Advancing Structural Engineering Through Data-Driven Methodologies: Seismic Vulnerability Assessment and Backbone Curve Determination

by

Niloofar Elyasi

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Civil Engineering

Waterloo, Ontario, Canada, 2025

Examining Committee Membership

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

External Examiner:	Muntasir Billah Associate Professor Department of Civil Engineering University of Calgary
Supervisor:	Kunho Eugene Kim Assistant Professor Department of Civil and Environmental Engineering University of Waterloo
Internal Member:	Mahesh Pandey Professor Department of Civil and Environmental Engineering University of Waterloo
Internal Member:	Chul Min Yeum Associate Professor Department of Civil and Environmental Engineering University of Waterloo
Internal-External Member:	Rodrigo Costa Assistant Professor Department 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 Contribution

This thesis comprises work written by Niloofar Elyasi under the supervision of Dr. Eugene Kim. Chapters 1, 2, and 6 were solely authored by Niloofar Elyasi, with guidance from Dr. Kim. Chapters 3, 4, and 5 were prepared for publication, and exceptions to sole authorship are as follows:

Chapter 3 is based on the following published journal paper:
N. Elyasi, E. Kim, and C. M. Yeum, "A machine-learning-based seismic vulnerability assessment approach for low-rise RC buildings," *Journal of Earthquake Engineering*, vol. 28, no. 3, pp. 760–776, 2024.

In this paper, Niloofar Elyasi served as the primary author, responsible for developing the methodology, conducting data collection, and implementing computer programming. Dr. Eugene Kim and Dr. Chul Min Yeum contributed by providing guidance on methodology development, conceptualizing the idea, and editing the manuscript.

Chapter 4 is based on the following journal paper submitted for publication:
N. Elyasi and E. Kim, "An enhanced machine learning-based rapid visual screening framework considering model uncertainty and decision threshold optimization," 2024.

In this paper, Niloofar Elyasi was the primary author, responsible for developing the methodology, data collection, and computer programming. Dr. Eugene Kim provided supervision, conceptualized the idea, and contributed to methodology refinement and manuscript editing.

Chapter 5 is based on the following conference paper and a journal paper currently under preparation:
N. Elyasi and E. Kim, "A machine learning-based methodology to determine the backbone curve of RC columns," in *Proc. 18th World Conf. Earthquake Eng.*, Milan, Italy, 2024.

For both the conference paper and the journal manuscript, Niloofar Elyasi was the primary author, responsible for developing the methodology, conducting data collection, implementing computer programming, and presenting the research at the conference. Dr. Eugene Kim guided the methodology development, contributed to conceptualizing the idea, and provided editorial support.

Abstract

Structural engineering has traditionally relied on analytical and experimental methods to ensure the safety of structures. These methods, while effective, often require significant resources, time, and expertise, limiting their applicability across diverse contexts. Meanwhile, vast amounts of data collected from surveys, experimental studies, and seismic events remain largely underutilized, providing a unique opportunity to develop advanced data-driven methodologies. This thesis aims to harness the potential of the available data repositories to address critical challenges in structural engineering, with a focus on seismic vulnerability assessment and backbone curve determination. Through the use of machine learning (ML), this thesis introduces innovative methodologies at both the system and component levels. A rapid visual screening (RVS) framework is developed to quickly assess the seismic vulnerability of low-rise reinforced concrete (RC) buildings. By incorporating ML models, this framework outperforms traditional evaluation methods with higher accuracy and broader applicability. Using post-earthquake survey datasets from a variety of seismic events, it proposes a region-independent tool, eliminating reliance on subjective judgments and region-specific calibrations. For backbone curve determination, used for analyzing the seismic behavior of RC columns, this thesis introduces a novel ML-based methodology. By employing experimental datasets and advanced regression techniques, it offers a practical and efficient alternative to the conventional methods. This approach not only predicts backbone curve parameters with high accuracy but also ensures accessibility for broader applications, especially in resource-limited environments. In summary, this thesis bridges system-level and component-level challenges, underscoring the potential of data-driven approaches in structural engineering. By providing a foundation for integrating innovative approaches into the field, this research advances both academic insights and practical applications. These contributions respond to the demand for efficient and reliable solutions, supporting safer structures and more effective resource management in modern structural engineering practices.

Acknowledgements

This PhD journey has been a path of both challenges and growth, and I am immensely grateful for the encouragement and support I have received from many remarkable people along the way.

First and foremost, I wish to express my deepest gratitude to my supervisor, Dr. Eugene Kim. His invaluable guidance, expertise, and dedication to my academic and personal growth have been essential throughout this journey. I am truly grateful for his support and the opportunities he has provided.

I would like to thank the members of my PhD committee, Dr. Muntasir Billah, Prof. Mahesh Pandey, Dr. Chul Min Yeum, and Dr. Rodrigo Costa, for their invaluable time, insightful feedback, and contributions to my research. Special thanks to Dr. Yeum and Dr. Costa for their expert advice and constructive criticism, which have enriched this thesis.

My heartfelt thanks go to my parents, Zahra and Ali, and my sisters, Naiereh and Nasim, for their constant support and encouragement. Their belief in my abilities and their sacrifices made this journey possible. I also want to express my deepest love for my nephews, Abtin and Arash, whose joy and laughter have brightened my life and inspired me to be my best self.

Last but certainly not least, I would like to express my sincere appreciation to my amazing friends for their support, encouragement, and the countless memories we created together. Their friendship has been a source of joy and strength throughout this path.

Dedication

To my beloved parents, Zahra and Ali,

&

To the courageous women of Iran who fight tirelessly against gender apartheid every day.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Research Motivation

In the field of structural engineering, analytical approaches have traditionally been the favored solutions for addressing a diverse range of problems. These approaches are commonly supported by experimental strategies to enhance their validation. While many of these analytical approaches have demonstrated impressive accuracy and applicability, a drawback lies in their time-consuming and cost-intensive nature. On the other hand, extensive volumes of data collected within the field over the years, through surveys and experimental endeavors, offer the opportunity for the development of data-driven methods. In recent years, machine learning (ML), a branch of artificial intelligence (AI), has become a powerful tool for creating data-driven solutions. By imitating the way humans learn, ML has made significant advancements not only in structural engineering [1] but also in fields such as medicine [2], entertainment [3], and agriculture [4]. It can analyze large datasets, find patterns, and make predictions, allowing for faster and more efficient solutions compared to traditional methods. However, despite the availability of numerous datasets, much of their potential remains unexplored. This is a primary motivation behind conducting this research aiming to propose data-driven methodologies utilizing the wealth of available datasets. Within the field of structural engineering, numerous domains offer rich datasets. The scope of this research focuses on two specific domains: seismic vulnerability assessment and backbone curve determination. These efforts aim to enhance the safety of structures through the development of efficient and innovative data-driven methods.

Structural failures are often the primary cause for human and economic losses in earthquakes. For instance, the 2016 Ecuador earthquake destroyed or severely damaged more than 35,000 masonry and reinforced concrete (RC) buildings resulting in 660 deaths and more than $2 billion in economic damage [5]. Many existing structures remain seismically vulnerable due to outdated design codes, insufficient consideration of seismic requirements in their construction, or the natural aging of their materials over time. Seismic vulnerability assessments of existing structures are essential to proactively ensure occupant safety and minimize losses. A large number of analytical methods have been developed so far to precisely assess the seismic performance of existing structures [6]. These methods often involve the examination of building portfolios using fragility functions which serve to establish the connection between seismic intensity and the probability of structural damage. Such assessments necessitate detailed evaluations that would be time-consuming and expensive if applied to every individual structure within a large inventory. On the other hand, there exists a wealth of large databases derived from surveys, some of which have remained unutilized leaving their full potential untapped. Reinforced concrete buildings, being among the most commonly constructed and occupied structures worldwide, require reliable and efficient assessments to guarantee their safety. The first part of this thesis focuses on developing a cost-effective and accurate data-driven methodology for seismic vulnerability assessment of low-rise RC buildings by utilizing the publicly available data inventories.

While the first part of this thesis focuses on developing a data-driven methodology for assessing RC buildings at a system level, the second part shifts its attention to one of the most critical components of RC structures, RC columns. Columns are essential load-bearing elements, and their seismic performance is pivotal to ensuring the overall stability of buildings during earthquakes. The seismic behavior of columns can be effectively represented and analyzed using backbone curves. These curves depict the relationship between the applied force and the resulting deformation of a structural element, providing critical insights for the design and seismic assessment. Conventional methods for deriving backbone curves, such as those outlined in FEMA 356 [7] and ASCE 41 [8], typically rely on analytical models or laboratorial testing. While these methods are reliable, they are often time-consuming. This makes their widespread application challenging, particularly where resources are limited. Therefore, there is a pressing need for alternative methods that can accurately and efficiently determine backbone curves without relying on costly procedures. Additionally, the availability of data on experimental test results of RC column in the literature presents an opportunity to leverage data-driven approaches. This research aims to address this need by developing a novel data-driven methodology for determining backbone curves of RC columns.

Although the two domains focused on this thesis (i.e., seismic vulnerability assessment of low-rise RC buildings and backbone curve determination for RC columns) differ in scale, they are inherently connected. Both aim to enhance structural resilience through innovative data-driven methodologies that utilize existing datasets. Together, these approaches address both system-level and component-level challenges, contributing to the overall improvement of the seismic performance of RC structures.

1.2 Research Objectives

The primary aim of this thesis is to employ ML to address the underutilization of available data in two critical domains: seismic vulnerability assessment and backbone curve determination. By emphasizing the transformative role of ML in advancing data-driven solutions, the research seeks to contribute to both academic knowledge and practical applications. This goal is achieved through the following specific objectives:

1. Utilize extensive data of low-rise RC buildings to develop a data-driven methodology for assessing their seismic vulnerability at the system level. This involves employing post-earthquake survey data and applying ML classification techniques to enhance accuracy and practicality.
2. Develop a data-driven methodology for determining the backbone curves of RC columns at the component level, utilizing experimental data and ML regression techniques. This aims to provide a practical and efficient alternative to conventional methods, reducing reliance on time-intensive analyses or laboratory testing.

While predictions of seismic performance are often required at the individual building level, conducting detailed evaluations for each building within a large inventory can be both time-consuming and costly. To address this challenge, rapid visual screening (RVS) methods have been developed to quickly identify at-risk buildings through straightforward observations, such as sidewalk surveys. These methods are widely used to assess large building inventories and prioritize actions for the most vulnerable structures. However, despite the availability of numerous RVS methodologies, there is a significant need for approaches that fully exploit untapped data while addressing limitations in accuracy and uncertainty. This thesis addresses this need by proposing a novel data-driven RVS methodology for low-rise RC buildings, presented in Chapters 3 and 4. Chapter 3 introduces a new ML-based RVS approach using classification techniques applied to post-earthquake survey data from major seismic events. While the method demonstrates reasonable accuracy and performance, challenges related to prediction uncertainty persist. Chapter 4 refines the methodology by incorporating uncertainty evaluation and minimizing misclassifications. This chapter proposes an enhanced and comprehensive methodology for RVS, aiming to deliver a robust and practical tool for large-scale seismic vulnerability assessment. The outcomes of this research contribute to safer RC building practices by enhancing RVS accuracy and applicability.

RC columns are critical structural components that play a pivotal role in the seismic stability of buildings. Their seismic behavior is captured through backbone curves, which provide essential insights into their performance under earthquake-induced forces. While conventional methods for deriving backbone curves exist, there is a need for data-driven approaches that can accurately and efficiently determine these curves without the reliance on time-intensive or costly procedures and analyses. To address this need, Chapter 5 focuses on developing an efficient data-driven methodology for determining backbone curves of RC columns. By utilizing existing experimental data and ML regression techniques, this research eliminates the need for simulations or laboratory tests. The goal is to create a practical, accurate, and efficient methodology that surpasses the limitations of traditional methods.

In summary, this thesis bridges system-level and component-level problems through innovative data-driven approaches, addressing both large-scale building assessments and focused structural component evaluations. By proposing novel methodologies for seismic vulnerability assessment and backbone curve determination, the thesis aims to significantly advance the field of structural engineering, providing tools that are both academically and practically impactful. The overview of the thesis is presented in Figure 1.1.

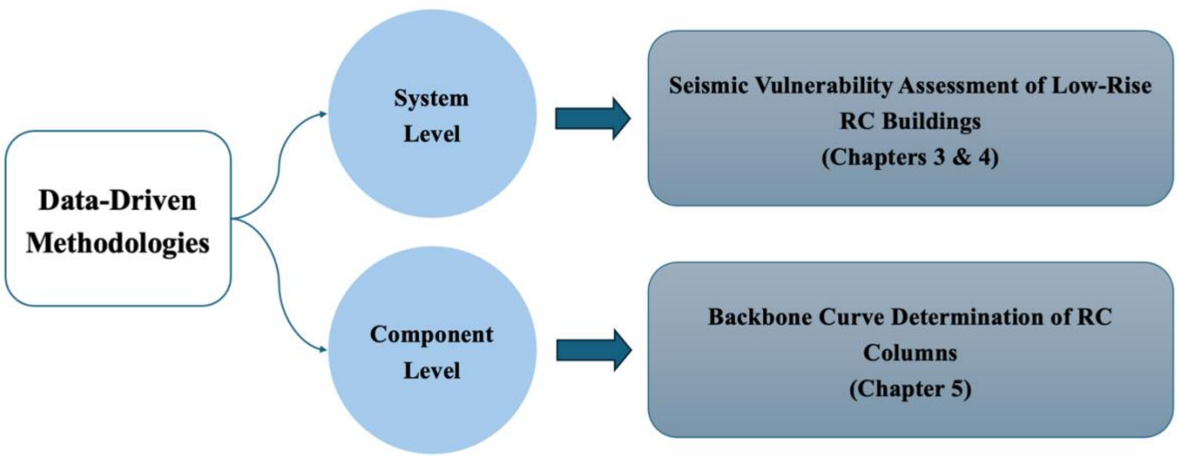


Figure 1.1: Overview of the thesis

1.3 Thesis organization

The rest of this thesis is organized as follows:

Chapter 2 reviews the main concepts and literature related to the studies developed in this thesis. It includes a review of ML techniques, existing studies on RVS, and approaches for determining backbone curves. The strengths and limitations of these methods are discussed, highlighting the gaps in existing research and the motivations for the work presented in this thesis.

Chapter 3 introduces an ML-based seismic vulnerability assessment methodology for low-rise RC buildings, developed using data from post-earthquake inspections. The proposed approach is compared with a widely recognized existing method, with a discussion of its advantages. This chapter is based on a published paper co-authored by the author, Dr. Kim, and Dr. Yeum [9].

Chapter 4 builds on the study presented in Chapter 3 by proposing an enhanced methodology that incorporates model uncertainty detection and threshold optimization. This work introduces a comprehensive three-grade ML-based RVS framework. The content of this chapter corresponds to a paper submitted for publication, co-authored by the author and Dr. Kim.

Chapter 5 focuses on developing a method to determine backbone curves for RC columns using experimental data and ML techniques. The proposed method identifies key points on the backbone curve and is compared to a similar approach from the literature. The results and their significance are discussed. This chapter is based on a conference paper and a journal paper currently under preparation, co-authored by the author and Dr. Kim.

Finally, Chapter 6 summarizes the main findings and contributions of the research presented in this thesis, discusses its limitations, and suggests directions for future work.

Chapter 2
Background and Literature Review

Building on the discussion in the previous chapter, this chapter highlights the transformative role of ML in advancing data-driven methodologies across various disciplines. It begins with an overview of widely used ML techniques and their applications in several studies with a focus on structural and earthquake engineering. Subsequently, the chapter delves into the existing methodologies employed in two selected domains, namely seismic vulnerability assessment and backbone curve determination. By examining their applications, inherent challenges, and potentials for advancement, this chapter aims to highlight opportunities for enhancing current methods and driving innovative methodologies in these domains.

2.1 ML Techniques

The literature exhibits a diverse spectrum of ML techniques used to develop data-driven solutions for addressing a multitude of problems. These techniques include traditional algorithms such as logistic regression, gradient boosting, classification and regression trees (CART), K-nearest neighbors, and random forest (RF), as well as representation learning algorithms like deep learning. Among the extensive range of ML algorithms available, this section provides an overview of the popular algorithms and those particularly relevant to the approaches adopted in this thesis, along with their applications in various studies within the field of structural engineering.

2.1.1 Logistic Regression

Logistic regression is a statistical model often used for classification problems. The logistic regression function is the sigmoid function $p(\mathbf{x}) = 1 / (1 + \exp(-f(\mathbf{x}))$ and $f(\mathbf{x})$ is a linear function of the form $f(\mathbf{x}) = b_0 + b_1x_1 + \cdots + b_nx_n$ where $x_1, x_2, \ldots, x_n$ are the input features, and $b_0, b_1, \ldots, b_n$ are the coefficients or the predicted weights. The logistic regression function predicts the probability that the output for a given set of input features, x, is equal to 1. Therefore, $1 - p(\mathbf{x})$ is the probability that the output is equal to 0. Logistic regression determines the best predicted weights $b_0, b_1, \ldots, b_n$ such that the function $p(\mathbf{x})$ is as close as possible to all actual responses, y_i, where $i = 1, \ldots, n$ and n is the number of observations. For a binary classification problem, the classification boundary is defined by the set of all points that satisfy $p(\mathbf{x}) = 0.5$ which would be given by $b_0 + b_1x_1 + \cdots + b_nx_n = 0$. The fundamentals of logistic regression can be found in resources such as [10]. Mitra et al. (2011) employed this straightforward algorithm for binary classification, distinguishing between non-ductile joint shear failure and ductile beam yielding failure in interior beam-column joints. The proposed model relies on five design parameters, including nominal joint shear stress demand, average bond stress demand, the ratio of joint transverse reinforcement strength to joint shear stress demand, column axial load ratio, and the ratio of beam top to bottom longitudinal reinforcement strength [11], [12]. In another study, Mangalathu and Jeon (2018) utilized logistic regression in addition to other ML classification algorithms to ascertain the failure mode of beam-column joints [13].

2.1.2 RF

RF is an ensemble learning algorithm developed based on the concept of CART. RF creates a large number of decision trees, each of which is trained using a random subset of the dataset (bootstrapping). A random subset of the input features is then considered for the nodes of every tree to split on. These steps can be repeated many times to build a forest with a wide variety of trees. To make a prediction from an RF model, the prediction from every individual tree is considered then an overall aggregated prediction is obtained. In classification problems, the final prediction is determined by a vote in which the most frequent prediction is the outcome. The random selection of the training set and features for each tree and each node makes RF more robust against overfitting compared to CART. Figure 2.1 illustrates an example of an RF classifier with n decision trees. RF has advantageous features including rapid training, scalability, relatively high accuracy with small datasets, and robustness against overfitting. Kohestani and Hassanlourad (2015) employed RF to predict the seismic liquefaction potential of soil based on cone penetration test data [14]. Furthermore, Mangalathu and Jeon (2019) introduced a framework for generating or updating fragility curves of bridges using RF algorithm [15].

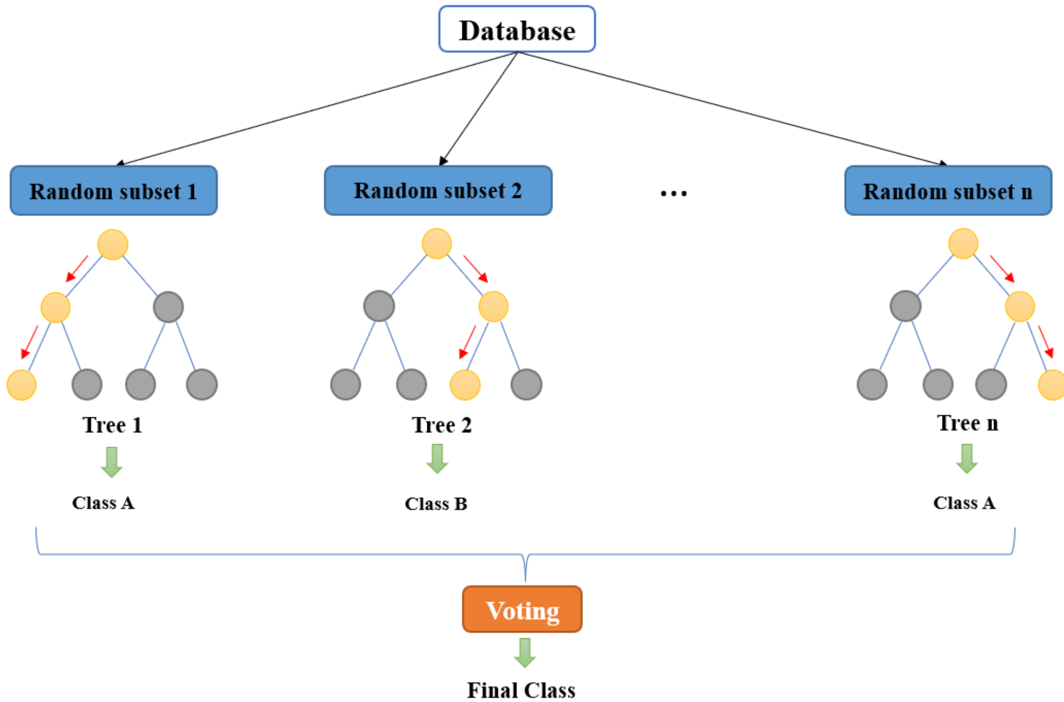


Figure 2.1: Example of an RF classifier [9]

2.1.3 CatBoost

CatBoost is a state-of-the-art, open-source gradient boosting framework developed in 2017 by Yandex, a leading multinational IT company. This framework combines multiple weak learners, typically decision trees, to construct a more accurate and robust model. It commences with a basic model and progressively refines it. New models correct the errors of the prior ones through Gradient Decent, an optimization algorithm that minimizes the loss function by updating parameters in the direction of the steepest descent. Predictions are weighted based on each model's performance. This iterative process persists, introducing new learners and adjusting their contributions, until convergence or a predefined number of iterations, ultimately enhancing its overall predictive capabilities. CatBoost possesses a range of distinct features that position it as a superior choice among boosting algorithms. These features include high model quality without the need for extensive parameter tunning, speed and scalability, efficient handling of categorial features, robustness against overfitting, and strong performance even with small datasets [16]. Chen et al. (2022) explored the application of ML algorithms, including CatBoost, to evaluate the seismic reliability of bridge networks by predicting their performance under seismic events [17]. Hossain et al. (2024) proposed a stochastic CatBoost-based approach for predicting shear wave velocity, a key parameter in seismic analysis [18].

2.1.4 Deep Learning

Deep learning is a technique that employs multiple processing layers to acquire the high-level representations of the data, and it is rooted in the concept of artificial neural networks (ANN) [19]. ANN is one of the most prominent ML techniques closely resembling the biological nervous systems found in the human brain. It can be described as a non-linear mathematical function that transforms a set of input variables from input layer into a set of output variables in the output layer, using hidden layers and a set of parameters known as weights. These weights are determined during the training process. Once the weights are learned and fixed, new data can be processed and tested through the trained network [20]. A typical architecture of ANN algorithm is depicted in Figure 2.2.

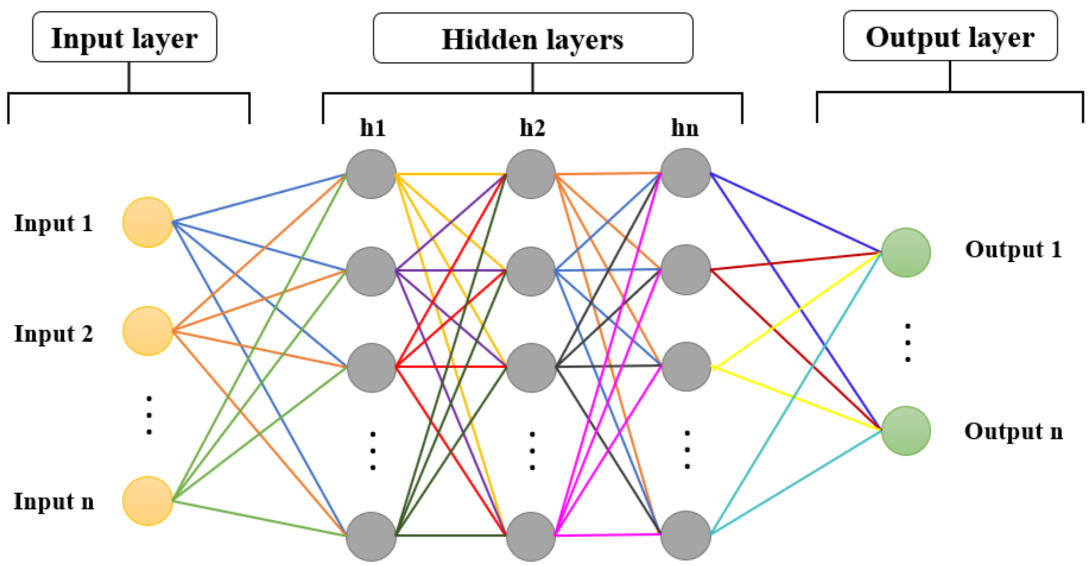


Figure 2.2: General ANN architecture

Deep learning has found application in addressing complex problems across various domains such as computer vision, climate prediction, speech recognition, and financial markets. Several deep learning architectures have been introduced so far encompassing convolutional neural networks, recurrent neural networks, and long short-term memory/gated recurrent unit. In structural and earthquake engineering, numerous researchers have harnessed this robust and widely used technique to devise innovative and effective solutions. For instance, Derakhshani and Foruzan (2019) developed deep neural network models for the estimation of time-domain ground motion parameters [21]. In another study, Lu et al. (2021) employed convolutional neural networks to propose a methodology for a regional post-earthquake seismic damage assessment [22]. Additionally, Yu et al. (2020) introduced an automated procedure utilizing deep learning techniques to identify soft-story buildings from street view images [23].

2.2 Seismic Vulnerability Assessment

As previously highlighted, seismic vulnerability assessments of existing buildings are essential in proactively safeguarding occupant safety and mitigating losses. Conventional analytical methods for seismic risk assessment commonly rely on portfolio-scale building evaluations using fragility functions that establish the correlation between earthquake intensity and the likelihood of damage. Risk assessment tools such as HAZUS categorize these fragility functions based on factors affecting seismic vulnerability such as the building's age and construction type [24]. Nevertheless, a prediction of seismic performance is often needed at the individual building level. However, conducting a detailed and comprehensive evaluation of each building within a large inventory can prove to be a time-consuming and expensive endeavor. This is the main motivation behind the development of RVS methods. These approaches are designed to quickly identify buildings at risk by relying on simple visual observations, such as sidewalk surveys [6]. In the following subsections, an overview of current RVS methods, achieved from the existing literature, is provided.

2.2.1 Scoring RVS Methods

Several scoring RVS methods have been proposed in the past few decades. FEMA 154 was the first sidewalk RVS approach introduced in 1988 to facilitate rapid vulnerability assessments [25]. This method has been revised twice in 2002 and 2015. The most recent version of FEMA 154 incorporates information such as material type, number of stories, seismic force resisting system type, soil type, and regional seismicity to compute a basic score [26]. Additional score modifiers are subsequently applied to account for factors like floor plan and vertical irregularities, redundancy, pounding, and retrofits. A final score is then computed providing an estimate of the building's seismic vulnerability in an earthquake. The scoring system in the FEMA P-154 procedure is related to the probability of collapse and relies on the seismic fragility assessment methodology in HAZUS. Buildings with a score below a cut-off score (2 is recommended in the approach) are prioritized to conduct further evaluation or mitigation strategies. Figure 2.3 presents the screening form from FEMA P-154, designed for high-seismicity regions and utilized in the RVS of a residential and commercial building. FEMA 310 was subsequently introduced as a supplementary document to FEMA 154 while it includes three general checklists: one for the structural system, another for the foundation and geotechnical hazards, and a third for the non-structural system. These checklists encompass several parameters including structural and non-structural components, building systems, lateral-force-resisting systems, diaphragms, material conditions, and connections [27]. Agencies in other countries also developed similar screening methodologies to be used in accordance with their local seismic circumstances. The European union, for instance, introduced EMS-98 and RISK-UE project as two RVS methodologies designed for application in Europe [27], [28], [29].

Rapid Visual Screening of Buildings for Potential Seismic Hazards **Level 1**
FEMA P-154 Data Collection Form **HIGH Seismicity**

Address: 5020 Ebony Drive
Anyplace Zip: 91011
Parcel Number:
Building Name:
Use: Residential and commercial
Latitude: Longitude:
S_S: S_1:
Screener(s): D. Taylor/A. Jones Date/Time: 2/28/14 1pm

No. Stories: Above Grade: 22 Below Grade: 2 Year Built: 2000 ☒ EST
Total Floor Area (sq. ft.): 712,800 Code Year:
Additions: ☒ None ☐ Yes, Year(s) Built:

Occupancy: Assembly, Commercial, Emer. Services, ☐ Historic, ☐ Shelter, Industrial, Office, School, ☐ Government, Utility, Warehouse, Residential, # Units: DNK

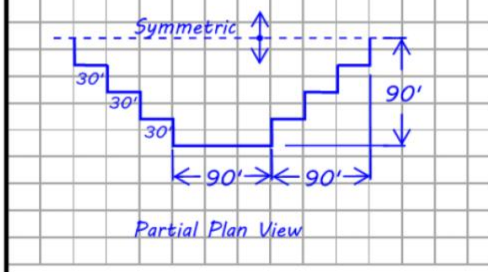
Soil Type: ☐A Hard Rock ☐B Avg Rock ☐C Dense Soil ☒D Stiff Soil ☐E Soft Soil ☐F Poor Soil DNK If DNK, assume Type D.

Geologic Hazards: None

Adjacency: ☐ Pounding ☐ Falling Hazards from Taller Adjacent Building

Irregularities: ☐ Vertical (type/severity)
☒ Plan (type) reentrant corners

Exterior Falling Hazards: ☐ Unbraced Chimneys ☐ Heavy Cladding or Heavy Veneer ☐ Parapets ☐ Appendages ☐ Other:



SKETCH

COMMENTS:
Year built is after benchmark year for cladding anchorage. Therefore, heavy cladding not a falling hazard.

Not apparent whether steel or concrete. Assume S1 or C1. Both are scored with similar results.

☐ Additional sketches or comments on separate page

BASIC SCORE, MODIFIERS, AND FINAL LEVEL 1 SCORE, S_{L1}

FEMA BUILDING TYPE	Do Not Know	W1	W1A	W2	S1 (MRF)	S2 (BR)	S3 (LM)	S4 (RC SW)	S5 (URM INF)	C1 (MRF)	C2 (SW)	C3 (URM INF)	PC1 (TU)	PC2	RM1 (FD)	RM2 (RD)	URM	MH
Basic Score		3.6	3.2	2.9	2.1	2.0	2.6	2.0	1.7	1.5	2.0	1.2	1.6	1.4	1.7	1.7	1.0	1.5
Severe Vertical Irregularity, V_{L1}		-1.2	-1.2	-1.2	-1.0	-1.0	-1.1	-1.0	-0.8	-0.9	-1.0	-0.7	-1.0	-0.9	-0.9	-0.9	-0.7	NA
Moderate Vertical Irregularity, V_{L1}		-0.7	-0.7	-0.7	-0.6	-0.6	-0.7	-0.6	-0.5	-0.5	-0.6	-0.4	-0.6	-0.5	-0.5	-0.5	-0.4	NA
Plan Irregularity, P_{L1}		-1.1	-1.0	-1.0	-0.8	-0.7	-0.9	-0.7	-0.6	-0.6	-0.8	-0.5	-0.7	-0.6	-0.7	-0.7	-0.4	NA
Pre-Code		-1.1	-1.0	-0.9	-0.6	-0.6	-0.8	-0.6	-0.2	-0.4	-0.7	-0.1	-0.5	-0.3	-0.5	-0.5	0.0	-0.1
Post-Benchmark		1.6	1.9	2.2	1.4	1.4	1.1	1.9	NA	1.9	2.1	NA	2.0	2.4	2.1	2.1	NA	1.2
Soil Type A or B		0.1	0.3	0.5	0.4	0.6	0.1	0.6	0.5	0.4	0.5	0.3	0.6	0.4	0.5	0.5	0.3	0.3
Soil Type E (1-3 stories)		0.2	0.2	0.1	-0.2	-0.4	0.2	-0.1	-0.4	0.0	0.0	-0.2	-0.3	-0.1	-0.1	-0.1	-0.2	-0.4
Soil Type E (> 3 stories)		-0.3	-0.6	-0.9	-0.6	-0.6	NA	-0.6	-0.4	-0.5	-0.7	-0.3	NA	-0.4	-0.5	-0.6	-0.2	NA
Minimum Score, S_{MIN}		*1.1*	*0.9*	*0.7*	*0.5*	*0.5*	*0.6*	*0.5*	*0.5*	*0.3*	*0.3*	*0.3*	*0.2*	*0.2*	*0.3*	*0.3*	*0.2*	*1.0*
FINAL LEVEL 1 SCORE, $S_{L1} \geq S_{MIN}$:					$S_{L1} = 2.7$					2.8								

EXTENT OF REVIEW
Exterior: ☐ Partial ☒ All Sides ☐ Aerial
Interior: ☒ None ☐ Visible ☐ Entered
Drawings Reviewed: ☐ Yes ☒ No
Soil Type Source:
Geologic Hazards Source:
Contact Person:

LEVEL 2 SCREENING PERFORMED?
☐ Yes, Final Level 2 Score, S_{L2} ______ ☒ No
Nonstructural hazards? ☐ Yes ☐ No

OTHER HAZARDS
Are There Hazards That Trigger A Detailed Structural Evaluation?
☐ Pounding potential (unless S_{L2} > cut-off, if known)
☐ Falling hazards from taller adjacent building
☐ Geologic hazards or Soil Type F
☐ Significant damage/deterioration to the structural system

ACTION REQUIRED
Detailed Structural Evaluation Required?
☐ Yes, unknown FEMA building type or other building
☐ Yes, score less than cut-off
☐ Yes, other hazards present
☒ No

Detailed Nonstructural Evaluation Recommended? (check one)
☐ Yes, nonstructural hazards identified that should be evaluated
☐ No, nonstructural hazards exist that may require mitigation, but a detailed evaluation is not necessary
☒ No, no nonstructural hazards identified ☐ DNK

Where information cannot be verified, screener shall note the following: EST = Estimated or unreliable data OR DNK = Do Not Know

Legend: MRF = Moment-resisting frame, BR = Braced frame, RC = Reinforced concrete, SW = Shear wall, URM INF = Unreinforced masonry infill, TU = Tilt up, MH = Manufactured Housing, LM = Light metal, FD = Flexible diaphragm, RD = Rigid diaphragm

Figure 2.3: FEMA P-154 level 1 data collection form for heigh seismicity regions [26]

EMS-98 categorizes buildings into six vulnerability classes and outlines five damage states based on straightforward and short descriptions. It takes into account factors like construction quality and workmanship, preservation condition, regularity, ductility, position with respect to other buildings, strengthening and retrofitting efforts, and earthquake resistance design as contributors to building vulnerability [28]. The RVS methodology introduced by the RISK-UE comprises three key steps to determine the vulnerability classes and damage grades: 1) calculation of the vulnerability index considering factors like building topology, regional characteristics, building properties, and seismic behavior; 2) determination of the mean damage grade based on the vulnerability index, topology, construction features, and the micro seismic intensity; and 3) Assessment of the damage distribution and the estimation of fragility curves [29].

