for all signal combinations, showing difficulties in effectively recognizing these emotions. The robust performance of combinations that incorporate *ECG* indicates its crucial role in enhancing the efficiency of emotion recognition. In Table 7.7, we revealed that the $RGB+ECG$ found the highest accuracy (96.13%), precision (96.18%), and F1-Score (96.19%) for *sad* emotion. For the same combination, the *neutral* emotion received the best accuracy of 92.49%. The best precision of 96.52% for *sad* emotion is achieved for the $RGB+OMG+EEG$ combination. While considering the *disgust* and *surprise* emotions, the performance for almost all signal combinations is relatively low, noting the significant variation in precision and F1-Scores. This outcome demonstrated that accurate recognition of these emotions is quite a challenge. In some cases, the *GSR* signal enhanced the performance for emotion recognition. For instance, $RGB+OMG+GSR$ recorded a comparatively higher F1-Score of 80.14% for *surprise* emotion. In summary, the findings emphasized the various effects

Table 7.6: Ablation analysis (%) of MSF-Net with different combinations of multimodal signals on the BioVid Emo DB dataset. We provide each emotion categories classification performances. The best performances are represented by bold front, while the second-best results are marked with underlined text.

Signals	Amusement				Anger				Disgust				Fear				Sad			
	A	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*
Image+tEMG+ECG	95.59	96.36	95.59	95.97	95.78	96.19	95.78	95.99	97.44	96.86	97.44	97.15	96.11	96.27	96.11	96.19	96.80	96.39	96.80	96.59
Image+tEMG+SCL	95.63	95.90	95.63	95.77	95.83	96.00	95.83	95.91	97.53	97.11	97.53	**97.32**	96.24	96.53	96.24	96.39	96.70	96.51	96.70	96.61
Image+ECG+SCL	95.75	96.06	95.75	95.90	95.60	95.99	95.60	95.80	97.38	97.12	97.38	97.25	96.12	96.24	96.12	96.18	96.81	96.44	96.81	96.62
Image+tEMG	95.65	96.07	95.65	95.86	95.79	95.90	95.79	95.85	97.23	**97.18**	97.23	97.21	96.48	96.28	96.48	96.38	96.63	96.54	96.63	96.58
Image+ECG	95.65	96.15	95.65	95.90	95.67	96.32	95.67	95.99	**97.57**	96.90	**97.57**	97.24	96.42	96.52	96.42	96.47	96.84	96.54	96.84	96.69
Image+SCL	95.99	95.90	95.99	95.94	95.68	96.27	95.68	95.98	97.08	97.04	97.08	97.06	96.17	95.96	96.17	96.07	96.77	96.61	96.77	96.69

Table 7.7: Ablation analysis (%) of MSF-Net with different combinations of multimodal signals on the MGEED dataset. We provide each emotion categories classification performances. The best performances are represented by bold front, while the second-best results are marked with underlined text.

Signals	Neutral				Happy				Sad				Angry				Fear				Disgust				Surprise			
	A	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*	*A*	*P*	*R*	*F1*
RGB+OMG+EEG+ECG	91.37	90.95	91.37	91.16	95.50	94.08	95.50	94.78	95.69	96.10	95.69	95.89	90.70	87.44	90.70	89.04	95.48	94.96	95.48	95.22	71.28	86.03	71.28	77.96	94.27	93.22	94.27	93.74
RGB+OMG+EEG+GSR	91.50	91.91	91.50	91.71	95.85	94.52	95.85	95.18	95.75	96.00	95.75	95.87	90.93	86.50	90.93	88.66	95.77	94.96	95.77	95.36	74.28	84.82	74.28	79.20	94.19	93.48	94.19	93.83
RGB+EEG+ECG+GSR	91.55	90.75	91.55	91.15	95.43	94.57	95.43	94.95	95.75	95.95	95.75	95.85	90.70	86.09	90.70	88.33	95.11	94.81	95.11	94.96	71.17	85.79	71.17	77.80	93.35	92.57	93.35	92.96
RGB+OMG+EEG	92.07	91.18	92.07	91.62	95.33	94.85	95.33	95.09	95.85	**96.52**	95.85	96.18	90.47	86.06	90.47	88.21	95.26	95.03	95.26	95.14	75.99	85.94	75.99	80.66	94.35	93.49	94.35	93.92
RGB+OMG+GSR	91.98	90.70	91.98	91.34	95.39	94.94	95.39	95.16	95.69	96.08	95.69	95.89	89.77	85.78	89.77	87.73	94.91	95.04	94.91	94.98	74.38	86.86	74.38	80.14	93.67	92.74	93.67	93.20
RGB+EEG+ECG	91.44	91.33	91.44	91.39	95.52	94.22	95.52	94.86	95.72	96.06	95.72	95.91	90.64	84.52	90.64	87.82	95.72	95.72	95.72	95.72	74.38	84.52	74.38	79.45	93.79	87.82	93.79	91.49
RGB+EEG+GSR	91.62	91.17	91.62	91.39	95.67	94.83	95.67	95.25	95.44	96.02	95.44	95.73	90.00	84.31	90.00	87.06	94.78	94.40	94.78	94.59	74.49	84.65	74.49	79.25	94.43	93.68	94.43	94.05
RGB+ECG+GSR	91.84	91.34	91.84	91.59	95.14	94.77	95.14	94.95	**96.13**	96.18	**96.13**	96.15	92.56	86.33	92.56	89.34	95.33	95.03	95.33	95.18	75.24	88.19	75.24	81.20	94.27	93.11	94.27	93.69
RGB+OMG	91.77	91.02	91.77	91.39	94.78	94.51	94.78	94.65	95.48	96.05	95.48	95.76	93.72	85.74	93.72	89.56	95.01	94.74	95.01	94.88	77.17	86.64	77.17	81.63	93.79	93.30	93.79	93.54
RGB+EEG	91.64	91.04	91.64	91.34	95.83	94.27	95.83	95.04	95.73	95.79	95.73	95.76	90.70	87.25	90.70	88.94	95.19	95.41	95.19	95.30	71.60	86.87	71.60	78.50	93.51	92.43	93.51	92.97
RGB+ECG	92.49	91.16	92.49	91.82	95.01	94.88	95.01	94.95	96.11	96.26	96.11	**96.19**	91.86	88.17	91.86	89.98	95.31	95.20	95.31	95.26	73.20	87.01	73.20	79.51	94.03	93.21	94.03	93.62
RGB+GSR	91.77	90.60	91.77	91.18	95.01	94.40	95.01	94.70	95.78	95.95	95.78	95.86	91.40	88.51	91.40	89.93	95.50	95.29	95.50	95.40	72.45	86.78	72.45	78.97	94.59	94.10	94.59	94.34



(a) (b)

Figure 7.6: Performance measures for IoT deployment in scalable healthcare scenario. (a) Latency (b) Response time.

of different signal combinations. In particular, the combination of *ECG* and *ECG* consistently improved the recognition of certain emotions, whereas *GSR* has limited but focused efficiency.

7.4.8 IoT Performance

We deployed our proposed multi-stage fusion architecture in scalable healthcare environment. We considered two performance measures, including latency and response time, to evaluate the scalability of the IoT network. Latency refers to the time to transmit the packets from the patients and the cloud server receives them. The response time is the total time to send the request from the patients and receives the classification results from the cloud server. This is the time measure for sending data from patients and receiving results, along with inference time. Inference time represents the time to make classifications through DL architecture on the data received from the patients.

Figure 7.6 shows the performance of the IoT deployment in scalable healthcare environment. We deployed $ViT + LBP + ORB + Attention + Bi\text{-}GRU$ model in cloud server that continuously takes input from multiple patients and makes classification for multimodal emotion recognition. It is found from Fig. 7.6 (a) that the latency for the MGEED data is comparatively higher than the BioVid Emo DB data. The reason behind this is that the MGEED dataset considered more physiological signals (70-channel EEG, 20-channel OMG, as well as one-channel ECG and GSR) compared to BioVid Emo DB dataset. The BioVid Emo DB dataset only considered the single

channel tEMG, ECG, and SCL. While considering the response time as shown in Fig. 7.6 (b), the response time in the case of MGEED dataset is slightly higher than the BioVid Emo DB database. For the fewer number of patients, the response time for both cases are mostly close. Moreover, the inference time for the MGEED data is relatively high. This is because of the large number of training parameters for the fusion architecture considering MGEED data. At the same time, the fusion network developed for MGEED data takes relatively more time to process a large number of input signals compared to BioVid Emo DB dataset. Overall, the deployed network is capable of handling an increasing number of patients. The performance metrics are increasing for the larger number of connected patients; however, all measures are relatively low to be applicable for real-world scalable healthcare scenario. There are few factors that affect the performance of IoT-enabled healthcare network, such as the size of the developed model, the cloud architecture, the hardware capabilities and the nature of the data.

7.5 Conclusion

The emotion recognition has grown in significance since it ensures specialized supports in a variety of contexts, including personalized healthcare. In this chapter, we introduced a MSF-Net for emotion recognition from multimodal data in scalable healthcare. In this fusion network, the visual features are extracted using transformer, LBP, and ORB, which are fused together and sent to the triplet attention module to capture cross-dimensional interactions from the deep features and computer vision-based features. The Bi-GRU network modeled the physiological signals to obtain the significant temporal dependencies. Finally, the retrieved facial patterns and temporal features are fused through feature-level fusion and fed into the FC layer for multimodal emotion recognition. The effectiveness of the proposed fusion network has been extensively tested using two benchmark databases: BioVid Emo DB and MGEED. The outcomes clearly demonstrate that the efficiency of multimodal discrete emotion recognition is significantly improved by the multi-stage fusion of the extracted features through feature-level fusion. Finally, we deployed the proposed multi-stage fusion architecture in IoT network to monitor multiple patients in scalable healthcare.

Chapter 8

Conclusions and Future Works

8.1 Conclusions

Smart healthcare provides an effective, and easily deployable health monitoring system that can ensure quality healthcare services at a fraction of the cost currently incurred by hospitals or assisted living centers. As a case study, we considered HAR and emotion recognition as it has become a prominent research area due to the fast development of various emerging computing paradigms and technologies as well as plays a crucial role in IoHT applications by monitoring the human behaviors.

In this thesis, we presented a fusion approach for multi-modal HAR that could handle different nature of data collected from both wearable and stationary devices in smart healthcare. The DL-based fusion technique used CNNs and ConvLSTM to extract high-level features from the raw images, and time-series sensors signals, respectively. In this fusion architecture, the extracted features are fused through SA mechanisms that reflect the prominent features of the original data. Moreover, we proposed a multi-level feature fusion approach for multimodal HAR that efficiently processes and recognizes human activities recorded by multiple sensors and stationary cameras embedded in a variety of IoT devices. Considering the multiple sensor information along with visual data recorded from various wearable sensors, a fusion approach is developed to enhance the performance of HAR in smart healthcare applications. The proposed system consists of several blocks for multimodal HAR. The multi-head CNN with CBAM extracts the channel information and spatial information from visual data. A CAM and SAM are added to each head of CNN to extract features in the channel and spatial dimensions. The extracted features from each head are fused to get all the attributes of the visual data. To handle the sequential data in the form of time-series collected by multiple sensors, ConvLSTM architecture is utilized that extracts the

temporal attributes from raw time-series data. Afterwards, the output of these two blocks is fused and sent to a fully connected layer through the GAP layer for recognition purposes.

Considering emotion recognition, we proposed a novel emotion recognition framework from multimodal data using DL architectures aiming to develop a connected healthcare platform to monitor the patient's health status. This work used two different data modalities, such as videos and physiological signals, to develop the framework. Initially, the frames are generated from the videos, and face recognition is performed on the extracted frames. To recognize the faces, we exploited two face detection techniques, HOG and ResNet-34, to see the effects of facial features on emotion recognition. To extract significant patterns from facial expressions, a DSCNN architecture is exploited. Moreover, physiological data including tEMG, ECG, and SCL are input into the Bi-LSTM network to retrieve the relevant patterns. The obtained attributes from the two data models are combined through the soft attention technique. The soft attention approach put importance on the most significant patterns while ignoring the relatively less significant features. Then, these features are passed through the GAP operation to the classification block for emotion classification. Further, we introduced a MSF-Net for multimodal emotion recognition in scalable healthcare. In this system, the physiological signals are input into the Bi-GRU network to extract the significant temporal features through both directions. Considering the visual data, the ViT retrieved the robust visual features. At the same time, the LBP and ORB feature descriptors are exploited to extract the computer vision-based important features from the facial expression. To emphasize the important visual modality in the fusion procedure, we simultaneously fused visual features extracted from the ViT module along with the LBP and ORB modules. To dynamically concentrate on more relevant features, we introduced to fuse multimodal visual features through attention mechanism, called triplet attention module that fuses the visual signals using three branch architecture. Afterwards, the MSF-Net aggregated the extracted multimodal features (both temporal and visual) and efficiently recognizes the emotional categories.

Finally, the experiments and evaluations have been done with large multimodal HAR and emotion recognition datasets to exhibit the robustness of the developed fusion approaches along with baseline architectures and existing SOTA methods to verify their performances. Lastly, we deploy the proposed fusion networks in the IoT network to facilitate real-time monitoring of the human behaviors in connected healthcare.

8.2 Future Works

In considering the future, we come across multiple possibilities for expanding the studies provided in this thesis. The major future research directions include the followings:

- **Considering more human behaviors in smart healthcare**: In this dissertation, we mainly focused on two key aspects of human behavior, including HAR and emotion recognition for monitoring the patient's health concerns in connected healthcare. In future work, we can include more health parameters to develop more efficient healthcare system using AI models and multimodal sensors. The other human behaviors that we can consider include sleep disorder detection, meal and medicine intake monitoring to assist chronic disease patients, particularly those with diabetes, distress level prediction to provide the patient's mental health status, and monitoring of some biophysical health behaviors (HR, BP, body temperature, and SpO_2).

- **Incorporate more data modalities for reliable recognition**: We mostly exploited visual and physiological data for HAR and emotion recognition in this thesis. We can investigate adding more data modalities in the future to enhance the reliability of fusion systems for recognizing human behaviors. For instance, integrating acoustic features and text data might help to provide more detailed insights into emotions and mental health disorders. A few more data modalities, such as ambient data (temperature and light) and wearable sensor data (GPS), could provide insights about emotion recognition and activity tracking.

- **Deploy LLMs for efficient recognition**: The next future study could involve using LLMs such as BERT and GPT to enhance the performance of recognition systems. LLMs are a preferable alternative for detecting high-level patterns in multimodal signals and developing context-aware recognition frameworks. These models are able to handle unstructured information, such as medical records or patient comments, and provide more personalized healthcare recommendations.

- **Cascaded TinyML architecture for smart healthcare**: Reduced power consumption is one of the advantages of using tiny systems. HAR and emotion recognition systems that can classify and recognize a large number of actions/status can have huge resource requirements. However, these systems do not need to be always-on. For example, in case of tiny video classification systems, if there is no person within the view of the camera, the larger action/status recognition model does not need to be active. Thus using a cascade system where a much lighter person detection model detects whether a person is present in the camera's point of view and controls the action/status classification model based on that information can reduce the total power consumption of the system. In the future, we will design and implement such power efficient cascading fusion systems for tiny devices.

- **Ensure lower communication bandwidth for data transfer in smart healthcare**: Another potential future research direction could be to reduce the communication bandwidth required to transfer data from IoT devices to cloud servers and vice versa. This could be

happen by deploying some approaches, such as data compression, edge computing, and federated learning to process the majority of the data locally rather than sending it all to a central server. Lightweight AI architectures and adaptive sampling methodologies could be deployed to process data locally, reducing the workload on cloud servers and making the healthcare system more scalable and efficient in resource-constrained environments.

- **Deploy fully functional smart healthcare system**: The final study could involve deploying a fully functional smart healthcare system to evaluate our proposed AI-driven fusion architectures in real-world applications such as smart homes, AAL, and hospitals. This deployment may arise a few practical concerns including reliability, user interface, and data security. By focusing on scalability and adaptability to real-time patient feedback, we can develop a system that effectively facilitates daily healthcare operations.

References

[1] World Health Organization. *World report on ageing and health*. World Health Organization, Geneva, Switzerland, 2015.

[2] Davide Calvaresi, Daniel Cesarini, Paolo Sernani, Mauro Marinoni, Aldo Franco Dragoni, and Arnon Sturm. Exploring the ambient assisted living domain: a systematic review. *Journal of Ambient Intelligence and Humanized Computing*, 8:239–257, 2017.

[3] Mike Patton. Us health care costs rise faster than inflation. *Retrieved from Forbes: https://scholar. google. com/scholar*, 2015.

[4] Kefeng Wei, Lincong Zhang, Yi Guo, and Xin Jiang. Health monitoring based on internet of medical things: architecture, enabling technologies, and applications. *IEEE Access*, 8:27468–27478, 2020.

[5] Erikson Júlio De Aguiar, Bruno S Faiçal, Bhaskar Krishnamachari, and Jó Ueyama. A survey of blockchain-based strategies for healthcare. *ACM Computing Surveys*, 53(2):1–27, 2020.

[6] Kun-Hsing Yu, Andrew L Beam, and Isaac S Kohane. Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10):719–731, 2018.

[7] Gonçalo Marques, Rui Pitarma, Nuno M. Garcia, and Nuno Pombo. Internet of things architectures, technologies, applications, challenges, and future directions for enhanced living environments and healthcare systems: a review. *Electronics*, 8(10):1081, 2019.

[8] Fatima Alshehri and Ghulam Muhammad. A comprehensive survey of the internet of things (iot) and ai-based smart healthcare. *IEEE Access*, 9:3660–3678, 2020.

[9] Shwet Ketu and Pramod Kumar Mishra. Internet of healthcare things: A contemporary survey. *Journal of Network and Computer Applications*, 192:103179, 2021.

[10] Zarlish Ashfaq, Abdur Rafay, Rafia Mumtaz, Syed Mohammad Hassan Zaidi, Hadia Saleem, Syed Ali Raza Zaidi, Sadaf Mumtaz, and Ayesha Haque. A review of enabling technologies for internet of medical things (iomt) ecosystem. *Ain Shams Engineering Journal*, 13(4):101660, 2022.

[11] Rytis Maskeliūnas, Robertas Damaševičius, and Sagiv Segal. A review of internet of things technologies for ambient assisted living environments. *Future Internet*, 11(12):259, 2019.

[12] Joan Cahill, Raul Portales, Sean McLoughin, Nithia Nagan, Braden Henrichs, and Sean Wetherall. Iot/sensor-based infrastructures promoting a sense of home, independent living, comfort and wellness. *Sensors*, 19(3):485, 2019.

