

A Deep-Learning Framework for Detecting and Predicting Clinical Events Using Continuous, Multimodal Physiological Signals

by

Sara Ross-Howe

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2023

© Sara Ross-Howe 2023

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: Dr. Sri Krishnan
Professor, Dept. of Electrical, Computer,
and Biomedical Engineering,
Toronto Metropolitan University
- Supervisor(s): Dr. Hamid Tizhoosh
Professor, Dept. of Systems Design Engineering,
University of Waterloo
Kimia Laboratory
- Internal-External Member: Dr. Otman Basir
Professor, Dept. of Electrical Engineering,
University of Waterloo
- Internal Member: Dr. Kumaraswamy Ponnambalam
Professor, Dept. of Systems Design Engineering,
University of Waterloo
- Internal Member: Dr. Siby Samuel
Assistant Professor, Dept. of Systems Design 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

All experimental results, figures, and text are generated from my own work during this Ph.D. research. Background research for this thesis is based on the early results that I have published:

1. S. Ross-Howe and H.R. Tizhoosh. *Atrial fibrillation detection using deep features and convolutional networks*. In IEEE International Conference on Biomedical and Health Informatics (BHI'19), May 19-22 2019.
2. S. Ross-Howe and H.R. Tizhoosh. *The effects of image pre- and post-processing, wavelet decomposition, and local binary patterns on u-nets for skin lesion segmentation*. In IEEE Conference on Computational Intelligence (WCCI 2018), July 8-13 2018.

Abstract

There are an estimated 313 million surgeries performed worldwide each year. Even with significant clinical and technical advances in perioperative research, many patients experience a major complication during the first 30 days following surgery. In recent years, there has been significant advancement in wearable technology and digital health platforms to support remote patient monitoring. Research into machine learning models for predicting adverse clinical events have predominantly focused on utilizing static, derived vital metrics extracted from Electronic Health Record (EHR) systems. However, many limitations have been identified with developing machine learning models from EHR data. Deep learning offers a solution to these challenges by analyzing the physiological signals directly to organize and automatically extract progressive layers of features. Preliminary research in deep learning has just begun for biomedical signal analysis and has been predominately limited to processing individual biometric channels and signal modalities.

This dissertation presents a novel deep learning framework, called the *BiometricNet*, for processing continuous, multimodal physiological signals for the detection and prediction of adverse clinical events. In the initial signal pre-processing stage of the *BiometricNet*, an integrated Generative Multiscale Wavelet De-Noising Autoencoder (Ψ -GANDAE) is utilized to remove common noise patterns encountered during ambulatory signal collection. Signal segmentation and feature extraction is performed in the second processing phase of the *BiometricNet*, where ResNet and BiLSTM architectures leverage post-processing attention layers (RBLAN) to support multiple signal sensor formats (e.g., electrocardiograms, photoplethysmograms, respiration, temperature, and arterial blood pressure), and collation of multiple synchronized channels. In the final stage of the *BiometricNet*, detection and prediction of adverse health events is achieved by a Siamese Neural Network (SNN), which produces a risk score based on the comparison of a baseline signal with samples taken through the patient event timeline.

Important considerations addressed in the proposed framework include optimizing the use of available data, given the lack of large publicly available labelled datasets in healthcare, and addressing the imbalanced nature of clinical data, where there is a chronic scarcity representing target adverse event conditions. The *BiometricNet* is flexible and adaptive to support application to a wide array of adverse clinical events, and it is demonstrated on continuous non-invasive blood pressure estimations from synchronized ECG and PPG signals, and myocardial infarction predication from ECG Lead I, II, and III channels. Lastly, this research honours important considerations regarding AI ethics and network interpretability and transparency to facilitate regulatory approval processes and garner confidence from clinical practitioners.

Acknowledgements

I would like to thank my supervisor Dr. Hamid Tizhoosh for his wisdom, patience, and constant support. He created an inclusive and magical space in the Kimia laboratory, with a mission to conduct meaningful research and improve the lives of patients everywhere. I would also like to thank the entire Kimia laboratory for their continual support, collaboration, and inspiration to push for excellence. Special thanks to Mitra for her kindness and help staying organized.

I would like to express my deep gratitude to Cloud DX for supporting my research through these many years. Without an alignment between my work in industry and my academic pursuits, this thesis never would have been possible. To all the talented people on my Research and Development team at Cloud DX, your unending encouragement and commitment has inspired me to keep going. Special thanks to Dr. Michael McGillion for his sage advice and for all the long conversations imagining a better future for post-surgical patients.

My profound gratitude goes to all of the patients that donated their time and experience to participate in the clinical trials that are critical to this research.

I would also like to thank Dejan Kocetkov for his graphic arts contributions on [Figure 7](#) and [Figure 2](#).

[Figure 8](#) and [Figure 26](#) have been designed using images from Flaticon.com:

- Risk icons created by ultimatearm.
- ECG monitor icons created by Marcus Christensen.

Dedication

This is dedicated to my children, Ainsley and Maclean, my husband, Scott, my parents, Carlene and Randolph, and my big brother Sean (who lost his battle with childhood Leukemia).

Your unwavering encouragement and unending sacrifices carried me up this mountain.

“Courage is going from failure to failure without losing enthusiasm.” - Winston Churchill

“Our greatest weakness lies in giving up. The most certain way to succeed is always to try just one more time.” – Thomas Edison

Table of Contents

List of Tables	x
List of Figures	xii
List of Abbreviations	xiv
1 Research Motivation	1
1.1 Thesis Organization	5
2 Literature Review	7
2.1 Deep Learning Architectures	7
2.1.1 Autoencoders	7
2.1.2 Recurrent Neural Networks	9
2.1.3 Long Short-Term Memory (LSTM) Networks	10
2.1.4 Convolutional Neural Networks	11
2.1.5 Siamese Neural Networks	12
2.1.6 Generative Adversarial Networks	12
2.1.7 Supervised, Unsupervised, and Reinforcement Learning	13
2.2 Detection & Prediction of Clinical Complications	13
2.2.1 Sepsis Prediction	13
2.2.2 Delirium Prediction	14

2.2.3	Patient Deterioration Prediction	15
2.2.4	Adverse Clinical Event Prediction	16
2.2.5	Mortality Prediction	18
2.3	Deep Learning for Physiological Signal Analysis	19
2.3.1	Cardiac Arrhythmia Classification	19
2.3.2	Signal Noise Filtering	21
2.3.3	Myocardial Infarction Detection and Prediction	23
2.3.4	Other Prediction Models	24
2.4	Summary and Analysis	25
3	The Biometric Neural Network	27
3.1	Thesis Contributions	28
3.2	Phase 1: Signal Pre-Processing	29
3.2.1	Prologue	29
3.2.2	Method	29
3.2.3	Generative Multiscale Wavelet De-Noising Autoencoder for Electrocardiography Signal De-noising	37
3.2.3.1	Data	37
3.2.3.2	Results	38
3.2.3.3	Analysis	39
3.3	Phase 2: Feature Extraction and Signal Segmentation	45
3.3.1	Prologue	45
3.3.2	Method	45
3.3.3	Continuous Blood Pressure Estimation with Deep Features and Attention	50
3.3.3.1	Data	54
3.3.3.2	Results	55
3.3.3.3	Analysis	57

3.4	Phase 3: Regression and Classification	65
3.4.1	Method	65
3.4.2	A Novel Siamese Network for Myocardial Infarction Prediction	68
3.4.2.1	Data	69
3.4.2.2	Results	70
3.4.2.3	Analysis	71
3.5	Summary	73
4	Conclusions and Future Directions	75
4.1	Conclusions	75
4.2	Limitations	78
4.3	Future Directions	79
	References	82
	Appendices	105
A	Literature Review of Machine Learning Research for Detection and Prediction of Clinical Complications	106
B	Results of k-Fold Cross-Validation for Ψ-GANDAE	118

List of Tables

3.1	Results of Leave-One-Out Cross-Validation for Ψ -GANDAE on Muscle Artifact	40
3.2	Results of Leave-One-Out Cross-Validation for Ψ -GANDAE on Electrode Motion Artifact	40
3.3	Results of Leave-One-Out Cross-Validation for Ψ -GANDAE on Baseline Wander	41
3.4	Performance of the RBLAN for Diastolic Blood Pressure Estimates	58
3.5	Performance of the RBLAN for Systolic Blood Pressure Estimates	58
3.6	Performance of the BiometricNet for Myocardial Infarction Prediction	71
B.1	Results of k-Fold Cross-Validation for Ψ -GANDAE on Muscle Artifact	119
B.2	Results of k-Fold Cross-Validation for Ψ -GANDAE on Electrode Motion Artifact	120
B.3	Results of k-Fold Cross-Validation for Ψ -GANDAE on Baseline Wander	121

List of Figures

1	Vitaliti TM Continuous Vital Sign Monitor (permission from Cloud DX) . . .	3
2	Remote Automated Monitoring Architecture	4
3	Vanilla Autoencoder	8
4	Recurrent Neural Network (Source: LeCun et al., 2015)	9
5	Recurrent Neural Network Unit Versus a Long Short-Term Memory Unit - (Source: Olah, C., August 27, 2015)	10
6	Architecture of the LeNet-5 (Source: LeCun, 1999)	11
7	ECG Cardiac Cycle	22
8	Deep Learning Architecture for the <i>BiometricNet</i>	28
9	Scalogram of ECG Segment	31
10	Wavelet Packet Decomposition Tree for ECG Signal	32
11	Generator Architecture of the Generative Multiscale Wavelet De-noising Autoencoder (Ψ -GANDAE)	35
12	Discriminator Architecture of the Generative Multiscale Wavelet De-noising Autoencoder (Ψ -GANDAE)	36
13	ECG Signals with Noise and Reconstructed Output from Generative Multiscale Wavelet Denoising Autoencoder (Ψ -GANDAE) k-Fold Cross-Validation	43
14	ECG Signals with Noise and Reconstructed Output from Generative Multiscale Wavelet Denoising Autoencoder (Ψ -GANDAE) Leave-One-Out Cross Validation	44

15	Convolutional Neural Network - Bi-Directional Long Short Term Memory Network with Attention Feature Extraction	46
16	Application of Convolutional Filter and Average Pooling on ECG Data Segment	46
17	Pulse Transit Time Measurements	54
18	Signal Segments from Pulse Transit Time PPG Dataset	56
19	Bland-Altman Plots for Difference in Diastolic Estimates to Reference Values	59
20	Bland-Altman Plots for Difference in Systolic Estimates to Reference Values	60
21	Scatter Plots for Diastolic Estimates	61
22	Scatter Plots for Systolic Estimates	62
23	Heat Map of Attention Mask for PPG Signal	64
24	Heat Map of Attention Mask for ECG Signal	64
25	Siamese Neural Network for Multimodal Biometric Signal-Based Classification	66
26	Prediction of Myocardial Infarction from ECG Signals Extracted from the Patient ICU Admission Timeline	69
27	Receiver Operator Characteristics Curves for Myocardial Infarction Prediction	72
28	Precision-Recall Curves for Myocardial Infarction Prediction	72

List of Abbreviations

AMI Acute Myocardial Infarction

AUROC Area Under the Receiver Operating Characteristics

cNIBP Continuous Non-Invasive Blood Pressure

CNN Convolutional Neural Network

DAE De-noising Autoencoder

ECG Electrocardiogram

EHR Electronic Health Records

EWS Early Warning Scores

GAN Generative Adversarial Network

LSTM Long Short-Term Memory

MI Myocardial Infarction

MINS Myocardial Injury Non-Surgical

PPG Photoplethysmogram

PTT Pulse Transit Time

RAM Remote Automated Monitoring

ResNet Residual Neural Network

RMSE Root Mean Squared Error

SNN Siamese Neural Network

SNR Signal-to-Noise Ratio

Chapter 1

Research Motivation

There are an estimated 313 million surgeries performed worldwide each year, with an additional 143 million more required to save lives and prevent disabilities in low-income and middle-income countries.¹ In Canada, that translates to 6,857 surgeries per 100,000 population annually.² Even with significant clinical and technical advances in perioperative research, more than 25% of adults over the age of 45 will have a major complication during the first 30 days following noncardiac surgery.³ Major complications following surgery include myocardial infarction (MI), myocardial injury non-surgical (MINS), sepsis, infection, bleeding, delirium and strokes.³⁻¹⁰ Tragically, on average 1.8% of adults worldwide that undergo surgery will die due to postoperative complications. Mortality rates for these complications range from as high as 6.4% in Africa to 1.1% in North America.³

Postoperative adverse events that have the highest attributable fraction to mortality include major bleeding (17%), MINS (15.9%), sepsis (12.0%), and infection without sepsis (2.8%).³ While stroke represents a small number of surgical complications, with an average incidence rate of 0.08%-0.7% (higher in head and neck surgery at 4.8%), it leads to mortality in 18%-26% of all cases.⁹ Postoperative Atrial Fibrillation (PAF), a cardiac arrhythmia that is characterized by frequent and irregular ventricular beats, is the most common complication arising after coronary bypass surgery (30% of all cases) and results in a twofold increase in the incidence of cerebral infarction in patients during an ICU stay (2.3% rate for non-PAF versus 5.3% for PAF group).¹¹ Additionally, delirium, which is a state of cognitive impairment and altered consciousness that can lead to patient injury and poor clinical outcomes, was observed in 24.4% of post-surgical patients with critical illnesses.¹⁰

A substantial number of patient deaths do not occur in the operating room (0.70%), but rather after surgery in the ICU and general wards (69.93%), or following discharge home (29.37%).³ Vital sign monitoring in the hospital during surgery is often done on a continuous basis with specialized equipment, but this is reduced on the general wards to measurements taken at intermittent and irregular intervals (2.0-5.4 readings/day).¹² The diligence of the monitoring frequency is also dependent on the vital metric type, with blood pressure being the most frequently taken measurement at 5.0 readings/day, heart rate at 4.4 readings/day, temperature at 4.2 readings/day, and respiration rate at only 1.0 reading/day.¹² The level of patient support and vital sign monitoring then often drops even more considerably following the patient's discharge from the hospital and transition home.

In recent years, there has been significant advancement in wearable technology and digital health platforms to support remote patient monitoring. These low power *Internet of Things* (IoT) devices offer the potential to capture multiple continuous biometric signals from patients outside of the confines of the hospital environment.^{13,14} One such *Mobile Health* (mHealth) technology is being developed by Cloud DX, a wearable Continuous Vital Sign Monitor (CVSM) called VitalitiTM (see [Figure 1](#)). This device is designed to continuously collect and derive the major vital signs through a chest-mounted wearable device with a detachable in-ear sensor. Parameters measured include: 5-Lead ECG (sampled at 1 kHz), heart rate and heart rate variability, respiration rate (sampled at 100 Hz), core body temperature (infrared sensor in ear sampled at 1 Hz), blood oxygen saturation (Red and IR channels sampled at 1 kHz), continuous non-invasive blood pressure (cNIBP), and pulse transit time.¹³



Figure 1: Vitaliti™ Continuous Vital Sign Monitor (permission from Cloud DX)

A *remote automated monitoring* (RAM) architecture for supporting patients in their home during their post-surgery convalescence is illustrated in Figure 2.^{13,15,16} The clinical grade wearable sensors are worn continuously by the patient throughout their recovery to derive key vital metrics, capture physiological signals (e.g., electrocardiograms), and provide real-time detection of cardiac arrhythmias (e.g., atrial fibrillation). These vital metrics and system diagnostic information are transported in near real-time through wireless transfer protocols, such as Bluetooth Low Energy (BLE), to a tablet or smartphone. The smartphone relays pertinent information to a HIPAA-compliant¹ cloud infrastructure through Wifi or Cellular networks. The cloud infrastructure provides distributed and secure storage for the patient data and analyzes trends to produce Early Warning Scores (EWS) indicative of patient deterioration and adverse clinical events. Notifications are generated for the clinical care team when the EWS fall beyond clinically established thresholds. The care team then further investigates these notifications, and where required provides meaningful clinical interventions.

A before-and-after study was conducted to investigate the impact of a RAM workflow for deteriorating patients in a general medical ward in the UK.¹⁷ A Philips IntelliVue Guardian Solution was applied to patients and spot-check measurements by the nurses (along with data from sensors) captured respiratory rate, blood pressure, heart rate, pulse oximetry, and temperature, and automatically transferred these metrics to the hospital Electronic Medical Record (EMR) system for processing. Three different EWS

¹Health Insurance Portability and Accountability Act

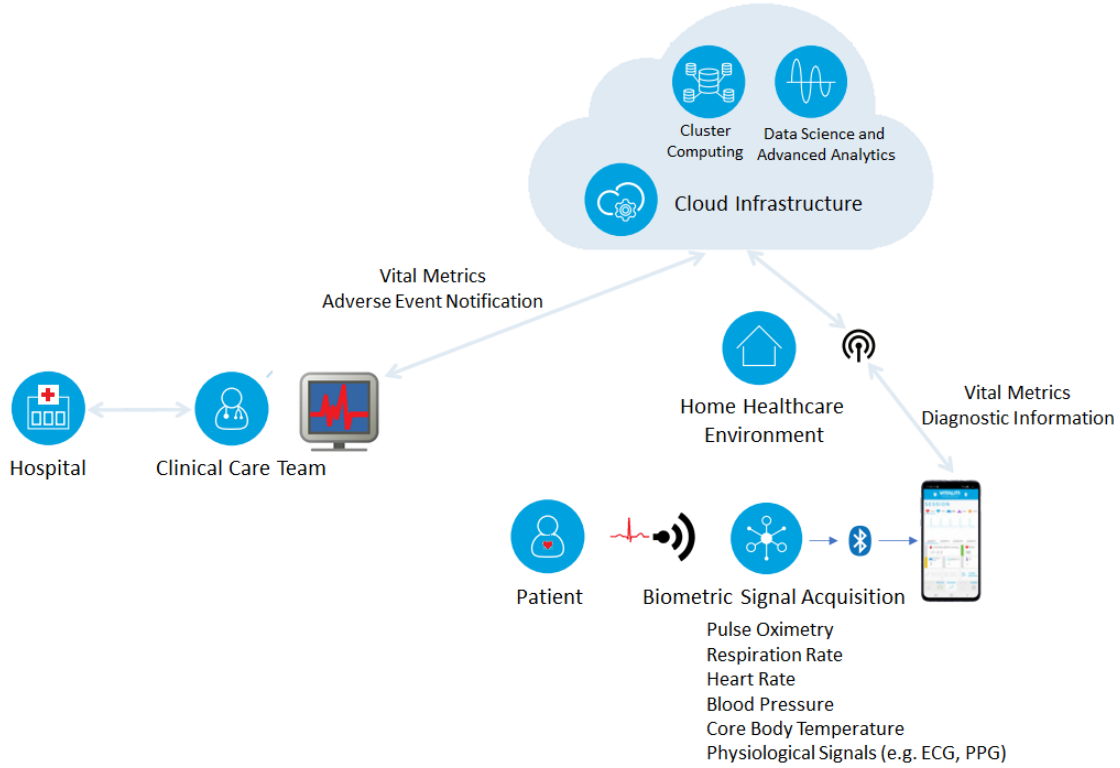


Figure 2: Remote Automated Monitoring Architecture

were derived from the patient vital metrics including the National Early Warning Score (NEWS), the Chronic Respiratory Early Warning Score (CREWS), and a palliative care algorithm. The study was conducted over a one-year control period where data was collected from 2,139 patients, and a subsequent one-year intervention phase with 2,263 patients. The study showed some promising early results for the application of RAM workflows, as demonstrated by a reduction of mortality from 8.09% in the control period to 6.50% in the intervention period ($p = 0.042$). A reduction was also seen in the number of major complications, with the incidence of severe sepsis reduced from 0.98% to 0.04% ($p < 0.001$) and cardiopulmonary arrest from 0.66% to 0.09% ($p = 0.002$). The severity of patients' illness was also reduced for those requiring ICU intervention with a mean Acute Physiology and Chronic Health Evaluation II score of 26 (SD 9) during the control period versus 18 (SD 8) during the intervention period.

Early Warning Scores (EWS) have been developed to identify serious clinical complications in a healthcare setting. These scores are often developed from gold standard-of-care and trial-and-error iterations of the designated weights and thresholds to balance identification with alarm fatigue.¹⁸ EWSs are typically represented by linear combinations of measured vital metrics and are limited in the complexity of events that can be detected. Machine learning methodologies are increasingly being leveraged in recent literature to develop more complex models to identify adverse clinical events. Deep learning topologies hold particular promise as they can predict more complicated adverse clinical states and are able to process the underlying physiological signals directly, without significant feature engineering efforts.

The initial intended commercial use for the outlined research will be to support patients during post-surgical convalescence in the home healthcare environment. Deep learning models will be architected to reside in the cloud infrastructure and provide inference on collected high fidelity physiological signals from the VitalitiTM clinical grade wearable device. Due to the cost of bandwidth on cellular networks, as well as the potential for delay in transferring signals through the mobile network, it is advantageous to move inference to the middleware layer, or on the device itself with edge-based computing for near real-time performance. Limitations on the resources available in the embedded development environment will constrain the size of the predictive models that can be supported and techniques such as network distillation will be explored.¹⁹ It is also advantageous to bring clinical decision-making to the point-of-care to eliminate the risk of lost or delayed data during transfer to the cloud infrastructure, and thus hinder notice for timely medical interventions. While the initial intended use scenario is for post-surgical patients, the proposed solution has garnered interest for Emergency Room triage as well as expanded “hospital-at-home” scenarios.

1.1 Thesis Organization

The focus of this thesis will be to present a novel deep-learning framework to support the processing of continuous, multimodal biometric signals in near real-time for the detection and prediction of complex adverse clinical events. Current research in machine learning, and more recently deep learning, for patient deterioration and adverse clinical event prediction will first be explored in chapter 2. A background on the machine-learning technologies that are leveraged throughout the thesis will be provided, along with an analysis of the current challenges and opportunities in the field of adverse clinical event detection and prediction. The deep learning architecture to support real-time processing

of continuous multimodal biometric signals for the identification and prediction of clinical events will then be developed in chapter 3, along with details of relevant experimentation to demonstrate the utility of each stage of the solution. The proposed deep learning framework will be separated into three phases to represent the stages of processing, namely: 1) Signal pre-processing, 2) Signal segmentation and feature extraction, and 3) Regression or classification. Lastly, the conclusions, limitations, and future directions for this research will be discussed in chapter 4.

Chapter 2

Literature Review

2.1 Deep Learning Architectures

An overview of deep networks is first provided to acquaint the reader with the architectures that were most commonly employed in recent literature. These deep topologies provide the essential components for developing models to predict the onset of adverse events and patient deterioration in a clinical setting. Readers familiar with these deep learning approaches are invited to skip to section 2.2 for the literature review.

2.1.1 Autoencoders

The concept of the autoencoder has been present in artificial neural network research for over 30 years.^{20–22} In early works, they were used predominately for “dimensionality reduction or feature learning” (Goodfellow et al., page 393).²³ In a paper published in 2006 by G. E. Hinton, and R. R. Salakhutdinov entitled “Reducing the Dimensionality of Data with Neural Networks” a method was outlined for initializing the weights for autoencoders to a sufficiently close approximation to allow the networks to converge and encode feature vectors in a manner that is better than principal component analysis.²⁴

An *autoencoder* is an unsupervised learning approach, where x_i represents both the input and output label y_i . For a basic, or *Vanilla Autoencoder*, a feedforward network is used and backpropagation training is applied to learn an output function where $\mathbf{h}$ represents the encoding function and $\mathbf{r}$ a reconstruction function:²³

$$\mathbf{h} = f(x), \tag{2.1}$$

$$\mathbf{r} = g(\mathbf{h}). \quad (2.2)$$

The loss function used represents the reconstruction error between the output and the original inputs as measured by the Mean Squared Error or L2 loss:

$$\mathcal{L}(x - \hat{x}) = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \hat{x}_i)^2. \quad (2.3)$$

The layers in this network are incomplete (the number of nodes in the hidden layer are fewer than in the input layer), thus forcing the network to compress the input to a better encoding representation (see [Figure 3](#) below).

A de-noising autoencoder (DAE) is used for input features that contain an additive noise distribution:

$$\tilde{x}_i = x_i + z_i \sim \mathcal{N}(\mu, \sigma). \quad (2.4)$$

The reconstruction loss is then represented by the input vectors without the noise. The goal of the encoding-decoding process is to reconstruct the input vectors without the signal noise.

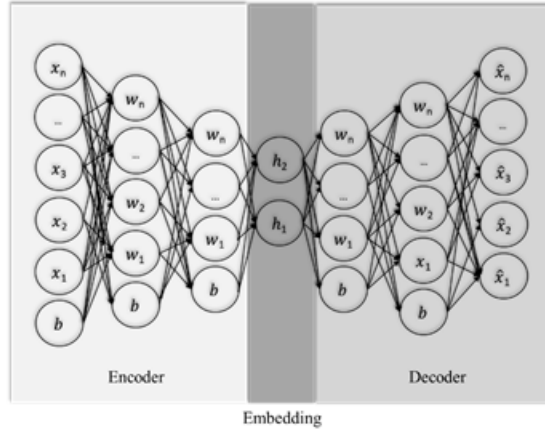


Figure 3: Vanilla Autoencoder

The vanilla autoencoder relies on a deterministic model for the network parameters, but the Variational Autoencoder (VAE) models the inputs, hidden units, and output as stochastic models.^{25,26} The loss function therefore uses a probabilistic formulation of the log-likelihood reconstruction loss, as well as a regularizer term based on Kullback-Leibler

divergence. Where the encoder network is represented by $Q(z|X)$, the decoder network by $P(X|z)$, and z is the encoded representation²⁷ :

$$\mathcal{L}(X) = \mathbb{E}_{Q(z|X)} \log P(X|z) - KL(Q(z|X) || P(z)). \quad (2.5)$$

One of the benefits of stochastic parameter modelling includes the ability for the network to better generalize on unseen data. The VAE can also be used to generate simulated data from the learned probability distribution. With a sparse autoencoder (SAE), the hidden layer over-represents the input vector with additional nodes, and penalty terms are used to encourage sparsity during training. Sparsity can be achieved by removing weights during the training process. The loss function used in training includes the reconstruction loss between the input and output vector, as well as a regularization, where λ is the sparsity penalty term, and $\mathbf{w}_j$ the model weights:

$$\mathcal{L}(x - \hat{x}) = \frac{1}{N} \sum_{i=0}^N (x_i - \hat{x}_i)^2 + \lambda \sum_{j=0}^M \|\mathbf{w}_j\| \quad (2.6)$$

2.1.2 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) were introduced by Rumelhart et al. in 1986 and are specifically designed for use with sequential or temporally related inputs.²⁸ The recurrent portion of the network are feedback loops that allow outputs from the previous step to be used as inputs for the next processing unit (See [Figure 4](#) below). A limitation of this structure is that for long sequences of input the network can experience the vanishing gradient problem due to repeated applications of the activation function.

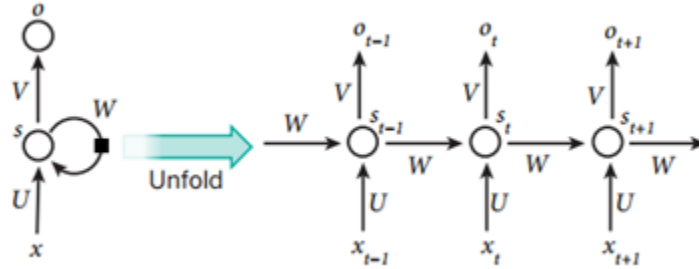


Figure 4: Recurrent Neural Network (Source: LeCun et al., 2015)

2.1.3 Long Short-Term Memory (LSTM) Networks

A solution to the vanishing gradient problem was proposed by Hochreiter and Schmidhuber in 1997 with their Long Short-Term Memory (LSTM) architecture²⁹ (see Figure 5 below). In this network architecture, the LSTM cell uses a forget gate to allow control of how many inputs to use from previous cells:³⁰

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \quad (2.7)$$

As well as what new inputs should be stored in the cell state (C_t) with an input gate, where:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (2.8)$$

$$\check{C}_t = \tanh(C_i \cdot [h_{t-1}, x_t] + b_c), \quad (2.9)$$

$$C_t = f_t * C_{t-1} + i_t * \check{C}_t, \quad (2.10)$$

An output gate is used to control how much of the cell state to provide in the output (where h_t represents the output of the cell):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (2.11)$$

$$h_t = o_t * \tanh(C_t). \quad (2.12)$$

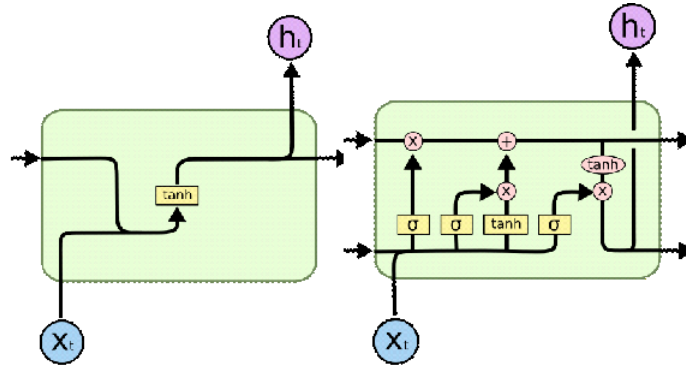


Figure 5: Recurrent Neural Network Unit Versus a Long Short-Term Memory Unit - (Source: Olah, C., August 27, 2015)

Long-Short Term Memory models have been applied successfully for natural language processing (NLP) tasks (character-based models;³¹ word-based models³²), as well as speech

recognition.³³ The LSTM model is able to identify sequential dependencies in data with the control of outputs from previous time steps. This makes them well suited to analysis of time series data where temporal connections can be identified. A variant of the LSTM is a bi-directional LSTM (BiLSTM) that processes the data in a forward and backward direction.

2.1.4 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) were formally introduced by LeCun et al. in 1999 through their paper titled “Object Recognition with Gradient-Based Learning”,³⁴ which outlined a new convolution-based architecture in the LeNet-5 network for digit recognition (See Figure 6). In this architecture, a bank of convolutional filters is first applied to the input matrix to produce multiple feature maps. These feature maps are then down sampled to further reduce the dimensionality of the inputs, with average pooling also helping to promote generalization for unseen data. The convolutional and average pooling steps are repeated and the resultant feature maps are flattened as inputs to two final fully connected layers. This method is well-suited for processing 2-dimensional matrices but can also be used against 1-dimensional vector inputs (such as biomedical signals).

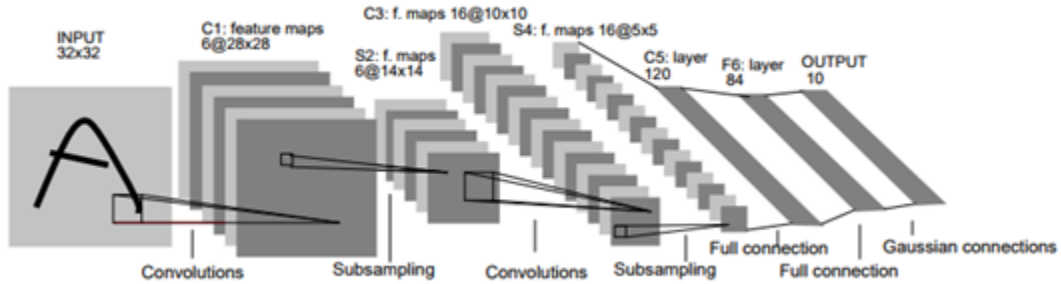


Figure 6: Architecture of the LeNet-5 (Source: LeCun, 1999)

Convolution is a mathematical operation that is applied against two functions, where one is flipped and shifted across the other while measuring the integral of the overlap. The convolution of $[f * g](t)$ is thus represented as:³⁵

$$[f * g](t) \stackrel{def}{=} \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau. \quad (2.13)$$

Convolution in the time domain is multiplication in the frequency domain, and this operation is thus very suitable for image and signal filters.

Convolutional Networks (CNN) provide a foundation from which many more complex network architectures have been based for both image classification and segmentation tasks.³⁶⁻³⁸ For example, Fully Connected Networks (FCNs) were one of the first network architectures to utilize CNNs for end-to-end, pixel-to-pixel semantic segmentation.³⁸ FCNs utilize a convolution-deconvolution architecture that does not contain any fully connected layers. AlexNet was introduced at the ImageNet 2012 competition, where it produced a top five test error of 15.4% on the difficult ImageNet data set, consisting of over 15 million images and representing over 22,000 categories (the next best competitor achieved a 26.2% top 5 error rate).³⁶

2.1.5 Siamese Neural Networks

Siamese Neural Networks were first introduced in 1993 by Bromley et al.³⁹ and were initially demonstrated for the task of signature verification. A Siamese Neural Network consists of twin sub-networks that share the same weights and process two inputs concurrently. After the processing stage, comparison layers are utilized to determine a difference or similarity metric between the two inputs. The Siamese Neural Network architecture has been demonstrated effectively on the task of one-shot image recognition.⁴⁰

2.1.6 Generative Adversarial Networks

Goodfellow et al. first introduced Generative Adversarial Networks (GANs) to the world in seminal work published in 2014.⁴¹ Since that time, GANs have had wide adoption in the machine learning community, setting new performance standards in the fields of large language models⁴² and image generation.^{41,43} Arguably, GANs have had the best translation to mainstream applications, and produced some of the most sensational results in the field. OpenAI is an industry leader on GAN technologies, with their GPT-4 text-or image-to-text model,⁴² and DALL-E natural language-to-image⁴³ commercial models.

GAN architectures consist of two distinct models, a generator and a discriminator. The generator creates content in the modality of choice, and the discriminator tries to determine if it is generated or real content. The tension between the generator and discriminator drives both to iterate to higher levels of performance. The discriminator model is typically discarded at the end of the training process, as it is only used as scaffolding to improve the performance of the generator model.

2.1.7 Supervised, Unsupervised, and Reinforcement Learning

Prevalent strategies for training machine learning models include supervised, unsupervised, and reinforcement techniques.²³ Supervised learning is associated with regression (continuous values or ordinal values) and classification problems (categorical or unordered), where labels exist to define desired network input-output mappings. Unsupervised techniques are utilized for data sets without associated labels and include clustering and generative modelling approaches.⁴⁴ Reinforcement learning defines a reward mechanism for networks to measure performance against and networks are defined by actions and states. Supervised learning is preferred for developing predictive models as it leverages gold standard case labels and represents the majority of research that has been done in this field.

2.2 Detection & Prediction of Clinical Complications

Machine learning is increasingly being leveraged in recent literature to develop more complex models to identify serious clinical complications. Much of this research utilizes derived vital metrics, typically measured during standard care in the hospital, to predict the onset of common adverse conditions, such as sepsis,⁴⁵⁻⁵¹ adverse clinical events,⁵²⁻⁶⁰ delirium,⁶¹⁻⁶⁵ patient deterioration,⁶⁶⁻⁷³ and mortality.⁷⁴⁻⁷⁷ See [Appendix A](#) for a summary of current research that has been conducted utilizing machine learning methodologies for detecting and predicting clinical complications from features readily available in the patient data captured in the EMR, such as vital metrics, laboratory results, patient symptoms, and diagnosis.