Japan building disaster prevention association (JBDPA) developed a more comprehensive RVS methodology compared to the previous approaches with a particular focus on considering the shear capacity of columns and walls. This method involves a comparison between the seismic index of the structure (I_S) and the seismic demand index (I_{S0}) to determine the structure's vulnerability. Several parameters including the basic seismic index of structural and non-structural elements, irregularity index, shear-modification factor, cumulative strength index, and time index are considered in the calculation of I_S [27], [30], [31]. It is worth noting that compared to the FEMA approach, the JBDPA method requires more time for implementation [32]. The New Zealand Society for Earthquake Engineering (NZSEE) has introduced an RVS method inspired by FEMA framework, with an emphasis on the impact of structural irregularities and building size on potential casualties [33]. The distinction between this methodology and FEMA lies in the types of structures considered and the scoring modifiers applied [34]. Another FEMA-based RVS methodology was introduced by the Earthquake Planning and Protection Organization of Greece known as OASP-0. Similar to FEMA P-154, this approach prioritizes buildings with a final score of 2 or lower for detailed vulnerability assessment [35].

National Research Council Canada (NRC) introduced an advanced structural seismic risk assessment approach customized for Canada's seismic characteristics, building types, and seismic design codes. This tailored approach builds upon the foundation of the FEMA method but goes further to incorporate additional score modifiers including building importance, building deterioration and age, and remaining occupancy time. To assess the seismic risk of the existing building more accurately, the NRCC method include different consequences of failure such as fatalities and serious injuries, spanning from low to high severity [36]. While the NRC method may require a longer assessment duration compared to the

FEMA approach, its heightened accuracy is a result of its incorporation of a broader spectrum of parameters and considerations. Figure 2.4 illustrates the flowchart of structural seismic risk scoring methodology developed by NRC.

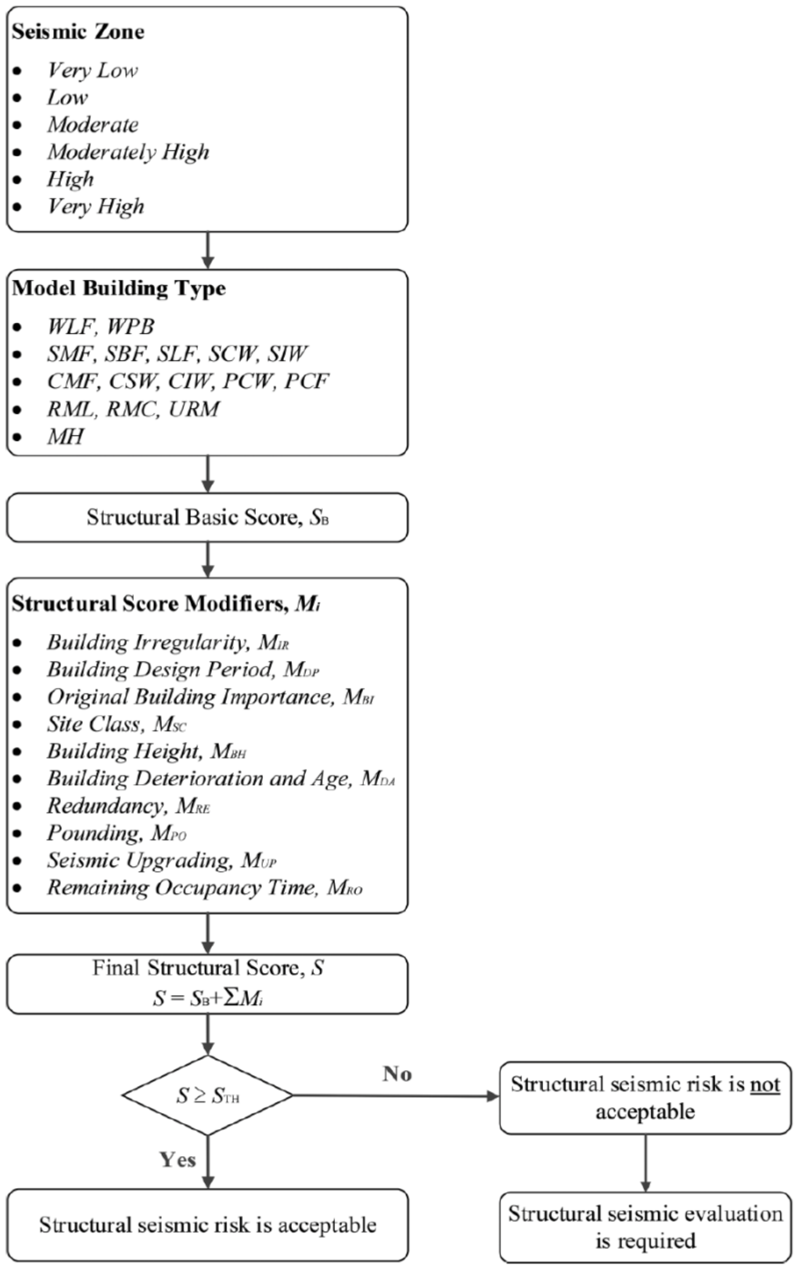


Figure 2.4: Flowchart of structural seismic risk scoring methodology developed by NRC [36]

The Indian Institute of Technology Kanpur (IITK) has also made contributions by adapting the FEMA-based RVS method to align with the conditions prevalent in India. This adaptation involves the modification of modifier values [37] [38]. Notably, a critical threshold in this approach is marked by a cut-off score of 0.7 which signals the necessity for a detailed vulnerability assessment [39]. Furthermore, the Indian standard BIS has introduced its own RVS methodology, known as IS13935, specifically designed for RC and masonry structures, taking into account India's distinct regional conditions [27], [40]. An additional noteworthy Indian RVS methodology was introduced by Murty et. al (2012). This method is based on the seismic performance rating and the safety index, offering an assessment that considers various factors such as material conditions, soil and foundation properties, maintenance quality, and details of structural system [27], [41], [42], [43].

Turkey also proposed its own distinct methodologies to evaluate the seismic vulnerability of structures. The earthquake master plan of Istanbul (EMPI) incorporates an RVS method that takes into account the seismic zone in which the building is located and the number of stories it comprises [44]. The final performance score is determined through the consideration of a basic score, parameters' modifiers, and a vulnerability score. Additionally, Turkey has implemented the RBTE-2019 methodology, a regional and simplified technique for assessing RC, Masonry, and mixed buildings. This method relies on a building performance score as a fundamental component of its assessment process [27], [45]. Despite the simplicity and usefulness of these scoring RVS methods, challenges persist in their application. Notably, certain input parameters such as soil type, building age, and retrofit status may be unavailable or difficult to quantify. Furthermore, defining a reliable cut-off score for distinguishing high-risk structures is challenging. The region or country-specific nature of these methods also introduces a level of limitation to their universal applicability.

2.2.2 Empirical RVS Methods

While the scoring RVS methods offer valuable insights, some input parameters may prove elusive or challenging to quantify (e.g., soil type, building age, retrofit status, etc.). Furthermore, establishing a reliable cut-off score to differentiate high-risk structures can present a formidable task. In addition to the scoring screening methods mentioned earlier, various empirically derived approaches are also at the disposal. Tesfamariam and Saatcioglu (2008), for instance, proposed a risked-based evaluation technique for RC buildings that utilizes a knowledge-based fuzzy rule modelling approach. Several parameters such as the structural system, irregularities, construction quality, and year of construction were incorporated as effective features for assessing seismic vulnerability. The proposed model's performance was validated using seismic damage data gathered from the RC buildings surveyed following the 1994 Northridge Earthquake [46]. Sen (2010) developed a soft computation approach that relies on a fuzzy logic model and inference system. This method is designed to categorize existing buildings into five hazard categories. It takes into account various parameters such as the number of stories, structural irregularities, building quality, potential for pounding, hill-slope effect, and peak ground velocity [47]. The method's effectiveness and accuracy were established through calibration using data from 1249 buildings in Istanbul. Ningthoujam and Nada (2018) conducted a study in the aftermath of the 2016 Manipur earthquake in Northeastern India. They performed a statistical regression analysis on a dataset comprising 396 damaged buildings and developed an RVS method based on six major failure parameters including soil type, construction quality, maintenance condition, building age, substantial overhang, and number of stories [48]. Coskun et al. (2020) employed ordinary least square regression and multivariate linear regression techniques to develop an RVS method for RC buildings. They conducted a detailed statistical analysis of 545 RC buildings located in Turkey [49]. Despite the high accuracy of their method, 14 different input variables are required to make predictions.

A major challenge encountered in most RVS methodologies lies in the accurate prediction of a structure's performance relying primarily on fundamental structural parameters and visual inspections. However, the availability of data and the continuous advancements in AI have paved the way for extracting valuable insights. These insights hold the potential to enhance and enrich traditional approaches in seismic risk assessment. Arslan et al. (2012), for instance, introduced an ANN- based method to assess the earthquake performance of RC buildings in Turkey. They incorporated 19 parameters including the number of stories, steel tension strength, average inertia of beams, spectrum intensity, and dominant frequency. To calibrate the algorithms, a dataset of 66 Turkish RC buildings with heights ranging from four to ten stories were utilized. An accuracy of 80% was achieved in predicting building performance, classifying them into 4-grade performance levels as stipulated by the Turkish Earthquake Code-2007 (TEC-2007) [50]. Zhang et al. (2018) proposed an ML-based framework to establish a connection between the response and damage patterns of a damaged structure and its residual collapse capacity, thus determining its overall safety state. The proposed framework was applied to a four-story RC special frame building and the results demonstrated a high level of prediction accuracy [51]. Mangalathu et al. (2020) harnessed the power of ML algorithms to predict building damage levels drawing upon several input characteristics such as spectral acceleration at a 0.3-

second period, fault distance, and building age. Data from the 2014 South Napa earthquake was utilized in their study, yielding results with an accuracy of 66% [24]. While this study demonstrated the potential of ML, it focused primarily on specific types of damage such as chimney, cripple wall, and porch damage. In another study, Mangalathu et al. (2019) proposed a post-earthquake RVS methodology for existing bridges using the capabilities of ML techniques. A wide range of input parameters from various bridge components including superstructure, interior and exterior bent, foundation, and expansion joints were incorporated as input features for their model. To evaluate the proposed methodology, three damage categories (green, yellow, and red) were introduced as earthquake-induced damage states employing Californian two-span box-girder bridges. Multiple ML algorithms were tested with RF exhibiting the best performance with an accuracy ranging from 73% to 82% [52]. Harirchian et al. (2020) employed SVM as a supervised ML algorithm to propose an RVS method. Twenty-two parameters including ground floor area, year of construction, system type, overhang area, and soft story index were incorporated as input features. The model was trained and tested using the damage data stemming from the 1999 Düzce Earthquake in Turkey. An accuracy rate of 52% was reported in this study [53]. Coskun and Aldemir (2023) introduced an RVS method for forecasting the damage level of masonry buildings employing ensemble learning which combines various supervised ML algorithms. Their models took into consideration several input parameters such as the number of stories, floor system type, visual damage (damaged or undamaged), wall material type, typical story height, vertical irregularity, typical plan area, and earthquake zone. To assess their ML models, a dataset of 543 masonry buildings in Turkey were utilized [54].

2.2.3 Hassan-Sozen RVS Method

Compared to all the abovementioned methods, the priority index (PI) proposed by Hassan and Sozen (1997) is by far the simplest screening method [55]. Developed specifically for low-rise RC buildings based on data from the 1992 Erzincan earthquake, the Hassan-Sozen PI only relies on two metrics computed from seven input parameters to identify and rank the vulnerability of low-rise RC buildings: the number of stories, floor area, column cross-sectional areas, and areas of concrete walls and masonry infill walls in two horizontal directions, east-west (E-W) and north-south (N-S). It is worth noting that unlike other seismic screening methods such as FEMA P-154, the PI does not take into account features such as soil type or geometric irregularities. The PI is computed as the sum of Wall Index (WI) and Column Index (CI):

$$\mathrm{PI} = \mathrm{WI} + \mathrm{CI} \qquad \text{(Equation 2.1)}$$

Where:

$$\mathrm{WI} = \frac{A_{cw} + A_{mw}/10}{A_{ft}} \times 100 \qquad \text{(Equation 2.2)}$$

$$\mathrm{CI} = \frac{A_{col}/2}{A_{ft}} \times 100 \qquad \text{(Equation 2.3)}$$

A_{cw} and A_{mw} are the total cross-sectional areas of RC walls and unreinforced masonry infill walls in a given horizontal direction at the base, respectively. A_{ft} is the total building floor area above the base and A_{col} is the total cross-sectional area of columns above the base. After calculating the PI in both horizontal directions, the minimum value is taken as the PI. Buildings with a lower PI are more seismically vulnerable and should be prioritized for further detailed investigation. The factor of 1/10 applied to A_{mw} in Equation (2) and the factor of 1/2 applied to A_{col} in Equation (3) were originally chosen based on data collected from the 1992 Erzincan earthquake to consider the differences in the stiffness and strength of columns, reinforced concrete walls, and masonry infill walls [55]. However, these values have not been validated in any other datasets.

To evaluate buildings for seismic damage risk based on the PI, Hassan and Sozen proposed a graphical procedure in which the values of WI (y-axis) are plotted against the CI (x-axis) and boundaries are established to define the relative vulnerability of the buildings in the inventory. The original example provided in Hassan and Sozen (1997) for the 1992 Erzincan earthquake is reproduced in Figure 2.4.

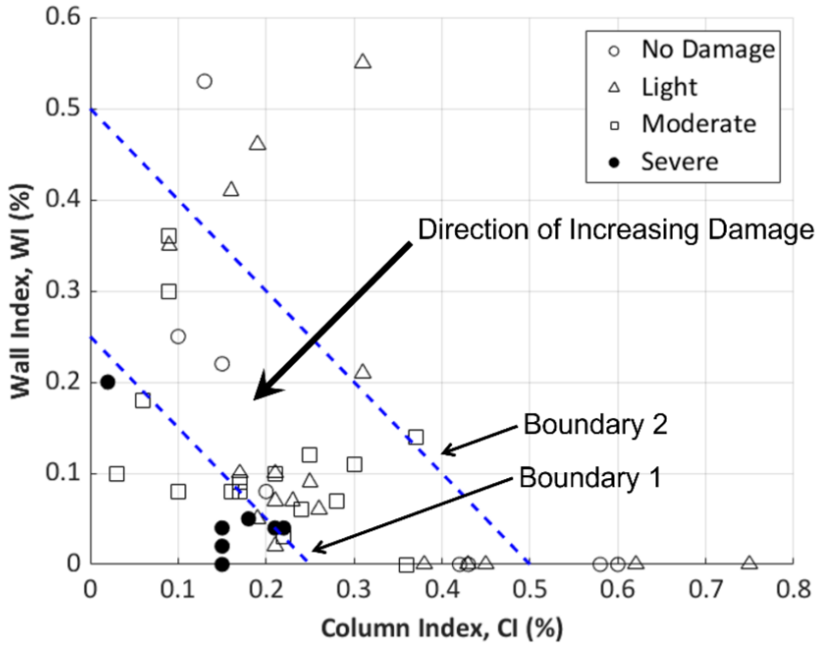


Figure 2.5: Graphical seismic vulnerability evaluation method proposed by Hassan and Sozen [55]

As can be seen, buildings that sustained severe damage in the Erzincan earthquake were closer to the origin (i.e., PI $\rightarrow$ 0) and fell under or close to Boundary 1. These buildings would be recommended for more detailed evaluation or remedial action. A large portion of the moderately damaged buildings are captured between Boundaries 1 and 2 while buildings above Boundary 2 shows light or no damage. Comparing these criteria with the actual damage observed after the earthquake shows the PI can be a reasonably good indicator of seismic vulnerability and the graphical evaluation method for assessing the relative vulnerability of buildings in a given region could be a valid RVS method. However, Hassan and Sozen note that there is no absolute basis for defining the boundaries. If retrofit decisions are to be made based on this approach, definition of the boundaries must reflect the seismic risk tolerance and the availability of resources. Furthermore, manually defining the boundaries for a large dataset could be challenging and boundaries calibrated on limited, region-specific data may not be directly applicable to building inventories in other regions. Additionally, although it is explainable that WI and CI are major parameters in predicting the seismic vulnerability of buildings, the optimal assignment of contributing factors (weights) of WI and CI into PI may not be uniformly equal as considered by Hassan and Sozen. Moreover, shaping the raw input data into two input features (WI and CI) may not be the best way to predict vulnerability. For example, although the number of stories is required to compute WI and CI, it is not considered as a raw input feature for the assessment.

2.2.4 Summary

Detailed vulnerability assessment methods, often paired with experimental validations, usually involve time-consuming and costly procedures. In contrast, RVS methods enable the rapid identification of at-risk buildings by gathering necessary data through visual observations. Based on the literature review provided, it can be deduced that despite the multitude of RVS methodologies introduced to date, there remains a need for a robust approach to maximize the utility of existing screening data. ML, which has streamlined the advancement of data-driven methodologies by identifying optimal relationships between input and output data, can offer a promising avenue for introducing a novel assessment approach. It is worth noting that unlike the existing methods, such an approach should not be constraints by regional or national boundaries.

2.3 Backbone Curve Determination

In structural engineering, a backbone curve represents the force-deformation relationship of structural elements, such as beams, columns, or walls, under cyclic loading conditions, such as those caused by earthquakes. It is typically derived from hysteresis loops generated during cyclic loading tests and provides a simplified summary of the element's behavior, including key parameters like elastic stiffness, yield strength, peak strength, post-peak degradation, and ultimate deformation capacity. The backbone curve serves as a critical tool in seismic design, performance-based assessments, and nonlinear dynamic analyses. The curve highlights distinct phases of structural response: the elastic range, where the structure deforms without sustaining permanent damage; the yield point, marking the transition to inelastic behavior; the peak strength, representing the maximum load-bearing capacity; and the descending branch, which captures strength degradation and residual capacity. These characteristics are essential for understanding stiffness reduction, strength loss, energy dissipation, and overall structural behavior during seismic events. Figure 2.5 illustrates a schematic hysteresis curve and its corresponding backbone curve.

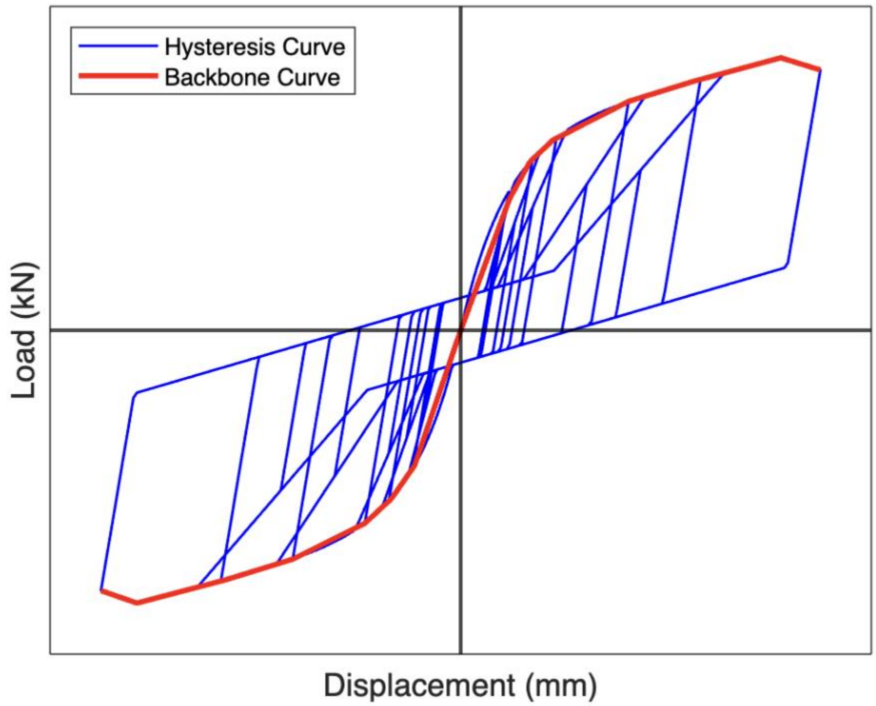


Figure 2.6: A schematic hysteresis curve and its corresponding backbone curve

Backbone curves are widely used in seismic design, seismic performance-based assessments, and retrofitting strategies to ensure that structures meet safety. Traditionally, they have been derived through analytical models and experimental methods, such as cyclic testing. However, recent advancements in data-driven techniques now enable the prediction of backbone curves using experimental datasets, offering a more efficient and cost-effective alternative. In the following subsection, and overview of the existing methods for backbone curve determination is provided.

2.3.1 Conventional Methods

Traditional methods for determining backbone curves have significantly contributed to understanding the seismic behavior of structural elements. FEMA 356, “Prestandard and Commentary for the Seismic Rehabilitation of Buildings” (2000), introduced methodologies and default models for constructing backbone curves for structural components, including beams, columns, shear walls, and connections, under seismic loading. These methodologies outline essential steps such as determining elastic stiffness, yield strength, post-yield behavior, ultimate capacity, and deformation limits, which are aligned with specific performance objectives like immediate occupancy and life safety. The default models offer predefined parameters, such as yield rotation and force-deformation relationships, tailored to standard structural configurations. While these models provide practical solutions for general applications, FEMA 356 underscores the necessity of incorporating experimental test data or conducting detailed analyses to develop more accurate and site-specific backbone curves [7].

Following FEMA, ASCE introduced the guideline, “Seismic Evaluation and Retrofit of Existing Buildings”, which provide a framework for deriving backbone curves based on experimental cyclic testing of structural components. In these tests, elements are subjected to reversed cyclic loading with progressively increasing displacement levels to capture their inelastic behavior. The resulting data is used to determine critical points on the force-deformation curve, which shape the backbone curve. The guideline includes generalized models, such as bilinear and trilinear curves, with predefined parameters like drift limits and strength degradation factors for typical configurations. These backbone curves are encouraged to be refined using detailed test data or analyses tailored to specific materials, geometries, and seismic performance objectives [8].

Elwood and Moehle (2006) further advanced these methodologies by developing an idealized backbone model for evaluating the seismic performance of RC columns with light transverse reinforcement. Their model focuses on critical damage states, including flexural yielding, shear failure, and axial failure. Using analytically derived drift capacities based on column properties such as geometry, material strength, and reinforcement details, the model provides a detailed representation of column behavior under seismic loading. Validation against experimental data, including shake table tests and real-world seismic events, demonstrated its improved accuracy over FEMA 356 criteria. However, the model's applicability is limited to lightly reinforced columns, restricting its broader use for other reinforcement configurations [56].

2.3.2 Data-Driven Methods

Data-driven approaches have become popular for determining backbone curves due to their ability to use large amounts of experimental data available. In a study by Riahi et al. (2009), a model to simulate the seismic behavior of confined masonry walls was developed by focusing on shear-dominated responses. The model defined a trilinear backbone curve representing key limit states: cracking, maximum shear strength, and ultimate deformation capacity. Derived from an extensive database of monotonic and cyclic tests, the model employed iterative linear regression to establish relationships between design variables and the backbone curve parameters. The study also examined the effects of wall openings, aspect ratios, and the presence of bed joint reinforcement on strength characteristics. Comparative analyses with existing predictive equations demonstrated the improved accuracy of the proposed model in reflecting the seismic performance of typical confined masonry walls. However, the model's applicability is limited to walls with atypical features or those subjected to high axial loads [57].

Luo and Paal (2018) developed an ML–Based Backbone Curve model to predict the pre-peak nonlinear behavior of RC columns under cyclic loading. Their proposed bi-linear representation of the backbone curve focused on the safe region before the peak point. Figure 2.7, extracted from their paper, illustrates a schematic comparison between the actual backbone curve and their proposed bi-linear model. Using a multioutput least-squares SVM, the model mapped 15 input parameters, such as column dimensions, material properties, and reinforcement details, to backbone curve outputs including yield and peak shear forces and drift ratios. A grid search optimization algorithm was utilized to fine-tune the model parameters. Trained on a database of 262 tested RC columns, the model achieved significant reductions in prediction errors compared to traditional methods. Despite its success in predicting pre-peak behavior, the study highlighted certain limitations. The model focused primarily on the elastic and peak strength phases, offering limited insights into post-peak degradation and residual capacity, which are critical for understanding structural collapse potential [58].

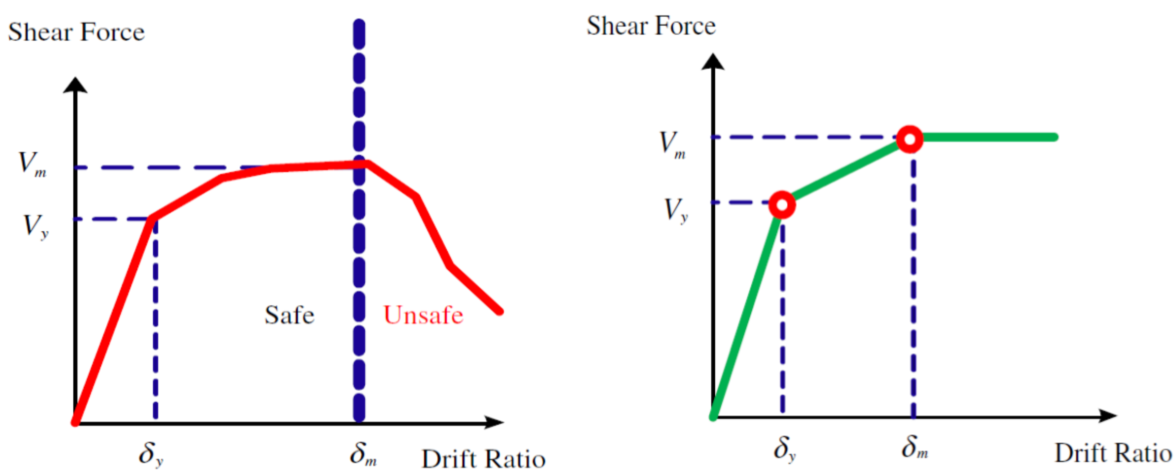


Figure 2.7: Schematic representation of the actual backbone curve and the bi-linear model proposed by Luo and Paal [58]

In another study, Deger and Taskin (2022) developed a novel ML model using Gaussian Process Regression to predict the cyclic backbone curves of RC shear walls. The study focused on seven critical points defining the backbone curve: shear at cracking, shear and displacement at yielding, peak shear and corresponding displacement, and ultimate shear with its associated displacement. Trained on a dataset of 384 experimentally tested RC shear walls, the model effectively captured complex relationships between design variables such as material properties, wall geometry, reinforcement ratios, and structural behavior. The model achieved high accuracy, with coefficients of determination (R^2) between 0.90 and 0.97 for the backbone curve parameters. However, it showed relatively high relative root mean squared error (RelRMSE) for certain parameters, indicating room for improvement [59].

Ma et al. (2023) introduced a novel approach to modeling the nonlinear behavior of RC walls under cyclic loading using genetic programming-based symbolic regression. The approach leveraged evolutionary algorithms to derive explicit mathematical expressions that effectively represent the backbone curves of RC walls. The study utilized experimental data from 388 RC walls, covering diverse design conditions and failure modes, including flexure, flexure-shear, shear, and shear-sliding. When evaluated against the ASCE 41-17 standard, the GP-SR model demonstrated superior accuracy across all failure modes. However, its moderate predictive performance suggests opportunities for further refinement and enhancement [60].

Yu et al. (2024) proposed an approach to identify the backbone curve of RC columns by framing the task as a multi-time series prediction problem. They developed a model, BC-ABiLSTM, which combines bidirectional long short-term memory (BiLSTM) networks with an attention mechanism to capture sequential relationships between points on the backbone curve, thereby improving prediction accuracy. In comparative analyses with other models and traditional identification methods, the BC-ABiLSTM demonstrated superior performance. Despite its promising accuracy, the model's reliance on advanced learning techniques necessitates substantial computational resources and specialized expertise to implement effectively. Furthermore, deep learning models like BC-ABiLSTM generally require large datasets to achieve optimal performance. Applying such models to a small dataset, as in this study, may result in overfitting or limited generalization [61].

2.3.3 Summary

The determination of backbone curves has undergone significant advancements, evolving from traditional analytical and experimental methods to more sophisticated data-driven approaches. While traditional methods have offered crucial insights into structural behavior under seismic loading, they are often resource-intensive, time-consuming, and reliant on generalized assumptions. Recent advancements in ML-based techniques have leveraged experimental datasets to predict backbone curves with improved accuracy and efficiency. However, the existing approaches still face challenges, such as inconsistent performance or difficulty in capturing post-peak part of the curve. These limitations emphasize the need for a new methodology that makes a balance between accuracy, efficiency, and applicability. Developing a robust approach can overcome the existing challenges, providing a more comprehensive and effective framework for determining backbone curves and understanding structural performance.

Chapter 3

A Machine Learning-Based Seismic Vulnerability Assessment Approach for Low-Rise RC Buildings

This chapter corresponds to the following published journal paper:

N. Elyasi, E. Kim, and C. M. Yeum, "A machine-learning-based seismic vulnerability assessment approach for low-rise RC buildings," *Journal of Earthquake Engineering*, vol. 28, no. 3, pp. 760–776, 2024.

Abstract

Seismic vulnerability evaluation of existing buildings is essential to minimize the destructive impacts of earthquakes. RVS methods are simple and effective vulnerability assessment techniques to help quickly identify high-risk buildings for more detailed evaluations. Among various RVS methods, the Hassan-Sozen PI is one of the simplest methods that can be used for low-rise RC buildings. The PI relates simple, easily attainable geometric features of a building including number of stories, floor area, column area, and wall area to damageability. However, the relationship is overly simplified and there is no absolute basis for defining damage classification boundaries that can be used to interpret the PI. Furthermore, given the lack of seismic parameters as inputs, the PI only allows for a relative evaluation of buildings in a specific region. To address these issues and develop a more broadly applicable RVS method, this study first proposes an improved PI evaluation method using ML techniques to define damage classification boundaries. Then, a new generalized RVS method is proposed that considers the PI input features and earthquake intensity measures to predict damage states. Data from six post-earthquake damage surveys (Duzce (1999), Bingol (2003), Nepal (2015), Taiwan (2016), Ecuador (2016), and Pohang (2017)) are used to train and evaluate the classification models. Two earthquake intensity features, modified Mercalli intensity (MMI) and peak ground acceleration (PGA), are introduced to develop a new earthquake intensity aware RVS. The results of the proposed methodologies show a considerable improvement from the original PI with no judgment needed to define the damage classification boundaries.

3.1 Introduction

The literature review in Chapter 2 provided a comprehensive overview of existing RVS methods. Among these, the PI proposed by Hassan and Sozen (1997) stands out as one of the simplest approaches, tailored specifically for low-rise RC buildings based on data from the 1992 Erzincan earthquake. This method relies on just two metrics derived from seven basic structural parameters: the number of stories, floor area, column cross-sectional areas, and areas of concrete walls and masonry infill walls in two horizontal directions, E-W and N-S. While the Hassan-Sozen PI offers a straightforward and efficient screening framework, its accuracy and applicability beyond the 1992 Erzincan earthquake remain largely unexplored. Additionally, challenges such as the lack of a definitive basis for damage classification boundaries and the region-specific nature of the model limit its broader application. These limitations were critically discussed in Chapter 2, highlighting the need for refinement.

In this study, the original Hassan-Sozen assessment method is first evaluated then improved upon using ML algorithms including logistic regression and RF. A large Hassan-Sozen PI building inventory from six earthquakes (Duzce (1999) [62], Bingol (2003) [62], Nepal (2015) [63], Taiwan (2016) [64], Ecuador (2016) [65], and Pohang (2017) [66]) is used to train and test the models. The research further introduces a novel ML-based rapid visual screening (RVS) framework that incorporates two additional features related to earthquake intensity: modified Mercalli intensity (MMI) and peak ground acceleration (PGA). The results demonstrate the capability of the developed RVS to accurately identify buildings with a high likelihood of being severely damaged.