[13] Argyro Mavrogiorgou, Athanasios Kiourtis, Konstantinos Perakis, Stamatios Pitsios, and Dimosthenis Kyriazis. Iot in healthcare: Achieving interoperability of high-quality data acquired by iot medical devices. *Sensors*, 19(9):1978, 2019.

[14] Hany Abdelghaffar and Mohamed Abousteit. Internet of things (iot) interoperability success criteria. *International Journal of Enterprise Information Systems*, 17(1):85–105, 2021.

[15] Zahra Nasiri Aghdam, Amir Masoud Rahmani, and Mehdi Hosseinzadeh. The role of the internet of things in healthcare: Future trends and challenges. *Computer Methods and Programs in Biomedicine*, 199:105903, 2021.

[16] Mohammad Nuruzzaman Bhuiyan, Md Mahbubur Rahman, Md Masum Billah, and Dipanita Saha. Internet of things (iot): A review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities. *IEEE Internet of Things Journal*, 8(13):10474–10498, 2021.

[17] Mamata Rath and Binod Pattanayak. Technological improvement in modern health care applications using internet of things (iot) and proposal of novel health care approach. *International Journal of Human Rights in Healthcare*, 12(2):148–162, 2019.

[18] Bikash Pradhan, Saugat Bhattacharyya, and Kunal Pal. Iot-based applications in healthcare devices. *Journal of Healthcare Engineering*, 2021(1):6632599, 2021.

[19] Lorian Hardcastle and Ubaka Ogbogu. Virtual care: Enhancing access or harming care? *Healthcare Management Forum*, 33(6):288–292, 2020.

[20] Marc Jayson Baucas, Petros Spachos, and Stefano Gregori. Internet-of-things devices and assistive technologies for health care: applications, challenges, and opportunities. *IEEE Signal Processing Magazine*, 38(4):65–77, 2021.

[21] Hongxu Zhu, Chung Kit Wu, Cheon Hoi Koo, Yee Ting Tsang, Yucheng Liu, Hao Ran Chi, and Kim-Fung Tsang. Smart healthcare in the era of internet-of-things. *IEEE Consumer Electronics Magazine*, 8(5):26–30, 2019.

[22] Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma, and Sandeep Kaushik. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 6(1):1–25, 2019.

[23] Omar Ali, Anup Shrestha, Jeffrey Soar, and Samuel Fosso Wamba. Cloud computing-enabled healthcare opportunities, issues, and applications: A systematic review. *International Journal of Information Management*, 43:146–158, 2018.

[24] Harleen Kaur, Mohd Atif, and Ritu Chauhan. An internet of healthcare things (ioht)-based healthcare monitoring system. In *Advances in Intelligent Computing and Communication: Proceedings of ICAC 2019*, pages 475–482. Springer, 2020.

[25] Gulraiz J Joyia, Rao M Liaqat, Aftab Farooq, and Saad Rehman. Internet of medical things (iomt): Applications, benefits and future challenges in healthcare domain. *Journal of Communications*, 12(4):240–247, 2017.

[26] Duarte Dias and João Paulo Silva Cunha. Wearable health devices—vital sign monitoring, systems and technologies. *Sensors*, 18(8):2414, 2018.

[27] F John Dian, Reza Vahidnia, and Alireza Rahmati. Wearables and the internet of things (iot), applications, opportunities, and challenges: A survey. *IEEE Access*, 8:69200–69211, 2020.

[28] Wenhao Zhao, Shulin Tian, Lei Huang, Ke Liu, Lijuan Dong, and Jinhong Guo. A smartphone-based biomedical sensory system. *Analyst*, 145(8):2873–2891, 2020.

[29] Tai Le, Floranne Ellington, Tao-Yi Lee, Khuong Vo, Michelle Khine, Sandeep Kumar Krishnan, Nikil Dutt, and Hung Cao. Continuous non-invasive blood pressure monitoring: a methodological review on measurement techniques. *IEEE Access*, 8:212478–212498, 2020.

[30] Sarmad Mehrdad, Yao Wang, and S Farokh Atashzar. Perspective: Wearable internet of medical things for remote tracking of symptoms, prediction of health anomalies, implementation of preventative measures, and control of virus spread during the era of covid-19. *Frontiers in Robotics and AI*, 8:610653, 2021.

[31] Hamid Mukhtar, Saeed Rubaiee, Moez Krichen, and Roobaea Alroobaea. An iot framework for screening of covid-19 using real-time data from wearable sensors. *International Journal of Environmental Research and Public Health*, 18(8):4022, 2021.

[32] Ivo Sérgio Guimarães Brites, Lídia Martins da Silva, Jorge Luis Victória Barbosa, Sandro José Rigo, Sérgio Duarte Correia, and Valderi Reis Quietinho Leithardt. Machine learning and iot applied to cardiovascular diseases identification through heart sounds: A literature review. *Informatics*, 8(4):73, 2021.

[33] Neha Sharma and Ashima Singh. Diabetes detection and prediction using machine learning/iot: A survey. In *Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I 2*, pages 471–479. Springer, 2019.

[34] Tatsuya Morita, Kenta Taki, Manato Fujimoto, Hirohiko Suwa, Yutaka Arakawa, and Keiichi Yasumoto. Ble beacon-based activity monitoring system toward automatic generation of daily report. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 788–793. IEEE, 2018.

[35] Aitor Almeida, Rubén Mulero, Piercosimo Rametta, Vladimir Urošević, Marina Andrić, and Luigi Patrono. A critical analysis of an iot—aware aal system for elderly monitoring. *Future Generation Computer Systems*, 97:598–619, 2019.

[36] Yeon-Hwan Park, Hee Kyung Chang, Min Hye Lee, and Seong Hyeon Lee. Community-dwelling older adults' needs and acceptance regarding the use of robot technology to assist with daily living performance. *BMC Geriatrics*, 19:1–9, 2019.

[37] Karthik Kumar Santhanaraj and Ramya MM. A survey of assistive robots and systems for elderly care. *Journal of Enabling Technologies*, 15(1):66–72, 2021.

[38] Toru Kobayashi, Kazushige Katsuragi, Taishi Miyazaki, and Kenichi Arai. Social media intermediation robot for elderly people using external cloud-based services. In *2017 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, pages 31–38. IEEE, 2017.

[39] David Fischinger, Peter Einramhof, Konstantinos Papoutsakis, Walter Wohlkinger, Peter Mayer, Paul Panek, Stefan Hofmann, Tobias Koertner, Astrid Weiss, Antonis Argyros, et al. Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. *Robotics and Autonomous Systems*, 75:60–78, 2016.

[40] Andrej Grgurić, Miran Mošmondor, and Darko Huljenić. The smarthabits: an intelligent privacy-aware home care assistance system. *Sensors*, 19(4):907, 2019.

[41] Xanthi S Papageorgiou, Georgia Chalvatzaki, Athanasios C Dometios, Costas S Tzafestas, and Petros Maragos. Intelligent assistive robotic systems for the elderly: two real-life use cases. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 360–365, 2017.

[42] Higinio Mora, David Gil, Rafael Munoz Terol, Jorge Azorín, and Julian Szymanski. An iot-based computational framework for healthcare monitoring in mobile environments. *Sensors*, 17(10):2302, 2017.

[43] Harishchandra Dubey, Admir Monteiro, Nicholas Constant, Mohammadreza Abtahi, Debanjan Borthakur, Leslie Mahler, Yan Sun, Qing Yang, Umer Akbar, and Kunal Mankodiya. Fog computing in medical internet-of-things: architecture, implementation, and applications. *Handbook of Large-Scale Distributed Computing in Smart Healthcare*, pages 281–321, 2017.

[44] Anand Paul, Hameed Pinjari, Won-Hwa Hong, Hyun Cheol Seo, and Seungmin Rho. Fog computing-based iot for health monitoring system. *Journal of Sensors*, 2018(1):1386470, 2018.

[45] Thaha Muhammed, Rashid Mehmood, Aiiad Albeshri, and Iyad Katib. Ubehealth: A personalized ubiquitous cloud and edge-enabled networked healthcare system for smart cities. *IEEE Access*, 6:32258–32285, 2018.

[46] Nourah Janbi, Iyad Katib, Aiiad Albeshri, and Rashid Mehmood. Distributed artificial intelligence-as-a-service (daiaas) for smarter ioe and 6g environments. *Sensors*, 20(20):5796, 2020.

[47] Syed Umar Amin and M Shamim Hossain. Edge intelligence and internet of things in healthcare: A survey. *IEEE Access*, 9:45–59, 2020.

[48] Arijit Saha, Ruhul Amin, Sourav Kunal, Satyanarayana Vollala, and Sanjeev K Dwivedi. Review on "blockchain technology based medical healthcare system with privacy issues". *Security and Privacy*, 2(5):e83, 2019.

[49] Aiqing Zhang and Xiaodong Lin. Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain. *Journal of Medical Systems*, 42(8):140, 2018.

[50] Hongyu Li, Liehuang Zhu, Meng Shen, Feng Gao, Xiaoling Tao, and Sheng Liu. Blockchain-based data preservation system for medical data. *Journal of Medical Systems*, 42:1–13, 2018.

[51] Yaxian Ji, Junwei Zhang, Jianfeng Ma, Chao Yang, and Xin Yao. Bmpls: Blockchain-based multi-level privacy-preserving location sharing scheme for telecare medical information systems. *Journal of Medical Systems*, 42:1–13, 2018.

[52] Sonal Gahlot, SRN Reddy, and Dinesh Kumar. Review of smart health monitoring approaches with survey analysis and proposed framework. *IEEE Internet of Things Journal*, 6(2):2116–2127, 2018.

[53] Mohammed Talal, AA Zaidan, BB Zaidan, Ahmed Shihab Albahri, Abdullah Hussein Alamoodi, Osamah Shihab Albahri, MA Alsalem, Chen Kim Lim, Kian Lam Tan, WL Shir, et al. Smart home-based iot for real-time and secure remote health monitoring of triage and priority system using body sensors: Multi-driven systematic review. *Journal of Medical Systems*, 43:1–34, 2019.

[54] Sumit Majumder, Tapas Mondal, and M Jamal Deen. Wearable sensors for remote health monitoring. *Sensors*, 17(1):130, 2017.

[55] Md Milon Islam, Saifuddin Mahmud, LJ Muhammad, Md Rabiul Islam, Sheikh Nooruddin, and Safial Islam Ayon. Wearable technology to assist the patients infected with novel coronavirus (covid-19). *SN Computer Science*, 1:1–9, 2020.

[56] Md Milon Islam, Ashikur Rahaman, and Md Rashedul Islam. Development of smart healthcare monitoring system in iot environment. *SN Computer Science*, 1:1–11, 2020.

[57] Konguvel Elango and Kannan Muniandi. A low-cost wearable remote healthcare monitoring system. In *Role of Edge Analytics in Sustainable Smart City Development: Challenges and Solutions*, pages 219–242. Wiley Online Library, 2020.

[58] Mustafa A Al-Sheikh and Ibrahim A Ameen. Design of mobile healthcare monitoring system using iot technology and cloud computing. *IOP Conference Series: Materials Science and Engineering*, 881(1):012113, 2020.

[59] Ajaegbu Chigozirim, Nwaocha Vivian, Nzenwata Uchenna, and Adesegun Oreoluwa. A patient monitoring system using internet of things technology. *Ingénierie des systèmes d Inf.*, 25(3):351–357, 2020.

[60] Subasish Mohapatra, Suchismita Mohanty, and Subhadarshini Mohanty. Smart healthcare: an approach for ubiquitous healthcare management using iot. In *Big data analytics for intelligent healthcare management*, pages 175–196. Elsevier, 2019.

[61] K Narendra Swaroop, Kavitha Chandu, Ramesh Gorrepotu, and Subimal Deb. A health monitoring system for vital signs using iot. *Internet of Things*, 5:116–129, 2019.

[62] Mohammed Al-Khafajiy, Thar Baker, Carl Chalmers, Muhammad Asim, Hoshang Kolivand, Muhammad Fahim, and Atif Waraich. Remote health monitoring of elderly through wearable sensors. *Multimedia Tools and Applications*, 78(17):24681–24706, 2019.

[63] Neelkamal Semwal, Mrinmay Mukherjee, Chanchal Raj, and Wasim Arif. An iot based smart e-health care system. *Journal of Information and Optimization Sciences*, 40(8):1787–1800, 2019.

[64] Abhishek Kumar, Gaurav Chattree, and Sasikumar Periyasamy. Smart healthcare monitoring system. *Wireless Personal Communications*, 101:453–463, 2018.

[65] Jie Wan, Munassar AAH Al-awlaqi, MingSong Li, Michael O'Grady, Xiang Gu, Jin Wang, and Ning Cao. Wearable iot enabled real-time health monitoring system. *EURASIP Journal on Wireless Communications and Networking*, 2018(1):1–10, 2018.

[66] Sumit Majumder and M Jamal Deen. Smartphone sensors for health monitoring and diagnosis. *Sensors*, 19(9):2164, 2019.

[67] Ruwini Edirisinghe, Andrew Stranieri, and Nilmini Wickramasinghe. A taxonomy for mhealth. In *Virtual and mobile healthcare: Breakthroughs in research and practice*, pages 823–842. IGI Global, 2020.

[68] Abhinav Bassi, Oommen John, Devarsetty Praveen, Pallab K Maulik, Rajmohan Panda, Vivekanand Jha, et al. Current status and future directions of mhealth interventions for health system strengthening in india: systematic review. *JMIR mHealth and uHealth*, 6(10):e11440, 2018.

[69] Gaobo Zhang, Zhen Mei, Yuan Zhang, Xuesheng Ma, Benny Lo, Dongyi Chen, and Yuanting Zhang. A noninvasive blood glucose monitoring system based on smartphone ppg signal processing and machine learning. *IEEE Transactions on Industrial Informatics*, 16(11):7209–7218, 2020.

[70] Andrea Nemcova, Ivana Jordanova, Martin Varecka, Radovan Smisek, Lucie Marsanova, Lukas Smital, and Martin Vitek. Monitoring of heart rate, blood oxygen saturation,

and blood pressure using a smartphone. *Biomedical Signal Processing and Control*, 59:101928, 2020.

[71] Maha Alafeef and Mohammad Fraiwan. Smartphone-based respiratory rate estimation using photoplethysmographic imaging and discrete wavelet transform. *Journal of Ambient Intelligence and Humanized Computing*, 11:693–703, 2020.

[72] Fatemehsadat Tabei, Jon Michael Gresham, Behnam Askarian, Kwanghee Jung, and Jo Woon Chong. Cuff-less blood pressure monitoring system using smartphones. *IEEE Access*, 8:11534–11545, 2020.

[73] Jishnu Dey, Aman Gaurav, and Vijay N Tiwari. Instabp: cuff-less blood pressure monitoring on smartphone using single ppg sensor. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 5002–5005. IEEE, 2018.

[74] Edward Jay Wang, Junyi Zhu, Mohit Jain, Tien-Jui Lee, Elliot Saba, Lama Nachman, and Shwetak N Patel. Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera. In *Proceedings of the 2018 CHI conference on human factors in computing Systems*, pages 1–9, 2018.

[75] Maha Alafeef. Smartphone-based photoplethysmographic imaging for heart rate monitoring. *Journal of Medical Engineering & Technology*, 41(5):387–395, 2017.

[76] Jean-Pierre Lomaliza and Hanhoon Park. A highly efficient and reliable heart rate monitoring system using smartphone cameras. *Multimedia Tools and Applications*, 76:21051–21071, 2017.

[77] Abdul Qayyum, Aamir Saeed Malik, Aliyu Nuhu Shuaibu, and Nadeem Nasir. Estimation of non-contact smartphone video-based vital sign monitoring using filtering and standard color conversion techniques. In *2017 IEEE Life Sciences Conference (LSC)*, pages 202–205. IEEE, 2017.

[78] Marcus AG Santos, Roberto Munoz, Rodrigo Olivares, Pedro P Rebouças Filho, Javier Del Ser, and Victor Hugo C de Albuquerque. Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook. *Information Fusion*, 53:222–239, 2020.

[79] Herbert Endres, Marta Indulska, and Arunava Ghosh. Unlocking the potential of industrial internet of things (iiot) in the age of the industrial metaverse: Business models and challenges. *Industrial Marketing Management*, 119:90–107, 2024.

[80] Wei Li, Yuanbo Chai, Fazlullah Khan, Syed Rooh Ullah Jan, Sahil Verma, Varun G Menon, fnm Kavita, and Xingwang Li. A comprehensive survey on machine learning-based big data analytics for iot-enabled smart healthcare system. *Mobile Networks and Applications*, 26:234–252, 2021.

[81] Kanchan Pradhan and Priyanka Chawla. Medical internet of things using machine learning algorithms for lung cancer detection. *Journal of Management Analytics*, 7(4):591–623, 2020.

[82] Worldometer. Covid-19 coronavirus pandemic. https://www.worldometers.info/coronavirus/, 2021. Accessed: Jan. 22, 2021.

[83] Xueyan Mei, Hao-Chih Lee, Kai-yue Diao, Mingqian Huang, Bin Lin, Chenyu Liu, Zongyu Xie, Yixuan Ma, Philip M Robson, Michael Chung, et al. Artificial intelligence–enabled rapid diagnosis of patients with covid-19. *Nature Medicine*, 26(8):1224–1228, 2020.

[84] Amanullah Asraf, Md Zabirul Islam, Md Rezwanul Haque, and Md Milon Islam. Deep learning applications to combat novel coronavirus (covid-19) pandemic. *SN Computer Science*, 1:1–7, 2020.

[85] Dac-Nhuong Le, Velmurugan Subbiah Parvathy, Deepak Gupta, Ashish Khanna, Joel JPC Rodrigues, and K Shankar. Iot enabled depthwise separable convolution neural network with deep support vector machine for covid-19 diagnosis and classification. *International journal of machine learning and cybernetics*, pages 1–14, 2021.

[86] Joseph Paul Cohen, Paul Morrison, Lan Dao, Karsten Roth, Tim Q Duong, and Marzyeh Ghassemi. Covid-19 image data collection: Prospective predictions are the future. *arXiv:2006.11988*, 2020.

[87] Alfonso P Ramallo-González, Aurora González-Vidal, and Antonio F Skarmeta. Ciotvid: Towards an open iot-platform for infective pandemic diseases such as covid-19. *Sensors*, 21(2):484, 2021.

[88] Imran Ahmed, Awais Ahmad, and Gwanggil Jeon. An iot-based deep learning framework for early assessment of covid-19. *IEEE Internet of Things Journal*, 8(21):15855–15862, 2020.

[89] Mwaffaq Otoom, Nesreen Otoum, Mohammad A Alzubaidi, Yousef Etoom, and Rudaina Banihani. An iot-based framework for early identification and monitoring of covid-19 cases. *Biomedical Signal Processing and Control*, 62:102149, 2020.

