ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

EARTHQUAKE RISK ASSESSMENT USING ARTIFICIAL INTELLIGENCE

Ph.D. THESIS

Bilal EIN LAROUZI

Department of Civil Engineering

Structure Engineering Programme

FEBRUARY 2026

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

EARTHQUAKE RISK ASSESSMENT USING ARTIFICIAL INTELLIGENCE

Ph.D. THESIS

Bilal EIN LAROUZI
(501202004)

Department of Civil Engineering

Structure Engineering Programme

Thesis Advisor: Prof. Dr. Yasin FAHJAN

FEBRUARY 2026

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YAPAY ZEKÂ KULLANILARAK DEPREM RISK DEĞERLENDIRMESI

DOKTORA TEZİ

Bilal EIN LAROUZI
(501202004)

İnşaat Mühendisliği Anabilim Dalı

Yapı Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Yasin FAHJAN

ŞUBAT 2026

Bilal EIN LAROUZI, a Ph.D. student of İTU Graduate School, student ID 501202004, successfully defended the thesis entitled "EARTHQUAKE RISK ASSESSMENT USING ARTIFICIAL INTELLIGENCE", which he prepared after fulfilling the requirements specified in the associated legislation, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Yasin FAHJAN** .....
Istanbul Technical University

Jury Members : **Prof. Dr. Gülüm TANIRCAN** .....
Boğaziçi University

Assoc. Prof. Dr. Barlas Özden ÇAĞLAYAN .....
Istanbul Technical University

Asst. Prof. Dr. Ahmet Anıl DINDAR .....
Gebze Technical University

Asst. Prof. Dr. Ömer Tuğrul TURAN .....
Istanbul Technical University

Date of Submission : 26 January 2026
Date of Defense : 16 Febraury 2026

To my Family,

FOREWORD

This doctoral dissertation marks the completion of several years of dedicated study, research, and personal development. Its successful completion would not have been possible without the guidance, support, and encouragement of many individuals and institutions, to whom I am sincerely grateful.

First and foremost, I would like to express my deep appreciation to Istanbul Technical University for providing an inspiring academic environment and the essential resources required to carry out this research. I am also grateful to the faculty and staff of the Civil Engineering Department for their continuous support throughout my doctoral studies.

I am profoundly indebted to my supervisor, Prof. Dr. Yasin FAHJAN, for his exceptional guidance, patience, and constant encouragement. His insightful feedback, constructive criticism, and commitment to academic excellence have played a fundamental role in shaping this research.

I sincerely thank the members of my Thesis Steering Committee, Assoc. Prof. Dr. Barlas Özden ÇAĞLAYAN and Asst. Prof. Dr. Ahmet Anıl DINDAR, for their valuable suggestions and continuous support during the preparation of this dissertation. Their constructive contributions significantly enhanced both the quality and scope of this work.

I also extend my appreciation to the members of the thesis jury, Prof. Dr. Gülüm TANIRCAN and Asst. Prof. Dr. Ömer Tuğrul TURAN, for their time, thoughtful evaluations, and insightful comments, which greatly improved the final version of this dissertation.

Furthermore, I would like to acknowledge my professors and colleagues whose discussions, collaboration, and friendship enriched my academic experience and broadened my perspective throughout this journey.

Finally, I owe my deepest gratitude to my parents and siblings for their unconditional love, understanding, and continuous encouragement. Their support has been a constant source of strength and motivation. This achievement belongs to them as much as it does to me.

This research was supported by the Research Fund of Istanbul Technical University (Project Number: 45808).

February 2026

Bilal EIN LAROUZI
Ph.D. Structural Engineering

TABLE OF CONTENTS

Page

ABBREVIATIONS

RF	**:** Random Forest Classifier
DTR	**:** Decision Tree Regressor
DTC	**:** Decision Tree Classifier
GB	**:** Gradiant-Boosting Classifier
XGB	**:** Extreme Gradiant-Boosting Classifier
HGB	**:** Hist-Gradiant Boosting Classifier
SD	**:** Sever Damage
MD	**:** Moderate Damage
LD	**:** Low Damage
ND	**:** No Damage
RC	**:** Reinforced Concrete
Rrup	**:** Rupture Distance From Fult Lin

TABLE OF TABLES

Page

TABLE OF FIGURES

Page

EARTHQUAKE RISK ASSESSMENT USING ARTIFICIAL INTELLIGENCE

SUMMARY

Earthquakes are among the most unpredictable and devastating natural hazards, posing serious risks to human life and to social and economic systems. Although earthquakes themselves cannot be prevented, the findings of this study indicate that effective disaster management—particularly when supported by rapid and dependable post-earthquake damage assessment—can substantially decrease casualties and mitigate social and economic impacts. Rapid estimation of structural damage plays a critical role in directing emergency response efforts, prioritizing field inspections, and ensuring efficient allocation of available resources. In this regard, machine learning has attracted growing interest over the past decade due to its strong capability to analyze large and complex datasets and to model nonlinear relationships among seismic, structural, and geospatial variables.

This thesis proposes and validates a machine learning–based framework for estimating building damage following earthquakes by utilizing real field inspection data that represent actual observed damage and structural performance. These field observations are combined with scenario-based ground motion parameters generated through REDAS software, which provide consistent seismic demand estimates at specific building locations. Furthermore, geospatial attributes obtained from open-source databases are incorporated to enhance the overall dataset. The developed framework is assessed through two significant earthquake case studies in Turkiye: the 2020 Izmir earthquake and the 2023 Kahramanmaras earthquake sequence.

The Izmir earthquake serves as a well-documented benchmark for model evaluation. Post-earthquake inspection data were combined with seismic and geospatial information to train and test multiple machine learning algorithms. Two dataset configurations were examined to assess the influence of feature selection. The results clearly demonstrate that including building age as an input parameter significantly improves model performance, increasing prediction accuracy to nearly 90% compared to approximately 80% when this feature is excluded. Confusion matrix analysis shows that models incorporating building age are far more effective at identifying severely damaged buildings, emphasizing the critical role of this parameter in representing structural vulnerability.

The Kahramanmaras earthquake sequence represents a severe seismic event that impacted a large geographic area and a diverse building inventory. The related dataset contains both masonry and reinforced concrete structures and is analyzed using three damage classification approaches: binary, three-class (traffic-light), and four-class systems, applied at the building, district, and city levels. Although the binary classification approach produces relatively high overall accuracy, it leads to an unacceptably high proportion of unsafe buildings being incorrectly identified as safe, which limits its practical applicability. The four-class classification scheme offers more detailed damage differentiation; however, it exhibits considerable confusion between neighboring damage states. In comparison, the three-class traffic-light classification provides the most balanced and reliable performance, showing minimal confusion between safe and unsafe categories. Among the evaluated machine learning algorithms, the Bagging classifier delivers the most stable and consistent performance under this classification framework.

At district and city scales, the proposed models successfully reproduce observed damage distributions, confirming their suitability for large-scale damage assessment and regional decision-making. Overall, this thesis presents a practical and transferable machine learning framework for rapid post-earthquake damage estimation based on real observed data. The findings indicate that the three-class traffic-light damage classification offers the best balance between accuracy, clarity, and operational relevance for emergency response. The proposed approach provides a realistic and scalable solution that can support disaster management efforts and contribute to reducing human and economic losses in earthquake-prone regions.

YAPAY ZEKÂ KULLANILARAK DEPREM RISK DEĞERLENDIRMESI

ÖZET

Depremler, insan yaşamı ve sosyo-ekonomik sistemler için ciddi tehditler oluşturan, en öngörülemez ve yıkıcı doğal afetler arasında yer almaktadır. Depremlerin kendileri önlenemese de, bu çalışma hızlı ve güvenilir deprem sonrası hasar tespitiyle desteklenen etkili afet yönetiminin, can kayıplarını önemli ölçüde azaltabileceğini ve sosyal ile ekonomik kayıpları sınırlayabileceğini ortaya koymaktadır. Zamanında gerçekleştirilen hasar tahmini; acil müdahale faaliyetlerinin başlatılması, bina incelemelerinin önceliklendirilmesi ve kaynakların etkin biçimde tahsis edilmesi açısından kritik öneme sahiptir. Bu bağlamda, makine öğrenmesi; büyük ve karmaşık veri kümelerini işleme yeteneği ile sismik, yapısal ve coğrafi değişkenler arasındaki doğrusal olmayan ve birbirine bağlı ilişkileri yakalayabilmesi sayesinde son on yılda giderek daha fazla ilgi görmektedir.

Bu tez, gerçek saha incelemelerine dayalı ve gözlenen gerçek hasar durumlarını ile yapısal davranışı yansıtan veriler kullanılarak geliştirilmiş bir deprem sonrası bina hasar tahminine yönelik makine öğrenmesi tabanlı bir çerçevenin geliştirilmesini ve doğrulanmasını sunmaktadır. Saha verileri, REDAS yazılımı kullanılarak senaryo temelli analizler yoluyla üretilen yer hareketi parametreleri ile entegre edilmiş ve her bina konumu için tutarlı sismik talep tahminleri elde edilmiştir. Ayrıca, açık kaynaklı veritabanlarından elde edilen coğrafi özellikler de veri setini zenginleştirmek amacıyla çalışmaya dahil edilmiştir. Önerilen çerçeve, Türkiye'de meydana gelen iki büyük deprem olayı üzerinden değerlendirilmiştir: 2020 İzmir depremi ve 2023 Kahramanmaras deprem dizisi.

İzmir depremi, geliştirilen modellerin performansını değerlendirmek için iyi belgelenmiş bir gerçek dünya referansı sunmaktadır. Deprem sonrası saha incelemelerinden elde edilen yapısal özellikler ve hasar durumları, sismik ve coğrafi verilerle birleştirilerek birden fazla makine öğrenmesi algoritmasının eğitilmesi ve test edilmesi için kullanılmıştır. Özellik seçiminin etkisini incelemek amacıyla iki farklı veri seti yapılandırması ele alınmıştır. Elde edilen sonuçlar, bina yaşının girdi parametresi olarak eklenmesinin model performansını önemli ölçüde artırdığını açıkça göstermektedir. Bina yaşının dahil edildiği modellerde doğruluk oranı yaklaşık %90'a ulaşırken, bu parametrenin dışlandığı modellerde doğruluk yaklaşık %80 seviyesinde kalmıştır. Karışıklık matrisi analizleri, bina yaşını içeren modellerin ağır hasarlı binaları doğru biçimde tanımlamada çok daha başarılı olduğunu ortaya koymakta ve bu parametrenin yapısal kırılganlığın temsilindeki kritik rolünü vurgulamaktadır.

Kahramanmaras deprem dizisi, geniş bir bölgeyi ve farklı yapı türlerini etkileyen son derece şiddetli bir sismik olayı temsil etmektedir. Bu çalışmada oluşturulan veri seti, yığma ve betonarme binaları kapsamakta ve bina, ilçe ve şehir ölçeklerinde ikili sınıflandırma, üç sınıflı (trafik ışığı) ve dört sınıflı hasar sınıflandırma yaklaşımları kullanılarak değerlendirilmiştir. İkili sınıflandırma görece yüksek doğruluk değerleri sunsa da, güvenli olmayan binaların önemli bir kısmının güvenli olarak sınıflandırılması nedeniyle uygulamada ciddi riskler barındırmaktadır. Dört sınıflı hasar sınıflandırması daha ayrıntılı bilgi sağlamasına rağmen, komşu hasar seviyeleri arasındaki yüksek karışma nedeniyle güvenilirliğini kaybetmektedir. Buna karşılık, üç sınıflı trafik ışığı sınıflandırması, güvenli ve güvensiz binalar arasındaki düşük karışma oranı ile en dengeli ve güvenilir performansı sunmaktadır. İncelenen algoritmalar arasında, bu sınıflandırma şeması için en tutarlı ve sağlam sonuçları Bagging sınıflandırıcısının sağladığı görülmüştür.