3.2 PI Building Inventories

The PI data from six different earthquakes, 1) Duzce, Turkey (1999), 2) Bingol, Turkey (2003), 3) Nepal (2015), 4) Taiwan (2016), 5) Ecuador (2016), and 6) Pohang, South Korea (2017), are used in this research. For each earthquake, post-earthquake reconnaissance teams collected input parameters for computing the PI of low-rise RC buildings. These datasets are available on Datacenterhub, an open data sharing platform developed by Purdue University and University of Nebraska (www.datacenterhub.org). A total of 658 buildings are included in the datasets. The earthquake events considered in this study are identified in the map shown in Figure 3.1.

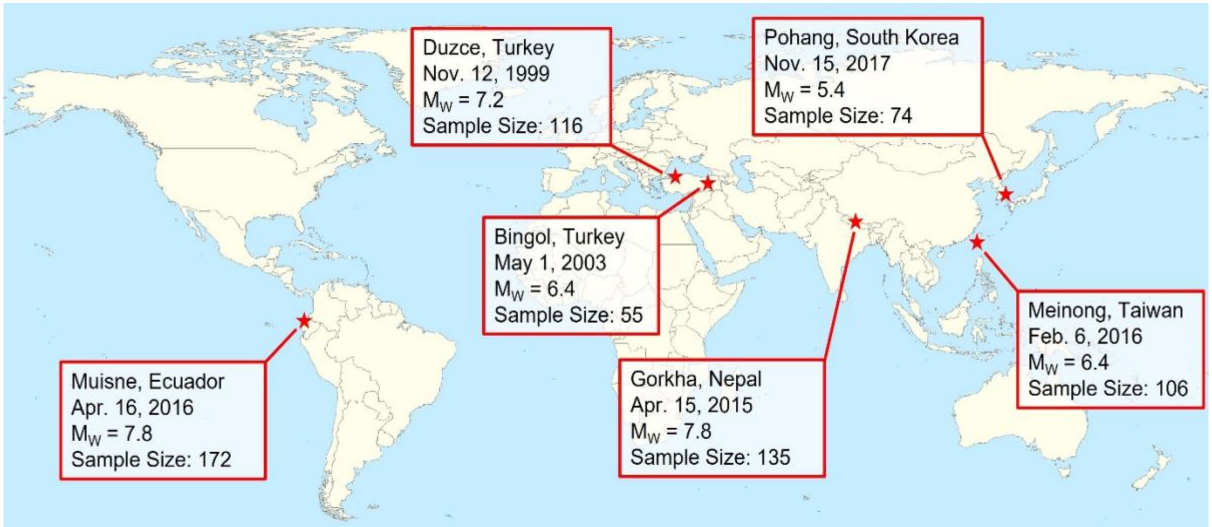


Figure 3.1: Earthquake datasets used in this study

The Duzce earthquake with a moment magnitude (M_w) of 7.2 occurred on the Duzce fault in Turkey on November 12, 1999, causing approximately 900 deaths and 3000 injuries. The 40 km long surface rupture passed through urban areas and resulted in thousands of severely damaged buildings [67]. Of the 116 samples in the Duzce dataset, 98 (approximately 85%) are one- to four-story buildings with a total floor area (A_{ft}) of less than 750 m^2. Almost 96% of the surveyed buildings have only RC walls and 4% have no walls at all.

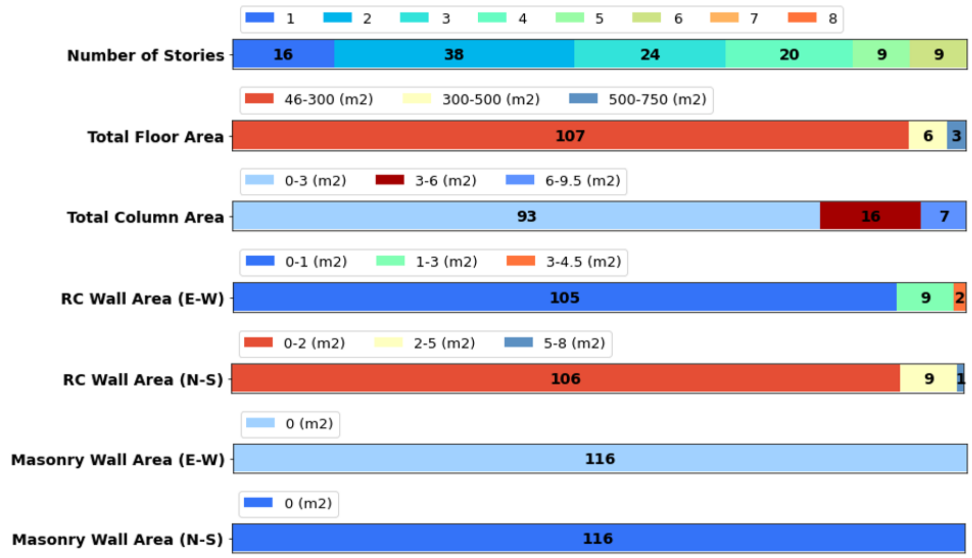
The Bingol earthquake (M_w = 6.4) occurred in the city of Bingol, Turkey on May 1, 2003. A total of 176 people were killed, and more than 600 buildings collapsed or were heavily damaged [68]. A key factor in the vast damage to buildings, especially school buildings constructed within 5-6 years of the earthquake, was poor concrete quality and reinforcement [69]. Of the 55 low-rise RC buildings in the Bingol dataset, more than 90% are three- to five-story buildings with a total floor area of less than 900 m^2. 63% of the samples have only masonry infill walls, 2% have only RC walls, and 33% have both masonry and RC walls.

The Nepal earthquake (M_w = 7.8), also known as Gorkha earthquake, occurred on 25 April 2015, and caused significant damage and triggered numerous landslides. This earthquake led to more than 8,500 casualties and thousands of severely damaged buildings. Several intense aftershocks caused additional damage to the rural areas in the northern part of central Nepal [63]. Of the 135 buildings in the Nepal dataset, 117 (more than 85%) are three- to six-story buildings with a total floor area of less than 550 m^2. Almost all buildings in the dataset have masonry infill walls.

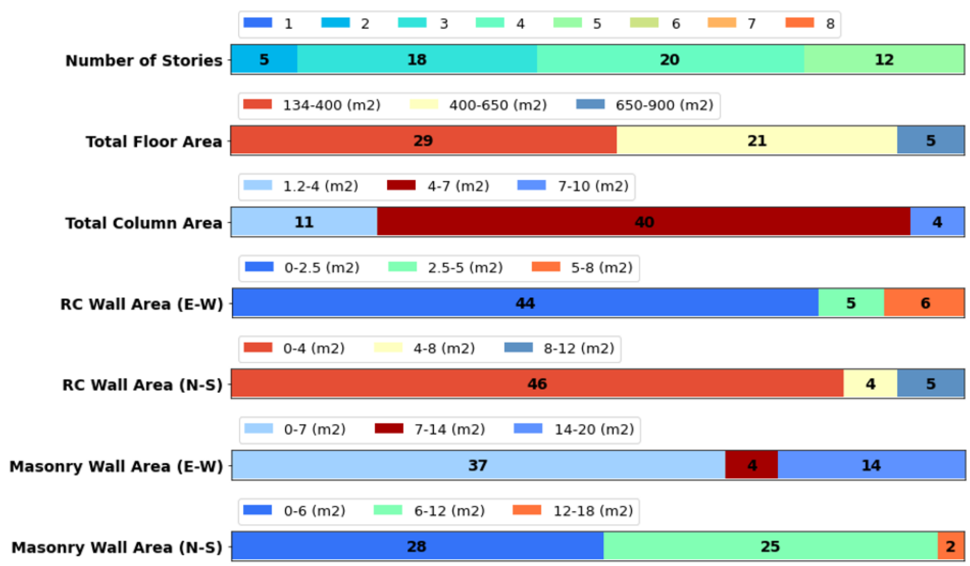
The Taiwan earthquake (M_w = 6.4) occurred on February 6, 2016, near Meinong in southern Taiwan. This event caused 117 casualties and widespread damage to infrastructure including 314 bridges [64]. The Meinong, Taiwan dataset used in this study consists of 106 low-rise RC buildings. More than 88% of the buildings are between two to four stories with a total floor area of less than 1,650 m^2. 60% of the samples have masonry infill walls, 17% have RC walls, and 22% have both masonry and RC walls.

One of the most destructive earthquakes whose data is used in this study is the Ecuador earthquake (M_w = 7.8) which occurred on April 16, 2016, near Muisne. More than 660 casualties and 35,000 severely damaged buildings were reported mostly across the coastal province of Manabí. Following this earthquake, the American Concrete Institute (ACI), with the help of the Escuela Superior Politécnica del Litoral (ESPOL) supported a group of researchers who recorded data from 172 RC buildings affected by the earthquake [65]. Almost 70% of the samples in the Ecuador dataset are two- or three-story residential buildings with a total floor area of less than 1,350 m^2. Almost all buildings in the dataset have masonry infill walls.

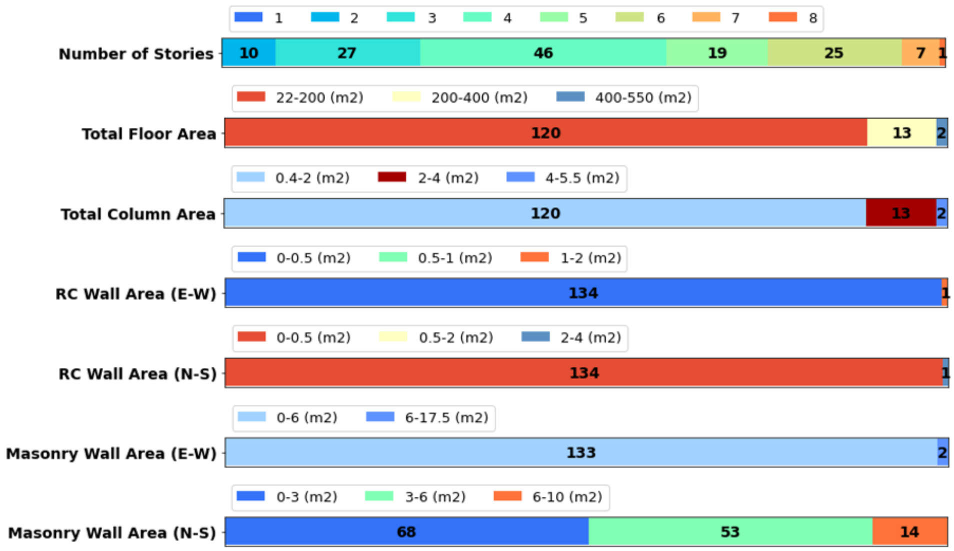
Lastly, the Pohang earthquake was a moderate earthquake with a moment magnitude of 5.4 that occurred near Heunghae and Pohang, in South Korea on November 15, 2017. Although no deaths were reported in this event, over a thousand residential buildings and infrastructure were severely damaged with the cost of approximately 100 million dollars [66]. Data from 74 low-rise RC buildings surveyed after the Pohang earthquake is used in this study. Nearly 90% of the samples are three- or four-story buildings with a total floor area of less than 1,000 m^2. 5% of the buildings have only masonry infill walls, 88% have only RC walls, and 4% have both masonry and RC walls. A summary of each dataset is shown in Figure 3.2.


1 2 3 4 5 6 7 8
Number of Stories
16 38 24 20 9 9
46-300 (m2) 300-500 (m2) 500-750 (m2)
Total Floor Area
107 6 3
0-3 (m2) 3-6 (m2) 6-9.5 (m2)
Total Column Area
93 16 7
0-1 (m2) 1-3 (m2) 3-4.5 (m2)
RC Wall Area (E-W)
105 9 2
0-2 (m2) 2-5 (m2) 5-8 (m2)
RC Wall Area (N-S)
106 9 1
0 (m2)
Masonry Wall Area (E-W)
116
0 (m2)
Masonry Wall Area (N-S)
116

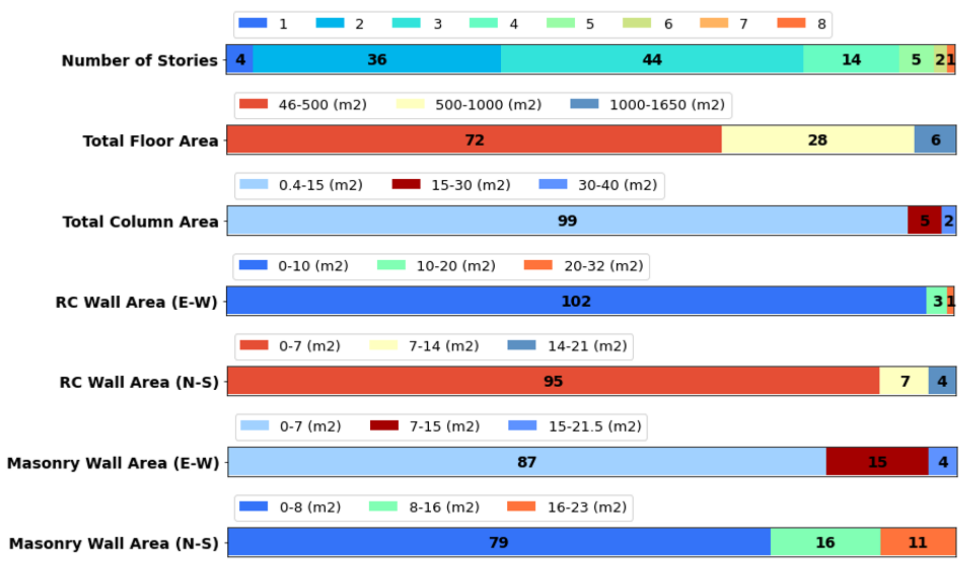
(a)


1 2 3 4 5 6 7 8
Number of Stories
5 18 20 12
134-400 (m2) 400-650 (m2) 650-900 (m2)
Total Floor Area
29 21 5
1.2-4 (m2) 4-7 (m2) 7-10 (m2)
Total Column Area
11 40 4
0-2.5 (m2) 2.5-5 (m2) 5-8 (m2)
RC Wall Area (E-W)
44 5 6
0-4 (m2) 4-8 (m2) 8-12 (m2)
RC Wall Area (N-S)
46 4 5
0-7 (m2) 7-14 (m2) 14-20 (m2)
Masonry Wall Area (E-W)
37 4 14
0-6 (m2) 6-12 (m2) 12-18 (m2)
Masonry Wall Area (N-S)
28 25 2

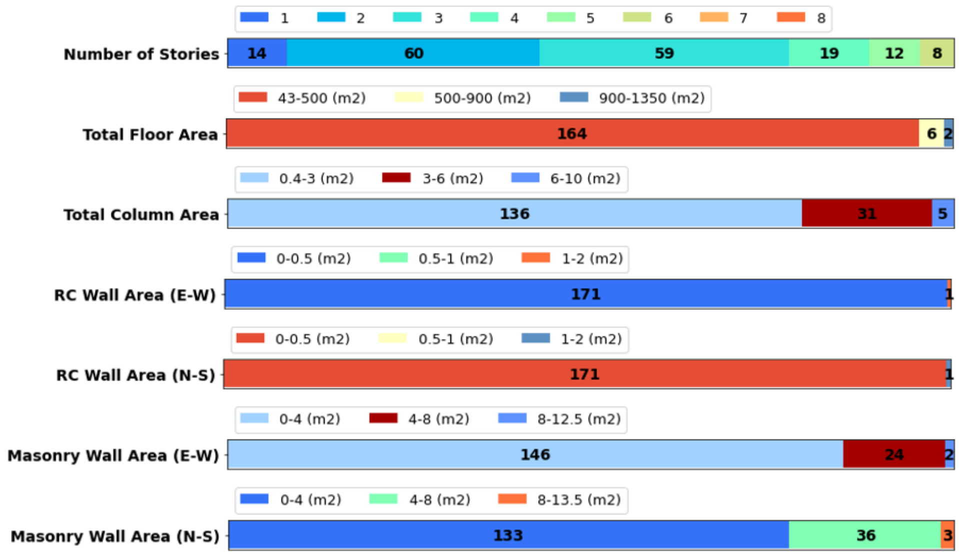
(b)


1 2 3 4 5 6 7 8
Number of Stories
10 27 46 19 25 7 1
22-200 (m2) 200-400 (m2) 400-550 (m2)
Total Floor Area
120 13 2
0.4-2 (m2) 2-4 (m2) 4-5.5 (m2)
Total Column Area
120 13 2
0-0.5 (m2) 0.5-1 (m2) 1-2 (m2)
RC Wall Area (E-W)
134 1
0-0.5 (m2) 0.5-2 (m2) 2-4 (m2)
RC Wall Area (N-S)
134 1
0-6 (m2) 6-17.5 (m2)
Masonry Wall Area (E-W)
133 2
0-3 (m2) 3-6 (m2) 6-10 (m2)
Masonry Wall Area (N-S)
68 53 14

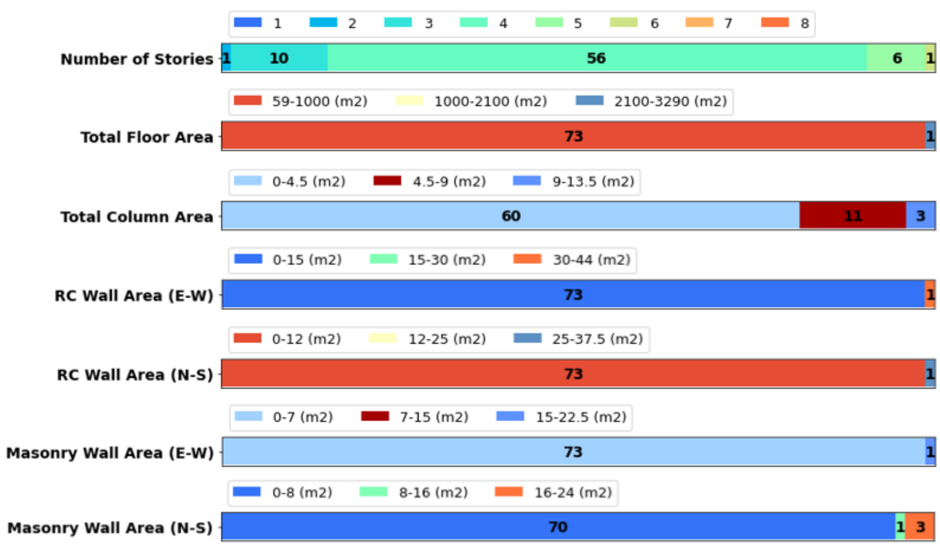
(c)


1 2 3 4 5 6 7 8
Number of Stories
4 36 44 14 5 2 1
46-500 (m2) 500-1000 (m2) 1000-1650 (m2)
Total Floor Area
72 28 6
0.4-15 (m2) 15-30 (m2) 30-40 (m2)
Total Column Area
99 5 2
0-10 (m2) 10-20 (m2) 20-32 (m2)
RC Wall Area (E-W)
102 3 1
0-7 (m2) 7-14 (m2) 14-21 (m2)
RC Wall Area (N-S)
95 7 4
0-7 (m2) 7-15 (m2) 15-21.5 (m2)
Masonry Wall Area (E-W)
87 15 4
0-8 (m2) 8-16 (m2) 16-23 (m2)
Masonry Wall Area (N-S)
79 16 11

(d)



(e)



(f)

Figure 3.2: Datasets summaries: (a) Duzce (1999), (b) Bingol (2003), (c) Nepal (2015), (d) Taiwan (2016), (e) Ecuador (2016), and (f) Pohang (2017)

In each dataset, inspectors classified the observed structural damage into three categories: light damage, moderate damage, and severe damage. A description of each damage level is presented in Table 3.1.

Table 3.1. Description of structural damage levels [65], [70]

Damage Level	**Description**
Light	Hairline cracks (< 0.25 mm) and/or flexural cracks.  
Moderate	Inclined cracks, spalling, or loss of concrete cover.  
Severe	Loss of concrete core, reinforcement buckling or fracture, permanent drift in structure visually apparent, shear failure, or failure of any structural element.  

Data from the Duzce, Bingol, Nepal and Taiwan earthquakes also include a "no damage" level but this class is absent in the remaining datasets. For the purposes of this study, buildings with no damage were combined into the light damage class. This consolidation reduces the risk of overfitting models to the datasets with the "no damage" class and does not affect the applicability or accuracy of the screening method since the no damage and light damage classes signify low seismic risk and the main goal is to identify buildings likely to be severely damaged. Figure 3.3 shows the distribution of the damage classes for the six datasets used in this study.

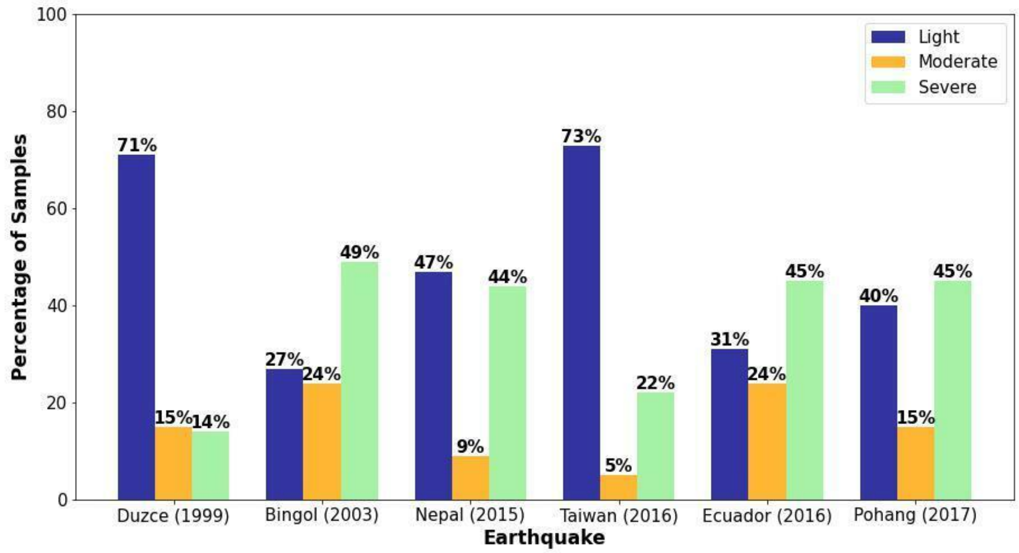


Figure 3.3: Distribution of the damage states in the datasets

As can be seen in Figure 3.3, the moderate damage class only accounts for a small portion of the building inventory. The small sample size makes explicitly classifying the moderate damage class challenging. In the scatter plots in Figure 3.4, it can also be seen that unlike data from the 1992 Erzincan earthquake used by Hassan and Sozen (1997) (Figure 2.4), the moderate damage class is difficult to clearly distinguish from the light and severe damage classes based on the PI. This was further corroborated by an initial investigation of a multi-class classifier with light, moderate and severe damage classes which yielded poor predictions of the moderate damage class. Therefore, buildings with moderate damage level were conservatively included in the severe damage class. This simplifies the RVS methods developed in this study to a binary classification between two damage classes: the non-severe damage class which includes no damage and light damage; and the severe damage class which combines moderate and severe damage. Although this results in loss of specificity in the damage states, it conservatively smooths out judgment errors that may have been made in the data collection. Furthermore, this binary classification is more representative of the descriptive power of the input features which are simple and non-technical.

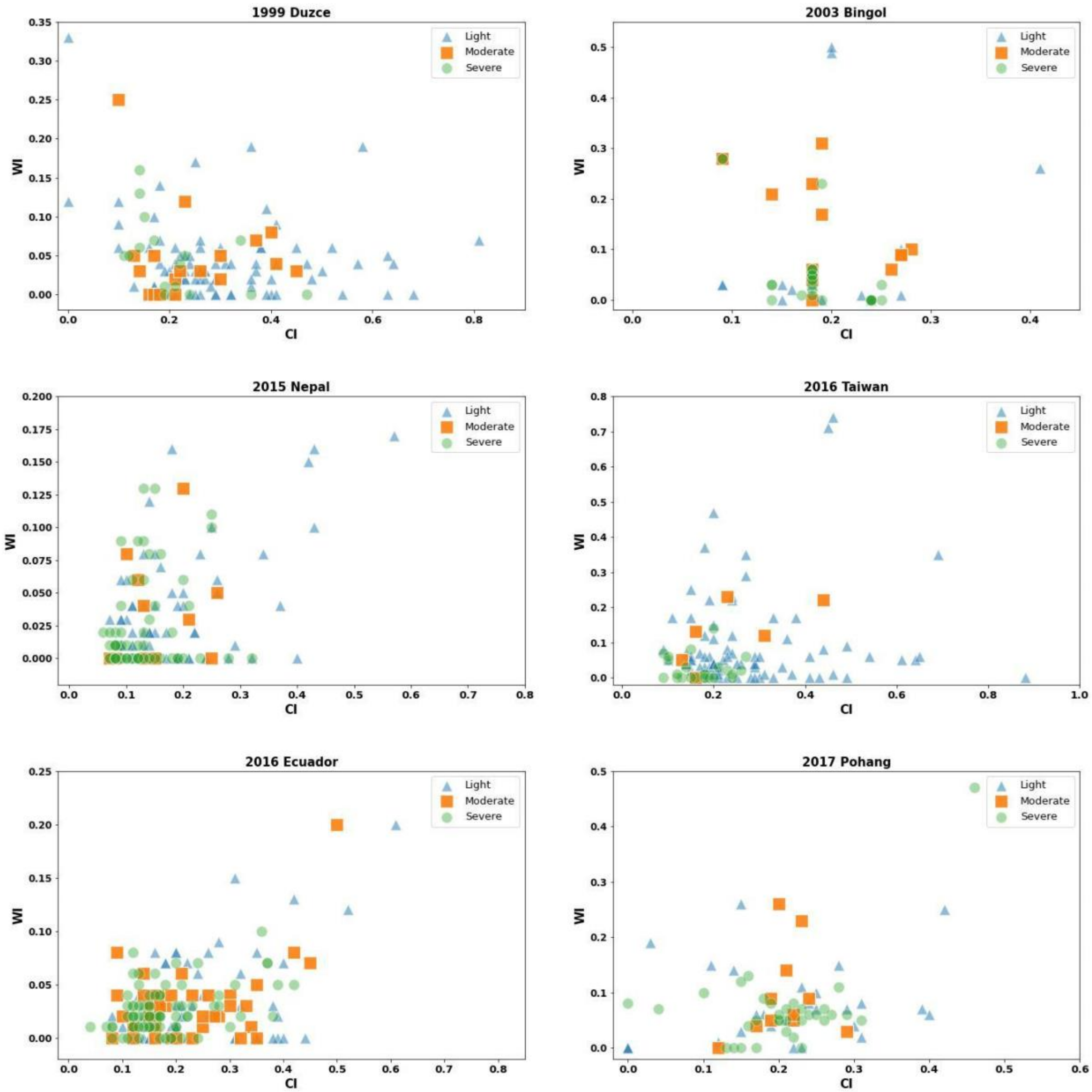


Figure 3.4: Distribution of damage in relation to the WI and CI

3.3 ML Techniques

In this study, logistic regression is applied to automate the process of finding the damage classification boundaries in the original graphical PI evaluation method proposed by Hassan and Sozen (1997) instead of establishing them manually based on judgment. Subsequently, RF as a nonlinear classification algorithm is used to propose a new RVS approach that directly predicts the damage class based on the PI input parameters instead of considering the CI and WI and graphically defining damage classification boundaries.

The accuracy of the developed ML models is evaluated using the weighted average F1-score in which the F1-score of each class is weighted by the number of samples in that class while averaging over all classes. The reason for utilizing F1-score in this study is that the damage classes of the datasets are imbalanced, and this metric takes into account the imbalanced data distribution. The F1-score is defined as follows:

$$\mathrm{F1} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \qquad \text{(Equation 3.1)}$$

where precision is the ratio of correctly predicted labels to all predicted labels while recall is the ratio of correctly predicted labels to all actual labels. Precision and recall are given by the following Equations.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \qquad \text{(Equation 3.2)}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad \text{(Equation 3.3)}$$

TP, FP, and FN indicate true positives, false positives, and false negatives, respectively. For a given damage class (severe or non-severe), TP is equal to the number of samples that are correctly predicted while FP is the number of samples that are incorrectly predicted to be in the damage class under consideration. FN is the number of the samples that actually belong to the damage class but are incorrectly identified to be in the other damage class. Further discussion on these accuracy metrics can be found in [71], [72].

To explain and interpret the developed ML models, SHapley Additive exPlanations (SHAP) is used. Introduced by Lundberg and Lee (2017), SHAP is a pivotal framework for interpreting ML models, built on cooperative game theory and Shapley values (Shapley, 1953). It assigns an importance value to each input feature, quantifying its contribution to a specific prediction. SHAP is a powerful tool in ML applications, facilitating model debugging by revealing feature influence, improving feature selection by identifying key variables, and enhancing transparency by clarifying the decision-making process [73], [74].

3.4 Results and Discussion

3.4.1 Evaluation of the PI

In the Hassan-Sozen method, buildings with lower PI are correlated to higher levels of damage. To evaluate the PI as a seismic vulnerability measure, logistic regression is first used to define a model which predicts the damage class based on the PI as the only input feature. Since the PI is region-specific and does not take into account any seismic features (e.g., regional seismicity, earthquake intensity, and site class), the model is trained and tested on each of the six datasets individually using 5-fold cross validation technique. In this technique, for each fold, the dataset is divided into five subsets, with four subsets used for training and one subset used as the test set. The accuracies of the models are computed using F1-score as summarized in Table 3.2.

Table 3.2: PI-based seismic damage prediction accuracies

1999 Duzce				**2003 Bingol**			
Damage Class	Precision	Recall	F1	**Damage Class**	Precision	Recall	F1
Non-Severe	0.8	0.48	0.6	Non-Severe	0.28	0.33	0.3
Severe	0.36	0.71	0.48	Severe	0.73	0.68	0.7
Weighted Average Accuracy: 0.54				**Weighted Average Accuracy: 0.56**			
2015 Nepal				**2016 Taiwan**			
Damage Class	Precision	Recall	F1	**Damage Class**	Precision	Recall	F1
Non-Severe	0.61	0.47	0.53	Non-Severe	0.87	0.55	0.67
Severe	0.6	0.73	0.66	Severe	0.42	0.79	0.55
Weighted Average Accuracy: 0.60				**Weighted Average Accuracy: 0.65**			
2016 Ecuador				**2017 Pohang**			
Damage Class	Precision	Recall	F1	**Damage Class**	Precision	Recall	F1
Non-Severe	0.42	0.49	0.45	Non-Severe	0.46	0.57	0.51
Severe	0.75	0.7	0.72	Severe	0.65	0.55	0.59
Weighted Average Accuracy: 0.64				**Weighted Average Accuracy: 0.53**			

As can be seen, the PI predicted the damage states with the weighted average accuracies of 54%, 56%, 60%, 65%, 64%, and 53% respectively for the Duzce, Bingol, Nepal, Taiwan, Ecuador, and

Pohang datasets. Despite the simplicity of the PI which is defined as the sum of the WI and CI with equal weights (Equation 2.1), the average accuracies exceeding 50% indicate that it can be a competent but inconsistent preliminary measure of seismic vulnerability in low-rise RC buildings.

To evaluate the relative importance of the WI and CI on the damage class prediction, the graphical evaluation approach proposed by Hassan and Sozen is applied to each of the datasets. For each dataset, WI is plotted against CI, and logistic regression is used to define the best fit classification boundaries. Since this is a binary classification problem with two input features (CI and WI), the classification boundary is given by a straight line. In the scatter plots shown in Figure 3.5, CI and WI are on the horizontal and vertical axes respectively, and the different marker types indicate non-severe and severe damage cases. The classification boundary is shown using red dashed lines and the weighted average accuracy is indicated in each plot.