2.2.1 Sepsis Prediction

Sepsis is a serious medical condition caused by a cascade immune response to an infection, which is associated with 5.5% of all deaths in Canada from 2009-2011.⁷⁸ Sepsis detection and prediction has been a recent focus of supervised machine learning research.^{45-51,79,80} All of the early research identified utilized key vital metrics as input features to machine learning models, with heart rate (HR), respiration rate (RR), peripheral capillary oxygen saturation (SpO₂), systolic blood pressure (SBP), and temperature included in all the classification models reviewed, and diastolic blood pressure (DBP), and Glasgow Coma Score (GCS) included in most models. Improved prediction accuracy was obtained by Henry et. al⁴⁶ by also including key laboratory results, patient co-morbidities, and more

advanced indexes (e.g. shock index HR/SBP) in their TREWScore algorithm, which was also aligned with the results from Fagerström et al. in 2019.⁵⁰ In 2020,⁷⁹ a more comprehensive event sequence was assembled containing approximately 5,000 features from information extracted from the Electronic Health Records (EHR) admissions system. This data included medication information, laboratory results, medical imaging, diagnostics and procedures from the national patient register, and sociodemographic information from a civil register system. This extensive feature vector was utilized as an input to a CNN-LSTM network, producing an Area Under the Receiver Operating Characteristics (AUROC) curve of 0.856 at three hours to onset. The most common machine learning approaches utilized in sepsis prediction models were gradient tree boosting and ensemble forests,^{47–49} but the best classification accuracy in the reviewed research utilized an LSTM network with an AUROC of 0.88 at four hours to onset.⁵⁰ All research reported the detection accuracy at onset and then at variable time increments up to a maximum prediction window of 72 hours before diagnosis. To address concerns noted with model generalization for single site studies, a large, international, multi-centre study was conducted on 136,478 unique ICU admissions collected in the US, the Netherlands, and Switzerland between 2001 and 2016.⁸⁰ A deep learning model with dense outer layers and causal self-attention inner layers produced an average AUROC of 0.846 (95% CI, 0.841–0.852), and when fine-tuned on a hold-out validation set, achieved a transfer performance to other sites with AUROC of 0.807 (95% CI, 0.801–0.813).⁸⁰

2.2.2 Delirium Prediction

Delirium is a common post-operative complication occurring in 24.4% of critically ill patients and is associated with a 25% hospital mortality rate, versus 6% for non-delirious patients ($p < 0.01$).¹⁰ Research efforts have focused on utilizing preoperative variables gathered during the patient’s admission assessment as inputs into a variety of machine learning models to predict post-operative delirium risk.^{61–65,81} These input attributes include continuous and discrete variables across a wide array of care domains, comprising demographics, socioeconomic, clinical, pharmacy, laboratory, and vital signs measurements. In this early phase of research, a diverse set of machine learning models, with both linear and non-linear solution spaces have been investigated. Modelling approaches utilized in the identified research include: Logistic Regression (LR), Generalized Additive Model (GAM), Random Forests (RF), Support Vector Machines (SVM) with radial basis functions (RBF), Artificial Neural Networks (ANN), Extreme Gradient Boosting (XGB), Gradient Boosting Machine (GBM), and Naïve Bayes (NB). A GBM model produced the best results in one study,⁸¹ with a resultant AUROC of 0.933 (95%

CI, 0.929 - 0.936) for delirium prediction, utilizing demographic data as inputs to the model, as well as a 56 item nursing care assessment that measures the abilities and impairments of patients in areas such as level of consciousness and motor skills. It is important to note that a comparison between the published research was difficult given there were no standard data sets used between the research groups.

State-of-the-art research has just begin to move beyond structured EMR data to leverage deep learning methods to process physiological signals directly for delirium prediction. Mulkey et al.⁸² utilized a vision transformer model that processed electroencephalogram (EEG) waveforms, which produced a 97% accuracy for predicting delirium. Ahmed et al.⁸³ utilized EMR data as well as actigraphy information from wrist-mounted sensors, with the resultant performance of the machine learning models (Random Forest, SVM and XGBoost) demonstrating an overall mean accuracy of 74% (10%), and F1 score of 65% (10%).

2.2.3 Patient Deterioration Prediction

Other research has focused on detecting general patient deterioration that is characterized by the need for life-saving interventions (LSI), transfer to ICU, cardiac arrest, and/or mortality.^{66-73,84} In the identified research, all models utilized a combination of discrete, continuous, or categorical variables extracted from patient demographics, laboratory results, and vital metrics gathered during the admission process, transport to the hospital, or during the in-patient stay. Key vital metrics, such as respiration rate, heart rate, and pulse rate, were captured at more frequent intervals of up to 2 seconds, and non-invasive BP sometimes captured at longer intervals of up to 2 hours.^{71,72} Patients were also monitored with intra-arterial catheters in some studies to directly measure arterial blood pressure.^{70,84} In Moss et al.,⁷² ECG data was also processed into Data Quality Indexes (DQIs) to provide indications of overall signal quality, and heart rate variability (HRV) measurements from R-R peak intervals extracted to provide information on cardiac arrhythmias and ectopic beats. Early research predominantly utilized Multivariate Logistic Regression,^{67,72,84} with more recent research including deeper model topologies such as Recurrent Neural Networks (RNNs),⁶⁶ Levenberg-Marquardt ANNs,⁶⁹ LSTMs, and residual temporal convolutional networks (TCNs).⁸⁵

Most recently, notable results were demonstrated by Mehrdad et al.⁸⁵ with a residual temporal convolutional network (TCN) that utilized vital signs (i.e. oxygen saturation, heart rate, and temperature) over a proceeding interval as network inputs, achieving an AUROC of 0.9336 on the detection of deterioration in patients with COVID-19. The

prediction window in the identified research ranged from 6-8 hours for Colopy et al.⁶⁸ with a clinically relevant false positive rate of 10%-15%, to the work from Boulitsakis Logothetis et al.,⁸⁶ which provided a 24-hour prediction window with an AUROC of 0.92. Many of the data sets used were limited in size due, in part, to the cost and effort to collect study data in a clinical setting when these metrics are not already part of standard care. Exceptions to those limitations are the collated data set by Churpek et al.⁶⁷ with data from 269,999 patients collected over five hospitals in Chicago, and Boulitsakis Logothetis et al.⁸⁶ with 118,886 patients that had unplanned admissions to Salford Royal Hospital in the UK. In Mehrdad et al.,⁸⁵ a substantial data set of 37,006 patients admitted to NYU Langone Health through the COVID-19 pandemic was utilized to support training of deeper network topologies for prediction of patient deterioration.

2.2.4 Adverse Clinical Event Prediction

Independent machine learning models have also been developed to detect and predict specific adverse clinical events, such as: hypertension (SBP > 120 and DBP $\geq$ 80 or MBP $\geq$ 105 mmHg), hypotension (SBP < 90 and DBP $\leq$ 60 or MBP $\leq$ 70 mmHg), tachycardia (HR > 100 beats per min.), bradycardia (HR < 60 beats per min.), tachypnea (RR > 17 breaths per min), bradypnea (RR < 12 breaths per min), hypoxia (SpO₂ < 93%), bleeding, renal failure, and heart failure^{52-60,87} as early precursors to mortality. Much of the identified research involved retrospective studies that utilized clinical features that were previously captured in the EHR, including: vital metrics (HR, SBP, DBP, MBP, RR, SpO₂), laboratory test results, arterial blood gases, and balance output (e.g. urine output). Tsur et al.⁵³ also included pulse pressure (PP=SBP-DBP) and relative cardiac output (CO=HR*PP) along with SBP, DBP, and HR measured at one minute intervals to build time series data and then extract statistical features (mean, median, standard deviation, variance, interquartile range, skewness, kurtosis and linear regression slope), wavelet features, and cross-correlation features from 30-240 minute observation windows. These features were used as inputs into an Extreme Gradient Boosting (XGBoost) ensemble network, which achieved an AUROC of 0.937 for predicting hypotensive events after 240 minutes of observation, with a 60-minute gap window, and a 30-minute target window.⁵³

Most recently, a deep stacked recurrent network was used to predict Acute Kidney Injury (AKI) to determine cases that required follow-up dialysis.⁸⁷ The AKI prediction model achieved a 90.2% accuracy with a 2-to-1 ratio of false alarms to true events, at an event horizon up to a 72 hour prediction window.⁸⁷ The window of prediction for these hemodynamic and cardio-respiratory events was typically 15-120 minutes,⁶⁰ which is much smaller than those achieved for general patient deterioration detection models.

The exception to this was a model to predict heart failure that utilized an extended 12-month prediction window.⁵⁷ Random Forests and XGBoost models were top performers for hypotension prediction,^{53,54} while Recurrent Neural Networks (RNNs) were used for predicting more complex medical states such as heart failure, renal failure, bleeding, and mortality.^{55,57}

Probabilistic models using Hidden Markov Models (HMMs) were developed with success to identify a broad array of cardio-respiratory events, such as hypertension, hypotension, tachycardia, bradycardia, tachypnea, bradypnea, and hypoxia with an average accuracy of 97.8% on clinically complex combinations of these events.⁵⁸ The work by Hatib et al.⁶⁰ was also notable, where a logistic regression model was developed to predict hypotension events 15 minutes before onset, achieving a sensitivity of 83.6%, specificity of 83.3%, and an AUROC of 0.91 on the external validation set. A set of 51 features was utilized as inputs to this hypotension prediction model, each extracted from individual beats in the arterial blood pressure waveform, separated into five phases (systolic phase, diastolic phase, systolic rise phase, systolic decay phase, and overall decay phase). Many researchers solely used the MIMIC-II (III) clinical ICU dataset for training and testing models, while Tomašev et al.⁸⁷ utilized information from the US Department of Veterans Affairs collected from 2011-2015 that included 703,782 patients records.

A prediction model for cardiac arrest in the pediatric ICU was developed at the Hospital for Sick Children in Toronto,⁸⁸ which utilized a CNN feature extraction layer followed by an LSTM. Individual time series were assembled for measurements taken at five second intervals for heart rate, respiration rate, pulse rate, oxygen saturation, arterial blood pressure, and standard deviation of R-R intervals. Each of these physiological signals was processed by the CNN+LSTM model for a given window of analysis to produce a prediction score, and then added together in an ensemble model with a linear weighting of each physiological signal as:

$$\text{Risk Score} = \frac{\sum_{i=1}^N (W_i * \text{Risk}_i) * \mathbb{1}_i}{\sum_{i=1}^N (W_i) * \mathbb{1}_i}, \quad (2.14)$$

where $\mathbb{1}_i = 0$ if signal i is missing and $\mathbb{1}_i = 1$ otherwise. A Risk_i was determined by training a small network to attribute for the age of the patient, with:

$$\text{Risk}_i = \text{Sigmoid}(r * \text{agescale}), \quad (2.15)$$

$$\text{agescale} = \text{ReLU}(\text{age} * W_{\text{age}} + b_{\text{age}}), \quad (2.16)$$

Where r represents a primary risk probability, W_{age} the network weights, and bias b_{age} . On a small test cohort of 20 patients with at least one incidence of cardiac arrest, the weighted

risk score demonstrated alarms above a 0.5 threshold at 30 minutes prior to event in all cases.

2.2.5 Mortality Prediction

An ultimate manifestation of prognostic clinical research has been towards mortality prediction for patients in a critical care setting. It is noted that this research is associated with more controversial ethical questions regarding the negative impact it could have on the quality of care that patients receive who are deemed to have a high mortality risk. As a counterargument, it might also have the potential to improve care standards by identifying patients that require more aggressive interventions to alter current patient trajectories. At the very least, research conducted in this field is transferable to predicting other clinical adverse events as it provides labelled classifications that are frequently available in many EHR systems.

Seminal work in the field of mortality prediction was conducted by Du et al. in 2016,⁷⁴ which utilized 73 clinical variables gathered from the MIMIC-II database that reflect patient demographics and severity scores (e.g. Simplified Acute Physiology Score-SAPS, Sequential Organ Failure Assessment-SOFA, and Glasgow Coma Score-GCS), co-morbidities, laboratory results during admission, vital signs (median values of first 24 hours), and interventions (e.g. use of mechanical ventilation, vasopressor, and sedative medications). These 73 features were inputs into a Deep Belief Network (DBN) with Genetic Algorithms (GA) that were used to determine an optimal network topology. This work was notable for the use of unsupervised learning methodologies in contrast to supervised learning methods employed almost exclusively by all the other research identified in the arena of health outcome prediction. The DBN demonstrated a high mortality prediction accuracy at 98% against a 28-day horizon.

The majority of research reviewed utilized a subset of clinical variables available through retrospective analysis from the EHR and associated data registries, with attribute categories including patient demographics, laboratory results, vital signs, interventions, care events (e.g. procedures), and diagnosis codes. Models typically employed include Random Forest (RF),^{75,76} Gradient Boosting Machines (GBM),^{89,90} and ensemble models⁹¹ (that contain a combination of Gradient Boosting Machine-GBM, Random Forest-RF, Support Vector Machines-SVM, Naïve Bayes Model-NBM). An LSTM model was utilized with a breadth of clinical features as inputs to achieve an AUROC of 0.9030 for a five year prediction window,⁹⁰ and a convolutional Neural Network (CNN) achieved an AUROC of 0.881 with a 60-hour mortality prediction window.⁷⁷

Novel research by Wang et al.,⁹² used Natural Language Processing (NLP) to extract features from patient demographic information and longitudinal clinical notes with an LSTM model to predict patient mortality against a 2-year event horizon with an AUROC of 0.943. In the most recent research, Sun et al.⁹³ used only ECG waveforms as inputs into a ResNet architecture, achieving an AUROC of 0.812 (95% CI, 0.808–0.816) for a one year prediction horizon, and an AUROC of 0.798 (95% CI, 0.792–0.803) for a five year event horizon.

2.3 Deep Learning for Physiological Signal Analysis

Deep learning is a more recent specialization of machine learning, which utilizes representation learning to organize and automatically extract progressive layers of features directly from raw data, without significant feature engineering efforts.⁴⁴ Deep learning architectures, such as Generative Adversarial Networks (GANs), Convolution Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) (specifically Long Short-Term Memory), have been demonstrated to surpass previous methods in the fields of machine vision,³⁶ speech recognition,⁹⁴ natural language processing,⁹⁵ and even beat world renowned human experts in games of logic such as Go.⁹⁶

Deep learning applications in healthcare have had slower adoption than in many other fields though, likely due to the scarcity of data because of concerns relating to the Personal Health Information Protection Act (PHIPPA), poorly labelled and unbalanced data for supervised learning techniques, and lack of precedent for approval from the FDA, Health Canada, and the European Commission. These regulatory challenges are being overcome with the first deep learning algorithm cleared by the FDA for analyzing cardiac MRI images in 2017.⁹⁷ Deep learning methods are a natural candidate to use in the detection of complex patterns from raw biomedical signals and offer the potential to surpass machine learning models that utilize only derived clinical metrics as network inputs. Deep learning architectures have been emerging in recent years for the analysis of physiological signals for cardiac arrhythmia classification,^{98–108} signal de-noising,^{109–111} and myocardial infarction detection and prediction.^{112,113}

2.3.1 Cardiac Arrhythmia Classification

Atrial fibrillation is a cardiac arrhythmia that affects an estimated 33.5 million people in the world, or approximately 0.5% of the global population.¹¹⁴ This arrhythmia is characterized

by rapid and irregular ventricular beats and the resultant chaotic heart rhythm is a key contributor to the occurrence of strokes. The Heart and Stroke Foundation of Canada estimates that atrial fibrillation is the cause of 1 in 3 strokes for people over the age of 60,¹¹⁵ and the CDC estimates that this cardiac arrhythmia is the cause of 130,000 deaths each year in the United States.¹¹⁶

Applications of deep learning towards the classification of cardiac arrhythmias in ECG signals have recently begun to emerge. In the 2017 PhysioNet/Computing in Cardiology (CinC) Challenge, a significant number of the solutions involved deep models for the classification of cardiac rhythms. Winning classification methods from the CinC competitors as demonstrated on a reference dataset included many successful deep learning topologies for atrial fibrillation (AFIB) detection.^{98–101} The common deep network architectures shared across this work includes CNNs and RNNs.^{98–100} Representative results from these methods demonstrated a sensitivity of 82.5% and specificity of 98.7% on the CinC dataset.¹⁰¹

Hybrid combinations of recurrent and convolutional networks have been proposed to capture the unique sequential properties and morphological features of biomedical signals.^{100–102} Techniques are also being introduced to capture both time and frequency domain characteristics of the signals as better inputs for machine learning models. A continuous wavelet transform (CWT) was used by He et al.¹⁰³ to bound frequency information to a time scale in a window surrounding an identified beat (from the annotated reference), which was then represented by a scalogram for subsequent processing by a CNN. This combined CWT and convolutional neural network approach achieved a sensitivity of 99.41% and specificity of 98.91% on the MIT-BIH AFIB database.

In seminal research by Hannun et al.,¹⁰⁴ a deep learning model for cardiac arrhythmia detection using convolutional networks was introduced. This work involved the collection of a significant data set for training and testing a deep convolutional network with 34 layers. Data was collected from 29,163 patients from a single lead ECG recording, sampled at 200 Hz, for a duration up to 14 days. The training and testing set was created from 30 second intervals extracted from these patient recordings for a total of 64,121 ECG samples (approximately 534 hours of ECG recordings). The deep network was trained to classify 12 different types of arrhythmias, along with sinus rhythm and noise. A committee of three cardiologists provided the ground truth for the test set and then the model performance was evaluated against annotations made by a second group of cardiologists. The deep model outperformed the average cardiologist in all but two classifications for sequence level metrics and all but one in set level metrics. This predicate research provides evidence that data sets at this scale are adequate for training deep models with biomedical signals for complex classification tasks.

A novel combination of CNNs and de-noising autoencoders (ConvDAE) was used by Ochiai et al.¹⁰⁵ for arrhythmia detection in 2-lead ECG signals. In this research, the input signal x was corrupted by a stochastic mapping to partially destroy the input signal and the loss function used for training was the mean squared error:

$$\tilde{x} \sim C(\tilde{x}|x). \quad (2.17)$$

Both 1-D and 2-D convolutional filters were investigated in this research and 2 leads of ECG signals were processed simultaneously by the system. Performance of the models were evaluated on the MIT-BIH Arrhythmia database for both ventricular ectopic beat (VEB) and supraventricular ectopic beat (SVEB) types. The 2-D ConvDAE produced better classification performance than the 1-D model, with an accuracy of 0.957, sensitivity of 0.847, specificity of 0.9666, and positive predictive value of 0.748 on VEB beat types. This compares favourably with predicate research by Zubair et al.¹⁰⁶ that utilized a 1-D CNN, which achieved an accuracy of 0.951, sensitivity of 0.801, specificity of 0.964, and positive predictive value of 0.740.

Daniel et al.¹⁰⁷ recently utilized a 2-D representation of the ECG segment to capture frequency and time domain information by applying the Morse Wavelet. Transfer learning was employed by training the ECG segments from ResNet 50 and AlexNet pre-trained models. An overall average accuracy of 99.2%, average sensitivity of 99.2%, and average specificity of 99.6%, was achieved on a test set from the MIT-BIH Arrhythmia database. A hybrid model utilizing CNN and LSTM layers, was most recently leveraged by Yildirim et al.¹⁰⁸ for cardiac arrhythmia classification. Their network produced an accuracy of 92.24%, a sensitivity of 80.15%, and specificity of 98.72% on seven arrhythmia patterns including atrial fibrillation and Supraventricular Tachycardia (SVT). Results were tested on a large ECG database representing over 10,000 patients and lead II configuration was found to produce the best results.

2.3.2 Signal Noise Filtering

ECG signals are susceptible to many different sources of noise, particularly during ambulatory collection. Traditional approaches to ECG signal de-noising include Band-Pass Filters (IIR or FIR),¹¹⁷ Adaptive Filters,¹¹⁸ Savitzky-Golay Filters, Empirical Mode Decomposition (EMD),^{119,120} and wavelets.^{121,122} These methods have many recognized challenges that limit digital signal processing performance on ECG signals. Techniques that rely on spectral power isolation of noise, such as band-pass filters, EMD, and wavelets, struggle with overlapping EMG noise regions with QRS complex activity in the 10Hz-20Hz

band, and P and T wave activity with electrode motion artifacts in the $<5\text{Hz}$ range. Electrode motion artifact is particularly difficult to mitigate, as it mimics the appearance of ectopic beats and cannot be easily removed with such filters. Current de-noising approaches can also alter beat morphology, impacting R-peak detection and diagnostic activities (see [Figure 7](#) for an illustration of the cardiac cycle as captured in an ECG signal).

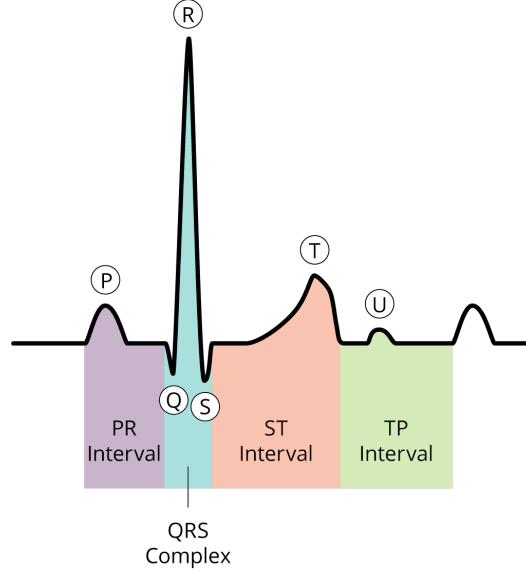


Figure 7: ECG Cardiac Cycle

Autoencoders have been demonstrated as an effective means of removing common noise patterns, such as DC baseline offset drift (BW), muscle artifact (EMG-MA), and electrode movement artifact (EM), within ECG signals.¹⁰⁹ In early work published in 2016,¹⁰⁹ an improved stacked contractive de-noising autoencoder (I-DAE) was demonstrated as more effective at removing common noise rhythms than established traditional filtering methods. This technique utilized a wavelet transform and soft threshold as an initial pre-processing step, before input into the autoencoder. The network was trained using noise rhythms obtained from the MIT-BIH Noise Stress Test database and added to recordings from the MIT-BIH Arrhythmia database at Signal-to-Noise-Ratio (SNR) levels of 0 dB, 1.25 dB, 5 dB. This approach achieved $\text{SNR}=22.83667$ dB and Root Mean Square Error (RMSE)=0.033333 on baseline wander noise corruption, a $\text{SNR}=22.1$ dB and $\text{RMSE}=0.034333$ on electrode movement artifacts, and a $\text{SNR}=22.19333$ dB and $\text{RMSE}=0.033667$ on muscle artifact.

Deep Recurrent Neural Networks (DRNN) were utilized by Antczak¹¹⁰ for ECG signal denoising. An LSTM-based RNN with a 4-layer architecture was initially trained on simulated ECG data and then fine-tuned with real patient data to assess the network’s ability for transfer learning. The network performance was better when simulated data was first used to pre-train the network weights, versus training on real patient data alone. Performance of the network was compared against a band-pass filter with a range of 0.05-150 Hz, and an Undecimated Wavelet Transform (UWT) with a Daubechies D6 mother wavelet. The testing data set contained 5,000 real samples with a SNR from -20 dB to 0 dB. The simple bandpass filter outperformed the UWT and DRNN when there was relatively little noise > -3 dB, while the UWT produced the best results in the -3 dB to -7 dB ranges. The DRNN outperformed the UWT and band-pass filter for noise levels < -7 dB, with a demonstrated stability across noise levels ranging from -9 dB to 0 dB.

More recently, Generative Adversarial Networks (GANs) have been developed for ECG signal denoising.¹¹¹ GANs consist of generator and discriminator sub-networks,¹²³ where the generator filters the incoming signal and the discriminator is trained to determine if the signal is indistinguishable from the ground truth noise-free signal. Once training is complete, the generator is used to filter the signal to remove noise and the discriminator is discarded. The GAN network performance was assessed utilizing the Massachusetts Institute of Technology and Beth Israel Hospital (MIT-BIH) Noise Stress database, where it produced improved Root Mean Square Error (RMSE) and Signal-to-Noise Ratio (SNR) metrics at 5 dB levels as compared to the previous state-of-the-art I-DAE.¹¹¹

2.3.3 Myocardial Infarction Detection and Prediction

Myocardial Infarction (MI), or most commonly referred to as a heart attack, occurs when the blood flow stops or is limited to an area of the heart. In the US, one million myocardial infarctions occur annually, resulting in 300,000 to 400,000 deaths per year.¹²⁴ Han et al.¹¹² utilized a residual network with three residual blocks that each contained convolutional layers, to detect MI in 12-lead ECG signals. This research utilized the PTB (Physikalisch-Technische Bundesanstalt) database with 12-lead ECG recordings captured at a sampling frequency of 1kHz with 16-bit resolution, over a range of ± 16.384 mV, for 13 patients with MI and 52 healthy control patients. Each ECG signal was segmented into four second intervals and processed independently by the residual blocks and then fused together with the features of the other leads into a flattened, fully connected layer. These fused features were then subsequently processed through softmax to the output layer with two classification nodes. This network topology produced a reported accuracy of 95.49%, sensitivity of 94.85%, and specificity of 97.37% on the task of MI detection.

In 2022, Gustafsson et al.¹¹³ assembled a dataset of 492,226 ECG recordings from 214,250 patients gathered in the emergency rooms of Sweden between 2007 and 2016. A ResNet (CNN with skip connections between the layers) was utilized to train a deep network to predict MI in patients. The classification was split into ST-elevation myocardial infarction (STEMI) and non-ST-elevation myocardial infarction (NSTEMI). The ResNet was trained utilizing ECG waveforms and the patient’s age and sex as inputs. The network demonstrated a C-statistic of 0.991 and Brier score of 0.001 on STEMI, and a C-statistic of 0.832 and Brier score of 0.008 on NSTEMIs.

2.3.4 Other Prediction Models

Innovative applications of deep learning models utilizing ECG signals have been developed for age and gender estimation,¹²⁵ emotion detection,^{126,127} and biometrics.^{128,129} Kanjo et al.¹²⁶ utilized multimodal data from the EnvBodySens dataset containing recordings from 40 subjects, composed of on-body raw sensor measurements for galvanic skin response (GSR), body temperature, accelerometer/gyroscope, and derived heart rate, environmental data such as noise levels, UV, air pressure, and location information, along with self-reported emotion levels using the five step Self-Assessment-Manikan (SAM) scale for valence. A CNN-LSTM architecture was utilized, which was able to achieve a 94.7% accuracy, 0.927 precision score, and 0.95 recall on emotion classification. In Attia et al.,¹²⁵ CNNs were trained using 10 second samples from 12-lead ECG signals for 499,727 patients and tested on data from 275,056 patients to predict age and gender. The model achieved a 90.4% classification accuracy for gender determination with an AUROC of 0.97, and age was estimated as a continuous variable with an average error of 6.9 ± 5.6 years ($R^2 = 0.7$). An interesting observation to note, was that in patients where the predicted error > 7 years from the chronological age there was a higher correlation to low ejection fraction, hypertension, and coronary disease ($P < 0.01$).

A deep autoencoder architecture was used by Eduardo et al.¹²⁸ to process 12-lead ECG data sampled at 500 Hz with 16-bit resolution for biometric security identification. Data was collected from 709 different subjects and manually labelled by a domain expert. Signal pre-processing of the ECG signals included a 150th order Finite Impulse Response (FIR) filter with a frequency range of 5 to 20 Hz and then heartbeats extracted from a fixed-length window of -200 to 400 ms surrounding the R-peak. The deep autoencoder demonstrated a 0.91% identification error on 330 test subjects. Most recently, Agrawal et al.¹²⁹ compared the performance of a CNN network to an LSTM network for biometric security identification based on 12-lead ECG. They utilized the ECG-ID Database from Physikalisch-Technische Bundesanstalt (PTB) containing 310 ECG recordings from 90

subjects. The LSTM architecture outperformed the CNN for patient identification with a 99.69% accuracy versus 98.34%, respectively .

2.4 Summary and Analysis

Machine learning models utilizing derived vital metrics and clinical features as inputs have demonstrated improved predictive accuracy for patient deterioration and adverse clinical event detection as compared to standard early warning scores. However, many limitations have been noted with developing machine learning models from static, derived vital metrics extracted from EHR data. Missing or incomplete data is a consistent issue in EHR information systems, where gaps in data quality and quantity can arise from clinician confirmation bias and non-uniform sampling intervals.¹³⁰ Vital signs are manually taken at intervals from 2.0-5.4¹² readings/day depending on the ward and institutional policies regarding measurement intervals, which can also lead to undersampling and cause brief transient events to not be reflected in captured vital measurements. A smoothing effect phenomenon has also been noted in acute clinical wards where the care team internally average measurements before recording them manually to compensate for artifacts and poor signal quality.¹³¹ In a study conducted on 212 patients across surgical and medical units, manually charted oxygen saturation values were on average 6.5% (95% CI, 4.0%-9.0%) higher than those recorded by the automated system.¹³¹

It has also been demonstrated in the critical care unit at the Hospital for Sick Kids in Toronto¹³² that the window used for sampling of vital metrics can significantly impact the variability across core vital measurements. In this study, the variability of systolic blood pressure, heart rate, and oxygen saturation measurements increased as the sampling window was increased from intervals of 10 seconds to 4 hours, which could mask the inherent variability in these metrics caused by the influence of the sympathetic and parasympathetic systems and could deride important pre-cursors of patient deterioration. Similar observations were noted by Tomlinson et al.¹³³ where the Limit of Agreement (LoA) between continuous data and manual observations was -0.57 (-5.63 to 4.48) breaths per minute for respiratory rate, -0.71 (-10.2 to 8.73) beats per minute for heart rate, and -0.07% (-2.67% to 2.54%) for SpO₂. Vital metrics also do not reflect the quality of the inherent physiological signal present when the manual measurements were collected. By integrating data quality indexes with heart rate into a logistic regression model for predicting life-saving interventions (LSIs), the AUROC increased from 0.73 to 0.86.⁸⁴ Model deployment and maintenance are also important to consider, with both internal and external validity concerns for vital metric consistency due to concept drift and shift

across clinical institutions.¹³⁰

Deep learning holds significant promise for analyzing continuous physiological signals and predicting the onset of adverse clinical events. Preliminary research in deep learning has just begun for biomedical signal analysis and has been predominately limited to processing individual biometric channels and signal modalities. The goal of this dissertation is to design, implement, and test a novel deep learning framework that supports the processing of continuous, multimodal biometric signals in near real-time for the detection and prediction of adverse clinical events. Important considerations that have been addressed in the proposed framework include integrated biometric signal de-noising for ambulatory collection, support for multiple signal sensor formats (e.g., electrocardiograms, photoplethysmograms, respiration, temperature, and arterial blood pressure), and collation of multiple synchronized channels. The framework is flexible and adaptive to support application to a wide array of adverse clinical events, and offer detection of these events as prediction into the future (given the current patient trajectory). It was also important that the framework optimize the use of available data, given the lack of large, publicly available labelled datasets in healthcare, primarily due to privacy and regulatory requirements. Considerations have also been given to the imbalanced nature of clinical data, where there is a chronic scarcity of data representing target adverse event conditions. Lastly, this research honours important considerations regarding AI ethics and network interpretability and transparency to facilitate regulatory approval processes and confidence from clinical practitioners.

Chapter 3

The Biometric Neural Network

A novel deep learning framework has been developed that supports the processing of continuous, multimodal biometric signals in near real-time for the detection of patient deterioration and prediction of adverse clinical events. The architecture of this framework is separated into three primary processing stages, namely: 1) Signal pre-processing utilizing a Generative Multiscale Wavelet Autoencoder for signal denoising (Ψ -GANDAE), 2) Signal segmentation and feature extraction utilizing ResNet and BiLSTM architectures with Attention (RBLAN), and 3) Regression or classification with a Siamese Neural Network (SNN). This deep learning framework in its entirety will be referred to as the *Biometric Neural Network* or *BiometricNet*. The high-level architecture of the *BiometricNet* for the analysis of continuous, multimodal physiological signals is illustrated for reference in [Figure 8](#).

The signal pre-processing phase of the *BiometricNet* will first be discussed in section 3.2, with details of the Ψ -GANDAE architecture provided along with experimentation results demonstrating effective signal de-noising on ECG data using the PhysioNet MIT-BIH Arrhythmia¹³⁴ and PhysioNet MIT-BIH Stress Test Databases.¹³⁵ The signal segmentation and feature extraction capabilities of the RBLAN will be presented in section 3.3, with results detailing experiments for continuous non-invasive blood pressure estimation utilizing the PhysioNet Pulse Transit Time (PTT) PPG Dataset.¹³⁶ Finally, in section 3.3 the SNN will be explained and holistic performance of the *BiometricNet* will be demonstrated on MI prediction using the PTB Diagnostic ECG Database¹³⁷ (PTB).

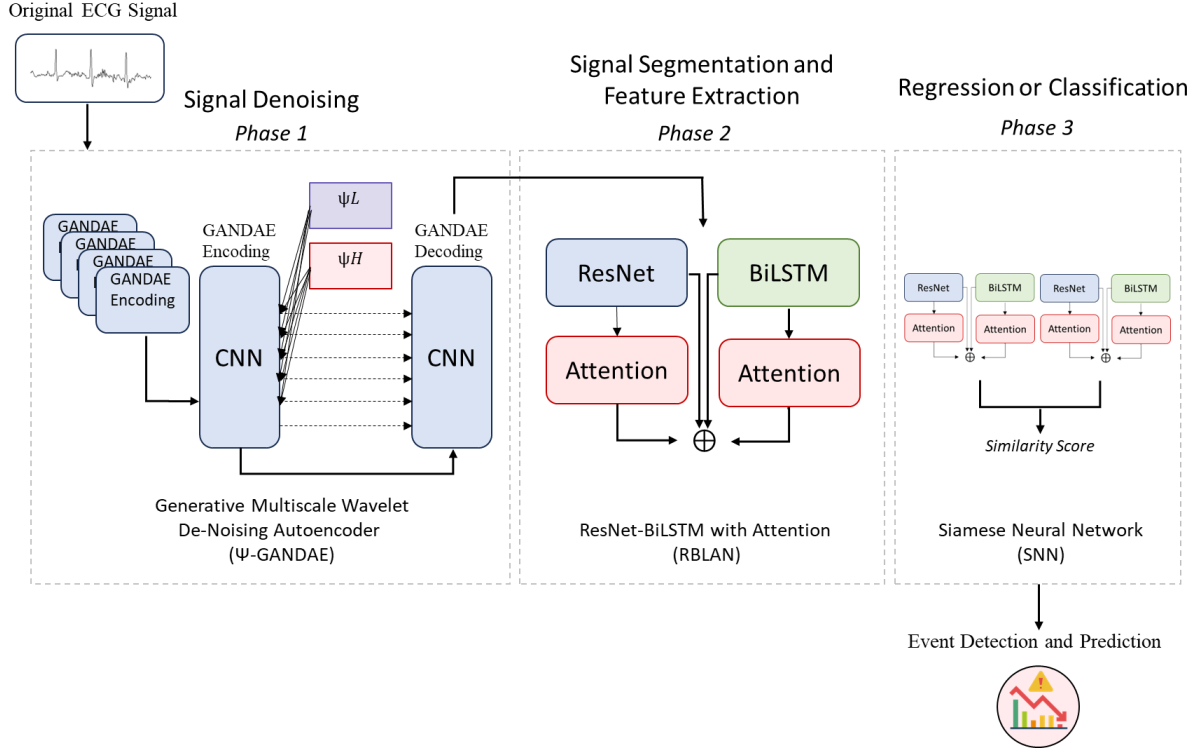


Figure 8: Deep Learning Architecture for the *BiometricNet*

3.1 Thesis Contributions

The key contribution of this dissertation is a novel deep learning framework to support the processing of continuous, multimodal biometric signals for the detection of patient deterioration and prediction of adverse clinical events. Development of the *BiometricNet* follows a system design perspective, with inherent signal pre-processing and feature extraction phases undertaken in relation to resultant regression and classification outcomes. Highlights of the original contributions of this research that have been integrated into the processing framework of the *BiometricNet* include the following:

- A novel Generative Multiscale Wavelet Autoencoder for physiological signal denoising (Ψ -GANDAE).
- A hybrid signal segmentation and feature extraction architecture for multimodal signals utilizing ResNet and BiLSTM architectures with Attention (RBLAN).

- An innovative application of a Siamese Neural Network (SNN) architecture for predicting patient deterioration and adverse clinical events from tandem baseline and sample physiological signal segments.