İlçe ve şehir ölçeklerinde yapılan değerlendirmelerde, önerilen modellerin gözlenen hasar dağılımlarını başarıyla yeniden üretebildiği ve büyük ölçekli hasar tahmini ile bölgesel karar verme süreçlerini destekleyebilecek kapasitede olduğu doğrulanmıştır. Sonuç olarak, bu tez gerçek gözlem verilerine dayalı, pratik ve farklı bölgelere uyarlanabilir bir makine öğrenmesi tabanlı deprem sonrası hasar tahmin çerçevesi sunmaktadır. Bulgular, üç sınıflı trafik ışığı hasar sınıflandırmasının doğruluk, açıklık ve operasyonel kullanılabilirlik açısından en uygun yaklaşım olduğunu göstermektedir. Önerilen yöntem, afet yönetimi çalışmalarını destekleyebilecek ve deprem riski yüksek bölgelerde can ve mal kayıplarının azaltılmasına katkı sağlayabilecek gerçekçi ve ölçeklenebilir bir çözüm sunmaktadır.

1. INTRODUCTION

Natural hazards represent significant threats to human life and socio-economic stability, leading not only to casualties but also to substantial economic losses due to the damage and collapse of buildings. Among these hazards, earthquakes are especially destructive because of the intense ground shaking they produce and their unpredictable nature. In October 2020, a strong earthquake occurred in the western part of Turkiye, with its epicenter located in the Aegean Sea near Izmir. The event caused considerable destruction, particularly in the Bayraklı district, where numerous reinforced concrete residential buildings either collapsed or sustained severe damage. As a result of this earthquake, 117 people lost their lives, more than 1,000 were injured, and thousands of buildings throughout the city were affected [1]. In early 2023, Turkiye experienced one of the most devastating seismic disasters in its history when two major earthquakes occurred in sequence in the southern region of the country. This catastrophic event resulted in more than 53,000 fatalities, approximately 107,000 injuries, and damage to nearly 4 million buildings across the affected regions [2]. These disasters clearly highlight the critical importance of developing effective methods for seismic risk assessment and rapid damage prediction.

Risk and damage assessment are essential components of earthquake risk reduction, as they support mitigation planning and enable rapid response actions that help minimize loss of life and injuries. Accurate estimation of building damage immediately after an earthquake allows authorities to activate emergency plans, prioritize affected areas, and allocate resources more effectively. In this context, disaster management plays a critical role in reducing human and economic losses, particularly during the post-earthquake response phase. Effective disaster management relies on rapid and reliable identification of different building damage states in near real time, which is crucial for guiding evacuation, search and rescue operations, and ensuring the safety of both occupants and

emergency responders. Timely damage assessment also helps prevent secondary risks by restricting access to unsafe structures and supporting informed decision-making during the critical hours following a seismic event.

Over the past decades, a wide range of procedures have been developed for building-scale seismic damage assessment, with the primary objective of evaluating the expected performance of individual structures under earthquake loading. These methods are typically based on structural analysis, empirical observations, or simplified mechanical models, and they provide valuable insight into building vulnerability. As earthquake risk mitigation increasingly requires city-wide and regional decision-making, regional-scale damage assessment approaches have gained significant importance. Consequently, many methodologies have been proposed to rapidly estimate damage distribution over large urban areas, supporting emergency response and disaster management efforts. Among the traditional approaches, Fragility curves are widely used as a standard approach for evaluating seismic damage. Fragility-based methodologies quantify the likelihood that particular damage states will be exceeded based on seismic demand, and they are routinely applied in risk assessment frameworks such as FEMA and HAZUS [3]. While these methods are well established and physically interpretable, they often rely on simplified assumptions and require extensive calibration, which can limit their accuracy when applied to complex building stocks or real-time post-earthquake scenarios.

In recent years, artificial intelligence (AI) has emerged as a strong complementary approach to conventional engineering methods. Over the last decade, machine learning (ML) algorithms been increasingly integrated into structural and earthquake engineering, giving a chance for data-driven modeling of complex, nonlinear relationships between seismic demand, building characteristics, and observed damage. These approaches have promising potential for rapid and automated damage prediction, particularly when large post-earthquake datasets are available. As a result, AI-based methods are becoming an important component of modern seismic risk assessment, offering improved speed, adaptability, and predictive performance compared to traditional approaches. In AI-based seismic risk and damage assessment, artificial intelligence techniques are applied to assist in evaluating earthquake-induced damage to buildings and the associated risks. By learning from past seismic events, AI-based models can identify relationships between

seismic demand, structural characteristics, and observed damage outcomes that are difficult to capture using conventional analytical methods. These approaches enable rapid estimation of damage levels across different spatial scales, from individual buildings to entire urban areas.

The importance of historical data collection has grown significantly with the increasing adoption of machine learning techniques. Because datasets form the foundation of machine learning models, the quality and source of the data used for training play a critical role in model development. Training data can be obtained from numerical simulations, which represent the idealized behavior of modeled structural elements; from laboratory experiments and testing, which capture structural responses under controlled conditions; or from field observations that record structural performance during real events and therefore provide the most realistic representation of behavior. In particular, data collected from post-earthquake field surveys can greatly enhance machine learning–based damage assessment models. Such datasets contain valuable information about the actual behavior of structures under real ground motion conditions, reflecting uncertainties related to construction quality, detailing practices, material properties, structural configuration, and ground motion characteristics. When these field observations are integrated with machine learning algorithms, they enable the development of predictive models capable of estimating earthquake-induced damage at both building and regional scales.

Although seismic risk assessment has advanced in recent years, a major challenge remains: quickly and reliably estimating building damage after an earthquake using real, event-specific data. Many existing methods are developed using simulated datasets that assume ideal structural behavior and construction quality. As a result, these approaches often do not reflect the complex and imperfect conditions found in real disaster situations. Remote sensing techniques, such as satellite imagery, can cover large areas efficiently, but their usefulness is often limited by weather conditions and image resolution. In addition, they commonly classify buildings only as collapsed or not collapsed, which fails to capture the full range of damage levels needed for effective emergency response and recovery planning. Traditional fragility analyses also have limitations. While they provide probabilistic estimates of damage, they do not directly translate into clear, actionable damage categories.

The main contribution of this thesis is the development of an integrated, multi-source real-world dataset that combines post-earthquake damage survey records with building structural characteristics, ground-motion parameters, and site-specific information. This comprehensive integration creates a robust and realistic dataset that allows machine learning models to capture actual damage mechanisms rather than relying on simplified or idealized structural responses. Notably, the framework is developed entirely using real post-earthquake observations, without incorporating any synthetic or numerically simulated damage data. Furthermore, the study proposes a multi-scale damage prediction framework capable of estimating earthquake-induced damage at the building, district, and city levels within a single methodological approach. This multi-scale capability supports post-earthquake disaster management by connecting detailed engineering-level assessments with broader emergency response and planning needs. To enhance the practical applicability of the results, three complementary damage classification schemes are introduced: (1) a detailed multi-level damage scale for engineering evaluation, (2) a simplified three-level traffic-light system designed for rapid safety tagging, prioritization of building inspections, and allocation of emergency resources, and (3) a binary classification that differentiates between safe and unsafe buildings. The proposed framework applies to both reinforced concrete and masonry structures, which constitute the most common building types in earthquake-prone regions. Although the study employs established ensemble machine learning algorithms, its novelty lies in the exclusive use of real post-earthquake data, the integration of multiple data sources, and the implementation of a multi-scale damage prediction strategy. Consequently, the framework provides a practical, data-driven, and disaster-oriented approach for rapid post-earthquake building damage assessment and decision support.

1.1 Purpose of Thesis

The primary objective of this thesis is to support post-earthquake disaster management by providing rapid and reliable information on building damage. To achieve this, the study develops a building damage estimation model based on machine learning techniques and artificial intelligence. The proposed model is trained on real post-earthquake field data

that reflect the actual structural behavior of buildings under seismic loading. To enhance robustness and general applicability, the primary dataset is integrated with multiple existing post-earthquake damage datasets.

A key focus of the research is to design a model that requires a minimal number of input features, avoiding complex structural analyses and detailed building information. This simplified approach allows the model to be applied quickly in the instantaneous consequence of the earthquake, when time and data availability are limited.

By enabling rapid damage estimation, the proposed model aims to assist disaster management authorities and emergency response teams in early decision-making processes, such as prioritizing building inspections, allocating emergency resources, and planning response and recovery actions. Ultimately, this thesis seeks to contribute to more effective post-earthquake disaster management and to the reduction of human and economic losses in earthquake-prone regions.

1.2 Literature Review

Machine learning techniques in earthquake engineering have increasingly emerged as a transformative alternative to traditional computationally intensive approaches, enabling faster and more accurate predictive frameworks that support post-disaster emergency response, urban resilience planning, and structural safety evaluation. Numerous studies have explored the application of machine learning models for seismic damage prediction and structural performance assessment. Kiani et al. examined classification-based machine learning approaches for deriving seismic fragility curves using 2,000 nonlinear dynamic analyses of an eight-story steel frame subjected to hazard-consistent ground motions. Ten classification algorithms, including Random Forest, Support Vector Machines, and neural networks, were evaluated in terms of dataset size and class imbalance effects. The results indicated that Random Forest and Quadratic Discriminant Analysis provided the best performance for imbalanced and balanced datasets, respectively, with prediction accuracies exceeding 90%. Logistic Regression demonstrated relatively stable performance regardless of dataset size, while other algorithms showed considerable variability, emphasizing the importance of selecting appropriate algorithms based on data characteristics [4]. Similarly, Kim et al.

demonstrated the potential of convolutional neural networks (CNNs) to replace computationally demanding nonlinear time-history analyses when predicting maximum responses of hysteretic systems. By converting ground motion records into image-based representations and combining them with structural parameters, the CNN model successfully predicted peak displacements with higher accuracy than traditional regression methods. The study highlighted the capability of deep learning architectures to capture complex frequency content and temporal features embedded in seismic records [5]. Wang et al. proposed a machine learning workflow to forecast the collapse probability of reinforced concrete columns already damaged during a mainshock when subjected to subsequent earthquakes. A synthetic dataset consisting of 10,000 numerical models was created using damage strain field theory and subjected to approximately 200,000 nonlinear time-history analyses representing various damage levels and aftershock scenarios. Seven machine learning algorithms were evaluated, including XGBoost, Random Forest, and Support Vector Machines. Among them, XGBoost achieved the highest performance, with F1-scores and accuracy exceeding 90%. Feature importance analysis revealed that residual displacement and the intensity of the subsequent earthquake were the most influential factors affecting collapse potential [6]. In another study, Zhou et al. investigated the cumulative effects of mainshock–aftershock sequences using 662 recorded sequences and 832 simulated structural systems representing eight single-degree-of-freedom models corresponding to different building types. Four ensemble algorithms—Random Forest, Extremely Randomized Trees, AdaBoost, and Gradient Boosting—were trained using 32 ground motion intensity measures. All models achieved correlation coefficients above 0.95, with Random Forest delivering the highest predictive accuracy. The results demonstrated that machine learning models can effectively capture complex cumulative damage patterns and energy dissipation mechanisms that are difficult to represent using conventional intensity measures alone [7]. Zhang et al. explored the relationship between structural damage patterns and residual collapse capacity using incremental dynamic analyses of a four-story reinforced concrete frame subjected to sequential earthquakes. Classification and Regression Trees (CART) and Random Forest models were used to classify buildings as safe or unsafe based on response parameters such as peak inter-story drift and residual drift. Random Forest