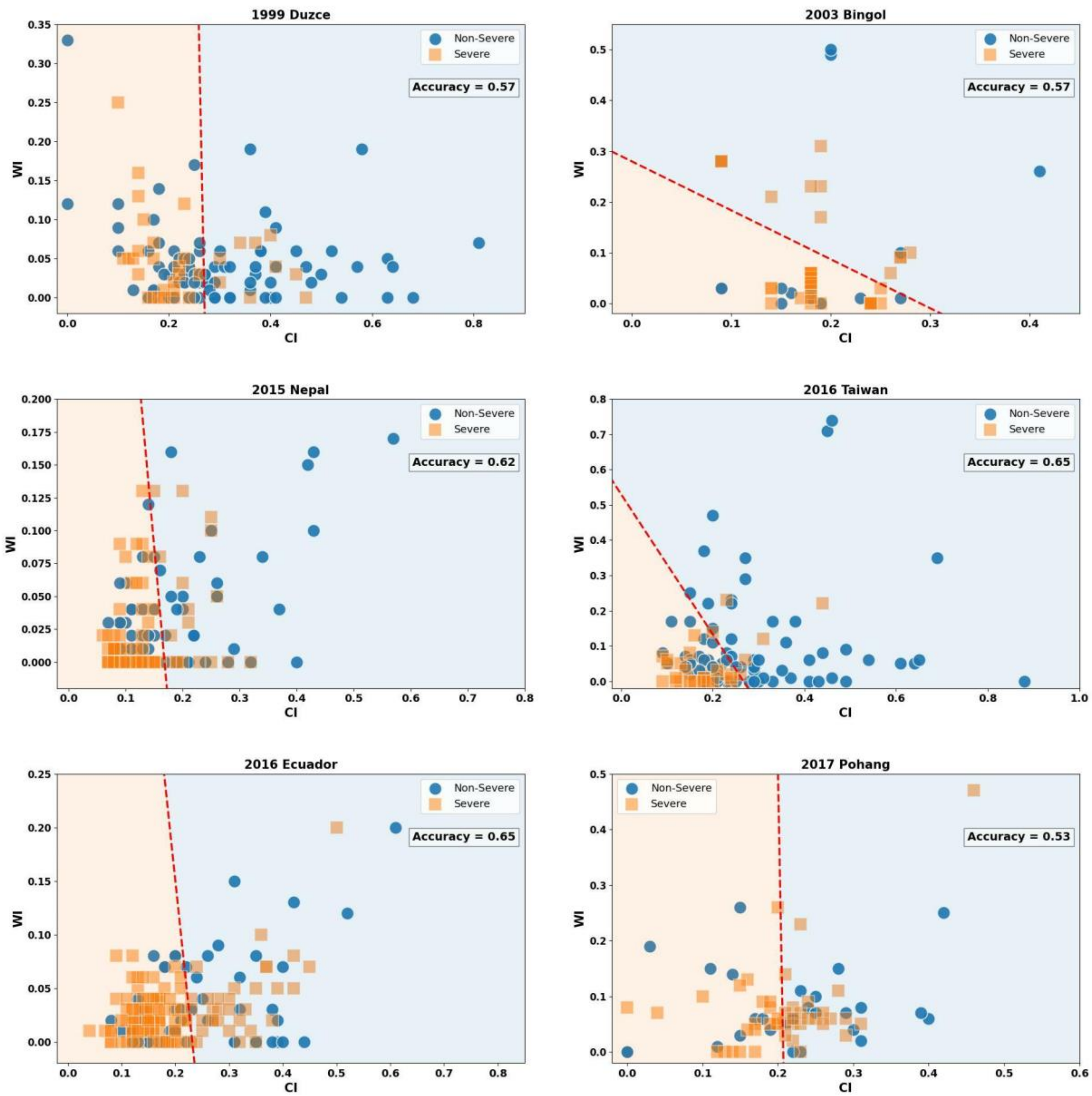


Figure 3.5: CI-WI scatter plots with classification boundaries defined by logistic regression

As can be seen, in none of the six datasets, the slope of the classification boundary is similar to the one proposed by Hassan and Sozen in Figure 2.4. Whereas the PI as defined in Equation (2.1) assumes equal contributions from CI and WI, the damage classification boundaries shown in Figure 3.5, suggest otherwise. The boundary is nearly vertical for Duzce (1999) and Pohang (2017) which suggests WI is essentially irrelevant in these cases when predicting seismic damage class using the PI. Hassan and Sozen recommended defining classification boundaries based on judgment taking into consideration risk tolerance and availability of resources. However, for region-specific screening, logistic regression can be used to easily customize the weighting on CI and WI and define the classification boundary that most accurately identifies buildings that are most likely to be severely damaged. Overall, adjusting the relative weighting of CI and WI resulted in only marginal improvements in accuracy. This finding suggests that the weights assigned to wall and column areas in WI and CI (Equations (2.2) and (2.3)) by Hassan and Sozen (1997) require further calibration to achieve optimal performance. Rather than taking an analytical approach to calibrate the weights for each region, a data-driven approach using logistic regression and RF was used to develop models that directly use the seven raw geometric parameters instead of considering the CI and WI.

3.4.2 ML-based Vulnerability Evaluation

The seven geometric parameters used in computing the PI are: the number of stories, floor area, column cross-sectional areas, and areas of concrete walls and masonry infill walls in two horizontal directions, E-W and N-S. A screening method using these parameters directly would be more generally applicable and robust than the Hassan and Sozen approach since it does not rely on the calibration of the CI, WI, and PI. The seven parameters are used to develop region-specific models using logistic regression and RF. 5-fold cross validation is used for training and testing the models. Table 3.3 shows the accuracies for the logistic regression models.

Table 3.3: Region-specific logistic regression model accuracies

1999 Duzce				**2003 Bingol**			
Damage Class	Precision	Recall	F1	**Damage Class**	Precision	Recall	F1
Non-Severe	0.86	0.72	0.78	Non-Severe	0.52	0.73	0.61
Severe	0.51	0.71	0.59	Severe	0.88	0.75	0.81
Weighted Average Accuracy: 0.73				**Weighted Average Accuracy: 0.76**			
2015 Nepal				**2016 Taiwan**			
Damage Class	Precision	Recall	F1	**Damage Class**	Precision	Recall	F1
Non-Severe	0.60	0.67	0.63	Non-Severe	0.92	0.57	0.70
Severe	0.67	0.59	0.63	Severe	0.43	0.86	0.57
Weighted Average Accuracy: 0.63				**Weighted Average Accuracy: 0.67**			

Model Accuracy-2016 Ecuador				Model Accuracy-2017 Pohang			
Damage Class	Precision	Recall	F1	**Damage Class**	Precision	Recall	F1
Non-Severe	0.48	0.75	0.59	Non-Severe	0.45	0.43	0.44
Severe	0.85	0.64	0.73	Severe	0.62	0.64	0.63
Weighted Average Accuracy: 0.69				**Weighted Average Accuracy: 0.55**			

In general, using the seven raw geometric parameters as input features greatly improved the accuracy of region-specific vulnerability predictions compared to predictions based on PI. As summarized in Table 3.3, logistic regression models yielded weighted average accuracies of 73%, 76%, 63%, 67%, 69%, and 55% for Duzce, Bingol, Nepal, Taiwan, Ecuador, and Pohang, respectively. The relatively low accuracy for the 2017 Pohang earthquake is attributable to the fact that piloti-type buildings (vertical irregularity) are common in the region, many of which were severely damaged due to the soft-story effect [75], [76].

RF is a robust ML algorithm with advantageous features such as fast training, scalability, and relatively high accuracy on small datasets. Its robustness against overfitting makes this algorithm particularly useful in this study. The process of training and testing the region-specific models is repeated using RF, with the accuracies summarized in Table 3.4.

Table 3.4: Region-specific RF model accuracies

1999 Duzce				**2003 Bingol**			
Damage Class	Precision	Recall	F1	**Damage Class**	Precision	Recall	F1
Non-Severe	0.86	0.67	0.75	Non-Severe	0.33	0.50	0.4
Severe	0.40	0.67	0.5	Severe	0.88	0.78	0.82
Weighted Average Accuracy: 0.72				**Weighted Average Accuracy: 0.78**			
2015 Nepal				**2016 Taiwan**			
Damage Class	Precision	Recall	F1	**Damage Class**	Precision	Recall	F1
Non-Severe	0.86	0.67	0.75	Non-Severe	0.68	0.93	0.79
Severe	0.54	0.78	0.64	Severe	0.67	0.25	0.63
Weighted Average Accuracy: 0.70				**Weighted Average Accuracy: 0.71**			
2016 Ecuador				**2017 Pohang**			
Damage Class	Precision	Recall	F1	**Damage Class**	Precision	Recall	F1

Non-Severe	0.44	0.80	0.57	Non-Severe	0.25	0.20	0.22
Severe	0.96	0.83	0.89	Severe	0.64	0.70	0.67
Weighted Average Accuracy: 0.71				**Weighted Average Accuracy: 0.60**			

As can be seen in Table 3.4, RF models are more accurate than logistic regression for region-specific screening. This was somewhat expected since RF can identify more complex and nonlinear relationships between the input features and the output. Overall, these results clearly demonstrate that while the original weighting used by Hassan and Sozen to define the CI, WI and PI require region-specific calibration, the raw input parameters on their own can be highly effective indicators for preliminary seismic vulnerability screening.

3.4.3 Generalized RVS

Although the ML-based models developed in the previous section are able to directly predict the seismic vulnerability of low-rise RC buildings, they are region-specific and are not interchangeable between earthquakes. Many factors could affect building performance including building age, structural details, construction quality, and site conditions. However, these factors are difficult to determine in inventory surveys or for rapid preliminary assessment purposes. Outside of structural and site conditions for which no information was available for the buildings considered in this study, earthquake intensity is another key parameter that can help explain the differences in damage outcomes in the datasets. The addition of earthquake intensity measures as input features in the RF model could make the classifier more accurate and widely applicable.

In order to develop a more generalized approach that can be applied in any region but still keep the simplicity of RVS methods, different intensity features including MMI, PGA, and peak ground velocity (PGV) are assessed as new features in addition to the seven PI input parameters. MMI is a seismic intensity scale developed by the American seismologists Harry Wood and Frank Neumann (1931) [77]. Instead of measurements or a mathematical basis, the MMI is a ranking based on the observed effects. PGA and PGV are the maximum ground acceleration and velocity during an earthquake shaking at a location. MMI, PGA, and PGV values at the site of each building in the datasets were obtained from US Geological Survey (USGS) ShakeMaps [78], [79], [80], [81], [82], [83], [84], [85]. The MMI for the six earthquakes considered in this study range between 6 and 9 while the PGA and PGV range between 0.15g and 0.89g, and 5 and 76.5 cm/s, respectively.

To assess the applicability of the intensity parameters as extra input features in the classifier, a leave-one-out cross validation approach is used in which five of the six datasets are used as a training set for an RF model and the remaining set is used as a testing set. The RF models are trained with and without considering the intensity measures as new input features and the model accuracies are compared. The six possible scenarios ($\binom{6}{1} = 6$) are as below:

- **Scenario 1:** Train on the Bingol, Nepal, Taiwan, Ecuador, Pohang datasets and test on the Duzce dataset
- **Scenario 2:** Train on the Duzce, Nepal, Taiwan, Ecuador, Pohang datasets and test on the Bingol dataset
- **Scenario 3:** Train on the Duzce, Bingol, Taiwan, Ecuador, Pohang datasets and test on the Nepal dataset
- **Scenario 4:** Train on the Duzce, Bingol, Nepal, Ecuador, Pohang datasets and test on the Taiwan dataset
- **Scenario 5:** Train on the Duzce, Bingol, Nepal, Taiwan, Pohang datasets and test on the Ecuador dataset
- **Scenario 6:** Train on the Duzce, Bingol, Nepal, Taiwan, Ecuador datasets and test on the Pohang dataset

The RF model accuracies for the six different scenarios with and without adding the new input features are summarized in Table 3.5. The mean and standard deviation across the six scenarios are also shown. Overall, incorporating earthquake intensity measures enhances the model accuracy by accounting for the seismic impact on building vulnerability. Notably, adding PGA as an extra input feature can markedly improve the mean accuracy of the model from 46% to 52%. However, the greatest improvement to prediction accuracy is observed when both MMI and PGA are provided with the mean accuracy and standard deviation of 54% and 4%, respectively.

Table 3.5: Model accuracies after adding new input features to the seven PI inputs

Scenario Number	**New Input Feature**							
	-	**MMI**	**PGA**	**PGV**	**MMI & PGA**	**MMI & PGV**	**PGV & PGA**	**MMI, PGA & PGV**
1	0.56	0.51	0.57	0.52	0.53	0.50	0.51	0.55

2	0.43	0.45	0.54	0.37	0.54	0.38	0.57	0.59
3	0.43	0.40	0.47	0.47	0.47	0.48	0.47	0.54
4	0.57	0.53	0.57	0.57	0.59	0.53	0.58	0.59
5	0.28	0.30	0.36	0.30	0.55	0.35	0.42	0.46
6	0.50	0.55	0.62	0.53	0.54	0.58	0.59	0.51
Mean	**0.46**	**0.46**	**0.52**	**0.46**	**0.54**	**0.47**	**0.52**	**0.54**
Standard Deviation	**0.11**	**0.09**	**0.09**	**0.10**	**0.04**	**0.09**	**0.07**	**0.05**

Comparing the mean and standard deviation values, PGV did not have a significant impact on the model accuracy. Although PGV is strongly correlated to earthquake intensity, studies have shown PGV is a more effective predictor of structural damage for medium- to long-period structures [78], [86]. Based on these results, MMI and PGA are selected as two additional features and a new RF model is trained on a mixed dataset containing all the buildings in the study. The accuracy of this model is highlighted in Table 3.6. Similar to the region-specific models, 5-fold cross validation is employed to train and test the model, with four subsets used for training and one subset used for testing.

Table 3.6: Accuracy of the new RF model trained on a mixed dataset with MMI and PGA inputs

Mixed Dataset			
Damage Class	Precision	Recall	F1
Non-Severe	0.71	0.62	0.66
Severe	0.69	0.77	0.73
Weighted Average Accuracy: 0.71			

This generalized model is not region-specific yet achieves a high accuracy, comparable to region-specific models, in rapidly identifying high-risk low-rise RC buildings. For a given building inventory, using the raw geometric features and including MMI and PGA as input parameters offers a more pragmatic way to reflect regional seismicity and earthquake risk tolerance in vulnerability screening than relying on the CI, WI, and a classification boundary based on judgment.

3.4.4 SHAP Analysis for the Non-Region-Specific RF Model

To interpret the generalized RVS, SHAP is applied to the RF model to assess the contribution of each input feature. Figure 3.6 presents the SHAP summary plot, quantifying the average impact of each feature on the model's predictions.

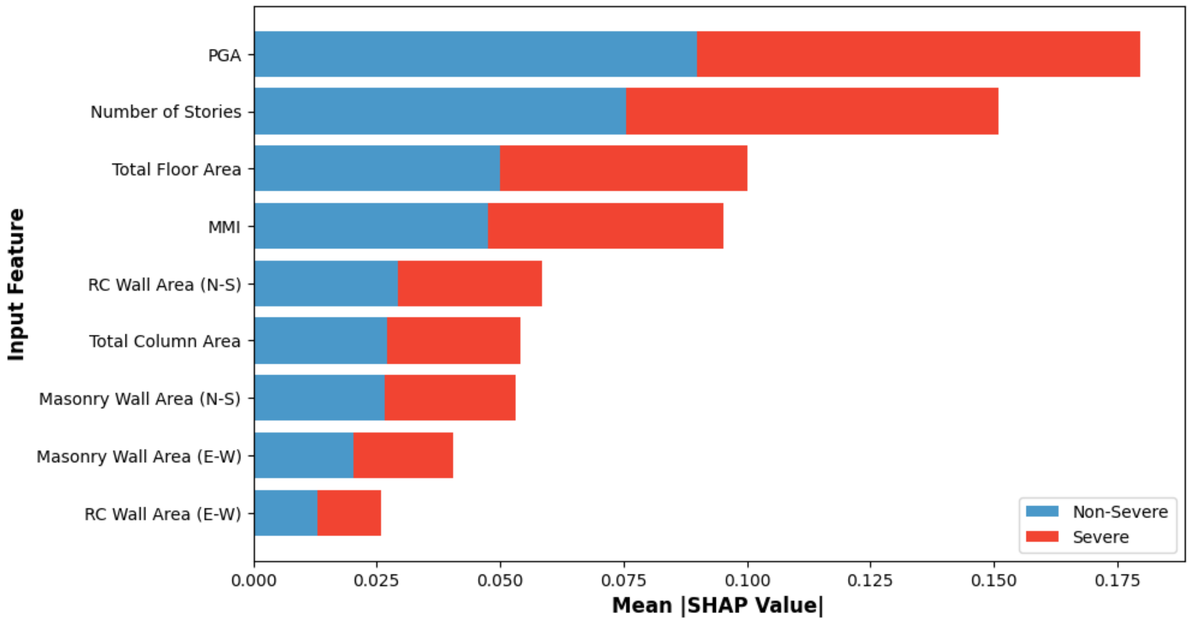


Figure 3.6: SHAP summary plot for the RF model

As can be seen, PGA, number of stories, total floor area, and MMI are the top influential features in damage classification. This highlights the significant role of seismic intensity and building size in determining structural performance. Specifically, higher PGA values and MMI levels correspond to more intense shaking, increasing the likelihood of severe damage. Additionally, taller buildings (number of stories) and larger structures (total floor area) tend to experience greater seismic demands, which can lead to higher vulnerability. Beyond these primary features, total column area, RC wall areas, and masonry wall areas also play critical roles in damage classification. Notably, walls oriented in the N-S direction show greater importance than those in the E-W direction. This trend may be due to several factors such as stronger reinforcement, fewer openings, building geometry, or the dominant direction of seismic forces in the studied regions.

3.5 Conclusions

RVS methods are commonly used to screen large building stocks to identify buildings that are highly vulnerable to earthquakes and prioritize actions. In this study, data from six earthquake events (Duzce (1999), Bingol (2003), Nepal (2015), Taiwan (2016), Ecuador (2016), and Pohang (2017)) were used to evaluate the effectiveness of the PI developed by Hassan and Sozen (1997) and ML techniques were employed to develop improved screening models for low-rise RC buildings. Analysis using logistic regression showed that PI, CI, and WI can be simple and useful indicators of seismic vulnerability. However, the weights applied to account for the stiffness of columns and walls in CI and WI, as well as the weights accounting for their relative contribution to overall building performance in PI are not universally applicable and need to be carefully calibrated for a specific region. Based on these observations, logistic regression and RF were used to train and test classification models that directly use the raw input building geometry parameters as features to predict seismic damage rather than the CI, WI, or PI. Region-specific screening models trained this way were able to accurately classify buildings likely to be severely and non-severely damaged without the need for an explicitly defined classification boundary. To further improve the ML-based RVS models and develop a generally applicable model that is not region-specific, seismic intensity parameters were introduced as additional input features. Incorporating MMI and PGA as input features in a non-region-specific model considerably improved the prediction accuracy and enables the consideration of regional seismicity and risk tolerance in the screening. An RF model was trained and tested on a mixed dataset containing all the buildings from the six datasets used in this study exhibited high accuracy. Furthermore, SHAP analysis was conducted to interpret the model, identifying PGA, number of stories, total floor area, and MMI as the most influential features in predicting the damage class. Outcomes from this study demonstrate the simple and easily attainable geometric parameters that form the basis of the PI are effective predictors of seismic vulnerability in low-rise RC buildings. With the use of ML techniques, the RVS based on these input features can be more widely applicable.

3.6 Limitations of the Study

While this study demonstrated significant advancements in seismic vulnerability assessment for low-rise RC buildings using ML, several limitations should be acknowledged. The datasets used were derived from six post-earthquake surveys, with their quality and consistency dependent on the expertise of the inspection teams. To simplify analysis, buildings with moderate damage were grouped into the severe damage class, which reduced model complexity but resulted in a loss of specificity and the ability to distinguish between moderate and severe damage states. The models primarily focused on structural parameters such as geometric features and seismic intensity, while other influential factors like construction quality, building age, and maintenance were excluded due to limited data availability. Misclassifications remain a critical concern: predicting a high risk building as low risk could lead to significant safety risks, while overestimating risk could unnecessarily increase costs. Additionally, uncertainty in the model's predictions highlights the need for further refinement to improve confidence and reliability in the assessments. Future research should aim to address these limitations by incorporating more diverse and comprehensive datasets, accounting for broader structural and non-structural factors, and expanding the models to other building types. Efforts to quantify and reduce prediction uncertainty will be crucial to enhance the robustness and practical applicability of this ML-based RVS method.

Chapter 4
An Enhanced Machine Learning-Based Rapid Visual Screening Framework Considering Model Uncertainty and Decision Threshold Optimization

This chapter corresponds to the following journal paper submitted for publication:

N. Elyasi and E. Kim, "An enhanced machine learning-based rapid visual screening framework considering model uncertainty and decision threshold optimization," 2024.

Abstract

RVS methods are commonly used to quickly filter large building inventories for at-risk structures, typically based on simple visual inspections, such as sidewalk surveys. In Chapter 3, an ML-based RVS method for low-rise RC buildings was developed, achieving an accuracy of 71% in identifying buildings at high risk of severe damage during an earthquake. However, uncertainty in the model's predictions remains a concern. This study refines the previously proposed RVS methodology by addressing model uncertainty and minimizing misclassifications. Two primary approaches are proposed: the first analyzes class probabilities from the ML-based screening model to assess the prediction uncertainty rather than relying on the final predicted damage class. With this approach, buildings for which the ML model shows high uncertainty can be prioritized for more detailed evaluation. The second approach aims to optimize the decision threshold used by the ML model to more accurately identify buildings at risk of severe damage. This is done by evaluating the relative cost of misclassifications, low-risk buildings identified as high-risk (false positives) and high-risk buildings identified as low-risk (false negatives). Building on the findings, this study proposes a comprehensive three-grade ML-based methodology for enhanced rapid seismic vulnerability assessments.

4.1 Introduction

The RVS method proposed in the previous chapter achieved a reasonable performance using only basic building features. While this result is promising, further refinements to minimize misclassifications are essential to enhance the proposed model's reliability. Improving prediction accuracy, particularly for the severe damage class, is vital given the potential consequences. Misclassifying a vulnerable building as safe could result in significant financial losses and pose serious risk to lives during an earthquake, while inaccurately labeling a structurally sound building as at risk would incur additional evaluation and analysis costs. To effectively persuade building owners to invest in necessary evaluations and retrofitting, it is crucial to develop a robust framework that clearly demonstrates how the costs of assessment and retrofitting can be significantly lower than the potential damages and rebuilding expenses that could arise from an earthquake. This urgency is particularly pronounced in high-risk seismic regions like California, where the probability of a major earthquake in the near future is exceedingly high. Identifying vulnerable buildings and implementing necessary actions, such as retrofitting, are critical to reducing these risks. However, this can only be achieved by convincing building owners through a reliable, cost-effective, and applicable procedure. Improving the model's accuracy and reliability will be essential in persuading stakeholders to adopt its recommendations, ultimately leading to safer buildings and more resilient communities.

This study aims to refine the previously proposed RVS methodology by addressing model uncertainty and reducing misclassifications in the severe damage class. Two approaches are proposed to achieve these objectives. The first approach focuses on model uncertainty. Model uncertainty refers to the uncertainty in predictions due to the limitations, assumptions, and approximations inherent in the model's structure. By analyzing the distribution of probabilities assigned by individual decision trees in the RF classifier to a building being in the severe damage class rather than relying solely on final predicted label, a more nuanced understanding of the model uncertainty is obtained. This involves reducing the maximum depth of decision trees in the RF classifier, limiting the longest path from the root node to a leaf node in the tree. This reduction helps avoid prediction purity, a condition where a leaf node contains samples that all belong to the same class, leading to more confident classifications that may indicate overfitting. By restricting the depth, class probabilities can be calculated from each decision tree instead of predicting a single class label. The overall probability for each damage class is then averaged across all trees in the forest. This strategy helps identify low, moderate, and high uncertainty levels in predictions. Buildings with high prediction uncertainty should be prioritized for further inspection and analysis, while those with low uncertainty can be classified with greater confidence. By focusing on probability distributions, the methodology enhances the reliability of seismic vulnerability assessments and helps mitigate potential risks associated with misclassifications. The second approach aims to optimize the decision threshold of the classifier to enhance model reliability by minimizing misclassifications for buildings at high risk. This is performed by considering the relative cost of misclassifying low-risk buildings as high-risk and those at high-risk as low-risk. A wide range of cost ratios are used to estimate the total misclassification cost and the effect of changing the decision threshold on prediction accuracy and the cost of misclassification is examined. Finally, the two refinement approaches are incorporated into the original ML model to propose a comprehensive three-grade ML-based RVS methodology.

4.2 Model Uncertainty Detection

Understanding and evaluating uncertainty is crucial for assessing the reliability of an ML model's predictions. There are two main types of uncertainty: data uncertainty and model uncertainty. Data uncertainty comes from the natural variation or noise in the data itself. This type of uncertainty is caused by factors that the model cannot account for, such as measurement errors, incomplete data, or randomness in data collection. In this study, differences in the quality or reporting standards of the six earthquake datasets used for training and testing the model represent examples of data uncertainty. Also, differences in how accurately the geometric features were measured (e.g., floor area, column and wall areas), and how the inspectors judged the level of damage for buildings in these datasets are also notable sources of data uncertainty. Model uncertainty comes from the limitations within the model itself. It can be caused by factors like the model's structure, assumptions made during its development, or the limited amount of training data. In the case of the RVS methodology developed by the authors, one source of model uncertainty is the fact that the model makes predictions based on a small set of coarse geometric features that do not directly represent the structural design or condition. Additionally, the model may not have clear decision boundaries for classification. For example, if the model encounters a building with a unique combination of features, it may be unsure whether to classify it as severe or non-severe, especially if there are no similar examples in the training data. Given that addressing data uncertainty is often difficult due to limited control over data quality and availability, this study focuses primarily on addressing model uncertainty.

In the ML-based RVS developed in the previous chapter, misclassifying a building that is likely to sustain non-severe damage as severe may simply lead to higher costs associated with a more detailed inspection and evaluation, incorrectly predicting a building at risk to be severely damaged in an earthquake as non-severe poses significant safety risks and could result in substantial losses in future earthquakes. Therefore, addressing the uncertainty in the classifier's predictions is essential for mitigating these risks. To address this uncertainty, the probabilities assigned to both damage classes (non-severe and severe) by the RF classifier during prediction are examined rather than relying solely on the final predicted label. Figure 4.1 shows how these probabilities are calculated, providing an overview of the RF classifier's structure.

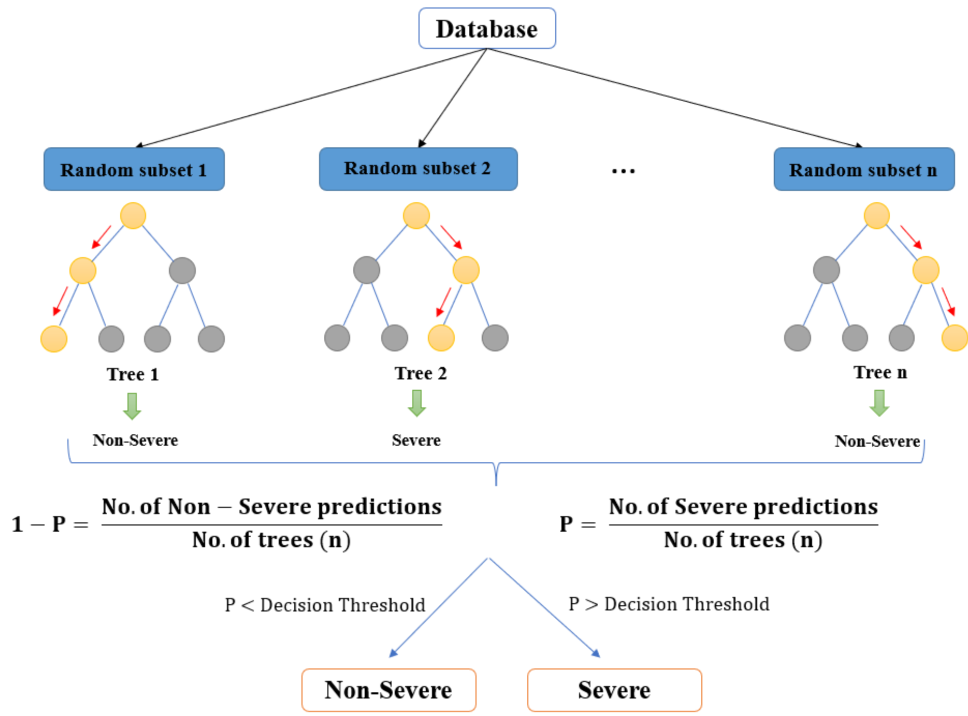


Figure 4.1: The process of calculating damage class probabilities in the RF classifier

Typically, the RF classifier makes predictions based on the class with the highest probability, using a decision threshold of 0.5 for binary classification. If the probability of a building being severely damaged exceeds this decision threshold, it is classified as severe; otherwise, it is categorized as non-severe. By focusing on probabilities rather than final labels, this approach moves beyond simple categorization of potential earthquake damage as severe or non-severe. It provides a more detailed understanding of the likelihood of each class and improves the overall assessment process by enabling more informed and careful decision-making regarding the building's vulnerability. To further illustrate this point, three examples are presented.

Table 4.1 outlines the characteristics of three dataset samples where the model incorrectly classified the damage class. The assigned probabilities for each class (non-severe and severe), the predicted damage class based on a 0.5 decision threshold, and the actual observed damage are provided in Table 4.2 This approach allows for a deeper examination of the model's uncertainty.

Table 4.1: Characteristics of the three buildings listed in Table 4.2

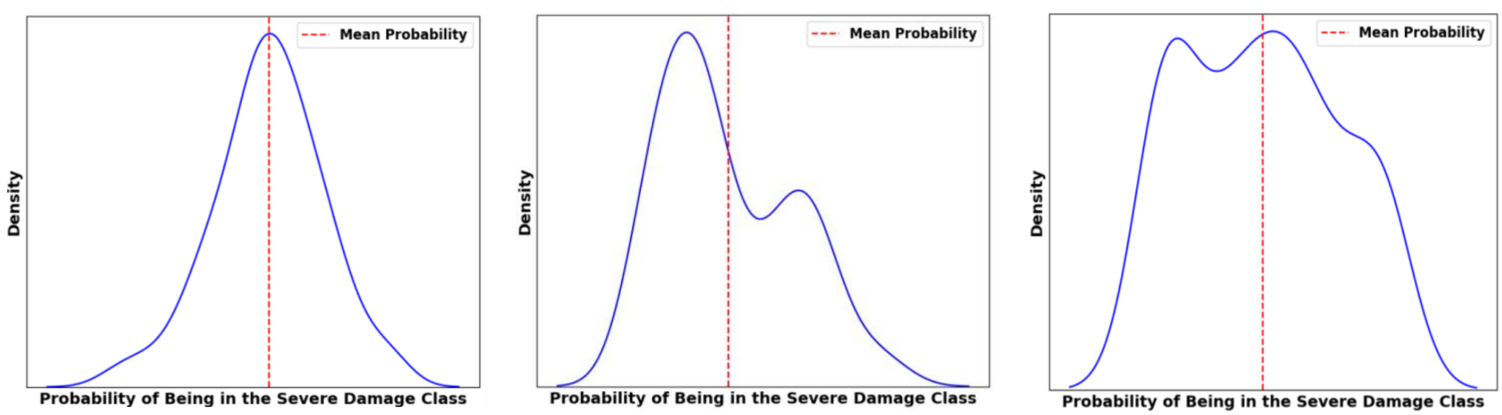
Sample	Earthquake	No. of Stories	Floor Area (m^2)	Column Area (m^2)	Concrete Wall Area (E-W) (m^2)	Concrete Wall Area (N-S) (m^2)	Masonry Wall Area (E-W) (m^2)	Masonry Wall Area (N-S) (m^2)	MMI	PGA
Building 1	Ecuador (2016)	4	148	2.6	0	0	3	4.8	7.3	0.32g
Building 2	Duzce (1999)	1	77	0.72	0	0	0	0	8.4	0.46g
Building 3	Pohang (2017)	4	246.25	3.24	0.416	1.05	0	0	6.2	0.26g

Table 4.2: Predicted class probabilities and labels for three test set samples

Sample	Probability of being in the non-severe damage class (1-P)	Probability of being in the severe damage class (P)	Predicted damage class	Actual damage class
Building 1	0.09	0.91	Severe	Non-Severe
Building 2	0.87	0.13	Non-Severe	Severe
Building 3	0.51	0.49	Non-Severe	Severe

In the case of Building 1, the model predicts a probability of severe damage as 0.91, giving it a final label of severe, although the actual damage was non-severe. This is a conservative but incorrect prediction by the model. Conversely, for Building 2, the model assigns a probability of severe damage as 0.13, classifying it as non-severe, while the actual damage in the earthquake was severe. Lastly, Building 3 presents a probability of severe damage at 0.49, marginally below the 0.5 threshold resulting in a final label of non-severe earthquake damage but the actual damage was severe. In a real application of the ML-based RVS model, the mislabeling of Building 1 would simply lead to detailed analysis of the building and possibly seismic retrofits. In contrast, the false predictions for Buildings 2 and 3 could mean significant losses in a future earthquake and potentially even put lives at risk. In cases like Building 3 where the probability of severe damage is high and close to the threshold of 0.5, it may be prudent to exercise judgment and classify such buildings as at risk of severe damage. These examples clearly illustrate that relying on predicted probabilities rather than solely on the final predicted label can lead to better decisions.

To address uncertainty in predictions for cases like Building 3, the maximum depth of the decision trees is reduced to avoid reaching final label predictions. In an RF model, decision trees typically continue to split until they reach leaf nodes that assign a definitive class label. By limiting the maximum depth, the trees are prevented from growing too complex and forcing a final classification at the leaf nodes. This adjustment allows the calculation of probabilities for each damage class from every tree in the forest, rather than just predicting a single class label. The probability distribution of a building being in the severe damage class can be visualized using histograms, where the height of each bar reflects the frequency (count) of probability estimates provided by the trees in the forest for a specific range (bin). A curve can be fitted to the histogram to represent a continuous probability density function, offering a visual representation of the overall shape of the probability distribution. A margin of 10% around 0.5 is selected in order to focus on cases like Building 3, so the test set samples with predicted probabilities of being in the severe class between 0.4 and 0.6, obtained from the original classifier, are examined. This margin can be adjusted according to user needs. Figure 4.2 illustrates three types of probability distributions for a building being in the severe damage class using the decision tree probabilities. The red dashed vertical line indicates the mean probability, which is the average of the probabilities predicted by all trees in the forest for a given building being classified as severely damaged.



a **b** **c**

Figure 4.2: Three types of probability distributions for a sample from the severe damage class

Figure 4.2.a shows a typical probability distribution with low uncertainty**.** The probability mass is concentrated around the mean probability, leading to a clear final class prediction. In contrast, in the probability distribution shown in Figure 4.2.b the probability mass is focused around two distinct points rather than around the mean probability. This bimodal distribution suggests moderate uncertainty, as the prediction is less clear. Lastly, the probability distribution in Figure 4.2.c is more uniform,

indicating that the probability mass is spread out rather than concentrated around the mean probability. This implies high uncertainty in the prediction, making the final class significantly less certain.