3.2 Phase 1: Signal Pre-Processing

3.2.1 Prologue

The content of this chapter was informed by background research developed in a conference paper published during the Ph.D. research:

S. Ross-Howe and H.R. Tizhoosh. *The effects of image pre- and post-processing, wavelet decomposition, and local binary patterns on u-nets for skin lesion segmentation*. In IEEE Conference on Computational Intelligence (WCCI 2018), July 8-13 2018.

3.2.2 Method

The initial processing stage of the *BiometricNet* is intended to remove common noise patterns from the signal to facilitate downstream feature extraction. Physiological signals are susceptible to many different sources of noise, particularly during ambulatory collection. A novel Generative Multiscale Wavelet De-noising Autoencoder (Ψ -GANDAE) architecture is proposed as a solution to improve signal de-noising performance for physiological signals. This Ψ -GANDAE is composed of encoder and decoder stages (Figure 11), consistent with de-noising autoencoders, and multiscale decomposition of the ECG signal. The network leverages the signal decomposition capabilities of wavelets, and benefits from elements of ResNet skip connections and Convolutional Networks embedded in U-Net contractile and expansive layers for feature extraction.

The architecture of the Ψ -GANDAE network is structured as a de-noising autoencoder, composed of encoder and decoder stages. The input is the corrupted input signal as represented by $\tilde{x} \in \mathbb{R}$, which is mapped to $h \in \mathbb{R}$ through function $h = f(\tilde{x})$ as the hidden layers of the encoder. The decoder layers then generate the reconstruction of the signal called $z \in \mathbb{R}$ through the function $z = g(h)$. With the goal of minimizing the mean squared reconstruction error of:

$$\frac{1}{N} \sum_{n=0}^{N-1} [x(n) - z(n)]^2, \quad (3.1)$$

Where $x(n)$ is the baseline signal (with all noise removed).

The ECG signal is segmented into a pre-determined length with a sliding sample window as input into the Ψ -GANDAE network. This input signal is then further segmented into four equally sized sample windows, with each sub-sample being processed by an independent instance of the encoder network. The intuition for this multiscale processing of the signal is that the longer segment provides information on the sequential cardiac rhythm for the network, while the nested sub-networks process information on independent beat morphology. All extracted features for the high and lower resolution processing of the signal are concatenated together before the final encoding layer.

The wavelet transform is a signal processing and image transformation approach that has been successfully applied to compression^{138,139} and de-noising problems.^{121,140} The wavelet transform is a time-frequency method where the optimal frequency band is adaptively determined to provide the best time-frequency localization. In the continuous wavelet transform ψ is a continuous function in both the time and frequency domain, often referred to as the *mother wavelet*. For a specific scale a and translation b the continuous wavelet function is represented as:¹⁴¹

$$T^{wav}f(a, b) = |a|^{-\frac{1}{2}} \int dt f(t) \psi \left(\frac{t - b}{a} \right). \quad (3.2)$$

The wavelet transform is a method of signal representation on a scale-time space, where each band or scale can be thought of as a result of a band-pass filter. The effects of applying the wavelet transform iteratively is to construct a multiresolution representation of the signal. When applied to physiological signals, in each subsequent layer of representation the detail residuals are stored along with the approximation.

In the Discrete Wavelet Transform (DWT), the inner product of the signal is taken with the wavelet function ψ :¹⁴¹

$$\psi_x^\psi(m, a^j) = \sum_{n=0}^{N-1} x[n] \psi_j^*[n - m], \quad (3.3)$$

$$\psi_j[m] = \frac{1}{\sqrt{a^j}} \psi \left(\frac{m}{a^j} \right) \quad (3.4)$$

A scalogram can be utilized to visualize the effect of the wavelet transform, with the sequential spectrum information visually encoded into progressive columns represented against a colormap. The scalogram captures the magnitude of the spectrum plotted against time and frequency.¹⁴¹

$$|\psi_x^\psi(m, a^j)|^2 \quad (3.5)$$

The DWT is an ideal method for extracting both time and frequency information from ECG signals as it performs well at localizing transients in non-stationary signals. This is demonstrated in Figure 9 where the QRS complex is clearly identified in the scalogram of the ECG segment (yellow shaded region), but also more detailed morphology is identified such as the P and T waves as illustrated by the shaded clusters in the 2-6 Hz frequency range. It is also particularly beneficial for AFIB detection to be able to capture the presence or absence of the P wave as a defining characteristic of this cardiac arrhythmia.

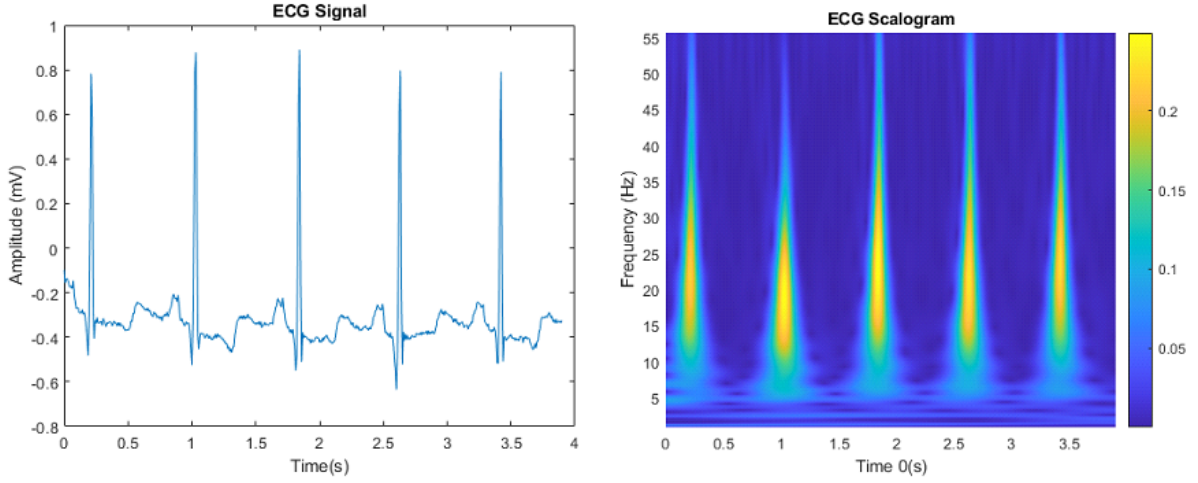


Figure 9: Scalogram of ECG Segment

In the Ψ -GANDAE, a dyadic grid is employed where the ECG signal is processed by a half-band high and low pass filter at each scale (low pass filter $\psi L_{f1, \frac{1}{2}N}$ and high pass filter $\psi H_{f\frac{1}{2}N, N}$ ¹⁴²). At each decomposition level, the high pass filter will represent the detail in the signal and the low pass filter the approximation of the signal. The wavelet packet analysis will be used as both the detail and approximation of the signal are split. See Figure 10 for a representative wavelet decomposition tree for a five second ECG segment, with a Daubechies 10 wavelet to 4 levels of scale. In the final implementation of the Ψ -GANDAE, a Daubechies 6 (db6) mother wavelet is utilized with a decomposition to 5 scales. The db6 wavelet was selected as it is asymmetric with a shape well-suited to the QRS complex in ECG signals. The effectiveness of the db6 for de-noising ECG signals has also been clearly demonstrated in predicate research,¹⁴³ as well as during experimentation with the Ψ -GANDAE.

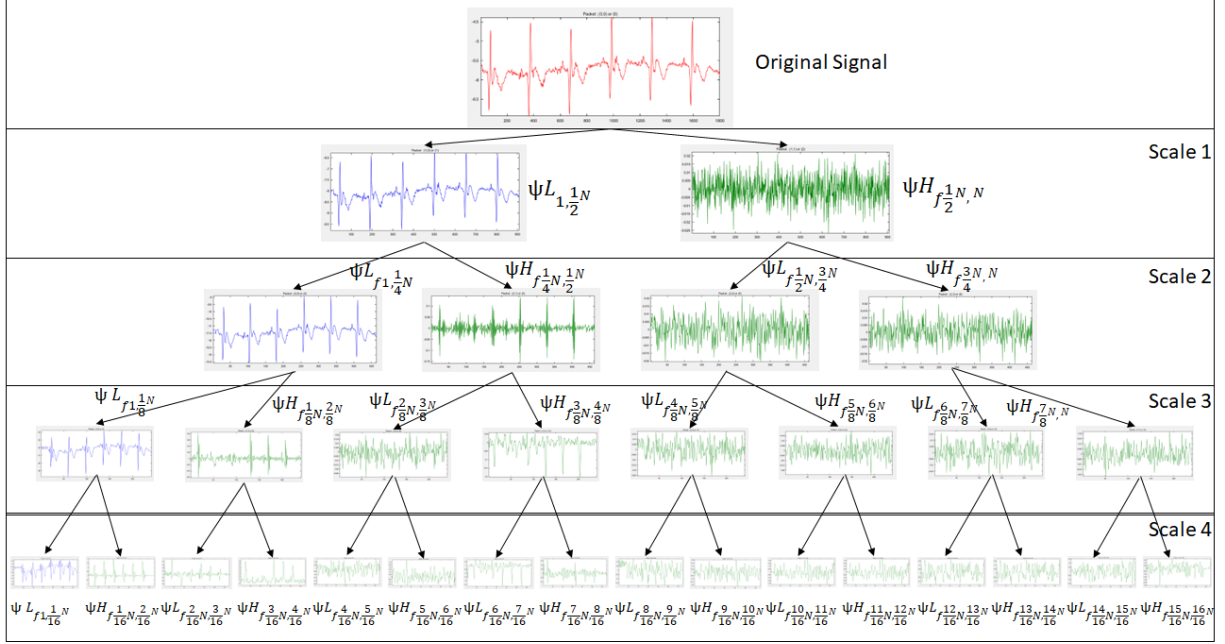


Figure 10: Wavelet Packet Decomposition Tree for ECG Signal

In the Ψ -GANDAE network, the wavelet transform is inherently integrated into the architecture by applying a high and low pass filter in tandem with each layer. The high and low pass content are then channel-wise concatenated to the output of the convolutional layer and it becomes a residual input to the decoder layer. The insight for this structure is that the network layers will then decompose the signal into time-frequency localized content bands to provide more distinction of frequency contributions and isolate common noise channels. The concept of the wavelet convolutional neural network was first introduced in 2018 by Fujieda et al.¹⁴⁴ as a method for extracting both spatial and spectral components from images. This network utilized high and low pass filters on each convolutional level to provide an inherent multiresolution decomposition. The wavelet convolutional neural network architecture demonstrated superior performance over pre-trained AlexNet and T-CNN models as applied to the kth-tips2-b and DTD texture databases with a classification accuracy of $74 \pm 1.2\%$ and $59.8 \pm 0.9\%$, respectively. Wavelet preprocessing of dermoscopy images was also demonstrated to improve the performance of U-Nets when used for lesion segmentation.¹⁴⁵

Deep residual networks (ResNets) were introduced by the Microsoft Research Group¹⁴⁶ as part of the ImageNet and COCO competitions in 2015, where they garnered first place in the categories of ImageNet detection, ImageNet localization, COCO detection, and COCO

segmentation. These ResNets were deeper than previous networks with up to 152 layers and offered a solution to the degradation problem that occurs in deep network structures, where a saturation and then rapid decrease in accuracy manifests during training. These degradations are not due to overfitting in the model, but rather are attributed to vanishing gradients. To address degradation problems in deep networks, ResNets utilize shortcuts where outputs from previous layers skip one or more stacked layers and bring content directly to deeper layers.

A special convolution-deconvolution architecture, referred to as U-Nets, was introduced as part of the ISBI cell tracking challenge in 2015.¹⁴⁷ These U-Nets address special challenges with medical image segmentation, such as limited numbers of labelled training and testing cases, and the importance of precision in boundary definition. The U-Net architecture contains a contracting path where image features are extracted and a subsequent expanding path that supports localization of the segmentation borders. The U-Nets were demonstrated as successful segmentation methods against a number of biomedical tasks, including segmentation of neuronal structures in electron microscope stacks, and cell segmentation in light microscopic images.

The Ψ -GANDAE architecture draws from key elements of both the U-Net and ResNet structures (but does not utilize the pre-trained U-Net model). Successive convolutional layers with skip connections and downsampling will be used through the decoder to produce a condensed feature vector. A stride of 2 on the 1-D convolutions was utilized to downsample the signal instead of a more common pooling approach, as it has been shown to be preferential for stability in GANs.¹⁴⁸ The decoder layer then utilizes transpose convolutional layers with upsampling to reconstruct the input signal without associated noise. Residual connections from the inputs to the output layers leverage on the observations that networks can be deeper and easier to train if the connections from the input to the output layers are shorter. A kernel size of 51 was used for the convolutional layers with a Parameterized ReLU (PReLU) activation function,¹¹¹ where a is an adaptive slope for $x < 0$, which helps to eliminate the “dying ReLU” issue when output is limited in the negative axis:

$$\text{PReLU}(x) = \begin{cases} ax & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \quad (3.6)$$

Generative Adversarial Networks have been demonstrated as a new state-of-the-art in many fields, including ECG signal de-noising.¹¹¹ The Ψ -GANDAE leverages a generator and discriminator architecture to advance model training and augment a traditional Mean Absolute Error (MAE) loss function. The GAN scaffolding is utilized during model training and the discriminator is discarded once complete. A byproduct of the process

is a discriminator that can evaluate an ECG segment and provide a determination if it has been sufficiently processed and noise removed. The generator loss function minimizes the mean absolute error between the reconstructed ECG signal $\hat{x}(n)$ and the ground truth ECG segment $x(n)$, in combination with minimizing the Discriminator's success for determining output produced by the Generator using a Least Squares Generative Adversarial Networks (LSGANs) loss.^{111,149} The LSGAN improves upon a Sigmoid Cross Entropy loss function by addressing vanishing gradients that can occur once a GAN output is on the correct side of a threshold, but requires further optimization:

$${}^G_{min} V_{LSGAN}(G) = \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - 1)^2] + \lambda * ||G(z) - x|| \quad (3.7)$$

Where x represents the samples of ground truth signal, taken from a set $p_x(x)$ of training data. And z sampled from a distribution of the noisy signal $p_z(z)$. A value of $\lambda = 100$ was utilized based on predicate research to balance the influence of MAE in the LSGAN.¹¹¹

The Discriminator utilizes the loss generated by the evaluation of the real and fake output for a representative LSGAN loss:¹¹¹

$$\begin{aligned} {}^D_{min} V_{LSGAN}(D) = & \frac{1}{2} \mathbb{E}_{x, p(x)} [(D(x) - 1)^2] \\ & + \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)))^2] \end{aligned} \quad (3.8)$$

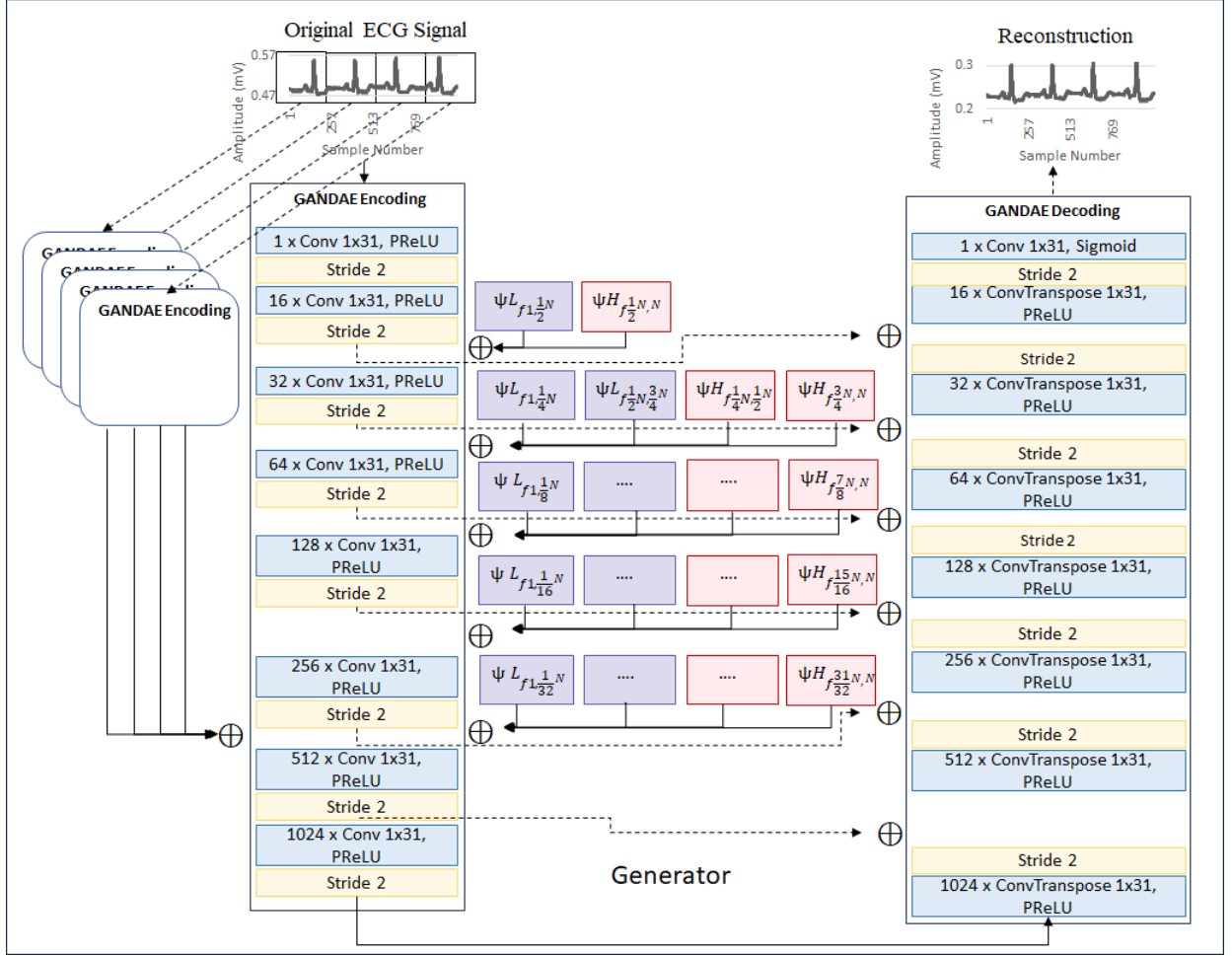


Figure 11: Generator Architecture of the Generative Multiscale Wavelet De-noising Autoencoder (Ψ-GANDAE)

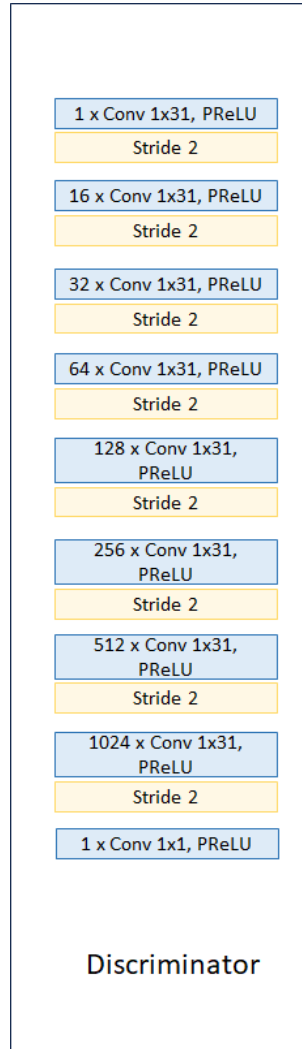


Figure 12: Discriminator Architecture of the Generative Multiscale Wavelet De-noising Autoencoder (Ψ -GANDAE)

3.2.3 Generative Multiscale Wavelet De-Noising Autoencoder for Electrocardiography Signal De-noising

3.2.3.1 Data

The PhysioNet MIT-BIH Arrhythmia Database contains 48 patient recordings of 30-minute duration captured using ECG ambulatory recorders at Boston’s Beth Israel Hospital (now the Beth Israel Deaconess Medical Center).¹³⁴ Each recording contains two channels representing a modified limb lead II and modified lead V1 configuration and are measured at a sampling frequency of 360 Hz with 11-bit resolution over a range of ± 10 millivolts. Two or more cardiologists independently annotated each of the recordings to identify 15 different ECG rhythms including: Atrial fibrillation, atrial flutter, normal sinus rhythm, and ventricular tachycardia.

The PhysioNet MIT-BIH Stress Test Database¹³⁵ contains 12 half-hour ECG recordings and 3 half-hour baseline noise recordings that represent predominant baseline wander (BW)(often caused by respiratory activity at 0-0.5 Hz), electrode motion artifact (EM)(~ 1 Hz), and EMG muscle artifact (MA)(resulting in 100-500 ms long distortions up to 10 kHz) for ECG signals.¹¹⁰ While these profiles all represent examples of additive noise, a deep learning strategy would also support de-noising of multiplicative noise profiles. An included script (nstdbgen) was used to generate noise at specified dB levels and added to the MIT-BIH Arrhythmia database¹³⁴ recordings. Noise was added after the first five minutes and then for two minutes, alternating with the original signal. The SNR during the noisy segments was set to 0 dB, 1.25 dB, and 5 dB.

To generate the training and testing data set, a 1024 sample (2.84 s) sliding, non-overlapping window was selected from the ECG signals. This segment duration was selected as a base of 2 to support wavelet decomposition and because the adult heart rate range is 30 bpm–200 bpm, resulting in a segment with 1-9 heartbeats. Lead II ECG channels were utilized exclusively for all patients, as this lead configuration is widely available in three-lead wearable ECG monitors. Ten patients were selected from the MIT-BIH Arrhythmia database¹³⁴ as identified in predicate research with representative arrhythmia rhythms, and then removed one at a time from the training set for leave-one-patient-out cross-validation (LOOCV). To match testing done by predicate research, k-fold cross validation was also conducted with an 80% training and 20% testing split for 5 different folds. For the LOOCV set 51,354 ECG segments were in the training set (40.58 hours) and 5,706 in the testing set (4.51 hours). Each fold contained 53,493 segments in the training set (42.27 hours) and 3,567 in the testing set (2.82 hours).

3.2.3.2 Results

The Ψ -GANDAE network was trained on NVIDIA GeForce GPUs for matrix acceleration. TensorFlow¹⁵⁰ and Keras¹⁵¹ were utilized to construct the network and facilitate the training and testing process, PyWavelets¹⁵² was used for wavelet decomposition, and Python¹⁵³ provided the framework for all workflow and implementation activities. A batch size of 32 was utilized to balance training time and network performance, and the networks were trained and evaluated over 300 epochs. A constant learning rate of 0.0001 was used with Root Mean Squared Propagation (RMSProp) for optimization.

Hyperparameters

Optimizer: Root Mean Squared Propagation (RMSprop)
 Learning Rate: 0.0001
 Epochs: 300
 Batch Size: 32
 Generator Loss Function: $V_{LSGAN}(G)$ (See Equation 3.7)
 Discriminator Loss Function: $V_{LSGAN}(D)$ (See Equation 3.8)

The performance metrics used to evaluate the network were the Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (\hat{x}_i - x_i)^2}, \quad (3.9)$$

and Signal-to-Noise-Ratio (SNR):

$$\text{SNR} = 10 \log \frac{\sum_{i=0}^{N-1} (x_i)^2}{\sum_{i=0}^{N-1} (\hat{x}_i - x_i)^2}. \quad (3.10)$$

The performance of the Ψ -GANDAE will be compared to predicate research, which include the S-Transform (ST),¹⁵⁴ an improved Wavelet Transform (WT),¹⁵⁵ an Improved Stacked Contractive De-Noising Autoencoder (I-DAE),¹⁰⁹ and most recently a de-noising GAN.¹¹¹ This predicate research presents model performance based on SNR and RMSE for ten pre-selected patients from the MIT-BIH Arrhythmia database¹³⁴ with defined noise signals from the MIT-BIH Stress Test Database.¹³⁵ All methods report results utilizing a k-fold cross validation approach. To compare the performance of the baseline GAN (BGAN) to the Ψ -GANDAE, this model was implemented as outlined by Singh et al.¹¹¹

and validated against the same training and testing data sets, using both k-fold and leave-one-out cross validation.

The testing results for the Ψ -GANDAE are presented in [Appendix B](#), [Table B.1](#), [Table B.2](#), and [Table B.3](#) for the k-fold cross-validation experiments. The performance metrics evaluated include both SNR and RMSE, and results are provided against the associated noise level of 0 dB, 1.25 dB, and 5 dB for each of the common ECG noise patterns (i.e., baseline wander, motion artifact, and electrode motion). The results presented for the ST, WT, and I-DAE represent performance as published in associated research by the original authors, and the BGAN and Ψ -GANDAE are constructed from experiments conducted with five randomly generated folds. The results in the tables for BGAN and Ψ -GANDAE represent the average of five folds. The columns in the table represent one of ten individual patients that were utilized in predicate research to evaluate performance, as these patients are representative of different cardiac arrhythmias (i.e., patient records 103, 105, etc.).

The performance of the BGAN and Ψ -GANDAE were also evaluated utilizing a leave-one-out cross-validation approach. The other predicate methods (i.e., ST, WT, I-DAE) did not have published results indicating performance using a LOOCV approach, therefore performance results presented are limited to the BGAN and Ψ -GANDAE models. Results for the LOOCV experiments are detailed in [Table 3.1](#), [Table 3.2](#), and [Table 3.3](#), for motion artifact, electrode motion, and baseline wander, respectively. For each of the LOOCV trials, one of the ten identified patients was sequentially removed from the training set and utilized as the testing set. Results presented are the performance of the model for each of the LOOCV trials on an individual patient. Performance metrics evaluated include SNR, RMSE, and an average for each metric, at each noise level.

3.2.3.3 Analysis

The results in [Appendix B](#), [Table B.1](#), [Table B.2](#), and [Table B.3](#) demonstrate the performance of the Ψ -GANDAE on k-fold cross validation as compared to predicate research, including performance of the state-of-the-art baseline GAN (BGAN). The Ψ -GANDAE produced an average across all noise levels of SNR=21.86 dB and RMSE=0.053 for BW, SNR=21.07 dB and RMSE=0.060 for EM, and SNR=21.27 dB and RMSE=0.062 for MA, with an overall performance of SNR=21.40 dB and RMSE=0.058 for all noise profiles. The results for the Ψ -GANDAE were consistent for BW, EM, and MA across all noise levels, which is ideal for a signal pre-processing stage for the *BiometricNet*. This compares with the overall performance metrics of the predicate approaches with BGAN demonstrating SNR=27.65 dB and RMSE=0.027, I-DAE with SNR=22.38 dB

Table 3.1: Results of Leave-One-Out Cross-Validation for Ψ -GANDAE on Muscle Artifact

Method	Level	Metric	103	105	111	116	122	205	213	219	223	230	Average
BGAN	0 dB	SNR	2.04	-1.77	4.13	7.79	5.49	1.28	0.77	6.39	7.03	0.05	3.32
		RMSE	0.346	0.540	0.197	0.499	0.507	0.326	0.653	0.397	0.310	0.421	0.420
	1.25 dB	SNR	2.21	-1.75	4.23	7.86	5.50	1.31	0.82	6.45	7.14	0.17	3.39
		RMSE	0.339	0.539	0.195	0.496	0.507	0.325	0.649	0.395	0.306	0.416	0.417
	5 dB	SNR	2.67	-1.74	4.46	7.99	5.51	1.39	0.92	6.57	7.38	0.39	3.55
		RMSE	0.321	0.539	0.190	0.490	0.507	0.322	0.641	0.390	0.298	0.408	0.411
GANDAE	0 dB	SNR	7.71	4.87	-0.04	6.46	5.05	3.49	4.34	6.07	8.50	-0.10	4.64
		RMSE	0.180	0.274	0.307	0.584	0.521	0.265	0.435	0.418	0.268	0.420	0.367
	1.25 dB	SNR	7.89	4.90	0.04	6.60	5.11	3.62	4.45	6.17	8.65	0.00	4.74
		RMSE	0.175	0.273	0.304	0.575	0.518	0.262	0.429	0.414	0.263	0.416	0.363
	5 dB	SNR	8.40	4.98	0.19	6.88	5.15	3.72	4.61	6.39	8.96	0.17	4.94
		RMSE	0.165	0.272	0.300	0.559	0.515	0.258	0.422	0.404	0.254	0.409	0.356

Table 3.2: Results of Leave-One-Out Cross-Validation for Ψ -GANDAE on Electrode Motion Artifact

Method	Level	Metric	103	105	111	116	122	205	213	219	223	230	Average
BGAN	0 dB	SNR	1.47	-1.50	3.56	7.49	5.04	0.82	0.56	5.92	6.65	-0.27	2.97
		RMSE	0.371	0.525	0.209	0.515	0.532	0.342	0.669	0.417	0.325	0.432	0.434
	1.25 dB	SNR	1.55	-1.51	3.65	7.58	5.06	0.83	0.62	6.00	6.76	-0.13	3.04
		RMSE	0.366	0.526	0.207	0.510	0.530	0.341	0.665	0.413	0.321	0.425	0.430
	5 dB	SNR	1.78	-1.58	3.95	7.81	5.14	0.88	0.77	6.17	7.04	0.14	3.21
		RMSE	0.356	0.530	0.199	0.499	0.526	0.338	0.653	0.405	0.310	0.413	0.423
GANDAE	0 dB	SNR	6.14	4.35	-2.09	6.20	4.61	2.73	3.99	5.49	7.80	-0.65	3.86
		RMSE	0.224	0.287	0.383	0.600	0.551	0.290	0.455	0.447	0.294	0.442	0.397
	1.25 dB	SNR	6.26	4.45	-2.07	6.26	4.70	2.78	4.07	5.58	7.96	-0.55	3.94
		RMSE	0.220	0.284	0.382	0.596	0.544	0.289	0.451	0.442	0.288	0.437	0.393
	5 dB	SNR	6.50	4.77	-2.07	6.49	4.92	2.74	4.33	5.81	8.26	-0.29	4.15
		RMSE	0.214	0.276	0.381	0.581	0.530	0.291	0.436	0.430	0.276	0.425	0.384

and RMSE=0.034, and SNR=17.23 dB and RMSE=0.125 for the WT. The Ψ -GANDAE demonstrated diminished performance as compared to the BGAN for BW, EM, MA at 0 dB, 1.25 dB, and 5 dB levels. The Ψ -GANDAE also showed a reduced overall average SNR differential of 0.98 dB and increased RMSE of 0.025 as compared to the I-DAE. However, when results are reviewed on an individual patient basis, the Ψ -GANDAE showed comparable alignment with the I-DAE for SNR, with improved performance over the I-DAE for 57% of the patients for BW, 50% of the patients for EM, and 57% of the patients for

Table 3.3: Results of Leave-One-Out Cross-Validation for Ψ -GANDAE on Baseline Wander

Method	Level	Metric	103	105	111	116	122	205	213	219	223	230	Average
BGAN	0 dB	SNR	2.74	-1.69	4.74	8.08	5.47	1.53	1.03	6.63	7.59	0.27	3.64
		RMSE	0.319	0.537	0.183	0.487	0.509	0.316	0.634	0.388	0.291	0.416	0.408
	1.25 dB	SNR	2.89	-1.69	4.80	8.06	5.47	1.57	1.05	6.65	7.64	0.36	3.68
		RMSE	0.314	0.537	0.182	0.488	0.508	0.315	0.633	0.387	0.289	0.413	0.407
	5 dB	SNR	3.25	-1.67	4.92	8.00	5.46	1.63	1.09	6.70	7.76	0.49	3.76
		RMSE	0.302	0.537	0.181	0.492	0.509	0.313	0.630	0.385	0.285	0.409	0.404
GANDAE	0 dB	SNR	8.48	4.89	0.12	7.03	5.20	3.75	4.65	6.51	9.15	0.14	4.99
		RMSE	0.163	0.275	0.304	0.552	0.512	0.256	0.420	0.398	0.248	0.411	0.354
	1.25 dB	SNR	8.68	4.99	0.10	7.10	5.20	3.81	4.70	6.55	9.24	0.19	5.06
		RMSE	0.160	0.272	0.305	0.548	0.512	0.255	0.417	0.397	0.246	0.409	0.352
	5 dB	SNR	9.04	5.13	-0.09	7.28	5.20	4.09	4.83	6.63	9.39	0.37	5.19
		RMSE	0.154	0.269	0.311	0.541	0.511	0.248	0.411	0.394	0.241	0.403	0.348

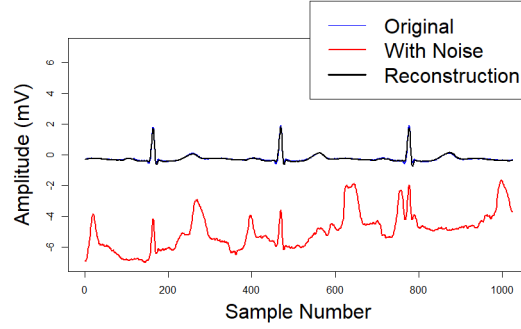
MA. The Ψ -GANDAE demonstrated better performance overall than the WT with an increased SNR of 4.17 dB and reduction of RMSE by 0.067, illustrating the benefit of hybrid deep learning methodologies that are integrated with traditional signal processing techniques such as the wavelet transform. It was also observed that the Ψ -GANDAE was removing inherent noise from the original MIT-BIH Arrhythmia database,¹³⁴ which would have penalized the SNR and RMSE values, but ultimately reflects positively on the abilities for signal pre-processing.

While the performance of the Baseline GAN surpassed the Ψ -GANDAE during k-fold cross-validation experiments, it is important to look at the performance for the methods on unseen cases to understand how results will generalize for a larger patient cohort. The results in Table 3.1, Table 3.2, and Table 3.3, present the performance of the BGAN and Ψ -GANDAE when evaluated utilizing a LOOCV approach. The overall average performance metrics demonstrated for the Ψ -GANDAE were SNR=4.61 dB and RMSE=0.368, and for the BGAN were SNR=3.40 dB and RMSE=0.417. It is demonstrated that the Ψ -GANDAE performance surpassed the BGAN for BW, EM, and MA, across all noise levels. The Ψ -GANDAE demonstrated the best performance for BW with average of all noise levels at SNR=5.08 dB and RMSE=0.351, followed by MA at SNR=4.77 dB and RMSE=0.362, and lastly EM at SNR=3.98 dB and RMSE=0.392. Performance was again consistent for the Ψ -GANDAE across all dB levels for each of the evaluated noise profiles.

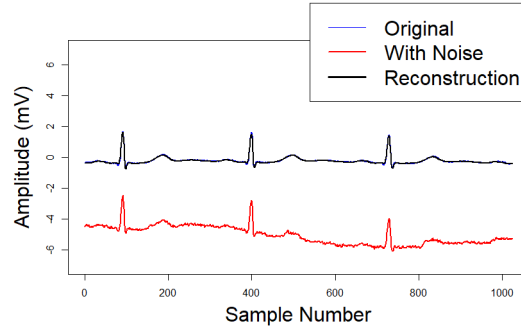
The Ψ -GANDAE will be performing signal pre-processing to prepare for successful signal segmentation and feature extraction in the second phase of processing for the

BiometricNet. It is the intention of this research to utilize the *BiometricNet* to detect and predict a wide range of adverse clinical events from data collected by clinical wearable devices. Thus, it is important that the Ψ -GANDAE can generalize for unseen patients that are not in the training set, and that it produce reliable and consistent signal reconstruction. The Ψ -GANDAE has demonstrated improved performance for ECG signal de-noising for baseline wander, electrode motion artifact, and motion artifact at 0 dB, 1.25 dB, and 5 dB noise levels on unseen cases as compared to the baseline GAN. It also demonstrates high levels of signal de-noising performance when it can benefit from samples of the patient’s ECG data. While the BGAN provided better performance when trained on a specific patient’s data, this is not a realistic requirement for use given regulations for machine learning software as a medical device. ECG is a very diverse physiological signal that is unique for every individual’s physiology and can be influenced by transient cardiac arrhythmias. It would require a large training dataset to sufficiently train the BGAN for suitable generalized performance, which is a challenge in healthcare given the scarcity of publicly available data and cost of large scale clinical trials. Therefore, the Ψ -GANDAE is a more suitable pre-processing model than the BGAN, given the demonstrated ability to provide essential performance on unseen cases.