demonstrated superior robustness and predictive capability, showing that multi-dimensional damage indicators can be used to reliably estimate post-earthquake safety conditions [8]. Using a sensor-based approach, Moscoso Alcantara and Saito developed a predictive framework based on approximately 60,000 nonlinear time-history analyses of 600 buildings with different structural characteristics subjected to spectrum-matched ground motions. Five machine learning algorithms—Linear Regression, Decision Tree, Random Forest, XGBoost, and Multilayer Perceptron—were applied to predict maximum inter-story drift ratios and damage states. Multilayer Perceptron and XGBoost achieved the highest prediction accuracy, demonstrating the feasibility of using machine learning models as efficient alternatives to time-consuming structural analyses and field inspections [9]. Focusing on real post-earthquake data, Tocchi et al. utilized nearly 60,000 inspection records from the 2009 L'Aquila earthquake to establish a link between component-level damage and overall building usability. Random Forest, Support Vector Machines, and K-Nearest Neighbors were used to classify buildings into usable, partially usable, or unusable categories. Among these methods, Random Forest achieved the best predictive performance, effectively capturing nonlinear relationships between individual component damage and overall building functionality [10]. In another study focusing on Turkish reinforced concrete buildings, Cosgun evaluated machine learning as a rapid alternative to traditional seismic assessment techniques. A dataset of 531 buildings incorporating structural properties such as number of stories, material strength, and architectural irregularities was analyzed using Random Forest, Support Vector Machines, Logistic Regression, and C4.5 Decision Tree algorithms. Random Forest achieved an accuracy of approximately 92%, and sensitivity analysis indicated that building height and column properties significantly influence seismic performance [11]. Zhang et al. generated 11,200 reinforced concrete frame models through nonlinear time-history analyses to train machine learning models for predicting different damage states. Random Forest, Extreme Gradient Boosting, and Active Learning approaches achieved prediction accuracies above 80%, with Active Learning demonstrating strong performance even when trained with relatively small datasets, highlighting its potential for rapid evaluation under limited computational resources [12]. To address computational constraints in traditional structural analysis, Tang et al. created synthetic datasets based on 4,800

reinforced concrete frame models subjected to 100 ground motion records. Five machine learning models—Random Forest, Support Vector Machines, K-Nearest Neighbors, AdaBoost, and Multilayer Perceptron—were used to predict maximum inter-story drift ratios and damage states using seismic intensity measures and structural parameters. Random Forest achieved the highest accuracy by effectively capturing complex nonlinear relationships between structural properties and seismic demand [13]. Bhatta and Dang combined structural parameters with ground motion characteristics derived from nonlinear time-history analyses of more than 10,000 reinforced concrete models. Random Forest, Support Vector Machines, and K-Nearest Neighbors algorithms were used to classify buildings into five damage levels ranging from no damage to collapse. Random Forest achieved prediction accuracy exceeding 90%, demonstrating that the integration of structural and seismic parameters leads to robust predictive performance across diverse scenarios [14]. Extending this work, Bhatta et al. developed machine learning models capable of simultaneously predicting ground motion parameters and structural damage states. Random Forest Classifier, Support Vector Machines, and K-Nearest Neighbors were applied to classify buildings into five damage categories. Random Forest again achieved the highest predictive accuracy and demonstrated strong generalization capability when validated against testing datasets and historical data from the Gorkha earthquake [15]. Large-scale validation using real-world data was conducted by Ghimire et al., who analyzed approximately 760,000 buildings affected by the 2015 Gorkha earthquake. The dataset included information on construction type, building age, height, and macro-seismic intensities obtained from USGS ShakeMap. Five algorithms—Random Forest, Gradient Boosting, Logistic Regression, Support Vector Regression, and Gradient Boosting Regression—were tested for both discrete and continuous damage prediction. Random Forest and Gradient Boosting demonstrated the highest predictive performance, particularly for extreme damage cases, although class imbalance in real-world datasets remained a challenge [16]. Beyond structural datasets, machine learning has also been applied to remote sensing–based damage detection. Wang et al. developed an automated framework for detecting building damage from satellite imagery under severe class imbalance conditions using the xBD dataset. The proposed two-stage approach included building localization followed by damage classification, supported by

customized loss functions and sampling strategies to address the scarcity of severely damaged samples. The method significantly improved F1-scores for minority damage classes, enabling rapid damage assessment useful for emergency response coordination [17]. Similarly, Braik and Koliou introduced an end-to-end mapping framework that integrates high-resolution satellite imagery from the xBD dataset with Geographic Information System (GIS) data to produce georeferenced damage maps. The automated workflow demonstrated high precision in detecting building footprints and classifying damage levels, showing that the integration of deep learning and spatial analysis tools can support large-scale disaster visualization and resource allocation during recovery operations [18]. Overall, these studies demonstrate that machine learning approaches provide effective and computationally efficient alternatives for various earthquake engineering applications, including fragility curve development, structural response prediction, post-earthquake safety evaluation, and large-scale damage mapping. Ensemble tree-based methods—particularly Random Forest—consistently show strong predictive performance and robustness across different datasets and problem types, while deep learning architectures are particularly effective for complex pattern recognition tasks such as time-series response prediction and image-based damage detection.

2. METHODOLOGY

The methodology proposed in this thesis aims to estimate post-earthquake building damage through the application of Artificial Intelligence (AI) and Machine Learning (ML) techniques. The overall workflow is organized into four primary stages: dataset development, data preprocessing, model training and validation, and performance evaluation across different spatial scales.

Additionally, the methodology incorporates several damage-state classification schemes in order to determine the most appropriate representation of building damage conditions. To evaluate predictive performance, multiple machine learning algorithms—specifically Random Forest, Histogram Gradient Boosting, and the Bagging Classifier—are implemented and compared to identify the model that delivers the most reliable and consistent damage predictions.

This systematic framework enables a thorough evaluation of model performance and contributes to the development of an efficient approach for rapid post-earthquake damage assessment.

2.1 Traditional Method of Building Damage Estimation

Estimating building damage under seismic loading is a fundamental step in understanding structural vulnerability and informing risk mitigation strategies. The traditional approach integrates several key components: the assessment of building capacity, the evaluation of structural response to earthquake ground motion, the probabilistic estimation of damage through fragility curves, and the application of vulnerability-based methods for predicting expected damage. Together, these elements provide a detailed framework for evaluating the seismic performance of structures at both the individual and regional scale [19 - 21].

2.1.1 Building a capacity curve

The building capacity curve illustrates the ability of a structure to resist lateral forces generated by earthquakes as horizontal displacement increases. It is commonly obtained through a pushover analysis, where incremental static lateral loads are applied to the structure while tracking the resulting roof displacement. To allow comparison between structural capacity and seismic demand, the base shear forces and roof displacement values are transformed into spectral acceleration and spectral displacement by utilizing the modal properties of the building.

Each capacity curve represents the expected seismic performance of a building type and is defined by two key points: yield capacity and ultimate capacity. The yield capacity indicates the building's effective lateral strength, reflecting design strength, structural redundancy, code conservatism, and expected material properties. The ultimate capacity identifies the maximum strength reached when the structure reaches the state of mechanism. Beyond this point, the building may continue to deform without collapse, but it does not provide additional resistance to lateral loads.

The capacity curve is assumed linear up to the yield point, transitions to inelastic behavior between yield and ultimate, and remains plastic beyond the ultimate point, without further increase in strength. This representation provides a simplified but practical model of structural performance under seismic forces [20].

2.1.2 Building response calculation

Building response calculation involves estimating how a structure reacts to earthquake ground motion in terms of forces and displacements. This is typically done by applying seismic demand to the structural model and determining key response parameters such as spectral displacement or spectral acceleration. These parameters are evaluated at the performance point as shown in (Figure 2.1), where the seismic demand intersects the structural capacity. The resulting response reflects the expected global behavior of the building during shaking and serves as an input for damage and fragility assessment [20].

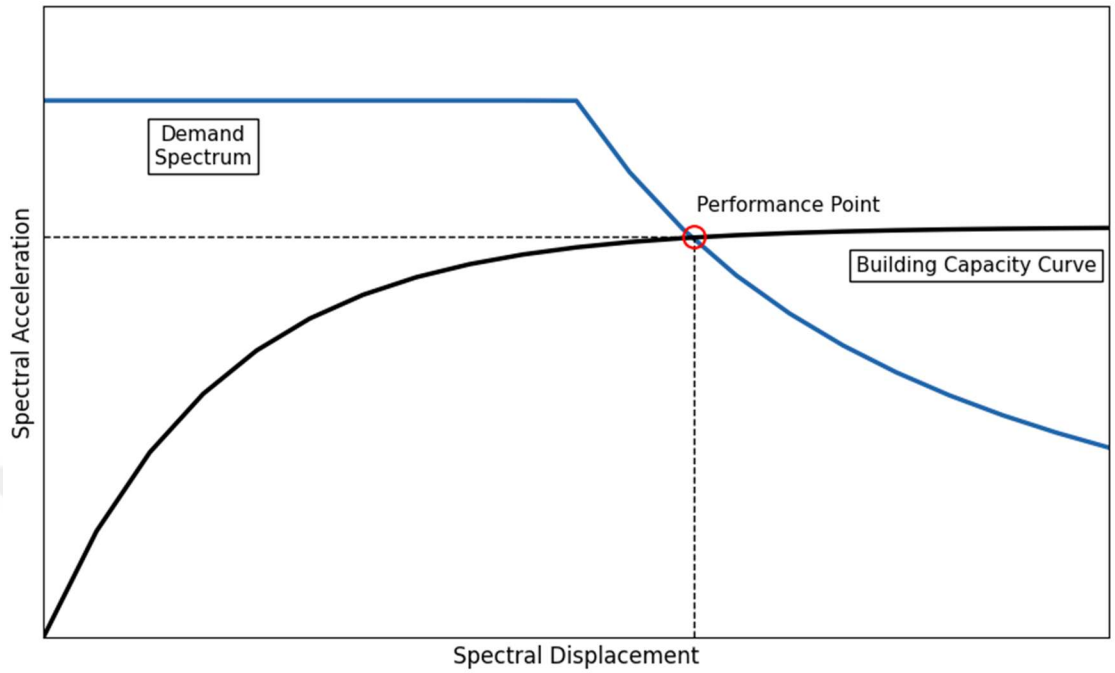

Figure 2.1: Building Performance Point [20].

2.1.3 Fragility curve

Fragility curves represent a probabilistic description of the risk that a building will attain or surpass particuler state of damage under a given level of seismic demand. Typically expressed as lognormal functions, these curves relate earthquake intensity measures—such as spectral displacement or spectral acceleration—to expected structural and nonstructural damage, while accounting for uncertainties in building capacity, material properties, construction quality, and ground shaking.

Fragility curves are defined for multiple states of damage—usually Slight, Moderate, Extensive, and Complete as illustrated in (Figure 2.2), allowing a detailed representation of building performance across different seismic intensities. For any given demand parameter, the probability of exceeding a particular damage state is obtained from the cumulative distribution function of the lognormal curve. Discrete damage-state probabilities are subsequently obtained by taking the difference between consecutive cumulative probabilities. By definition, the sum of all damage-state probabilities equals 100%, making them suitable for estimating expected damage distributions in individual buildings or across an entire inventory.

The median value of the demand parameter corresponds to the intensity at which there is a 50% probability of attaining or surpassing a particular damage state, while the associated dispersion reflects variability in building response. Fragility curves, therefore, provide a practical and widely used framework for seismic vulnerability assessment, scenario-based risk analyses, and post-earthquake planning [20].

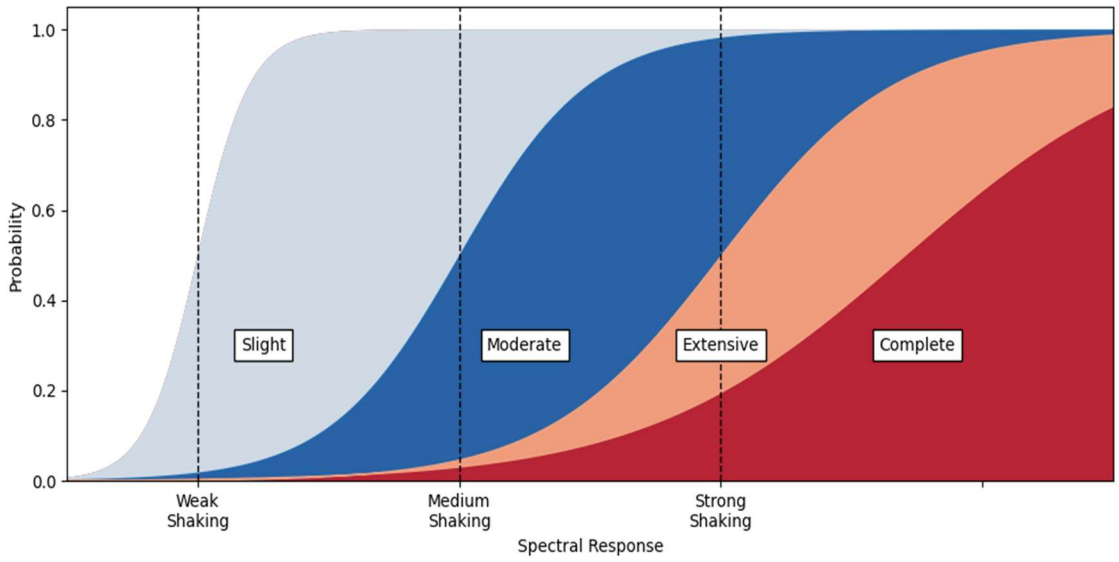

Figure 2.2: Fragility Curve [20].