4.3 Threshold Optimization

In the context of ML classification, a decision threshold is a specific probability value that determines how predicted probabilities from a model are converted into class labels. However, a default threshold of 0.5 may not always be optimal for every classification problem. Adjusting the threshold can enhance model performance and reduce misclassifications, especially when the consequences of misclassifying one class are more critical than another. This can make the model more sensitive to important classes, thereby reducing the impact of costly errors. In this study, optimization of the decision threshold is explored as a means to address model uncertainty and minimize misclassifications. To identify this optimal threshold, the decision threshold is changed between 0 and 1 at increments of 0.05 and the cost of misclassifications is estimated for the predictions made based on each threshold. The total misclassification cost is calculated as follows:

$$\text{Total misclassification cost} = \mathrm{N_{FP}} \times \mathrm{C_{FP}} + \mathrm{N_{FN}} \times \mathrm{C_{FN}} \quad \text{(Equation 4.1)}$$

Where $\mathrm{N_{FP}}$ and $\mathrm{C_{FP}}$ respectively are the number of cases and cost associated with low-risk buildings incorrectly classified as high risk of severe earthquake damage (false positives) while $\mathrm{N_{FN}}$ and $\mathrm{C_{FN}}$ are the number of cases and cost associated with high-risk buildings misclassified as low risk (false negatives). The value of $\mathrm{C_{FP}}$ includes expenses such as inspections, analyses, temporary housing, and business interruption while $\mathrm{C_{FN}}$ would encompass costs related to severe earthquake damage including loss of life, building repair or reconstruction, and broader social impacts such as disruptions to the economy, education, healthcare, and the displacement of communities. It is important to note that the values of $\mathrm{C_{FP}}$ and $\mathrm{C_{FN}}$ are not universal; they vary based on regional factors and specific building characteristics. Therefore, ranges of relative costs are considered instead of precise values in this study, but values can be specified by users according to their particular goals and the unique characteristics of their region. This cost-based adjustment of the decision threshold ensures optimizing the seismic screening through trade-offs between the costs of false positives and false negatives, ultimately leading to more robust and cost-effective predictions in seismic vulnerability assessments.

Due to the focus of this study on accurately predicting the severe damage class, an accuracy metric is required to assess the performance of the severe class as the threshold varies. To maintain consistency with the original RF classifier evaluation, the F1-score is used to measure the accuracy of the severe class. Additionally, the precision equation is modified to incorporate $\mathrm{C_{FP}}$ and $\mathrm{C_{FN}}$:

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \frac{\mathrm{C_{FP}}}{\mathrm{C_{FN}}} \cdot \mathrm{FP}} \qquad \text{(Equation 4.2)}$$

The objective of this approach is to identify the optimal decision threshold that minimizes total misclassification cost while ideally maximizing or achieving high accuracy for the severe damage class.

4.4 Results and Discussion

4.4.1 Results on Model Uncertainty

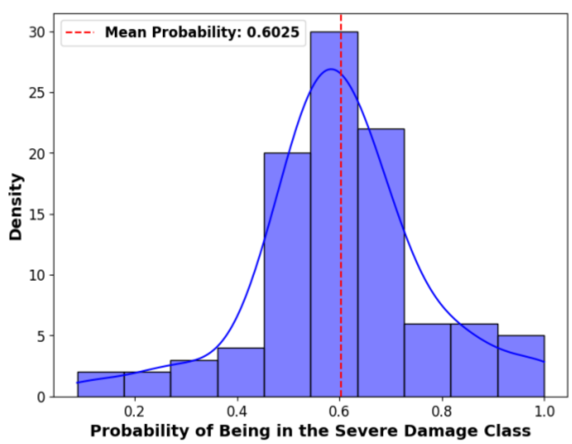
In analyzing model uncertainty using the dataset from the previous chapter, a 5-fold cross-validation technique is employed. Given the impracticality of showing probability distributions for all buildings in the test set, four building examples are presented for each level of uncertainty. These examples were carefully selected from the test set to ensure that all possible scenarios within their respective uncertainty levels are represented. The characteristics of these buildings are detailed in Table 4.3. Figures 4.3, 4.4, and 4.5 illustrate representative building examples of different levels of uncertainty.

Table 4.3: Characteristics of the building examples illustrated in Figures 4.3, 4.4, and 4.5

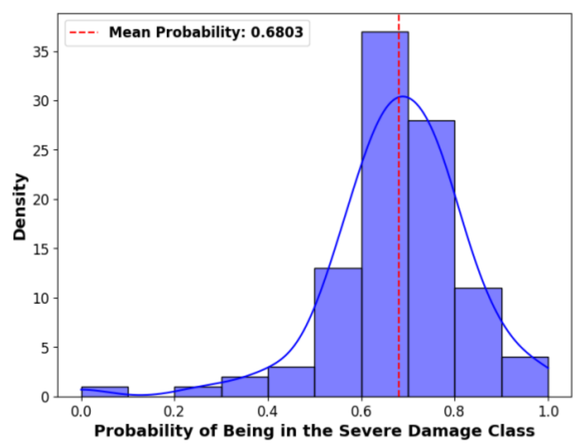
Uncertainty Level	Example Building Number	No. of Stories	Floor Area (m^2)	Column Area (m^2)	Concrete Wall Area (East-West) (m^2)	Concrete Wall Area (North-South) (m^2)	Masonry Wall Area (East-West) (m^2)	Masonry Wall Area (North-South) (m^2)	MMI	PGA
Low	**1**	4	192	2.84	0	0	0	3.4	8.1	0.31g
	2	3	235	1.8	0	0	0	2.7	7.3	0.32g
	3	2	175	1.1	0	0	2.7	3.1	7.6	0.46g
	4	4	134	1.87	0.54	0.79	0	0	7.7	0.48g
Moderate	**5**	1	114	1	0	0	0.9	1.6	7.3	0.32g
	6	2	300	3.7	0	0	3	3.2	7.4	0.34g
	7	3	124	3.75	0.2	0.11	0	0	7.7	0.48g

	8	4	193	3	0	0	0	5	6.5	0.16g
High	**9**	2	317	2.3	0	0	3.1	3.3	7.8	0.52g
	10	5	830	8	0	9.6	0	0	6.5	0.17g
	11	2	366	3.1	0	0	2	5.9	7.4	0.34g
	12	3	650	3.9	1.8	8.6	1.8	2.4	6.2	0.25g

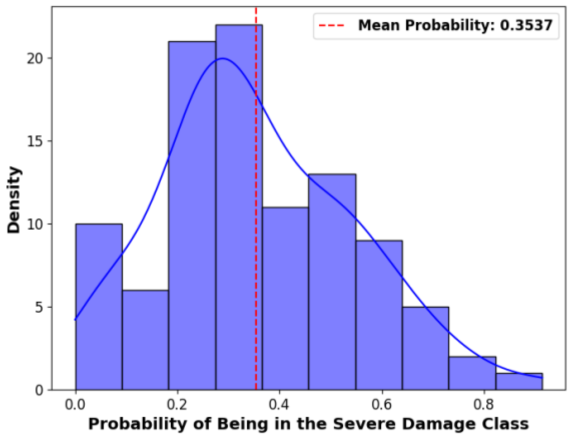
Low Uncertainty (Figure 4.3): In scenarios of low uncertainty, the probability distribution is sharply concentrated around the mean probability or a value very close to it. This concentration allows for a more confident classification based on the mean probability. However, despite this high confidence, some predictions may still be incorrect due to the model's inherent error. This error arises because there will always be a margin of misclassification due to factors such as data uncertainty and inherent variability in the dataset. In the RF model, errors can also result from its simplicity and the fact that seismic vulnerability of a building is influenced by many other parameters not typically accounted for in RVS methodologies. For instance, Example buildings 1 and 2 each have respectively mean probability of approximately 0.6 and 0.68, leading to their classification as severe since they exceed the threshold of 0.5. In contrast, Example building 3, with a mean probability around 0.35, falls below this threshold and is therefore classified as non-severe. Example building 4, however, presents a more complex case: its mean probability is about 0.58, resulting in a classification of severe, even though its actual damage class is non-severe. This highlights the inevitable misclassification errors that can occur in the model, even in situations of low uncertainty.



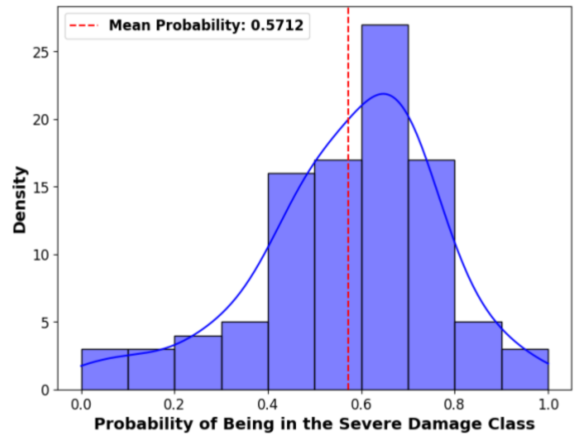
Example building 1) Actual damage class: Severe



Example building 2) Actual damage class: Severe



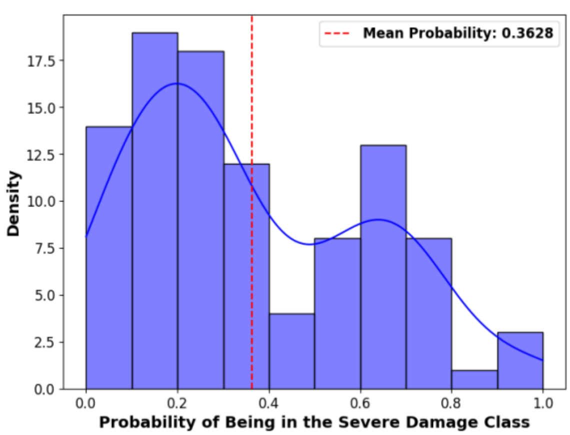
Example building 3) Actual damage class: Non-Severe



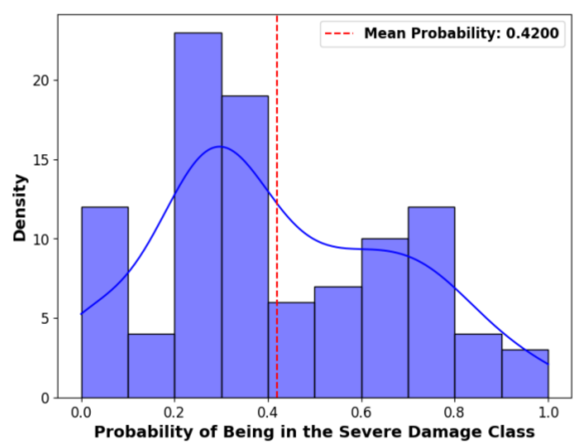
Example building 4) Actual damage class: Non-Severe

Figure 4.3: Example buildings for low uncertainty

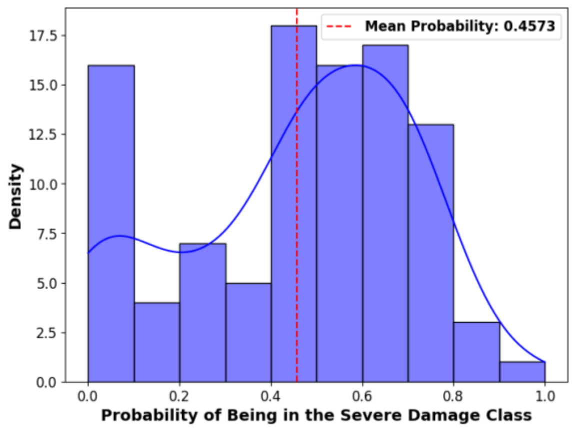
Moderate Uncertainty (Figure 4.4): Example buildings 5-8 exhibit probability distributions concentrated around two distinct points rather than a single mean probability. This bimodal distribution pattern introduces complications in classification and reduces the model's confidence in its predictions. In each Example building, one concentration point lies below the threshold of 0.5, while the other exceeds it, and both are positioned at a distance from the mean probability. These two-point concentrations indicate that the classifier's decision-making is not straightforward, making it difficult to rely solely on mean probabilities for accurate predictions.



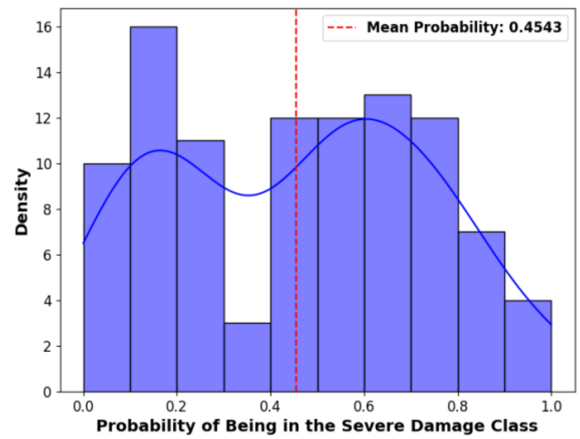
Example building 5) Actual damage class: Severe



Example building 6) Actual damage class: Severe



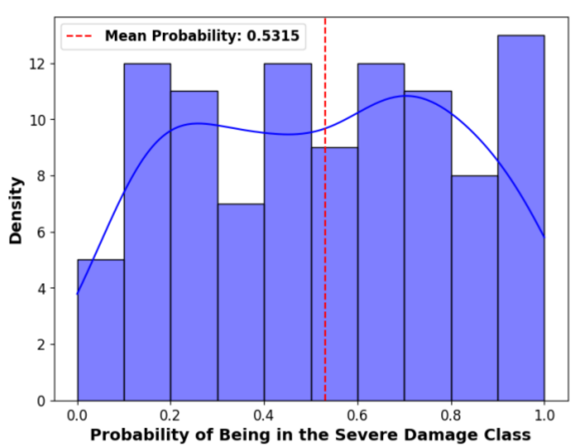
Example building 7) Actual damage class: Non-Severe



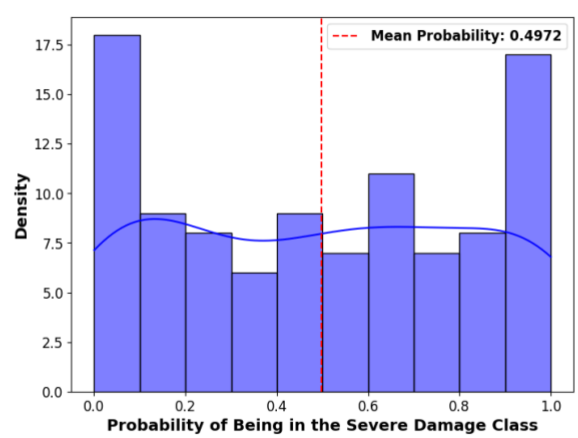
Example building 8) Actual damage class: Non-Severe

Figure 4.4: Example buildings for moderate uncertainty

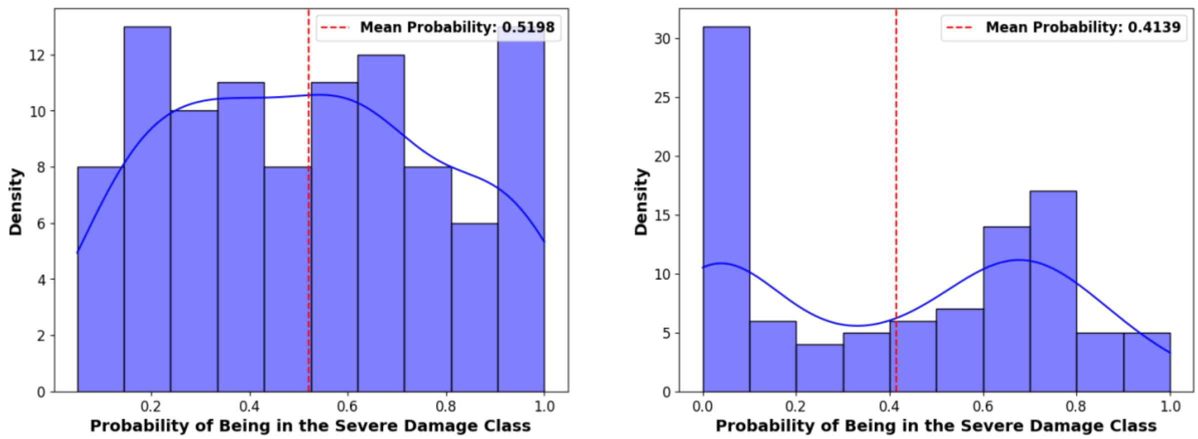
High Uncertainty (Figure 4.5): As indicated by the probability distributions the damage state predictions for Example buildings 9-12 show high uncertainty. The probability distributions are nearly uniform, meaning that the model provides nearly equal likelihoods for both non-severe and severe classifications. This signifies that the classifier is uncertain and struggles to differentiate between the classes. In such cases, it becomes nearly impossible to make a definitive classification, and detailed evaluation is required to resolve the ambiguity. To effectively manage cases of moderate and high uncertainty, it is essential to categorize these samples separately for further action.



Example building 9) Actual class: Severe



Example building 10) Actual class: Severe



Example building 11) Actual class: Non-Severe

Example building 12) Actual class: Severe

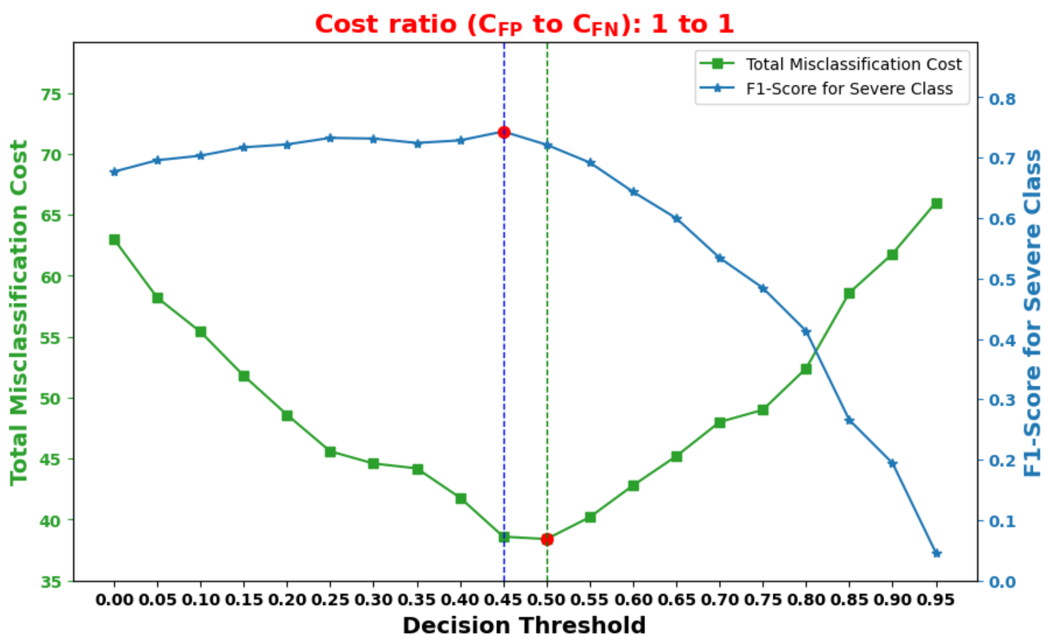
Figure 4.5: Example buildings for high uncertainty

Determining the uncertainty level of damage class predictions for buildings based on probability distributions is an effective method to prioritize high uncertainty cases and enhance the RVS method. However, manually analyzing and categorizing all these distributions based on judgment can be tedious and subjective. Therefore, the alternative approach of threshold optimization is discussed in the next section.

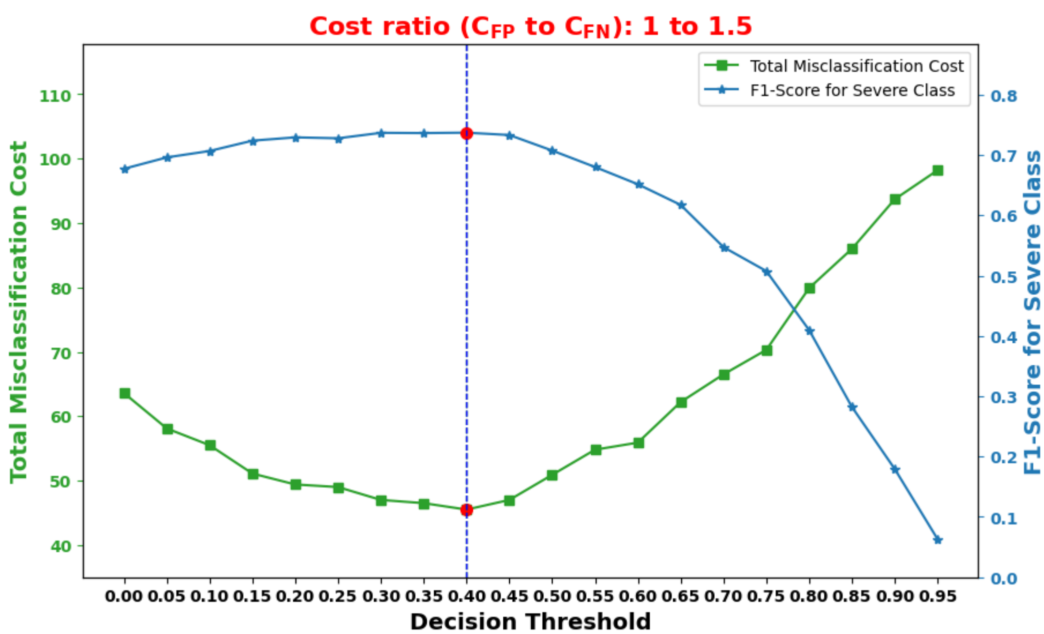
4.4.2 Results on Threshold Optimization

To address model uncertainty in classification, an alternative approach involves adjusting the decision threshold. Instead of using the traditional value of 0.5 for classification, an optimal threshold is determined to better suit the proposed classifier and its purpose. This adjustment is particularly important for reducing misclassifications in the severe damage class, which has more significant adverse consequences compared to misclassifications in the non-severe damage class. The goal is to select a threshold that minimizes the cost of false predictions while maintaining high accuracy for the model, especially for the severe damage class. To find this optimal threshold, a cost sensitivity analysis is conducted, considering the total cost of misclassifications by the model. For each fold of the 5-fold cross validation, the model is trained on the corresponding training set from the dataset used in the previous chapter. During training, cost coefficients, C_{FP} and C_{FN}, are applied as weights to samples in the non-severe and severe damage classes, respectively. These weights adjust the model's focus, guiding the decision trees to prioritize samples from the class with a higher associated cost. This ensures the model learns to account for the varying costs of false positives and false negatives during training. In the context of seismic vulnerability assessments, $\mathrm{C_{FN}}$ typically carries greater weight than $\mathrm{C_{FP}}$. In regions at high risk of significant earthquakes, where such assessments are conducted, the potential consequences of building collapse or severe damage are substantial. Therefore, it is reasonable to consider $\mathrm{C_{FN}}$ as equal to or greater than $\mathrm{C_{FP}}$ in these scenarios. Ratios higher than 1:5 are not considered because when $\mathrm{C_{FN}}$ is significantly greater than $\mathrm{C_{FP}}$, the optimal threshold tends to be very close to zero which indicates that the cost of incorrectly predicting a severely damaged building as non-severe (false negative) is so high that a detailed seismic evaluation and retrofit is preferred for every single building in the population.

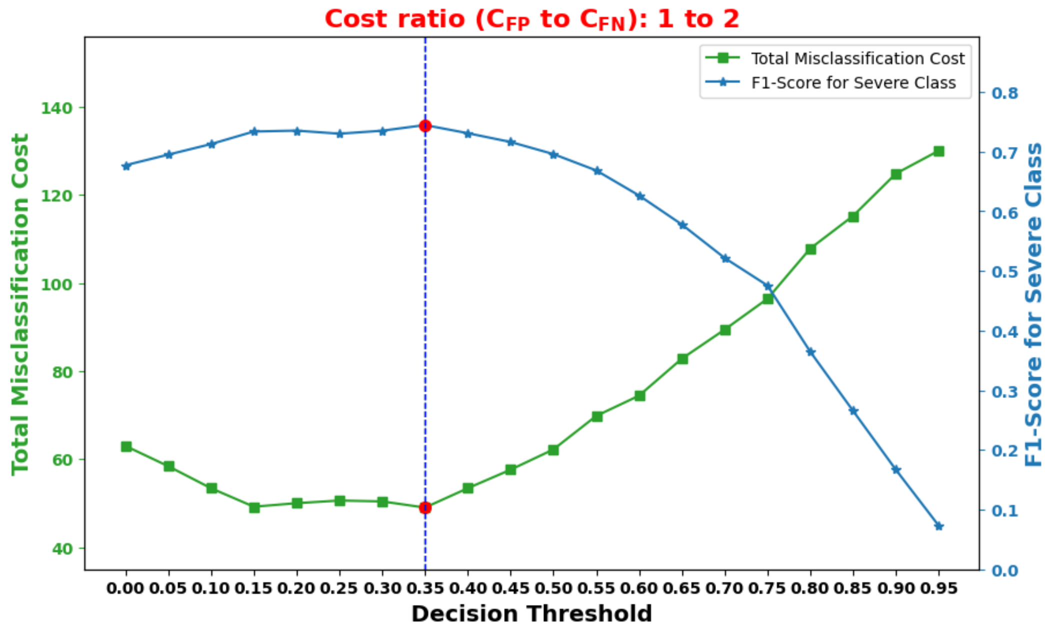
In this study, a range of $\mathrm{C_{FP}}$ to $\mathrm{C_{FN}}$ ratios (1:1, 1:1.5, 1:2, 1:2.5 up to 1:5) were explored to estimate the total misclassification cost instead of using exact cost figures since the data encompasses different regions and various building characteristics. This range was selected to reflect realistic conditions where $\mathrm{C_{FN}}$ is at least equal to $\mathrm{C_{FP}}$, aligning with seismic vulnerability assessment practices. Ratios beyond 1:5 were excluded, as optimizing the decision threshold becomes impractical when $\mathrm{C_{FN}}$ is significantly higher than $\mathrm{C_{FP}}$. However, if needed, additional ratios and a broader range can be explored using the same approach. For each $\mathrm{C_{FP}}$ to $\mathrm{C_{FN}}$ ratio, the trained model was evaluated on a test set using decision thresholds varying from 0 to 1 in increments of 0.05. The total misclassification cost and the F1-score for prediction accuracy of the severe damage class were computed for each case. These results are plotted in Figure 4.6. The optimal threshold is defined as the point where the misclassification cost is minimized and the F1-score for the severe damage class is maximized or is sufficiently high. The plots in Figure 4.6 provide a visual representation of how the decision threshold impacts the misclassification cost and the F1-score for the severe damage class, aiding in the identification of the optimal threshold. Each plot displays two curves: one representing the misclassification cost and the other representing the F1-score for the severe damage class. Vertical dashed lines highlight the threshold values where the misclassification cost is minimized and where the F1-score for the severe damage class is maximized. In most cases, the threshold that maximizes the prediction accuracy also minimizes the misclassification cost. In general, as the $\mathrm{C_{FP}}$ to $\mathrm{C_{FN}}$ ratio decreases, the optimum threshold decreases which means the ML model will predict more buildings to be severely damaged to reduce costs associated with false negatives. This also means that the model will yield more false positives (low-risk buildings identified as high-risk), but it will increase the F1-score for the severe damage class.

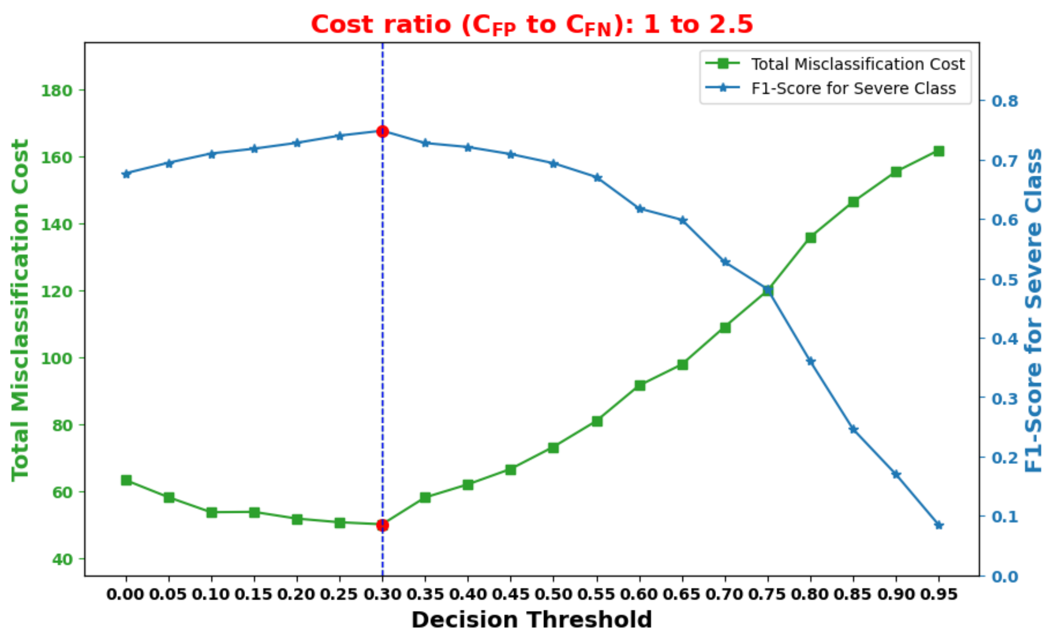


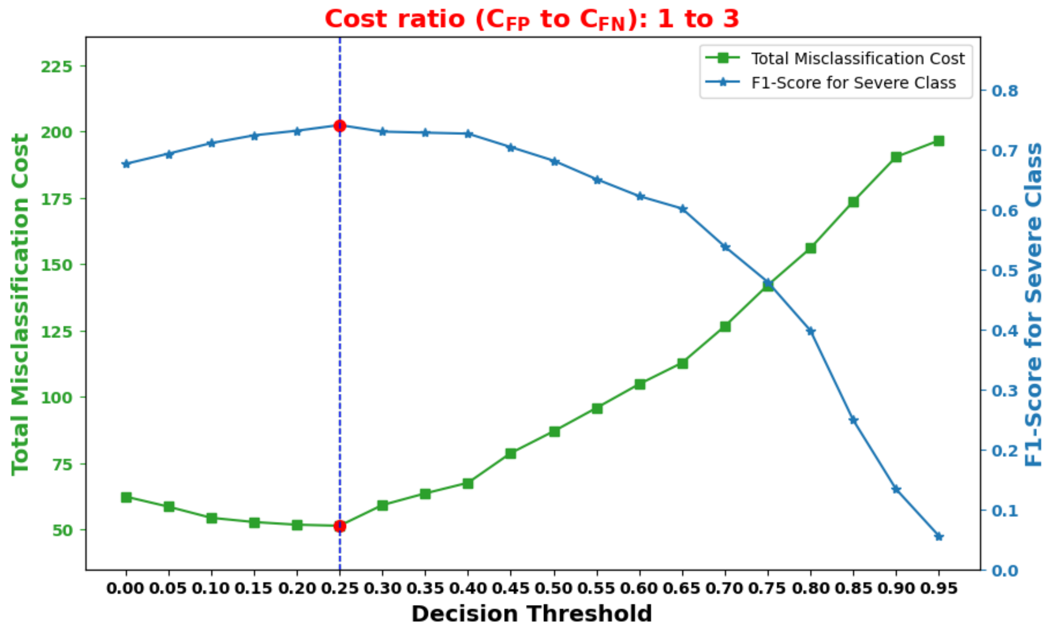
Cost ratio (C_{FP} to C_{FN}): 1 to 1
Total Misclassification Cost
F1-Score for Severe Class
Decision Threshold

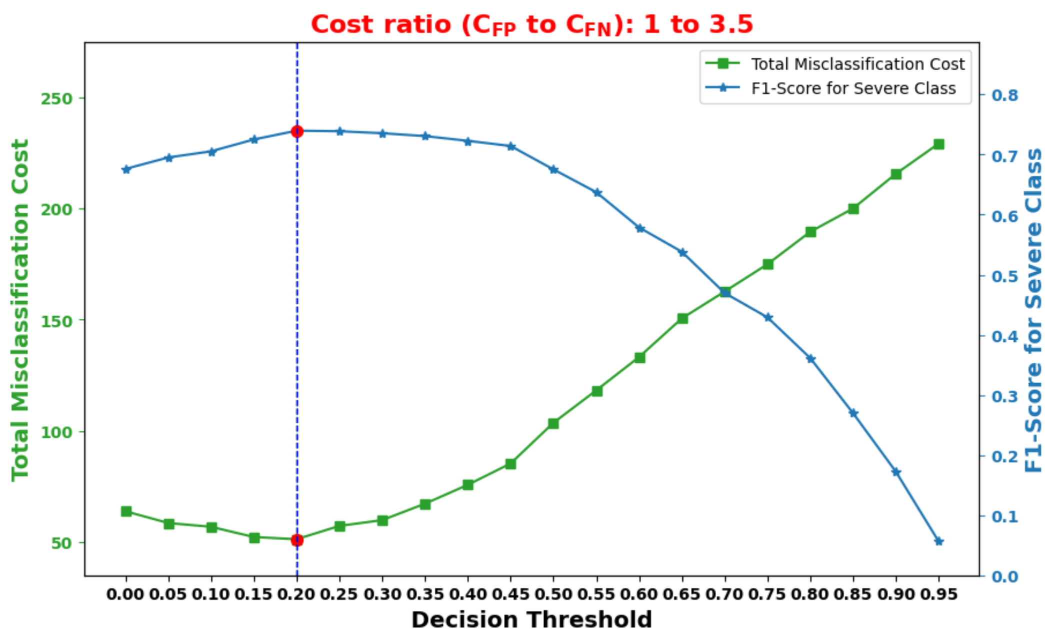


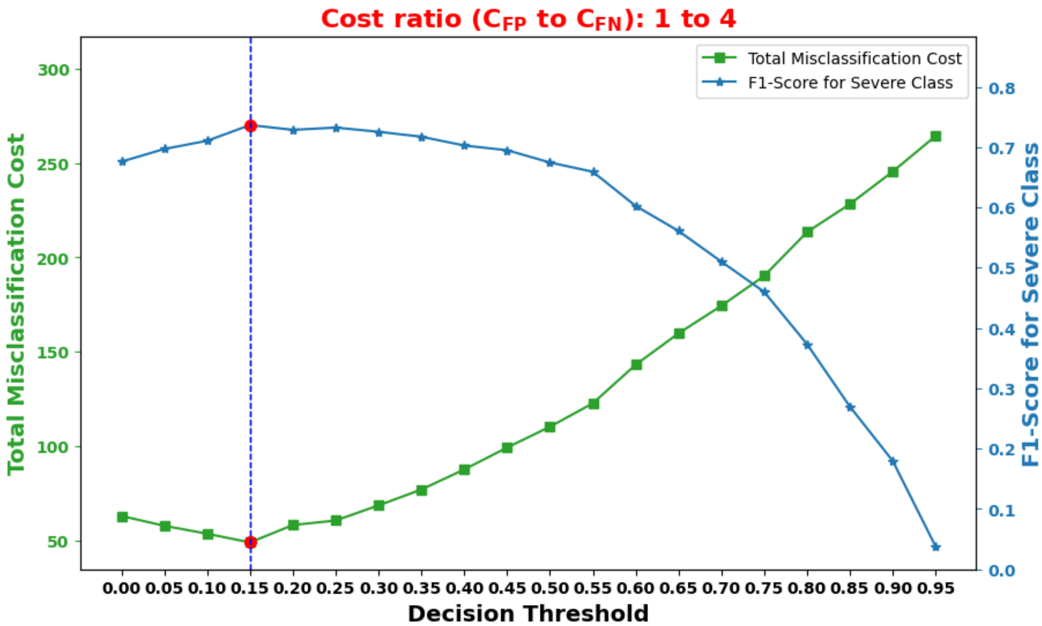
Cost ratio (C_{FP} to C_{FN}): 1 to 1.5
Total Misclassification Cost
F1-Score for Severe Class
Decision Threshold