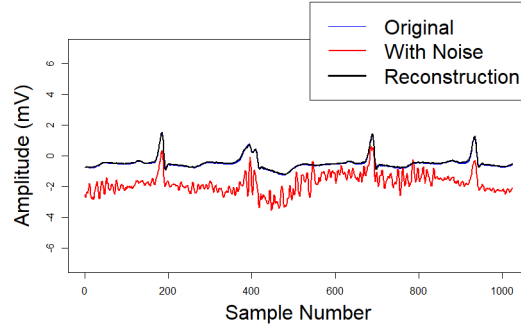
In [Figure 13](#), a number of samples of the ground truth ECG signal, along with reconstructed signal, from the Ψ -GANDAE are illustrated. These samples demonstrate the associated ECG tracings of the noisy signal with electrode motion, baseline wander, and motion artifact injected at 5 dB noise levels. These results were obtained from the Ψ -GANDAE trained through a k-fold cross validation approach. As a qualitative analysis, it is demonstrated that the reconstructed signal has an almost identical alignment with the ground truth signal. In [Figure 14](#), samples of EM, BW, and MA at 5 dB noise levels are again presented, but in this case the Ψ -GANDAE was trained utilizing a LOOCV approach. On visual inspection, it can be seen that the Ψ -GANDAE has still captured the intrinsic attributes of the QRS complex and beat-to-beat rhythm, while removing the muscle artifact, electrode motion, and correcting for baseline wander.



(a) Patient 103 with Electrode Motion Artifact at 5 dB

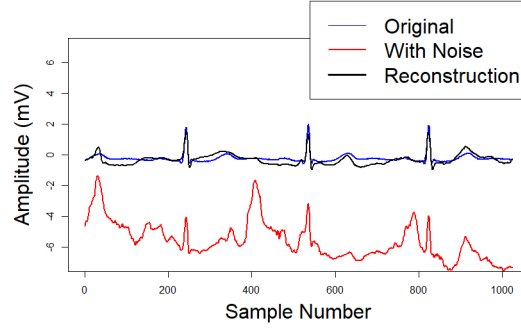


(b) Patient 103 with Baseline Wander Artifact at 5 dB

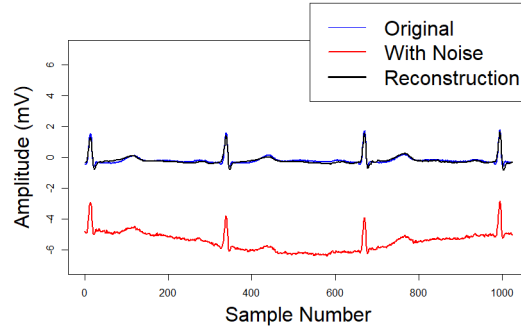


(c) Patient 223 with Motion Artifact at 5 dB

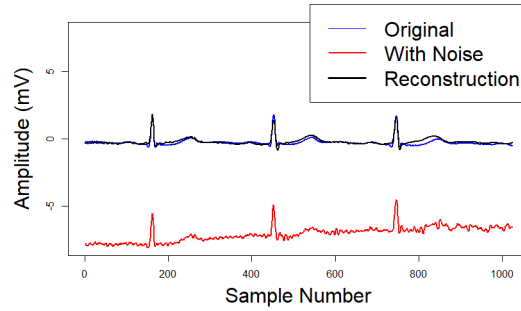
Figure 13: ECG Signals with Noise and Reconstructed Output from Generative Multiscale Wavelet Denoising Autoencoder (Ψ -GANDAE) k-Fold Cross-Validation



(a) Patient 103 with Electrode Motion Artifact at 5 dB



(b) Patient 103 with Baseline Wander Artifact at 5 dB



(c) Patient 223 with Motion Artifact at 5 dB

Figure 14: ECG Signals with Noise and Reconstructed Output from Generative Multiscale Wavelet Denoising Autoencoder (Ψ -GANDAE) Leave-One-Out Cross Validation

3.3 Phase 2: Feature Extraction and Signal Segmentation

3.3.1 Prologue

The content of this chapter was based in part on a conference paper published during the Ph.D. research:

S. Ross-Howe and H.R. Tizhoosh. *Atrial fibrillation detection using deep features and convolutional networks*. In IEEE International Conference on Biomedical and Health Informatics (BHI'19), May 19-22 2019.

3.3.2 Method

Biometric waveforms are non-stationary signals that contain frequency content that changes over time. Important information regarding the biometric signal can thus be represented in both the time and frequency domains. The proposed network for feature extraction and signal segmentation in the *BiometricNet* will utilize ResNet layers for feature extraction, bi-directional LSTM layers for sequence encoding, and attention to identify regions of semantic importance for segmentation. The architecture of the proposed ResNet, BiLSTM, and Attention Network (RBLAN) is depicted in [Figure 15](#).

Convolutional Neural Networks (CNN) have been clearly demonstrated as an effective architecture for extracting important features from common real-world images,^{[34,36–38](#)} specialized medical images,^{[145,147](#)} and biometric signals.^{[103–105,156,157](#)} In a CNN architecture, a bank of convolutional filters is first applied to the input matrix to produce multiple feature maps. These feature maps are then downsampled to further reduce the dimensionality of the inputs, and the average pooling also helps to promote generalization for unseen data. The convolution and average pooling steps are repeated and hyperparameter tuning done to optimize network performance and determine optimal topologies. Convolutional filters are able to extract both time and frequency features and have been successfully applied for deep feature extraction for biometric signal analysis.^{[103–105,156,157](#)} In [Figure 16](#), the application of successive convolutional filters and average pooling layers on an ECG segment is illustrated, resulting in an emphasis on elements of the beat morphology. A ResNet is a special variant of a CNN with skip connections into deeper layers of the network.^{[146](#)} The ResNet architecture mitigates the vanishing gradient problem caused by repeated application of the activation functions and allows for networks with an increased number of layers.

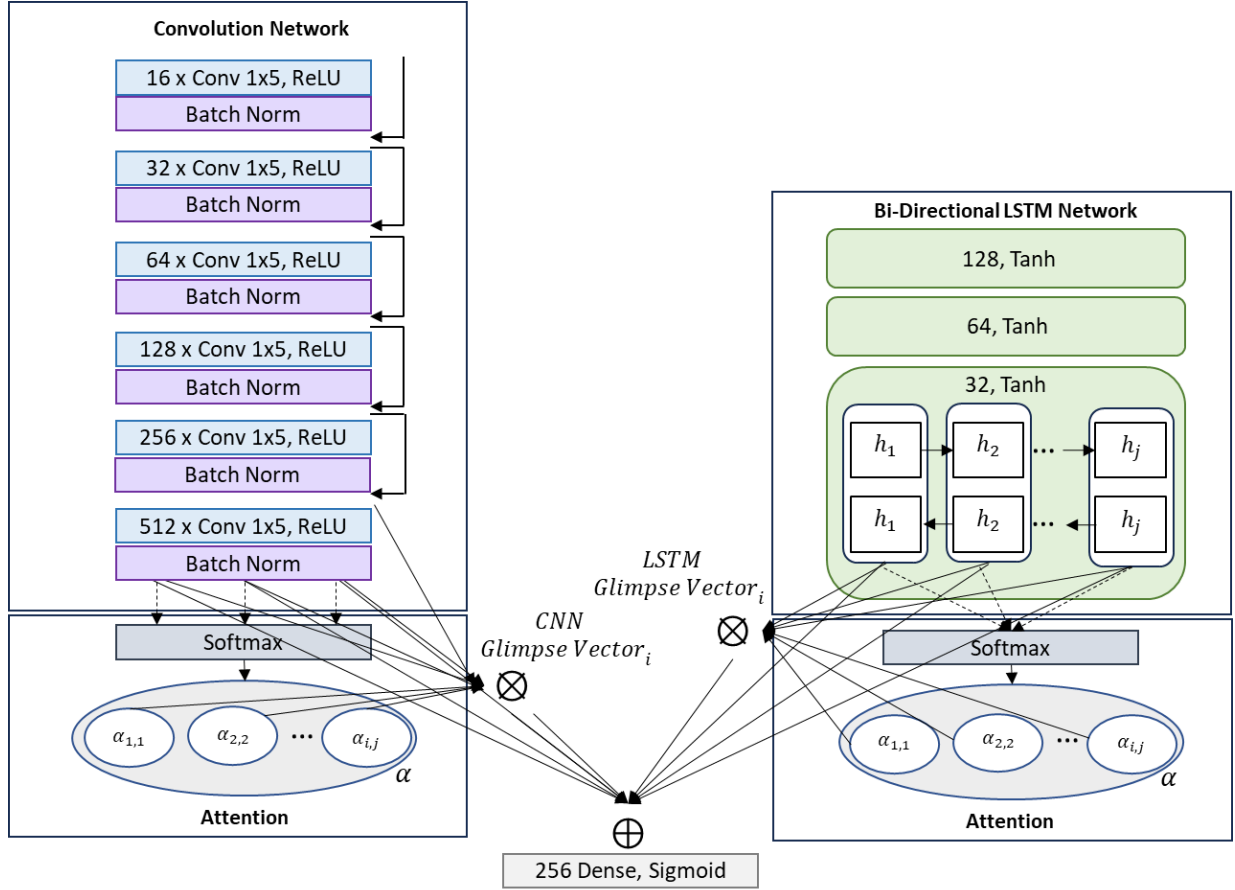


Figure 15: Convolutional Neural Network - Bi-Directional Long Short Term Memory Network with Attention Feature Extraction

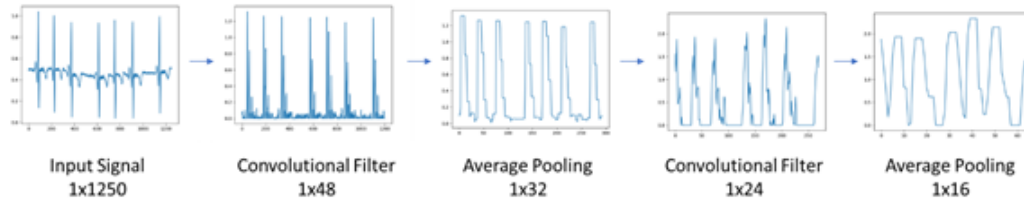


Figure 16: Application of Convolutional Filter and Average Pooling on ECG Data Segment

The RBLAN contains a dedicated sub-network with a ResNet architecture to extract temporal and frequency-based features from the pre-processed biometric signal. A 1-D

convolutional filter bank is arranged in successive layers to extract features in increasing levels of abstraction. A kernel size of 1x5 was determined empirically to be optimal for extracting features from biometric signals, and the filter size increases from 16 and doubles every layer to 512. Average pooling, or convolutional strides, were not included in the RBLAN, rather the system generates an increasing number of features in each layer that are then prioritized and reduced by attention layers and a dense output layer. The features extracted by RBLAN include those from the time domain, frequency domain, or the time-frequency domain. A ReLU activation function was utilized and bias employed on each layer. The ResNet architecture utilizes skip connections between each successive layer to improve system performance and to mitigate for vanishing gradients. Batch Normalization is utilized on the output of each CNN layer to accelerate network convergence and to support generalization on unseen cases.¹⁵⁸ The feature output from the ResNet is represented as c_i :

$$c_i = ReLU(\omega_i x + b) \quad (3.11)$$

Where ω_i are the weights, and b the bias.

$$ReLU(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \quad (3.12)$$

Recurrent neural network (RNN) architectures are well-suited to feature extraction from sequential and temporally related data.^{31,33,128} RNNs utilize a recurrent portion of the network as feedback loops that allow outputs from the previous step to be used as inputs for the next processing unit. Long-Short Term Memory models (LSTM) are a specialized version of the RNN, where each LSTM cell provides control to determine how many inputs to use from a previous cell, how many to store within the cell, and how many to provide as inputs to the next cell. LSTMs have been used successfully for natural language processing (NLP) tasks,³² speech recognition,³³ and biometric signal analysis.¹²⁹

Given that physiological signals are represented by a time series with inherent sequential dependencies, the RBLAN utilizes an LSTM sub-network to augment feature extraction from the ResNet sub-network. Bi-directional LSTM (BiLSTM) layers process in a forward and backward direction to determine temporal dependencies in either orientation. A stacked BiLSTM architecture is used to make this a deep learning methodology and improve the representation of the model. Each BiLSTM layer identifies features that can then be utilized in subsequent layers to build a higher level of abstraction, which extends the distance of dependencies that can be identified. The RBLAN uses three stacked BiLSTM

layers with 128, 64, and 32 units in each layer, respectively, and tanh as an activation function. The output of the network (h_t) is 2x32 features derived from the forward and backward processing of the BiLSTM.

Attention-based recurrent network architectures have been successfully applied to speech recognition,¹⁵⁹ machine translation,¹⁶⁰ and physiological signal processing.^{161,162} An attention network allows the output vector to be of a varying length and independent of the length of the input vector. In speech recognition applications for instance, the attention mechanism identifies the varying length phonemes that are contained within independent words. In the RBLAN model, attention networks can be utilized to identify regions of interest within a physiological signal, where outputs from feature extraction by the ResNet and BiLSTM sub-networks are subsequently passed through attention layers. This step aids in feature reduction within the biometric signal frame, but also across multimodal biometric channels. An attention-based approach integrates network interpretability into the *BiometricNet* and ultimately aids in supporting regulatory approval processes and clinical confidence in the network results. In the RBLAN, input samples $x \in \mathbb{R}$, are mapped to a feature vector by the encoder portion of these models, $h \in \mathbb{R}$ in the BiLSTM sub-network, and $c \in \mathbb{R}$ in the ResNet sub-network. The attention layer then produces an attention score by applying a dense layer with a tanh activation function on the feature output of the ResNet or BiLSTM:

$$s_i = \tanh(\omega c_i + b) \quad (3.13)$$

The output from the attention layer are passed through a softmax layer resulting in output probabilities:

$$\alpha_i = \frac{\exp(s_i)}{\sum_{i=1}^N \exp(s_i)} \quad (3.14)$$

The glimpse vector is the application of attention on the input features, represented as:

$$g_j = \sum_{i=1}^N \alpha_i h_j. \quad (3.15)$$

In this representation of the glimpse vector a deterministic soft attention mechanism has been employed. A deterministic hard attention was also investigated where:¹⁵⁹

$$p(t_i = 1 | t_{k < i}, h) = \alpha_i, \quad (3.16)$$

$$\hat{g}_j = \sum_{i=1}^N t_i h_j. \quad (3.17)$$

It was hypothesized that the hard attention mechanism would have the benefit of cropping beat regions from the signal to be passed into the feature extraction layer as a signal segmentation pathway. However, in experimentation it inhibited model convergence and the soft attention mechanism provided superior performance for the RBLAN.

Attention-based networks have been utilized to detect ECG abnormalities¹⁶¹ and atrial fibrillation.¹⁶² In one such application, called the ECGNET,¹⁶² a deep recurrent convolutional neural network (DRCNN) initially processes five second segments of the ECG signal and then translates this information into a wavelet power spectrum. A separate attention network is used to independently identify regions of interest in the ECG. The output of the attention and DRCNN are averaged together and then passed through a softmax layer to produce output probabilities for atrial fibrillation. This network achieved an impressive accuracy of 99.40%, sensitivity of 99.53%, and specificity of 99.26% on the detection of atrial fibrillation on the MIT-BIH Atrial Fibrillation database.

Hybrid models utilizing CNN, LSTM, and attentions layers have been increasingly utilized in recent signal processing research to capture both time and frequency features in waveforms, with attention reducing the feature space. Liu et al.¹⁶³ utilized a CNN-LSTM with attention to de-noise accelerometer and gyroscope signals from an IMU. This model utilized CNN layers to generate features that were passed through a dense layer into BiLSTM layers. The BiLSTM layers in turn produce features that are passed to the attention layer and a glimpse vector is generated. Features from the BiLSTM and the glimpse vector are concatenated together and passed through a final dense layer. Xu et al.¹⁶⁴ employed a CNN-LSTM for seizure detection in EEG signals. This model also utilized CNN layers to generate features, which are then passed through a dense layer into the LSTM layers. From the LSTM layers, the features are processed through a final dense layer and softmax is used to generate a probabilistic classification.

In the RBLAN, the features c_i output from the ResNet are passed through a flatten layer and concatenated with the glimpse vector g_j generated from the ResNet attention layer. The BiLSTM processes the signal inputs in parallel to the ResNet and produces a feature vector based on output states of h_k . The BiLSTM output features are then also passed through an attention layer and a glimpse vector g_k generated. The BiLSTM output features are flattened and concatenated with the glimpse vector g_k . The concatenated output from both the ResNet and BiLSTM are then concatenated together to form a comprehensive set of output features f_l . These features are passed through a dense output layer with a sigmoid activation function. In the final dense layer the individual features from the CNN and BiLSTM sub-networks are prioritized along with the glimpse vectors. The glimpse vectors identify important features from both sub-networks and reduce the feature space between multimodal signal channels. The RBLAN is trained independently for each physiological

signal, and the features generated for all signal channels are processed utilizing a dense layer with a linear activation function for regression. For classification tasks, a Siamese Neural Network architecture comprised of RBLAN twin networks is substituted in the final phase of the *BiometricNet* (see Section 3.4).

3.3.3 Continuous Blood Pressure Estimation with Deep Features and Attention

The World Health Organization estimates that hypertension causes 7.5 million deaths annually and accounts for almost 12.8% of total global deaths.¹⁶⁵ The prevalence of hypertension in adults (aged 30-79 years) worldwide is estimated at 32% in women and 34% in men, accounting for nearly 1.28 billion people in 2019.¹⁶⁶ The definition provided by the Public Health Agency of Canada to diagnosis hypertension is a systolic rate (pressure during contraction) of ≥ 140 mmHg and a diastolic rate (pressure during heart relaxation) of ≥ 90 mmHg.¹⁶⁷ Hypertension can lead to coronary artery disease, heart failure, transient ischemic attacks (TIA), stroke, dementia, kidney failure, and eye blood vessel damage, amongst other complications.¹⁶⁸ The gold standard for blood pressure measurement in a clinical setting involves an auscultatory method used in conjunction with a sphygmomanometer. The clinical professional applies a blood pressure cuff (usually on the brachial artery) and inflates past the systolic maximum threshold. A stethoscope is used to listen for Korotkoff sounds as the cuff is deflated, which signal the transition points for systolic and diastolic levels. The oscillometric technique is an alternative blood pressure measurement method that uses an electronic pressure gauge to quantify the pressure oscillations during cuff deflation. Systolic and diastolic measures are then extracted using empirically derived algorithms based on the shape of the recorded pulse wave envelope.⁶ The oscillometric technique has been proven to be comparable to the auscultatory method, and intra-arterial catheter pressures, in studies of commercially available devices.⁷

Challenges with the gold standard auscultatory and oscillometric approaches for blood pressure measurement include operator biases and inconsistencies, as well as being limited to intermittent and discrete measurements.¹⁶⁹ Ambulatory blood pressure measurement technologies have been evaluated that allow patients to take multiple timed readings throughout their daily activities.¹⁷⁰ Numerous studies have shown that these ambulatory measurements are more reliable predictors of cardiovascular disease risk than individual readings taken in a clinical environment.^{171,172} These ambulatory techniques still require compression of the arteries and subsequent discomfort to the patient and are often still discrete measurements. But with the increasing adoption of wearable technologies, the

opportunity now also exists to record critical biometric waveforms in a continuous and non-invasive manner. It would be highly advantageous if these continuous biometric signals could be translated to clinical grade blood pressure estimates captured passively throughout daily activity.

The foundations of work in continuous blood pressure estimation are based on the principals of biomechanics and specifically the Moens-Korteweg formula,¹⁷³ which establishes the relationship between the pulse wave velocity and characteristics of the neighbouring vessel walls:

$$v = \sqrt{\frac{gEa}{\rho d}} \quad (3.18)$$

Where v is the pulse wave velocity, g is the gravitational constant, E is the elastic modulus of the vessel, a is the wall thickness, ρ is the density of blood, and d is the interior diameter of the vessel.

In early work by Chen et al.,¹⁷⁴ the Moens-Korteweg equation was combined with the derivation by Hughes et al.,¹⁷⁵ which establishes the relationship between the elastic modulus and blood pressure:

$$E = E_0 e^{\gamma \rho} \quad (3.19)$$

Where E_0 is the elastic modulus at zero pressure, ρ is the blood pressure (mmHg) and γ is an empirically derived coefficient ($mmHg^{-1}$). With the substitution of elastic modulus into the Moens-Korteweg formula, the systolic blood pressure (SBP) can be defined by:

$$SBP = SBP_0 - \frac{2}{\gamma PTT_0} (PTT - PTT_0) \quad (3.20)$$

This method requires a calibration step to establish constants for the baseline systolic rate (SBP_0) and pulse transit time (PTT_0), and the value of the constant coefficient γ is selected based on the type of vessel being measured. The PTT is typically measured through the pulse wave propagation time from the ECG R-Peak to the first derivation peak in the PPG at a peripheral location on the body (e.g. finger, toe, ear, or forehead). Chen et al.¹⁷⁴ reported an error rate of $\pm 10\%$ over 97.8% of the reporting period and 38.8% with zero error, when utilizing γ derived from data collected from 20 patients:

$$\gamma = 0.016 - 0.018 \quad (mmHg^{-1}) \quad (3.21)$$

Extensions to this work were continued by Poon et al.¹⁷⁶ with the addition of estimations for diastolic blood pressures. Diastolic and systolic measurements are related based on the heuristic relationship to mean blood pressure (MBP) as outlined as:

$$MBP = \frac{1}{3}SBP + \frac{2}{3}DBP \quad (3.22)$$

When this heuristic is applied with an DBP substitution for SBP in equation 3.20, the DPB and SBP can be expressed by equations:¹⁷⁷

$$DBP = \frac{1}{3}SBP_0 + \frac{2}{3}DBP_0 + \frac{2}{\gamma} \ln \left(\frac{PTT_0}{PTT} \right) - \frac{SBP_0 - DBP_0}{3} \left(\frac{PTT_0}{PTT} \right)^2 \quad (3.23)$$

$$SBP = DBP + (SBP_0 - DBP_0) \left(\frac{PTT_0}{PTT} \right)^2 \quad (3.24)$$

A calibration step is again required to establish the baseline systolic rate (SBP_0), diastolic rate (DBP_0), pulse transit time (PTT_0), and the constant coefficient was revised based on empirical results by the investigators.¹⁷⁷

$$\gamma = 0.031 \quad (mmHg^{-1}) \quad (3.25)$$

In testing against 10 healthy patient volunteers, Poon et al.¹⁷⁶ reported overall accuracy rates for SBP of 0.6 ± 9.8 mmHg and for DBP of 0.9 ± 5.6 mmHg with this method.

The most recent extension to this method was conducted by Ding et al.¹⁷⁸ with the introduction of the PPG intensity ratio (PIR) as a new correlating attribute for DBP estimations. The PIR relates to changes in the arterial diameter and thus traces the low frequency variation in BP. The investigators combined the PIR and PTT models for a composite approach to model both the high and low frequency contributions to BP variation:

$$DBP = DBP_0 \frac{PIR_0}{PIR} \quad (3.26)$$

$$SBP = DBP_0 \frac{PIR_0}{PIR} + PP_0 \left(\frac{PTT_0}{PTT} \right)^2 \quad (3.27)$$

This method also requires a calibration step to establish the baseline diastolic rate (DBP_0), PPG intensity ratio (PIR_0), pulse transit time (PTT_0), and pulse pressure (PP_0). Testing was conducted with 27 healthy adults fitted with a finger cuff on the left thumb, brachial cuff on the left upper arm, ECG electrodes on left and right arms, and PPG sensor on the left index finger. Reported accuracy rates of -0.37 ± 5.21 mmHg for SBP, 0.08 ± 4.06 mmHg for DBP, and -0.18 ± 4.13 mmHg for mean BP were provided for this research.¹⁷⁸

While these PTT-based estimation methods provide accurate estimations of both SBP and DBP, they require an initial calibration step with a BP measurement device, and both ECG and PPG baseline signals against which to measure PTT. The calibration interval is also limited with measurable degradation in accuracy after a 24-hour period.¹⁷⁹ Recent research initiatives have focused on BP estimation methods that eliminate the requirement

for a calibration process or are dependant on analysis of the PPG waveforms alone. In Miao et al.,¹⁸⁰ characteristics of both the ECG and PPG waveform were extracted as features to be used as inputs into a multivariate linear regression and support vector regression models for estimating both SBP and DBP without the need for calibration. SBP variables that were identified with high feature importance included: PTT (from R-Peak to the maximum of the first derivative of the PPG), PPG characteristic value PPG_K , and PIR:

$$FI_i = \frac{\sum_{n=1}^N w_{i,n}}{N} \quad (3.28)$$

Where N is the number of patients in the experiment and $w_{i,n}$ is the weight of feature i for the n th patient. DBP variables that were identified with a high feature importance include: PIR, time width of the second derivate of the PPG waveform, foot intensity of the second derivate PPG waveform, and PPG_K :

$$PPG_k = \frac{p_m - p_d}{p_s - p_d} \quad (3.29)$$

Where p_m is the PPG amplitude at the mean blood pressure (MBP), p_d is the PPG amplitude at the DBP, and p_s is the PPG amplitude at the SBP. Testing to validate the method was conducted on 73 healthy adults and a reported overall accuracy of -0.001 ± 3.102 mmHg for SBP and -0.004 ± 2.199 mmHg for DBP was obtained.

Work by Xing et al.¹⁸¹ focused on processing the PPG waveform to estimate both SBP and DBP, without the use of accompanying ECG. The PPG signal was processed with an FFT and amplitude and phase features were extracted as input features for an Artificial Neural Network (ANN). The ANN was structured with a single hidden layer and two output nodes for SBP and DBP. Output from the trained network was validated against data from 69 patients in the MIMIC II waveform database and supplemented with data collected from 23 additional volunteers. The reported accuracy for this method was 0.06 ± 5.57 mmHg for SBP and 0.01 ± 3.69 mmHg for DBP. Most recently, Kyung et al.¹⁸² employed a CNN with attention for continuous blood pressure estimates from PPG and pressure sensor signals. The CNN network with attention achieved a 0.43 ± 9.35 mmHg for SBP and 0.21 ± 7.72 mmHg for DBP estimations.

Methods reviewed utilized a number of reference points on the PPG pulse for calculating the PTT in relation to the ECG R-peak. In Figure 17, the relationship between the ECG signal and the PPG signal landmarks is demonstrated. The PTT_f relates to the foot of the PPG pulse, the PTT_d to the maximum point of the first derivative, and the PTT_p to the pulse peak.

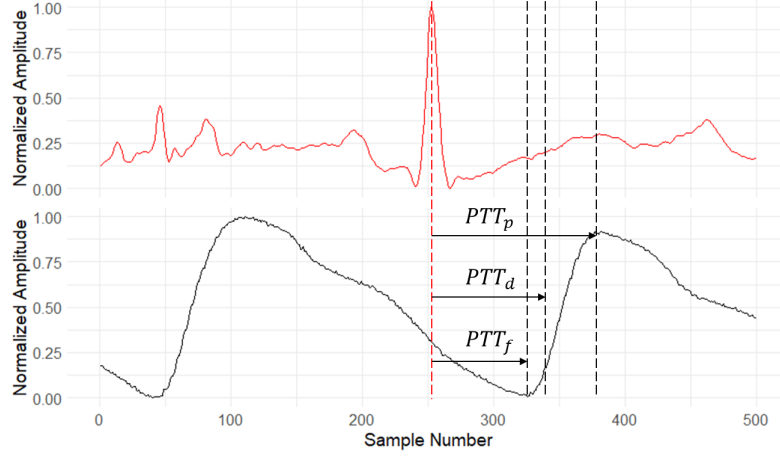


Figure 17: Pulse Transit Time Measurements

3.3.3.1 Data

The Pulse Transit Time (PTT) PPG Dataset is a recent addition to the PhysioNet clinical repository.¹³⁶ It contains data from 22 healthy subjects captured during 3 different physical activities, namely: sitting, walking, and running. Multiple sensors were utilized to capture time synchronized channels from different locations on the participant’s body. Signals include ECG, PPG (red $\lambda = 660 \pm 10$ nm, IR $\lambda = 880 + 2010$ nm, and Green $\lambda = 537 + 47$ nm), tri-axial accelerometer, and tri-axial gyroscope. All the sensors are connected through an ARM Cortex-M4 microcontroller running at 180 MHz, to allow the signals to be measured within a 2 ms window. This tight time resolution allows the experimentation on PTT algorithms related to ECG and PPG signal analysis. The blood pressure reference measures were taken with an OMRON HEM-7322 monitor, with an oscillometric reading taken at the start of the test session and again at the end. All the waveforms were captured at a 500 Hz sampling frequency. Temperature and peripheral pulse oxygen saturation levels (SpO2) are also provided. A 500 sample tracing of the physiological signals present in the dataset is illustrated in Figure 18, which include: ECG, PPG-red wavelength, PPG-IR wavelength, PPG-green wavelength, and tri-axial accelerometer and gyroscope.

The average PTT time for a finger tip-mounted sensor was determined in experimentation to be 269 ± 6 ms.¹⁸³ To ensure a data segment was extracted that would contain the peak of the PPG signal (and thus relate the PTT), a one second analysis window was selected. This window centers around the ECG R-peak, which is provided in the

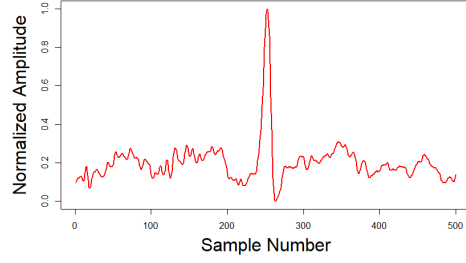
annotation track of the PhysioNet Pulse Transit Time (PTT) PPG Dataset.¹³⁶ The ECG signal and the PPG IR channel were extracted and separated into five training and testing folds with an 80%-20% split. The IR channel was selected amongst the PPG channels for processing, as it has the longest wavelength and thus the ability to penetrate deeper into the vessel bed for a higher quality signal. This is due to the transmittance-based orientation of the LEDs and the photo diodes in the finger-mounted optical sensor. The ECG and PPG signals were pre-processed with a rolling average mean filter and a Butterworth bandpass filter of 0.05 Hz-150 Hz for the ECG and 0.75 Hz-5 Hz for the PPG signal. All data segments were normalized to a range of 0 to 1 from local minimum and maximum values. First and second order derivatives were computed for the ECG and PPG signal from the normalized segment. Two reference BP measurements are made on each reading, at the beginning of the test session, and again at the end. A linear interpolation was used to estimate the reference BP for each time point between the beginning and the end reference measurements.

3.3.3.2 Results

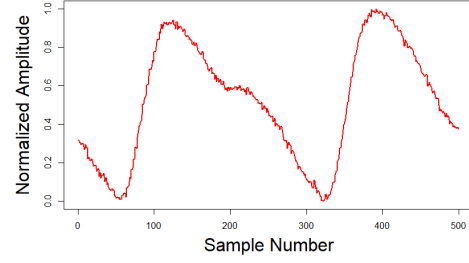
The RBLAN network was trained on NVIDIA GeForce GPUs for matrix acceleration. TensorFlow¹⁵⁰ and Keras¹⁵¹ were utilized to implement the network and facilitate the training and testing process, SciPy¹⁸⁴ was used for signal pre-processing methods, and Python¹⁵³ provided the programming framework for all implementation activities. A batch size of 32 was utilized and the networks were trained and evaluated over 100 epochs. The model was trained separately for diastolic and systolic estimations. A constant learning rate of 0.0001 was used with Adaptive Moment Estimation (Adam) for the optimizer. Mean Squared Error (MSE) was used for the loss function:

$$MSE = \frac{1}{N} \sum_{i=0}^{N-1} (\hat{BP}_i - BP_i)^2 \quad (3.30)$$

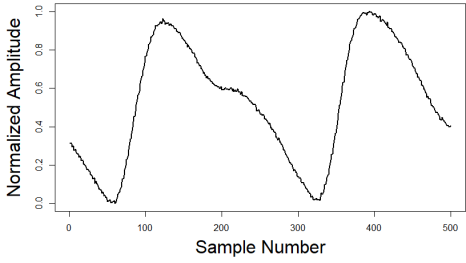
Where $\hat{BP}_i$ is the systolic or diastolic estimate and BP_i is the reference measurement.



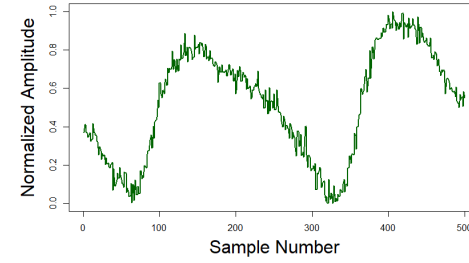
(a) ECG Signal



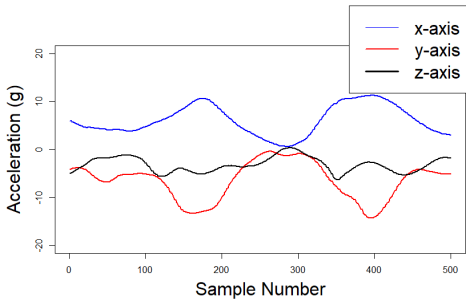
(b) PPG Signal (Red)



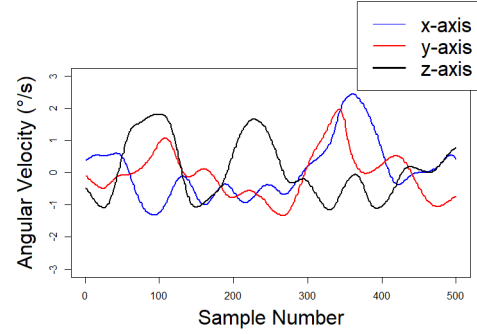
(c) PPG Signal (IR)



(d) PPG Signal (Green)



(e) Accelerometer Signals



(f) Gyrometer Signals

Figure 18: Signal Segments from Pulse Transit Time PPG Dataset

Hyperparameters

Optimizer: Adaptive Moment Estimation optimizer (Adam)

Learning Rate: 0.0001

Epochs: 100

Batch Size: 32

Loss Function: Mean Squared Error (See Equation 3.30)

The performance of the RBLAN model were assessed against the ability to estimate both SBP and DBP within the tolerances of established international standards for medical devices. The International Standards Organization (ISO) 81060-2:2018/Amd 1:2020 “Non-invasive sphygmomanometers — Part 2: Clinical investigation of intermittent automated measurement type” requires blood pressure estimation to produce mean absolute error levels $<5 \pm 8$ mmHg. The British Hypertension Society (BHS) defines device quality gradings based on the percentage of readings with error within 5, 10, and 15 mmHg. To the author’s knowledge, there are no publications to report BP estimation accuracy from the PhysioNet Pulse Transit Time PPG Dataset at the time this was published.

The results for the validation of the RBLAN model for continuous, non-invasive blood pressure estimation are presented in [Table 3.4](#) and [Table 3.5](#). The results detail performance on each of the five folds used for cross validation on random testing sets, with diastolic estimates presented in [Table 3.4](#) and systolic estimates presented in [Table 3.5](#). Metrics provided in these tables include the count of samples in each test fold, the mean and standard deviation of error in mmHg, as well as the absolute mean and standard deviation of error in mmHg. Predicate research often reports mean and standard deviation of error, but ISO 81060-2:2018/Amd 1:2020 requires error to be reported with absolute mean and standard deviation of error, so both have been evaluated. The percentage of readings with an error that is < 5 mmHg, < 10 mmHg, and < 15 mmHg are also defined, which allows the performance of the RBLAN to be evaluated based on the British Hypertension Society’s grading categories. These metrics are also segregated based on the activity performed by the participant during data collection, which includes sitting, walking, or running.