[90] Semantic Scholar. Covid-19 open research dataset (cord-19). *Retrieved March*, 28:2020, 2020.

[91] Nora El-Rashidy, Shaker El-Sappagh, SM Riazul Islam, Hazem M El-Bakry, and Samir Abdelrazek. End-to-end deep learning framework for coronavirus (covid-19) detection and monitoring. *Electronics*, 9(9):1439, 2020.

[92] Armagan Altun and Okan Erdogan. A chest x-ray image: total opacification of the hemithorax. *Cardiology in Review*, 11(6):301–302, 2003.

[93] Swapnili Karmore, Rushikesh Bodhe, Fadi Al-Turjman, R Lakshmana Kumar, and Sofia K Pillai. Iot-based humanoid software for identification and diagnosis of covid-19 suspects. *IEEE Sensors Journal*, 22(18):17490–17496, 2020.

[94] Dan Cacovean, Irina Ioana, and Gabriela Nitulescu. Iot system in diagnosis of covid-19 patients. *Informatica Economica*, 24(2):75–89, 2020.

[95] Kumar KR, Niveditha VR, S Magesh, G Magesh, and Shanmugasundaram Marappan. Monitoring and analysis of the recovery rate of covid-19 positive cases to prevent dangerous stage using iot and sensors. *International Journal of Pervasive Computing and Communications*, 18(4):365–375, 2022.

[96] Hager Ahmed, Eman MG Younis, Abdeltawab Hendawi, and Abdelmgeid A Ali. Heart disease identification from patients' social posts, machine learning solution on spark. *Future Generation Computer Systems*, 111:714–722, 2020.

[97] Stephen Kaptoge, Lisa Pennells, Dirk De Bacquer, Marie Therese Cooney, Maryam Kavousi, Gretchen Stevens, Leanne Margaret Riley, Stefan Savin, Taskeen Khan, Servet Altay, et al. World health organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *The Lancet Global Health*, 7(10):e1332–e1345, 2019.

[98] Simanta Shekhar Sarmah. An efficient iot-based patient monitoring and heart disease prediction system using deep learning modified neural network. *IEEE Access*, 8:135784–135797, 2020.

[99] UCI Machine Learning Repository. Heart disease data set. https://archive.ics.uci.edu/ml/datasets/heart+Disease, 2020. Accessed: Jan. 15, 2020.

[100] Farman Ali, Shaker El-Sappagh, SM Riazul Islam, Daehan Kwak, Amjad Ali, Muhammad Imran, and Kyung-Sup Kwak. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Information Fusion*, 63:208–222, 2020.

[101] Omer Deperlioglu, Utku Kose, Deepak Gupta, Ashish Khanna, and Arun Kumar Sangaiah. Diagnosis of heart diseases by a secure internet of health things system based on autoencoder deep neural network. *Computer Communications*, 162:31–50, 2020.

[102] Wenjie Zhang, Jiqing Han, and Shiwen Deng. Heart sound classification based on scaled spectrogram and tensor decomposition. *Expert Systems with Applications*, 84:220–231, 2017.

[103] Huseyin Coskun, Tuncay Yigit, and Omer Deperlioglu. Effect of filter selection on classification of extrasystole heart sounds via mobile devices. In *2016 IEEE 10th international conference on application of information and communication technologies (AICT)*, pages 1–5. IEEE, 2016.

[104] Mohammad Ayoub Khan and Fahad Algarni. A healthcare monitoring system for the diagnosis of heart disease in the iomt cloud environment using msso-anfis. *IEEE Access*, 8:122259–122269, 2020.

[105] Kaggle. Framingham heart study dataset. https://www.kaggle.com/amanajmera1/framingham-heart-study-dataset, 2020. Accessed: Jan. 15, 2020.

[106] Mohammad Ayoub Khan. An iot framework for heart disease prediction based on mdcnn classifier. *IEEE Access*, 8:34717–34727, 2020.

[107] Shreshth Tuli, Nipam Basumatary, Sukhpal Singh Gill, Mohsen Kahani, Rajesh Chand Arya, Gurpreet Singh Wander, and Rajkumar Buyya. Healthfog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated iot and fog computing environments. *Future Generation Computer Systems*, 104:187–200, 2020.

[108] Thanh-Hai Nguyen, Thanh-Nghia Nguyen, and Thanh-Tam Nguyen. A deep learning framework for heart disease classification in an iots-based system. In *A handbook of internet of things in biomedical and cyber physical system*, pages 217–244. Springer, 2020.

[109] M Ganesan and N Sivakumar. Iot based heart disease prediction and diagnosis model for healthcare using machine learning models. In *2019 IEEE international conference on system, computation, automation and networking (ICSCAN)*, pages 1–5. IEEE, 2019.

[110] UCI Machine Learning Repository. Statlog (heart) data set. http://archive.ics.uci.edu/ml/datasets/statlog+(heart), 2019. Accessed: Jan. 20, 2019.

[111] Pouya Saeedi, Inga Petersohn, Paraskevi Salpea, Belma Malanda, Suvi Karuranga, Nigel Unwin, Stephen Colagiuri, Leonor Guariguata, Ayesha A Motala, Katherine Ogurtsova, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas. *Diabetes Research and Clinical Practice*, 157:107843, 2019.

[112] Pramod Bide and Abhishek Padalkar. Survey on diabetes mellitus and incorporation of big data, machine learning and iot to mitigate it. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 1–10. IEEE, 2020.

[113] Riya Biswas, Souvik Pal, Nguyen Ha Huy Cuong, and Arindam Chakrabarty. A novel iot-based approach towards diabetes prediction using big data. In *Intelligent Computing in Engineering: Select Proceedings of RICE 2019*, pages 163–170. Springer, 2020.

[114] Amine Rghioui, Assia Naja, Jaime Lloret Mauri, and Abedlmajid Oumnad. An iot based diabetic patient monitoring system using machine learning and node mcu. *Journal of Physics: Conference Series*, 1743(1):012035, 2021.

[115] Viswanatha Reddy Allugunti, C Kishor Kumar Reddy, NM Elango, and PR Anisha. Prediction of diabetes using internet of things (iot) and decision trees: Sldps. In *Intelligent Data Engineering and Analytics: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Volume 2*, pages 453–461. Springer, 2021.

[116] Md Iftekharul Alam Efat, Shoaib Rahman, and Tasnim Rahman. Iot based smart health monitoring system for diabetes patients using neural network. In *Cyber Security and Computer Science: Second EAI International Conference, ICONCS 2020, Dhaka, Bangladesh, February 15-16, 2020, Proceedings 2*, pages 593–606. Springer, 2020.

[117] Amine Rghioui, Assia Naja, and Abedlmajid Oumnad. Diabetic patients monitoring and data classification using iot application. In *2020 International Conference on Electrical and Information Technologies (ICEIT)*, pages 1–6. IEEE, 2020.

[118] Amine Rghioui, Jaime Lloret, Sandra Sendra, and Abdelmajid Oumnad. A smart architecture for diabetic patient monitoring using machine learning algorithms. *Healthcare*, 8(3):348, 2020.

[119] Brahmaji Godi, Sangeeta Viswanadham, Appala Srinuvasu Muttipati, Om Prakash Samantray, and Sasi Rekha Gadiraju. E-healthcare monitoring system using iot with machine learning approaches. In *2020 international conference on computer science, engineering and applications (ICCSEA)*, pages 1–5. IEEE, 2020.

[120] Kaggle. Pima indians diabetes database. https://www.kaggle.com/uciml/pima-indians-diabetes-database, 2020. Accessed: May. 15, 2020.

[121] P Punitha Ponmalar and CR Vijayalakshmi. Aggregation in iot for prediction of diabetics with machine learning techniques. In *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2019)*, pages 789–798. Springer, 2020.

[122] Parampreet Kaur, Neha Sharma, Ashima Singh, and Bob Gill. Ci-dpf: A cloud iot based framework for diabetes prediction. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 654–660. IEEE, 2018.

[123] Gonçalo Marques and Rui Pitarma. An indoor monitoring system for ambient assisted living based on internet of things architecture. *International Journal of Environmental Research and Public Health*, 13(11):1152, 2016.

[124] Sandra Bedaf, Patrizia Marti, Farshid Amirabdollahian, and Luc de Witte. A multi-perspective evaluation of a service robot for seniors: the voice of different stakeholders. *Disability and Rehabilitation: Assistive Technology*, 13(6):592–599, 2018.

[125] Sukhan Lee and Ahmed M Naguib. Toward a sociable and dependable elderly care robot: design, implementation and user study. *Journal of Intelligent & Robotic Systems*, 98(1):5–17, 2020.

[126] Pieter Simoens, Mauro Dragone, and Alessandro Saffiotti. The internet of robotic things: A review of concept, added value and applications. *International Journal of Advanced Robotic Systems*, 15(1):1729881418759424, 2018.

[127] David Portugal, Paulo Alvito, Eleni Christodoulou, George Samaras, and Jorge Dias. A study on the deployment of a service robot in an elderly care center. *International Journal of Social Robotics*, 11:317–341, 2019.

[128] Liyakathunisa Syed, Saima Jabeen, S Manimala, and Abdullah Alsaeedi. Smart healthcare framework for ambient assisted living using iomt and big data analytics techniques. *Future Generation Computer Systems*, 101:136–151, 2019.

[129] Stefanos Stavrotheodoros, Nikolaos Kaklanis, and Dimitrios Tzovaras. A personalized cloud-based platform for aal support to cognitively impaired elderly people. In *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18-21 November 2017*, pages 87–91. Springer, 2018.

[130] Gillian Toms, Fiona Verity, and Alison Orrell. Social care technologies for older people: Evidence for instigating a broader and more inclusive dialogue. *Technology in Society*, 58:101111, 2019.

[131] Kiyoaki Komai, Manato Fujimoto, Yutaka Arakawa, Hirohiko Suwa, Yukitoshi Kashimoto, and Keiichi Yasumoto. Beacon-based multi-person activity monitoring system for day care center. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–6. IEEE, 2016.

[132] Thomas Tegou, Ilias Kalamaras, Markos Tsipouras, Nikolaos Giannakeas, Kostantinos Votis, and Dimitrios Tzovaras. A low-cost indoor activity monitoring system for detecting frailty in older adults. *Sensors*, 19(3):452, 2019.

[133] Gonçalo Marques, Ivan Miguel Pires, Nuno Miranda, and Rui Pitarma. Air quality monitoring using assistive robots for ambient assisted living and enhanced living environments through internet of things. *Electronics*, 8(12):1375, 2019.

[134] Giovanni Diraco, Alessandro Leone, and Pietro Siciliano. A radar-based smart sensor for unobtrusive elderly monitoring in ambient assisted living applications. *Biosensors*, 7(4):55, 2017.

[135] Banos Oresti, Garcia Rafael, AH Juan, Damas Miguel, Pomares Hector, Rojas Ignacio, Saez Alejandro, and Villalonga Claudia. mhealthdroid: a novel framework for agile development of mobile health applications. *Ambient Assisted Living and Daily Activities*, pages 91–98, 2014.

[136] Yuri Quintana, FAHY Darren, Bradley Crotty, JAIN Ruchira, Eli Kaldany, Maxwel Gorenberg, Lewis Lipsitz, Diane Engorn, Jorge Rodriguez, Alex Orfanos, et al. Infosage: Supporting elders and families through online family networks. In *AMIA Annual Symposium Proceedings*, volume 2018, page 932. American Medical Informatics Association, 2018.

[137] Stefanos Stavrotheodoros, Nikolaos Kaklanis, Konstantinos Votis, and Dimitrios Tzovaras. A smart-home iot infrastructure for the support of independent living of older adults. In *Artificial Intelligence Applications and Innovations: AIAI 2018 IFIP WG 12.5 International Workshops, SEDSEAL, 5G-PINE, MHDW, and HEALTHIOT, Rhodes, Greece, May 25-27, 2018, Proceedings 14*, pages 238–249. Springer, 2018.

[138] Jayneel Vora, Sudeep Tanwar, Sudhanshu Tyagi, Neeraj Kumar, and Joel JPC Rodrigues. Faal: Fog computing-based patient monitoring system for ambient assisted living. In *2017 IEEE 19th international conference on e-health networking, applications and services (Healthcom)*, pages 1–6. IEEE, 2017.

[139] Mahmoud Nasr, Fakhri Karray, and Yuri Quintana. Human machine interaction platform for home care support system. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 4210–4215. IEEE, 2020.

[140] David Loza-Matovelle, Alexandra Verdugo, Eduardo Zalama, and Jaime Gómez-García-Bermejo. An architecture for the integration of robots and sensors for the care of the elderly in an ambient assisted living environment. *Robotics*, 8(3):76, 2019.

[141] Francisco Gomez-Donoso, Félix Escalona, Francisco Miguel Rivas, Jose Maria Cañas, and Miguel Cazorla. Enhancing the ambient assisted living capabilities with a mobile robot. *Computational Intelligence and Neuroscience*, 2019(1):9412384, 2019.

[142] David Portugal, Luís Santos, Paulo Alvito, Jorge Dias, George Samaras, and Eleni Christodoulou. Socialrobot: An interactive mobile robot for elderly home care. In *2015 IEEE/SICE International Symposium on System Integration (SII)*, pages 811–816. IEEE, 2015.

[143] Ha Manh Do, Minh Pham, Weihua Sheng, Dan Yang, and Meiqin Liu. Rish: A robot-integrated smart home for elderly care. *Robotics and Autonomous Systems*, 101:74–92, 2018.

[144] Joao Quintas, Gonçalo S Martins, Luis Santos, Paulo Menezes, and Jorge Dias. Toward a context-aware human–robot interaction framework based on cognitive development. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(1):227–237, 2018.

[145] Horst-Michael Gross, Andrea Scheidig, Steffen Müller, Benjamin Schütz, Christa Fricke, and Sibylle Meyer. Living with a mobile companion robot in your own apartment-final implementation and results of a 20-weeks field study with 20 seniors. In *2019 international conference on robotics and automation (ICRA)*, pages 2253–2259. IEEE, 2019.

[146] Alessandro Vercelli, Innocenzo Rainero, Ludovico Ciferri, Marina Boido, and Fabrizio Pirri. Robots in elderly care. *DigitCult-Scientific Journal on Digital Cultures*, 2(2):37–50, 2018.

[147] Gajo Petrović, Vladimir Dimitrieski, and Hamido Fujita. Cloud-based health monitoring system based on commercial off-the-shelf hardware. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 003713–003718. IEEE, 2016.

[148] Abdelrahman Rashed, Ahmed Ibrahim, Ahmed Adel, Bishoy Mourad, Ayman Hatem, Mostafa Magdy, Nada Elgaml, and Ahmed Khattab. Integrated iot medical platform for

remote healthcare and assisted living. In *2017 Japan-Africa Conference on Electronics, Communications and Computers (JAC-ECC)*, pages 160–163. IEEE, 2017.

[149] Mahmoud Nasr Mohamed. Scalable human-machine interaction system for real-time care in the internet of health things. Master's thesis, University of Waterloo, 2020.

[150] Bahar Farahani, Farshad Firouzi, Victor Chang, Mustafa Badaroglu, Nicholas Constant, and Kunal Mankodiya. Towards fog-driven iot ehealth: Promises and challenges of iot in medicine and healthcare. *Future Generation Computer Systems*, 78:659–676, 2018.

[151] Freddy Feria, Octavio J Salcedo Parra, and Brayan S Reyes Daza. Design of an architecture for medical applications in iot. In *Cooperative Design, Visualization, and Engineering: 13th International Conference, CDVE 2016, Sydney, NSW, Australia, October 24–27, 2016, Proceedings 13*, pages 263–270. Springer, 2016.

[152] Yuri Quintana, Bradley Crotty, Darren Fahy, Alex Orfanos, Ruchira Jain, Eli Kaldany, Lewis Lipsitz, Diane Engorn, Jorge Rodriguez, Frank Pandolfe, et al. Infosage: use of online technologies for communication and elder care. In *Building Capacity for Health Informatics in the Future*, pages 280–285. IOS Press, 2017.

[153] Rih-Teng Wu and Mohammad Reza Jahanshahi. Data fusion approaches for structural health monitoring and system identification: Past, present, and future. *Structural Health Monitoring*, 19(2):552–586, 2020.

[154] Amarjit P Kene and Sounak K Choudhury. Analytical modeling of tool health monitoring system using multiple sensor data fusion approach in hard machining. *Measurement*, 145:118–129, 2019.

[155] Rachel C King, Emma Villeneuve, Ruth J White, R Simon Sherratt, William Holderbaum, and William S Harwin. Application of data fusion techniques and technologies for wearable health monitoring. *Medical Engineering & Physics*, 42:1–12, 2017.

[156] Kai Lin, Yihui Li, Jinchuan Sun, Dongsheng Zhou, and Qiang Zhang. Multi-sensor fusion for body sensor network in medical human–robot interaction scenario. *Information Fusion*, 57:15–26, 2020.

[157] Minghua Wang, Xu Wang, Laurence T Yang, Xianjun Deng, and Lingzhi Yi. Multi-sensor fusion based intelligent sensor relocation for health and safety monitoring in bsns. *Information Fusion*, 54:61–71, 2020.

[158] Rustem Dautov, Salvatore Distefano, and Rajkumaar Buyya. Hierarchical data fusion for smart healthcare. *Journal of Big Data*, 6(1):1–23, 2019.

[159] Jigna J Hathaliya and Sudeep Tanwar. An exhaustive survey on security and privacy issues in healthcare 4.0. *Computer Communications*, 153:311–335, 2020.

[160] Xuancheng Guo, Hui Lin, Yulei Wu, and Min Peng. A new data clustering strategy for enhancing mutual privacy in healthcare iot systems. *Future Generation Computer Systems*, 113:407–417, 2020.

[161] Haik Kalantarian, Costas Sideris, Bobak Mortazavi, Nabil Alshurafa, and Majid Sarrafzadeh. Dynamic computation offloading for low-power wearable health monitoring systems. *IEEE Transactions on Biomedical Engineering*, 64(3):621–628, 2016.

[162] Danhua Zhou, Zhanying Li, Jiali Zhu, Haichuan Zhang, and Lin Hou. State of health monitoring and remaining useful life prediction of lithium-ion batteries based on temporal convolutional network. *IEEE Access*, 8:53307–53320, 2020.

[163] Muhammad Babar, Ataur Rahman, Fahim Arif, and Gwanggil Jeon. Energy-harvesting based on internet of things and big data analytics for smart health monitoring. *Sustainable Computing: Informatics and Systems*, 20:155–164, 2018.

[164] Fatemeh Mansourkiaie, Loay Sabry Ismail, Tarek Mohamed Elfouly, and Mohamed H Ahmed. Maximizing lifetime in wireless sensor network for structural health monitoring with and without energy harvesting. *IEEE Access*, 5:2383–2395, 2017.