2.1.4 Building damage estimation-based vulnerability method

The building damage estimation–based vulnerability approach assesses the susceptibility of structures to earthquake damage by emphasizing expected damage outcomes rather than relying solely on analytical representations of structural capacity. Within this framework, buildings are categorized into discrete damage states, typically ranging from no damage to severe or complete damage. Damage estimation is carried out by relating seismic intensity measures—such as peak ground acceleration or spectral acceleration—to key building characteristics, including structural system, construction material, building height, and construction period. To determine the probability of each damage state under a given level of ground motion, empirical relationships, damage probability matrices, or vulnerability functions are commonly applied.

This method enables vulnerability assessment at both individual building and aggregated urban scales. At the building level, it identifies structures more likely to experience significant damage under specific seismic scenarios. At the regional level, it supports the generation of damage distribution maps, which are essential for emergency planning, loss estimation, and prioritization of mitigation measures. Given that the required input parameters are generally available or can be reasonably estimated, the method is particularly suitable for wide-scale implementation [21].

2.2 Machine Learning

Machine Learning (ML) is a central area of Artificial Intelligence (AI) that focuses on teaching computers to learn from experience in a way that resembles how people learn from examples. Instead of giving a computer detailed instructions for every situation, ML allows it to study large amounts of data, notice meaningful patterns, and use what it has learned to make decisions or predictions.

The learning process often begins with training, where the model is shown many examples so it can understand the relationships between different pieces of information. As it encounters more data, the model gradually adjusts and improves itself, becoming more accurate and dependable over time. This ability to grow and adapt without constant human guidance is one of the reasons ML has become so valuable.

Today, machine learning is used in many areas because it can handle both simple tasks and highly complex problems. Whether recognizing patterns, classifying information, or making predictions, ML provides practical and flexible tools that help systems behave more intelligently and respond more effectively to real-world challenges.

2.2.1 Decision tree

Decision trees are supervised learning methods applicable to both classification and regression problems, commonly known as Decision Tree Classifiers (DTC) and Decision Tree Regressors (DTR), respectively. Their main objective is to estimate the target variable by learning a sequence of decision rules derived from input features. Conceptually, a decision tree represents a piecewise constant approximation of the data

structure, where the feature space is partitioned into multiple regions, each associated with a constant predicted output value.

One of the major strengths of decision trees is their straightforward structure and interpretability. They require little data preprocessing, can naturally accommodate both numerical and categorical variables, and do not depend on feature scaling. Furthermore, they are capable of handling multi-output tasks, which enhances their flexibility across different applications. Despite these advantages, decision trees are prone to overfitting, particularly when the tree becomes excessively deep and complex. In such cases, the model may learn noise and minor variations in the training data instead of capturing the underlying general patterns. Additionally, since decision trees generate stepwise predictions rather than smooth functions, their ability to represent continuous relationships may be limited.

A decision tree classifier constructs a hierarchical tree structure that progressively divides the dataset into increasingly homogeneous subsets. The process starts with a root node representing the entire training dataset, as illustrated in Figure 2.3. At each internal node, the algorithm selects the most informative feature for splitting the data. This selection is typically based on statistical measures such as information gain or Gini impurity, which evaluate how effectively a feature separates the data into distinct classes. After selecting the optimal feature, the dataset is divided into subsets according to defined threshold values. The procedure is repeated recursively, forming successive branches until a predefined stopping criterion is met. Common stopping conditions include limiting the maximum tree depth, specifying a minimum number of samples per node, or terminating when no further improvement in class purity is achieved [22].

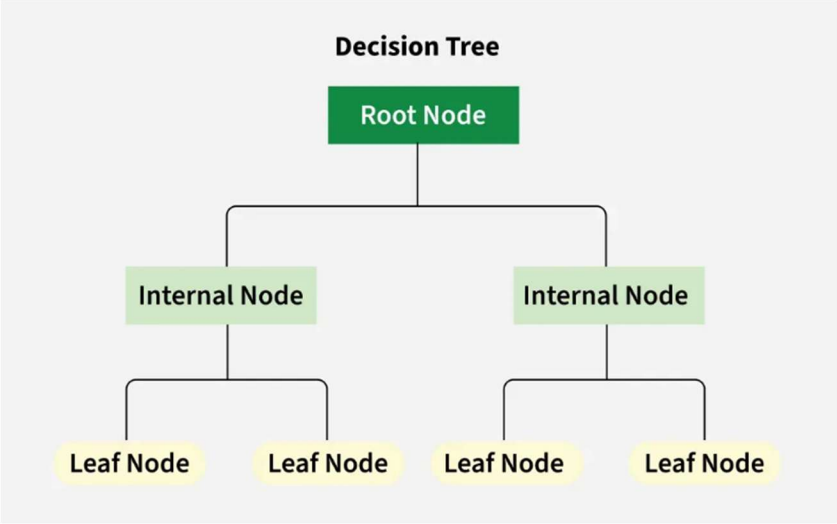

Figure 2.3: Decision Tree Algorithm Explanation [23].

2.2.2 Random forest

The Random Forest algorithm is an ensemble learning technique that combines multiple decision trees to achieve improved accuracy and more consistent predictions. Rather than relying on a single decision tree, which may be highly sensitive to variations in the training data, Random Forest constructs a collection of trees, each trained on different subsets of the data and features. This introduced randomness encourages diversity among the individual trees, helping to reduce correlated errors and enhancing the overall robustness of the model.

As each tree is built, it selects the best feature for splitting the data at every step. While an individual tree may capture noise or overfit to small patterns in the dataset, the ensemble effect allows the Random Forest to smooth out these errors. By averaging the outputs of many trees, the final prediction becomes more reliable and less affected by the weaknesses of any single tree. This reduces variance and helps the model generalize better, even at the cost of a minimal increase in bias.

In the scikit-learn implementation, each tree generates a probability for each state, and the Random Forest averages these probabilities to form its final prediction, as illustrated in (Figure 2.4). This method offers more nuanced results than a simple majority vote. Although training many trees can be computationally demanding, the combined model

benefits from the Law of Large Numbers: the more trees that are added, the more consistent and robust the predictions become [24,25].

Overall, Random Forest is known for its reliability, its resistance to overfitting, and its ability to handle complex datasets. These qualities make it a widely used and trusted machine learning algorithm in many practical applications [26].

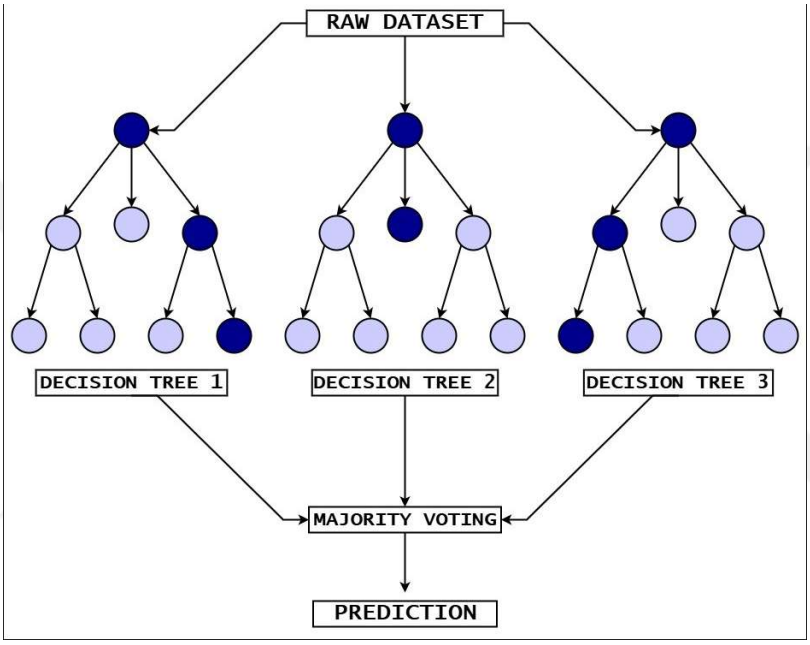

Figure 2.4: Random Forest Algorithm Explanation [27].

2.2.3 Gradient boosting and extreme gradient boosting

Gradient-boosted (GB) trees are advanced ensemble learning models that are used in wide areas for both classification and regression problems because of their strong predictive performance. Although gradient-boosted trees share some structural similarities with random forests—particularly in their reliance on decision trees as base learners—their fundamental difference lies in the learning strategy used to build the ensemble. (Figure 2.5) shows the gradient boosting method, which follows a sequential learning approach, in which trees are added one at a time, where every subsequent tree is designed to correct the mistakes of the existing ensemble.

In gradient boosting, the model begins with a simple initial prediction and iteratively improves it by minimizing a predefined loss function. In every iteration, a new decision

tree is built using the residuals of the preceding model as the target. This process can be interpreted as performing gradient descent in the function space, where each tree represents a step toward reducing the overall prediction error. As a result, the ensemble gradually becomes more accurate, with later trees focusing on the most difficult-to-predict samples [28].

Extreme Gradient Boosting (XGBoost), which is shown in (Figure 2.6), is an optimized application of gradient-boosted decision trees that introduces additional techniques to further enhance performance and efficiency. Compared to standard decision trees, XGBoost incorporates regularization mechanisms to control model complexity and reduce overfitting, as well as optimized tree construction and parallel processing capabilities. When sufficient computational resources are available, XGBoost models often achieve higher accuracy and better generalization performance than traditional decision tree-based methods, making them particularly suitable for complex and large-scale datasets [29].

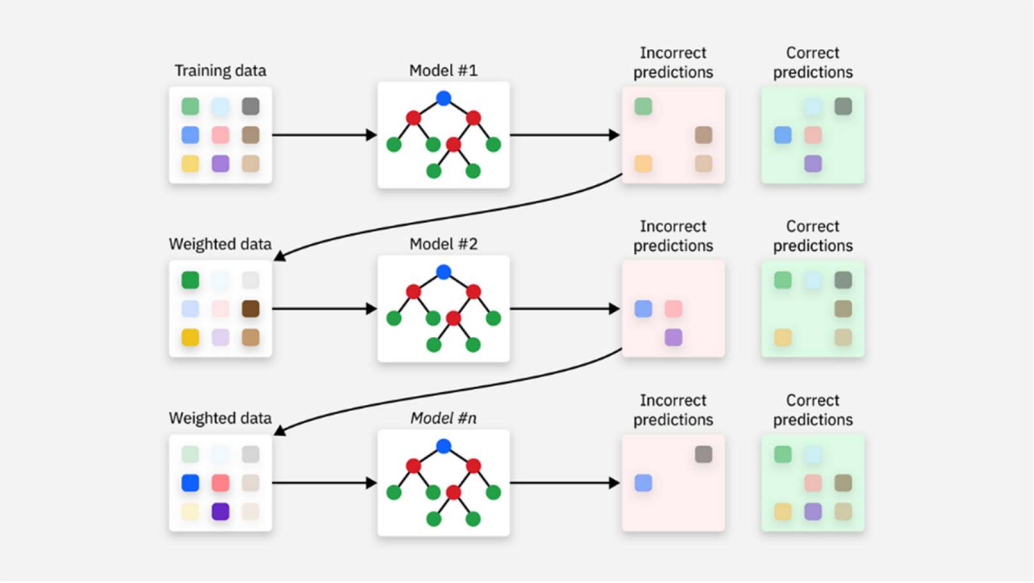

Figure 2.5: Gradient-Boosting Algorithm Explanation [30]**.**

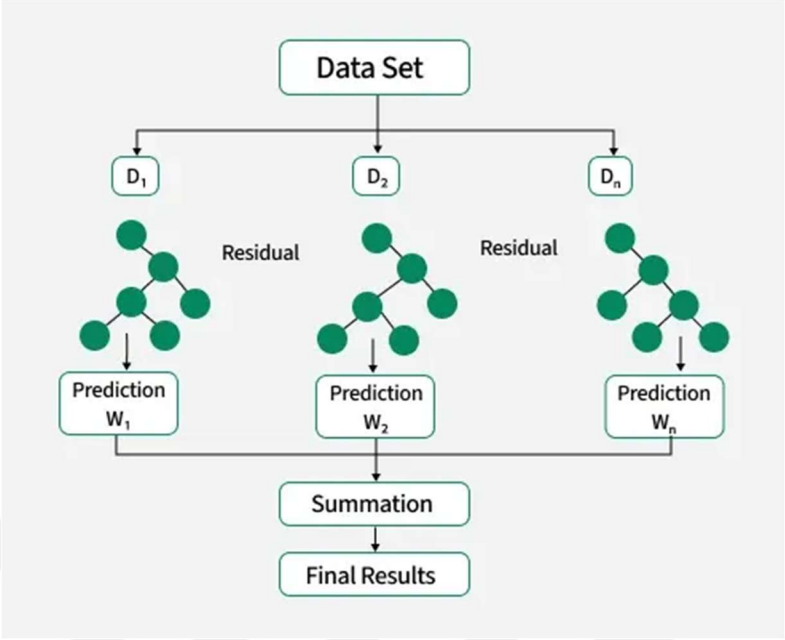

Figure 2.6: Extreme Gradient-Boosting Algorithm Explanation [31].