The individual paired reading errors between reference measurements and RBLAN estimations are presented in Bland-Altman plots in [Figure 19](#) and [Figure 20](#) for diastolic and systolic measurements, respectively. The Bland-Altman plots show limits of agreement, and can be utilized to identify systemic bias in the model and for outlier detection. Scatter plots have also been generated with detailed Pearson’s correlation coefficients to show correlation between reference and estimated systolic and diastolic measurements. The scatter plots for diastolic measurements can be found in [Figure 21](#) and for systolic measurements in [Figure 22](#). All Bland-Altman and scatter plots report results independently for each of the three collection positions (i.e., sitting, walking, and running).

3.3.3.3 Analysis

The RBLAN model was tested for performance of continuous, non-invasive blood pressure estimations against data that was collected in both sedentary (i.e., sitting), as well as

Table 3.4: Performance of the RBLAN for Diastolic Blood Pressure Estimates

Position	Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
Sitting	Count (N)	2718	2817	2701	2792	2694	2744.40
	Mean (SD)	0.95(2.20)	1.69(2.31)	1.53(2.34)	2.11(2.53)	2.67(3.03)	1.79(2.48)
	Mean Absolute (SD)	1.85(1.52)	2.28(1.74)	2.22(1.71)	2.63(1.99)	3.18(2.49)	2.43(1.89)
	BHS 5(%)	96.14	93.89	93.00	87.71	79.14	89.98
	BHS 10(%)	99.85	99.57	99.78	99.71	98.14	99.41
	BHS 15(%)	100.00	99.96	99.96	100.00	99.89	99.96
Walking	Count (N)	3232	3121	3253	3136	3166	3181.60
	Mean (SD)	1.40(2.97)	1.98(3.18)	1.63(3.25)	2.31(3.42)	2.82(3.64)	2.03(3.29)
	Mean Absolute (SD)	2.49(2.14)	2.87(2.41)	2.79(2.32)	3.25(2.53)	3.67(2.78)	3.02(2.44)
	BHS 5(%)	88.89	85.00	85.95	79.94	73.72	82.70
	BHS 10(%)	99.04	98.11	98.56	97.70	96.87	98.06
	BHS 15(%)	99.91	99.94	99.82	99.87	99.84	99.87
Running	Count (N)	3293	3176	3367	3236	3287	3271.80
	Mean (SD)	1.35(3.22)	1.76(3.51)	1.76(3.40)	2.29(3.54)	2.40(3.90)	1.91(3.51)
	Mean Absolute (SD)	2.61(2.31)	2.94(2.60)	2.95(2.44)	3.27(2.66)	3.48(2.98)	3.05(2.60)
	BHS 5(%)	86.88	83.41	84.14	78.52	77.55	82.10
	BHS 10(%)	98.48	97.32	97.95	97.40	95.71	97.37
	BHS 15(%)	99.91	99.81	99.88	99.75	99.33	99.74

Table 3.5: Performance of the RBLAN for Systolic Blood Pressure Estimates

Position	Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
Sitting	Count (N)	2718	2817	2701	2792	2694	2744.40
	Mean (SD)	0.90(3.77)	1.09(4.29)	1.38(3.88)	1.50(3.79)	1.91(4.35)	1.36(4.01)
	Mean Absolute (SD)	2.86(2.61)	3.25(3.00)	2.91(2.92)	2.95(2.81)	3.35(3.36)	3.06(2.94)
	BHS 5(%)	84.29	79.59	83.30	82.31	79.84	81.87
	BHS 10(%)	98.20	97.23	97.11	96.99	95.95	97.10
	BHS 15(%)	99.56	99.43	99.63	99.50	99.00	99.42
Walking	Count (N)	3232	3121	3253	3136	3166	3181.60
	Mean (SD)	1.18(4.42)	0.92(5.40)	1.32(4.53)	1.61(4.84)	1.40(4.80)	1.29(4.80)
	Mean Absolute (SD)	3.33(3.13)	4.15(3.58)	3.54(3.12)	3.78(3.43)	3.73(3.33)	3.71(3.32)
	BHS 5(%)	79.15	67.86	76.58	72.58	73.09	73.85
	BHS 10(%)	96.57	93.98	96.00	95.03	94.98	95.31
	BHS 15(%)	98.79	98.40	99.20	98.41	98.86	98.73
Running	Count (N)	3293	3176	3367	3236	3287	3271.80
	Mean (SD)	1.57(4.75)	1.41(5.30)	2.02(4.72)	2.13(4.90)	1.92(4.93)	1.81(4.92)
	Mean Absolute (SD)	3.57(3.50)	4.07(3.68)	3.78(3.47)	3.89(3.66)	3.86(3.62)	3.83(3.59)
	BHS 5(%)	75.46	69.52	72.82	71.97	72.77	72.51
	BHS 10(%)	94.72	94.02	94.24	94.00	93.52	94.10
	BHS 15(%)	98.60	98.58	98.72	98.45	98.78	98.63

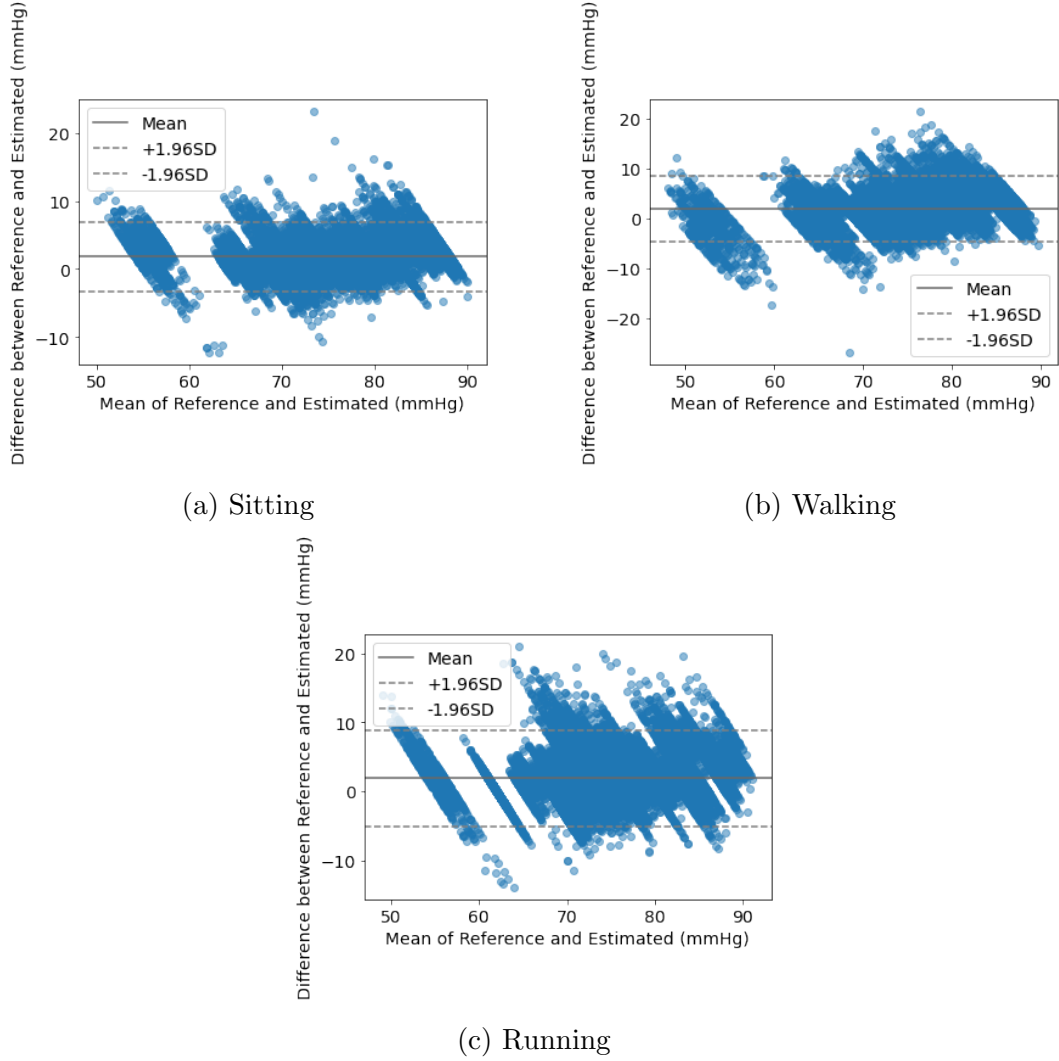


Figure 19: Bland-Altman Plots for Difference in Diastolic Estimates to Reference Values

ambulatory conditions (i.e., walking and running). Details of the errors of determination between the reference oscillometric measurements and the RBLAN model estimations are detailed in [Table 3.4](#) and [Table 3.5](#). The RBLAN demonstrated an overall mean error on all positions of 1.91 ± 3.10 mmHg and absolute error of 2.83 ± 2.31 mmHg for diastolic estimations, and mean error of 1.48 ± 4.58 mmHg with mean absolute error of 3.53 ± 3.28 mmHg for systolic estimations. As anticipated, the position impacted the accuracy of estimations for both systolic and diastolic estimates, with sitting demonstrating the lowest

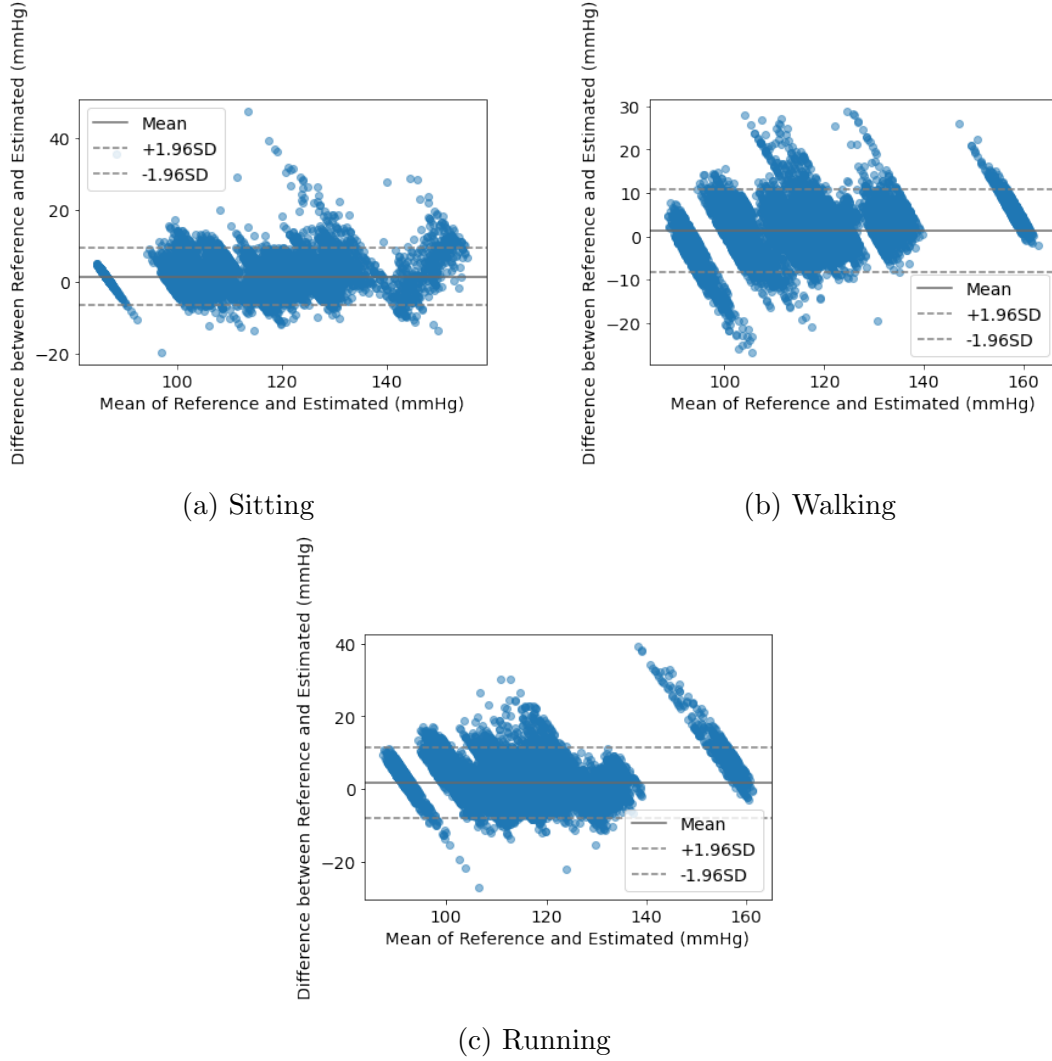


Figure 20: Bland-Altman Plots for Difference in Systolic Estimates to Reference Values

error, then walking, and running produced the highest errors. The RBLAN met the ISO 81060-2:2018/Amd 1:2020 device requirements of $<5 \pm 8$ mmHg for mean absolute error for both diastolic and systolic measurements, across all folds, and subject positions. This is an exciting result, as many previous validation studies have been restricted to evaluating performance in sedentary positions when the patient was seated or supine in a hospital bed,^{181,182} and it lends confidence to the regulatory approval pathways for ambulatory cNBIP measurements. It is also notable that no calibration based on individual

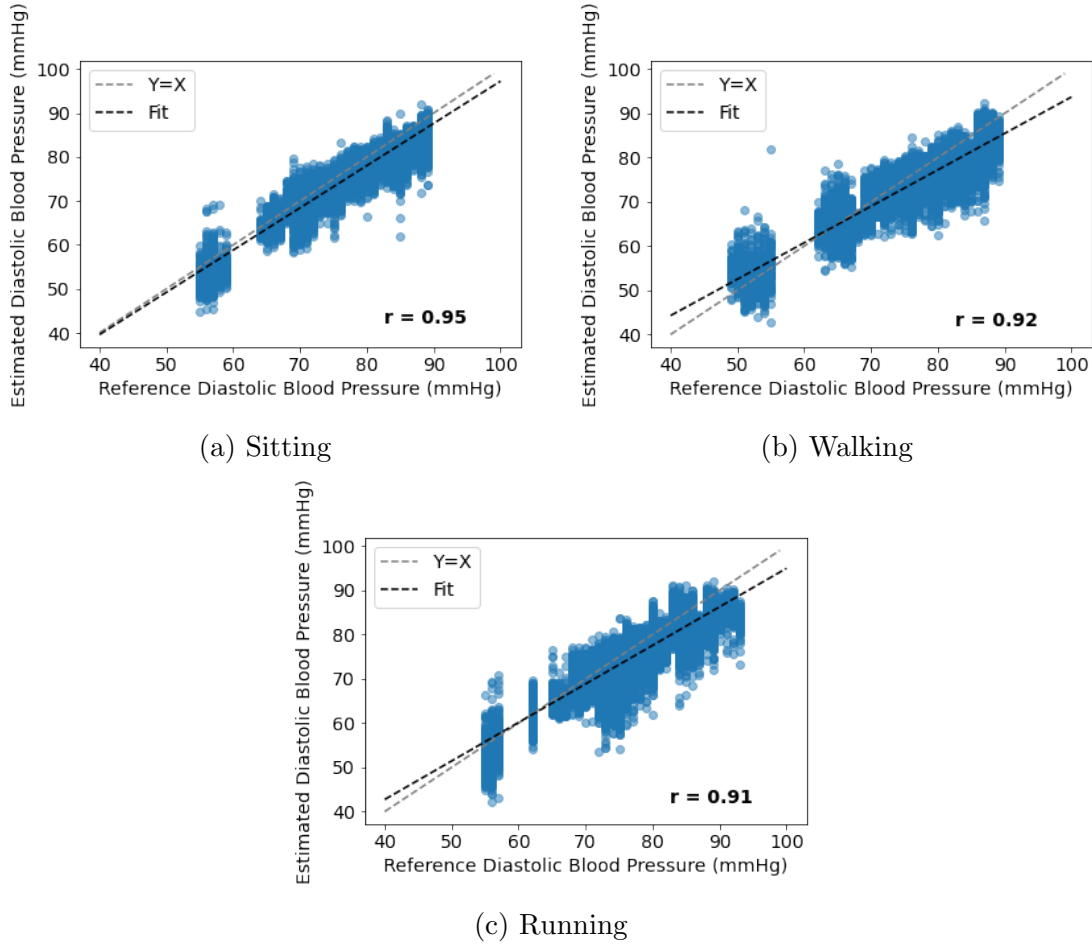


Figure 21: Scatter Plots for Diastolic Estimates

patient data was required for the RBLAN, as compared to many of the early cNBIP methods.^{174,176,178}

The performance of the RBLAN was also evaluated against the British Hypertension Society (BHS) thresholds for the percentage of measurement errors <5 mmHg, <10 mmHg, and <15 mmHg. The RBLAN performs at the top tier designation (Grade A) for BHS standards, which requires 60% of readings to have measurements with errors <5 mmHg, 85% to be <10 mmHg, and 95% to be <15 mmHg.¹⁷⁰ It is difficult to compare the results in a meaningful manner to other predicate research, given the disparity in data sets and use of data collection methods predominately limited to static positions. However,

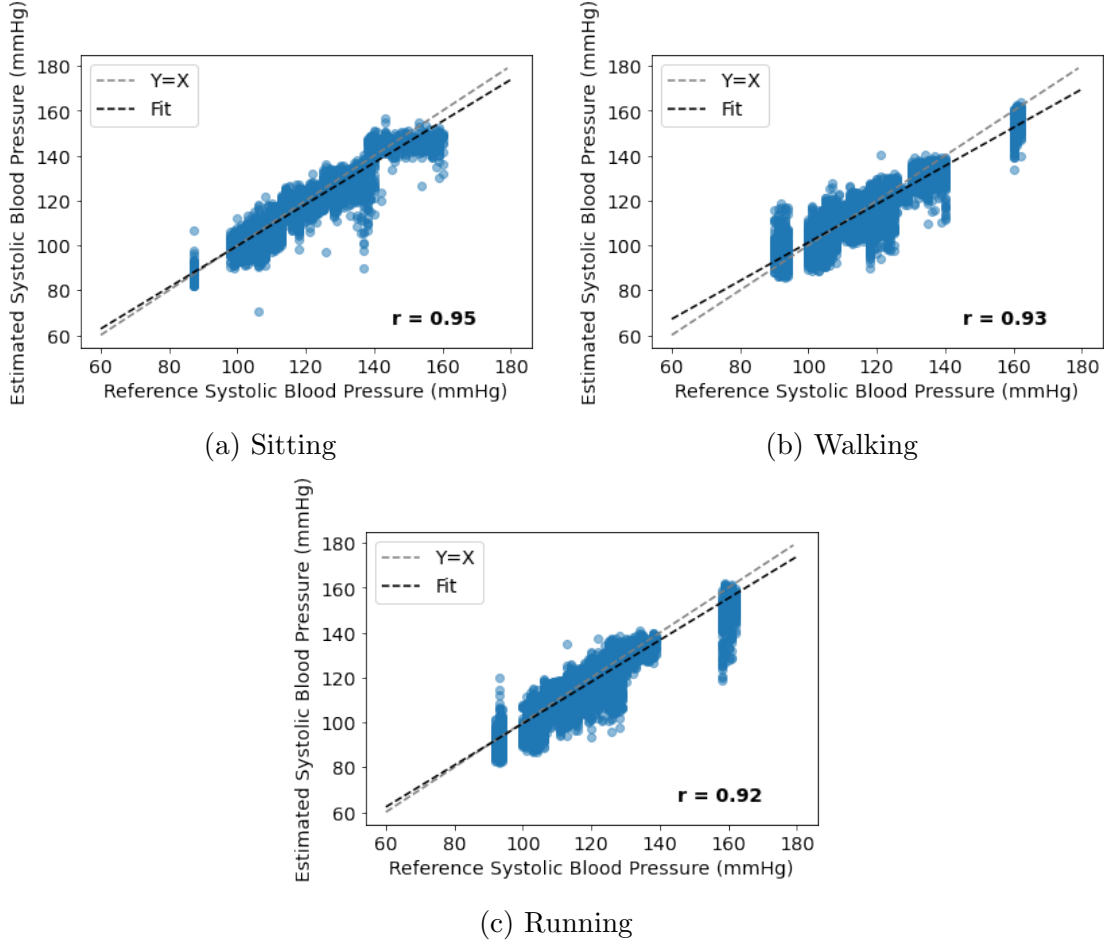


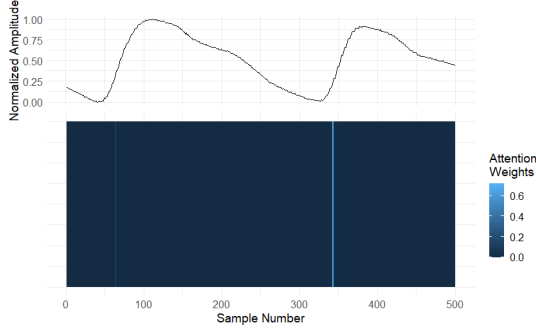
Figure 22: Scatter Plots for Systolic Estimates

the RBLAN results are in the close alignment with reported state-of-the-art methods for cNIBP estimation performance in a sitting position that use ECG and PPG and which do not require calibration. In comparison with the work by Miao et al.,¹⁸⁰ the RBLAN demonstrated 1.36 ± 4.01 mmHg for SBP versus -0.001 ± 3.102 mmHg, and 1.79 ± 2.48 mmHg for DBP versus -0.004 ± 2.199 mmHg. The RBLAN also shows substantial performance improvements when compared with the methods that utilized PPG signals alone, with recent research reporting 0.43 ± 9.35 mmHg for SBP and 0.21 ± 7.72 mmHg¹⁸² for DBP.

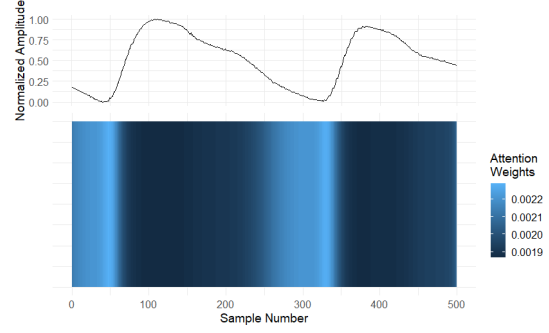
Further analysis of the RBLAN performance was accomplished by reviewing the Bland-Altman plots in Figure 19 and Figure 20. In these Bland-Altman plots, the limits

of agreement are identified along with points representing errors of determination for individual paired readings. Outliers were detected for systolic measurements in all positions and estimation errors > 38 mmHg were qualitatively investigated. It was anticipated that the errors would be due to noise in the PPG channel, because it is an optical sensor and more subject to movement artifact than the electrical sensors, but this was not determined to be the case. In the readings reviewed, some represented cases with ECG recordings that had significant noise levels to the point that R-peaks could not be determined, but the majority of these errors were issues with the RBLAN ECG R-peak detection that shifted the PTT calculations. This error could in part be caused because the ECG segment was always centred on the annotated peak, and the RBLAN did not review enough variability in the signal for robust performance. Some systemic bias was also noted in the high and low BP regions for systolic estimations during walking and running. This may be attributable to the lack of readings in these high and low BP regions and would be mitigated in larger, more diverse datasets with a more uniform hemodynamic distribution. Scatter plots were also generated for the RBLAN model performance for systolic and diastolic estimations (see [Figure 21](#) for diastolic estimates and [Figure 22](#) for systolic measurements). The Pearson’s correlation coefficients were found to be high for both diastolic and systolic estimations, with values ranging from 0.95 for sitting, to 0.91 for diastolic estimates during running.

The attention layers of the ResNet and BiLSTM are independent of one another, and produce a weighting score for each time step in the input series. These attention weights have been illustrated through a heatmap with an intensity scale in [Figure 23](#) and [Figure 24](#) for results from inference on one test sample processed through the RBLAN systolic estimation model. [Figure 23](#) shows the attention weights for a PPG segment and [Figure 24](#) for an ECG segment. Beyond providing a mechanism for segmenting features across multimodal channels, attention weights also help to interpret model performance and provide transparency into the decision-making process. Through the heat maps, it is demonstrated that the ResNet focuses on very narrow regions of both the ECG and PPG signal with high attention weights. The ResNet attention weights are primarily focused on the R-peak and negative S-peak of the QRS complex, and in the PPG the PTT_d of the following and proceeding pulse. This focus is in alignment with established PTT measurement locations and lends confidence to the underlying model. It also provides an interesting insight that referencing both the current and preceding peak of the PPG can aid in PTT measurements. The BiLSTM demonstrates low attention weights across both the ECG and PPG signal, indicating it is not contributing substantially to the RBLAN performance. This is not unexpected, given the small window size and limited sequential patterns represented in this short duration of analysis window. For other applications of adverse event detection, it is hypothesized that the BiLSTM will contribute more

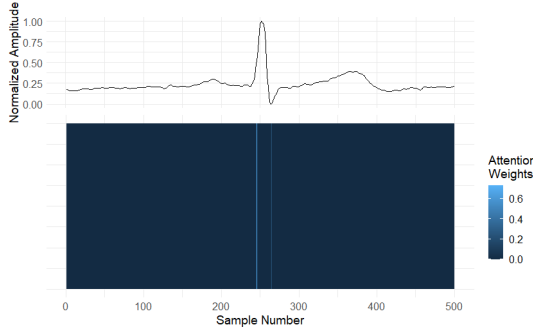


(a) ResNet Attention Mask

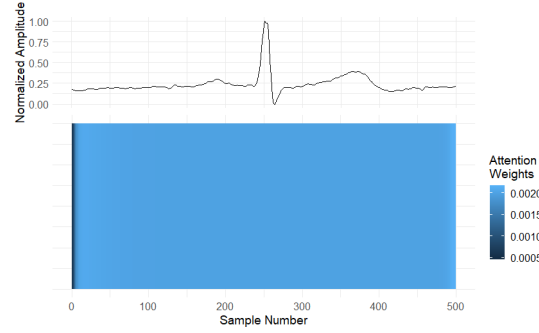


(b) BiLSTM Attention Mask

Figure 23: Heat Map of Attention Mask for PPG Signal



(a) ResNet Attention Mask



(b) BiLSTM Attention Mask

Figure 24: Heat Map of Attention Mask for ECG Signal

substantially when longer time frames are represented in the samples. The attention weights for the BiLSTM in the ECG are uniform across the signal, which could in part be due to the fact that the ECG pulse is centered in the window segment, and thus does not provide much information for PTT calculations. The attention weights for the BiLSTM on the PPG signal show elevated intensity through the PTT_f to PTT_d region.

3.4 Phase 3: Regression and Classification

3.4.1 Method

The final stage of analysis in the *BiometricNet* is an ensemble layer to fuse the output of all individual biometric channels together and produce a holistic patient risk metric (see architecture in [Figure 25](#)). Each individual biometric channel is processed in parallel and RBLAN output features (f_l) become inputs into a fusion model that performs either regression or classification. Cardio-respiratory physiological signals that are ideal to this feature extraction approach include: Respiration, PPG, ECG lead I, II, III, V, and continuous arterial blood pressure (ABP). Patient activity information can be integrated through the inclusion of accelerometer/gyroscope signal analysis to derive an overall activity index for gross motor assessments, or are also well-suited to more detailed time and frequency feature extraction through the RBLAN architecture.¹⁶³ If additional clinical time series are available in certain patient monitoring situations, such as continuous temperature measurements, these could be integrated directly by concatenating with the RBLAN output (f_l), without the requirement of extensive signal pre-filtering and parallel ResNet or BiLSTM modelling steps.

The fusion model can be adapted to produce an output that is relevant for the clinical application under investigation. For a regression task, such as that demonstrated for cNIBP in phase 2, a dense layer of 1 unit with a linear activation function is utilized to provide continuous values over a specified range. However, this is easily adjusted to utilize a sigmoid activation function in the output layer to achieve a binary classification, or softmax for multiple category classification. These regression and classification outcomes can be utilized for the detection of adverse clinical status and events, such as hypertension, hypotension, myocardial infarction, or the onset of cardiac arrhythmias.

For predicting patient deterioration or adverse clinical events, the *BiometricNet* utilizes a Siamese Neural Network (SNN) structure to learn highly discriminative properties of the RBLAN features between different points in the clinical care timeline to predict onset. SNNs first emerged for signature verification¹⁸⁵ and have been demonstrated as effective one-shot learning approaches to image recognition.⁴⁰ They have also been applied in healthcare for determining clinically relevant differentials in medical imaging.¹⁸⁶ Application of the SNN across a patient trajectory for predicting adverse clinical events is an innovative use of this network topology. The SNN is a twin network with shared parameters between two network instances. Output layers then utilize a Euclidean distance

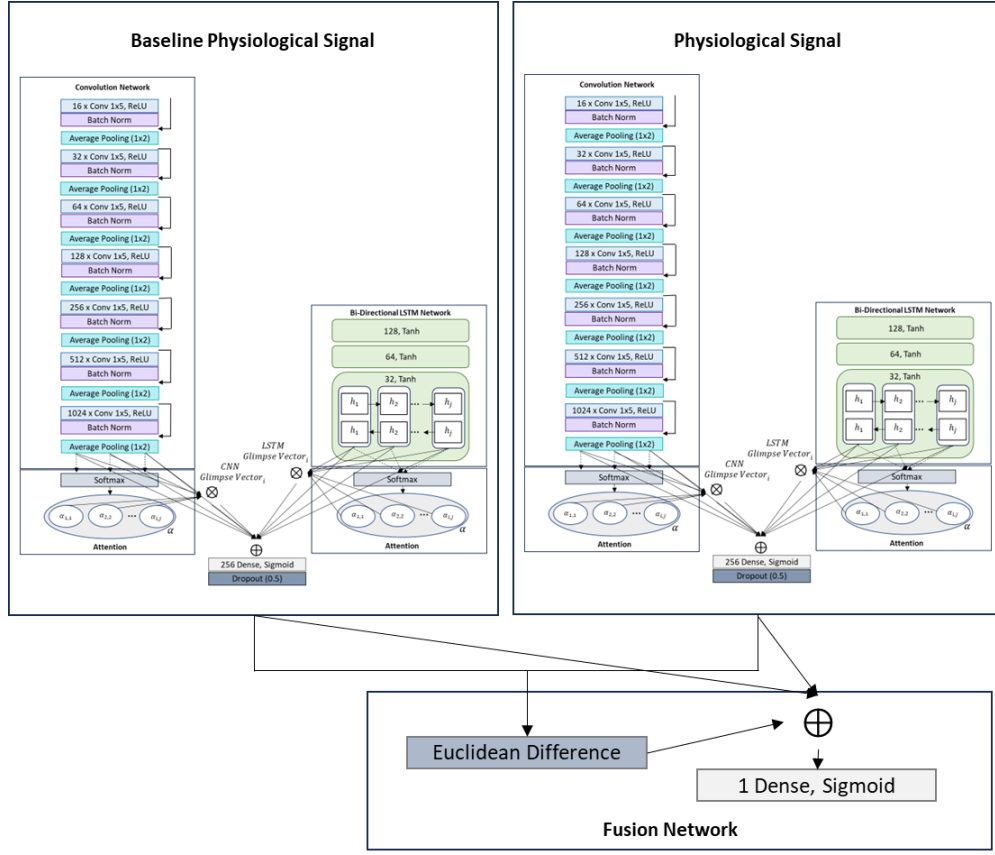


Figure 25: Siamese Neural Network for Multimodal Biometric Signal-Based Classification

to measure the difference between the feature vectors:

$$d_{Euclidean}(f_{1,l}, f_{2,l}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_{1,li} - f_{2,li})^2} \quad (3.31)$$

In the case of physiological signals, the SNN is trained to produce a similarity metric from RBLAN features that represents the differential in patient status between two chronological time points in the event horizon. The SNN thus magnifies changes in the patient's physiological signature across a clinical timeline to provide advanced notification of adverse clinical events. A benefit of an SNN is the demonstrated ability to support one-shot learning for unseen classifications. The resultant similarity metric is translated in a dense layer with a sigmoid activation function to q_i , a patient risk score on a normalized scale that represents

the patient trajectory at a specific point in a timeline. This q_i patient risk score is well suited to clinical care workflows that already utilize static Early Warning Scores.

It was determined empirically that the performance of the SNN for classification was improved if the features of the RBLAN were constrained. This is in contrast with the use of the RBLAN for regression, where an expanded feature space was beneficial. Regularization techniques were thus employed to constrain the complexity of the model during training, resulting in better generalization on unseen cases. Average pooling layers were added at the output of each convolutional layer, and dropout was added with a value of 0.5 at the dense output layer of the RBLAN. An additional layer was added to the CNN with a filter size of 1024.

A Binary Cross Entropy (BCE) loss function is utilized by the SNN, as it is the most commonly used loss function for binary classification tasks, where p_i is the predicted probability and y_i the reference classification:

$$BCE_{loss} = \frac{1}{N} \sum_{i=1}^N -(y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)) \quad (3.32)$$

Unbalanced datasets are a challenge for machine learning applications in healthcare. The incidence of the positive (e.g. disease) classification is far less prevalent than the control group. To address imbalance in the dataset, the SNN uses an adaptive weighting for the classifications based on the distribution of patients into the positive and negative groupings. These class weightings are calculated for a binary classification problem as:

$$weight_for_negative = \left(\frac{1}{negative_patients} \right) * \left(\frac{total_patients}{2.0} \right) \quad (3.33)$$

$$weight_for_positive = \left(\frac{1}{positive_patients} \right) * \left(\frac{total_patients}{2.0} \right) \quad (3.34)$$

The classification weighting vector cw is defined as:

$$cw = y_i * weight_for_positive + (1 - y_i) * weight_for_negative \quad (3.35)$$

The weighted BCE loss to address class imbalance becomes:

$$WBCE_{loss} = cw * BCE_{loss} \quad (3.36)$$

3.4.2 A Novel Siamese Network for Myocardial Infarction Prediction

Acute Myocardial Infarction (AMI) is a leading cause of mortality and morbidity in Canada, with an estimated 63,200 adults experiencing their first heart attack each year (2012/2013 data).¹⁸⁷ Diagnosis of AMI clinically is done by analyzing ECG recordings in conjunction with blood tests that measure troponin levels that signal cardiac injury. Time is of the essence for AMI detection to minimize the extent of cardiac injury, so solutions that can utilize clinical wearable devices at point-of-care can improve on the time it takes to seek essential treatment. AMIs are often categorized as ST Elevation Myocardial Infarction (STEMI) or Non-ST-Elevation Myocardial Infarction (NSTEMI), based on ECG criteria.¹¹³ Analysis of ECG signals for the prediction of AMI is a meaningful application for the *BiometricNet* and a demonstration of the ability to predict adverse clinical events at a future time horizon.

The *BiometricNet* utilizes a baseline ECG recording for the patient and then identifies changes over time to signal early pre-cursors as the AMI progresses. The ECG signal has been demonstrated to change as the patient progresses through the early stages of an AMI,¹⁸⁸ which makes it an ideal application of the SNN approach. Starting with hyperacute T waves (large, peaked T waves) that start immediately after onset of the infarction and are present for 5 to 30 minutes. These are followed by ST segment elevation that develops a few hours after the first symptoms, and is often the first discernible change in ECG for an AMI. As the AMI proceeds, the R-wave will lose amplitude and there will be an onset of pathological Q waves (abnormally wide or deep), which can occur as early as one to two hours after symptom onset, or take up to 24 hours to develop. The ST segment then begins to resolve and the T-waves begin to invert. ST segment elevation can remain for up to two weeks post-AMI and T-waves can remain for months.¹⁸⁸

Research into deep learning methods for Myocardial Infarction prediction has shown promising results in current literature.^{112,113} In Han et al.,¹¹² a ResNet architecture was utilized to analyze 12-lead ECG signals from the PTB (Physikalisch-Technische Bundesanstalt) database. This research reported an accuracy of 95.49%, sensitivity of 94.85%, and specificity of 97.37% on the task of MI detection. Most recently, Gustafsson et al.¹¹³ utilized a ResNet architecture to analyze 12-lead ECG signals, but also augmented the input parameters with age and sex of the patient. Performance of this model was reported on a custom dataset of 492,226 ECG recordings from 214,250 patients gathered in the emergency rooms of Sweden, as well as a subset of the PTB-XL ECG dataset (PTB-XL).¹⁸⁹ This model produced a reported overall C-statistic of 0.991 on STEMI, and a C-statistic of 0.832 on NSTEMI.