[165] David Pollreisz and Nima TaheriNejad. Detection and removal of motion artifacts in ppg signals. *Mobile Networks and Applications*, 27(2):728–738, 2022.

[166] Syed Khairul Bashar, Dong Han, Apurv Soni, David D McManus, and Ki H Chon. Developing a novel noise artifact detection algorithm for smartphone ppg signals: Preliminary results. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 79–82. IEEE, 2018.

[167] Dong Yang, Yongqiang Cheng, Jin Zhu, Dongfei Xue, Grant Abt, Hangyang Ye, and Yonghong Peng. A novel adaptive spectrum noise cancellation approach for enhancing heartbeat rate monitoring in a wearable device. *IEEE Access*, 6:8364–8375, 2018.

[168] Dhamanpreet Kaur, Matthew Sobiesk, Shubham Patil, Jin Liu, Puran Bhagat, Amar Gupta, and Natasha Markuzon. Application of bayesian networks to generate synthetic health data. *Journal of the American Medical Informatics Association*, 28(4):801–811, 2021.

[169] Dhamanpreet Kaur, Ralph J Panos, Omar Badawi, Sanika S Bapat, Li Wang, and Amar Gupta. Evaluation of clinician interaction with alerts to enhance performance of the tele-critical care medical environment. *International Journal of Medical Informatics*, 139:104165, 2020.

[170] Yujin Oh, Sangjoon Park, and Jong Chul Ye. Deep learning covid-19 features on cxr using limited training data sets. *IEEE Transactions on Medical Imaging*, 39(8):2688–2700, 2020.

[171] Jiangdian Song, Hongmei Wang, Yuchan Liu, Wenqing Wu, Gang Dai, Zongshan Wu, Puhe Zhu, Wei Zhang, Kristen W Yeom, and Kexue Deng. End-to-end automatic differentiation of the coronavirus disease 2019 (covid-19) from viral pneumonia based on chest ct. *European Journal of Nuclear Medicine and Molecular Imaging*, 47:2516–2524, 2020.

[172] Rutwik Gulakala, Bernd Markert, and Marcus Stoffel. Generative adversarial network based data augmentation for cnn based detection of covid-19. *Scientific Reports*, 12(1):19186, 2022.

[173] G Thippa Reddy, M Praveen Kumar Reddy, Kuruva Lakshmanna, Dharmendra Singh Rajput, Rajesh Kaluri, and Gautam Srivastava. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evolutionary Intelligence*, 13:185–196, 2020.

[174] Azhar Hussein Alkeshuosh, Mariam Zomorodi Moghadam, Inas Al Mansoori, and Moloud Abdar. Using pso algorithm for producing best rules in diagnosis of heart disease. In *2017 international conference on computer and applications (ICCA)*, pages 306–311. IEEE, 2017.

[175] Anna Karen Gárate-Escamila, Amir Hajjam El Hassani, and Emmanuel Andrès. Classification models for heart disease prediction using feature selection and pca. *Informatics in Medicine Unlocked*, 19:100330, 2020.

[176] P Punithavathi, S Geetha, Marimuthu Karuppiah, SK Hafizul Islam, Mohammad Mehedi Hassan, and Kim-Kwang Raymond Choo. A lightweight machine learning-based authentication framework for smart iot devices. *Information Sciences*, 484:255–268, 2019.

[177] Hengjin Ke, Dan Chen, Tejal Shah, Xianzeng Liu, Xinhua Zhang, Lei Zhang, and Xiaoli Li. Cloud-aided online eeg classification system for brain healthcare: A case study of depression evaluation with a lightweight cnn. *Software: Practice and Experience*, 50(5):596–610, 2020.

[178] Yanxing Song, F Richard Yu, Li Zhou, Xi Yang, and Zefang He. Applications of the internet of things (iot) in smart logistics: A comprehensive survey. *IEEE Internet of Things Journal*, 8(6):4250–4274, 2020.

[179] Jeretta Horn Nord, Alex Koohang, and Joanna Paliszkiewicz. The internet of things: Review and theoretical framework. *Expert Systems with Applications*, 133:97–108, 2019.

[180] Abhishek Khanna and Sanmeet Kaur. Internet of things (iot), applications and challenges: a comprehensive review. *Wireless Personal Communications*, 114:1687–1762, 2020.

[181] Lalit Chettri and Rabindranath Bera. A comprehensive survey on internet of things (iot) toward 5g wireless systems. *IEEE Internet of Things Journal*, 7(1):16–32, 2019.

[182] Parvaneh Asghari, Amir Masoud Rahmani, and Hamid Haj Seyyed Javadi. Internet of things applications: A systematic review. *Computer Networks*, 148:241–261, 2019.

[183] Abhishek Singh, Ashish Payal, and Sourabh Bharti. A walkthrough of the emerging iot paradigm: Visualizing inside functionalities, key features, and open issues. *Journal of Network and Computer Applications*, 143:111–151, 2019.

[184] Anna Triantafyllou, Panagiotis Sarigiannidis, and Thomas D Lagkas. Network protocols, schemes, and mechanisms for internet of things (iot): Features, open challenges, and trends. *Wireless Communications and Mobile Computing*, 2018(1):5349894, 2018.

[185] Muhammad Mahtab Alam, Hassan Malik, Muhidul Islam Khan, Tamas Pardy, Alar Kuusik, and Yannick Le Moullec. A survey on the roles of communication technologies in iot-based personalized healthcare applications. *IEEE Access*, 6:36611–36631, 2018.

[186] Stephanie B Baker, Wei Xiang, and Ian Atkinson. Internet of things for smart healthcare: Technologies, challenges, and opportunities. *IEEE Access*, 5:26521–26544, 2017.

[187] Hamza Baali, Hamza Djelouat, Abbes Amira, and Faycal Bensaali. Empowering technology enabled care using iot and smart devices: A review. *IEEE Sensors Journal*, 18(5):1790–1809, 2017.

[188] Pallavi Sethi and Smruti R Sarangi. Internet of things: architectures, protocols, and applications. *Journal of Electrical and Computer Engineering*, 2017(1):9324035, 2017.

[189] Bhagya Nathali Silva, Murad Khan, and Kijun Han. Internet of things: A comprehensive review of enabling technologies, architecture, and challenges. *IETE Technical review*, 35(2):205–220, 2018.

[190] Alem Čolaković and Mesud Hadžialić. Internet of things (iot): A review of enabling technologies, challenges, and open research issues. *Computer Networks*, 144:17–39, 2018.

[191] Subramanian Balaji, Karan Nathani, and Rathnasamy Santhakumar. Iot technology, applications and challenges: a contemporary survey. *Wireless Personal Communications*, 108:363–388, 2019.

[192] Abdul Ahad, Mohammad Tahir, and Kok-Lim Alvin Yau. 5g-based smart healthcare network: architecture, taxonomy, challenges and future research directions. *IEEE Access*, 7:100747–100762, 2019.

[193] Mrinai M Dhanvijay and Shailaja C Patil. Internet of things: A survey of enabling technologies in healthcare and its applications. *Computer Networks*, 153:113–131, 2019.

[194] Jun Qi, Po Yang, Geyong Min, Oliver Amft, Feng Dong, and Lida Xu. Advanced internet of things for personalised healthcare systems: A survey. *Pervasive and Mobile Computing*, 41:132–149, 2017.

[195] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, et al. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372, 2021.

[196] Muhammad Uzair Khan, Salman Sherin, Muhammad Zohaib Iqbal, and Rubab Zahid. Landscaping systematic mapping studies in software engineering: A tertiary study. *Journal of Systems and Software*, 149:396–436, 2019.

[197] Jolle W Jolles. Broad-scale applications of the raspberry pi: A review and guide for biologists. *Methods in Ecology and Evolution*, 12(9):1562–1579, 2021.

[198] Andrii Polianytsia, Olena Starkova, and Kostiantyn Herasymenko. Survey of hardware iot platforms. In *2016 Third International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T)*, pages 152–153. IEEE, 2016.

[199] Vladimir Vujović and Mirjana Maksimović. Raspberry pi as a wireless sensor node: Performances and constraints. In *2014 37th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, pages 1013–1018. IEEE, 2014.

[200] Hari Kishan Kondaveeti, Nandeesh Kumar Kumaravelu, Sunny Dayal Vanambathina, Sudha Ellison Mathe, and Suseela Vappangi. A systematic literature review on prototyping with arduino: Applications, challenges, advantages, and limitations. *Computer Science Review*, 40:100364, 2021.

[201] Dareen K Halim, Tang Chong Ming, Ng Mow Song, and Dicky Hartono. Arduino-based ide for embedded multi-processor system-on-chip. In *2019 5th International Conference on New Media Studies (CONMEDIA)*, pages 135–138. IEEE, 2019.

[202] Sung-Min Kim, Yosoon Choi, and Jangwon Suh. Applications of the open-source hardware arduino platform in the mining industry: A review. *Applied Sciences*, 10(14):5018, 2020.

[203] João Mesquita, Diana Guimarães, Carlos Pereira, Frederico Santos, and Luís Almeida. Assessing the esp8266 wifi module for the internet of things. In *2018 IEEE 23rd international conference on emerging technologies and factory automation (ETFA)*, volume 1, pages 784–791. IEEE, 2018.

[204] Indra Gunawan, Aris Sudianto, and Muhamad Sadali. Measuring body temperature based internet of things (iot) using esp8266 and firebase. *Sisfotenika*, 11(1):91–100, 2021.

[205] Tansu Zafer Asici, Burak Karaduman, Raheleh Eslampanah, Moharram Challenger, Joachim Denil, and Hans Vangheluwe. Applying model driven engineering techniques to the development of contiki-based iot systems. In *2019 IEEE/ACM 1st International Workshop on Software Engineering Research & Practices for the Internet of Things (SERP4IoT)*, pages 25–32. IEEE, 2019.

[206] Yousaf Bin Zikria, Muhammad Khalil Afzal, Farruh Ishmanov, Sung Won Kim, and Heejung Yu. A survey on routing protocols supported by the contiki internet of things operating system. *Future Generation Computer Systems*, 82:200–219, 2018.

[207] Challouf Sabri, Lobna Kriaa, and Saidane Leila Azzouz. Comparison of iot constrained devices operating systems: A survey. In *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, pages 369–375. IEEE, 2017.

[208] Fawwad Jaskani, Saba Manzoor, Muhammad Amin, Muhammad Asif, and Muntaha Irfan. An investigation on several operating systems for internet of things. *EAI Endorsed Transactions on Creative Technologies*, 6(18), 2019.

[209] Emmanuel Baccelli, Cenk Gündoğan, Oliver Hahm, Peter Kietzmann, Martine S Lenders, Hauke Petersen, Kaspar Schleiser, Thomas C Schmidt, and Matthias Wählisch. Riot: An

open source operating system for low-end embedded devices in the iot. *IEEE Internet of Things Journal*, 5(6):4428–4440, 2018.

[210] Burak Karaduman, Moharram Challenger, Raheleh Eslampanah, Joachim Denil, and Hans Vangheluwe. Platform-specific modeling for riot based iot systems. In *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops*, pages 639–646, 2020.

[211] Christopher Hannon, Jiaqi Yan, Yuan-An Liu, and Dong Jin. A distributed virtual time system on embedded linux for evaluating cyber-physical systems. In *Proceedings of the 2019 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, pages 37–48, 2019.

[212] H Ben Mansour, L Chaarabi, and K Jelassi. Embedded linux hardware/software architecture for electrical measurement acquisition. In *Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18), Vol. 2*, pages 167–176. Springer, 2020.

[213] Charles Bell. *Windows 10 for the Internet of Things*. Springer, 2016.

[214] Zulfahmi Syahputa, M Zarlis, and Beny B Nasution. Testing real-time applications on windows 10 iot using the nyquist theory. *Journal of Physics: Conference Series*, 1255(1):012066, 2019.

[215] Mihail Lyaskov, Grisha Spasov, and Galidiya Petrova. A practical implementation of smart home energy data storage and control application based on cloud services. In *2017 XXVI International Scientific Conference Electronics (ET)*, pages 1–4. IEEE, 2017.

[216] M Udin Harun Al Rasyid, M Husni Mubarrok, and Jauari Akhmad Nur Hasim. Implementation of environmental monitoring based on kaa iot platform. *Bulletin of Electrical Engineering and Informatics*, 9(6):2578–2587, 2020.

[217] A. Kumar and A. Kumar. *6 IoT-based platform for smart farming – Kaa*, pages 109–130. De Gruyter, 2021.

[218] Lawrence Oriaghe Aghenta and Tariq Iqbal. Design and implementation of a low-cost, open source iot-based scada system using esp32 with oled, thingsboard and mqtt protocol. *AIMS Electronics and Electrical Engineering*, 4(1):57–86, 2019.

[219] Matthew Henschke, Xinzhou Wei, and Xiaowen Zhang. Data visualization for wireless sensor networks using thingsboard. In *2020 29th Wireless and Optical Communications Conference (WOCC)*, pages 1–6. IEEE, 2020.

[220] William Tärneberg, Vishal Chandrasekaran, and Marty Humphrey. Experiences creating a framework for smart traffic control using aws iot. In *Proceedings of the 9th International Conference on Utility and Cloud Computing*, pages 63–69, 2016.

[221] Luca Calderoni. Preserving context security in aws iot core. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*, pages 1–5, 2019.

[222] Paola Pierleoni, Roberto Concetti, Alberto Belli, and Lorenzo Palma. Amazon, google and microsoft solutions for iot: Architectures and a performance comparison. *IEEE Access*, 8:5455–5470, 2019.

[223] Billy Grados and Hector Bedon. Software components of an iot monitoring platform in google cloud platform: a descriptive research and an architectural proposal. In *International Conference on Applied Technologies*, pages 153–167. Springer, 2019.

[224] Nivedita Mishra and Sharnil Pandya. Internet of things applications, security challenges, attacks, intrusion detection, and future visions: A systematic review. *IEEE Access*, 9:59353–59377, 2021.

[225] Abdulrahman Alreshidi and Aakash Ahmad. Architecting software for the internet of thing based systems. *Future Internet*, 11(7):153, 2019.

[226] Fatma Alshohoumi, Mohammed Sarrab, Abdulla AlHamadani, and Dawood Al-Abri. Systematic review of existing iot architectures security and privacy issues and concerns. *International Journal of Advanced Computer Science and Applications*, 10(7), 2019.

[227] Wafa'a Kassab and Khalid A Darabkh. A–z survey of internet of things: Architectures, protocols, applications, recent advances, future directions and recommendations. *Journal of Network and Computer Applications*, 163:102663, 2020.

[228] Ian Zhou, Imran Makhdoom, Negin Shariati, Muhammad Ahmad Raza, Rasool Keshavarz, Justin Lipman, Mehran Abolhasan, and Abbas Jamalipour. Internet of things 2.0: Concepts, applications, and future directions. *IEEE Access*, 9:70961–71012, 2021.

[229] Kun Liu, Yunrui Bi, and Di Liu. Internet of things based acquisition system of industrial intelligent bar code for smart city applications. *Computer Communications*, 150:325–333, 2020.

[230] Rui Liu, Jie-ran Shen, Feng Jiao, and Ming-hao Ding. Deployment optimization of perception layer nodes in the internet of things based on nb-iot technology. In *Advanced Hybrid Information Processing: 4th EAI International Conference, ADHIP 2020, Binzhou, China, September 26-27, 2020, Proceedings, Part II 4*, pages 202–210. Springer, 2021.

[231] Partha Pratim Ray. A survey on internet of things architectures. *Journal of King Saud University-Computer and Information Sciences*, 30(3):291–319, 2018.

[232] Pranjal Upadhyay and PD Upadhyay. Internet of things-a survey. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 54(12):417–438, 2021.

[233] Khaled Salah Mohamed and Khaled Salah Mohamed. Iot networking and communication layer. In *The Era of Internet of Things: Towards a Smart World*, pages 49–70. Springer, 2019.

[234] Jasenka Dizdarević, Francisco Carpio, Admela Jukan, and Xavi Masip-Bruin. A survey of communication protocols for internet of things and related challenges of fog and cloud computing integration. *ACM Computing Surveys*, 51(6):1–29, 2019.

[235] Lozan M Abdulrahman, SR Zeebaree, Shakir Fattah Kak, MA Sadeeq, AZ Adel, Baraa Wasfi Salim, and Karzan Hussein Sharif. A state of art for smart gateways issues and modification. *Asian Journal of Research in Computer Science*, 7(4):1–13, 2021.

[236] Arun Kumar, Ming Zhao, Kai-Juan Wong, Yong Liang Guan, and Peter Han Joo Chong. A comprehensive study of iot and wsn mac protocols: Research issues, challenges and opportunities. *IEEE Access*, 6:76228–76262, 2018.

[237] Jing Zhang, Yunying Ye, Chunjin Hu, and Bin Li. Architecture design and demand analysis on application layer of standard system for ubiquitous power internet of things. *Global Energy Interconnection*, 4(3):304–314, 2021.

[238] Xiaoqing Li, Yu Lu, Xianghua Fu, and Yingjian Qi. Building the internet of things platform for smart maternal healthcare services with wearable devices and cloud computing. *Future Generation Computer Systems*, 118:282–296, 2021.

[239] Yi Qian, Dan Wu, Wei Bao, and Pascal Lorenz. The internet of things for smart cities: Technologies and applications. *IEEE Network*, 33(2):4–5, 2019.

[240] Fotios Zantalis, Grigorios Koulouras, Sotiris Karabetsos, and Dionisis Kandris. A review of machine learning and iot in smart transportation. *Future Internet*, 11(4):94, 2019.

[241] Muhammad Ayaz, Mohammad Ammad-Uddin, Zubair Sharif, Ali Mansour, and El-Hadi M Aggoune. Internet-of-things (iot)-based smart agriculture: Toward making the fields talk. *IEEE Access*, 7:129551–129583, 2019.

[242] Shancang Li, Li Da Xu, and Shanshan Zhao. 5g internet of things: A survey. *Journal of Industrial Information Integration*, 10:1–9, 2018.

[243] Marco Lombardi, Francesco Pascale, and Domenico Santaniello. Internet of things: A general overview between architectures, protocols and applications. *Information*, 12(2):87, 2021.

[244] Sudip Misra, Chandana Roy, and Anandarup Mukherjee. *Introduction to industrial internet of things and industry 4.0*. CRC Press, 2021.

[245] Muath A Obaidat, Suhaib Obeidat, Jennifer Holst, Abdullah Al Hayajneh, and Joseph Brown. A comprehensive and systematic survey on the internet of things: Security and privacy challenges, security frameworks, enabling technologies, threats, vulnerabilities and countermeasures. *Computers*, 9(2):44, 2020.