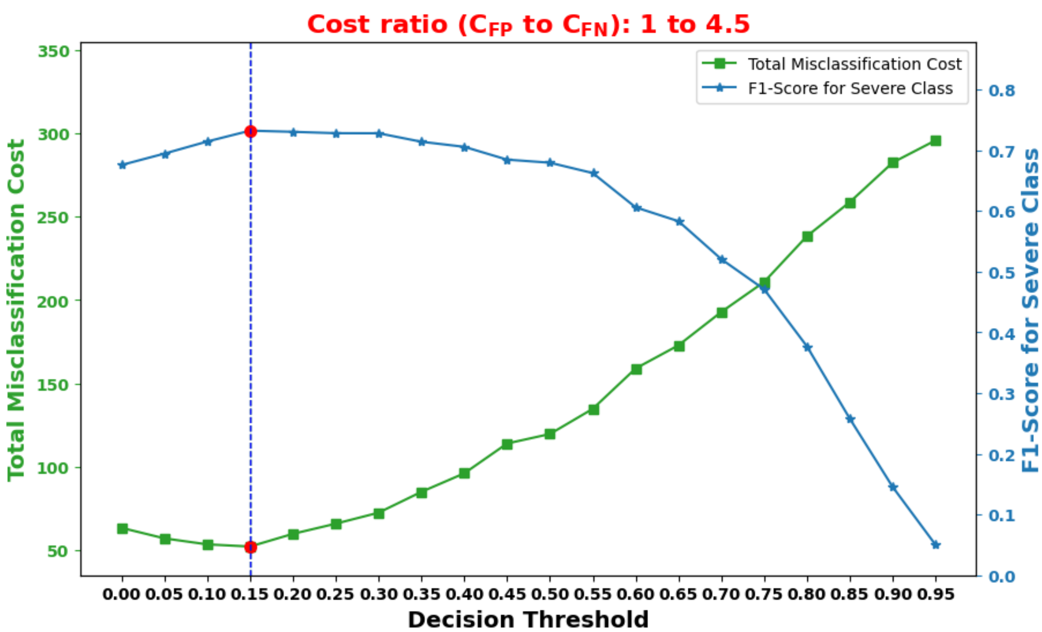

Cost ratio (C_{FP} to C_{FN}): 1 to 2
Total Misclassification Cost
F1-Score for Severe Class
Decision Threshold
Total Misclassification Cost
F1-Score for Severe Class
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95
40
60
80
100
120
140
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8


Cost ratio (C_{FP} to C_{FN}): 1 to 2.5
Total Misclassification Cost
F1-Score for Severe Class
Decision Threshold
Total Misclassification Cost
F1-Score for Severe Class
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95
40
60
80
100
120
140
160
180
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8


Cost ratio (C_{FP} to C_{FN}): 1 to 3
Total Misclassification Cost
F1-Score for Severe Class
Decision Threshold
225
200
175
150
125
100
75
50
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95


Cost ratio (C_{FP} to C_{FN}): 1 to 3.5
Total Misclassification Cost
F1-Score for Severe Class
Decision Threshold
250
200
150
100
50
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95


Cost ratio (C_{FP} to C_{FN}): 1 to 4
Total Misclassification Cost
F1-Score for Severe Class
Decision Threshold
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95


Cost ratio (C_{FP} to C_{FN}): 1 to 4.5
Total Misclassification Cost
F1-Score for Severe Class
Decision Threshold
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95

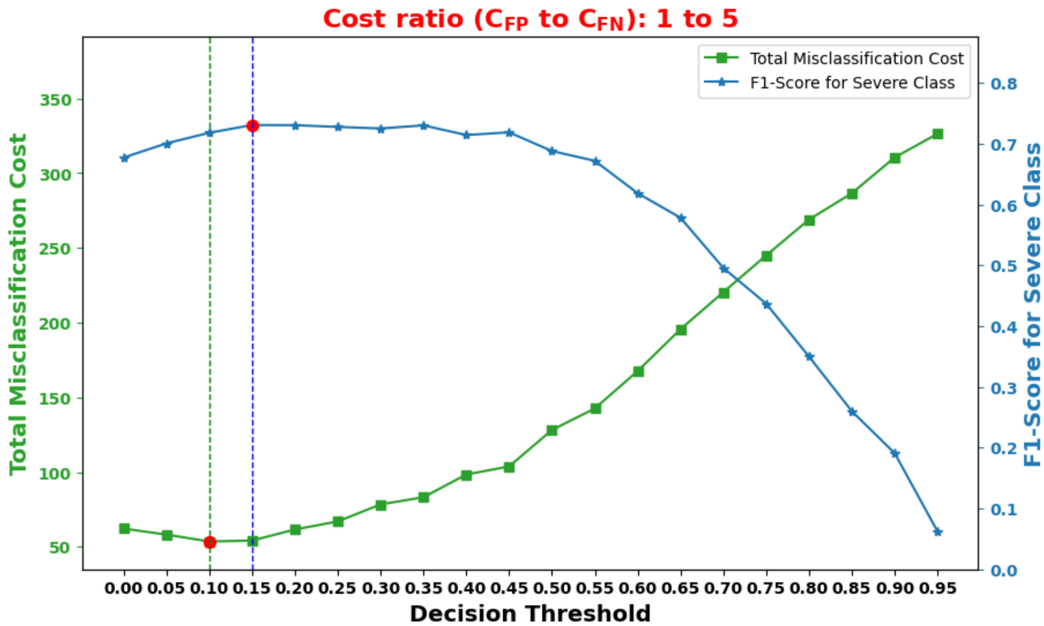


Figure 4.6: Threshold optimization based on misclassification cost and F1-score for the severe damage class

For a 1:1 cost ratio, the threshold value with minimum misclassification cost is 0.5. However, the F1-score for the severe damage class is maximized at the threshold of 0.45. As the cost ratio increases, the threshold that minimizes misclassification cost, and maximizes F1-score for the severe damage class, decreases, approaching zero. This adjustment improves the F1-score because when the cost of false negatives is higher, the model lowers the threshold to classify more buildings as severely damaged. This reduces the number of severe damage cases that are incorrectly classified as non-severe (false negatives), thereby minimizing the overall misclassification cost. The optimal thresholds for cost ratios of 1:1.5, 1:2, 1:2.5, and up to 1:5 are presented in Table 4.4. For the 1:5 cost ratio, a threshold of 0.1 minimizes the misclassification cost. The F1-score is close but not at its maximum with this threshold. This occurrence, observed for cost ratios of 1:1 and 1:5, is because the metrics for total misclassification cost and F1-score for the severe damage class prioritize different objectives, even though cost coefficients are incorporated into their calculations. Misclassification cost minimization focuses on reducing the overall weighted penalty for false positives and false negatives, directly influenced by the cost ratio. The F1-score, however, balances cost-weighted precision and recall, which may result in a slightly different threshold. For cost ratios exceeding 1:5 (i.e., cost of false negatives five times greater than that of false positives), the optimal threshold nears zero. This implies that almost all buildings should be subject to detailed inspection and seismic evaluation because the cost of false negatives is so significant.

To assess how the optimal decision threshold compares to the conventional threshold of 0.5, the modified ML-based RVS model was used on a random test set consisting of 132 buildings. 5-fold cross validation was performed, and the results were averaged across all five folds to ensure a comprehensive evaluation. The primary goal was to determine whether adjusting the threshold improves the performance of the ML model, particularly in accurately detecting buildings likely to be severely damaged. Table 4.4 presents the number of non-severe samples incorrectly classified as severe (false positives) and the number of severe samples incorrectly classified as non-severe (false negatives) for both the optimal thresholds and the conventional threshold of 0.5. When the optimal threshold is used to screen the buildings instead of 0.5, a notable reduction in the number of false negatives is observed, thereby improving the accuracy for the severe class. However, this improvement is accompanied by an increase in the number of false positives. For cost ratios of 1:1.5, 1:2, and 1:2.5, using the optimal threshold results in a reduction in false negatives that is approximately balanced by an increase in false positives. Conversely, for higher cost ratios with lower optimal thresholds, since the cost of misclassifying a severe case is equal to or greater than three times that of a non-severe case, adopting the optimal threshold over the conventional 0.5 is still considered an improvement. For example, with a cost ratio of 1:4, using the optimal threshold of 0.15 instead of 0.5 removes 18 out of 20 false negatives, correctly classifying these buildings as severe. However, this change also results in 26 more false positives. Given that the misclassification cost of each false negative is four times that of each false positive, this adjustment still enhances the classifier's performance. Notably, at a cost ratio of 1:5 with an optimal threshold of 0.1, nearly all false negatives are eliminated, significantly improving the classifier's accuracy in identifying severe cases.

Table 4.4: Comparison of false positives and false negatives at optimal thresholds versus the conventional Threshold of 0.5 for different cost ratios

Cost ratio	Conventional threshold of 0.5		Optimal threshold			Increase in number of false positives	Decrease in number of false negatives
	Number of false positives	Number of false negatives	Selected Threshold	Number of false positives	Number of false negatives		
1:1.5	22	20	0.40	30	11	8	9
1:2			0.35	33	9	11	11
1:2.5			0.3	36	7	14	13
1:3			0.25	40	4	18	16
1:3.5			0.2	45	3	23	17
1:4 & 1:4.5			0.15	48	2	26	18
1:5			0.1	52	1	30	19

By adopting this threshold optimization and cost sensitivity analysis approach, the model ensures that the decision-making process in seismic vulnerability assessments is both cost-effective and robust, particularly in minimizing the severe damage class misclassifications. This approach effectively manages the trade-offs between the costs of false positives and false negatives.

4.5 A Three-Grade ML-Based RVS Framework

The RVS approach previously proposed, which relies on basic building features, has demonstrated promise. However, this study introduced further refinements aimed at minimizing the model's misclassifications of high-risk buildings. This section presents a three-grade ML-based RVS framework that offers users the flexibility to choose between the original method and two enhanced alternatives, based on their specific needs and available resources for conducting an RVS. It is worth noting that while the proposed enhancements improve the model's reliability, the original ML model without them is still a robust method with 71% accuracy.

Grade 1: Initial Rapid Assessment

This grade involves the application of the RVS proposed in the previous chapter. Utilizing an RF classifier, users can rapidly identify low-rise RC buildings at high risk of severe damage. This step is recommended when resources are limited, aiming for a quick assessment due to budget and time constraints. This initial assessment provides a reasonable accuracy of approximately 71%, allowing for a rapid evaluation.

Grade 2: Enhanced Assessment with Model Uncertainty Detection

For users with sufficient budget and time, and who seek a more accurate assessment, this grade includes performing the model uncertainty detection proposed in this study. By analyzing the probability distribution of predictions, it becomes possible to identify the level of uncertainty: low, moderate, or high. Buildings with moderate or high uncertainty should be categorized for further actions, such as detailed inspections or additional analyses, with a priority given to those with high uncertainty. The decision to investigate both groups or only those with high uncertainty should be based on the available resources, time, and regional conditions.

Grade 3: Threshold Optimization for Misclassification Cost Reduction

This grade incorporates decision threshold optimization of the classifier to account for the cost of misclassifications, making it the most conservative of the three grades. Instead of using the conventional threshold of 0.5, an optimal threshold is determined to minimize the cost of false predictions by the model while maintaining high accuracy, especially for the severe damage class. Real costs or the relative cost of false positives ($\mathrm{C_{FP}}$) and false negatives ($\mathrm{C_{FN}}$) can be used to assign values to misclassifications. These coefficients are applied as weights to the samples in the non-severe and severe damage classes, guiding the model to prioritize samples from the class with a higher associated cost. The primary goal is to minimize the misclassification of the severe class, necessitating sufficient resources for a larger number of buildings detected as probable to be severely damaged. This comprehensive approach ensures a more cautious and thorough evaluation, aimed at significantly reducing the risk of overlooking severely damaged buildings. The enhanced screening approach is illustrated in Figure 4.7, highlighting the step-by-step process for each grade.

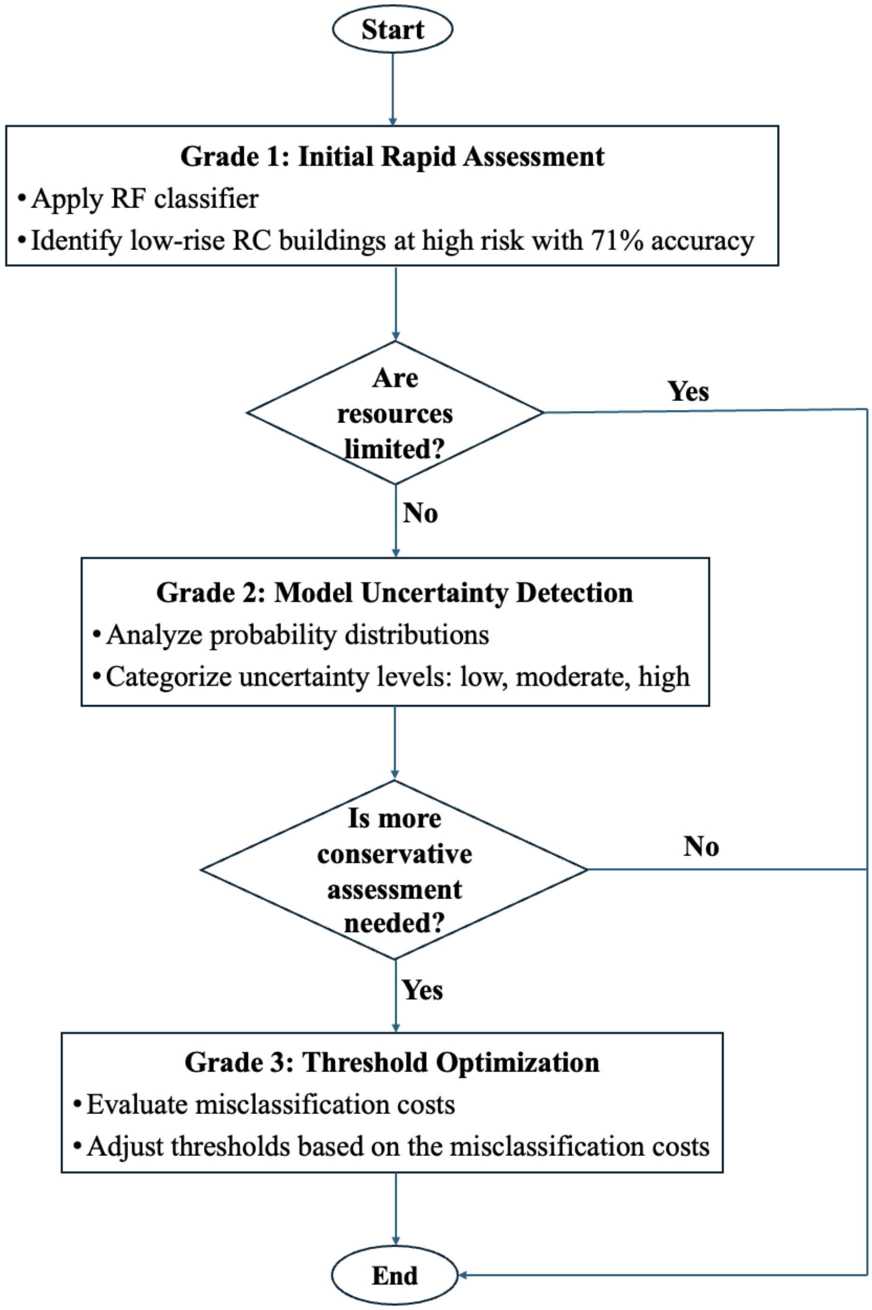


Figure 4.7: Flowchart of three-grade ML-based RVS framework

4.6 Conclusions

This study presented a comprehensive approach to RVS through a three-grade ML-based methodology, emphasizing the importance of model uncertainty detection and threshold optimization. The initial ML-based seismic screening model developed in Chapter 3 can identify low-rise reinforced concrete buildings likely to be severely damaged in earthquakes with 71% accuracy, but two enhancements were investigated in this work to provide a more nuanced understanding of prediction confidence and improve its performance. The first enhancement involved calculating probabilities for each damage class from every tree in the forest, rather than solely predicting a single class label. This allowed for visualizing the probability distribution for a building belonging to each damage class, offering clearer insights into the model's predictions. Buildings were then categorized into low, moderate, or high uncertainty groups based on these distributions. This categorization helps prioritize further detailed investigations for buildings with moderate to high uncertainty. This approach is essential in ensuring that the potential model uncertainty is addressed, particularly for structures at greater risk, thereby enhancing overall safety. Finally, the threshold optimization approach refined the decision-making process by considering the relative costs associated with misclassifications. This strategy was vital for reducing the misclassification of buildings at high risk of severe damage as low risk, as such errors could result in substantial financial and safety implications. By evaluating misclassification costs across various decision thresholds for the classifier, this methodology improved the overall reliability of seismic vulnerability assessments. The enhanced three-grade ML-based RVS methodology offers a structured and effective framework for accurately identifying buildings at risk of severe damage. By systematically addressing model uncertainty and optimizing decision thresholds, this framework helps improve prediction outcomes while being mindful of the associated costs, ultimately contributing to safer built environments and informed decision-making for stakeholders involved in seismic risk management.

4.7 Limitations of the Study

While this study addressed model uncertainty and misclassification costs in the proposed ML-based RVS method in Chapter 3, determining appropriate cost ratios for false positives and false negatives presents notable challenges. These ratios are inherently context-dependent, varying significantly based on regional economic conditions, building characteristics, seismic hazard levels, and stakeholder priorities. For example, in high-risk seismic zones, the cost of a false negative (misclassifying a severely vulnerable building as low risk) can extend beyond structural repair or rebuilding to include loss of life, economic disruptions, and social impacts. Conversely, in regions with lower seismic risks, the cost of a false positive (misclassifying a low risk building as high risk) can result in unnecessary expenditures for detailed inspections, temporary relocations, or retrofits. The inherently subjective nature of assigning monetary values to these costs complicates the development of universally applicable cost ratios. Effective calibration of cost ratios must consider local needs, available resources, and stakeholder goals, making it more challenging to apply the methodology across different contexts.

Chapter 5

A Data-Driven Methodology to Determine the Backbone Curve of RC Columns

This chapter corresponds to the following conference paper and a journal paper currently under preparation:

N. Elyasi and E. Kim, "A machine learning-based methodology to determine the backbone curve of RC columns," in *Proc. 18th World Conf. Earthquake Eng.*, Milan, Italy, 2024.

Abstract

RC columns are critical structural components, and their seismic performance during earthquakes plays a pivotal role in overall structural resilience. The backbone curve of an RC column, usually derived from its hysteresis behavior under cyclic loading, provides essential insights into its load-displacement relationship. This curve is crucial for the design and seismic assessment of RC columns. However, existing conventional methods for determining backbone curves, such as nonlinear analyses or laboratory tests, are often time-intensive and resource-demanding. In this study, a data-driven methodology is proposed to efficiently determine the backbone curves of RC columns. The methodology employs CatBoost regression models to predict the key points on the backbone curve (yield, peak, and ultimate points), using a dataset of 451 rectangular RC columns sourced from various references. Eight geometric and material properties are utilized as input features for the models. Initially, six independent CatBoost models are developed to predict the key points. To improve prediction accuracy, model chains are employed, using sequential learning to account for potential interdependencies among the key points. The proposed CatBoost-based regression chains demonstrate significant improvements in accuracy over independent models, achieving high R^2 and reduced error metrics (RMSE and RelRMSE). Moreover, the methodology is compared to the existing models from the literature, and the results show that the CatBoost chains outperform their models, particularly in terms of accuracy and efficiency. The findings of this study demonstrate the practicality and effectiveness of the proposed ML-based methodology for determining backbone curves of RC columns. Despite the existing conventional methods, the proposed methodology can readily determine the complete backbone curve of RC columns based on their dimensional and structural features and without the need for performing experimental tests or simulations.

5.1 Introduction

In structural engineering, the backbone curve represents the load-displacement relationship of a structure, illustrating its resistance to deformation under various loading conditions. For RC columns, which play a pivotal role in a structure's load-bearing capacity, the backbone curve provides essential performance information. This information is invaluable for design and assessment, making the determination of this curve a key task in structural engineering. Conventional methods for deriving backbone curves include analytical and experimental approaches such as those outlined in FEMA 356 [7], ASCE 41 [8], and the idealized model by Elwood and Moehle (2006) [56]. However, these methods often require nonlinear analysis or laboratory tests, which can be both time-consuming and expensive. Consequently, there is a growing need for methods that can reliably determine backbone curves without the need for complex analyses or experimental tests, thereby saving time and resources. ML offers a promising alternative for efficiently determining backbone curves. Recent studies, including those by Deger and Taskin (2022) [59], Yu et al. (2024) [61], and Luo and Paal (2018) [58], have explored ML-based approaches for identifying the backbone curves of RC columns. These studies are reviewed in detail in Chapter 2. While these methodologies have shown promise, there remains a need for a simpler and more easily applicable method to determine the backbone curve, especially for identifying the key points on this curve.

This study proposes an ML-based methodology to determine the backbone curve of RC columns, focusing on identifying three key points on the curve: yield, peak, and ultimate. The methodology utilizes CatBoost, a gradient-boosting regression algorithm, to predict these key points. A dataset comprising experimental data from 451 rectangular RC columns, sourced from various references, is used to train and test the ML models. Ten geometric and structural parameters are selected as input features for the models. Initially, six independent and separate CatBoost models are developed to predict the six output variables corresponding to the key points. While these independent models demonstrate reasonable performance, further enhancement is pursued by employing model chains to capture potential dependencies among the output variables. Model chains, a sequential learning technique, utilize the predictions of one model as additional inputs for subsequent models in the sequence. The results highlight the significant improvements in prediction accuracy achieved by model chains compared to independent models. The developed CatBoost chains are also compared to the models proposed by Luo and Paal [58], and the results demonstrate that the proposed chains performed better overall, proving their practicality and effectiveness. In summary, the outcomes of this study demonstrate the simplicity and efficiency of the proposed ML-based approach for establishing the backbone curve of RC columns. The proposed methodology eliminates the need for resource-intensive experimental testing or complex analyses, making it a practical and effective tool for structural analysis and design.

5.2 Dataset

The dataset utilized in this study comprises a total of 451 rectangular RC columns from various sources. Specifically, it includes 131 columns collected by Fei and Lu [87] from Chinese literature, 43 columns gathered by Hart et al. [88], and 277 columns compiled by Ghannoum et al. [89], [90]. This dataset encompasses general information, geometric and material properties, reinforcement details, and backbone curves obtained from laboratorial testing. Additionally, it includes information regarding the key points on these curves.

The following parameters of RC columns are selected from the dataset as the input features for the purpose of this study:

- Axial load ratio ($P / A_g.f_c^{'}$)
- Shear span to effective depth ratio (L / d)
- Cross-sectional area (A_g)
- Number of longitudinal reinforcing bars (N_l)
- Number of transverse reinforcing bars in the cross section (N_t)
- Axial compressive strength of concrete ($f_c^{'}$)
- Yield strength of longitudinal reinforcement (f_{yl})
- Yield strength of transverse reinforcement (f_{yt})

A detailed presentation of these features, including their respective statistical ranges, is provided in Table 5.1. Figure 5.1 illustrates the data distribution for each of these features, providing insights into their variability and statistical spread.

Table 5.1: Input features and their statistical ranges

Feature	Minimum	Maximum	Mean	Standard deviation
$P/A_g.f_c'$	0.24	0.8	0.5	0.16
L / d	1.36	8.9	4	1.49
A_g (mm^2)	23226	490000	124110	85213
N_l	2	32	6.42	5.49
N_t	2	6	3.02	0.93
f_c' (MPa)	13.1	118	43	25.1
f_{yl} (MPa)	324.1	667	449.4	71.7
f_{yt} (MPa)	235	1423.6	466.2	180

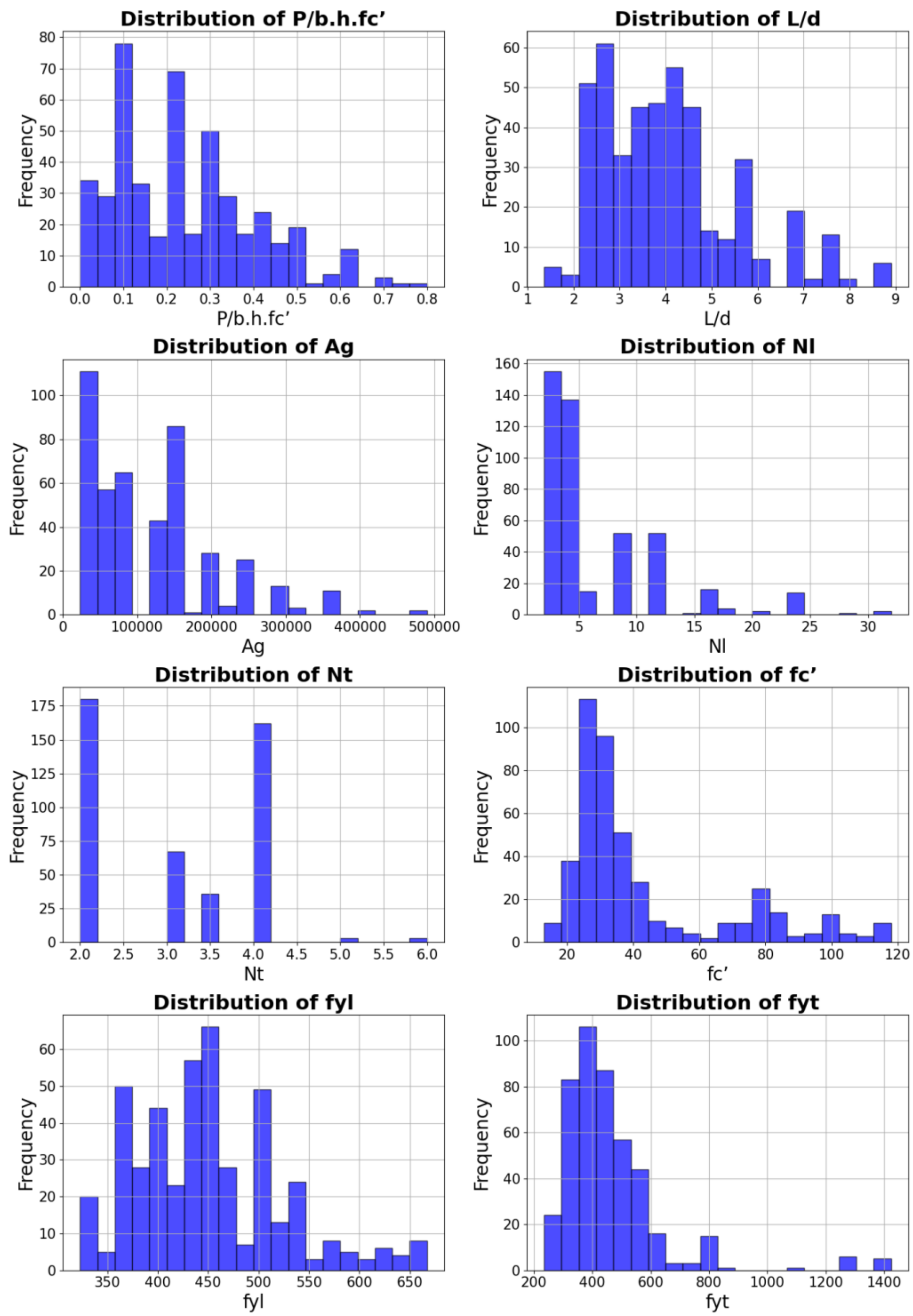


Figure 5.1: Distribution of the input features listed in Table 5.1

The dataset includes three key points of the samples' backbone curves: yield, peak, and ultimate, along with the complete experimental backbone curves for other samples. To extract these key points from the samples containing complete backbone curves, the following strategies employed by Ghannoum et al. [89], [90] are utilized:

- **Peak point:** the peak load (V_p) is determined as the maximum load on the curve. The corresponding displacement is then designated as the peak displacement (Δ_p).
- **Yield point:** the yield displacement (Δ_y) is determined by the intersection of a secant stiffness line passing through 0.7 of V_p and the horizontal line that passes through V_p. The load corresponding to Δ_y on the curve is selected as the yield load (V_y).
- **Ultimate point:** the ultimate load (V_u) is defined as 0.8 of V_p. The corresponding displacement is then designated as the ultimate displacement (Δ_u).

Figure 5.2 displays a sample backbone curve from the dataset, highlighting the strategies employed to identify the key points. Six key variables (Δ_y, Δ_p, Δ_u, V_y, V_p, and V_u) are extracted for each sample in the database. These variables serve as the output parameters for constructing the proposed backbone curve in this study. A detailed summary of these variables, along with their statistical ranges, is provided in Table 5.2. Figure 5.3 also shows the data distribution for each of these output variables, providing insights into their variability and statistical spread. The distributions indicate a higher concentration of smaller magnitude values for these output variables.