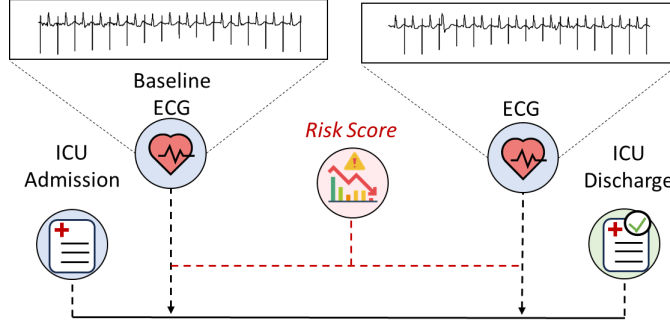


Figure 26: Prediction of Myocardial Infarction from ECG Signals Extracted from the Patient ICU Admission Timeline

3.4.2.1 Data

The PTB Diagnostic ECG Database¹³⁷ (PTB) is an open source database (available through PhysioNet)¹⁹⁰ that contains 549 ECG recordings captured from 290 patients. The labelled subset of PTB represents a number of cardiac conditions: Myocardial Infarction (148), Cardiomyopathy/Heart Failure (18), Bundle Branch Block (15), Dysrhythmia (14), Myocardial Hypertrophy (7), Valvular Heart Disease (6), Myocarditis (4), Miscellaneous (e.g., Angina, 4), and Healthy Controls (52). There are 15 simultaneously recorded ECG channels, including a standard 12 lead configuration (I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6), as well as three Frank leads (Vx, Vy, and Vz). The cases of MI represented in this dataset are limited to incidence of STEMI, and do not represent NSTEMI events. The ECG recordings are high fidelity, sampled at 1 kHz with 16 bit resolution and an input voltage range of ± 16 mV. Each patient record contains one to five ECG waveforms of ten second duration across all 15 ECG channels. A larger extension of the PTB dataset (PTB-XL)¹⁸⁹ is also available at PhysioNet, containing 21,837 12-lead ECG recordings captured from 18,885 patients. For this preliminary experiment, the PTB dataset was utilized with plans to extend the analysis to the larger dataset in future work.

A leave-k-out cross-validation strategy was applied with five generated training and testing folds. Each training fold was randomly assigned 80% of the patients, and 20% to the testing set. The representation of the *MI* versus *Other* classifications was balanced within the training and testing folds. Only the lead I, II, and III channels were included in the analysis, as these represented common configurations available on a three-lead clinical wearable device. The ten second ECG segment for each recording was split into a five second baseline and five second analysis window. Pre-processing of the ECG segments was done with the ψ -GANDAE (see section 3.2 for details). The ψ -GANDAE was pre-trained

using all 48 patient records within the PhysioNet MIT-BIH Arrhythmia Database¹³⁴ with noise injected from the PhysioNet MIT-BIH Stress Test Database¹³⁵ at 0 dB, 1.25 dB, and 5 dB levels.

3.4.2.2 Results

The *BiometricNet* was trained on NVIDIA GeForce GPUs for matrix acceleration. TensorFlow¹⁵⁰ and Keras¹⁵¹ were utilized to construct the network and implement the training and testing process, SciPy¹⁸⁴ was used for signal pre-processing support with the ψ -GANDAE, and Python¹⁵³ provided the programming framework for design and implementation activities. A batch size of 8 was utilized and the networks were trained and evaluated over 15 epochs. A constant learning rate of 0.00001 was used with Adaptive Moment Estimation (Adam) for the optimizer and $WBCE_{loss}$ for the loss function.

Hyperparameters

Optimizer: Adaptive Moment Estimation optimizer (Adam)
 Learning Rate: 0.00001
 Epochs: 15
 Batch Size: 8
 Loss: Weighted Binary Cross Entropy Loss (See equation 3.36)

The performance of the *BiometricNet* on AMI prediction was assessed against the C-Statistic or Area Under the Receiver Operator Characteristics curve (AUROC). The AUROC curves show the true positive rate (Sensitivity) versus the false positive rate (1-Specificity), where Sensitivity and Specificity are calculated as:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3.37)$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad (3.38)$$

The Average Precision (AP) metric will also be evaluated for the model, which represents the Area Under the Precision-Recall Curve (PR Curve). Where precision and recall are defined as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (3.39)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3.40)$$

The performance of the *BiometricNet* on the task of myocardial infarction prediction for the PTB Diagnostic ECG Database are summarized in Table 3.6. Validation of the model was done through Leave-k-Out cross validation, with all records for an individual patient segregated to either the training or testing set, to better assess the system’s ability to generalize on unseen cases. The performance for each fold of training/testing is detailed in Table 3.6, including number of samples in the fold, the resultant C-Statistic, and Average Precision (AP).

Table 3.6: Performance of the BiometricNet for Myocardial Infarction Prediction

Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
Total Samples (N)	88	111	110	112	102	104.60
C-Statistic (AUROC)	0.87	0.88	0.86	0.83	0.88	0.86
Average Precision	0.94	0.93	0.90	0.89	0.94	0.92

3.4.2.3 Analysis

The results in Table 3.6 demonstrate the performance of the *BiometricNet* for predicting myocardial infarction. The model produced an overall average C-Statistic of 0.86 (0.02) and Average Precision of 0.92 (0.02) across all validation folds. A comparison of predicate research is difficult, given that all methods reviewed analyzed all 12 leads of ECG for prediction, and the *BiometricNet* utilized only leads I, II, III. This restriction in lead configurations was made to assess suitability for use with clinical wearable devices that are limited in auxiliary leads and constrained by memory and flash storage. Recent work by Gustafsson et al.¹¹³ utilized all 12 leads of ECG, as well as age and sex, as inputs into their ResNet, which produced a C-Statistic of 0.932 (0.904–0.959) and AP of 0.955 (0.934–0.972) on a subset of the PTB-XL database. While performance of the *BiometricNet* is lower than those reported by Gustafsson et al.,¹¹³ it is hypothesized that this is due to differences in the dataset and inputs utilized for the model training, and not the overarching system architecture. This will be further validated in future work, where the *BiometricNet* will be trained on the extended PTB-XL database, and a baseline with 12-leads produced to compare directly with predicate research.¹⁸⁹

Performance of the *BiometricNet* was also assessed by constructing graphs for the Receiver Operator Characteristics curves (Figure 27), and Precision-Recall curves (Figure 28). The Receiver Operator Characteristics curves illustrated in Figure 27 show

the performance curve for each of the folds in the testing set with the Sensitivity on the y-axis and the false positive rate (1-Specificity) on the x-axis. The dotted line in Figure 27 demonstrates theoretical performance of a random classifier. The Precision-Recall curves in Figure 28 show Precision on the y-axis and Recall on the x-axis. It is noted on the Precision-Recall curve that there are “steps” in the profile, which is attributed to the smaller size and distribution of the testing set from the PTB database. This will be resolved when using the PTB-XL database with a larger training and testing set.

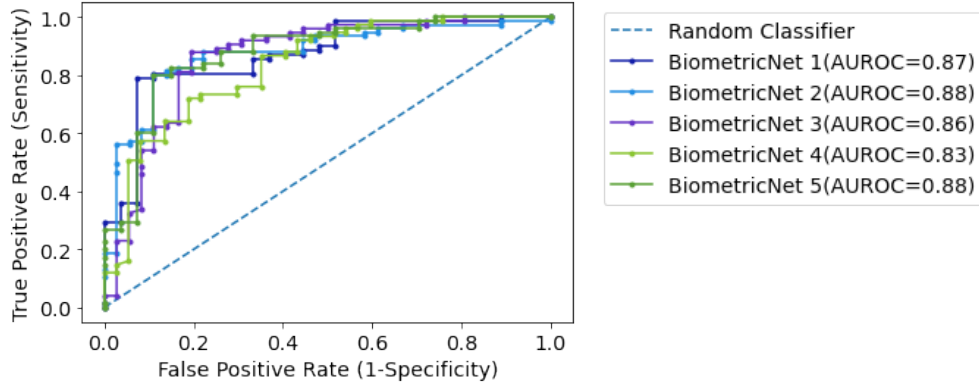


Figure 27: Receiver Operator Characteristics Curves for Myocardial Infarction Prediction

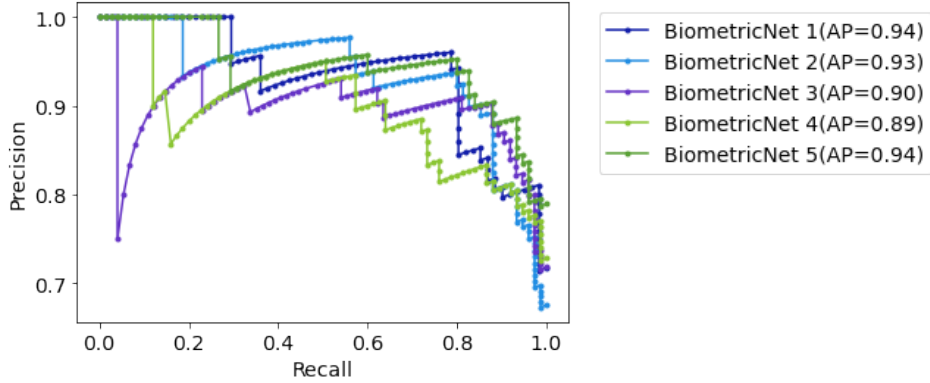


Figure 28: Precision-Recall Curves for Myocardial Infarction Prediction

3.5 Summary

In chapter 3, a novel deep learning framework has been presented to support the processing of continuous, multimodal biometric signals in near real-time for the detection of patient deterioration and prediction of adverse clinical events. The performance of the *BiometricNet* architecture for this framework has been analyzed for each of the three primary processing stages, 1) Signal pre-processing, 2) Signal segmentation and feature extraction, and 3) Regression or classification. The design, implementation, and testing of the *BiometricNet* architecture has been explored, with experimental results presented for each processing phase.

In section 3.2, the first phase of the *BiometricNet* architecture was investigated, which accomplishes signal pre-processing with a Generative Multiscale Wavelet De-noising Autoencoder (Ψ -GANDAE). The Ψ -GANDAE utilizes a GAN structure, with generator and discriminator networks. The generator leverages important elements of U-Net and ResNet architectures, and integrates wavelet transform decomposition into successive layers of this deep network. The Ψ -GANDAE also provides multiscale resolution, with enhanced processing on individual beat morphology, as well as higher levels of review for sequential information of the signal rhythm. The performance of the Ψ -GANDAE was assessed on the PhysioNet MIT-BIH Arrhythmia Database¹³⁴ with noise introduced from the PhysioNet MIT-BIH Stress Test Database,¹³⁵ which represents ECG baseline wander, muscle artifact, and electrode motion artifact at 0 dB, 1.25 dB, and 5 dB levels. While the Ψ -GANDAE showed lower SNR and RMSE metrics as compared to the Baseline GAN when compared utilizing k-fold cross validation, it showed an overall average improvement of 1.22 dB for SNR and 0.049 for RMSE when evaluated with leave-one-out cross validation.

In section 3.3, a ResNet-BiLSTM-Attention model (RBLAN) was presented for signal segmentation and feature extraction within the *BiometricNet* framework. The RBLAN utilizes a ResNet sub-network to extract time domain, frequency domain, and time-frequency domain features of the physiological signals, and a BiLSTM sub-network to identify inherent sequential dependencies. An attention post-processing layer is provided that produces a glimpse vector, which is the attention weights applied to the sub-network outputs. This attention layer supports segmentation of the features across multiple biometric channels, as well as providing transparency on the features being utilized within the network for interpretability and clinical confidence in the underlying model. The RBLAN was demonstrated on the task of cNBIP estimation using the Pulse Transit Time (PTT) PPG Dataset.¹³⁶ Both systolic and diastolic estimates were determined by the RBLAN, using data collected in both sedentary (i.e., sitting) and ambulatory positions (i.e., walking and running). The RBLAN demonstrated a mean absolute error $< 5 \pm 8$

mmHg for systolic and diastolic estimates in all positions, achieving compliance with ISO 81060-2:2018/Amd 1:2020 for certification of blood pressure measurement devices. It also achieved the highest rating of Grade A for blood pressure derivations as defined by the British Hypertension Society, which is based on the percentage of readings within 5, 10, and 15 mmHg of the reference measurements.

In section 3.4, the ability of the *BiometricNet* architecture to achieve both regression and classification tasks was demonstrated. In this last stage of processing, the *BiometricNet* utilizes a fusion network that reduces the feature space, and produces a continuous value output using a linear activation for regression. For classification tasks, the *BiometricNet* uses a Siamese Neural Network (SNN) to compare a patient’s physiological signature across the clinical timeline. The SNN is a twin network, with a differentiating output layer that computes the Euclidean distance between feature vectors at two different chronological points in time. The demonstrated end-to-end performance of the *BiometricNet* was evaluated on the task of myocardial infarction prediction using the PTB Diagnostic ECG Database.¹³⁷ The experiment used a pre-trained version of the Ψ -GANDAE for ECG signal pre-processing, a modified RBLAN for signal segmentation and feature extraction, and a SNN to produce a risk score for MI. The *BiometricNet* analyzed only leads I, II, and III for predicting MI, to better represent the application of this technology to three-lead clinical wearable devices. The *BiometricNet* demonstrated a promising overall average C-Statistic of 0.86 (0.02) and Average Precision of 0.92 (0.02) on the task of AMI prediction.

Chapter 4

Conclusions and Future Directions

In this dissertation, a novel deep learning framework has been presented for processing continuous, multimodal physiological signals for the detection and prediction of adverse clinical events. The design, implementation, and testing of the *BiometricNet* architecture was investigated and demonstrations provided for each of the three primary processing stages, namely: 1) Signal pre-processing utilizing a novel Generative Multiscale Wavelet De-noising Autoencoder (Ψ -GANDAE), 2) Signal segmentation and feature extraction with a ResNet-BiLSTM-Attention model (RBLAN), and 3) Classification or regression from an ensemble of multimodal biometric signals, where a Siamese Neural Network (SNN) is used to support classification outcomes.

4.1 Conclusions

The Ψ -GANDAE is a novel Generative Multiscale Wavelet De-noising Autoencoder that performs signal pre-processing in the *BiometricNet*. The Ψ -GANDAE utilizes a generative architecture, with a discriminator and generator to improve model performance during training and testing. It also harnesses key elements of U-Net and ResNet architectures, and inherently integrates wavelet transform decomposition into the model by separating layers into different time-frequency sub-bands. The Ψ -GANDAE provides different scales of processing to analyze both aspects of the individual beat morphology in biometric signals, as well as beat-to-beat rhythm. The performance of the Ψ -GANDAE was demonstrated on ECG recordings from the PhysioNet MIT-BIH Arrhythmia Database¹³⁴ with common noise injected from the PhysioNet MIT-BIH Stress Test Database,¹³⁵ which represent baseline wander, muscle artifact, and electrode motion artifact at 0 dB, 1.25

dB, and 5 dB levels. When evaluated utilizing k-fold cross validation, the Ψ -GANDAE demonstrated an average signal quality for the reconstructed ECG across all noise levels of SNR=21.86 dB and RMSE=0.053 for BW, SNR=21.07 dB and RMSE=0.060 for EM, and SNR=21.27 dB and RMSE=0.062 for MA, with an overall performance of SNR=21.40 dB and RMSE=0.058. The performance of the Ψ -GANDAE showed an overall average improvement of 1.22 dB for SNR and 0.049 for RMSE when compared to the baseline GAN method using leave-one-out cross validation. This demonstrates that the Ψ -GANDAE is a suitable pre-processing stage for physiological signals that offers more reliable de-noising when applied to unseen cases. As a deep learning model, the Ψ -GANDAE improves on performance from traditional Wavelet Transform methods alone, and benefits from continued pre-training on representative ECG datasets.

The second stage of the *BiometricNet* is a ResNet-BiLSTM-Attention model (RBLAN) for signal segmentation and feature extraction. The RBLAN consists of ResNet and BiLSTM sub-networks that are supported by attention output layers. The ResNet network extracts time domain, frequency domain, and time-frequency domain features of the physiological signals, and the BiLSTM sub-network identifies inherent temporal and sequential dependencies. An attention post-processing layer is provided that supports segmentation of the features across multiple biometric channels, as well as providing transparency on the features being utilized within the network for interpretability and clinical confidence in the underlying model. This is a flexible and adaptive feature extraction method to support different signal modalities and applications to adverse clinical event detection and prediction. The RBLAN’s performance was investigated using the Pulse Transit Time (PTT) PPG Dataset¹³⁶ for cNIBP estimation of systolic and diastolic blood pressure metrics. The results showed an overall mean error of 1.91 (3.10) and absolute error of 2.83 (2.31) for diastolic estimations, and mean error of 1.48 (4.58) with mean absolute error of 3.53 (3.28) for systolic estimations. These are especially promising results for a cNIBP method that does not require pre-calibration, and given the data was collected in both sedentary (i.e., sitting) and ambulatory positions (i.e., walking and running). In conclusion, the performance of the RBLAN demonstrated compliance with ISO 81060-2:2018/Amd 1:2020 for systolic and diastolic estimates in all positions, meeting the requirements for a mean absolute error $<5 \pm 8$ mmHg. The RBLAN also met the requirements from the British Hypertension Society with a Grade A criteria, based on the percentage of readings within 5, 10, and 15 mmHg.

The final stage of the *BiometricNet* is an ensemble layer that utilizes a fusion network to bring all individual biometric channels together to produce a holistic patient risk metric. For regression tasks, the fusion network contains a dense layer with a linear activation, which reduces the features generated by the RBLAN and produces a continuous value

output within a specified range. For classification tasks, a Siamese Neural Network (SNN) architecture is utilized to amplify subtle changes in the patient’s physiological signature across a timeline to provide advanced notification of patient deterioration. The SNN is a twin network with shared parameters between network instances. A differentiating layer then utilizes a Euclidean distance to measure the difference between the feature vectors and augments the RBLAN features for processing by the output layer. In the output network, a dense layer is then utilized with a sigmoid activation function to produce a probabilistic score for a binary classification.

The application of the full *BiometricNet* architecture was demonstrated for the task of myocardial infarction prediction on the PTB Diagnostic ECG Database.¹³⁷ This experiment utilized the Ψ -GANDAE for signal pre-processing of ECG signals, a modified RBLAN for feature extraction, and a SNN for classification. The ECG data was limited to lead I, II, and III configurations to assess the performance of the system for three-lead clinical wearable devices. The *BiometricNet* produced a C-Statistic of 0.86 (0.02) and Average Precision of 0.92 (0.02) during Leave-k-Out cross validation on the PTB Diagnostic ECG Database.¹³⁷ This performance is a clear advancement on current processes for MI triage, with estimates indicating there is currently only a 54% accuracy in identifying cases within an ER setting.¹⁹¹ These results are promising, and offer the potential for the *BiometricNet* framework to be utilized to triage STEMI identification in ECG recordings from three-lead clinical wearable devices.

Research into the design, implementation, and application of the *BiometricNet* has honoured important considerations regarding AI ethics and network interpretability and transparency. This research has a strong ethical alignment, with an objective to improve health outcomes for all post-surgical patients by removing access barriers with advanced remote patient monitoring technologies. Ultimately, adoption of these technologies will help drive equity and inclusion in healthcare. An additional concern with AI systems is the “black box” nature of the solution, where it is unclear how and why outputs are being generated. To provide *transparency* as an AI system, the *BiometricNet* framework incorporates architectural elements to support model *explainability* and provide evidence to understand the decision-making process. The *BiometricNet* utilizes attention layers within the RBLAN feature extraction processing stage to provide a window through which areas of the input signal that are important for the output signal can be visualized. In the experiment conducted for this research on continuous non-invasive blood pressure estimations, the attention layers clearly identified the R-peak and negative S-peak in the QRS complex of the ECG, and maximum of the first derivative in the PPG signal, as the primary focus for analysis. These areas are established PTT landmarks, which give confidence in the model’s performance and thus *explainability*. The attention mechanism

was also used to investigate cases of model failure, and aid in network *interpretability* for determining predictability of system performance. Transparency is a crucial element of an AI system in healthcare, and key to building confidence from clinical practitioners and regulatory agencies.

4.2 Limitations

The Ψ -GANDAE is the pre-processing stage of the *BiometricNet*. In this dissertation, the performance of the Ψ -GANDAE was demonstrated on ECG data from the PhysioNet MIT-BIH Arrhythmia Database¹³⁴ with common noise injected from the PhysioNet MIT-BIH Stress Test Database.¹³⁵ In the next stages of research for the *BiometricNet*, the Ψ -GANDAE will be required to filter many different signal modalities and remove common noise patterns from each. This will require labelled datasets for each signal type, with adequate samples of all noise profiles at quantified dB levels, as well as selection of wavelets that are optimal for decomposing each signal type.

The Ψ -GANDAE utilizes a fixed and pre-determined window length for signal segmentation. While a multiscale decomposition of the segmented window is inherently supported within the Ψ -GANDAE’s architecture, the primary resolution of the data analysis is fixed and pre-determined. In future work involving multiple signal modalities, the Ψ -GANDAE would benefit from an adaptive signal segmentation strategy that automatically determines an optimal analysis window for the specified problem domain. A transformer architecture may hold promise for achieving adaptive segmentation within the Ψ -GANDAE, given it is well-suited for sequence-to-sequence encodings.¹⁹²

The second stage of the *BiometricNet* is an ResNet-BiLSTM-Attention model (RBLAN) for signal segmentation and feature extraction. The RBLAN produced promising performance for cNIBP estimation as demonstrated on the Pulse Transit Time (PTT) PPG Dataset.¹³⁶ The data for the cNIBP experiment was segmented into windows for processing that centered around annotated R-peaks from the ECG segment. This segmentation strategy introduces compounding error from the R-peak detection algorithm, and it contributes to a lack of resilience in the model for ECG variability. There was also some systemic bias witnessed in systolic estimations captured during walking and running, which is potentially due to the lack of representation of reading data in high and low blood pressure regions.

In the third stage of the *BiometricNet* a Siamese Neural Network (SNN) was applied to compare a baseline for a patient against a point on the care timeline to predict

patient deterioration and adverse clinical events. The SNN was demonstrated on the PTB Diagnostic ECG Database¹³⁷ (PTB) on the task of MI prediction. Performance of the model was difficult to compare with predicate research, due to the differing number of ECG leads evaluated and datasets utilized. In ongoing experiments, the *BiometricNet* is being evaluated utilizing the extension of the PTB dataset (PTB-XL)¹⁸⁹ with a 12-lead configuration for a better comparison to predicate research. The SNN also currently provides a determination on each 10 second segment of ECG, and with even with a low false positive rate this will produce a considerable number of false alarms in a clinical setting that could lead to alarm fatigue. As a triage tool, the *BiometricNet*'s performance will need to be tuned through the adaptive class weighting mechanism to ensure a sensitivity of 100% and thus identification of all patients experiencing an acute myocardial infarction, which will also negatively impact specificity and increase false positives. A secondary network is to be investigated that will process individual determinations from the SNN and consider them over time to reduce the false alarm rate to a practicable level.

An important aspect to consider for AI ethics is the constitution of the datasets utilized for training and testing to identify underlying biases. These datasets must be fully representative of the entire population they are supporting or there will be inequalities in performance for the predictive models. All three stages of the *BiometricNet* were evaluated utilizing public datasets, which are limited in size and diversity. The Pulse Transit Time (PTT) PPG Database¹³⁶ utilized for the continuous non-invasive blood pressure estimation experiments for instance, lacks gender (27% female), age (mean of 28.52 years), and health diversities (all healthy subjects). The MIT-BIH Arrhythmia Database¹³⁴ is more gender diverse (46% female), but with data primarily restricted to an elderly population (mean age in years 63.67 ± 17.58). While cardiac arrhythmias are more prevalent in an elderly population, it is critical that all aspects of a population are adequately represented in the training and testing datasets to ensure there is consistent model performance for all patients. An important next step for the *BiometricNet* will be to evaluate performance on large, multi-site datasets to ensure that performance can generalize to diverse, unseen populations.

4.3 Future Directions

As a comprehensive framework for detecting and predicting clinical events the *BiometricNet* has many potential areas for further application. In this dissertation, the *BiometricNet* has been demonstrated on the application of de-noising of ECG signals, the detection of continuous, non-invasive blood pressure measurements, and the prediction of myocardial

infarction. Future research directions will look to demonstrate the utility of the *BiometricNet* for de-noising additional physiological signal modalities, as well as expanding the types of clinical events that can be detected and predicted, while expanding the horizon of event prediction.

The Ψ -GANDAE is the signal pre-processing network for the framework, which was demonstrated in this thesis on the task of ECG signal denoising. Transfer learning will be further investigated to demonstrate the application of the pre-trained Ψ -GANDAE to other physiological signals. It is postulated that the non-stationary, sparse characteristics of photoplethysmogram and continuous intra-arterial blood pressure measurements will benefit from a model pre-trained on ECG data, with a similarity in artifacts due to motion and respiration.

The *BiometricNet* utilizes a Siamese Neural Network to compare a baseline for a patient against a point on the care timeline to predict patient deterioration and adverse clinical events. This baseline is constructed from features extracted from physiological signals by ResNet and BiLSTM architectures with Attention (RBLAN). In future work, the feature extraction from physiological signals can be augmented with other data modalities, including medical imaging and acoustic recordings of lung or heart sounds. The baseline for the patient has also been envisioned in the initial implementation of the *BiometricNet* as a single timeframe, which aligns well with a pre-surgical consultation period prior to convalescence monitoring. However, a multiscale framework for reference time frames will be investigated in future work, which would incorporate many reference points for comparison at intervals of minutes, hours, or even days to amplify changes and increase the prognostication window for adverse clinical events.

Sepsis is a critical medical condition that occurs as an immune response to an infection and is a significant contributor to mortality in an ICU setting. The *BiometricNet* is an ideal framework for developing a predictive model to detect and predict sepsis based on the analysis of continuous, multimodal physiological signals. The MIMIC-III Clinical Database¹⁹³ contains records on 4,652 ICU admissions that last > 72 hours for 4,086 distinct adult patients with ICD-9 codes consistent with sepsis. Clinically, the Systemic Inflammatory Response Syndrome (SIRS) Criteria defines vital metric measurements and laboratory results where ≥ 2 indicate the onset of sepsis.¹⁹⁴ The SIRS criteria can be utilized in connection with the matched waveforms for each patient from the MIMIC-III Waveform Database Matched Subset database¹⁹³ to determine an onset time that can be utilized to extract a timeline of adverse event data to create a meaningful training and testing set. The *BiometricNet* will then be trained to determine the ability to predict sepsis across a 72 hour event horizon.

The absence of large data sets containing labelled, synchronized, continuous physiological waveforms, remains a significant limitation to developing predictive models of adverse health events. To address this limitation, a 20,000 patient study, called VISION-2 has been developed by researchers at the Population Health Research Institute and McMaster University that will follow patients for 30 days during post-surgical convalescence, with high fidelity, multi-parameter vital sign monitoring.¹⁴ This will be an international, multi-site study, with careful study design to ensure diversity and inclusion in patient recruitment. The dataset captured during this study will serve to continue this research with the application of the *BiometricNet* to predict common post-surgical adverse events, such as myocardial infarction/injury, bleeding, sepsis, and delirium.

References

- [1] J.G. Meara, A.J.M. Leather, L. Hagander, B.C. Alkire, N. Alonso, E.A. Ameh, S.W. Bickler, L. Conteh, A.J. Dare, J. Davies, E.D. Mérisier, S. El-Halabi, P.E. Farmer, A. Gawande, R. Gillies, S.L.M. Greenberg, C.E. Grimes, R.L. Gruen, E.A. Ismail, T.B. Kamara, C. Lavy, G. Lundeg, N.C. Mkandawire, N.P. Raykar, J.N. Riesel, E. Rodas, J. Rose, N. Roy, M.G. Shrimé, R. Sullivan, S. Verguet, D. Watters, T.G. Weiser, I.H. Wilson, G. Yamey, and W. Yip. Global surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *The Lancet*, 386(9993):569–624, Aug 2015.
- [2] T. Weiser, A. Haynes, G. Molina, S.R. Lipsitz, M.M. Esquivel, T. Uribe-Leitz, R. Fu, T. Azad, T. Chao, W. Berry, and A.A. Gawande. Size and distribution of the global volume of surgery in 2012. *Bulletin of the World Health Organization*, 94:201–209, 2016.
- [3] The Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators. Association between complications and death within 30 days after noncardiac surgery. *CMAJ*, 191(30):E830–837, 2019.
- [4] P.J. Devereaux, B.M. Bickard, A. Sigamani, and et al. Association of postoperative high-sensitivity troponin levels with myocardial injury and 30-day mortality among patients undergoing noncardiac surgery. *JAMA*, 317:1642–1651, 2017.
- [5] P.J. Devereaux, M.T.V Chan, P. Alonso-Coello, and et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA*, 307:2295–2304, 2012.
- [6] I. Harris, A. Madan, J. Naylor, and S. Chong. Mortality rates after surgery in new south wales. *ANZ journal of surgery*, 82, 2012.

- [7] J.L.W. Ng, M.T.V. Chan, and A.W. Gelb. Perioperative stroke in noncardiac, nonneurosurgical surgery. *Anesthesiology*, 115(4):879–890, 2011.
- [8] A.S. Nagre. Perioperative stroke – prediction, prevention, and protection. *Indian Journal of Anaesthesia*, 62(738), 2018.
- [9] F. Macellari, M. Paciaroni, G. Agnelli, and V. Caso. Perioperative stroke risk in nonvascular surgery. *Cerebrovasc Dis*, 34:175–181, 2012.
- [10] O. Chaiwat, M. Chanidnuan, W. Pancharoen, and et al. Postoperative delirium in critically ill surgical patients: incidence, risk factors, and predictive scores. *BMC Anesthesiol*, (39), 2019.
- [11] G.H. Almassi, T. Schowalter, A.C. Nicolosi, A. Aggarwal, T.E. Moritz, W.G. Henderson, R. Tarazi, A.L. Shroyer, G.K. Sethi, F.L. Grover, and K.E. Hammermeister. Atrial fibrillation after cardiac surgery: a major morbid event? *Ann Surg*, 226(4):501–511, 1997.
- [12] C.H. Leuwan and I. Mitchell. Missed opportunities? an observational study of vital sign measurements. *Crit Care Resusc*, 10:111–115, 2008.
- [13] M.H. McGillion, E. Duceppe, K. Allan, M. Marcucci, S. Yang, AP Johnson, S. Ross-Howe, E. Peter, T. Scott, C. Ouellette, S. Henry, Y. Le Manach, G. Paré, B. Downey, SL Carroll, J. Mills, A. Turner, W. Clyne, N. Dvirnik, S. Mierdel, L. Poole, M. Nelson, V. Harvey, A. Good, S. Pettit, K. Sanchez, P. Harsha, D. Mohajer, S. Ponnambalam, S. Bhavnani, A. Lamy, R. Whitlock, P.J. Devereaux, and on behalf of the PROTECT Network Investigators. Remote automated postoperative monitoring: Need for and state of the science. *The Canadian Journal of Cardiology*, 34(7), 2018.
- [14] M.H. McGillion, K. Allan, S. Ross-Howe, W. Jiang, M. Graham, M. Marcucci, A. Johnson, T. Scott, C. Ouellette, D. Kocetkov, J. Lounsbury, M. Bird, P. Harsha, K. Sanchez, V. Harvey, J. Vincent, F.K. Borges, S.L. Carroll, E. Peter, A. Patel, S. Bergh, and P.J. Devereaux. Beyond wellness monitoring: Continuous multiparameter remote automated monitoring of patients. *Can. J. Cardiol.*, 38(2):267–278, February 2022.
- [15] M. Al-khafajiy, T. Baker, C.Chalmers, M. Asim, H. Kolivand, M. Fahim, and A. Waraich. Remote health monitoring of elderly through wearable sensors. multimedia tools and applications. *Multimedia Tools and Applications*, 1(26), 2019.

- [16] L.A. Durán-Vega, P. C. Santana-Mancilla, R. Buenrostro-Mariscal, J. Contreras-Castillo, J. L. E. Anido-Rifón, M. A. García-Ruiz, and F. Estrada-González. An iot system for remote health monitoring in elderly adults through a wearable device and mobile application. *Geriatrics*, 4(2):34, 2019.
- [17] C.P. Subbe, B. Duller, and R. Bellomo. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. *Crit Care*, 21(52), 2017.
- [18] G. Smith, D. Prytherch, P. Meredith, P. Schmidt, and P. Featherstone. The ability of the national early warning score (news) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation*, 84, 2013.
- [19] G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. *ArXiv*, 2015.
- [20] Y. Lecun. *PhD thesis: Modeles connexionnistes de l'apprentissage (connectionist learning models)*. PhD thesis, 1987.
- [21] H. Bourlard and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*, 59:291–294, 1988.
- [22] G. Hinton and R. Zemel. Autoencoders, minimum description length, and helmholtz free energy. In *NIPS' 1993*, 1991.
- [23] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. The MIT Press, 2016.
- [24] G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(28), 2006.
- [25] D. Kingma and M. Welling. Auto-encoding variational bayes. *ArXiv*, 2013.
- [26] D. Rezende, S. Mohamed, and D. Wierstra. Stochastic backpropagation and approximate inference in deep generative models. *ArXiv*, 2014.
- [27] M. Kusner, B. Paige, and J. Grammar Hernández-Lobato. Grammar variational autoencoder. *ArXiv*, 2017.
- [28] D. Rumelhart, G. Hinton, and R. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

- [29] S. Hochreiter and S. Schmidhuber. Long short-term memory. *Neural Computation*, 9:533–536, 1997.
- [30] C. Olah. Understanding lstm networks. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>, August 27 2015. [Online; accessed 12-December-2023].
- [31] I. Suskever, J. Martens, and G. Hinton. Generating text with recurrent neural networks. ICML 2011, pages 1017–1024.
- [32] D. Gillick, C. Brunk, O. Vinayals, and A. Subramanya. Multilingual language processing from bytes. *ArXiv*, 2016.
- [33] L. Deng and D. Yu. Deep learning – methods and applications. *Foundations and Trends in Signal Processing*, 2014.
- [34] Y. Lecun, P. Haffner, and Y. Bengio. Object recognition with gradient-based learning. *ATT Shannon Lab, 100 Schultz Drive, Red Bank NJ 07701, USA*, 2000.
- [35] A.V. Oppenheim, A.S. Willsky, and I.T. Young. *Signals and Systems*. Prentice-Hall, 1983.
- [36] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Proc. Advances in Neural Information Processing Systems*, volume 25, pages 1090–1098, 2012.
- [37] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. volume 86 of *Proceedings of the IEEE*, 1998.
- [38] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional models for semantic segmentation. In *CVPR 2015*, 2015.
- [39] J. Bromley, I. Guyon, Yann Y. LeCun, E. Säckinger, and R. Shah. Signature verification using a “siamese” time delay neural network. In *Proceedings of the 6th International Conference on Neural Information Processing Systems, NIPS’93*, page 737–744, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc.
- [40] G. Koch, R. Zemel, and R. Salakhutdinov. Siamese neural networks for one-shot image recognition. In *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, 2015.