[246] Jong-Hyun Park. An intelligent service middleware based on sensors in iot environments. *International Journal of Software Engineering and Knowledge Engineering*, 30(04):523–536, 2020.

[247] Brij B Gupta and Megha Quamara. An overview of internet of things (iot): Architectural aspects, challenges, and protocols. *Concurrency and Computation: Practice and Experience*, 32(21):e4946, 2020.

[248] Jingbin Zhang, Meng Ma, Ping Wang, and Xiao-dong Sun. Middleware for the internet of things: A survey on requirements, enabling technologies, and solutions. *Journal of Systems Architecture*, 117:102098, 2021.

[249] Matteo Palmaccio, Grazia Dicuonzo, and Zhanna S Belyaeva. The internet of things and corporate business models: A systematic literature review. *Journal of Business Research*, 131:610–618, 2021.

[250] Armir Bujari, Marco Furini, Federica Mandreoli, Riccardo Martoglia, Manuela Montangero, and Daniele Ronzani. Standards, security and business models: key challenges for the iot scenario. *Mobile Networks and Applications*, 23:147–154, 2018.

[251] Jonathan de C Silva, Joel JPC Rodrigues, Jalal Al-Muhtadi, Ricardo AL Rabêlo, and Vasco Furtado. Management platforms and protocols for internet of things: A survey. *Sensors*, 19(3):676, 2019.

[252] Venetis Kanakaris and George A Papakostas. Internet of things protocols-a survey. *International Journal of Humanitarian Technology*, 1(2):101–117, 2020.

[253] Muhammad Ashar Tariq, Murad Khan, Muhammad Toaha Raza Khan, and Dongkyun Kim. Enhancements and challenges in coap—a survey. *Sensors*, 20(21):6391, 2020.

[254] Neven Nikolov. Research of mqtt, coap, http and xmpp iot communication protocols for embedded systems. In *2020 XXIX International Scientific Conference Electronics (ET)*, pages 1–4. IEEE, 2020.

[255] Godfrey A Akpakwu, Gerhard P Hancke, and Adnan M Abu-Mahfouz. Cacc: Context-aware congestion control approach for lightweight coap/udp-based internet of things traffic. *Transactions on Emerging Telecommunications Technologies*, 31(2):e3822, 2020.

[256] Sharu Bansal and Dilip Kumar. Distance-based congestion control mechanism for coap in iot. *IET Communications*, 14(19):3512–3520, 2020.

[257] Shih-Ping Hsu, Yu-Ming Chang, Shun-Ren Yang, and Ping-Han Hsieh. The design and implementation of a lightweight coap-based iot framework with smart contract security guarantee. In *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pages 1–6. IEEE, 2020.

[258] Silvio Quincozes, Tubino Emilio, and Juliano Kazienko. Mqtt protocol: fundamentals, tools and future directions. *IEEE Latin America Transactions*, 17(09):1439–1448, 2019.

[259] Biswajeeban Mishra and Attila Kertesz. The use of mqtt in m2m and iot systems: A survey. *IEEE Access*, 8:201071–201086, 2020.

[260] Manikkam Veeramanikandan and Suresh Sankaranarayanan. Publish/subscribe based multi-tier edge computational model in internet of things for latency reduction. *Journal of Parallel and Distributed Computing*, 127:18–27, 2019.

[261] Dan Dinculeană and Xiaochun Cheng. Vulnerabilities and limitations of mqtt protocol used between iot devices. *Applied Sciences*, 9(5):848, 2019.

[262] Eyhab Al-Masri, Karan Raj Kalyanam, John Batts, Jonathan Kim, Sharanjit Singh, Tammy Vo, and Charlotte Yan. Investigating messaging protocols for the internet of things (iot). *IEEE Access*, 8:94880–94911, 2020.

[263] M Faiqurahman, MM Madani, and DR Akbi. Performance of xmpp-based gateway for iot device communication services. *Jurnal Teknologi dan Sistem Komputer*, 7(4):127–133, 2019.

[264] Chi-Shiang Cho, Wei-Guang Chen, Chun-Kuei Chang Liao, Jen-Li Liao, Fung-Fei Chen, and Teh-Chen Shen. Building on the distributed energy resources iot based iec 61850 xmpp for tpc. In *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, pages 61–66. IEEE, 2019.

[265] Carlos A Garcia, Jose E Naranjo, and Marcelo V Garcia. Analysis of amqp for industrial internet of things based on low-cost automation. In *Brazilian Technology Symposium*, pages 235–244. Springer, 2019.

[266] Purvi Bhimani and Gaurang Panchal. Message delivery guarantee and status update of clients based on iot-amqp. In *Intelligent Communication and Computational Technologies: Proceedings of Internet of Things for Technological Development, IoT4TD 2017*, pages 15–22. Springer, 2018.

[267] Girija P Naik and A Umesh Bapat. A brief comparative analysis on application layer protocols of internet of things: Mqtt, coap, amqp and http. *nternational Journal of Computer Science and Mobile Computing*, 9(9):135–141, 2020.

[268] Firat Karatas and Ibrahim Korpeoglu. Fog-based data distribution service (f-dad) for internet of things (iot) applications. *Future Generation Computer Systems*, 93:156–169, 2019.

[269] Tanweer Alam, Baha Rababah, Arshad Ali, and Shamimul Qamar. Distributed intelligence at the edge on iot networks. *Annals of Emerging Technologies in Computing (AETiC)*, 4(5):1–18, 2020.

[270] Ridlo Savvidina Auliva, Ruev-Kai Sheu, Deron Liang, and Wei-Jen Wang. Iiot testbed: a dds-based emulation tool for industrial iot applications. In *2018 International conference on system science and engineering (ICSSE)*, pages 1–4. IEEE, 2018.

[271] Cenk Gündoğan, Peter Kietzmann, Thomas C Schmidt, and Matthias Wählisch. Designing a lowpan convergence layer for the information centric internet of things. *Computer Communications*, 164:114–123, 2020.

[272] Meet K Shah and LK Sharma. Study on 6lowpan routing protocols with sd aspects in iot. In *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2018 2nd International Conference on*, pages 60–65. IEEE, 2018.

[273] Praveen Kumar Malik, Naveen Bilandi, Anish Gupta, et al. Narrow band-iot and long-range technology of iot smart communication: Designs and challenges. *Computers & Industrial Engineering*, 172:108572, 2022.

[274] Emmanuel M Migabo, Karim D Djouani, and Anish M Kurien. The narrowband internet of things (nb-iot) resources management performance state of art, challenges, and opportunities. *IEEE Access*, 8:97658–97675, 2020.

[275] Riccardo Marini, Konstantin Mikhaylov, Gianni Pasolini, and Chiara Buratti. Low-power wide-area networks: Comparison of lorawan and nb-iot performance. *IEEE Internet of Things Journal*, 9(21):21051–21063, 2022.

[276] Kais Mekki, Eddy Bajic, Frederic Chaxel, and Fernand Meyer. A comparative study of lpwan technologies for large-scale iot deployment. *ICT Express*, 5(1):1–7, 2019.

[277] Mario H Castañeda Garcia, Alejandro Molina-Galan, Mate Boban, Javier Gozalvez, Baldomero Coll-Perales, Taylan Şahin, and Apostolos Kousaridas. A tutorial on 5g nr v2x communications. *IEEE Communications Surveys & Tutorials*, 23(3):1972–2026, 2021.

[278] Federica Rinaldi, Alessandro Raschella, and Sara Pizzi. 5g nr system design: A concise survey of key features and capabilities. *Wireless Networks*, 27(8):5173–5188, 2021.

[279] Xingqin Lin, Jingya Li, Robert Baldemair, Jung-Fu Thomas Cheng, Stefan Parkvall, Daniel Chen Larsson, Havish Koorapaty, Mattias Frenne, Sorour Falahati, Asbjorn Grovlen, et al. 5g new radio: Unveiling the essentials of the next generation wireless access technology. *IEEE Communications Standards Magazine*, 3(3):30–37, 2019.

[280] A Phani Kumar Reddy, M Sheeba Kumari, Vishakha Dhanwani, Akhil Kalpesh Bachkaniwala, Navin Kumar, Kasturi Vasudevan, Srinivasan Selvaganapathy, Sendil Kumar Devar, Punit Rathod, and Vinosh Babu James. 5g new radio key performance indicators evaluation for imt-2020 radio interface technology. *IEEE Access*, 9:112290–112311, 2021.

[281] Sherali Zeadally, Farhan Siddiqui, and Zubair Baig. 25 years of bluetooth technology. *Future Internet*, 11(9):194, 2019.

[282] Eduardo Garcia-Espinosa, Omar Longoria-Gandara, Ioseth Pegueros-Lepe, and Arturo Veloz-Guerrero. Power consumption analysis of bluetooth low energy commercial products and their implications for iot applications. *Electronics*, 7(12):386, 2018.

[283] Kang Eun Jeon, James She, Perm Soonsawad, and Pai Chet Ng. Ble beacons for internet of things applications: Survey, challenges, and opportunities. *IEEE Internet of Things Journal*, 5(2):811–828, 2018.

[284] Yehia R Hamdy and Ahmed I Alghannam. Evaluation of zigbee topology effect on throughput and end to end delay due to different transmission bands for iot applications. *Journal of Communications Software and Systems*, 16(3):254–259, 2020.

[285] Hamza Zemrane, Youssef Baddi, and Abderrahim Hasbi. Internet of things industry 4.0 ecosystem based on zigbee protocol. In *Advances on Smart and Soft Computing: Proceedings of ICACIN 2020*, pages 249–260. Springer, 2021.

[286] Malvika Gupta and Shweta Singh. A survey on the zigbee protocol, it's security in internet of things (iot) and comparison of zigbee with bluetooth and wi-fi. In *Applications of artificial intelligence in engineering: proceedings of first global conference on artificial intelligence and applications (GCAIA 2020)*, pages 473–482. Springer, 2021.

[287] Mostafa Haghi Kashani, Mona Madanipour, Mohammad Nikravan, Parvaneh Asghari, and Ebrahim Mahdipour. A systematic review of iot in healthcare: Applications, techniques, and trends. *Journal of Network and Computer Applications*, 192:103164, 2021.

[288] Ghulam Muhammad, Fatima Alshehri, Fakhri Karray, Abdulmotaleb El Saddik, Mansour Alsulaiman, and Tiago H Falk. A comprehensive survey on multimodal medical signals fusion for smart healthcare systems. *Information Fusion*, 76:355–375, 2021.

[289] Sureshkumar Selvaraj and Suresh Sundaravaradhan. Challenges and opportunities in iot healthcare systems: a systematic review. *SN Applied Sciences*, 2(1):139, 2020.

[290] Yazdan Ahmad Qadri, Ali Nauman, Yousaf Bin Zikria, Athanasios V Vasilakos, and Sung Won Kim. The future of healthcare internet of things: a survey of emerging technologies. *IEEE Communications Surveys & Tutorials*, 22(2):1121–1167, 2020.

[291] Vankamamidi S Naresh, Suryateja S Pericherla, Pilla Sita Rama Murty, and Sivaranjani Reddi. Internet of things in healthcare: Architecture, applications, challenges, and solutions. *Computer Systems Science & Engineering*, 35(6), 2020.

[292] Ashikur Rahaman, Md Milon Islam, Md Rashedul Islam, Muhammad Sheikh Sadi, Sheikh Nooruddin, et al. Developing iot based smart health monitoring systems: A review. *Rev. d'Intelligence Artif.*, 33(6):435–440, 2019.

[293] Gonçalo Marques, Jagriti Saini, Ivan Miguel Pires, Nuno Miranda, and Rui Pitarma. Internet of things for enhanced living environments, health and well-being: technologies, architectures and systems. In *Handbook of Wireless Sensor Networks: Issues and Challenges in Current Scenario's*, pages 616–631. Springer, 2020.

[294] Nada Y Philip, Joel JPC Rodrigues, Honggang Wang, Simon James Fong, and Jia Chen. Internet of things for in-home health monitoring systems: Current advances, challenges and future directions. *IEEE Journal on Selected Areas in Communications*, 39(2):300–310, 2021.

[295] Hossein Ahmadi, Goli Arji, Leila Shahmoradi, Reza Safdari, Mehrbakhsh Nilashi, and Mojtaba Alizadeh. The application of internet of things in healthcare: a systematic literature review and classification. *Universal Access in the Information Society*, 18:837–869, 2019.

[296] Taiyang Wu, Jean-Michel Redouté, and Mehmet Yuce. A wearable, low-power, real-time ecg monitor for smart t-shirt and iot healthcare applications. In *Advances in body area networks I: Post-conference proceedings of bodynets 2017*, pages 165–173. Springer, 2019.

[297] Tamanna Shaown, Imam Hasan, Md Muradur Rahman Mim, and Md Shohrab Hossain. Iot-based portable ecg monitoring system for smart healthcare. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pages 1–5. IEEE, 2019.

[298] Hamza Djelouat, Mohamed Al Disi, Issam Boukhenoufa, Abbes Amira, Faycal Bensaali, Christos Kotronis, Elena Politi, Mara Nikolaidou, and George Dimitrakopoulos. Real-time ecg monitoring using compressive sensing on a heterogeneous multicore edge-device. *Microprocessors and Microsystems*, 72:102839, 2020.

[299] Manju Lata Sahu, Mithilesh Atulkar, Mitul Kumar Ahirwal, and Afsar Ahamad. Iot-enabled cloud-based real-time remote ecg monitoring system. *Journal of Medical Engineering & Technology*, 45(6):473–485, 2021.

[300] Samik Basu, Anwesha Sengupta, Anindita Das, Mahasweta Ghosh, and Soma Barman. Iot-based real-time remote ecg monitoring system. In *Computers and Devices for Communication: Proceedings of CODEC 2019*, pages 23–28. Springer, 2021.

[301] Francesco Lamonaca, Eulalia Balestrieri, Ioan Tudosa, Francesco Picariello, Domenico Luca Carnì, Carmelo Scuro, Francesco Bonavolontà, Vitaliano Spagnuolo, Gioconda Grimaldi, and Antonio Colaprico. An overview on internet of medical things in blood pressure monitoring. In *2019 IEEE international symposium on medical measurements and applications (MeMeA)*, pages 1–6. IEEE, 2019.

[302] L. R. Krakoff. Home blood pressure monitoring. *Hypertension*, 63(4):670–671, Apr. 2014.

[303] Qin Xin and Jianping Wu. A novel wearable device for continuous, non-invasion blood pressure measurement. *Computational Biology and Chemistry*, 69:134–137, 2017.

[304] Bharat Singh, Shabana Urooj, Sakshi Mishra, and Surojeet Haldar. Blood pressure monitoring system using wireless technologies. *Procedia Computer Science*, 152:267–273, 2019.

[305] Puput Dani Prasetyo Adi and Akio Kitagawa. Zigbee radio frequency (rf) performance on raspberry pi 3 for internet of things (iot) based blood pressure sensors monitoring. *International Journal of Advanced Computer Science and Applications*, 10(5):18–27, 2019.

[306] Norlezah Hashim, Nurbahirah Norddin, Fakrulradzi Idris, SNIM Yusoff, and Madiha Zahari. Iot blood pressure monitoring system. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(3):1384–1390, 2020.

[307] Amina El Attaoui, Salma Largo, Abdelilah Jilbab, and Abdennaser Bourouhou. Wireless medical sensor network for blood pressure monitoring based on machine learning for real-time data classification. *Journal of Ambient Intelligence and Humanized Computing*, 12(9):8777–8792, 2021.

[308] T Mohanraj, B Arunkumar, S ShahulHammed, and R Santhosh. A review on internet of things (iot) based pulse rate, blood pressure, body temperature monitoring system. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pages 78–79. IEEE, 2020.

[309] Hiroki Ota, Minghan Chao, Yuji Gao, Eric Wu, Li-Chia Tai, Kevin Chen, Yasutomo Matsuoka, Kosuke Iwai, Hossain M Fahad, Wei Gao, et al. 3d printed "earable" smart devices for real-time detection of core body temperature. *ACS Sensors*, 2(7):990–997, 2017.

[310] Nor Aini Zakaria, Fatin Nadia Binti Mohd Saleh, and Mohd Azhar Abdul Razak. Iot (internet of things) based infant body temperature monitoring. In *2018 2nd international conference on biosignal analysis, processing and systems (ICBAPS)*, pages 148–153. IEEE, 2018.

[311] Asif A Rahimoon, Mohd Noor Abdullah, and Ishkrizat Taib. Design of a contactless body temperature measurement system using arduino. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(3):1251–1258, 2020.

[312] Wasana Boonsong, Narongrit Senajit, and Piya Prasongchan. Contactless body temperature monitoring of in-patient department (ipd) using 2.4 ghz microwave frequency via the internet of things (iot) network. *Wireless Personal Communications*, 124(3):1961–1976, 2022.

[313] Bin Fang, Fuchun Sun, Zhou Quan, Huaping Liu, and Jianhua Shan. [retracted] smart bracelet system for temperature monitoring and movement tracking analysis. *Journal of Healthcare Engineering*, 2021(1):8347261, 2021.

[314] Tiago M Fernández-Caramés, Iván Froiz-Míguez, Oscar Blanco-Novoa, and Paula Fraga-Lamas. Enabling the internet of mobile crowdsourcing health things: A mobile fog computing, blockchain and iot based continuous glucose monitoring system for diabetes mellitus research and care. *Sensors*, 19(15):3319, 2019.

[315] Katarzyna Dziergowska, Magdalena Beata Łabowska, Marlena Gąsior-Głogowska, Barbara Kmiecik, and Jerzy Detyna. Modern noninvasive methods for monitoring glucose levels in patients: A review. *Bio-Algorithms and Med-Systems*, 15(4):20190052, 2019.

[316] Tuan Nguyen Gia, Mai Ali, Imed Ben Dhaou, Amir M Rahmani, Tomi Westerlund, Pasi Liljeberg, and Hannu Tenhunen. Iot-based continuous glucose monitoring system: A feasibility study. *Procedia Computer Science*, 109:327–334, 2017.

[317] MA Aizat Rahmat, ELM Su, M Mohd Addi, and CF Yeong. Gluqo: Iot-based non-invasive blood glucose monitoring. *Journal of Telecommunication, Electronic and Computer Engineering*, 9(3-9):71–75, 2017.

[318] Antonio Alarcón-Paredes, Victor Francisco-García, Iris P Guzmán-Guzmán, Jessica Cantillo-Negrete, René E Cuevas-Valencia, and Gustavo A Alonso-Silverio. An iot-based non-invasive glucose level monitoring system using raspberry pi. *Applied Sciences*, 9(15):3046, 2019.