2.2.4 Histogram gradient boosting

Histogram Gradient Boosting is an optimized variant of Gradient Boosted Decision Trees (GBDT), an ensemble learning technique that constructs a series of decision trees sequentially. In this approach, each new tree is trained to reduce the errors made by the previously built trees, allowing the model to progressively improve its predictions. This iterative learning strategy enables the algorithm to capture complex relationships within the data and achieve high predictive accuracy for both regression and classification tasks, especially when applied to structured tabular datasets.

A defining characteristic of Histogram Gradient Boosting is its approach to handling continuous features. Instead of evaluating every possible split point across the full continuous range, the algorithm groups similar feature values into a fixed number of discrete bins. These bins convert continuous values into integer representations, allowing the model to use fast, integer-based data structures. This design greatly reduces computational cost while preserving the essential structure of the data.

By restricting split decisions to bin boundaries, the algorithm significantly accelerates the training process with minimal loss in accuracy. This efficiency makes Histogram Gradient

Boosting considerably faster than traditional gradient boosting methods while maintaining competitive predictive performance.

An additional advantage of the method is its inherent ability to handle missing values. The algorithm can process incomplete data without requiring explicit imputation, simplifying the preprocessing workflow and improving overall practicality, where the illustration of the method can be shown in Figure 2.7). Histogram Gradient Boosting offers a combination of processing efficiency, accuracy, and flexibility, which makes it a highly appropriate model for big and complex problems[32].

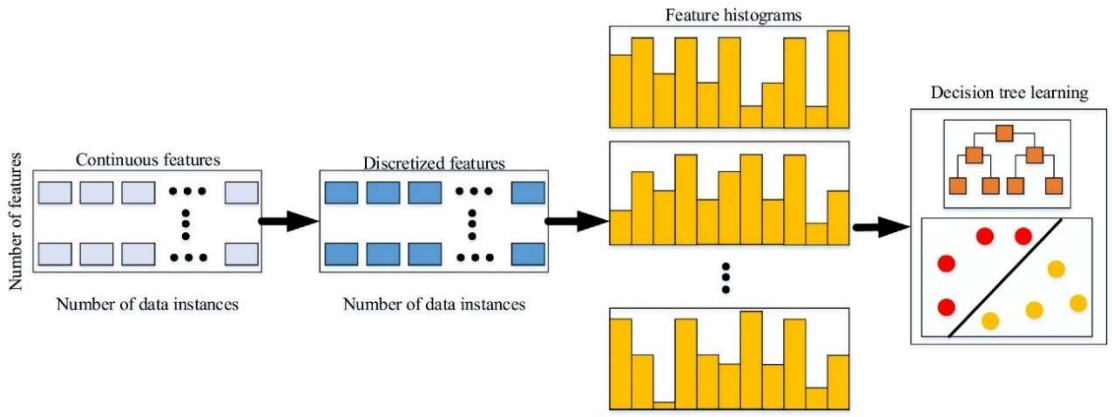

Figure 2.7: Hist-Gradiant Boosting Algorithm Explanation [33].

2.2.5 Bagging classifier

Bagging, an abbreviation for Bootstrap Aggregating, is an ensemble learning technique developed to improve the reliability and accuracy of predictive models. The method operates by training multiple instances of a base learner—commonly a decision tree—on different randomly generated subsets of the original dataset, as illustrated in Figure 2.8. For classification problems, the final prediction is determined through majority voting among the individual models, whereas for regression tasks, the outputs of the models are averaged to produce the final result.

When the subsets of data are drawn with replacement, the method is specifically called Bagging. If the subsets are drawn without replacement, the approach is referred to as Pasting. Variations of this method exist that introduce randomness in the features as well:

selecting random subsets of features is known as the Random Subspaces method, while combining random samples and random feature subsets is called Random Patches.

Bagging does not represent a single predictive model on its own; rather, it is a workflow for generating multiple distinctive models and integrating their outputs. By aggregating the results of these individual models, Bagging decreases the variance inherent in complex or "black-box" models, such as decision trees, and enhances the overall generalization ability of the ensemble. This makes it particularly effective for improving model reliability and robustness in various machine learning applications [34]

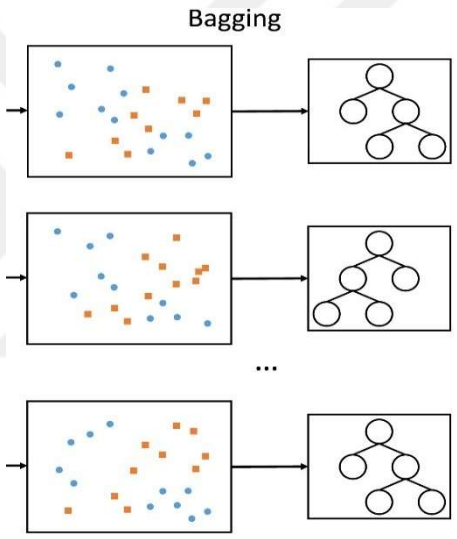

Figure 2.8: Bagging Classifier Algorithm Explanation [35].

2.2.6 Comparative analysis of different machine learning algorithms

The machine learning algorithms used in this study—Decision Tree, Random Forest, Gradient Boosting (including XGBoost), Histogram Gradient Boosting, and Bagging Classifier—represent varying levels of complexity and learning strategies. The Decision Tree is the fundamental component of the ensemble methods, offering high interpretability and straightforward implementation; however, it is prone to overfitting, particularly when applied to complex datasets. Ensemble techniques address this limitation by combining multiple decision trees to improve stability, reduce variance, and enhance generalization performance.

Random Forest and Bagging are parallel learning approaches in which multiple trees are trained independently using randomly sampled subsets of the data, primarily reducing variance and improving robustness against noise and outliers. Random Forest further increases model diversity by introducing random feature selection at each split, which strengthens overall generalization. In contrast, Gradient Boosting methods follow a sequential learning process, where each new tree focuses on correcting the errors of previous trees, enabling the model to capture complex nonlinear relationships more effectively; however, this approach typically requires careful hyperparameter tuning and greater computational resources.

Histogram Gradient Boosting enhances efficiency by applying feature binning, which accelerates training while maintaining competitive predictive performance, making it particularly suitable for large-scale datasets. Overall, Bagging and Random Forest are generally preferred when robustness and ease of tuning are priorities, whereas Gradient Boosting and XGBoost are better suited for tasks requiring high predictive accuracy and detailed error correction. In addition, model performance may vary depending on data quality, class imbalance, and feature characteristics. Therefore, algorithm selection depends on the trade-off between interpretability, computational cost, dataset complexity, and predictive accuracy, as summarized in Table 2.1.

Table 2.1: Comparison Between Utilized Models.

Model	Learning Strategy	Key Strengths	Main Limitations
Decision Tree	Single model	High interpretability, low preprocessing	Overfitting, low generalization
Random Forest	Bagging + feature randomness	Strong generalization, handles high-dimensional data	Less interpretable
Gradient Boosting	Sequential ensemble	High accuracy, captures complex patterns	Sensitive to tuning, higher cost
XGBoost	Optimized gradient boosting	Regularization, high performance	Computational demand
Histogram Gradient Boosting	Binned gradient boosting	Fast training, handles missing data	Reduced split resolution
Bagging Classifier	Parallel ensemble	Variance reduction, robustness	Limited bias reduction

2.3 Dataset Creation

The creation of a robust and representative dataset is one of the most essential steps in any machine learning–based prediction framework. In data-driven structural engineering studies, the effectiveness of a model is directly limited by the accuracy, completeness, and realistic nature of the input data. Therefore, in this research, the dataset development process was not considered a simple preparatory task, but rather a fundamental part of the overall methodology.

This study adopts a multi-source data integration strategy, combining (i) post-earthquake field survey data, (ii) ground motion parameters, and (iii) publicly available geospatial databases. The objectives of this integration are: capturing both structural characteristics and seismic demand parameters influencing damage formation and constructing a feature-rich, spatially consistent dataset suitable for supervised machine learning applications.

By systematically merging various data sources, the resulting dataset reflects the complex interaction between structural vulnerability, local site effects, and earthquake intensity measures. This integrated approach enhances predictive capability while preserving the physical meaning of each feature.

2.3.1 Data collection from the field

Post-earthquake field data form the main foundation of this research. Real damage records show how buildings actually behaved during an earthquake. They reflect real structural responses that cannot be fully captured using only simulations or analytical models. Using real inspection data ensures that the developed prediction model reflects actual differences in construction quality, materials, workmanship, and site conditions found in urban areas.

The field dataset used in this study was obtained from the Ministry of Environment, Urbanization, and Climate Change. It includes official post-earthquake building inspection records prepared by certified engineering teams. These inspections were carried out according to national damage assessment guidelines. The use of standardized procedures helps maintain consistency between different inspection teams and reduces subjectivity in assigning damage levels.

For each building, the dataset provides important structural information such as structural type, number of stories, construction material, construction age, and the observed damage state. In addition, each building has precise geographic coordinates. This allows the building data to be linked with ground motion parameters and other spatial information. The availability of geographic location is important because it makes it possible to associate each building with the specific seismic intensity and site conditions it experienced.

Damage assessments were based on detailed visual inspections carried out by experienced engineers. Inspectors examined structural elements, non-structural components, visible cracks, deformations, and signs of instability.

Using official inspection data provides several important advantages. First, the evaluations were performed by qualified engineers, which increases data reliability compared to informal or crowdsourced reports. Second, the dataset reflects real damage patterns observed under actual earthquake loading conditions. Third, the large number of inspected buildings allows the model to learn from different structural systems and site conditions, which improves its ability to generalize. Finally, the clearly defined damage states provide a suitable target variable for both multi-class and binary classification models.

2.3.2 Ground motion parameters (REDAS Software)

The ground motion parameters are numerical values assigned to specific geographic locations within an earthquake-affected area, offering detailed information about the spatial distribution and propagation of seismic waves from the epicenter.

REDAS is an integrated earthquake support system developed within the REDACt project to enhance preparedness and response efforts in Southeastern Europe and the Black Sea region. The platform consists of three main components: REDA.p, a desktop-based application for seismic damage assessment; a mobile application designed for rapid field data collection and updates; and an Educational Hub (Edu.Hub) aimed at strengthening user knowledge and technical capacity.

REDA.p supports both scenario-based and near real-time damage evaluations for different infrastructure systems, including buildings, pipelines, and geotechnical structures. It

operates by integrating standardized ground motion prediction models with a consistent building classification framework, ensuring uniformity in assessment procedures. By incorporating real-time information from regional seismic networks, REDAS can produce spatial damage distribution maps that assist emergency teams in prioritizing response actions immediately after an earthquake.

Beyond immediate disaster response, REDAS also plays a role in long-term urban resilience planning. Its capability to combine real-time seismic data, standardized structural information, and simulation-based scenarios makes it a comprehensive tool for both operational decision-making and strategic disaster preparedness initiatives [36].