- [41] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [42] OpenAI. Gpt-4 technical report. *ArXiv*, 2023.
- [43] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. Hierarchical text-conditional image generation with clip latents. *ArXiv*, abs/2204.06125, 2022.
- [44] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521:436–444, 2015.
- [45] M.M. Islam, T. Nasrin, B.A. Walther, C-Wu, H-C Yang, and Y-C Li. Prediction of sepsis patients using machine learning approach: A meta-analysis. *Computer Methods and Programs in Biomedicine*, 170:1–9, 2019.
- [46] K.E. Henry, D.N. Hager, P. J. Pronovost, and S. Saria. A targeted real-time early warning score (trewscore) for septic shock. *Science translational medicine*, 7, 2015.
- [47] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, and R. Das. Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach. *JMIR Medical Informatics*, 4(3):e28, 2016.
- [48] A. Mitra and K. Ashraf. Sepsis prediction and vital signs ranking in intensive care unit patients. *ArXiv*, 2018.
- [49] C. Barton, U. Chettipally, Y. Zhou, Z. Jiang, A. Lynn-Palevsky, S. Le, J. Calvert, and R. Das. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Computers in Biology and Medicine*, 109:79–84, 2019.
- [50] J. Fagerström, M. Bång, D. Wilhelms, and et al. Liseplstm: A machine learning algorithm for early detection of septic shock. *Sci Rep*, 9, 2019.
- [51] Q. Mao, M. Jay, J. L. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, and R. Das. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and icu. *BMJ Open*, 8(1), 2018.
- [52] A. Roederer, J. Weimer, J. Dimartino, J. Gutsche, and I. Lee. Towards non-invasive monitoring of hypovolemia in intensive care patients. In *6th Workshop on Medical Cyber-Physical Systems (MedicalCPS 2015)*, 2015.

- [53] E. Tsur, M. Last, V. F. Garcia, R. Udassin, M. Klein, and E. Brotfain. Hypotensive episode prediction in icus via observation window splitting. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 472–487, 2019.
- [54] A. Forkan, I. Khalil, and M. Atiquzzaman. Visibid: A learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data. *Computer Networks*, 113:244–257.
- [55] A. Meyer, D. Zverinski, B. Pfahringer, J. Kempfert, T. Kuehne, S. Sündermann, C. Stamm, T. Hofmann, V. Falk, and C. Eickhoff. Machine learning for real-time prediction of complications in critical care: a retrospective study. *The Lancet Respiratory Medicine.*, 6, 2018.
- [56] B. Fritz, Y. Chen, T. Murray-Torres, S. Gregory, A.B. Abdallah, A. Kronzer, S. McKinnon, T. Budelier, D. Helsten, T. Wildes, A. Sharma, and M. Avidan. Using machine learning techniques to develop forecasting algorithms for postoperative complications: Protocol for a retrospective study. *BMJ Open*, 8, 2018.
- [57] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun. Using recurrent neural network models for early detection of heart failure onset. *J. Am. Med. Inform. Assoc.*, 292:344–350, 2016.
- [58] A. Forkan and I. Khalil. A probabilistic model for early prediction of abnormal clinical events using vital sign correlations in home-based monitoring. In *Conference: 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–9, 2016.
- [59] B. Saugel, K. Kouz, P. Hoppe, K. Maheshwari, and T.W.L. Scheeren. Predicting hypotension in perioperative and intensive care medicine. *Best Practice and Research: Clinical Anaesthesiology.*, 2019.
- [60] F. Hatib, Z. Jian, S. Buddi, and et al. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology*, 2018.
- [61] H. Lindroth, L. Bratzke, S. Purvis, and et al. Systematic review of prediction models for delirium in the older adult inpatient. *BMJ Open*, 8, 2018.
- [62] T. Davoudi, A. Ozrazgat-Baslanti, A. C. Ebadi, A. Bihorac Bursian, and P. Rashidi. Delirium prediction using machine learning models on predictive electronic health

records data. In *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 568–573, Washington, DC, 2017.

- [63] A. Wong, A.T. Young, A.S. Liang, R. Gonzales, V.C. Douglas, and D. Hadley. Development and validation of an electronic health record–based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. *JAMA Netw Open*, 1(4), 2018.
- [64] H.N. Mufti, G.M. Hirsch, S.R. Abidi, and S.S.R. Abidi. Exploiting machine learning algorithms and methods for the prediction of agitated delirium after cardiac surgery: Models development and validation study. *JMIR Med. Inform.*, 7(4):e14993, October 2019.
- [65] J. Corradi, S. Thompson, J. Mather, C. Waszynski, and R. Dicks. Prediction of incident delirium using a random forest classifier. *Journal of Medical Systems*, 42, 2018.
- [66] B. Shickel, T.J. Loftus, L. Adhikari, and et al. Deepsofa: A continuous acuity score for critically ill patients using clinically interpretable deep learning. *Sci Rep*, 9:1879, 2019.
- [67] M.M. Churpek, T.C. Yuen, C. Winslow, D.O. Meltzer, M.W. Kattan, D.P. Edelson, and K. M.W. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Critical Care Medicine*, 44(2):368–374, 2016.
- [68] G.W. Colopy, M.A.F. Pimentel, S.J. Roberts, and D.A. Clifton. Bayesian gaussian processes for identifying the deteriorating patient. In *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, pages 5311–5314, 2016.
- [69] T.I.A.L. Al-Shwaheen and W. Hau. A new model for tracking and detection of deterioration of vital signs based on artificial neural network. *Journal of Theoretical and Applied Information Technology*, 31(14), 2019.
- [70] T. Ahmed, S. Ahmed, and F.E. Chowdhury. Taking meredith out of grey’s anatomy: Automating hospital icu emergency signaling. In *ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 1886–1890, 2016.

- [71] M. Guillame-Bert, A. Dubrawski, L. Chen, M. Hravnak, G. Clermont, and M. Pinsky. Forecasting escalation of cardio-respiratory instability using noninvasive vital sign monitoring data. *Intensive Care Medicine Experimental*, 3(S1), 2015.
- [72] T.J. Moss, D.E. Lake, J.F. Calland, K.B. Enfield, J.B. Delos, K.D. Fairchild, and J.R. Moorman. Signatures of subacute potentially catastrophic illness in the icu: Model development and validation. *Crit Care Med*, 44(9):1639–1648, 2016.
- [73] J.H. Yoon and M.R. Pinsky. Predicting adverse hemodynamic events in critically ill patients. *Current Opinion in Critical Care*, 24:196–203.
- [74] H. Du, M M. Ghassemi, and M. Feng. The effects of deep network topology on mortality prediction. In *In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2602–2605, 2016.
- [75] A. Awad, M. Bader-El-Den, J. McNicholas, and J. Briggs. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *International Journal of Medical Informatics*, 108, 2017.
- [76] M. Hsieh, M. Hsieh, C-M Chen, C-C Hsieh, C-M Chao, and C-C Lai. Comparison of machine learning models for the prediction of mortality of patients with unplanned extubation in intensive care units. *Scientific Reports*, 2018.
- [77] S.Y. Kim, S. Kim, J. Cho, and et al. A deep learning model for real-time mortality prediction in critically ill children. *Crit Care*, 23:279, 2019.
- [78] T. Navaneelan, S. Alam, P.A. Peters, and O. Phillips. Deaths involving sepsis in canada. *Statistics Canada*, 2016.
- [79] S.M. Lauritsen, M. Ellersgaard Kalør, E. Lund Kongsgaard, K. Meyer Lauritsen, M. Johansson Jørgensen, J. Lange, and B. Thiesson. Early detection of sepsis utilizing deep learning on electronic health record event sequences. *Artificial Intelligence in Medicine*, 104:101820, 2020.
- [80] M. Moor, N. Bennett, D. Plecko, M. Horn, B. Rieck, N. Meinshausen, P. Bühlmann, and K. Borgwardt. Predicting sepsis using deep learning across international sites: a retrospective development and validation study. *eClinicalMedicine*, 62:102124, 08 2023.
- [81] T.R. Spiller, E. Tufan, H. Petry, S. Böttger, S. Fuchs, O. Duek, Z. Ben-Zion, N. Korem, I. Harpaz-Rotem, R. von Känel, and J. Ernst. Delirium screening

in an acute care setting with a machine learning classifier based on routinely collected nursing data: A model development study. *Journal of Psychiatric Research*, 156:194–199, 2022.

- [82] M. Mulkey, H. Huang, T. Albanese, S. Kim, and B. Yang. Supervised deep learning with vision transformer predicts delirium using limited lead eeg. *Scientific Reports*, 13, 05 2023.
- [83] A. Ahmed, A. Garcia Agundez Garcia, I. Petrovic, F. Radaei, J. Fife, J. Zhou, H. Karas, S. Moody, J. Drake, R. Jones, C. Eickhoff, and M. Reznik. Delirium detection using wearable sensors and machine learning in patients with intracerebral hemorrhage. *Frontiers in neurology*, 14:1135472, 06 2023.
- [84] N.T. Liu, J.B. Holcomb, C.E. Wade, M.I. Darrah, and J. Salinas. Data quality of a wearable vital signs monitor in the pre-hospital and emergency departments for enhancing prediction of needs for life-saving interventions in trauma patients. *Journal of Medical Engineering and Technology*, 39(6):316–321, 2015.
- [85] S. Mehrdad, F.E. Shamout, Y. Wang, and S.F. Atashzar. Deep learning for deterioration prediction of covid-19 patients based on time-series of three vital signs. *Scientific reports*, 13(1):9968, June 2023.
- [86] S. Boulitsakis Logothetis, D. Green, M. Holland, and N. Al Moubayed. Predicting acute clinical deterioration with interpretable machine learning to support emergency care decision making. *Scientific Reports*, 13(1):13563, Aug 2023.
- [87] N. Tomašev, X. Glorot, J. Rae, M. Zielinski, H. Askham, A. Saraiva, A. Mottram, C. Meyer, S. Ravuri, I. Protsyuk, A. Connell, C. Hughes, A. Karthikesalingam, J. Cornebise, H. Montgomery, G. Rees, C. Laing, C. Baker, K. Peterson, and S. Mohamed. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*, 572:116, 08 2019.
- [88] S. Tonekaboni, M. Mazwi, P. Laussen, D. Eytan, R. Greer, S. Goodfellow, A. Goodwin, M. Brudno, and A. Goldenberg. Prediction of cardiac arrest from physiological signals in the pediatric icu. In *Proceedings of Machine Learning Research*, 2018.
- [89] H. Mohamdlou, S. Panchavati, J. S. Calvert, A. Lynn-Palevsky, C. W. Barton, G. Fletcher, and R. Das. Multicenter validation of a machine learning algorithm for 48 hour all-cause mortality prediction. *BioRxiv Preprint*, 872, 2018.

- [90] H-C Thorsen-Meyer, A.B. Nielsen, A.P. Nielsen, B.S. Kaas-Hansen, P. Toft, J. Schierbeck, T. Strøm, P.J. Chmura, M. Heimann, L. Dybdahl, L. Spangsege, P. Hulsen, K. Belling, S. Brunak, and A. Perner. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. *Lancet Digit Health*, 2(4):e179–e191, March 2020.
- [91] J. Allyn, N. Allou, P. Augustin, I. Philip, O. Martinet, M. Belghiti, S. Provenchère, P. Montravers, and C. Ferdynus. A comparison of a machine learning model with euroscore ii in predicting mortality after elective cardiac surgery: A decision curve analysis. *PLOS ONE*, 12, 2017.
- [92] L. Wang, L. Sha, J. Lakin, J. Bynum, D. Bates, P. Hong, and L. Zhou. Development and validation of a deep learning algorithm for mortality prediction in selecting patients with dementia for earlier palliative care interventions. *JAMA Netw Open*, 2(7), 2019.
- [93] W. Sun, S. Kalmady, N. Sepehrvand, A. Salimi, Y. Nademi, K. Bainey, J. Ezekowitz, R. Greiner, A. Hindle, F. McAlister, R. Sandhu, and P. Kaul. Towards artificial intelligence-based learning health system for population-level mortality prediction using electrocardiograms. *npj Digital Medicine*, 6:21, 02 2023.
- [94] G. Hinton and et al. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 29:82–97, 2012.
- [95] I. Sutskever, O. Vinyals, and Q. Le. Sequence to sequence learning with neural networks. In *Proc. Advances in Neural Information Processing Systems*, volume 27, pages 3104–3112, 2014.
- [96] D. Silver and et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.
- [97] B. Marr. First fda approval for clinical cloud-based deep learning in healthcare. <https://www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare/#4006934e161c>, January 1 2017. [Online; accessed 12-December-2023].
- [98] Z. Xiong, M. Stiles, and J. Zhao. Robust ecg signal classification for the detection of atrial fibrillation using novel neural networks. In *Computing in Cardiology Conference*, 2017.

- [99] S. Ghiasi, M. Abdollahpur, N. Madani, K. Kiyani, and A. Ghaffari. Atrial fibrillation detection using feature based algorithm and deep conventional neural network. In *Computing in Cardiology Conference*, 2017.
- [100] Z. Xiong, M. Nash, E. Cheng, V. Fedorov, M. Stiles, and J. Zhao. Ecg signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network. *Physiological Measurement*, August 2018.
- [101] M. Limam and F. Precioso. Atrial fibrillation detection and ecg classificaiton based on convolutional recurrent neural network. In *2017 Computing in Cardiology Conference*, 2017.
- [102] P. Warrick and M. Nabhan Homsy. Ensembling convolutional and long short-term memory networks for electrocardiogram arrhythmia detection. *Physiological Measurement*, 39(11), 2018.
- [103] R. He, K. Wang, N. Zhao, Y. Liu, Y. Yuan, Q. Li, and H. Zhang. Automatic detection of atrial fibrillation based on continuous wavelet transform and 2d convolutional neural networks. *Frontiers in Physiology*, 9, August 2018.
- [104] A. Hannun, P. Rajpurkar, M. Haghpanahi, G. Tison, C. Bourn, M. Turakhia, and A. Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25, 2019.
- [105] K. Ochiai, S. Takahashi, and Y. Fukazawa. Arrhythmia detection from 2-lead ecg using convolutional denoising autoencoders. In *KDD’18 Deep Learning Day*, London, UK, August 2018.
- [106] M. Zubair, J. Kim, and C. Yoon. An automated ecg beat classification system using convolutional neural networks. In *6th International Conference on IT Convergence and Security (ICITCS)*, pages 1–5. IEEE, 2016.
- [107] Y. Daniel, B. Thamineni, and A. Dawud. Cardiac arrhythmia detection using deep learning approach and time frequency representation of ecg signals. *BMC Medical Informatics and Decision Making*, 23, 10 2023.
- [108] O. Yildirim, M. Talo, E.J. Ciaccio, R.S. Tan, and U.R. Acharya. Accurate deep neural network model to detect cardiac arrhythmia on more than 10,000 individual subject ecg records. *Computer Methods and Programs in Biomedicine*, 197:105740, 2020.

- [109] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, and X. Liu. Ecg signal enhancement based on improved denoising auto-encoder. *Engineering Applications of Artificial Intelligence*, 52:194–202, 2016.
- [110] K. Antczak. Deep recurrent neural networks for ecg signal denoising. *ArXiv*, 2019.
- [111] P. Singh and G. Pradhan. A new ecg denoising framework using generative adversarial network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(2):759–764, 2021.
- [112] C. Han and L. Shi. Ml-resnet: A novel network to detect and locate myocardial infarction using 12 leads ecg. *Computer methods and programs in biomedicine*, 185:105138, 2020.
- [113] S. Gustafsson, D. Gedon, E. Lampa, A. Ribeiro, M.J. Holzmann, and T. B. Schön. Development and validation of deep learning ecg-based prediction of myocardial infarction in emergency department patients. *Scientific reports*, 12(1):19615, 2022.
- [114] S. Chugh, R. Haymoeller, K. Narayanan, D. Singh, M. Rienstra, E. Benjamin, R. Gillum, Y-H Kim, J. McAnulty, Z-J Zheng, M. Forouzanfar, M. Naghavi, G. Mensah, M. Ezzati, and C. Murray. Worldwide epidemiology of atrial fibrillation: A global burden of disease 2010 study. *Circulation*, 129(8):837–87, 2013.
- [115] Heart and Stroke Foundation of Canada. Atrial fibrillation. <https://www.heartandstroke.ca/heart/conditions/atrial-fibrillation>. [Online; accessed 12-December-2023].
- [116] CDC. Atrial fibrillation fact sheet. https://www.cdc.gov/dhdsdp/data_statistics/fact_sheets/fs_atrial_fibrillation.htm, August 13 2015. [Online; accessed 12-December-2023].
- [117] A. Fedotov. Selection of parameters of bandpass filtering of the ecg signal for heart rhythm monitoring systems. *Biomedical Engineering*, 50, 09 2016.
- [118] Z. Islam, G.M.S. Sajjad, H. Rahman, A. Dey, M. Biswas, and A. Hoque. Performance comparison of modified lms and rls algorithms in denoising of ecg signals. *International Journal of Engineering and Technology*, 2:466–468, 03 2012.
- [119] S. Pal and M. Mitra. Empirical mode decomposition based ecg enhancement and qrs detection. *Computers in Biology and Medicine*, 42(1):83–92, 2012.

- [120] M.A. Kabir and C. Shahnaz. Denoising of ecg signals based on noise reduction algorithms in emd and wavelet domains. *Biomedical Signal Processing and Control*, 7:481–489, 09 2012.
- [121] B.N. Singh and A.K. Tiwari. Optimal selection of wavelet basis function applied to ecg signal denoising. *Digit. Signal Process.*, 16:275–287, 2006.
- [122] M. Awal, M. Ahmad, I. Daut, E. Che Mid, and M. Rashid. Wavelet based distortion measurement and enhancement of ecg signal. In *2012 International Conference on Biomedical Engineering, ICoBE 2012*, 02 2012.
- [123] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *ArXiv*, 2014.
- [124] E.J. Benjamin, S.S. Virani, C.W. Callaway, and et al. Heart disease and stroke statistics 2018 at-a-glance: A report of the american heart association. *Circulation*, 137, 2018.
- [125] Z. Attia, P. Friedman, P. Noseworthy, F. Lopez-Jimenez, D. Ladewig, G. Satam, P. Pellikka, T. Munger, S. Asirvatham, C. Scott, R. Carter, and S. Kapa. Age and sex estimation using artificial intelligence from standard 12-lead ecgs. *Circulation: Arrhythmia and Electrophysiology*, 12.
- [126] E. Kanjo and et al. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Information Fusion*, 49:46–56, 2019.
- [127] P. Sarkar and A. Etemad. Self-supervised ecg representation learning for emotion recognition. *IEEE Transactions on Affective Computing*, PP:1–1, 08 2020.
- [128] A. Eduardo, H. Aidos, and A. Fred. Ecg-based biometrics using a deep autoencoder for feature learning - an empirical study on transferability. *ICPRAM*, pages 463–470, 2017.
- [129] Vibhav Agrawal, Mehdi Hazratifard, Haytham Elmiligi, and Fayez Gebali. Electrocardiogram (ecg)-based user authentication using deep learning algorithms. *Diagnostics*, 13, 2023.
- [130] M. Ghassemi, T. Naumann, P. Schulam, A.L. Beam, and R. Ranganath. Opportunities in machine learning for healthcare. *ArXiv*, 2018.

- [131] A. Taenzer, J. Pyke, M. Herrick, T. Dodds, and S. McGrath. A comparison of oxygen saturation data in inpatients with low oxygen saturation using automated continuous monitoring and intermittent manual data charting. *Anesthesia and analgesia*, 118, 2013.
- [132] D. Eytan, A. Jegatheeswaran, M. Mazwi, A. Assadi, A. Goodwin, R. Greer, S. Goodfellow, and P. Laussen. Temporal variability in the sampling of vital sign data limits the accuracy of patient state estimation. *Pediatric Critical Care Medicine*, 20, 2019.
- [133] H.R. Tomlinson, M.A.F. Pimentel, S. Gerry, D.A. Clifton, L. Tarassenko, and P.J. Watkinson. Smoothing effect in vital sign recordings: Fact or fiction? a retrospective cohort analysis of manual and continuous vital sign measurements to assess data smoothing in postoperative care. *Anesthesia and Analgesia*, 127(4):960–966, 2018.
- [134] G.B. Moody and R.G. Mark. The impact of the mit-bih arrhythmia database. *IEEE Eng in Med and Biol*, 20(3):45–50, 2001.
- [135] G.B. Moody, W.E. Muldrow, and R.G. Mark. A noise stress test for arrhythmia detectors. *Computers in Cardiology*, 11:381–384, 1984.
- [136] P. Mehrgardt, M. Khushi, S. Poon, and A. Withana. Pulse transit time ppg dataset (version 1.0.0). *PhysioNet*, 2022.
- [137] R. Bousseljot, D. Kreiseler, and A. Schnabel. Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. 40(s1):317–318, Dec 1995.
- [138] J.J. Akhtar and M.Y. Younus Javed. Image compression with different types of wavelets. In *IEEE—ICET 2006 2nd International Conference on Emerging Technologies*, Peshawar, Pakistan, 13-14 November 2006.
- [139] H.T. Wei and V. Jeoti. A wavelet footprints-based compression scheme for ecg signals. In *2004 IEEE Region 10 Conference TENCN*, page 283, 2004.
- [140] Z. Xizhi. The application of wavelet transform in digital image processing. In *MMIT '08 Proceedings of the 2008 International Conference on MultiMedia and Information Technology*, pages 326–329, December 30-31, 2008 208.
- [141] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.

- [142] P.R. Gomes, F.O. Soares, J. H. Correia, and C.S. Lima. Ecg data-acquisition and classification system by using wavelet-domain hidden markov models. In *Proceedings of the Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference*, pages 4670–4673, Buenos Aires, 2010.
- [143] J. Gao, H. Zhang, P. Lu, and Z. Wang. An effective LSTM recurrent network to detect arrhythmia on imbalanced ECG dataset. *J Healthc Eng.* 2019 Oct 13;2019:6320651. doi:, 10., October 2019.
- [144] S. Fujieda, K. Takayama, and T. Hachisuka. Wavelet convolutional neural networks. *ArXiv*, 2018.
- [145] S. Ross-Howe and H.R. Tizhoosh. The effects of image pre- and post-processing, wavelet decomposition, and local binary patterns on u-nets for skin lesion segmentation. In *IEEE Conference on Computational Intelligence (WCCI 2018)*, Rio de Janeiro, Brazil, 8-13 July 2018.
- [146] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [147] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing, 2015.
- [148] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *ArXiv*, 2016.
- [149] X. Mao, Q. Li, H. Xie, R.Y.K. Lau, Z. Wang, and S.P. Smolley. Least squares generative adversarial networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2813–2821, 2016.
- [150] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems. <https://www.tensorflow.org/>, 2015. Software available from tensorflow.org.

- [151] F. Chollet et al. Keras. <https://keras.io>, 2015. [Online; accessed 12-December-2023].
- [152] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O’Leary. PyWavelets: A python package for wavelet analysis. *J. Open Source Softw.*, 4(36):1237, April 2019.
- [153] G. van Rossum. Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.
- [154] S. Ari, M.K. Das, and A. Chacko. Ecg signal enhancement using s-transform. *Computers in Biology and Medicine*, 43(6):649–660, 2013.
- [155] G. Reddy, M. Muralidhar, and V. Sourirajan. Ecg denoising using improved thresholding based on wavelet transforms. *International Journal of Computer Science and Network Security*, 9, 01 2009.
- [156] F. Oliver, Y. Hagiwara, T.J. Hong, O.S. Lih, and U.R. Acharya. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, 161:1–13, 2018.
- [157] S. Ross-Howe and H.R. Tizhoosh. Atrial fibrillation detection using deep features and convolutional networks. In *IEEE International Conference on Biomedical and Health Informatics (BHI’19)*, Chicago, USA, May 19-22 2019.
- [158] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, volume 37, 2015.
- [159] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio. Attention-Based Models for Speech Recognition. *ArXiv*, page arXiv:1506.07503, June 2015.
- [160] M.T. Luong, H. Pham, and C. D. Manning. Effective Approaches to Attention-based Neural Machine Translation. *ArXiv*, page arXiv:1508.04025, August 2015.
- [161] Y. Liu, R. He, K. Wang, Q. Li, Q. Sun, N. Zhao, and H. Zhang. Automatic detection of ecg abnormalities by using an ensemble of deep residual networks with attention. *ArXiv*, 2019.
- [162] S. Mousavi, F. Afghah, A. Razi, and U.R. Acharya. Ecgnet: Learning where to attend for detection of atrial fibrillation with deep visual attention. *ArXiv*, 2019.

- [163] Y. Liu, J. Cui, and W. Liang. A hybrid learning-based stochastic noise eliminating method with attention-Conv-LSTM network for low-cost MEMS gyroscope. *Front Neurorobot*, 16:993936, December 2022.
- [164] G. Xu, T. Ren, Y. Chen, and W. Che. A one-dimensional cnn-lstm model for epileptic seizure recognition using eeg signal analysis. *Frontiers in Neuroscience*, 14, 12 2020.
- [165] World Health Organization. Indicator metadata registry details. <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/3155>, 21 November 2012 2012. [Online; accessed 12-December-2023].
- [166] B. Zhou and et al. Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *The Lancet*, 398:957 – 980, 2021.
- [167] Public Health Agency of Canada. High blood pressure. <https://www.canada.ca/en/public-health/services/chronic-diseases/cardiovascular-disease/hypertension.html>, Oct 2013. [Online; accessed 12-December-2023].
- [168] Mayo Foundation for Medical Education and Research. How high blood pressure can affect the body. <https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/high-blood-pressure/art-20045868>, Jan 2022. [Online; accessed 12-December-2023].
- [169] G. Ogedegbe and T. Pickering. Principles and techniques of blood pressure measurement. *Cardiology Clinics*, 28(4):571–586, 2010.
- [170] E. O’Brien, B. Waeber, G. Parati, J. Staessen, and M. Myers. Blood pressure measuring devices: Recommendations of the european society of hypertension. *BMJ (Clinical research ed.)*, 322:531–6, 04 2001.
- [171] Y.T. Zhang, Y.L. Zheng, W.H. Lin, H.Y. Zhang, and X.L. Zhou. Challenges and opportunities in cardiovascular health informatics. *IEEE Trans. Biomed. Eng.*, 60(3):633–642, Mar 2013.
- [172] R.A. Phillips and J.A. Diamond. Ambulatory blood pressure monitoring and echocardiography–noninvasive techniques for evaluation of the hypertensive patient. *Progress in Cardiovascular Diseases*, 41(6):397–440, 1999 May-Jun.

- [173] J.W. Remington. The physiology of the aorta and major arteries. *Handb, Physiol.*, pages 799–838, 1961.
- [174] W. Chen, T. Kobayashi, S. Ichikawa, Y. Takeuchi, and T. Togawa. Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration. *Medical and Biological Engineering and Computing*, 38(5):569–574, 2000.
- [175] D.J. Hughes, C.F. Geddes, L.A. Geddes, and J.D. Bourland. Measurement of young’s modulus of elasticity of the canine aorta with ultrasound. *Ultrasonic Imaging*, 1:799–838, 1979.
- [176] C.C.Y. Poon and Y.T. Zhang. Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time. In *Proc. 27th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.(Shanghai, China)*, pages 5877–5880, 2006.
- [177] Y. Zheng, B. Yan, Y-T Zhang, and C. Poon. An armband wearable device for overnight and cuff-less blood pressure measurement. *IEEE Transactions on Biomedical Engineering*, 04 2014.
- [178] X.R. Ding, Y.T. Zhang, J. Liu, W.X. Dai, and H. K. Tsang. Continuous cuffless blood pressure estimation using pulse transit time and photoplethysmogram intensity ratio. *IEEE Trans. Biomed. Eng.*, 63(5):964–972, May 2016.
- [179] X. Ding, B.P. Yan, Y-T Zhang, J. Liu, N. Zhao, and H.K. Tsang. Pulse transit time based continuous cuffless blood pressure estimation: A new extension and a comprehensive evaluation. *Scientific Reports*, 7(1), 2017.
- [180] F. Miao, N. Fu, Y t Zhang, X. Ding, X. Hong, Q. He, and Y. Li. A novel continuous blood pressure estimation approach based on data mining techniques. *IEEE Journal of Biomedical and Health Informatics*, 21:1730–1740, 2017.
- [181] X. Xing and M. Sun. Optical blood pressure estimation with photoplethysmography and fft-based neural networks. *Biomed. Opt. Express*, 7(8):3007–3020, Aug 2016.
- [182] J. Kyung, J-Y Yang, J-H Choi, J-H Chang, S. Bae, J. Choi, and Y. Kim. Deep-learning-based blood pressure estimation using multi channel photoplethysmogram and finger pressure with attention mechanism. *Scientific Reports*, 13, 06 2023.
- [183] R. Block, M. Yavarimanesh, K. Natarajan, A. Carek, A. Mousavi, A. Chandrasekhar, C-S Kim, J. Zhu, G. Schifitto, L. Mestha, O. Inan, J-O Hahn, and R. Mukkamala.

Conventional pulse transit times as markers of blood pressure changes in humans. *Scientific Reports*, 10, 10 2020.

- [184] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, Í. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [185] J. Bromley, J. Bentz, B.W. Leon, I. Guyon, Y. LeCun, C. Moore, E. Sackinger, and R. Shah. Signature verification using a siamese time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(4):669–688, 1993.
- [186] M. Li, K. Chang, B. Bearce, C. Chang, A. Huang, J. Campbell, J. Brown, P. Singh, K. Hoebel, D. Erdogmus, S. Ioannidis, W. Palmer, M. Chiang, and J. Kalpathy-Cramer. Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging. *npj Digital Medicine*, 3, 12 2020.
- [187] Public Health Agency of Canada. Heart disease in canada: Highlights from the canadian chronic disease surveillance system. <https://www.canada.ca/en/public-health/services/publications/diseases-conditions/heart-disease-canada-fact-sheet.html>, 2017. [Online; accessed 12-December-2023].
- [188] F. Morris. ABC of clinical electrocardiography: Acute myocardial infarction—part I. *BMJ*, 324(7341):831–834, April 2002.
- [189] P. Wagner, N. Strodthoff, R-D Bousseljot, D. Kreiseler, F.I. Lunze, W. Samek, and T. Schaeffter. PTB-XL, a large publicly available electrocardiography dataset. *Sci. Data*, 7(1):154, May 2020.
- [190] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C.h. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C-K Peng, and H.E. Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(2), 2000.

- [191] S.F. Sanders and H.A. DeVon. Accuracy in ed triage for symptoms of acute myocardial infarction. *Journal of emergency nursing*, 42(4):331–337, Dec 2016.
- [192] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *ArXiv*, abs/1706.03762, 2017.
- [193] A.E.W. Johnson, T.J. Pollard, L. Shen, L. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, LA Celi, and RG Mark. Mimic-iii, a freely accessible critical care database. scientific data. 2016.
- [194] R.K. Chakraborty and B. Burns. Systemic inflammatory response syndrome. *StatPearls Publishing*, May 2023.
- [195] S. Majumder, T. Mondal, and M.J. Deen. Wearable sensors for remote health monitoring. *Sensors*, 17, 2017.
- [196] R. Villar and et al. Validation of the hexoskin wearable vest during lying, sitting, standing, and walking activities. *Applied Physiology, Nutrition, and Metabolism*, 40(10):1019–1024, 2015.
- [197] T. Beltrame, R. Amelard, A. Wong, and R.L. Hughson. Extracting aerobic system dynamics during unsupervised activities of daily living using wearable sensor machine learning models. *Journal of applied physiology*, 124(2):473–481, 2018.
- [198] J. Oh, D. Cho, J. Park, S. Na, J. Kim, J. Heo, C. Shin, J-J Kim, J. Park, and B. Lee. Prediction and early detection of delirium in the intensive care unit by using heart rate variability and machine learning. *Physiological Measurement*, 39, 2018.
- [199] W. Ge, J-W Huh, Y.R. Park, J.H. Lee, Y.H. Kim, and A. Turchin. An interpretable icu mortality prediction model based on logistic regression and recurrent neural networks with lstm units. In *Annual Symposium proceedings*, pages 460–469, AMIA Symposium, 2018. AMIA.
- [200] A. Johnson and R. Mark. Real-time mortality prediction in the intensive care unit. In *Annual Symposium proceedings*, AMIA Symposium, pages 994–1003. AMIA, 2018.
- [201] B. Rush, L.A. Celi, and D.J. Stone. Applying machine learning to continuously monitored physiological data. *Journal of Clinical Monitoring and Computing*, 2018.
- [202] S. Yang, L.G. Stansbury, P. Rock, T. Scalea, and P.F. Hu. Linking big data and prediction strategies: Tools, pitfalls, and lessons learned. *Critical Care Medicine*, 47(6):840–848, 2019.

- [203] C.M. Ruminski, M.T. Clark, D. E. Lake, R.R. Kitzmiller, J. Keim-Malpass, M.P. Robertson, and J.F. Calland. Impact of predictive analytics based on continuous cardiorespiratory monitoring in a surgical and trauma intensive care unit. *Journal of Clinical Monitoring and Computing*, 33(4):703–711, 2019.
- [204] M. Churpek, R. Adhikari, and D. Edelson. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation*, 102, 2016.
- [205] K. Tateno and L. Glass. A method for detection of atrial fibrillation using rr intervals. *Computers in Cardiology*, 27:391–394, 2000.
- [206] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, and Y. Zhang. Automatic online detection of atrial fibrillation based on symbolic dynamics and shannon entropy. *Biomedical Engineering Online*, pages 13–18, 2014.
- [207] G. Moody and R. Mark. A new method for detection atrial fibrillation using r-r intervals. *Computers in Cardiology*, 10:227–230, 1983.
- [208] A. Goodwin, S. Goodfellow, D. Eytan, R. Greer, M. Mazwi, and P. Laussen. Classification of atrial fibrillation using multidisciplinary features and gradient boosting. In *2017 Computing in Cardiology Conference*, 2017.
- [209] P. Vincent, H. Larochelle, Y. Bengio, and P-A Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1096–1103, 2008.
- [210] S.S. Thakur, S.S. Abdul, H-Y.S. Chiu, R.B. Roy, P-Y Huang, S. Malwade, A.A. Nursetyo, and Y-C.J. Li. Artificial-intelligence-based prediction of clinical events among hemodialysis patients using non-contact sensor data. *Sensors*, 18:2833, 2018.
- [211] C-C Wu, W-D Hsu, Md. Islam, T. Poly, H-C Yang, P-A Nguyen, Y-C Wang, and Y-C Li. An artificial intelligence approach to early predict non-st-elevation myocardial infarction patients with chest pain. *Computer Methods and Programs in Biomedicine*, page 173, 2019.
- [212] H.R. Tizhoosh. Opposition-based learning: A new scheme for machine intelligence. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC’06)*, volume 1, pages 695–701, 2005.

- [213] E. Ng, Y. Chengyang, T.O. Shultz, S. Ross-Howe, and S. Wang. Magneto-nanosensor smartphone platform for the detection of hiv and leukocytosis at point-of-care. *Nanomedicine: Nanotechnology, Biology, and Medicine*, 16:10–19, 2019.
- [214] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger. Densely connected convolutional networks. In *CVPR*, volume 1, page 3, 2017.
- [215] O. Russakovsky*, J. Deng*, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei (* = equal contribution). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015.
- [216] Pedregosa and et al. Scikit-learn: Machine learning in python. *JMLR*, 12:2825–2830, 2011.
- [217] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In *3rd International Conference for Learning Representations*, San Diego, 2015.
- [218] JM. Lilly and S.C. Olhede. Generalized morse wavelets as a superfamily of analytic wavelets. *IEEE Transactions on Signal Processing*, 60(11):6036–6041, 2012.
- [219] The MathWorks Inc. Matlab version 9.7.0.1190202 (r2019b), 2019.
- [220] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. *ArXiv*, 2015.
- [221] A.V. Oppenheim and R.W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [222] P. Liu, H. Zhang, W. Lian, and W. Zuo. Multi-level wavelet convolutional neural networks, 2019.
- [223] S. Santucci, E.M. Cates, G.D. James, and et al. A comparison of two ambulatory blood pressure monitors, the del mar avionics pressurometer iv and the spacelabs 90202. *Am J Hypertens*, 2(10):797–799, 1989.
- [224] W. Qiu, H. Chen, A. Dincer, S. Lundberg, M. Kaeberlein, and S-I Lee. Interpretable machine learning prediction of all-cause mortality. *Communications Medicine*, 2:125, 10 2022.