[319] Francisco Valenzuela, Armando García, Erica Ruiz, Mabel Vazquez, Joaquin Cortez, and Adolfo Espinoza. An iot-based glucose monitoring algorithm to prevent diabetes complications. *Applied Sciences*, 10(3):921, 2020.

[320] Siti Nur Shahidah Anis and Rozlan Alias. A portable non-invasive blood glucose monitoring device with iot. *Evolution in Electrical and Electronic Engineering*, 2(1):36–44, 2021.

[321] Wentao Zhong, Zhong Ji, and Changlong Sun. A review of monitoring methods for cerebral blood oxygen saturation. *Healthcare*, 9(9):1104, 2021.

[322] Samik Basu, Sinjoy Saha, Soumya Pandit, and Soma Barman. Smart health monitoring system for temperature, blood oxygen saturation, and heart rate sensing with embedded processing and transmission using iot platform. In *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*, pages 81–91. Springer, 2020.

[323] Lai Phuoc Son, Nguyen Thi Anh Thu, and Nguyen Trung Kien. Design an iot wrist-device for spo2 measurement. In *2017 international conference on advanced technologies for communications (ATC)*, pages 144–149. IEEE, 2017.

[324] RR Adiputra, Sugondo Hadiyoso, and Y Sun Hariyani. Internet of things: Low cost and wearable spo2 device for health monitoring. *International Journal of Electrical and Computer Engineering*, 8(2):939, 2018.

[325] Bedjo Utomo, Sari Luthfiyah, Urip Mudjiono, Torib Hamzah, et al. The real-time vital sign monitor for heart rate and spo2 parameter using internet of things technology. *Journal of Physics: Conference Series*, 1373(1):012028, 2019.

[326] Qingguo Chen and Liqin Tang. A wearable blood oxygen saturation monitoring system based on bluetooth low energy technology. *Computer Communications*, 160:101–110, 2020.

[327] Qingxue Zhang, David Arney, Julian M Goldman, Eric M Isselbacher, and Antonis A Armoundas. Design implementation and evaluation of a mobile continuous blood oxygen saturation monitoring system. *Sensors*, 20(22):6581, 2020.

[328] Alaa A Qaffas, Rakibul Hoque, and Najah Almazmomi. The internet of things and big data analytics for chronic disease monitoring in saudi arabia. *Telemedicine and e-Health*, 27(1):74–81, 2021.

[329] Tlija Amira, Istrate Dan, Bennani Az-eddine, Hoai Huong Ngo, Gattoufi Said, and Wegrzyn-Wolska Katarzyna. Monitoring chronic disease at home using connected devices. In *2018 13th Annual Conference on System of Systems Engineering (SoSE)*, pages 400–407. IEEE, 2018.

[330] Syed Tauhid Ullah Shah, Faizan Badshah, Faheem Dad, Nouman Amin, and Mian Ahmad Jan. Cloud-assisted iot-based smart respiratory monitoring system for asthma patients. In *Applications of intelligent technologies in healthcare*, pages 77–86. Springer, 2019.

[331] Osama Rehman, Zaroon Farrukh, Asiya Al-Busaidi, Kyungjin Cha, Simon Park, and Ibrahim Rahman. Iot powered cancer observation system. In *The 9th International Conference on Smart Media and Applications*, pages 313–318, 2020.

[332] Ashif Newaz Shihab, Miftahul Jannat Mokarrama, Rezaul Karim, Sumi Khatun, and Mohammad Shamsul Arefin. An iot-based heart disease detection system using rnn. In *Image Processing and Capsule Networks: ICIPCN 2020*, pages 535–545. Springer, 2021.

[333] Madhuri Kerappa Gawali and C Rambabu. Iot model for heart disease detection using machine learning (ml) techniques. In *Techno-Societal 2020: Proceedings of the 3rd International Conference on Advanced Technologies for Societal Applications—Volume 1*, pages 399–409. Springer, 2021.

[334] Aman Hebbale, GHR Vinay, BVN Vamsi Krishna, and Jalpa Shah. Iot and machine learning based self care system for diabetes monitoring and prediction. In *2021 2nd global conference for advancement in technology (GCAT)*, pages 1–7. IEEE, 2021.

[335] PR Anisha and C Kishor Kumar Reddy. Early detection of diabetes using machine learning algorithms and internet of things: Adpa. In *Communication Software and Networks: Proceedings of INDIA 2019*, pages 79–85. Springer, 2021.

[336] Anoop Kumar Prasad. Smart asthma alert using iot and predicting threshold values using decision tree classifier. In *Applications of Internet of Things: Proceedings of ICCCIOT 2020*, pages 141–150. Springer, 2020.

[337] Adeniyi Onasanya and Maher Elshakankiri. Smart integrated iot healthcare system for cancer care. *Wireless Networks*, 27(6):4297–4312, 2021.

[338] Jinbo Huang, Xianjun Wu, Wendong Huang, Xiaoli Wu, and S Wang. Internet of things in health management systems: A review. *International Journal of Communication Systems*, 34(4):e4683, 2021.

[339] Kei Hong So, Cheuk Wun Ting, Chui Ping Lee, Teddy Tai-Ning Lam, Sau Chu Chiang, and Yin Ting Cheung. Medication management service for old age homes in hong kong using information technology, automation technology, and the internet of things: Pre-post interventional study. *JMIR Medical Informatics*, 9(2):e24280, 2021.

[340] S Aakash Bharadwaj, Divyank Yarravarapu, Sadiparala Charan Kumar Reddy, Thirumalaraju Prudhvi, KSP Sandeep, and Obulam Siva Dheeraj Reddy. Enhancing healthcare using m-care box (monitoring non-compliance of medication). In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pages 352–356. IEEE, 2017.

[341] Hsiu-Ling Tsai, Chun Hsiang Tseng, Long-Cian Wang, and Fuh-Shyang Juang. Bidirectional smart pill box monitored through internet and receiving reminding message from remote relatives. In *2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pages 393–394. IEEE, 2017.

[342] M Srinivas, P Durgaprasadarao, and V Naga Prudhvi Raj. Intelligent medicine box for medication management using iot. In *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, pages 32–34. IEEE, 2018.

[343] Ghazanfar Latif, Achyut Shankar, Jaafar M Alghazo, V Kalyanasundaram, CS Boopathi, and M Arfan Jaffar. I-cares: advancing health diagnosis and medication through iot. *Wireless Networks*, 26(4):2375–2389, 2020.

[344] PA Harsha Vardhini, M Shiva Harsha, P Naga Sai, and P Srikanth. Iot based smart medicine assistive system for memory impairment patient. In *2020 12th international conference on computational intelligence and communication networks (CICN)*, pages 182–186. IEEE, 2020.

[345] B Ayshwarya and R Velmurugan. Intelligent and safe medication box in health iot platform for medication monitoring system with timely remainders. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 1828–1831. IEEE, 2021.

[346] Shah Muhammad Azmat Ullah, Md Milon Islam, Saifuddin Mahmud, Sheikh Nooruddin, SM Taslim Uddin Raju, and Md Rezwanul Haque. Scalable telehealth services to combat novel coronavirus (covid-19) pandemic. *SN Computer Science*, 2:1–8, 2021.

[347] Christian M Graham and Nory Jones. Impact of iot on geriatric telehealth. *Working with Older People*, 24(3):231–243, 2020.

[348] Divya Ganesh, Gayathri Seshadri, Sumathi Sokkanarayanan, Panjavarnam Bose, Sharanya Rajan, and Mithileysh Sathiyanarayanan. Autoimpilo: smart automated health machine using iot to improve telemedicine and telehealth. In *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, pages 487–493. IEEE, 2020.

[349] Kostas M Tsiouris, Dimitrios Gatsios, Vassilios Tsakanikas, Athanasios A Pardalis, Ioannis Kouris, Thelma Androutsou, Marilena Tarousi, Natasa Vujnovic Sedlar, Iason Somarakis, Fariba Mostajeran, et al. Designing interoperable telehealth platforms: bridging iot devices with cloud infrastructures. *Enterprise Information Systems*, 14(8):1194–1218, 2020.

[350] Tharaka Hewa, An Braeken, Mika Ylianttila, and Madhusanka Liyanage. Multi-access edge computing and blockchain-based secure telehealth system connected with 5g and iot. In *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pages 1–6. IEEE, 2020.

[351] Md Rokonuzzaman, Md Imran Hossain, Tariqul Islam, Partha Pratim Sarkar, Md Robiul Islam, and Nowshad Amin. Design and implementation of telehealth device: Linking iot

sensors to cloud networks. In *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 281–285. IEEE, 2021.

[352] Hocine Hamil, Zahia Zidelmal, Mohamed Salah Azzaz, Samir Sakhi, Redouane Kaibou, Salem Djilali, and Djaffar Ould Abdeslam. Design of a secured telehealth system based on multiple biosignals diagnosis and classification for iot application. *Expert Systems*, 39(4):e12765, 2022.

[353] Rajesh Gupta, Sudeep Tanwar, Sudhanshu Tyagi, and Neeraj Kumar. Tactile-internet-based telesurgery system for healthcare 4.0: An architecture, research challenges, and future directions. *IEEE Network*, 33(6):22–29, 2019.

[354] Francesk Mulita, Georgios-Ioannis Verras, Christos-Nikolaos Anagnostopoulos, and Konstantinos Kotis. A smarter health through the internet of surgical things. *Sensors*, 22(12):4577, 2022.

[355] Patrick Barba, Joshua Stramiello, Emily K Funk, Florian Richter, Michael C Yip, and Ryan K Orosco. Remote telesurgery in humans: a systematic review. *Surgical Endoscopy*, 36(5):2771–2777, 2022.

[356] N Shabana and G Velmathi. Advanced tele-surgery with iot approach. In *Intelligent Embedded Systems: Select Proceedings of ICNETS2, Volume II*, pages 17–24. Springer, 2018.

[357] Yiming Miao, Yingying Jiang, Limei Peng, M Shamim Hossain, and Ghulam Muhammad. Telesurgery robot based on 5g tactile internet. *Mobile Networks and Applications*, 23:1645–1654, 2018.

[358] Pedram Fekri, Peyman Setoodeh, Fariba Khosravian, AA Safavi, and Mehrdad H Zadeh. Towards deep secure tele-surgery. In *Proceedings of the international conference on scientific computing (CSC)*, pages 81–86. The Steering Committee of TheWorld Congress in Computer Science, 2018.

[359] Rajesh Gupta, Sudeep Tanwar, Sudhanshu Tyagi, Neeraj Kumar, Mohammad S Obaidat, and Balqies Sadoun. Habits: Blockchain-based telesurgery framework for healthcare 4.0. In *2019 international conference on computer, information and telecommunication systems (CITS)*, pages 1–5. IEEE, 2019.

[360] Sohail Iqbal, Shahzad Farooq, Khuram Shahzad, Asad Waqar Malik, Mian M Hamayun, and Osman Hasan. Securesurginet: A framework for ensuring security in telesurgery. *International Journal of Distributed Sensor Networks*, 15(9):1550147719873811, 2019.

[361] Ghulam Muhammad and M Shamim Hossain. Emotion recognition for cognitive edge computing using deep learning. *IEEE Internet of Things Journal*, 8(23):16894–16901, 2021.

[362] José Paulo Lousado and Sandra Antunes. Monitoring and support for elderly people using lora communication technologies: Iot concepts and applications. *Future Internet*, 12(11):206, 2020.

[363] Ahmed Abdelgawad, Kumar Yelamarthi, and Ahmed Khattab. Iot-based health monitoring system for active and assisted living. In *Smart Objects and Technologies for Social Good: Second International Conference, GOODTECHS 2016, Venice, Italy, November 30–December 1, 2016, Proceedings 2*, pages 11–20. Springer, 2017.

[364] Olivier Debauche, Saïd Mahmoudi, Pierre Manneback, and Abdessamad Assila. Fog iot for health: A new architecture for patients and elderly monitoring. *Procedia Computer Science*, 160:289–297, 2019.

[365] J Boobalan and M Malleswaran. A novel and customizable framework for iot based smart home nursing for elderly care. In *Emerging Trends in Computing and Expert Technology*, pages 27–38. Springer, 2020.

[366] Aparajith Srinivasan, Nithya Natarajan, Raj Vignesh Karunakaran, Ramya Elangovan, Abirami Shankar, Padmanaaban M Sabharish, BS Sreeja, and S Radha. Elder care system using iot and machine learning in aws cloud. In *2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)*, pages 92–98. IEEE, 2020.

[367] Rajdeep Kumar Nath and Himanshu Thapliyal. Wearable health monitoring system for older adults in a smart home environment. In *2021 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pages 390–395. IEEE, 2021.

[368] Gregory MP O'Hare, Dominic Carr, and Michael J O'Grady. Collaborative sensing: Delivering an intelligent and adaptive sensing infrastructure. In *2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 13–17. IEEE, 2015.

[369] Shibo He, Kun Shi, Chen Liu, Bicheng Guo, Jiming Chen, and Zhiguo Shi. Collaborative sensing in internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 24(3):1435–1474, 2022.

[370] Lingzhi Yi, Xianjun Deng, Laurence T Yang, Hengshan Wu, Minghua Wang, and Yi Situ. Reinforcement-learning-enabled partial confident information coverage for iot-based bridge structural health monitoring. *IEEE Internet of Things Journal*, 8(5):3108–3119, 2020.

[371] Fan Yang, Lei Shu, Yuli Yang, Guangjie Han, Simon Pearson, and Kailiang Li. Optimal deployment of solar insecticidal lamps over constrained locations in mixed-crop farmlands. *IEEE Internet of Things Journal*, 8(16):13095–13114, 2021.

[372] Kun Shi, Zhiguo Shi, Chaoqun Yang, Shibo He, Jiming Chen, and Anjun Chen. Road-map aided gm-phd filter for multivehicle tracking with automotive radar. *IEEE Transactions on Industrial Informatics*, 18(1):97–108, 2021.

[373] Bing Dong, Vishnu Prakash, Fan Feng, and Zheng O'Neill. A review of smart building sensing system for better indoor environment control. *Energy and Buildings*, 199:29–46, 2019.

[374] Imad Belkacem, Aymen Gammoudi, Nasredine Cheniki, Yacine Sam, and Nizar Messai. Collaborative internet of things (c-iot) data analysis for enhancing activity recognition and preventing serious health problems. In *2021 IEEE 30th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pages 149–152. IEEE, 2021.

[375] Abdul Jaleel, Tayyeb Mahmood, Muhammad Awais Hassan, Gulshan Bano, and Syed Khaldoon Khurshid. Towards medical data interoperability through collaboration of healthcare devices. *IEEE Access*, 8:132302–132319, 2020.

[376] Minh Tuan Nguyen. Distributed compressive and collaborative sensing data collection in mobile sensor networks. *Internet of Things*, 9:100156, 2020.

[377] Youcef Djenouri, Asma Belhadi, Anis Yazidi, Gautam Srivastava, Pushpita Chatterjee, and Jerry Chun-Wei Lin. An intelligent collaborative image-sensing system for disease detection. *IEEE Sensors Journal*, 23(2):947–954, 2022.

[378] Mehedi Masud, Gurjot Singh Gaba, Salman Alqahtani, Ghulam Muhammad, Brij B Gupta, Pardeep Kumar, and Ahmed Ghoneim. A lightweight and robust secure key establishment protocol for internet of medical things in covid-19 patients care. *IEEE Internet of Things Journal*, 8(21):15694–15703, 2020.

[379] Mehedi Masud, Gurjot Singh Gaba, Karanjeet Choudhary, M Shamim Hossain, Mohammed F Alhamid, and Ghulam Muhammad. Lightweight and anonymity-preserving

user authentication scheme for iot-based healthcare. *IEEE Internet of Things Journal*, 9(4):2649–2656, 2021.

[380] Bhabendu Kumar Mohanta, Debasish Jena, Somula Ramasubbareddy, Mahmoud Daneshmand, and Amir H Gandomi. Addressing security and privacy issues of iot using blockchain technology. *IEEE Internet of Things Journal*, 8(2):881–888, 2020.

[381] Zhengping Ruan. Blockchain technology for security issues and challenges in iot. In *2023 International Conference on Computer Simulation and Modeling, Information Security (CSMIS)*, pages 572–580. IEEE, 2023.

[382] Ramon Sanchez-Iborra and Antonio F Skarmeta. Tinyml-enabled frugal smart objects: Challenges and opportunities. *IEEE Circuits and Systems Magazine*, 20(3):4–18, 2020.

[383] Lachit Dutta and Swapna Bharali. Tinyml meets iot: A comprehensive survey. *Internet of Things*, 16:100461, 2021.

[384] Vangelis Gazis. A survey of standards for machine-to-machine and the internet of things. *IEEE Communications Surveys & Tutorials*, 19(1):482–511, 2016.

[385] Shuo Wan, Jiaxun Lu, Pingyi Fan, and Khaled B Letaief. Toward big data processing in iot: Path planning and resource management of uav base stations in mobile-edge computing system. *IEEE Internet of Things Journal*, 7(7):5995–6009, 2019.

[386] Dante D Sánchez-Gallegos, Diana Carrizales-Espinoza, Hugo G Reyes-Anastacio, JL Gonzalez-Compean, Jesus Carretero, Miguel Morales-Sandoval, and Alejandro Galaviz-Mosqueda. From the edge to the cloud: A continuous delivery and preparation model for processing big iot data. *Simulation Modelling Practice and Theory*, 105:102136, 2020.

[387] Gabriel Signoretti, Marianne Silva, Pedro Andrade, Ivanovitch Silva, Emiliano Sisinni, and Paolo Ferrari. An evolving tinyml compression algorithm for iot environments based on data eccentricity. *Sensors*, 21(12):4153, 2021.

[388] Keith Nolan and Mark Kelly. Ipv6 convergence for iot cyber–physical systems. *Information*, 9(4):70, 2018.

[389] Teemu Savolainen, Jonne Soininen, and Bilhanan Silverajan. Ipv6 addressing strategies for iot. *IEEE Sensors Journal*, 13(10):3511–3519, 2013.

[390] Ahmed E Khaled, Abdelsalam Helal, Wyatt Lindquist, and Choonhwa Lee. Iot-ddl–device description language for the "t" in iot. *IEEE Access*, 6:24048–24063, 2018.

[391] Miguel Zúñiga-Prieto, Daniel Rodríguez, Juan Rodríguez, Lizandro Solano, Emilio Insfran, and Silvia Abrahao. Iot-adl: An adl for describing cloud iot applications. *Journal of Computers*, 29(6):264–273, 2018.

[392] Anju S Pillai, Gaddam S Chandraprasad, Ahmed S Khwaja, and Alagan Anpalagan. A service oriented iot architecture for disaster preparedness and forecasting system. *Internet of Things*, 14:100076, 2021.

[393] Sudhanshu Maurya and Kuntal Mukherjee. An energy efficient architecture of iot based on service oriented architecture (soa). *Informatica*, 43(1), 2019.