2.3.3 Online databases

Several geospatial datasets are publicly available online, and integrating these resources with other data can significantly improve the quality and richness of a dataset. By combining multiple sources, it is possible to capture more detailed information and enhance the overall accuracy of analysis and modeling.

- *OpenStreetMap (OSM):* OpenStreetMap is a worldwide collaborative mapping initiative in which volunteers contribute to building a free and openly editable map of the Earth. It provides extensive geospatial data, including information on roads, buildings, land use, and natural features, and is regularly updated through sources such as GPS traces, aerial imagery, and publicly available datasets. Distributed under the Open Database License (ODbL), the data can be freely used, modified, and shared for both commercial and non-commercial purposes, provided that it remains open and appropriately attributed. Due to its broad spatial coverage and open accessibility, OpenStreetMap serves as a valuable resource for research activities, disaster management, urban development, and various other location-based applications [37].

- *Open-Elevation API:* The Open-Elevation API is a publicly available service that delivers elevation information based on geographic coordinates (latitude and longitude), relying on global digital elevation models such as NASA's SRTM and ASTER datasets. Its straightforward design and open-access structure make it a

practical tool for applications in environmental research, transportation planning, and geospatial analysis. When integrated with mapping datasets such as those from OpenStreetMap, it enables detailed terrain representation, elevation profiling, and hydrological assessments without requiring extensive local data processing. Although lightweight in design, the Open-Elevation API substantially improves the capabilities of geospatial systems for both academic research and real-world applications [38].

2.4 Damage State Classification Schemes

Damage states are an essential element in disaster management, providing critical information about the extent and severity of structural or infrastructural harm following an earthquake. Accurate knowledge of damage states is essential for prioritizing emergency response, coordinating rescue operations, and allocating resources efficiently in the direct aftermath of an earthquake.

In this thesis, three distinct damage state classification schemes are introduced, each designed to capture varying levels of structural damage with clarity and precision. These schemes serve as the foundation for developing a robust predictive model that can assess post-earthquake conditions more effectively. By integrating multiple classification approaches, the proposed workflow aims to improve the reliability of damage predictions, enabling decision-makers to respond rapidly and strategically. Ultimately, understanding and classifying damage states not only improves the accuracy of post-earthquake assessments but also strengthens the global resilience of communities and infrastructure against future seismic hazards.

2.4.1 Actual damage classification schemes

The classification of building damage was carried out by experienced engineering teams as part of the comprehensive post-earthquake assessment process. Each building was thoroughly inspected, with assessments based on visual observations, structural indicators, and expert engineering judgment. The aim was to categorize buildings

according to the severity of damage, providing essential information to guide emergency response, resource allocation, and recovery planning.

Buildings were classified into six distinct damage levels:

- *Collapsed Class:* This category includes buildings that experienced total or partial collapse, representing the most severe structural failure and posing immediate hazards to life and surrounding structures.
- *Required Immediate Demolishing:* Buildings in this class, while not entirely collapsed, were deemed unsafe and required urgent demolition to prevent potential accidents or further structural failures.
- *Severe-Damaged Class (SD):* These buildings persisted standing but suffered significant structural damage. They may pose safety risks and generally require extensive repair or reinforcement before they can be used safely.
- *Moderate-Damaged Class (MD):* Buildings in this class showed moderate damage that could affect structural integrity or functionality. Further detailed inspections are necessary to determine whether they can be repaired and safely reoccupied.
- *Low-Damaged Class (LD):* Structures in this group sustained minor, often superficial, damage such as cracks or small spalling. While these buildings remain largely safe, additional inspections and minor repairs are recommended to maintain long-term integrity.
- *Non-Damaged Class (ND):* This category includes buildings that showed no observable damage following the earthquake, indicating resilience to the seismic event and minimal risk for occupants.

This classification framework provides a systematic approach to assess and prioritize buildings for emergency action, repair, or demolition. It serves as a foundational dataset for developing predictive models that can estimate damage patterns across urban areas in future seismic events, improving both preparedness and response strategies.

2.4.2 Proposed damage state classification schemes

Damage classification represents one of the most critical components in post-earthquake risk evaluation, directly influencing emergency response strategies, recovery planning, and long-term resilience assessment. While several standardized frameworks exist worldwide—most notably the Federal Emergency Management Agency HAZUS methodology—practical post-disaster operations often require classification systems that are not only technically robust but also adaptable to varying decision-making contexts.

In this thesis, a novel multi-level damage state framework is proposed as a primary contribution to the literature. Unlike conventional single-scheme approaches, the proposed methodology introduces three hierarchical classification schemes (four-class, three-class, and two-class systems) derived from the same field-based damage observations. This structure enables flexible adaptation of damage categorization depending on operational, analytical, or policy-driven needs. The proposed framework is conceptually aligned with the damage progression logic of the HAZUS methodology but differs in its operational philosophy. While HAZUS is primarily designed for probabilistic loss estimation and regional-scale modeling, the proposed scheme is specifically structured to enhance machine learning-based predictive modeling and real-time decision-making applications. Table 2.2 presents a comparative mapping between the proposed framework and HAZUS damage states [3].

Table 2.2: Comparison Between Proposed Classification Scheme and HAZUS Scheme.

<table>
<tr><td>Damage State Schemes</td><td colspan="5">Damage States</td></tr>
<tr><td rowspan="2">HAZUS</td><td>NONE (DS0)</td><td>SLIGHT (DS1)</td><td>MODERATE (DS2)</td><td>EXTENSIVE (DS3)</td><td>COMPLETE (DS4)</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Proposed 4 Damage states</td><td>NO DAMAGE</td><td>LOW DAMAGE</td><td>MODERATE DAMAGE</td><td colspan="2">SEVERE DAMAGE</td></tr>
<tr><td>Proposed 3 Damage states</td><td>SAFE</td><td colspan="2">NEED TO BE CHECKED</td><td colspan="2">UNSAFE</td></tr>
<tr><td>Proposed 2 Damage states</td><td colspan="3">SAFE</td><td colspan="2">UNSAFE</td></tr>
</table>

The principal innovation of this framework lies in the following aspects:

- Hierarchical Flexibility

Instead of fixing the dataset to a single damage scale, the proposed system enables dynamic aggregation:

  - A detailed four-class model for high-resolution predictive modeling.
  - A three-class (traffic-light) system for operational prioritization.
  - A simplified binary classification for rapid emergency decisions.

This hierarchical structure allows the same dataset to be used across different modeling objectives without data restructuring or re-labeling.

- Machine Learning Optimization

Multi-class classification problems often suffer from class imbalance, overlapping feature distributions, and reduced predictive stability. By systematically aggregating damage states into three and two classes, this study enables:

  - Improved class separability.
  - Reduced misclassification between adjacent damage states.
  - Enhanced model robustness under imbalanced data conditions.
  - Comparative evaluation of model performance across different resolution levels.

This approach allows the investigation of how classification granularity influences predictive accuracy—an aspect rarely explored in existing earthquake damage literature.

- Operational Decision Alignment

The proposed classification system is directly mapped to actionable emergency decisions:

  - Structural intervention urgency
  - Re-occupation permissions
  - Evacuation prioritization
  - Resource allocation strategies

This ensures that model outputs are not purely academic but directly translatable into disaster management practice.

- Bridging Field Observations and Predictive Analytics

The framework preserves engineering-based field inspection logic while restructuring it into machine-learning-compatible formats. This reduces the conceptual gap between:

  - Qualitative engineering judgment
  - Quantitative predictive modeling

- *Four Damage Classes:*

The four-class system offers a detailed categorization of building conditions based on field inspections and engineering evaluations. It provides a comprehensive view of the extent of damage:

  - *Severe-Damaged Class (SD):* Buildings in this category have sustained extensive structural damage. Some may remain standing, while others may have partially or completely collapsed. These structures pose significant risks and typically require urgent intervention.
  - *Moderate-Damaged Class (MD):* Buildings showing moderate damage fall into this category. They may have structural impairments that need careful assessment before reoccupation is permitted.
  - *Low-Damaged Class (LD):* This category includes buildings with minor but visible damage, such as superficial cracks or small spalling. While these structures generally remain safe, they require inspections and minor repairs to ensure long-term functionality.
  - *Non-Damaged Class (ND):* Buildings classified as non-damaged did not exhibit any observable structural issues following the earthquake, indicating their resilience to the event.

- *Three Damage Classes (Traffic Light System):*

For practical disaster management and rapid decision-making, a three-level or "traffic light" classification is often applied:

  - *Unsafe:* Buildings considered severely damaged (SD) and requiring immediate attention.

- *Need To Be Checked:* Buildings in moderate- or low-damaged categories (MD or LD), which require detailed inspections before determining usability or reoccupation.
- *Safe:* Non-damaged buildings (ND) that can be safely occupied.

This system simplifies prioritization for emergency responders, helping allocate resources efficiently and identify buildings that need urgent inspection or evacuation.

- *Two Damage Classes:*

For rapid operational decisions, especially during large-scale disasters, a simplified binary classification is sometimes used:

- *Unsafe:* Buildings classified as severely or moderately damaged (SD or MD). Immediate evacuation or restriction of access is recommended.
- *Safe:* Buildings with low damage or no damage (LD or ND), which can generally be re-occupied with minimal risk.

This binary approach supports fast, on-the-ground decision-making, particularly in scenarios where quick evacuation or occupancy determinations are critical.

Overall, these damage classification schemes provide a structured framework for assessing post-earthquake building conditions. By translating complex structural assessments into actionable categories, they enable authorities, engineers, and emergency managers to make informed decisions efficiently, ensuring both public safety and effective disaster response.

2.5 Machine Learning Algorithms Utilization

This section describes the complete machine learning workflow adopted in this study, from data preparation to model assessment. Each step is designed to ensure data quality, reliable model training, and unbiased performance assessment. The procedures include data cleaning and preprocessing, dataset splitting, parameter tuning, and model validation strategies. Particular attention is given to reducing uncertainty, avoiding overfitting, and

preserving the integrity of post-earthquake damage data. Together, these steps establish a robust and transparent framework for developing reliable damage prediction models.

2.5.1 Data clearance and filtering

Data cleaning and filtering are critical stages in any machine learning workflow. Their main purpose is to remove issues in the dataset that could negatively affect how the model learns and performs. When the data is properly cleaned, the model is able to focus on meaningful, high-quality information, which reduces noise and leads to more reliable predictions. Skipping this step and training directly on raw data often results in problems such as overfitting, underfitting, or biased predictions toward certain classes.

Cleaning is also crucial for handling illogical or unrealistic values—such as negative ages or impossible elevation values—which can disrupt the learning process and harm the model's capacity to generalize. In this study, data cleaning was carried out both manually and automatically. For example, data points lying outside the Kahramanmaras region were manually removed to ensure a geographically focused dataset. Automatic filtering was also applied to detect and eliminate outliers, especially when a feature fell outside the acceptable range of the dataset.

There are several techniques for identifying and removing outliers, including Z-score analysis, predefined thresholds, and duplicate removal. These methods help maintain a consistent and trustworthy feature space for the model.

In addition to outlier removal, handling missing data is another important part of the cleaning process. Missing values were imputed to ensure that each data point had complete information across all features. Data filtering also plays a key role: irrelevant or contradictory data points were removed so the model could better concentrate on the specific problem being studied. This targeted approach helps reduce uncertainty and improve the model's generalization within the defined scope of the task.

2.5.2 Data preprocessing

Data preprocessing is a fundamental step in improving a model's performance. Feeding raw data directly into a model can lead to errors, unrealistic outputs, and overall unreliable results.

Once the data is cleaned, preprocessing steps follow. In many cases, the data may need to be scaled using techniques such as Min–Max Scaling, Max–Abs Scaling, or Standard Scaling. The choice of scaling method depends on the characteristics and requirements of the dataset. Scaling transforms the features to a consistent range, preventing any variable with larger numerical values from dominating the learning process.

Another important preprocessing step is encoding categorical data. Since many features may appear as text—such as labels, categories, or descriptions—they must be converted into numerical values, as most machine learning models cannot interpret strings. Encoding can be performed using techniques like label encoding or one-hot encoding, which help convert categorical variables into model-friendly numerical formats.