- [225] M.D. Aldhoayan and Y. Aljubran. Prediction of ICU patients' deterioration using machine learning techniques. *Cureus*, 15(5):e38659, May 2023.
- [226] M.S. Rahman, C. Karmakar, I. Natgunanathan, J. Yearwood, and M. Palaniswami. Robustness of electrocardiogram signal quality indices. *Journal of The Royal Society Interface*, 19, 04 2022.
- [227] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek. Deep learning for ecg analysis: Benchmarks and insights from ptb-xl. *IEEE journal of biomedical and health informatics*, PP, 09 2020.

Appendices

Appendix A

Literature Review of Machine Learning Research for Detection and Prediction of Clinical Complications

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Henry et al.	2015	Sepsis Prediction	MIMIC-II	Shock index (HR/SBP), SBP, Time since first antibiotics, Urine 6hr, BUN/CR, GCS, Arterial pH, RR, PaO ₂ , BUN, Chronic liver disease and cirrhosis, HR, Hepatic SOFA, FiO ₂ , Cardiac surgery patient, Immunocompromised, SIRS, WBC, Hemotological malignancy, Chronic heart failure, Chronic organ insufficiency, Diabetes, Renal SOFA, Neurologic SOFA, Platelets, Metastatic carcinoma	Cox Proportional Hazard Model	Before onset AUC=0.83
Desautels et al.	2016	Sepsis Prediction	MIMIC-III	GCS, HR, RR, SpO ₂ , Temp, NIDiasABP, NISysABP, SysABP, DiasABP	Gradient Boosting Tree	Hour to onset Sensitivity=80%, Specificity=55%, AUROC=0.75 (SD 0.008)
Mitra et al.	2018	Sepsis Prediction	MIMIC-III	HR, RR, SpO ₂ , Temp, SBP, DBP	Logistic Regression, Random Forest, and XGBoost, feedforward neural network or Perceptron	4 hours to onset Ensemble of all ML models Sepsis: AUC=0.80 Severe: AUC=0.81 Shock: AUC=0.84
Barton et al.	2019	Sepsis Prediction	UCSF and BIDMC	HR, RR, SpO ₂ , Temp, SBP, DBP	Gradient Boosted Trees	48 hours to onset AUROC=0.83
Fagerström et al.	2019	Sepsis Prediction	MIMIC-III	Shock index (HR/SBP), SBP, Time since first antibiotics, Urine 6hr, BUN/CR, GCS, Arterial pH, RR, PaO ₂ , BUN, Chronic liver disease and cirrhosis, HR, Hepatic SOFA, FiO ₂ , Cardiac surgery patient, Immunocompromised, SIRS, WBC, Hemotological malignancy, Chronic heart failure, Chronic organ insufficiency, Diabetes, Renal SOFA, Neurologic SOFA, Platelets, Metastatic carcinoma	LSTM	48 hours to onset AUROC=0.88
Lauritsen et al.	2020	Sepsis Prediction	Data collected for admissions to ICU and other health wards in four Danish municipalities (Odder, Hedensted, Skanderborg, and Horsens) from 2010 to 2017.	Electronic Health Record admission and medication events, laboratory results, medical imaging, national patient register for patient admissions and diagnosis, and civil registration system for sociodemographic information	Gradient boosting, Multilayer Perceptron, and CNN-LSTM	AUROC 0.856 at 3 hours to onset. AUROC 0.756 at 24 hours to onset.

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Moor et al.	2023	Sepsis Prediction	Four large databases collected from ICU admissions in the US, the Netherlands, and Switzerland between 2001 and 2016	Vital measurements, laboratory results, and static patient variables	A deep network with dense outer layers and causal self-attention inner layers	AUROC of 0.846 (95% confidence interval [CI], 0.841–0.852) internal site validation and an AUROC of 0.807 (95% CI, 0.801–0.813) across sites with transfer learning.
Davoudi et al.	2017	Delirium Prediction	University of Florida EHR data	70 preoperative features in the HER, including: demographics, socioeconomic, administrative, clinical, pharmacy, and laboratory variables	Logistic regression (LR), generalized additive model (GAM), random forests (RF), support vector machine (SVM), neural network (NN), extreme gradient boosting (XGB), Naïve Bayes (NB)	GAM model accuracy=0.80, AUC=0.86, sensitivity=0.75, and specificity=0.81
Oh et al.	2018	Delirium Prediction	Custom data set collected at a University Hospital in Korea	Mean HR, Mean RR, SDNN, SDSD, RMSSD, pNN20, pNN50, nLF Average of normalized low frequency component (0.04–0.15 Hz), nHF Average of normalized high frequency component (0.15–0.4 Hz), LF/HF, ApEn15, ApEn20, SampEn15, SampEn20, SD1, SD2, SD1/SD2, TTNN	Linear support vector machine (SVM), SVM with radial basis function (RBF) kernels, linear extreme learning machine (ELM), ELM with RBF kernels, linear	71.49% accuracy in patient-based classification using SVM with RBF kernels
Wong et al.	2018	Delirium Prediction	UCSF EHRs	796 clinical variables including: admission diagnoses, medications, laboratory values, vital signs, and demographic and nursing data obtained during the admission assessment (i.e., mobility, visual and hearing function, Glasgow Coma Scale, lines and tubes)	Penalized logistic regression (GLMNET), gradient machine boosting (GBM), artificial neural network with a single hidden layer (NN), linear support vector machine (e1071), and random forest (RF).	GBM performed the best with AUC of 0.855. Setting specificity at 90% the sensitivity was 59.7%
Corradi et al.	2018	Delirium Prediction	Hartford Hospital HER (N=64038)	Demographic data, comorbidities, medications, procedures, and physiological measures	Random Forest (RF).	AUROC= 0.909

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Mufti et al.	2019	Delirium Prediction	Maritime Heart Center (MHC) (N=5584)	Pre-operative variables	logistic regression, neural networks (ANN), support machines (SVM), Bayesian belief networks (BBN), naïve Bayes, random forest, and decision trees	ANN performed the best with AUROC= 0.782
Spiller et al.	2022	Delirium Prediction	Adults hospitalized between 2014-2018 at a single tertiary teaching hospital in Zurich, Switzerland (N=48,840)	Demographic data, as well as 56 item nursing care assessment that measures the abilities and impairments of patients in areas such as level of consciousness and motor skills	Non-regularized logistic regression (LR; glm), penalized logistic regression (PLR; glmnet), random forest (RF; randomForest), linear support vector machine (SVM; e1071), gradient boosting machine (GBM; gbm)	Gradient boosting machine performed the best with an AUC of 0.933 (95% CI, 0.929 - 0.936)
Mulkey et al.	2023	Delirium Prediction	Adults admitted to three intensive care units in a large rural academic medical center in North Carolina between 2019-2020	EEG waveforms	Random forest, stepwise discriminant analysis, support vector machines, and Vision Transformer	Vision Transformer model produced a 97% accuracy for predicting Delirium
Ahmed et al.	2023	Delirium Prediction	Patients with moderate-to-severe acute intracerebral hemorrhage (ICH) admitted to Neurocritical Care and Stroke Units at an academic medical center (N=39).	Utilized EMR data as well as actigraphy information from wrist-mounted sensors	Random Forest, SVM and XGBoost	Activity data increase the performance of the machine learning models (Random Forest, SVM and XGBoost), to an overall accuracy 74% (10%) (mean (SD)), and F1 score 65% (10%)

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Guillame-Bert et al.	2015	Patient Deterioration	Step Down Unit (SDU) N=156	HR, RR, SpO2 samples every 20s and BP every 2 hours	N/A	AUC 0.801 and specificity to 100% with a sensitivity of 37%
Liu et al.	2015	Patient Deterioration	University of Texas Health Science Center in Houston (N=104)	Heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP), respiratory rate (RR), blood oxygenation (SpO2), shock index (SI=HR/SBP) and pulse pressure (PP=SBP -DBP). All demographic information, physical exam results, Glasgow coma scale (GCS) scores (Motor, Verbal, Eye) and associated Data Quality Indexes (DQI) to detect field and ED Life Saving Interventions	Multivariate Logistic Regression	AUC 0.99
Moss et al.	2016	Patient Deterioration	University of Virginia Health System (N=9232)	Heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), respiratory rate (RR), blood oxygenation (SpO2) taken every 2 seconds. Mean and STDEV of HR, RR, SpO2 every 30 minutes, and HR variability (HRV), detrended fluctuation analysis (DFA), and the coefficient of sample entropy (COSEn) from ECG.	Binary Logistic Regression	Prediction up to 24 hours in advanced with C-statistics of 0.61 to 0.88
Colopy et al.	2016	Patient Deterioration	Surgical-trauma SDU at the University of Pittsburgh Medical Center Presbyterian Hospital N=333	HR, SBP, DBP, RR, SpO2	Gaussian Regression (GPR)	6-8 hours advanced warning with a clinically viable 10%-15% false positive rate
Ahmed et al.	2016	Patient Deterioration	ICU data from the National Heart Foundation Hospital and Research Institute (N=3)	37 vital statistics including: HR, PR, SBP, DBP, MAP, Body Temp, SYS Pulmonary Arterial Pressure, DIA Pulmonary Arterial Pressure, Mean Pulmonary Arterial Pressure, Oxygen and Carbon Dioxide Pressure, Sodium, Potassium, and Chloride Ion Concentration Levels.	Kernel-based Feature Transformation	Demonstrated high detection metric for all of the three studied patients and associated anomalous events

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Churpek et al.	2017	Patient Deterioration	Data from five hospitals, which included a tertiary academic center, two suburban teaching hospitals, and two community hospitals affiliated with the University of Chicago N=269999	Demographic variables, laboratory values, and vital signs to predict cardiac arrest, intensive care unit transfer, or death	Logistic Regression (LR), Random Forest (RF), K Nearest Neighbour (KNN), SVM, ANN	Random Forest model achieved the best results at 0.80 AUC
Shickel et al.	2019	Patient Deterioration	UFHealth (N=27660) MIMIC (N=35993)	13 variables representing six different organ systems (cardiovascular, respiratory, nervous, liver, coagulation, and renal) and uses their worst measurements over a given interval (typically 24 hours) in conjunction with static value thresholds to assign numerical scores for each component	Recurrent Neural Network (RNN)	AUC for entire ICU stay of 0.90
Ibrahim et al.	2019	Patient Deterioration	MIMIC-II (N=14)	HR, SBP, DBP, Body Temperature, SpO2	Levenberg-Marquardt ANN	R ² overall of 0.95
Boulitsakis Logothetis et al.	2022	Patient Deterioration	Unplanned admissions to Salford Royal Hospital, UK (N=118886)	Critical events including outcomes, vitals, supplemental observations, labs, and service utilization	Logistic regression (LR), gradient-boosted decision trees (GBDT), and support vector machines (SVM)	LightGBM model achieved an AUROC of 0.92 for a 24 hour prediction horizon.
Aldhoayan et al.	2023	Patient Deterioration	A retrospective review of patients' charted data from ICU at a tertiary hospital in 2019 (N=653)	blood pressure, oxygen saturation, temperature and heart rate	logistic regression, support vector machine classifier, k-nearest neighbours (KNN), gradient boosting classifier, and Naïve Bayes classifier	The gradient boosting classifier was able to predict the patients' deterioration or survival with a 88.43% accuracy.
Mehrdad et al.	2023	Patient Deterioration	Data collected from COVID-19 patients at NYU Langone Health in New York, USA (N=37006)	Routine triadic vital signs, oxygen saturation, heart rate, and temperature.	LSTM and residual temporal convolutional network (TCN)	The TCN network achieved an AUROC of 0.8844-0.9336 for 3 to 24 hour prediction horizon.

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Choi et al.	2016	Adverse Events: Heart Failure	Sutter Palo Alto Medical Foundation (Sutter-PAMF) primary care patients (N=3884) HF incident cases and N=28 903 controls	Demographics, tobacco and alcohol consumption, clinical and laboratory values, International Classification of Disease version 9 (ICD-9) codes associated with encounters, orders, and referrals, procedure information in Current Procedural Terminology (CPT) codes, and medication prescription information	Recurrent Neural Network (RNN), Logistic Regression multilayer perceptron (MLP), Support Vector Machine (SVM), and K-nearest neighbour (KNN)	The RNN model demonstrated an AUC of 0.777 at 12 months to onset
Forkan et al.	2016	Adverse Events: Hypertension, Hypotension, Tachycardia, Bradycardia, Bradypnea, Hypoxia	MIMIC-II (N=1000) records	Heart rate (HR), systolic, diastolic and mean blood pressure (SBP, DBP, MBP), respiratory rate (RR) and blood oxygen saturation (SpOx)	Hidden Markov Model (HMM)	97.8% accuracy over five class combinations of the five adverse events
Forkan et al.	2017	Adverse Events: Hypotension	MIMIC-II (N=5266) records	heart rate (HR), systolic, diastolic and mean blood pressure (SBP, DBP, MBP), respiratory rate (RR) and blood oxygen saturation (SpOx)	J48 decision tree (J48), Naive Bayes (NB), Bayes Net (BN), Multi Layer Perceptron (MLP), Sequential Minimum Optimization (SMO), Decision Table (DT), Random Forest (RF) and k-Nearest Neighbour (KNN)	The Random Classifier produced a 95.18% classification accuracy 120 minutes before onset

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Meyer et al.	2018	Adverse Events: Bleeding, Mortality, Renal Failure	MIMIC-III (N=9269)	Patient information: Age, sex, height, weight Information relating to initial surgery: Anaesthesia type, American Society of Anesthesiologists Score, cardioplegic solution, aortic cross-clamp time, cardiopulmonary bypass time, anaesthetic monitoring time, surgery duration, surgery type, urgency Vital signs: Systolic, mean, and diastolic arterial pressure; systolic, mean, and diastolic pulmonary artery pressure; central venous pressure; ventilator FiO2 setting; heart and respiratory frequency; body temperature Arterial blood gas: Bicarbonate, glucose, haemoglobin, oxygen saturation, partial pressure of carbon dioxide and oxygen, pH level, potassium, sodium Laboratory results: Albumin, bilirubin, urea, C-reactive protein, creatine kinase, -glutamyltransferase glutamic oxaloacetic transaminase, haemoglobin, haematocrit, international normalised ratio, creatinine, white blood cell count, lactate dehydrogenase, magnesium, partial thromboplastin time, platelets, prothrombin time Balance output: Bleeding rate, urine flow rate	Recurrent Neural Network (RNN)	PPV 0.90 and sensitivity 0.85 for mortality, 0.87 and 0.94 for renal failure, and 0.84 and 0.74 for bleeding
Hatib et al.	2019	Adverse Events: Hypotension	Edwards Database (N=1358) for training, MIMIC-II Database (N=326) for training, University of California at Irvine Medical Center (N=204) for validation	51 features extracted from the arterial blood pressure waveform. This was separated into beats and each beat into five phases (systolic phase, diastolic phase, systolic rise phase, systolic decay phase, and overall decay phase). For each phase, several features were analyzed and calculated: arterial blood pressure waveform time, amplitude, area, and slope features; FloTrac algorithm features (FloTrac; Edwards Lifesciences, Irvine, CA); CO-Trek features obtained from the ClearSight system (Edwards Lifesciences); complexity features; baroreflex features; variability features; spectral features; "delta change" features and combinatorial features	Logistic Regression	The model predicted a hypotension event 15 minutes before onset from features extracted from the arterial pulse waveform with a sensitivity of 83.6%, specificity of 83.3%, and AUC=0.91 on an external validation set

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Tsur et al.	2019	Adverse Events: Hypotension	Soroka Medical Center in Beer Sheva (N=4757), Hadassah Hospital, Ein-Karem, Jerusalem (N=8366), and MIMIC II (N=5266)	Vital Sign: SBP, DBP, and HR, Pulse Pressure (PP = SBP - DBP), Relative Cardiac Output (CO=HR*PP) sampled at 1 measurement per minute Statistical Features: Mean, median, standard deviation (Std), variance, interquartile range (Iqr), skewness, kurtosis and linear regression slope are calculated for each of the vital signs Wavelet Features Cross-Correlation Features	Extreme Gradient Boosting (XGBoost)	AUC = 0.937 with a 240-minute observation window and then 60 minute gap window, and 30 minute target window
Tomašev et al.	2020	Adverse Events: Acute Kidney Injury	703,782 patients collected by the US Department of Veterans Affairs between 2011 and 2015	Outpatient visits and admissions, diagnosis codes, procedure codes, laboratory results, medications and prescriptions, orders, vital signs, health factors and note titles	Deep recurrent Network	The model achieved a 90.2% accuracy detection of all acute kidney injuries with a false alarm ratio of 2 to 1.
Han et al.	2020	Adverse Events: Myocardial Infarction	PTB (Physikalisch-Technische Bundesanstalt) database	ECG recordings captured at 1kHz, 16-bit resolution, over a range of ± 16.384 mV	ResNet	Accuracy of 95.49%, sensitivity of 94.85%, and specificity of 97.37% on Myocardial Infarction detection
Tsur et al.	2022	Adverse Events: Myocardial Infarction	492,226 ECG recordings from 214,250 patients gathered in the emergency rooms of Sweden between 2007 and 2016	ECG waveforms and the patient's age and sex	ResNet (Convolutional Network with skip connections between the layers)	C-statistic of 0.991 and Brier score of 0.001 on ST-elevation myocardial infarction (STEMI), and a C-statistic of 0.832 and Brier score of 0.008 on non-ST-elevation myocardial infarction (NSTEMI)

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Du et al.	2016	Mortality Prediction	MIMIC-II (N=15647)	Demographic and Severity: age, gender, weight, critical service type (medical/surgical/cardiac/recovery), Simplified Acute Physiology Score (SAPS), Sequential Organ Failure Assessment (SOFA) Score, and Glasgow Coma Score (GCS) Co-morbidity: all 23 co-morbidities used to calculate the Elixhauser score Laboratory Results: Albumin, ALP (Alkaline phosphatase), ALT (Alanine transaminase, AST (Aspartate transaminase), Bilirubin, BUN (Blood urea nitrogen), Cholesterol, Creatinine, Glucose, HCO3 (Serum bicarbonate, HCT (Hematocrit), K (Serum potassium, Lactate, Mg (Serum magnesium), Na (Serum sodium), PaCO2 (partial pressure of arterial CO2, PaO2 (Partial pressure of arterial O2, pH, Platelets, Troponin, WBC (White Blood Cell count), Vital Signs (Median Values of first 24 hours): Heart Rate, Temperature, Mean Arterial Blood Pressure, Oxygen Saturation and Spontaneous Respiration Rate Interventions: Usage of mechanical Ventilation, Vasopressor and sedative medications	Deep Belief Network (DBN) with GA for optimal network topology, Gradient Boosting Model (GBM), SVM	DBN achieved a 98% prediction accuracy on 28-day mortality
Allyn et al.	2017	Mortality Prediction	Paris University Hospital (N=6520)	Demographic Data: Age, Sex, Weight, Height, BMI Comorbidities Preoperative Treatment Laboratory Results Surgery Characteristics	Logistic Regression Gradient Machine Boosting (GBM), Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes Model (NBM)	Ensemble of ML methods produced an AUC of 0.795
Awad et al.	2017	Mortality Prediction	MIMIC II (N=11722)	Demographic Variables: Age Vital Signs: Minimum and Maximum of Heart Rate, SBP, Temp, RR Examination Variable: Glasgow Coma Scale (GCS) Lab Test Results Disease Variables	Random Forest, Decision Trees, Naive Bayes and rule-based Projective Adaptive Resonance Theory (PART)	Random Forest at 6-hour predictive window produced an AUC of 0.82

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Hsieh et al.	2018	Mortality Prediction	Chi-Mei Medical Center (N=341)	16 input features including subject sex, age, APACHE II scores, Glasgow Coma Scale (GCS), TISS scale, comorbidities, ICU length of stay in days, and hospital length of stay in days	Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR)	Random Forest achieved F1=0.860 Precision=0.882, and recall=0.850 and AUROC curve of 0.910 prediction of mortality for patients with unplanned extubation
Mohammadlou et al.	2018	Mortality Prediction	Stanford Medical Center (N=515452) University of California, San Francisco Medical Center (N=95869) University of Washington Medical Center (N=32936)	Heart rate, respiratory rate, peripheral oxygen saturation (SpO2), temperature, systolic blood pressure, diastolic blood pressure, and Glasgow Coma Score (GCS)	Extreme Gradient Boosting (XGBoost)	On the Stanford data set a 48-hour mortality prediction with sensitivity=0.798, specificity=0.960 and AUROC=0.948
Ge et al.	2018	Mortality Prediction	Asan Medical Center (AMC) (N=4896)	Diagnosis: Do Not Resuscitate, Disseminated intravascular coagulation, Gastroesophageal reflux with esophagitis, Heart failure, Chronic kidney disease (stage 4), Acidosis, Delirium, Drug rash, Aspiration pneumonia, Liver transplant, Aspiration pneumonia Medications: Prednisolone, Peramivir, Adenosine Vitals: Oxygen saturation, pulse, respiratory rate Laboratory Results: Arterial blood gas pH Demographics: Age (71-80)	Recurrent Neural Network (RNN) with Long-Short Term Memory (LSTM)	AUROC=0.7614 for predicting mortality during the ICU stay
Kim et al.	2019	Mortality Prediction	Severance Hospital (Seoul, Korea) (N=1445) Samsung Medical Center (N=364)	Vital signs: Systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MBP), heart rate (HR), respiratory rate (RR), peripheral capillary oxygen saturation (SpO2), and body temperature (BT) Demographic Data: patient age and body weight	Convolutional Neural Network (CNN), Gradient Boosting Trees Long Short-Term Memory (LSTM)	CNN achieved an AUROC of 0.881 at 60 hours predictive window

<i>Authors</i>	<i>Year</i>	<i>Application</i>	<i>Data Set</i>	<i>Features</i>	<i>Machine Learning Model</i>	<i>Performance</i>
Wang et al.	2019	Mortality Prediction	Partners HealthCare System, an integrated care delivery system in Boston, Massachusetts (N=26921)	Patient demographic data and longitudinal clinical notes	Natural Language Processing (NLP) and Long Short-Term Memory (LSTM)	LSTM model achieved an AUC of 0.943 at a 2-year predictive window
Thorsen-Meyer et al.	2020	Mortality Prediction	National Health and Nutrition Examination Survey (NHANES) data set (N=47261 samples)	Demographic, laboratory, examination, and questionnaire features for a total of 151 features	Gradient Trees Boosted	Mortality risk calculated at a time horizon of 0-72 hours and at discharge. The network achieved an AUROC of 0.88 (95% CI 0.87-0.89) at discharge.
Qiu et al.	2022	Mortality Prediction	Data from patients admitted to four ICUs in the Capital Region, Denmark, between 2011 and 2016 (N=15615 ICU admissions of 12616 patients)	Demographic and high-frequency data collected from medical equipment (e.g. monitors, ventilators, and infusion pumps). Nationwide registry containing procedures and diagnosis codes and laboratory results.	LSTM	One, three, five, and ten year mortality risk calculated. The five-year mortality risk was on the model using all features (AUROC=0.9030)
Sun et al.	2023	Mortality Prediction	605,268 ECGs collected from healthcare episodes of Alberta, Canada over the period of 2007-2020 (N=748773 of 244077 patients)	ECG recordings, extracted ECG features, age, sex, and laboratory tests	Gradient boosted tree ensembles (XGB) models, and ResNets	Thirty day, one year, and five year mortality horizons. The ResNet models achieved an AUROC of 0.812 (0.808-0.816) for the one year horizon, and an AUROC of 0.798 (0.792-0.803) for the five year horizon.

Appendix B

Results of k-Fold Cross-Validation for Ψ -GANDAE

Table B.1: Results of k-Fold Cross-Validation for Ψ -GANDAE on Muscle Artifact

Method	Level	Metric	103	105	111	116	122	205	213	219	223	230	Average
S-T	0 dB	SNR	10.41	10.02	8.21	8.19	9.20	8.32	8.79	10.05	9.95	8.70	9.18
		RMSE	0.302	0.316	0.389	0.389	0.347	0.384	0.363	0.314	0.318	0.367	0.349
	1.25 dB	SNR	10.89	10.42	8.66	8.51	9.67	8.61	9.67	10.61	10.56	9.01	9.66
		RMSE	0.286	0.301	0.369	0.375	0.328	0.371	0.329	0.295	0.297	0.355	0.331
	5dB	SNR	12.63	12.76	9.94	9.75	11.69	9.91	12.53	12.89	13.44	10.15	11.57
		RMSE	0.234	0.23	0.318	0.325	0.26	0.32	0.236	0.227	0.213	0.311	0.267
WT	0dB	SNR	19.66	22.09	20.02	12.44	6.71	21.23	11.83	7.33	18.58	18.03	15.79
		RMSE	0.044	0.040	0.050	0.100	0.182	0.044	0.096	0.149	0.055	0.066	0.083
	1.25 dB	SNR	16.87	22.49	19.69	14.40	7.43	20.24	13.27	8.68	20.32	18.86	16.23
		RMSE	0.060	0.039	0.052	0.080	0.167	0.049	0.081	0.127	0.045	0.060	0.076
	5 dB	SNR	15.79	24.11	18.81	19.15	11.14	16.51	18.94	14.55	21.41	21.14	18.16
		RMSE	0.067	0.032	0.057	0.046	0.109	0.075	0.042	0.065	0.040	0.046	0.058
I-DAE	0 dB	SNR	21.38	24.72	23.15	19.22	19.57	24.23	19.59	18.80	22.91	22.58	21.62
		RMSE	0.034	0.030	0.035	0.045	0.040	0.031	0.038	0.039	0.032	0.038	0.036
	1.25 dB	SNR	22.41	24.86	23.27	20.22	20.02	24.49	19.78	19.63	23.41	22.60	22.07
		RMSE	0.031	0.029	0.034	0.040	0.038	0.030	0.037	0.034	0.030	0.038	0.034
	5 dB	SNR	23.33	25.13	23.33	22.41	20.63	24.67	20.63	21.97	24.21	22.63	22.89
		RMSE	0.027	0.028	0.034	0.031	0.036	0.030	0.034	0.027	0.028	0.038	0.031
BGAN	0 dB	SNR	28.68	24.20	22.03	27.91	28.83	27.28	26.13	28.62	31.57	27.84	27.33
		RMSE	0.016	0.032	0.023	0.051	0.035	0.018	0.037	0.030	0.020	0.019	0.028
	1.25 dB	SNR	29.29	25.28	22.01	28.54	29.04	27.22	26.13	28.34	31.43	28.46	27.55
		RMSE	0.015	0.027	0.024	0.042	0.034	0.018	0.036	0.032	0.020	0.017	0.026
	5 dB	SNR	29.01	25.23	22.44	28.21	28.85	27.26	26.30	28.92	31.50	28.86	27.61
		RMSE	0.016	0.028	0.022	0.045	0.035	0.018	0.035	0.029	0.019	0.017	0.026
GANDAE	0 dB	SNR	22.14	17.96	13.09	21.64	19.20	20.32	21.65	22.30	26.28	25.52	21.01
		RMSE	0.032	0.056	0.064	0.111	0.111	0.038	0.063	0.061	0.033	0.027	0.059
	1.25 dB	SNR	22.63	18.27	13.13	23.02	19.63	20.53	21.53	22.59	26.11	27.01	21.44
		RMSE	0.030	0.053	0.066	0.085	0.105	0.037	0.061	0.061	0.035	0.021	0.055
	5 dB	SNR	22.87	18.29	11.04	22.78	19.75	20.70	22.19	22.53	26.40	26.97	21.35
		RMSE	0.029	0.053	0.229	0.088	0.099	0.039	0.056	0.059	0.033	0.018	0.070

Table B.2: Results of k-Fold Cross-Validation for Ψ -GANDAE on Electrode Motion Artifact

Method	Level	Metric	103	105	111	116	122	205	213	219	223	230	Average
S-T	0 dB	SNR	6.41	6.13	5.45	5.47	5.87	5.59	5.83	6.05	6.21	6.29	5.93
		RMSE	0.478	0.493	0.534	0.533	0.509	0.525	0.51	0.498	0.489	0.484	0.505
	1.25 dB	SNR	7.47	7.35	6.4	6.32	6.96	6.47	7.06	7.17	7.45	7.48	7.01
		RMSE	0.423	0.429	0.478	0.483	0.449	0.475	0.444	0.438	0.424	0.423	0.447
	5 dB	SNR	10.32	10.4	8.54	8.32	9.60	8.55	10.12	10.04	10.74	10.45	9.71
		RMSE	0.305	0.302	0.374	0.384	0.331	0.374	0.312	0.315	0.290	0.300	0.329
WT	0 dB	SNR	9.51	21.47	9.34	9.63	9.71	18.29	15.02	13.15	18.11	11.99	13.62
		RMSE	0.138	0.044	0.171	0.138	0.128	0.063	0.067	0.076	0.058	0.132	0.102
	1.25 dB	SNR	10.37	23.08	9.65	10.56	9.51	20.30	15.80	13.85	21.10	12.93	14.72
		RMSE	0.125	0.037	0.165	0.124	0.132	0.050	0.061	0.070	0.040	0.119	0.092
	5 dB	SNR	13.07	28.37	14.96	13.77	8.65	21.36	19.20	16.20	21.05	15.91	17.55
		RMSE	0.091	0.020	0.090	0.086	0.145	0.043	0.040	0.054	0.029	0.084	0.068
I-DAE	0 dB	SNR	22.75	23.70	23.39	21.34	17.70	23.47	19.33	18.38	23.17	22.40	21.56
		RMSE	0.029	0.033	0.034	0.035	0.050	0.033	0.040	0.041	0.031	0.039	0.037
	1.25 dB	SNR	22.97	23.94	23.57	21.82	18.76	23.57	19.79	19.07	23.55	22.54	21.96
		RMSE	0.029	0.033	0.033	0.033	0.042	0.033	0.037	0.038	0.030	0.038	0.035
	5 dB	SNR	23.45	24.66	23.65	23.08	20.81	23.66	20.69	21.01	24.00	22.81	22.78
		RMSE	0.027	0.030	0.033	0.030	0.035	0.030	0.034	0.030	0.028	0.037	0.031
BGAN	0 dB	SNR	28.84	24.52	21.55	27.86	28.69	27.68	25.88	28.17	31.02	27.97	27.22
		RMSE	0.017	0.033	0.027	0.050	0.035	0.017	0.038	0.032	0.024	0.021	0.030
	1.25 dB	SNR	29.35	25.03	22.43	28.39	28.85	27.42	26.11	28.46	31.37	28.95	27.64
		RMSE	0.016	0.031	0.026	0.043	0.035	0.016	0.037	0.032	0.022	0.020	0.028
	5 dB	SNR	28.56	24.69	22.00	27.70	28.75	26.72	25.77	28.27	31.17	27.35	27.03
		RMSE	0.017	0.033	0.026	0.051	0.036	0.019	0.038	0.032	0.022	0.024	0.030
GANDAE	0 dB	SNR	22.20	17.77	12.77	21.79	17.61	20.73	21.02	22.28	25.37	25.78	20.73
		RMSE	0.033	0.058	0.066	0.112	0.140	0.037	0.067	0.063	0.041	0.024	0.064
	1.25 dB	SNR	22.64	17.96	13.13	22.66	19.23	20.27	21.49	22.73	25.80	26.47	21.24
		RMSE	0.030	0.056	0.068	0.088	0.113	0.038	0.062	0.061	0.037	0.023	0.058
	5 dB	SNR	22.74	18.20	13.05	22.59	19.43	20.41	21.82	22.43	25.86	25.87	21.24
		RMSE	0.029	0.055	0.069	0.097	0.111	0.038	0.060	0.059	0.036	0.023	0.058

Table B.3: Results of k-Fold Cross-Validation for Ψ -GANDAE on Baseline Wander

Method	Level	Metric	103	105	111	116	122	205	213	219	223	230	Average
ST	0 dB	SNR	11.4	11.56	9.22	9.01	10.58	9.31	11.57	11.55	12.14	11.68	10.80
		RMSE	0.269	0.264	0.346	0.354	0.296	0.342	0.264	0.265	0.247	0.261	0.291
	1.25 dB	SNR	12.06	12.23	9.61	9.35	11.17	9.7	12.22	12.22	12.95	12.35	11.39
		RMSE	0.249	0.245	0.331	0.341	0.276	0.327	0.245	0.245	0.225	0.241	0.273
	5 dB	SNR	13.54	13.77	10.41	10.04	12.38	10.46	13.77	13.65	14.8	13.89	12.67
		RMSE	0.210	0.205	0.302	0.315	0.240	0.300	0.205	0.208	0.182	0.202	0.237
WT	0 dB	SNR	14.87	31.53	18.41	20.00	9.12	22.64	20.83	18.69	17.34	22.23	19.57
		RMSE	0.074	0.014	0.060	0.042	0.138	0.037	0.034	0.040	0.063	0.041	0.054
	1.25 dB	SNR	14.88	31.91	18.42	20.06	8.54	22.73	20.47	20.22	17.40	22.22	19.69
		RMSE	0.074	0.013	0.060	0.042	0.147	0.037	0.036	0.034	0.063	0.041	0.055
	5 dB	SNR	14.90	32.71	18.43	20.10	8.22	22.91	19.11	21.44	17.51	22.18	19.75
		RMSE	0.074	0.012	0.060	0.041	0.153	0.036	0.042	0.029	0.062	0.041	0.055
I-DAE	0 dB	SNR	23.78	25.40	23.31	23.51	20.07	20.07	21.30	23.02	24.25	22.72	22.74
		RMSE	0.026	0.028	0.034	0.027	0.050	0.050	0.032	0.024	0.027	0.037	0.034
	1.25 dB	SNR	23.82	25.42	23.32	23.59	20.08	20.08	21.36	23.31	24.41	22.74	22.81
		RMSE	0.026	0.028	0.034	0.027	0.050	0.050	0.032	0.023	0.027	0.037	0.033
	5 dB	SNR	23.89	25.45	23.35	23.76	20.08	20.08	21.46	24.08	24.64	22.79	22.96
		RMSE	0.025	0.027	0.034	0.026	0.050	0.050	0.031	0.021	0.026	0.004	0.033
BGAN	0 dB	SNR	29.99	25.08	22.70	28.54	29.40	27.76	26.61	28.82	32.00	29.09	28.06
		RMSE	0.013	0.028	0.021	0.043	0.033	0.015	0.035	0.031	0.017	0.015	0.025
	1.25 dB	SNR	30.35	26.01	22.85	28.65	29.40	28.32	27.20	28.87	31.88	29.10	28.32
		RMSE	0.012	0.026	0.022	0.048	0.033	0.015	0.032	0.030	0.018	0.015	0.025
	5 dB	SNR	29.96	25.60	22.71	28.25	29.41	27.64	26.95	28.56	31.72	28.84	28.04
		RMSE	0.013	0.028	0.022	0.051	0.033	0.017	0.032	0.030	0.018	0.016	0.026
GANDAE	0 dB	SNR	22.82	18.42	12.82	22.88	20.30	21.07	21.64	22.45	26.08	25.54	21.40
		RMSE	0.030	0.053	0.065	0.089	0.095	0.032	0.064	0.061	0.034	0.027	0.055
	1.25 dB	SNR	23.91	19.14	13.69	23.59	21.23	22.12	22.18	22.55	26.00	26.89	22.13
		RMSE	0.025	0.049	0.061	0.095	0.085	0.029	0.056	0.062	0.036	0.018	0.051
	5 dB	SNR	23.77	18.79	13.44	23.43	20.58	21.96	22.29	22.65	26.32	27.12	22.04
		RMSE	0.026	0.051	0.063	0.089	0.093	0.030	0.056	0.058	0.032	0.018	0.051