[394] Kyriaki Panagidi, Christos Anagnostopoulos, A Chalvatzaras, and Stathes Hadjiefthymiades. To transmit or not to transmit: Controlling communications in the mobile iot domain. *ACM Transactions on Internet Technology*, 20(3):1–23, 2020.

[395] Xiong Xiong, Kan Zheng, Lei Lei, and Lu Hou. Resource allocation based on deep reinforcement learning in iot edge computing. *IEEE Journal on Selected Areas in Communications*, 38(6):1133–1146, 2020.

[396] Xin Wei, Jialin Zhao, Liang Zhou, and Yi Qian. Broad reinforcement learning for supporting fast autonomous iot. *IEEE Internet of Things Journal*, 7(8):7010–7020, 2020.

[397] Shahirah Mohamed Hatim, S Jamel Elias, Norkhushaini Awang, Mohd Yusof Darus, et al. Vanets and internet of things (iot): A discussion. *Indonesian Journal of Electrical Engineering and Computer Science*, 12(1):218–224, 2018.

[398] Kamal Azghiou, Manal El Mouhib, Mohammed-Amine Koulali, and Abdelhamid Benali. An end-to-end reliability framework of the internet of things. *Sensors*, 20(9):2439, 2020.

[399] Liudong Xing. Reliability in internet of things: Current status and future perspectives. *IEEE Internet of Things Journal*, 7(8):6704–6721, 2020.

[400] Oscar Novo. Blockchain meets iot: An architecture for scalable access management in iot. *IEEE Internet of Things Journal*, 5(2):1184–1195, 2018.

[401] Asad Javed, Avleen Malhi, Tuomas Kinnunen, and Kary Främling. Scalable iot platform for heterogeneous devices in smart environments. *IEEE Access*, 8:211973–211985, 2020.

[402] Mahda Noura, Mohammed Atiquzzaman, and Martin Gaedke. Interoperability in internet of things: Taxonomies and open challenges. *Mobile Networks and Applications*, 24:796–809, 2019.

[403] Partha Pratim Ray, Nishant Thapa, and Dinesh Dash. Implementation and performance analysis of interoperable and heterogeneous iot-edge gateway for pervasive wellness care. *IEEE Transactions on Consumer Electronics*, 65(4):464–473, 2019.

[404] Inés Sittón-Candanedo, Ricardo S Alonso, Óscar García, Lilia Muñoz, and Sara Rodríguez-González. Edge computing, iot and social computing in smart energy scenarios. *Sensors*, 19(15):3353, 2019.

[405] Neha Gupta, Suneet K Gupta, Rajesh K Pathak, Vanita Jain, Parisa Rashidi, and Jasjit S Suri. Human activity recognition in artificial intelligence framework: a narrative review. *Artificial Intelligence Review*, pages 1–54, 2022.

[406] Santosh Kumar Yadav, Kamlesh Tiwari, Hari Mohan Pandey, and Shaik Ali Akbar. A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. *Knowledge-Based Systems*, 223:106970, 2021.

[407] Rex Liu, Albara Ah Ramli, Huanle Zhang, Erik Henricson, and Xin Liu. An overview of human activity recognition using wearable sensors: Healthcare and artificial intelligence. In *International Conference on Internet of Things*, pages 1–14. Springer, 2021.

[408] Fatemeh Serpush, Mohammad Bagher Menhaj, Behrooz Masoumi, and Babak Karasfi. Wearable sensor-based human activity recognition in the smart healthcare system. *Computational Intelligence and Neuroscience*, 2022, 2022.

[409] Shibo Zhang, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng, and Nabil Alshurafa. Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, 22(4):1476, 2022.

[410] Binh Nguyen, Yves Coelho, Teodiano Bastos, and Sridhar Krishnan. Trends in human activity recognition with focus on machine learning and power requirements. *Machine Learning with Applications*, 5:100072, 2021.

[411] E Ramanujam, Thinagaran Perumal, and S Padmavathi. Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sensors Journal*, 21(12):13029–13040, 2021.

[412] Pranjal Kumar, Siddhartha Chauhan, and Lalit Kumar Awasthi. Human activity recognition (har) using deep learning: Review, methodologies, progress and future research directions. *Archives of Computational Methods in Engineering*, 31(1):179–219, 2024.

[413] Sakorn Mekruksavanich and Anuchit Jitpattanakul. Lstm networks using smartphone data for sensor-based human activity recognition in smart homes. *Sensors*, 21(5):1636, 2021.

[414] Zhixuan Wu, Nan Ma, Yongqiang Yao, and Guoping Zhang. Survey of action recognition based on attention mechanism. In *Proceedings of 2021 Chinese Intelligent Automation Conference*, pages 555–563. Springer, 2022.

[415] Carlos Betancourt, Wen-Hui Chen, and Chi-Wei Kuan. Self-attention networks for human activity recognition using wearable devices. In *2020 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 1194–1199. IEEE, 2020.

[416] Farhad Nazari, Darius Nahavandi, Navid Mohajer, and Abbas Khosravi. Human activity recognition from knee angle using machine learning techniques. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 295–300. IEEE, 2021.

[417] Tan-Hsu Tan, Ching-Jung Huang, Munkhjargal Gochoo, and Yung-Fu Chen. Activity recognition based on fr-cnn and attention-based lstm network. In *2021 30th Wireless and Optical Communications Conference (WOCC)*, pages 146–149. IEEE, 2021.

[418] Sakorn Mekruksavanich, Ponnipa Jantawong, and Anuchit Jitpattanakul. A lightweight deep convolutional neural network with squeeze-and-excitation modules for efficient human activity recognition using smartphone sensors. In *2021 2nd International Conference on Big Data Analytics and Practices (IBDAP)*, pages 23–27. IEEE, 2021.

[419] Satya P Singh, Madan Kumar Sharma, Aimé Lay-Ekuakille, Deepak Gangwar, and Sukrit Gupta. Deep convlstm with self-attention for human activity decoding using wearable sensors. *IEEE Sensors Journal*, 21(6):8575–8582, 2020.

[420] Shaohua Wan, Lianyong Qi, Xiaolong Xu, Chao Tong, and Zonghua Gu. Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications*, 25(2):743–755, 2020.

[421] Nidhi Dua, Shiva Nand Singh, and Vijay Bhaskar Semwal. Multi-input cnn-gru based human activity recognition using wearable sensors. *Computing*, 103(7):1461–1478, 2021.

[422] Sravan Kumar Challa, Akhilesh Kumar, and Vijay Bhaskar Semwal. A multibranch cnn-bilstm model for human activity recognition using wearable sensor data. *The Visual Computer*, pages 1–15, 2021.

[423] Xiaokang Zhou, Wei Liang, I Kevin, Kai Wang, Hao Wang, Laurence T Yang, and Qun Jin. Deep-learning-enhanced human activity recognition for internet of healthcare things. *IEEE Internet of Things Journal*, 7(7):6429–6438, 2020.

[424] Ricardo Espinosa, Hiram Ponce, Sebastián Gutiérrez, Lourdes Martínez-Villaseñor, Jorge Brieva, and Ernesto Moya-Albor. A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the up-fall detection dataset. *Computers in Biology and Medicine*, 115:103520, 2019.

[425] Thomas Wiatowski and Helmut Bölcskei. A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Transactions on Information Theory*, 64(3):1845–1866, 2017.

[426] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[427] Jingyu Zhao, Feiqing Huang, Jia Lv, Yanjie Duan, Zhen Qin, Guodong Li, and Guangjian Tian. Do rnn and lstm have long memory? In *International Conference on Machine Learning*, pages 11365–11375. PMLR, 2020.

[428] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wangchun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 2015.

[429] Hengshuang Zhao, Jiaya Jia, and Vladlen Koltun. Exploring self-attention for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10076–10085, 2020.

[430] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[431] Lourdes Martínez-Villaseñor, Hiram Ponce, Jorge Brieva, Ernesto Moya-Albor, José Núñez-Martínez, and Carlos Peñafort-Asturiano. Up-fall detection dataset: A multimodal approach. *Sensors*, 19(9):1988, 2019.

[432] Lourdes Martínez-Villaseñor, Hiram Ponce, and Ricardo Abel Espinosa-Loera. Multimodal database for human activity recognition and fall detection. *Proceedings*, 2(19):1237, 2018.

[433] Lourdes Martínez-Villaseñor, Hiram Ponce, and Karina Perez-Daniel. Deep learning for multimodal fall detection. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 3422–3429. IEEE, 2019.

[434] Santosh Kumar Yadav, Achleshwar Luthra, Kamlesh Tiwari, Hari Mohan Pandey, and Shaik Ali Akbar. Arfdnet: An efficient activity recognition & fall detection system using latent feature pooling. *Knowledge-Based Systems*, 239:107948, 2022.

[435] Yilin Yang, Haocong Wang, Ruizhe Jiang, Xiaonan Guo, Jerry Cheng, and Yingying Chen. A review of iot-enabled mobile healthcare: Technologies, challenges, and future trends. *IEEE Internet of Things Journal*, 9(12):9478–9502, 2022.

[436] Henry Friday Nweke, Ying Wah Teh, Ghulam Mujtaba, and Mohammed Ali Al-Garadi. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion*, 46:147–170, 2019.

[437] Martin Gjoreski, Vito Janko, Gašper Slapničar, Miha Mlakar, Nina Reščič, Jani Bizjak, Vid Drobnič, Matej Marinko, Nejc Mlakar, Mitja Luštrek, et al. Classical and deep learning methods for recognizing human activities and modes of transportation with smartphone sensors. *Information Fusion*, 62:47–62, 2020.

[438] Wenbo Zheng, Lan Yan, Chao Gou, and Fei-Yue Wang. Meta-learning meets the internet of things: Graph prototypical models for sensor-based human activity recognition. *Information Fusion*, 80:1–22, 2022.

[439] Jun Qi, Po Yang, Lee Newcombe, Xiyang Peng, Yun Yang, and Zhong Zhao. An overview of data fusion techniques for internet of things enabled physical activity recognition and measure. *Information Fusion*, 55:269–280, 2020.

[440] Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit LeDuc, and Ioannis Kanellos. A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning. *Sensors*, 21(18):6037, 2021.

[441] Dong Chen, Sira Yongchareon, Edmund MK Lai, Quan Z Sheng, and Veronica Liesaputra. Locally-weighted ensemble detection-based adaptive random forest classifier for sensor-based online activity recognition for multiple residents. *IEEE Internet of Things Journal*, 9(15):13077–13085, 2022.

[442] Sen Qiu, Hongkai Zhao, Nan Jiang, Zhelong Wang, Long Liu, Yi An, Hongyu Zhao, Xin Miao, Ruichen Liu, and Giancarlo Fortino. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, 80:241–265, 2022.

[443] Zhen Qin, Yibo Zhang, Shuyu Meng, Zhiguang Qin, and Kim-Kwang Raymond Choo. Imaging and fusing time series for wearable sensor-based human activity recognition. *Information Fusion*, 53:80–87, 2020.

[444] Marcin Straczkiewicz, Peter James, and Jukka-Pekka Onnela. A systematic review of smartphone-based human activity recognition methods for health research. *npj Digital Medicine*, 4(1):1–15, 2021.

[445] Preksha Pareek and Ankit Thakkar. A survey on video-based human action recognition: recent updates, datasets, challenges, and applications. *Artificial Intelligence Review*, 54(3):2259–2322, 2021.

[446] Ying He, Yan Chen, Yang Hu, and Bing Zeng. Wifi vision: Sensing, recognition, and detection with commodity mimo-ofdm wifi. *IEEE Internet of Things Journal*, 7(9):8296–8317, 2020.

[447] Jingjing Cao, Wenfeng Li, Congcong Ma, and Zhiwen Tao. Optimizing multi-sensor deployment via ensemble pruning for wearable activity recognition. *Information Fusion*, 41:68–79, 2018.

[448] Gábor Csizmadia, Krisztina Liszkai-Peres, Bence Ferdinandy, Ádám Miklósi, and Veronika Konok. Human activity recognition of children with wearable devices using lightgbm machine learning. *Scientific Reports*, 12(1):1–10, 2022.

[449] Zehua Sun, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, Gang Wang, and Jun Liu. Human action recognition from various data modalities: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[450] Yirui Wu, Hao Cao, Guoqiang Yang, Tong Lu, and Shaohua Wan. Digital twin of intelligent small surface defect detection with cyber-manufacturing systems. *ACM Transactions on Internet Technology*, pages 1–20, 2022.

[451] Lorin Werthen-Brabants, Geethika Bhavanasi, Ivo Couckuyt, Tom Dhaene, and Dirk Deschrijver. Split birnn for real-time activity recognition using radar and deep learning. *Scientific Reports*, 12(1):1–11, 2022.

[452] Xiaochun Yin, Zengguang Liu, Deyong Liu, and Xiaojun Ren. A novel cnn-based bi-lstm parallel model with attention mechanism for human activity recognition with noisy data. *Scientific Reports*, 12(1):1–11, 2022.

[453] Chao-Lung Yang, Zhi-Xuan Chen, and Chen-Yi Yang. Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images. *Sensors*, 20(1):168, 2019.

[454] Alana de Santana Correia and Esther Luna Colombini. Attention, please! a survey of neural attention models in deep learning. *Artificial Intelligence Review*, pages 1–88, 2022.

[455] Enrique Garcia-Ceja, Carlos E Galván-Tejada, and Ramon Brena. Multi-view stacking for activity recognition with sound and accelerometer data. *Information Fusion*, 40:45–56, 2018.

[456] Qimeng Li, Raffaele Gravina, Ye Li, Saeed H Alsamhi, Fangmin Sun, and Giancarlo Fortino. Multi-user activity recognition: Challenges and opportunities. *Information Fusion*, 63:121–135, 2020.

[457] Mohammed A. A. Al-qaness, Abdelghani Dahou, Mohamed Abd Elaziz, and A. M. Helmi. Multi-resatt: Multilevel residual network with attention for human activity recognition using wearable sensors. *IEEE Transactions on Industrial Informatics*, 19(1):144–152, 2023.

[458] Jianchao Lu, Xi Zheng, Michael Sheng, Jiong Jin, and Shui Yu. Efficient human activity recognition using a single wearable sensor. *IEEE Internet of Things Journal*, 7(11):11137–11146, 2020.

[459] Mohamed Abdel-Basset, Hossam Hawash, Victor Chang, Ripon K Chakrabortty, and Michael Ryan. Deep learning for heterogeneous human activity recognition in complex iot applications. *IEEE Internet of Things Journal*, 9(8):5653–5665, 2022.

[460] Mohamed Abdel-Basset, Hossam Hawash, Ripon K Chakrabortty, Michael Ryan, Mohamed Elhoseny, and Houbing Song. St-deephar: Deep learning model for human activity recognition in ioht applications. *IEEE Internet of Things Journal*, 8(6):4969–4979, 2020.

[461] Haoxi Zhang, Zhiwen Xiao, Juan Wang, Fei Li, and Edward Szczerbicki. A novel iot-perceptive human activity recognition (har) approach using multihead convolutional attention. *IEEE Internet of Things Journal*, 7(2):1072–1080, 2019.

[462] Heilym Ramirez, Sergio A Velastin, Ignacio Meza, Ernesto Fabregas, Dimitrios Makris, and Gonzalo Farias. Fall detection and activity recognition using human skeleton features. *IEEE Access*, 9:33532–33542, 2021.

[463] Anitha Rani Inturi, VM Manikandan, and Vignesh Garrapally. A novel vision-based fall detection scheme using keypoints of human skeleton with long short-term memory network. *Arabian Journal for Science and Engineering*, pages 1–13, 2022.

[464] Fang Lin, Zhelong Wang, Hong yu Zhao, Sen Qiu, Xin Shi, Lina Wu, Giancarlo Fortino, and Raffaele Gravina. Adaptive multimodal fusion framework for activity monitoring of people with mobility disability. *IEEE Journal of Biomedical and Health Informatics*, 26(8):4314–4324, 2022.

[465] Caetano M Ranieri, Patricia A Vargas, and Roseli AF Romero. Uncovering human multimodal activity recognition with a deep learning approach. In *2020 International joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2020.

[466] Caetano Mazzoni Ranieri, Scott MacLeod, Mauro Dragone, Patricia Amancio Vargas, and Roseli Aparecida Francelin Romero. Activity recognition for ambient assisted living with videos, inertial units and ambient sensors. *Sensors*, 21(3):768, 2021.

[467] Zan Gao, Hai-Zhen Xuan, Hua Zhang, Shaohua Wan, and Kim-Kwang Raymond Choo. Adaptive fusion and category-level dictionary learning model for multiview human action recognition. *IEEE Internet of Things Journal*, 6(6):9280–9293, 2019.

[468] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018.

[469] Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9):2352–2449, 2017.

[470] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[471] Crystal L Park, Laura D Kubzansky, Sandra M Chafouleas, Richard J Davidson, Dacher Keltner, Parisa Parsafar, Yeates Conwell, Michelle Y Martin, Janel Hanmer, and Kuan Hong Wang. Emotional well-being: What it is and why it matters. *Affective Science*, 4(1):10–20, 2023.

[472] M Maithri, U Raghavendra, Anjan Gudigar, Jyothi Samanth, Prabal Datta Barua, Murugappan Murugappan, Yashas Chakole, and U Rajendra Acharya. Automated emotion recognition: Current trends and future perspectives. *Computer Methods and Programs in Biomedicine*, 215:106646, 2022.

[473] M Kalpana Chowdary, Tu N Nguyen, and D Jude Hemanth. Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Computing and Applications*, 35:23311–23328, 2023.

[474] Smith K Khare, Victoria Blanes-Vidal, Esmaeil S Nadimi, and U Rajendra Acharya. Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. *Information Fusion*, 102:102019, 2023.

[475] Zijian Zhou, Muhammad Adeel Asghar, Daniyal Nazir, Kamran Siddique, Mohammad Shorfuzzaman, and Raja Majid Mehmood. An ai-empowered affect recognition model for healthcare and emotional well-being using physiological signals. *Cluster Computing*, 26(2):1253–1266, 2023.

[476] Yan Wang, Wei Song, Wei Tao, Antonio Liotta, Dawei Yang, Xinlei Li, Shuyong Gao, Yixuan Sun, Weifeng Ge, Wei Zhang, et al. A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, 83:19–52, 2022.

[477] Felipe Zago Canal, Tobias Rossi Müller, Jhennifer Cristine Matias, Gustavo Gino Scotton, Antonio Reis de Sa Junior, Eliane Pozzebon, and Antonio Carlos Sobieranski. A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*, 582:593–617, 2022.