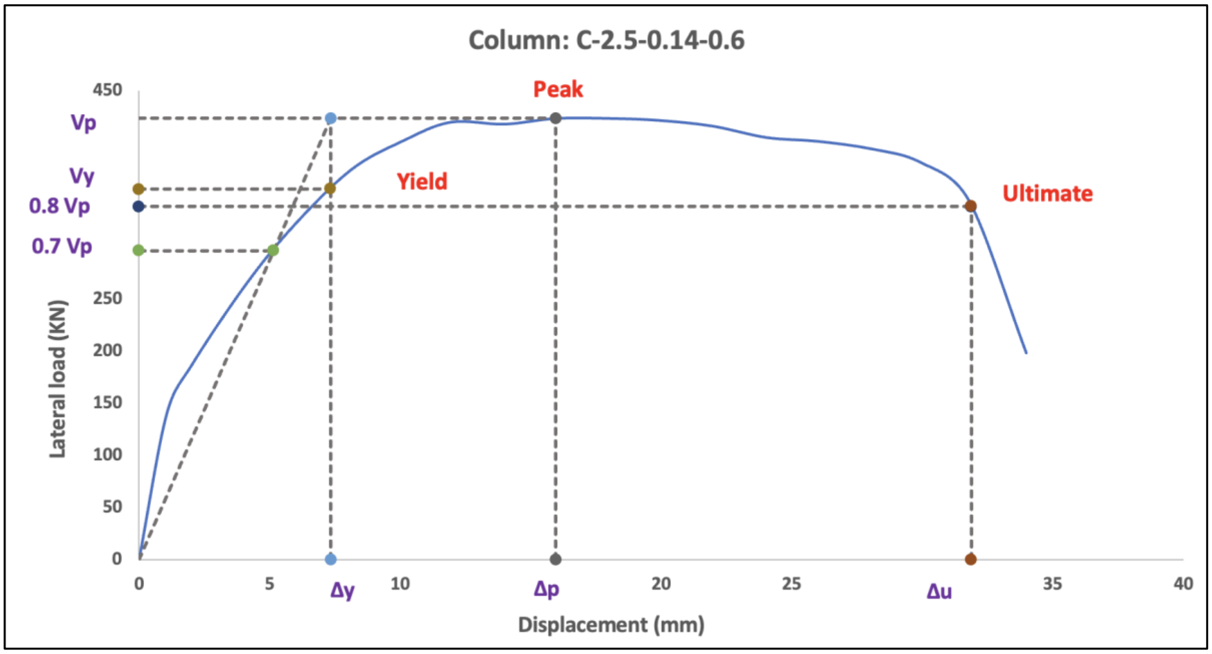


Figure 5.2: Experimental Backbone curve incorporating three key points for a column sample

Table 5.2: Output variables and their statistical ranges

Variable	Minimum	Maximum	Mean	Standard deviation
Δ_y (mm)	0.48	51.1	6.2	7.5
Δ_p (mm)	0.92	174	16	21.5
Δ_u (mm)	1.6	245.7	31.7	37.6
V_y (KN)	26	1356.7	215.8	181.7
V_p (KN)	36.9	1339	261	215.9
V_u (KN)	8.1	1138	200	179.1

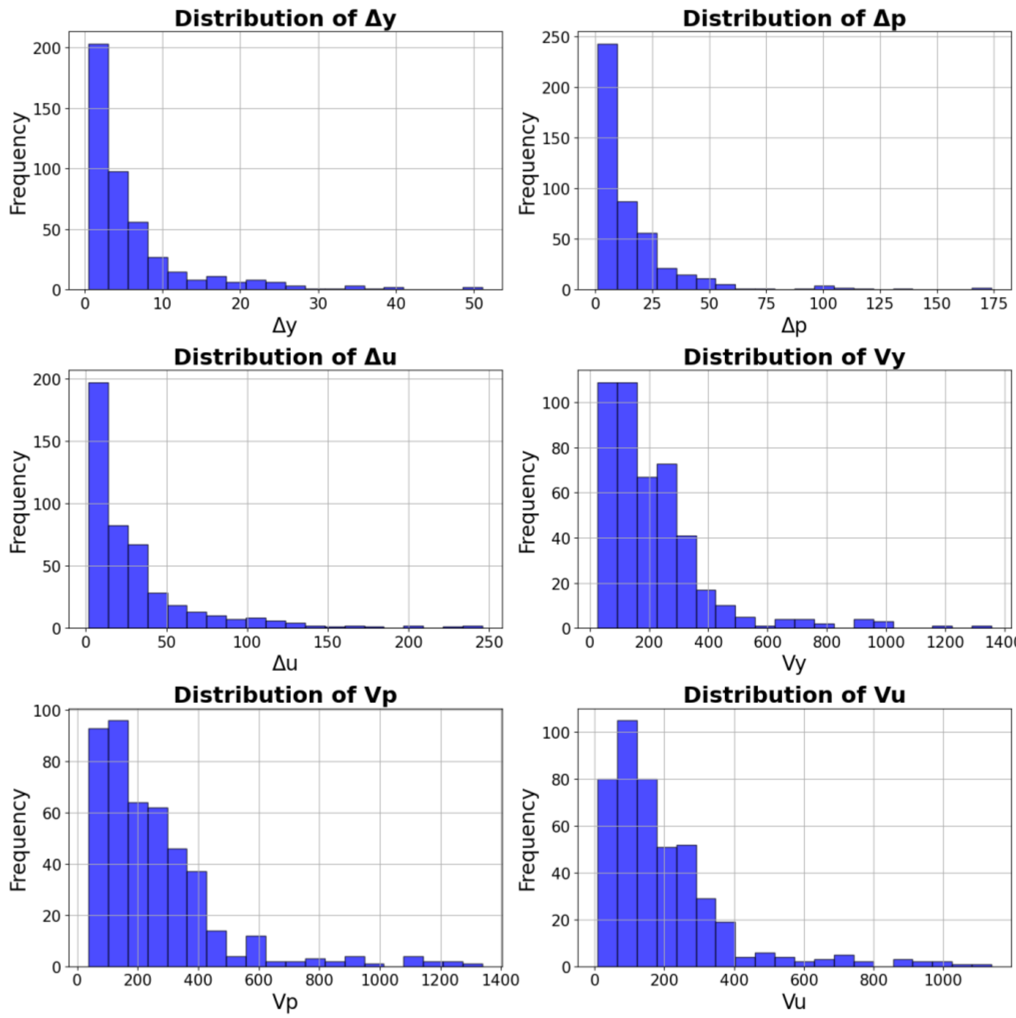


Figure 5.3: Distribution of the output variables listed in Table 5.2

5.3 Proposed Method

In this study, the emphasis is on identifying three key points on the backbone curve: yield, peak, and ultimate. These points are essential as they provide significant insights into the column's behavior, while the rest of the curve is less significant. To simplify the representation, straight lines are used to connect these key points, similar to the approach by Luo and Paal [58] for the pre-peak segment of the backbone curve. This representation also includes the post-peak segment, enhancing the backbone curve model proposed by Luo and Paal. Figure 5.4 illustrates the proposed backbone curve model for the sample presented in Figure 5.2, demonstrating the application of this approach.

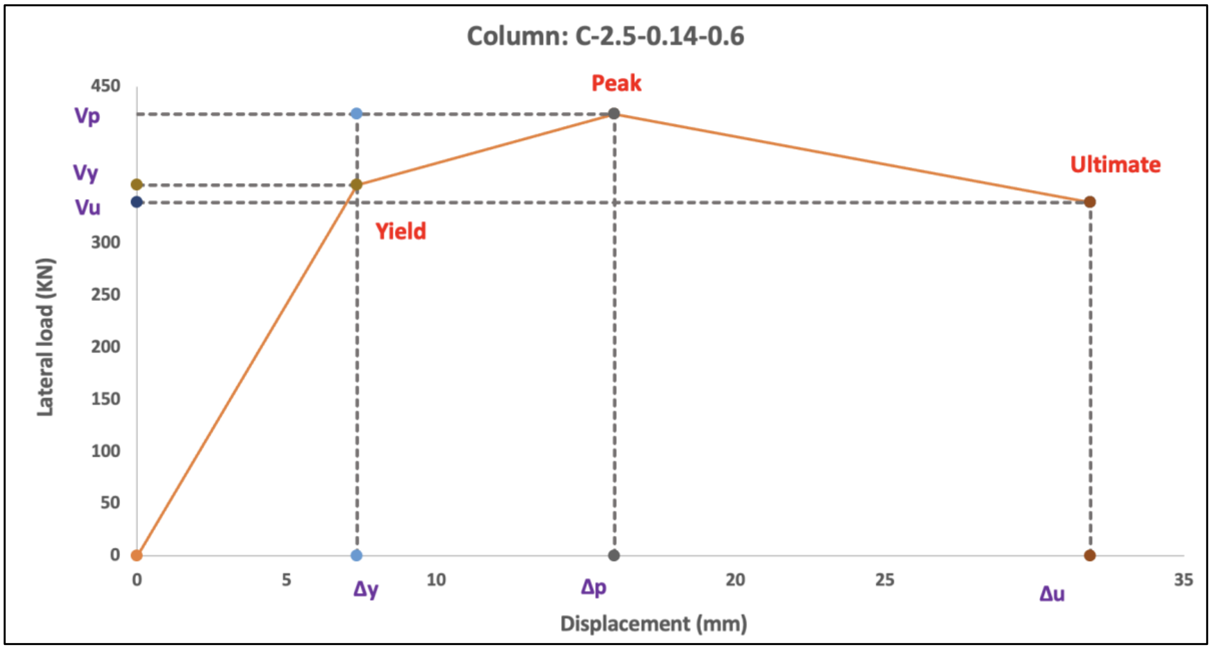


Figure 5.4: proposed backbone curve incorporating three key points for a column sample

To estimate the three key points of the proposed backbone curve, a CatBoost regression algorithm is used to predict the six output variables, as detailed in Table 5.2. Models are trained and evaluated using the input features listed in Table 5.1. CatBoost is selected due to its several advantages, including high model accuracy with minimal need for parameter tuning, fast execution, scalability, effective handling of categorical features, robustness against overfitting, and strong performance even with small datasets. The dataset is divided into an 80% training set and a 20% validation set. A 5-fold cross validation technique is employed using the training set to identify the best hyperparameters for tuning the models. Initially, six independent models are developed to predict the six output variables: Δ_y, Δ_p, Δ_u, V_y, V_p, and V_u. These six models operate under the assumption that all output variables are entirely independent, not capturing any potential interdependencies. However, it is known that dependencies may exist, particularly among V_y, V_p, and V_u., as V_p played a role in determining V_y and Vu from the experimental backbone curves. To account for these possible dependencies, two model chains are then developed: one for Δ_y, Δ_p, Δ_u, and another for V_y, V_p, V_u. A model chain is a sequential learning technique wherein the outputs of one model serve as additional inputs for the subsequent model in the chain. Each model chain in this study comprises three regression models. Each model within a chain is trained to predict one of the three output variables in its respective category (Displacement: Δ_y, Δ_p, and Δ_u, or Load: V_y, V_p, and V_u), utilizing both the input features listed in Table 5.1 and the predicted values from previous models in the sequence. The performance of the model chains is compared with that of the six independent models to assess any potential improvements in prediction.

In regression problems, the evaluation of ML models is a critical step, accomplished by utilizing coefficient and error metrics such as coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and relative root mean squared error (RelRMSE). In this study, the metrics selected for model evaluation are R^2, RMSE, and RelRMSE as these metrics are easy to interpret, sensitive to large errors, and useful for model comparison.

R^2 measures the proportion of the variance observed in the target variable explained by the model. This metric is employed to gauge the goodness of fit, with a scale from 0 to 1, with 1 indicating a perfect fit [91]. The formal definition of R^2 is as follows:

$$R^2 = 1 - \frac{\sum(\text{actual value} - \text{predicted value})^2}{\sum(\text{actual value} - \text{mean ofactual values})^2} \qquad \text{(Equation 5.1)}$$

RMSE is a metric that quantifies the average magnitude of the errors between predicted and actual values. A notable advantage of RMSE is its unit consistency with the target variable, facilitating its interpretability [92]. The formal definition of RMSE is outlined as follows.

$$\text{RMSE} = \sqrt{\frac{\sum(\text{actual value} - \text{predicted value})^2}{\text{number of data points}}} \qquad \text{(Equation 5.2)}$$

RelRMSE is a variation of RMSE that is normalized by the magnitude of the actual values. This normalization allows for the metric to be expressed as a proportion, making it easier to compare different models with varying scales. The formal definition of RelRMSE is as follows.

$$\text{RelRMSE} = \frac{\text{RMSE}}{\text{Mean of actual values}} \qquad \text{(Equation 5.3)}$$

5.4 Results and Discussion

5.4.1 Independent Regression Models

Six independent and separate CatBoost models are trained to predict Δ_y, Δ_p, Δ_u, V_y, V_p, and V_u using the parameters listed in Table 5.1, including axial load ratio ($P / A_g.f_c^{'}$), shear span to effective depth ratio (L / d), cross-sectional area (A_g), number of longitudinal reinforcing bars (N_l), number of transverse reinforcing bars in the cross section (N_t), axial compressive strength of concrete ($f_c^{'}$), yield strength of longitudinal reinforcement (f_{yl}), and yield strength of transverse reinforcement (f_{yt}). The results from evaluating these six models are presented in Table 5.3.

Table 5.3: Performance of independent CatBoost models

Independent CatBoost Models				
Model Number	**Output Variable**	**Evaluation Metrics**		
		R^2	RMSE	RelRMSE
1	**Δ_y (mm)**	0.92	2.08	0.34
2	**Δ_p (mm)**	0.91	6.07	0.39
3	**Δ_u (mm)**	0.92	10.15	0.33
4	**V_y (KN)**	0.94	44.15	0.21
5	**V_p (KN)**	0.96	41.77	0.16
6	**V_u (KN)**	0.91	45.63	0.24

As can be seen, the models demonstrated high R^2 values, ranging from 0.91 to 0.96, indicating good fit to the data. The model for V_p achieved the highest R^2 value of 0.96. In terms of RMSE, the models showed varied values as RMSE gives values in unit of the output variable. When considering RelRMSE, which normalizes the RMSE by the mean of the observed values, the models for V_y and V_p performed particularly well with RelRMSE values of 0.21 and 0.16, respectively. This suggests that the prediction errors for these models are relatively low compared to the mean values of the observed data. The models for Δ_y, Δ_p, Δ_u, and V_u had RelRMSE values of 0.34, 0.39, 0.33, and 0.24, respectively, indicating relatively high prediction errors. While the models overall demonstrate fair performance, the high prediction errors in certain cases underscore the need for further refinement to enhance accuracy and reliability.

5.4.2 Regression Model Chains

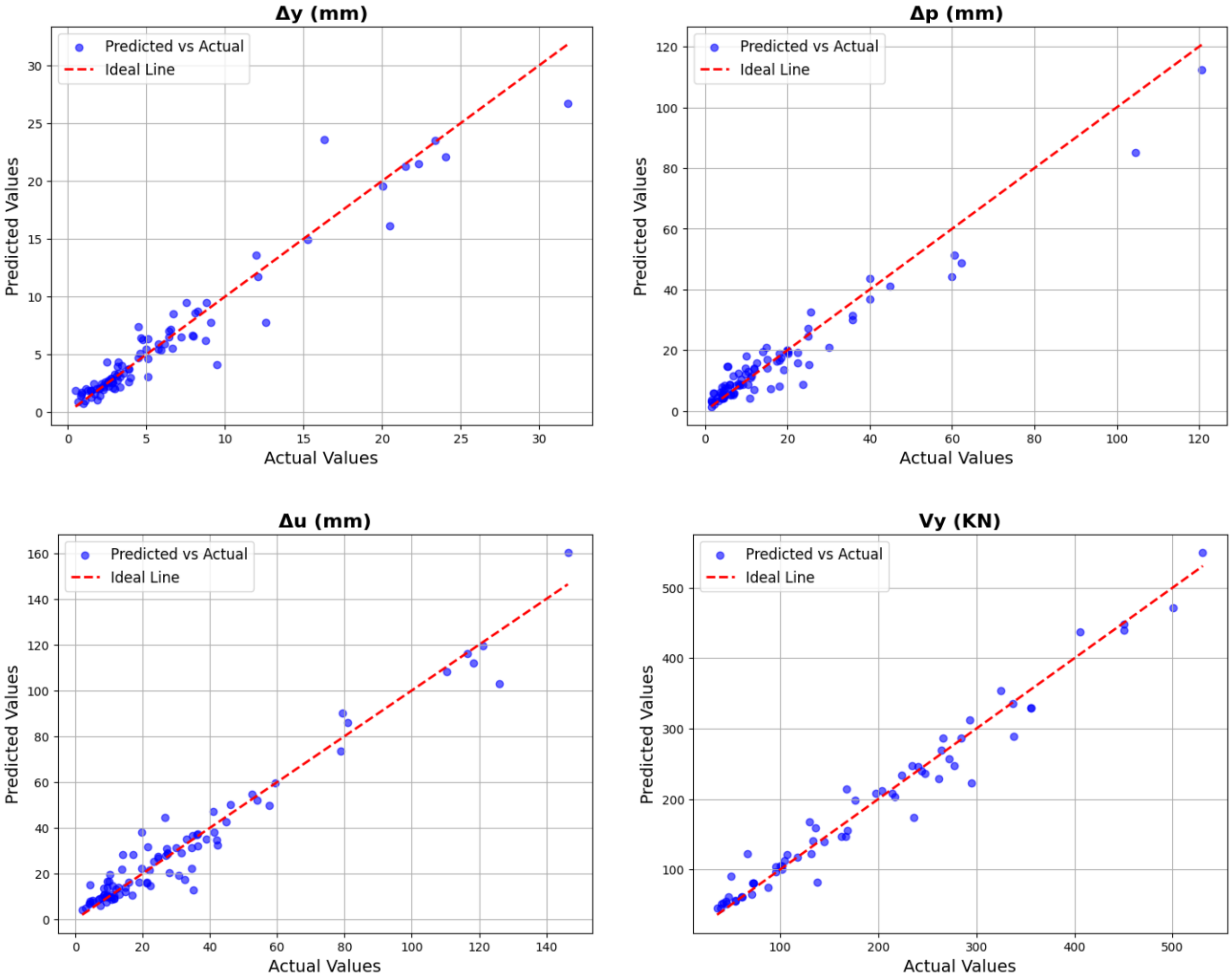
To address the relatively high prediction errors observed in the independent CatBoost models, particularly for Δ_y, Δ_p, Δ_u, and V_u, two model chains are developed for the displacement and load variables. Each chain employs a sequential training approach, using both the input features listed in Table 5.1 and the predicted outputs from earlier models in the chain. For instance, a chain is constructed for the displacement variables (Δ_y, Δ_p, and Δ_u), consisting of three sequential models. The first model is trained to predict Δ_y using only the input features from Table 5.1. The second model, tasked with predicting Δ_p, incorporates the input features alongside the predicted values of Δ_y from the first model. Similarly, the third model is trained to predict Δ_u, using the input features as well as the predicted values of Δ_y and Δ_p from the previous models. A similar approach is adopted for the load chain. Table 5.4 summarizes the performance of these chains, and the percentage improvement achieved compared to the independent regression models outlined in Table 5.3.

Table 5.4: Performance and improvements of CatBoost chains over independent models

CatBoost Chains					**Improvement (%)**		
Chain Category	**Output Variable**	**Evaluation Metrics**			**Evaluation Metrics**		
		R^2	RMSE	RelRMSE	R^2	RMSE	RelRMSE
Displacement	**Δ_y (mm)**	0.93	1.55	0.27	1.1	25.5	20.6
	Δ_p (mm)	0.93	5.09	0.32	2.2	16.1	17.9
	Δ_u (mm)	0.95	6.79	0.23	3.3	33.1	30.3
Load	**V_y (KN)**	0.96	23.22	0.13	2.1	47.4	38.1
	V_p (KN)	0.98	22.08	0.10	2.1	47.1	37.5
	V_u (KN)	0.98	20.11	0.12	7.7	55.9	50

The results demonstrate the significant advantages of sequential learning in improving prediction accuracy. While the independent models achieved acceptable R^2 values ranging from 0.91 to 0.96, the model chains showed even better performance, with R^2 values reaching 0.95 for Δ_u and 0.98 for V_y and V_u. Notable reductions in RMSE were also observed, particularly for the load variables: RMSE for V_y decreased from 44.15 to 23.22 (47.4% improvement), for V_p from 41.77 to 22.08 (47.1% improvement), and for V_u from 45.63 to 20.11 (55.9% improvement). Similarly, RelRMSE values were improved notably, with V_y decreasing from 0.21 to 0.13, V_p from 0.16 to 0.10, and V_u from 0.24 to 0.12. Among displacement variables, Δ_u showed the most significant gains, with a 33.1% and 30.3% reduction in RMSE and RelRMSE respectively, highlighting its enhanced predictability. However, Δ_y and Δ_p showed moderate improvement.

Figure 5.5 illustrates the predicted versus actual values for all output variables on the test set. The plots reveal a reasonable agreement between predictions and actual values, particularly for smaller magnitudes. This is primarily because the dataset contains a higher concentration of smaller magnitude values, especially for the displacement variables. The abundance of data in this range allows the models to perform more accurately. However, predictions for larger magnitudes, where data are relatively sparse, remain more challenging.



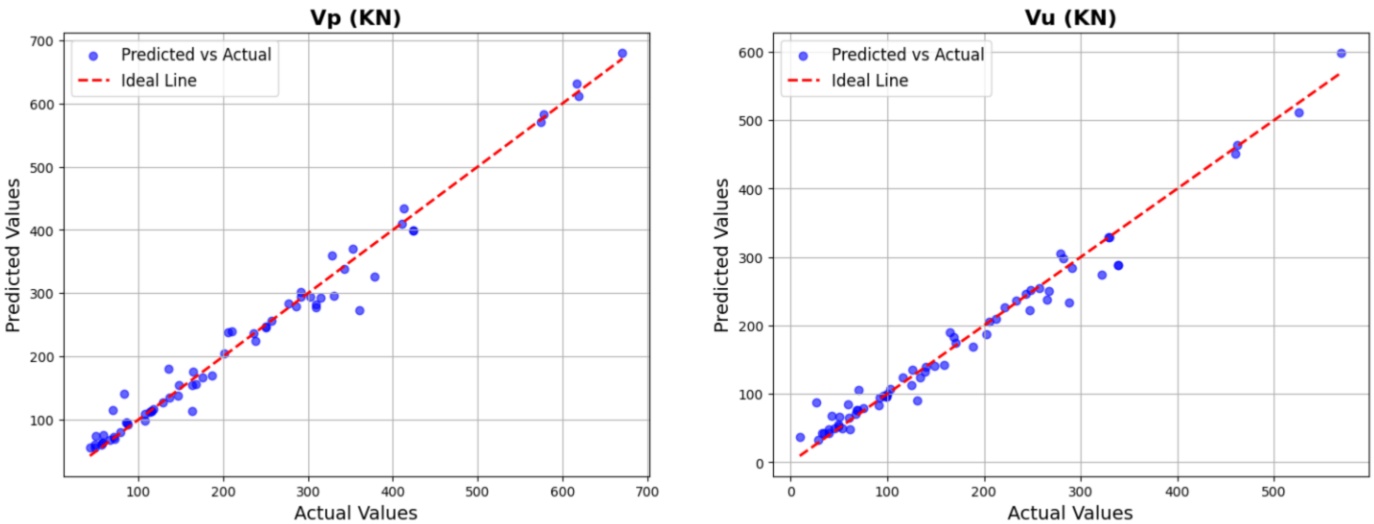
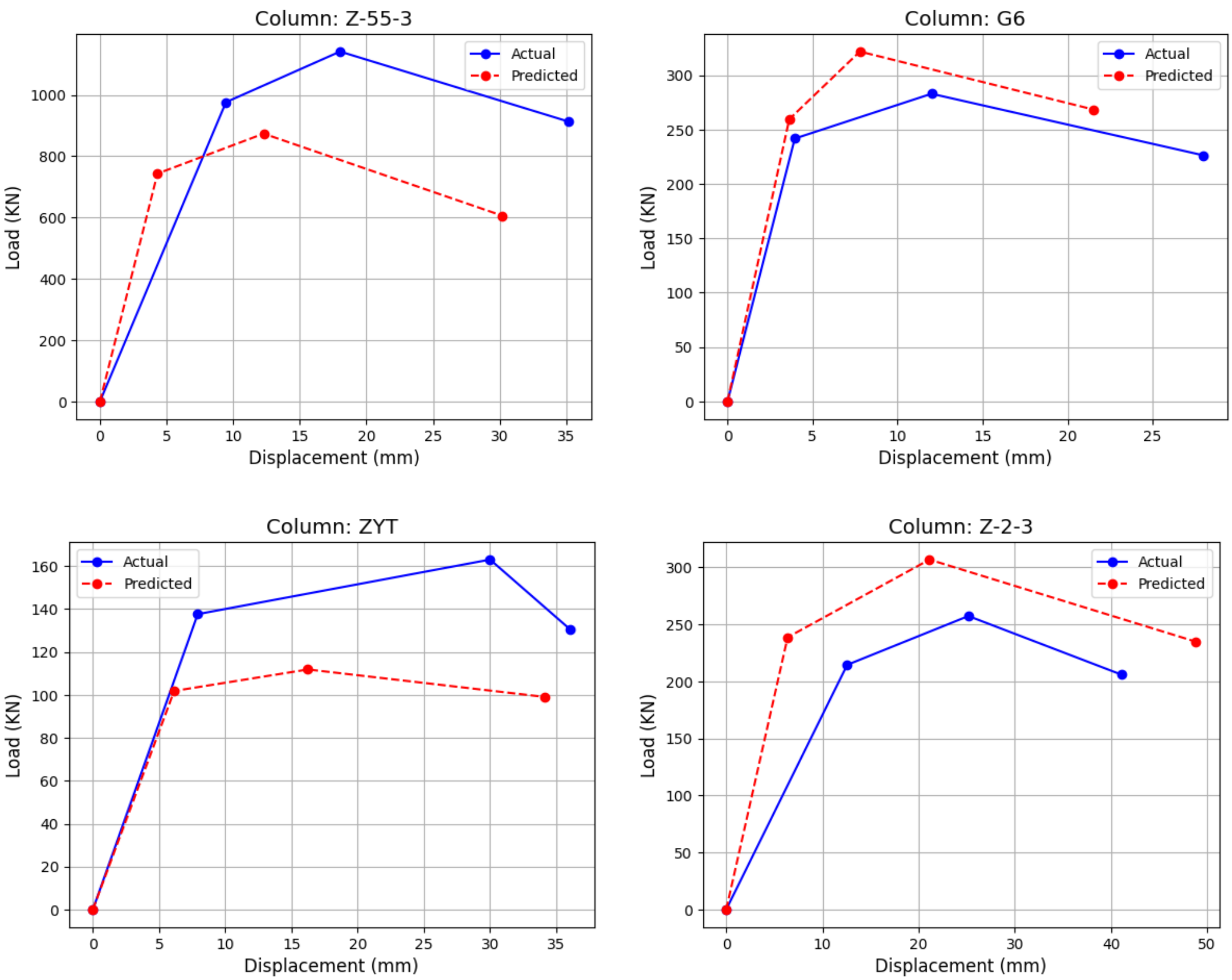


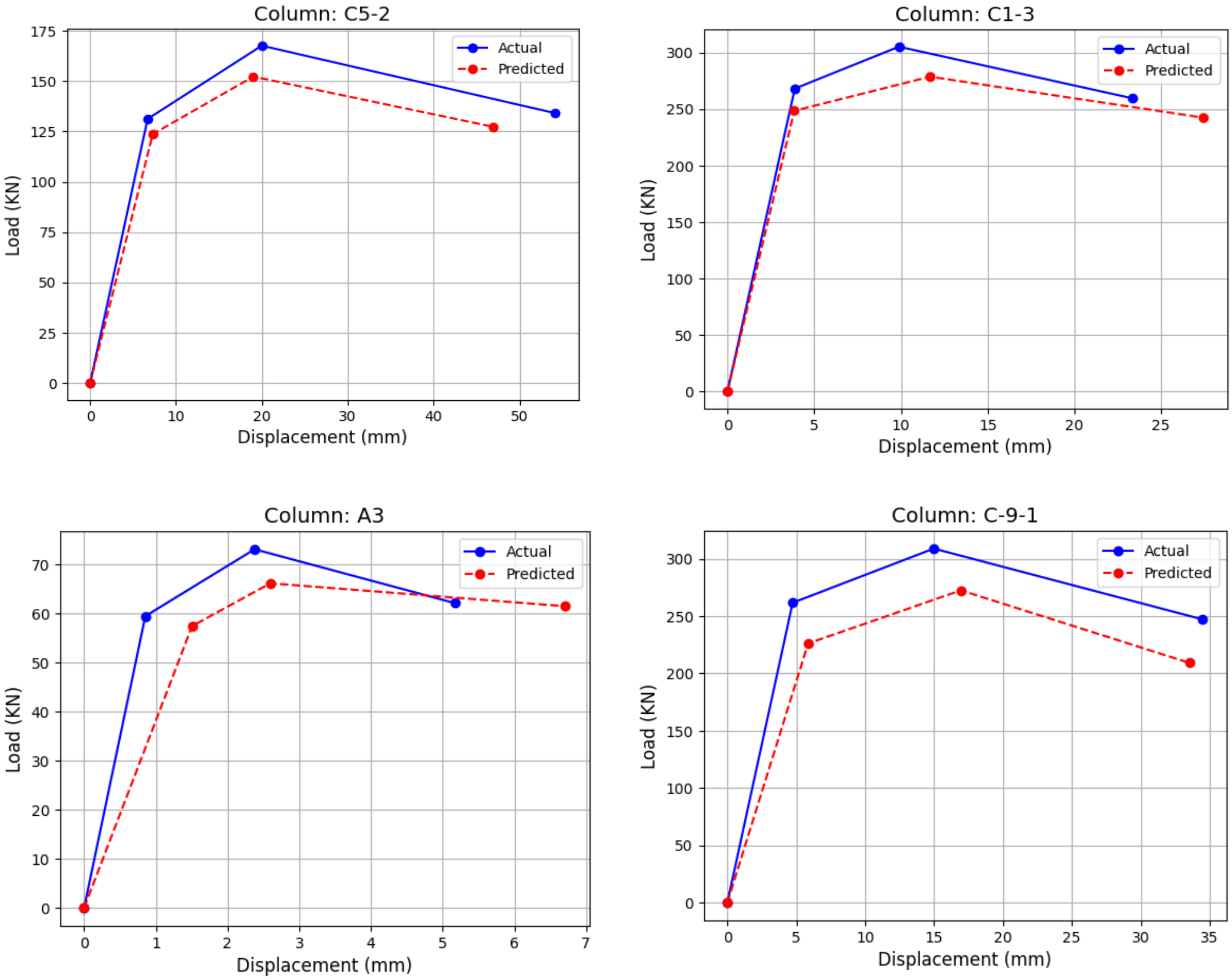
Figure 5.5: Predicted vs. actual values for all output variables on the test set

Overall, the developed CatBoost chains delivered strong performance for predicting load variables. This is due to the interdependence of load variables, which the sequential learning approach effectively captured. For the displacement variables, the sequential model chain produced fair predictions but did not achieve similar improvements. This can be explained by the weaker relationships among displacement variables. The higher concentration of smaller magnitude values for displacement variables, as shown in Figure 5.5, further influenced the chain's predictive accuracy. The proposed CatBoost chains offer a promising method for rapidly estimating the backbone curves of RC columns, relying solely on simple column parameters. However, their performance could be further improved with access to larger, more comprehensive, and better-balanced datasets. Future studies are encouraged to explore these improvements to optimize the predictive capabilities of the proposed models.

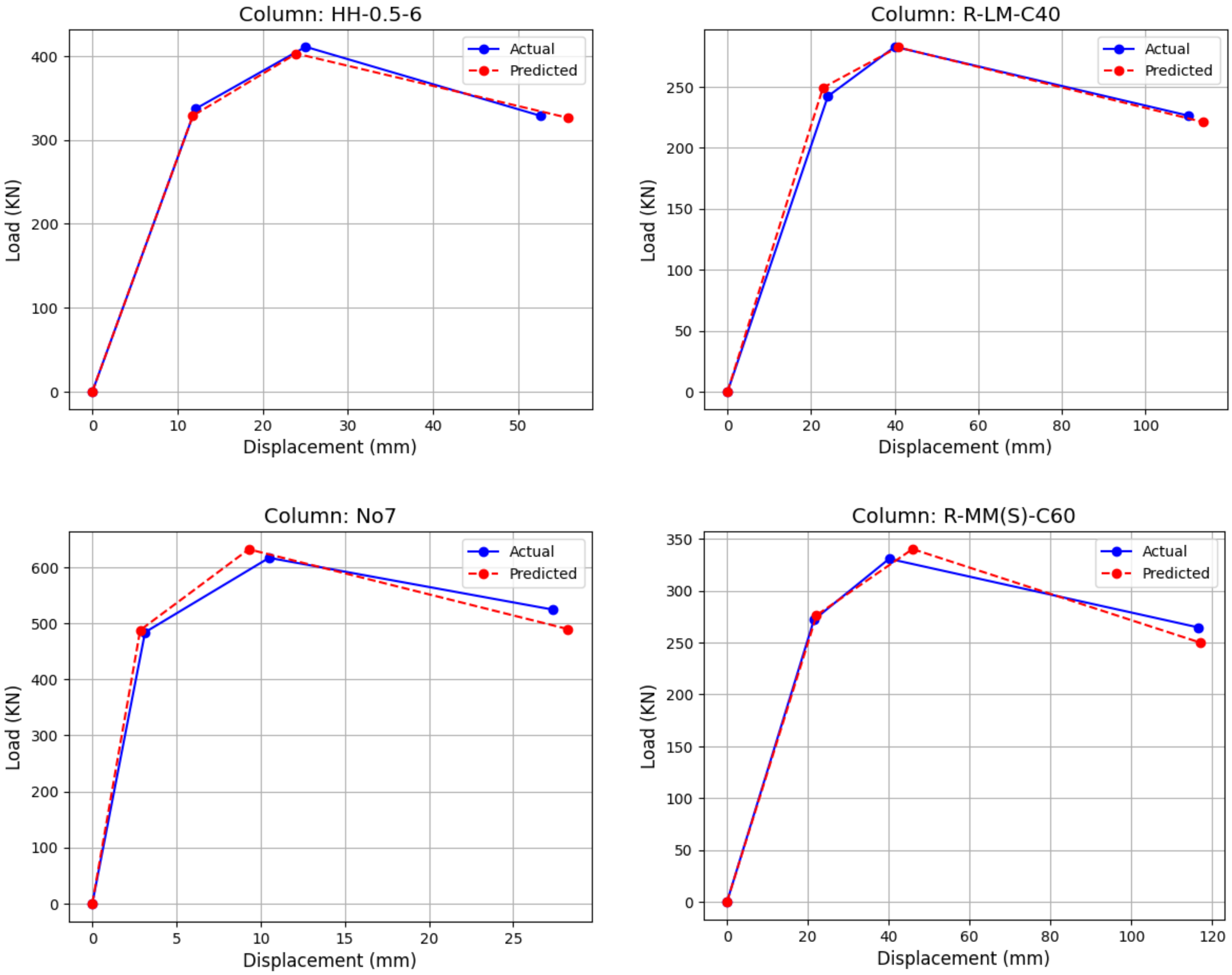
To more illustrate the performance of the developed CatBoost chains, Figure 5.6 compares the actual and predicted trilinear backbone curves for 12 column samples from the test set, categorized by different levels of prediction accuracy. Part (a) highlights four samples where the chains exhibited weak predictive performance, with significant differences between the actual and predicted curves. Part (b) shows intermediate performance, where the predicted curves align moderately well with the actual ones. Part (c) demonstrates good performance, as the predicted curves closely overlap with the actual ones, indicating high predictive accuracy. These observations reveal a clear trend: the chains perform well for some samples, while their predictions are less accurate for others. This discrepancy can be attributed to factors such as the quality of the experimental data, the small size of the dataset leading to overfitting, and the absence of important features like failure mode in the dataset. These results underscore the potential, limitations, and scope for enhancement of the proposed approach.