2.5.3 Data splitting

To develop the model effectively, the dataset was partitioned into training, validation, and test subsets. Generally, 80–90% of the data was used for model development (training and validation), while the remaining portion was reserved for final performance assessment. Stratified sampling was applied to maintain the original class distribution within each subset, which is particularly important for imbalanced datasets to prevent biased learning. Allocating a larger share of data to training allows the model to learn more patterns, whereas the validation set provides an unbiased measure of performance during development.

An independent validation set, evaluated after training is completed, offers a realistic estimate of how the model is expected to perform in practical applications. In this framework, the training data were used to fit the model parameters, and the test data were employed exclusively for final evaluation to avoid information leakage.

To enhance robustness and minimize overfitting, 10-fold cross-validation was implemented. In this approach, the training dataset is randomly divided into ten equal

subsets. Each subset is used once as a temporary validation set, while the remaining nine subsets are used for training. This procedure ensures that every observation contributes to both training and validation, resulting in a more reliable and stable estimate of model performance.

2.5.4 Model parameters

Model parameters play a key role in shaping how a model behaves during training. They set boundaries on the learning process, influencing how far the model can generalize, how much computational power it requires, and how quickly it should learn. Each parameter affects the model in a specific way, and because most models have several parameters—each with multiple possible values—tuning them becomes an essential step in achieving good performance.

Manually adjusting parameters can be extremely time-consuming and often inefficient. Just as cross-validation helps evaluate model performance across different data splits, similar systematic approaches are used to search for the best parameter combinations. Common techniques include:

- *Grid Search:* Tries every possible combination of predefined parameter values, ensuring thorough but computationally expensive exploration.
- *Random Search:* Samples random combinations from predefined ranges, often finding good results faster than grid search.
- *Bayesian Optimization:* A more intelligent method that learns from previous trials and predicts which parameter sets are likely to perform better next, reducing unnecessary evaluations.

In many cases, certain parameters only work well when paired with specific values of other parameters, adding another layer of complexity. This makes automated parameter tuning even more valuable, helping identify configurations that manual tuning might easily miss.

2.5.5 K-folds and cross-validation

K-fold cross-validation is a technique designed to provide a reliable and unbiased assessment of model performance. As illustrated in Figure 2.9, the available dataset is randomly divided into *k* equal-sized subsets, commonly referred to as folds. In each iteration, the model is trained using *k–1* folds, while the remaining fold is used for validation. This procedure is repeated *k* times so that every fold serves once as the validation set. The overall model performance is then calculated by averaging the results obtained from all folds, thereby reducing the influence of any single data split. Consequently, k-fold cross-validation enhances the stability and generalizability of performance estimates, especially when dealing with limited post-earthquake damage datasets.

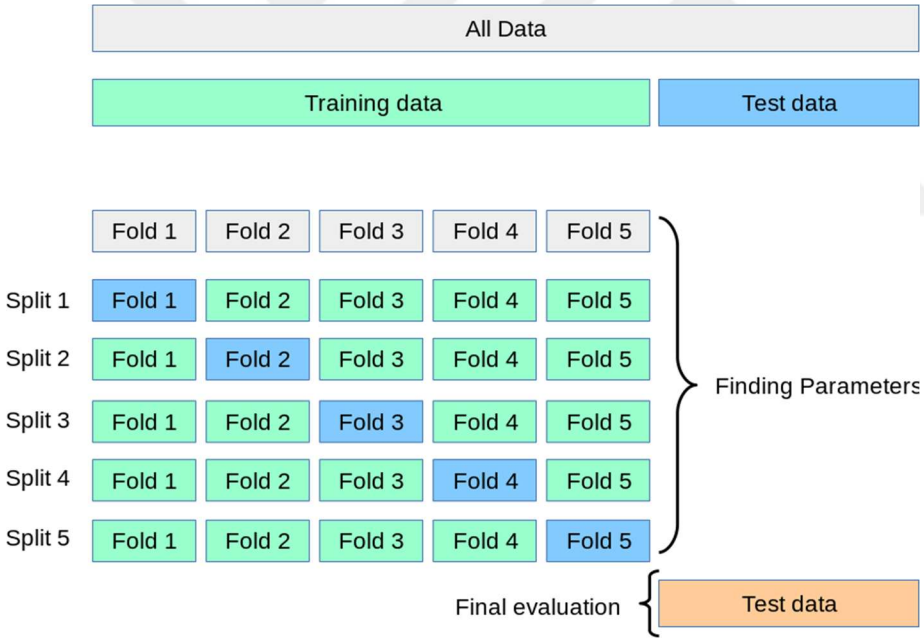

Figure 2.9: K-folds Procedure [39].

2.5.6 Model training

Model training is arguably the most significant stage in the entire machine learning workflow. In unsupervised learning has no labels at all—the model must discover hidden structures or patterns on its own.

Training a machine learning model typically involves feeding it input features, letting it generate an initial prediction using randomly initialized parameters, and then comparing that prediction to the true value. A loss function measures how far off the prediction is. The model then adjusts its parameters—usually through optimization techniques like gradient descent—to reduce this error. This cycle repeats over many iterations, known as epochs, until the model's performance improves and becomes stable.

Depending on the dataset size and the model level of complexity, training can require substantial time and computational resources. Large models perform millions of calculations across multiple iterations, which can be demanding even on modern hardware. Techniques such as efficient data processing, careful model tuning, early stopping, or transfer learning can help make the process faster and more efficient.

2.5.7 Model validation

Model validation represents a fundamental phase in machine learning, as it evaluates the ability of a trained model to perform accurately on new, previously unseen data. This process helps confirm that the model is not merely memorizing the training dataset—an issue known as overfitting—but is capable of producing reliable predictions in real-world scenarios.

Typically, validation is conducted using a separate dataset that was not involved in the training process. The selection of evaluation metrics depends on the nature of the problem. For classification tasks, commonly used metrics include accuracy, precision, recall, F1-score, and confusion matrices. In contrast, regression problems are generally assessed using indicators such as mean squared error (MSE). Selecting appropriate performance metrics is essential to ensure that the evaluation accurately represents the model's effectiveness for the specific application.

During the validation process, various aspects of the model can be adjusted to improve performance. This includes tuning hyperparameters, choosing the most relevant input features, or modifying the model architecture. Systematic approaches like grid search, random search, or Bayesian optimization help in an efficient manner to identify the best parameter settings. It is also essential to avoid data leaks, which happen when external information from outside the training dataset unintentionally affects the model, leading to

very optimistic performance estimates. Finally, after a model demonstrates strong performance on the validation set, it should be evaluated on a separate test set or in real-world conditions—such as through A/B testing or shadow deployment—to confirm its reliability before full-scale implementation.

2.5.8 Choosing the best model performance

Selecting the right machine learning model is one of the most critical steps in building a reliable prediction system. The goal is to find a model that can perform well while training on the dataset and at the same time generalize well to unseen data, to ensure dependable predictions when applied in real-world scenarios.

The definition of the "best" model depends heavily on the type of problem being solved—whether it's regression, ranking, or classification—since each task relies on different evaluation metrics. For classification problems in particular, metrics such as accuracy, precision, recall, and F1-score help reveal different aspects of model performance. Examining learning curves can also provide valuable insights, such as whether the model is overfitting, underfitting, or achieving appropriate generalization.

Although selecting the model witch exhibit highest accuracy or F1-score may seem like the obvious choice, model evaluation often requires a deeper look. Confusion matrices, for example, can highlight how well the model separates classes, revealing misclassifications that simple metrics might overlook. To determine the most suitable train–test split, the configuration that produced the lowest degree of class confusion across all folds was selected, ensuring clearer decision boundaries and stronger generalization.

In the end, the "best" model is not just the one with the top score—it is the one that strikes the right balance between strong performance and practical suitability for the specific task at hand.

2.5.9 Feature importance analysis

In machine learning, feature importance analysis is used to identify the input variables that have the greatest impact on model predictions. This technique helps clarify how different factors influence the output, thereby improving the transparency and

interpretability of the model. In engineering and structural applications, for example, feature importance can reveal which material properties, geometric characteristics, or structural parameters most significantly affect performance or damage outcomes.

Besides improving interpretability, feature importance can guide the model development process. By highlighting the most influential variables, it can help simplify the model without sacrificing accuracy and focus attention on the factors that truly matter. It also provides valuable understanding of the primary patterns in the dataset, supporting theoretical assumptions and informing practical decisions for design, monitoring, or disaster mitigation. In this way, feature importance analysis not only strengthens the reliability of machine learning models but also connects data-driven results to real-world engineering understanding.

2.6 Feature Selection and Justification

The dataset utilized in this study comprises a diverse set of features that are grouped into three primary categories: structural attributes, ground-motion parameters, and site-related conditions. Collectively, these variables offer a comprehensive representation of each building and the seismic environment to which it was subjected during the earthquake.

- Structural Characteristics

The first category includes variables that represent the physical properties of each building. Building age is a particularly significant factor, as older structures may exhibit higher seismic vulnerability due to material deterioration, previous construction practices, and the lack of contemporary earthquake-resistant design standards. The number of floors is another important parameter, since structural response to seismic loading varies with building height. Multi-story buildings are typically more affected by ground motions with longer dominant periods and may experience greater dynamic amplification, especially under high Peak Ground Acceleration (PGA) levels. Building footprint areas were originally determined using data from OpenStreetMap by linking each building's geographic coordinates with its corresponding mapped boundary.

- Ground-Motion Parameters

Ground-motion variables describe the intensity and characteristics of seismic shaking at each building location. Peak Ground Acceleration (PGA) represents the highest recorded acceleration during the earthquake and is especially critical for the response of low-rise structures. Peak Ground Velocity (PGV) provides complementary information related to the energy content of ground motion and has a stronger influence on structures with lower natural frequencies. Seismic intensity offers a broader, qualitative indication of shaking severity, typically based on observed impacts on buildings and the surrounding environment.

Another key parameter is the rupture distance (Rrup), defined as the shortest distance between the building site and the fault rupture. This distance plays a major role in determining the level of seismic demand, as locations closer to the fault generally experience stronger shaking and higher-frequency ground motions. All ground-motion parameters in this study were produced using REDAS software through a scenario-based simulation representing the February 7, 2023, earthquake event.

- Site-Specific Conditions

The third group consists of features related to the local soil and terrain. VS30, which is the average shear-wave velocity in the upper 30 meters of the soil profile, is one of the most widely used indicators of site stiffness. Stiffer soils with higher VS30 values can significantly influence how ground motions are amplified or dampened. Elevation was also included, as variations in topography can affect how seismic waves travel and can lead to localized amplification or attenuation of shaking.

2.7 Performance Evaluation

The evaluation of the machine learning models developed in this thesis is carried out using two primary approaches, each aligned with a different scale of assessment. These methods deliver a deep understanding of how the models perform both at the individual building level and at a broader, city- or district-wide scale. By examining performance across multiple levels, the study ensures that the model is not only accurate in predicting isolated

cases but also reliable when applied to large, real-world datasets used in regional damage assessment.

2.7.1 Performance for building scale

The building-level evaluation examines the model's predictive performance for individual structures within the dataset. This assessment is conducted using a confusion matrix, which presents the model's capability to correctly classify different damage categories. Based on this matrix, essential evaluation metrics such as accuracy, recall, precision, and F1-score are derived. These indicators offer a comprehensive view of the model's performance, demonstrating not only its overall correctness but also its effectiveness in identifying each specific damage class. Evaluating performance at the building scale is crucial to ensure consistent and reliable predictions at the most detailed level, where accurate results are particularly important for informed decision-making.

- *Confusion matrices*

In machine learning, assessing the predictive model's performance is essential to understanding how reliably it can classify new, unseen data. A confusion matrix is considered one of these tools, which provides a clear and organized summary of the comparison to the actual labels. By breaking down the results into distinct categories, the confusion matrix allows researchers to pinpoint not only how often the model is correct, but also the specific types of mistakes it makes.