[478] Ying Tan, Zhe Sun, Feng Duan, Jordi Solé-Casals, and Cesar F Caiafa. A multimodal emotion recognition method based on facial expressions and electroencephalography. *Biomedical Signal Processing and Control*, 70:103029, 2021.

[479] Qi Li, Yunqing Liu, Fei Yan, Qiong Zhang, and Cong Liu. Emotion recognition based on multiple physiological signals. *Biomedical Signal Processing and Control*, 85:104989, 2023.

[480] Yingying Jiang, Wei Li, M Shamim Hossain, Min Chen, Abdulhameed Alelaiwi, and Muneer Al-Hammadi. A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition. *Information Fusion*, 53:209–221, 2020.

[481] Sze Chit Leong, Yuk Ming Tang, Chung Hin Lai, and CKM Lee. Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing. *Computer Science Review*, 48:100545, 2023.

[482] Hanwei Liu, Huiling Cai, Qincheng Lin, Xiwen Zhang, Xuefeng Li, and Hui Xiao. Feda: Fine-grained emotion difference analysis for facial expression recognition. *Biomedical Signal Processing and Control*, 79:104209, 2023.

[483] K Tawsif, Nor Azlina Ab Aziz, J Emerson Raja, J Hossen, and MZH Jesmeen. A systematic review on emotion recognition system using physiological signals: Data acquisition and methodology. *Emerging Science Journal*, 6(5):1167–1198, 2022.

[484] Ambuj Mehrish, Navonil Majumder, Rishabh Bharadwaj, Rada Mihalcea, and Soujanya Poria. A review of deep learning techniques for speech processing. *Information Fusion*, page 101869, 2023.

[485] Dias Issa, M Fatih Demirci, and Adnan Yazici. Speech emotion recognition with deep convolutional neural networks. *Biomedical Signal Processing and Control*, 59:101894, 2020.

[486] K Ezzameli and H Mahersia. Emotion recognition from unimodal to multimodal analysis: A review. *Information Fusion*, 99:101847, 2023.

[487] Masoumeh Sharafi, Mohammadreza Yazdchi, Reza Rasti, and Fahimeh Nasimi. A novel spatio-temporal convolutional neural framework for multimodal emotion recognition. *Biomedical Signal Processing and Control*, 78:103970, 2022.

[488] Jose Maria Garcia-Garcia, Maria Dolores Lozano, Victor MR Penichet, and Effie Lai-Chong Law. Building a three-level multimodal emotion recognition framework. *Multimedia Tools and Applications*, 82(1):239–269, 2023.

[489] Alaa Alslaity and Rita Orji. Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. *Behaviour & Information Technology*, 43(1):139–164, 2022.

[490] Jianhua Zhang, Zhong Yin, Peng Chen, and Stefano Nichele. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*, 59:103–126, 2020.

[491] Md Rabiul Islam, Mohammad Ali Moni, Md Milon Islam, Md Rashed-Al-Mahfuz, Md Saiful Islam, Md Kamrul Hasan, Md Sabir Hossain, Mohiuddin Ahmad, Shahadat Uddin, Akm Azad, et al. Emotion recognition from eeg signal focusing on deep learning and shallow learning techniques. *IEEE Access*, 9:94601–94624, 2021.

[492] Manoj Kumar Singh, Mona Singh, et al. A deep learning approach for subject-dependent & subject-independent emotion recognition using brain signals with dimensional emotion model. *Biomedical Signal Processing and Control*, 84:104928, 2023.

[493] Wang Kay Ngai, Haoran Xie, Di Zou, and Kee-Lee Chou. Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources. *Information Fusion*, 77:107–117, 2022.

[494] Sihem Nita, Salim Bitam, Matthieu Heidet, and Abdelhamid Mellouk. A new data augmentation convolutional neural network for human emotion recognition based on ecg signals. *Biomedical Signal Processing and Control*, 75:103580, 2022.

[495] Cunhang Fan, Heng Xie, Jianhua Tao, Yongwei Li, Guanxiong Pei, Taihao Li, and Zhao Lv. Icaps-reslstm: Improved capsule network and residual lstm for eeg emotion recognition. *Biomedical Signal Processing and Control*, 87:105422, 2023.

[496] Wafa Mellouk and Wahida Handouzi. Cnn-lstm for automatic emotion recognition using contactless photoplythesmographic signals. *Biomedical Signal Processing and Control*, 85:104907, 2023.

[497] R Gnana Praveen, Patrick Cardinal, and Eric Granger. Audio-visual fusion for emotion recognition in the valence-arousal space using joint cross-attention. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 5(3):360–373, 2023.

[498] Shiqing Zhang, Yijiao Yang, Chen Chen, Ruixin Liu, Xin Tao, Wenping Guo, Yicheng Xu, and Xiaoming Zhao. Multimodal emotion recognition based on audio and text by using hybrid attention networks. *Biomedical Signal Processing and Control*, 85:105052, 2023.

[499] Himanshu Kumar and A Martin. Artificial emotional intelligence: Conventional and deep learning approach. *Expert Systems with Applications*, 212:118651, 2022.

[500] Andrius Dzedzickis, Artūras Kaklauskas, and Vytautas Bucinskas. Human emotion recognition: Review of sensors and methods. *Sensors*, 20(3):592, 2020.

[501] Değer Ayata, Yusuf Yaslan, and Mustafa E Kamasak. Emotion recognition from multimodal physiological signals for emotion aware healthcare systems. *Journal of Medical and Biological Engineering*, 40:149–157, 2020.

[502] Tao Zhang, Minjie Liu, Tian Yuan, and Najla Al-Nabhan. Emotion-aware and intelligent internet of medical things toward emotion recognition during covid-19 pandemic. *IEEE Internet of Things Journal*, 8(21):16002–16013, 2020.

[503] M Shamim Hossain and Ghulam Muhammad. An audio-visual emotion recognition system using deep learning fusion for a cognitive wireless framework. *IEEE Wireless Communications*, 26(3):62–68, 2019.

[504] Jinyan Xie, Xiangmin Xu, and Lin Shu. Wt feature based emotion recognition from multichannel physiological signals with decision fusion. In *2018 first asian conference on affective computing and intelligent interaction (ACII Asia)*, pages 1–6. IEEE, 2018.

[505] Amr Mostafa, Mahmoud I Khalil, and Hazem Abbas. Emotion recognition by facial features using recurrent neural networks. In *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, pages 417–422. IEEE, 2018.

[506] Rahasya Barkur, Devishi Suresh, Mahesh Kumar TN, AV Narasimhadhan, et al. Ensemblewave: An ensembled approach for automatic speech emotion recognition. In *2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pages 1–6. IEEE, 2022.

[507] Zhongmin Wang, Xiaoxiao Zhou, Wenlang Wang, and Chen Liang. Emotion recognition using multimodal deep learning in multiple psychophysiological signals and video. *International Journal of Machine Learning and Cybernetics*, 11(4):923–934, 2020.

[508] Akshi Kumar, Kapil Sharma, and Aditi Sharma. Memor: A multimodal emotion recognition using affective biomarkers for smart prediction of emotional health for people analytics in smart industries. *Image and Vision Computing*, 123:104483, 2022.

[509] Yifeng Zhao and Deyun Chen. Expression eeg multimodal emotion recognition method based on the bidirectional lstm and attention mechanism. *Computational and Mathematical Methods in Medicine*, 2021:1–12, 2021.

[510] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[511] Daniel Haase and Manuel Amthor. Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14600–14609, 2020.

[512] Binh-Son Hua, Minh-Khoi Tran, and Sai-Kit Yeung. Pointwise convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 984–993, 2018.

[513] Lovejit Singh. Deep bi-directional lstm network with cnn features for human emotion recognition in audio-video signals. *International Journal of Swarm Intelligence*, 7(1):110–122, 2022.

[514] Mona Algarni, Faisal Saeed, Tawfik Al-Hadhrami, Fahad Ghabban, and Mohammed Al-Sarem. Deep learning-based approach for emotion recognition using electroencephalography (eeg) signals using bi-directional long short-term memory (bi-lstm). *Sensors*, 22(8):2976, 2022.

[515] Soumyya Kanti Datta, Mohammad Abuzar Shaikh, Sargur N Srihari, and Mingchen Gao. Soft attention improves skin cancer classification performance. In *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data*, pages 13–23. Springer, 2021.

[516] Mohammad Abuzar Shaikh, Tiehang Duan, Mihir Chauhan, and Sargur N Srihari. Attention based writer independent verification. In *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 373–379. IEEE, 2020.

[517] Lin Zhang, Steffen Walter, Xueyao Ma, Philipp Werner, Ayoub Al-Hamadi, Harald C Traue, and Sascha Gruss. "biovid emo db": A multimodal database for emotion analyses validated by subjective ratings. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–6. IEEE, 2016.

[518] AV Geetha, T Mala, D Priyanka, and E Uma. Multimodal emotion recognition with deep learning: Advancements, challenges, and future directions. *Information Fusion*, 105:102218, 2024.

[519] Stanisław Saganowski, Bartosz Perz, Adam G. Polak, and Przemysław Kazienko. Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review. *IEEE Transactions on Affective Computing*, 14(3):1876–1897, 2023.

[520] Lucia Pepa, Luca Spalazzi, Marianna Capecci, and Maria Gabriella Ceravolo. Automatic emotion recognition in clinical scenario: A systematic review of methods. *IEEE Transactions on Affective Computing*, 14(2):1675–1695, 2023.

[521] Bei Pan, Kaoru Hirota, Zhiyang Jia, and Yaping Dai. A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods. *Neurocomputing*, 561:126866, 2023.

[522] Pragya Singh Tomar, Kirti Mathur, and Ugrasen Suman. Unimodal approaches for emotion recognition: A systematic review. *Cognitive Systems Research*, 77:94–109, 2023.

[523] A Aruna Gladys and V Vetriselvi. Survey on multimodal approaches to emotion recognition. *Neurocomputing*, 556:126693, 2023.

[524] Juan Antonio Domínguez-Jiménez, Kiara Coralia Campo-Landines, Juan C Martínez-Santos, Enrique J Delahoz, and Sonia H Contreras-Ortiz. A machine learning model for emotion recognition from physiological signals. *Biomedical Signal Processing and Control*, 55:101646, 2020.

[525] Reem Hamed Aljuhani, Areej Alshutayri, and Shahd Alahdal. Arabic speech emotion recognition from saudi dialect corpus. *IEEE Access*, 9:127081–127085, 2021.

[526] Nazish Azam, Tauqir Ahmad, and Nazeef Ul Haq. Automatic emotion recognition in healthcare data using supervised machine learning. *PeerJ Computer Science*, 7:e751, 2021.

[527] Yujian Cai, Xingguang Li, and Jinsong Li. Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review. *Sensors*, 23(5):2455, 2023.

[528] Suci Dwijayanti, Muhammad Iqbal, and Bhakti Yudho Suprapto. Real-time implementation of face recognition and emotion recognition in a humanoid robot using a convolutional neural network. *IEEE Access*, 10:89876–89886, 2022.

[529] Muhammad Hussain, Emad-Ul-Haq Qazi, Hatim A. AboAlSamh, and Ihsan Ullah. Emotion recognition system based on two-level ensemble of deep-convolutional neural network models. *IEEE Access*, 11:16875–16895, 2023.

[530] Ioannis Kansizoglou, Evangelos Misirlis, Konstantinos Tsintotas, and Antonios Gasteratos. Continuous emotion recognition for long-term behavior modeling through recurrent neural networks. *Technologies*, 10(3):59, 2022.

[531] Bhuvaneshwari Chakravarthi, Sin-Chun Ng, MR Ezilarasan, and Man-Fai Leung. Eeg-based emotion recognition using hybrid cnn and lstm classification. *Frontiers in Computational Neuroscience*, 16:1019776, 2022.

[532] Rajesh Singh, Sumeet Saurav, Tarun Kumar, Ravi Saini, Anil Vohra, and Sanjay Singh. Facial expression recognition in videos using hybrid cnn & convlstm. *International Journal of Information Technology*, 15(4):1819–1830, 2023.

[533] Simrith C Prince and Nirmal Varghese Babu. Advancing multiclass emotion recognition with cnn-rnn architecture and illuminating module for real-time precision using facial expressions. In *2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)*, pages 1–11. IEEE, 2024.

[534] Yan Zhao, Ming Guo, Xuehan Sun, Xiangyong Chen, and Feng Zhao. Attention-based sensor fusion for emotion recognition from human motion by combining convolutional neural network and weighted kernel support vector machine and using inertial measurement unit signals. *IET Signal Processing*, 17(4):e12201, 2023.

[535] Samson Akinpelu and Serestina Viriri. Speech emotion classification using attention based network and regularized feature selection. *Scientific Reports*, 13(1):11990, 2023.

[536] Yekta Said Can, Bhargavi Mahesh, and Elisabeth André. Approaches, applications, and challenges in physiological emotion recognition—a tutorial overview. *Proceedings of the IEEE*, 111(10):1287–1313, 2023.

[537] Aseel Mahmoud, Khalid Amin, Mohamad Mahmoud Al Rahhal, Wail S Elkilani, Mohamed Lamine Mekhalfi, and Mina Ibrahim. A cnn approach for emotion recognition via eeg. *Symmetry*, 15(10):1822, 2023.

[538] Md Asadur Rahman, Anika Anjum, Md Mahmudul Haque Milu, Farzana Khanam, Mohammad Shorif Uddin, and Md Nurunnabi Mollah. Emotion recognition from eeg-based relative power spectral topography using convolutional neural network. *Array*, 11:100072, 2021.

[539] Mehmet Akif Ozdemir, Murside Degirmenci, Elif Izci, and Aydin Akan. Eeg-based emotion recognition with deep convolutional neural networks. *Biomedical Engineering/Biomedizinische Technik*, 66(1):43–57, 2021.

[540] Md Rabiul Islam, Md Milon Islam, Md Mustafizur Rahman, Chayan Mondal, Suvojit Kumar Singha, Mohiuddin Ahmad, Abdul Awal, Md Saiful Islam, and Mohammad Ali Moni. Eeg channel correlation based model for emotion recognition. *Computers in Biology and Medicine*, 136:104757, 2021.

[541] Ketan Sarvakar, R Senkamalavalli, S Raghavendra, J Santosh Kumar, R Manjunath, and Sushma Jaiswal. Facial emotion recognition using convolutional neural networks. *Materials Today: Proceedings*, 80:3560–3564, 2023.

[542] Shervin Minaee, Mehdi Minaei, and Amirali Abdolrashidi. Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors*, 21(9):3046, 2021.

[543] M Kalpana Chowdary, Tu N Nguyen, and D Jude Hemanth. Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Computing and Applications*, 35(32):23311–23328, 2023.

[544] Zi-Yu Huang, Chia-Chin Chiang, Jian-Hao Chen, Yi-Chian Chen, Hsin-Lung Chung, Yu-Ping Cai, and Hsiu-Chuan Hsu. A study on computer vision for facial emotion recognition. *Scientific Reports*, 13(1):8425, 2023.

[545] Gaurav Meena, Krishna Kumar Mohbey, Ajay Indian, Mohammad Zubair Khan, and Sunil Kumar. Identifying emotions from facial expressions using a deep convolutional neural network-based approach. *Multimedia Tools and Applications*, 83:15711–15732, 2024.

[546] Farah Muhammad, Muhammad Hussain, and Hatim Aboalsamh. A bimodal emotion recognition approach through the fusion of electroencephalography and facial sequences. *Diagnostics*, 13(5):977, 2023.

[547] Adrian R Aguiñaga, Daniel E Hernandez, Angeles Quezada, and Andrés Calvillo Téllez. Emotion recognition by correlating facial expressions and eeg analysis. *Applied Sciences*, 11(15):6987, 2021.

[548] Ying Tan, Zhe Sun, Feng Duan, Jordi Solé-Casals, and Cesar F Caiafa. A multimodal emotion recognition method based on facial expressions and electroencephalography. *Biomedical Signal Processing and Control*, 70:103029, 2021.

[549] Bogdan Mocanu, Ruxandra Tapu, and Titus Zaharia. Multimodal emotion recognition using cross modal audio-video fusion with attention and deep metric learning. *Image and Vision Computing*, 133:104676, 2023.

[550] Dilnoza Mamieva, Akmalbek Bobomirzaevich Abdusalomov, Alpamis Kutlimuratov, Bahodir Muminov, and Taeg Keun Whangbo. Multimodal emotion detection via attention-based fusion of extracted facial and speech features. *Sensors*, 23(12):5475, 2023.

[551] Asif Iqbal Middya, Baibhav Nag, and Sarbani Roy. Deep learning based multimodal emotion recognition using model-level fusion of audio–visual modalities. *Knowledge-Based Systems*, 244:108580, 2022.

[552] Dou Hu, Xiaolong Hou, Lingwei Wei, Lianxin Jiang, and Yang Mo. Mm-dfn: Multimodal dynamic fusion network for emotion recognition in conversations. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7037–7041. IEEE, 2022.

[553] Panagiotis Tzirakis, Jiaxin Chen, Stefanos Zafeiriou, and Björn Schuller. End-to-end multimodal affect recognition in real-world environments. *Information Fusion*, 68:46–53, 2021.

[554] Hoai-Duy Le, Guee-Sang Lee, Soo-Hyung Kim, Seungwon Kim, and Hyung-Jeong Yang. Multi-label multimodal emotion recognition with transformer-based fusion and emotion-level representation learning. *IEEE Access*, 11:14742–14751, 2023.

[555] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv:2010.11929*, 2020.

[556] Matti Pietikäinen and Guoying Zhao. Two decades of local binary patterns: A survey. In *Advances in independent component analysis and learning machines*, pages 175–210. Elsevier, 2015.

[557] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. IEEE, 2011.

[558] Diganta Misra, Trikay Nalamada, Ajay Uppili Arasanipalai, and Qibin Hou. Rotate to attend: Convolutional triplet attention module. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3139–3148, 2021.

[559] Bo-Hao Su, Chun-Min Chang, Yun-Shao Lin, and Chi-Chun Lee. Improving speech emotion recognition using graph attentive bi-directional gated recurrent unit network. In *Interspeech*, pages 506–510, 2020.

[560] Yiming Wang, Hui Yu, Weihong Gao, Yifan Xia, and Charles Nduka. Mgeed: A multimodal genuine emotion and expression detection database. *IEEE Transactions on Affective Computing*, 15(2):606–619, 2024.

[561] Saira Bano, Nicola Tonellotto, Pietro Cassarà, and Alberto Gotta. Fedcmd: A federated cross-modal knowledge distillation for drivers' emotion recognition. *ACM Transactions on Intelligent Systems and Technology*, 15(3):1–27, 2024.

[562] Zi Cheng, Lin Shu, Jinyan Xie, and CL Philip Chen. A novel ecg-based real-time detection method of negative emotions in wearable applications. In *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pages 296–301. IEEE, 2017.