(a)


Column: C5-2
Load (KN)
Displacement (mm)
Actual
Predicted
Column: C1-3
Load (KN)
Displacement (mm)
Actual
Predicted
Column: A3
Load (KN)
Displacement (mm)
Actual
Predicted
Column: C-9-1
Load (KN)
Displacement (mm)
Actual
Predicted

(b)



(c)

Figure 5.6: Comparison of actual and predicted trilinear backbone curves for 12 column samples with (a) weak, (b) moderate, and (c) good prediction performances

5.4.3 Comparison with Luo and Paal (2018)

The performance of the developed CatBoost-based regression chains in this study is compared to the models proposed by Luo and Paal [58]. Since their work focuses solely on the pre-peak section of the backbone curve, the comparison is limited to this section, and the post-peak results (i.e., Δ_u and V_u) are not included. It is important to highlight that the proposed models in this study utilize eight input features, while Luo and Paal's models use 15 input features, including parameters like column's failure mode, which were not available in the dataset used in this study. Consequently, implementing their method on the same dataset for a direct comparison was not feasible. Instead, the comparison relies on the results reported in their paper. Another challenge in the comparison arises from Luo and Paal's use of drift ratios (displacement divided by column length) instead of absolute displacements for the x-axis of the backbone curve, which is a common approach. To ensure comparability, the displacement-based results from this study were converted into drift ratios for the evaluation. Although the dataset used in this study differs from that of Luo and Paal, their statistical ranges are fairly similar to provide a reasonable basis for comparison. The comparison results are summarized in Table 5.5.

Table 5.5: Comparison of performance between the proposed CatBoost Chains and Luo and Paal's models [58]

Models	**Proposed by Luo and Paal**		**Proposed CatBoost Chains**	
	R^2	RMSE	R^2	RMSE
δ_y (%)	0.91	0.10	0.93	0.12
δ_p (%)	0.88	0.5	0.93	0.38
V_y (KN)	0.96	27.48	0.96	23.22
V_p (KN)	0.97	26.84	0.98	22.08

The results indicate that the proposed CatBoost Chains generally outperform the models developed by Luo and Paal, particularly in terms of lower RMSE values for δ_p, V_y, and V_p. Although the RMSE for the yield drift ratio (δ_y) is slightly higher in this study, the difference is marginal and does not significantly impact the overall performance. In addition to improved accuracy, the CatBoost Chains demonstrate greater versatility by predicting the post-peak section of the backbone curve, which was not addressed in the study by Luo and Paal. Furthermore, these models achieve superior performance using only eight input features, compared to the 15 features required by Luo and Paal's models. This highlights the efficiency and practical applicability of the proposed approach.

5.5 Conclusions

This study introduced an efficient ML-based methodology for determining the backbone curves of rectangular RC columns. Using a dataset of 451 RC columns, CatBoost regression models were developed to predict six variables associated with three key points on the backbone curve: the yield point (Δ_y, V_y), the peak point (Δ_p, V_p), and the ultimate point (Δ_u, V_u). While these initial models demonstrated reasonable performance, prediction accuracy was further enhanced by implementing model chains. This sequential learning approach incorporated predictions from earlier models as inputs for subsequent ones, effectively capturing interdependencies among the parameters. The proposed CatBoost chains achieved notable improvements over independent regression models, particularly for load-related variables. For example, R^2 values reached 0.98 for both V_p and V_u, while RMSE values were significantly reduced from 41.77 to 22.08 KN for V_p and from 45.63 to 20.11 KN for V_u, representing improvements of 57.1% and 55.9%, respectively. Displacement variables (Δ_y, Δ_p, and Δ_u) also exhibited gains, with Δ_u achieving an R^2 value of 0.95 and a 33.1% reduction in RMSE. Compared to the models proposed by Luo and Paal [58], the CatBoost chains delivered superior accuracy while requiring fewer input features. Additionally, unlike their study, the proposed methodology predicts both pre- and post-peak behavior, providing a more comprehensive understanding of column performance. By eliminating the need for resource-intensive experimental tests and analyses, this study offered a practical, efficient, and cost-effective alternative for determining backbone curves based on readily available geometric and material properties.

5.6 Limitations of the Study

This study, while demonstrating the potential of an ML-based approach for determining backbone curves of RC columns, had several limitations that should be noted. First, the dataset used, although extensive for this type of study, was relatively small (451 samples) and primarily focused on rectangular RC columns within a specific range of geometric and material properties. This limited scope restricts the generalizability of the findings to other column shapes, such as circular or irregular sections, or to columns with parameters beyond the statistical ranges observed in the dataset. Additionally, the dataset contained a higher concentration of smaller magnitude values for displacement variables, potentially affecting the models' accuracy in predicting larger magnitudes. Expanding the dataset to include a broader and more balanced representation of column types and performance characteristics would enhance the robustness and applicability of the proposed methodology. Another limitation lied in the exclusion of certain potentially influential input features, such as failure modes or other design-specific parameters, which were not available in the dataset. These features could improve the accuracy of the models. Furthermore, while the CatBoost chains demonstrated significant improvements in accuracy for load-related variables, the gains for displacement variables were less significant, highlighting the potential need for alternative approaches. Future studies should focus on addressing these limitations by incorporating larger, more diverse datasets, exploring alternative or complementary modeling techniques, and including a wider range of input features to refine the methodology and expand its practical applicability.

Chapter 6
Conclusions and Contributions

6.1 Thesis Summary

In the field of structural engineering, ensuring the stability of structures against seismic events is a critical challenge. Earthquakes are among the most destructive natural disasters, often resulting in significant human and economic losses. Structural failures during earthquakes underscore the need for efficient and accurate assessment methodologies to evaluate the seismic performance of buildings and their components. Traditional methods, while reliable, are often time-intensive and resource-demanding. In contrast, the growing availability of datasets and advancements in ML provide an opportunity to develop innovative data-driven approaches that address these limitations.

This thesis aimed to harness the potential of available datasets to propose efficient and accurate data-driven methodologies for seismic vulnerability assessment and backbone curve determination of RC structures. By utilizing data and ML techniques, the research sought to provide tools that are both practical and scalable, addressing critical challenges in structural engineering. Two primary objectives guided this work: the development of a data-driven methodology for assessing the seismic vulnerability of low-rise RC buildings at the system level and the creation of an efficient framework for determining backbone curves of RC columns at the component level. The following sections summarize how these objectives were addressed in this thesis.

6.1.1 Objective 1: Utilize extensive data of low-rise RC buildings to develop a data-driven methodology for assessing their seismic vulnerability at the system level

Seismic vulnerability assessments are essential for identifying at-risk buildings and prioritizing retrofitting or replacement efforts. Existing methods often rely on simplified relationships between structural characteristics and vulnerability. While these methods are straightforward, they often lack adaptability, making them less effective for large-scale applications or regions with varying seismic conditions. To overcome these challenges, this research proposed a novel ML-based RVS methodology.

In Chapter 3, ML classification models, including logistic regression and RF, were developed using data from six post-earthquake surveys. These models utilized simple geometric parameters including the number of stories, floor area, column area, and wall area, to predict building vulnerability. By directly employing raw input data rather than calculated indices like the PI proposed by Hassan and Sozen (1997) [55], the models eliminated the need for region-specific calibration and achieved significantly improved prediction accuracy. Additionally, the inclusion of seismic intensity measures, such as MMI and PGA, allowed for a more comprehensive evaluation of vulnerability, accommodating regional variations in seismic conditions.

Chapter 4 further refined this methodology by addressing prediction uncertainty and optimizing decision thresholds. A three-grade ML-based RVS framework was introduced, categorizing buildings based on their vulnerability levels. The first enhancement ensured that buildings with high uncertainty in predictions were prioritized for detailed inspections, improving the reliability and practicality of the screening process. The optimized decision thresholds reduced misclassification risks, particularly for highly vulnerable buildings, and provided a decision-making tool for stakeholders.

By utilizing ML techniques, this research developed an adaptable and reliable RVS methodology for low-rise RC buildings. The proposed framework not only improved prediction outcomes but also offered a practical solution for large-scale seismic screening, enhancing structural safety.

6.1.2 Objective 2: Develop a data-driven methodology for determining the backbone curves of RC columns at the component level

RC columns are critical components in structural systems, and their seismic performance is pivotal to the overall stability of buildings during earthquakes. Backbone curves, which represent the relationship between force and displacement, are essential for understanding the seismic behavior of columns. Conventional methods for deriving backbone curves, such as nonlinear analyses and experimental tests, are often resource-intensive and time-consuming for large-scale applications. This research addressed these limitations by proposing a data-driven approach using ML to determine backbone curves of RC columns.

Chapter 5 introduced CatBoost regression models trained on a dataset of 451 RC columns to predict six key variables of the backbone curve: yield displacement (Δ_y), peak displacement (Δ_p), ultimate displacement (Δ_u), yield load (V_y), peak load (V_p), and ultimate load (V_u). The initial models demonstrated reasonable accuracy, using 10 geometric and material properties as input features. To further improve predictive performance, regression chains were employed to capture interdependencies among the backbone curve key variables. Compared to existing methodologies, such as those proposed by Luo and Paal (2018) [58], the CatBoost regression chains achieved superior performance. Additionally, unlike their models, the proposed framework predicted post-peak behaviors as well, providing a more comprehensive understanding of column performance under seismic loads. These advancements made the methodology a practical, efficient, and cost-effective alternative to the conventional methods.

In summary, this thesis successfully achieved its objectives by introducing data-driven methodologies for seismic vulnerability assessment and backbone curve determination. By addressing system-level challenges through an enhanced RVS framework and component-level challenges through efficient backbone curve determination, the thesis provided tools that advance the field of structural engineering.

6.2 Thesis Limitations

The limitations of the methodologies proposed in Chapters 3, 4, and 5 were discussed in detail within their respective chapters. A summary of these limitations is provided here to highlight areas that require further attention and refinement.

In Chapter 3, the ML-based seismic vulnerability assessment methodology relied on datasets from six post-earthquake surveys, with variability in quality due to differences in factors such as data collection tools, inspection protocols, and geographical contexts. To simplify the classification, moderate and severe damage classes were grouped into a single category, which reduced model complexity but limited the ability to differentiate between these damage classes. The models did not take into account critical factors like construction quality, building age, and maintenance, focusing solely on structural parameters and seismic intensity due to data constraints. Misclassifications posed significant concerns, as classifying high-risk buildings as low-risk can compromise safety, while the opposite leads to unnecessary costs. prediction uncertainty underscored the need for further refinement to improve confidence and reliability.

In Chapter 4, the enhanced RVS methodology incorporated model uncertainty detection and misclassification cost optimization. However, determining appropriate cost ratios for false positives and false negatives proved challenging due to their strong dependency on regional contexts, including economic conditions, building characteristics, and seismic hazard levels. The subjective nature of assigning monetary values to these costs made it difficult to develop universally applicable cost ratios. Effective calibration requires local considerations, limiting the methodology's broader applicability of the methodology.

In Chapter 5, the ML-based approach for determining backbone curves of RC columns used a relatively small dataset of 451 rectangular columns, limiting its generalizability to other shapes or parameter ranges outside the dataset’s statistical range. Data imbalances, with a higher concentration of smaller magnitude values for displacement variables, affected accuracy for larger magnitudes. Additionally, the exclusion of potentially influential features like failure modes or design-specific parameters reduced the predictive capabilities of the models. While the proposed CatBoost chains significantly improved the accuracy of load-related variables, the gains for displacement variables were less significant, highlighting the need for further enhancement or complementary approaches.

These limitations highlight challenges related to data constraints, regional variability, and the refinement of modeling techniques. Addressing these issues in future research can improve the accuracy and applicability of the proposed methodologies. The next section provides recommendations to help overcome these limitations and guide further advancements.

6.3 Recommendations for Future Work

Considering the limitations highlighted in this thesis, several pathways are suggested to improve and extend the proposed methodologies. These recommendations aim to enhance the accuracy, reliability, and broader application of the methods.

Expanding and improving datasets: The datasets used in this thesis, while valuable, were limited in size and scope. Future research should focus on expanding these datasets by incorporating new data as it becomes available. Data collection is an ongoing process in structural engineering, especially with the occurrence of major earthquakes and the subsequent post-earthquake surveys. Modern data collection tools, such as drones and high-resolution imaging, offer opportunities to gather more comprehensive and higher-quality data. These tools can capture additional structural details, such as material degradation or construction quality, that were not available in older datasets. Extending the datasets can improve the performance and reliability of the proposed ML models in this thesis. Moreover, with larger datasets, advanced algorithms such as deep learning, which perform better with extensive data, can be explored. Additionally, increasing the data available on damage states in RVS could allow moderate damage to be classified as a separate category, rather than grouping it with severe damage.

Incorporating additional features: Future research should aim to add more detailed features to improve the training and evaluation of the ML models, as additional data becomes available. Features like failure modes, construction quality, material aging, and environmental conditions are important for better understanding how structures behave under seismic forces. Adding these features could improve the accuracy and reliability of the proposed models for seismic vulnerability assessment and backbone curve determination. Expanding the range of input features can make the models more practical and applicable to a variety of real-world situations.

Exploring other building and structural types: This thesis primarily focused on low-rise RC buildings and rectangular RC columns. Future research could expand this focus to include a wider range of structures, such as high-rise buildings, steel buildings, and masonry structures. Similarly, exploring different column geometries, including circular or irregular sections, could improve the applicability of the proposed methodologies. Such extensions would necessitate incorporating new datasets and features specific to these additional types. These efforts could provide deeper insights into structural behaviors and broaden the practical use of the developed methodologies in structural engineering contexts.

Incorporating regional variations: Future work can focus on the development of an adaptable framework for cost ratio determination that considers the diverse and region-specific factors influencing seismic vulnerability assessments. This framework must integrate regional economic conditions, building characteristics, seismic hazard levels, and stakeholder priorities to ensure practical and tailored applicability. For instance, high-risk zones with densely populated areas may prioritize minimizing false negatives to prevent catastrophic losses, whereas low-risk regions may focus on reducing false positives to avoid unnecessary expenses. To enhance adaptability, methodologies that incorporate dynamic cost ratio adjustments based on real-time data should be explored.

Enhancing backbone curve determination methodology: The proposed method for determining backbone curves showed limitations, particularly in accurately predicting displacement variables. Future research should aim to refine these models to improve their accuracy and minimize prediction errors. This could include exploring alternative modeling techniques to better capture the relationships among the variables. Additionally, performing detailed simulations of RC columns, paired with comparisons to experimental data, could provide valuable insights into the quality of the datasets used for training and evaluating the ML models. This comparison could identify inconsistencies in the data, further guiding model refinement. Expanding these efforts could lead to more robust and reliable methods, making the backbone curve determination process more practical and broadly applicable across various structural contexts.

6.4 Publications and Presentations

6.4.1 Journal Papers

- N. Elyasi, E. Kim, and C. M. Yeum, "A machine-learning-based seismic vulnerability assessment approach for low-rise RC buildings," *Journal of Earthquake Engineering*, vol. 28, no. 3, pp. 760–776, 2024.
- N. Elyasi and E. Kim, "An enhanced machine learning-based rapid visual screening framework considering model uncertainty and decision threshold optimization," 2024. (Submitted, Under review)

6.4.2 Conference Papers and Presentations

- N. Elyasi, M. Midwinter, E. Kim, and C. M. Yeum, "Rapid seismic vulnerability assessment of low-rise RC buildings using machine learning," in *Proc. 12th Nat. Conf. Earthquake Eng.*, Salt Lake City, UT, USA, 2022.
- N. Elyasi, W. Carofilis, E. Kim, and C. M. Yeum, "Numerical validation of a seismic rapid visual screening method," in *Proc. Canadian-Pacific Conf. Earthquake Eng.*, Vancouver, BC, Canada, 2023.
- N. Elyasi, I. Crant, E. Kim, and M. Pandey, "A machine learning-based tool for assessing the condition of wood utility poles," in *Proc. CSCE Annu. Conf.*, Niagara Falls, ON, Canada, 2024.
- N. Elyasi and E. Kim, "A machine learning-based methodology to determine the backbone curve of RC columns," in *Proc. 18th World Conf. Earthquake Eng.*, Milan, Italy, 2024.

References

[1] H.-T. Thai, “Machine learning for structural engineering: A state-of-the-art review,” in *Structures*, Elsevier, 2022, pp. 448–491.

[2] R. C. Deo, “Machine learning in medicine,” *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[3] M. Berno *et al.*, “A machine learning-based approach for advanced monitoring of automated equipment for the entertainment industry,” in *2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0&IoT)*, IEEE, 2021, pp. 386–391.

[4] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” *Sensors*, vol. 18, no. 8, p. 2674, 2018.

[5] International Federation of Red Cross and Red Crescent Societies (IFRC), “A report on emergency plan of action operation for 2016 Ecuador earthquake,” 2018.

[6] M. M. Kassem, F. M. Nazri, and E. N. Farsangi, “The seismic vulnerability assessment methodologies: A state-of-the-art review,” *Ain Shams Engineering Journal*, vol. 11, no. 4, pp. 849–864, 2020.

[7] F. E. FEMA 356, “Prestandard and commentary for the seismic rehabilitation of buildings,” *Federal Emergency Management Agency: Washington, DC, USA*, 2000.

[8] ASCE, *ASCE/SEI, 41-17, seismic evaluation and retrofit of existing buildings.* 2017.

[9] N. Elyasi, E. Kim, and C. M. Yeum, “A machine-learning-based seismic vulnerability assessment approach for low-rise RC buildings,” *Journal of Earthquake Engineering*, vol. 28, no. 3, pp. 760–776, 2024.

[10] D. G. Kleinbaum and M. Klein, “Analysis of matched data using logistic regression,” *Logistic regression: A self-learning text*, pp. 227–265, 2002.

[11] Y. Xie, M. Ebad Sichani, J. E. Padgett, and R. DesRoches, “The promise of implementing machine learning in earthquake engineering: A state-of-the-art review,” *Earthquake Spectra*, vol. 36, no. 4, pp. 1769–1801, 2020.

[12] N. Mitra, S. Mitra, and L. N. Lowes, “Probabilistic model for failure initiation of reinforced concrete interior beam–column connections subjected to seismic loading,” *Eng Struct*, vol. 33, no. 1, pp. 154–162, 2011.

[13] S. Mangalathu and J.-S. Jeon, “Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques,” *Eng Struct*, vol. 160, pp. 85–94, 2018.

[14] V. R. Kohestani, M. Hassanlourad, and A. Ardakani, “Evaluation of liquefaction potential based on CPT data using random forest,” *Natural Hazards*, vol. 79, no. 2, pp. 1079–1089, 2015.

[15] S. Mangalathu and J. Jeon, “Stripe-based fragility analysis of multispan concrete bridge classes using machine learning techniques,” *Earthq Eng Struct Dyn*, vol. 48, no. 11, pp. 1238–1255, 2019.

[16] “https://catboost.ai.”

[17] M. Chen, S. Mangalathu, and J.-S. Jeon, “Machine learning–based seismic reliability assessment of bridge networks,” *Journal of Structural Engineering*, vol. 148, no. 7, p. 06022002, 2022.

[18] T. M. Hossain, M. Hermana, and J. O. Olutoki, “A Novel Stochastic CatBoost Based Shear Wave Velocity Prediction and Uncertainty Analysis in Sandstone Reservoir using Multi-Seismic Attributes,” *IEEE Access*, 2024.

[19] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] C. M. Bishop, “Neural networks and their applications,” *Review of scientific instruments*, vol. 65, no. 6, pp. 1803–1832, 1994.

[21] A. Derakhshani and A. H. Foruzan, “Predicting the principal strong ground motion parameters: A deep learning approach,” *Appl Soft Comput*, vol. 80, pp. 192–201, 2019.

[22] X. Lu, Y. Xu, Y. Tian, B. Cetiner, and E. Taciroglu, “A deep learning approach to rapid regional post-event seismic damage assessment using time-frequency distributions of ground motions,” *Earthq Eng Struct Dyn*, vol. 50, no. 6, pp. 1612–1627, 2021.

[23] Q. Yu *et al.*, “Rapid visual screening of soft-story buildings from street view images using deep learning classification,” *Earthquake Engineering and Engineering Vibration*, vol. 19, pp. 827–838, 2020.

[24] S. Mangalathu, H. Sun, C. C. Nweke, Z. Yi, and H. V Burton, “Classifying earthquake damage to buildings using machine learning,” *Earthquake Spectra*, vol. 36, no. 1, pp. 183–208, 2020.

[25] C. Rojahn, *Rapid visual screening of buildings for potential seismic hazards: a handbook*, vol. 21. Federal Emergency Management Agency, 1988.

[26] F. E. M. A. (US), *Rapid visual screening of buildings for potential seismic hazards: a handbook*. Government Printing Office, 2017.

[27] N. Bektaş and O. Kegyes-Brassai, “Conventional RVS methods for seismic risk assessment for estimating the current situation of existing buildings: A state-of-the-art review,” *Sustainability*, vol. 14, no. 5, p. 2583, 2022.

[28] G. Grünthal, “European macroseismic scale 1998 (EMS-98),” 1998.

[29] Z. V Milutinovic and G. S. Trendafiloski, “Risk-UE An advanced approach to earthquake risk scenarios with applications to different european towns,” *Contract: EVK4-CT-2000-00014, WP4: Vulnerability of current buildings*, pp. 1–111, 2003.

[30] S. Otani, “Seismic vulnerability assessment methods for buildings in Japan,” *Earthquake Engineering and Engineering Seismology*, vol. 2, no. 2, pp. 47–56, 2000.

[31] Japan Building Disaster Prevention Association (JBDPA), “Standard for Evaluation of Seismic Capacity of Existing Reinforced Concrete Buildings.,” 2017.

[32] V. J. Moseley, S. E. Dritsos, and D. L. Kolaksis, “Pre-earthquake fuzzy logic and neural network based rapid visual screening of buildings,” *Structural engineering and mechanics: An international journal*, vol. 27, no. 1, pp. 77–97, 2007.

[33] NZSEE, “Assessment and Improvement of the Structural Performance of Buildings in Earthquake, Recommendations of a NZSEE Study Group on Earthquake Risk Buildings,” 2006.

[34] P. Kapetana and S. Dritsos, “Seismic assessment of buildings by rapid visual screening procedures,” *Earthquake Resistant Engineering Structures VI*, vol. 93, p. 409, 2007.

[35] O. E. Planning, “Protection Organization (2000) Provisions for Pre-Earthquake Vulnerability Assessment of Public Buildings (Part A),” *OASP (Greek Earthquake Planning and Protection Organization), Athens, Greece*.

[36] R. Fathi-Fazl, Z. Cai, E. Jacques, and W. L. Cortés-Puentes, “Methodology for seismic risk screening of existing buildings in Canada: Structural scoring system,” *Canadian Journal of Civil Engineering*, vol. 48, no. 3, pp. 250–262, 2021.

[37] R. Sinha and A. Goyal, “A national policy for seismic vulnerability assessment of buildings and procedure for rapid visual screening of buildings for potential seismic vulnerability,” *Report to Disaster Management Division, Ministry of Home Affairs, Government of India, Hindistan*, pp. 1–9, 2004.

[38] D. C. Rai, “Review of documents on seismic evaluation of existing buildings,” *Department of Civil Engineering, Indian Institute of Technology Kanpur India*, 2005.

[39] S. K. Jain, K. Mitra, M. Kumar, and M. Shah, “A proposed rapid visual screening procedure for seismic evaluation of RC-frame buildings in India,” *Earthquake Spectra*, vol. 26, no. 3, pp. 709–729, 2010.

[40] Bureau of Indian Standards (BIS), “IS 13935: Indian Standard Guidelines for Repair and Seismic Strengthening of Buildings.,” 2009.

[41] C. V. R. Murty *et al.*, “A methodology for documenting housing typologies in the moderate-severe seismic zones,” in *proceedings of the 15th world conference on earthquake engineering, Lisbon, Portugal*, 2012, pp. 24–28.

[42] A. Bhalkikar and R. P. Kumar, “A comparative study of different rapid visual survey methods used for seismic assessment of existing buildings,” in *Structures*, Elsevier, 2021, pp. 1847–1860.

[43] G. of I. National Disaster Management Authority of Home Affairs, “A Primer on Rapid Visual Screening (RVS) Consolidating Earthquake Safety Assessment Efforts in India,” 2020.

[44] A. , Ansal *et al.*, “Earthquake Master Plan for Istanbul.,” 2003.

[45] Ministry for Environment and Urban Planning, “Principles for Identifying Risky Buildings.,” 2019.

[46] S. Tesfamariam and M. Saatcioglu, “Risk-based seismic evaluation of reinforced concrete buildings,” *Earthquake Spectra*, vol. 24, no. 3, pp. 795–821, 2008.

[47] Z. Şen, “Rapid visual earthquake hazard evaluation of existing buildings by fuzzy logic modeling,” *Expert Syst Appl*, vol. 37, no. 8, pp. 5653–5660, 2010.

[48] M. C. Ningthoujam and R. P. Nanda, “Rapid visual screening procedure of existing building based on statistical analysis,” *International journal of disaster risk reduction*, vol. 28, pp. 720–730, 2018.

[49] O. Coskun, A. Aldemir, and M. Sahmaran, “Rapid screening method for the determination of seismic vulnerability assessment of RC building stocks,” *Bulletin of Earthquake Engineering*, vol. 18, pp. 1401–1416, 2020.

[50] M. H. Arslan, M. Ceylan, and T. Koyuncu, “An ANN approaches on estimating earthquake performances of existing RC buildings,” *Neural Network World*, vol. 22, no. 5, p. 443, 2012.

[51] Y. Zhang, H. V Burton, H. Sun, and M. Shokrabadi, “A machine learning framework for assessing post-earthquake structural safety,” *Structural safety*, vol. 72, pp. 1–16, 2018.

[52] S. Mangalathu, S.-H. Hwang, E. Choi, and J.-S. Jeon, “Rapid seismic damage evaluation of bridge portfolios using machine learning techniques,” *Eng Struct*, vol. 201, p. 109785, 2019.

[53] E. Harirchian, T. Lahmer, V. Kumari, and K. Jadhav, “Application of support vector machine modeling for the rapid seismic hazard safety evaluation of existing buildings,” *Energies (Basel)*, vol. 13, no. 13, p. 3340, 2020.

[54] O. Coskun and A. Aldemir, “Machine learning network suitable for accurate rapid seismic risk estimation of masonry building stocks,” *Natural Hazards*, vol. 115, no. 1, pp. 261–287, 2023.

[55] A. F. Hassan and M. A. Sozen, “Seismic vulnerability assessment of low-rise buildings in regions with infrequent earthquakes,” *ACI Struct J*, vol. 94, no. 1, pp. 31–39, 1997.

[56] K. J. Elwood and J. P. Moehle, “Idealized backbone model for existing reinforced concrete columns and comparisons with FEMA 356 criteria,” *The Structural Design of Tall and Special Buildings*, vol. 15, no. 5, pp. 553–569, 2006.

[57] Z. Riahi, K. J. Elwood, and S. M. Alcocer, “Backbone model for confined masonry walls for performance-based seismic design,” *Journal of structural engineering*, vol. 135, no. 6, pp. 644–654, 2009.

[58] H. Luo and S. G. Paal, “Machine learning–based backbone curve model of reinforced concrete columns subjected to cyclic loading reversals,” *Journal of Computing in Civil Engineering*, vol. 32, no. 5, p. 04018042, 2018.

[59] Z. T. Deger and G. Taskin, “A novel GPR-based prediction model for cyclic backbone curves of reinforced concrete shear walls,” *Eng Struct*, vol. 255, p. 113874, 2022.

[60] G. Ma, Y. Wang, and H.-J. Hwang, “Genetic programming-based backbone curve model of reinforced concrete walls,” *Eng Struct*, vol. 283, p. 115824, 2023.

[61] D. Yu, T. Gai, S. Yang, S. Zeng, and J. C.-W. Lin, “An effective multi-time series model of RC column backbone curve identification,” *Case Studies in Construction Materials*, vol. 20, p. e03183, 2024.

[62] C. Sim, C. Song, N. R. Skok, A. Irfanoglu, S. Pujol, and M. Sozen, “Database of low-rise reinforced concrete buildings with earthquake damage,” 2016. [Online]. Available: https://datacenterhub.org

[63] P. Shah, S. Pujol, A. Puranam, and L. Laughery, “2015 Nepal Earthquake Building Performance Database,” 2015. [Online]. Available: https://datacenterhub.org

[64] NCREE, “2016 Taiwan (Meinong) Earthquake,” 2016. [Online]. Available: https://datacenterhub.org

[65] E. Villalobos, C. Sim, J. P. Smith-Pardo, P. Rojas, S. Pujol, and M. E. Kreger, “The 16 April 2016 Ecuador earthquake damage assessment survey,” *Earthquake Spectra*, vol. 34, no. 3, pp. 1201–1217, 2018, [Online]. Available: https://datacenterhub.org

[66] C. Sim, L. Laughery, T. C. Chiou, and P. W. Weng, “Pohang Earthquake—Reinforced Concrete Building Damage Survey; DEEDS, Purdue University Research Repository: Lafayette, IN, USA, 2018,” 2017. [Online]. Available: https://datacenterhub.org

[67] H. S. Akyuz *et al.*, “Surface rupture and slip distribution of the 12 November 1999 Duzce earthquake (M 7.1), North Anatolian fault, Bolu, Turkey,” *Bulletin of the Seismological Society of America*, vol. 92, no. 1, pp. 61–66, 2002.

[68] Ö. Aydan, R. Ulusay, and M. Miyajima, “The bingöl earthquake of may 1, 2003,” *Japan Society of Civil: Tokyo, Japan*, 2003.

[69] H. Kaplan, S. Yilmaz, H. Binici, E. Yazar, and N. Çetinkaya, “May 1, 2003 Turkey—Bingöl earthquake: damage in reinforced concrete structures,” *Eng Fail Anal*, vol. 11, no. 3, pp. 279–291, 2004.

[70] M. G. Johnson and D. R. Fick, “Generalized trends of severe damage observed from building surveys of seven different Earthquakes,” in *Eleventh US National Conference on Earthquake Engineering, Los Angeles, California, June*, 2018.

[71] N. Jardine and C. J. van Rijsbergen, “The use of hierarchic clustering in information retrieval,” *Information storage and retrieval*, vol. 7, no. 5, pp. 217–240, 1971.

[72] L. Derczynski, “Complementarity, F-score, and NLP Evaluation,” in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, 2016, pp. 261–266.

[73] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Adv Neural Inf Process Syst*, vol. 30, 2017.

[74] L. S. Shapley, “A value for n-person games,” 1953.

[75] T. Kim, Y. Chu, S. R. Kim, and D. Bhandari, “Seismic behavior of domestic piloti-type buildings damaged by 2017 Pohang earthquake,” *Journal of the Earthquake Engineering Society of Korea*, vol. 22, no. 3, pp. 161–168, 2018.

[76] J. Shin, I. Choi, and J. Kim, “Rapid decision-making tool of piloti-type RC building structure for seismic performance evaluation and retrofit strategy using multi-dimensional structural parameter surfaces,” *Soil Dynamics and Earthquake Engineering*, vol. 151, p. 106978, 2021.

[77] H. O. Wood and F. Neumann, “Modified Mercalli intensity scale of 1931,” *Bulletin of the Seismological Society of America*, vol. 21, no. 4, pp. 277–283, 1931.

[78] D. J. Wald, V. Quitoriano, T. H. Heaton, H. Kanamori, C. W. Scrivner, and C. B. Worden, “TriNet ‘ShakeMaps’: Rapid generation of peak ground motion and intensity maps for earthquakes in southern California,” *Earthquake Spectra*, vol. 15, no. 3, pp. 537–555, 1999.

[79] D. J. Wald, B. C. Worden, V. Quitoriano, and K. L. Pankow, “ShakeMap manual: technical manual, user’s guide, and software guide,” 2005.

[80] USGS, “1999 Duzce Earthquake,” https://earthquake.usgs.gov/earthquakes/eventpage/usp0009hev/shakemap.

[81] USGS, “2003 Bingol Earthquake,” https://earthquake.usgs.gov/earthquakes/eventpage/usp000bwbc/shakemap.

[82] USGS, “2015 Nepal Earthquake,” https://earthquake.usgs.gov/earthquakes/eventpage/us20002926/shakemap.

[83] USGS, “2016 Taiwan Earthquake,” https://earthquake.usgs.gov/earthquakes/eventpage/us20004y6h/shakemap.

[84] USGS, “2016 Ecuador Earthquake,” https://earthquake.usgs.gov/earthquakes/eventpage/us20005j32/shakemap.

[85] USGS, “2017 Pohang Earthquake,” https://earthquake.usgs.gov/earthquakes/eventpage/us2000bnrs/shakemap.

[86] K. Kostinakis, A. Athanatopoulou, and K. Morfidis, “Correlation between ground motion intensity measures and seismic damage of 3D R/C buildings,” *Eng Struct*, vol. 82, pp. 151–167, 2015.

[87] Y. L. X. Fei, “‘RC column backbone curve dataset’, in RC column backbone curve dataset [Version 3]. DesignSafe-CI.,” 2020. doi: https://doi.org/10.17603/ds2-eppg-5133.

[88] M. Tazarv, K. Won, Y. Jang, K. Hart, and E. Greeneway, “Post-earthquake serviceability assessment of standard RC bridge columns using computer vision and seismic analyses,” *Eng Struct*, vol. 272, p. 115002, 2022.

[89] Balaji Sivaramakrishnan, “Non - Linear Modeling Parameters for Reinforced Concrete Columns Subjected to Seismic Loads,” The University of Texas at Austin, 2010.

[90] W. Ghannoum *et al.*, "NEES: ACI 369 Rectangular Column Database," 2015. [Online]. Available: https://datacenterhub.org/resources/255

[91] A. Di Bucchianico, "Coefficient of determination (R 2)," *Encyclopedia of statistics in quality and reliability*, 2008.

[92] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature," *Geosci Model Dev*, vol. 7, no. 3, pp. 1247–1250, 2014.