A standard confusion matrix, which can be shown in Figure 2.10, is built around four key components:

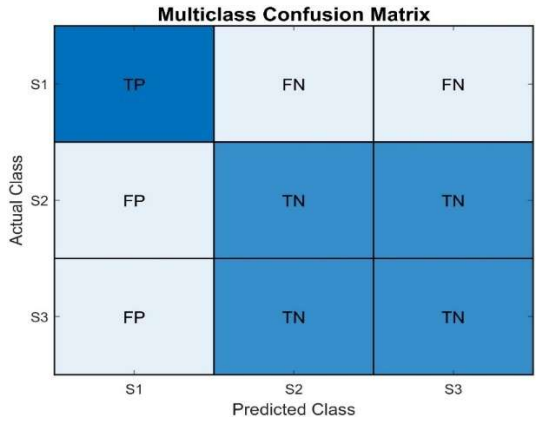

Figure 2.10: Multiclass Confusion Matrix Example.

- True Positives (TP): Cases where the model correctly identifies the presence of a particular class or condition.
- True Negatives (TN): Cases where the model correctly predicts that a class or condition is absent.
- False Positives (FP) – Type I Error: Instances where the model predicts a class or condition that is not actually present.
- False Negatives (FN) – Type II Error: Instances where the model fails to detect a class or condition that truly exists.

These four values form the foundation for calculating several important performance metrics that describe different aspects of model behavior. While the confusion matrix gives an overview, the derived metrics—such as accuracy, precision, recall, and F1-score—provide deeper insight into the quality and reliability of the model's predictions. For example, a model may show high accuracy overall, but if it has low recall, it means it is missing many positive cases, which could be critical in applications such as damage detection or medical diagnosis.

- *Accuracy*

Accuracy measures the proportion of correctly predicted instances—both positive and negative—relative to the total number of cases. While it gives an overall indication of model performance, it can be misleading when the dataset is highly imbalanced.

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{2.1}$$

- *Precision*

Precision quantifies the proportion of the model's correct positive predictions. A high-precision model produces very few false positives. This metric is particularly critical in scenarios where false alarms are costly.

$$\frac{TP}{TP + FP} \tag{2.2}$$

- *Recall*

Recall—also referred to as sensitivity or the true positive rate—assesses a model's ability to correctly identify actual positive cases. A high recall indicates that the model successfully captures most true positive instances, minimizing the likelihood of overlooking important cases.

$$\frac{TP}{TP + FN} \tag{2.3}$$

- *F1-Score*

The F1-score combines precision and recall into a single metric using their harmonic mean. It is particularly useful for imbalanced datasets, where accuracy alone can be misleading. A strong F1-score requires the model to achieve a good balance between correctly identifying positive cases and avoiding false positives.

$$2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2.4}$$

The confusion matrix and these evaluation metrics provide a comprehensive framework for assessing the performance of classification models. They help identify strengths, reveal weaknesses, and guide improvements, ultimately ensuring that the model is dependable and effective for real-world applications.

2.7.2 Performance for the regional scale

For larger-scale evaluations—such as at the district or city-wide level—the model's performance is assessed by making a comparison of the actual number of buildings in each damage state with the predicted number of buildings in those same categories. Instead of focusing on individual buildings, this approach looks at how well the model captures overall damage patterns across broader geographic areas.

This large-scale comparison is particularly important for disaster management, as it helps determine whether the model can correctly identify the most heavily affected districts or neighborhoods after an earthquake. By analyzing how closely the predicted counts match the observed damage distribution, researchers can evaluate the model's ability to give reliable insights for regional planning, resource allocation, and emergency response.

In simple terms, while building-level evaluation checks if the model is correct for every single structure, district- or city-level evaluation checks if the model understands the big picture of which areas were impacted the most. This helps ensure that the model is not only accurate on a small scale but also useful for practical, real-world decision-making during large earthquake events.

3. STUDY CASES

This chapter presents the real-world earthquake events used to evaluate the proposed Machine learning–based framework for post-earthquake building damage evaluation. Two major seismic events in Turkiye—the 2020 Izmir earthquake and the 2023 Kahramanmaras earthquake sequence—are selected as representative case studies due to their contrasting seismic characteristics, damage patterns, and spatial scales. These events provide diverse structural, geotechnical, and ground-motion conditions for testing model robustness and generalization. For each case, the seismic context, observed damage distribution, and assembled datasets are systematically introduced. The use of multiple case studies allows an in-depth evaluation of model performance across different building types and damage classification schemes.

3.1 Izmir Earthquake

This section presents the seismic event, study area, and dataset used to develop and validate the proposed machine learning framework for post-earthquake damage assessment. It first describes the characteristics and impacts of the 30 October 2020 Izmir earthquake, followed by an overview of the spatial distribution of damage. Finally, the composition, sources, and statistical characteristics of the dataset are introduced to establish the foundation for subsequent modeling and analysis. The workflow can be summarized in (Figure 3.1).

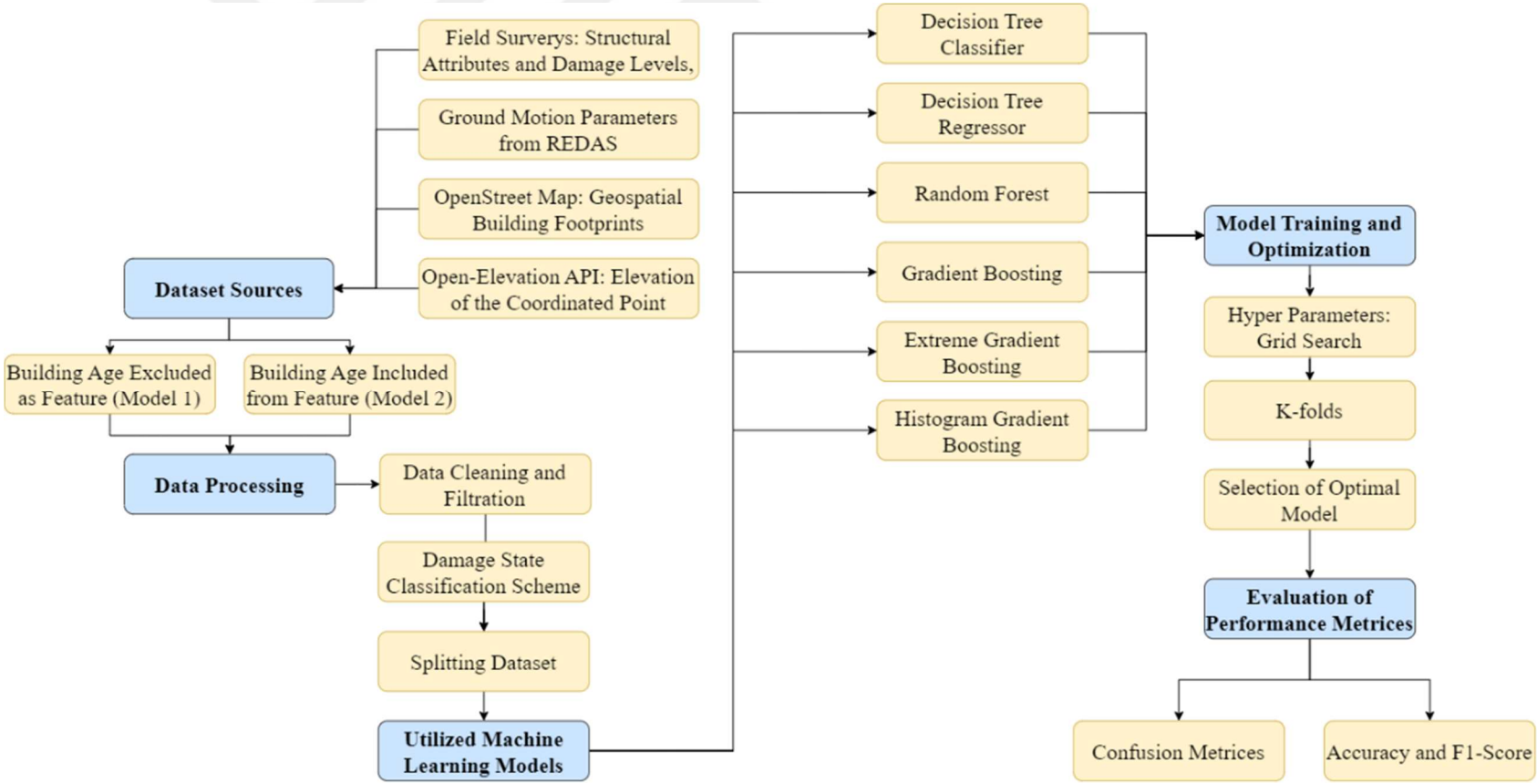

Figure 3.1: Izmir Study Case, Methodology Workflow.

3.1.1 Location and seismic event

A strong earthquake struck the Aegean region of Turkiye on 30 October 2020, with its epicenter located offshore in the Aegean Sea, north of the Greek island of Samos and west of the İzmir metropolitan area. The mainshock had a moment magnitude of Mw 6.9 and occurred at a shallow depth, resulting in intense ground shaking along the Izmir coastline. The earthquake was associated with normal faulting within the Aegean extensional tectonic regime and did not produce a clear surface rupture on land, as the fault rupture was primarily submarine.

The impacts of the earthquake were concentrated in the city of Izmir, particularly in the districts of Bayraklı and Bornova, where numerous mid-rise reinforced concrete residential buildings suffered severe damage or collapse. Although the affected area was considerably smaller than that of large continental earthquakes, the event caused disproportionate structural losses due to local site effects, including soft soil conditions that amplified ground motion. The earthquake resulted in significant human and economic losses, including fatalities, injuries, widespread building damage, and disruption to urban infrastructure. In addition, localized coastal flooding and minor liquefaction were observed along the Izmir Bay, further emphasizing the susceptibility of the built environment in the region [40–42].

3.1.2 The distribution of damage in Izmir

The 30 October 2020 earthquake occurred offshore in the eastern Aegean Sea, west of İzmir, but its shallow depth and proximity resulted in severe damage within the metropolitan area. Although the epicenter was located offshore, strong ground shaking was concentrated in specific districts due to local site effects. The Bayraklı and Bornova districts experienced the most significant damage, as they consist of dense residential developments built predominantly on soft alluvial soils. Many mid-rise reinforced concrete buildings in these areas collapsed or were heavily damaged, most of which were constructed prior to the implementation of modern seismic design codes. The earthquake caused more than 110 fatalities, largely concentrated in these heavily affected districts.

From an engineering perspective, the earthquake exposed several structural vulnerabilities. Numerous collapsed buildings lacked adequate seismic detailing, including insufficient transverse reinforcement and limited ductile capacity. Soft-story configurations, commonly created by open ground floors used for commercial purposes, were a dominant failure mechanism. In addition, strong soil amplification effects significantly increased ground motion intensity in Bayraklı, contributing to the uneven spatial distribution of damage. Construction quality deficiencies, such as low-strength concrete and poor reinforcement detailing, were frequently observed in older buildings. Furthermore, many severely damaged structures were 5–8 stories tall, corresponding to the dominant vibration periods of the ground motion, which amplified structural demands through resonance effects [43,44].

Overall, the 2020 Izmir earthquake highlights the crucial role of local soil conditions, outdated construction practices, and structural irregularities in driving urban earthquake losses. The event emphasized the need for systematic seismic assessment and retrofitting of existing buildings, particularly in areas with unfavorable site conditions, to enhance urban seismic resilience and reduce future losses.

3.1.3 Dataset

The primary dataset used in this study is based on real post-earthquake field data collected after the 30 October 2020 Izmir earthquake in Turkiye. These data were obtained through comprehensive on-site inspections and form the core foundation of the proposed machine learning framework. To enhance the completeness and reliability of the dataset, the field observations were further enriched by integrating information from two external open-source platforms: building footprint data extracted from OpenStreetMap (OSM), which were used to calculate building areas, and elevation data referenced to sea level, obtained through the Open-Elevation API. The dataset is represented in Figure 3.2.

The buildings included in the dataset are representative of the typical reinforced concrete building stock found across the wider Aegean region of Turkiye. They share similar structural characteristics, such as construction materials, building typologies, and floor configurations. In addition, these buildings are exposed to comparable geotechnical and seismological conditions, including soil classifications, elevation profiles, and regional

seismicity levels. This level of consistency enhances the representativeness of the dataset and supports the potential transferability of the developed model to other urban areas within the Aegean region that are subject to similar seismic hazards.

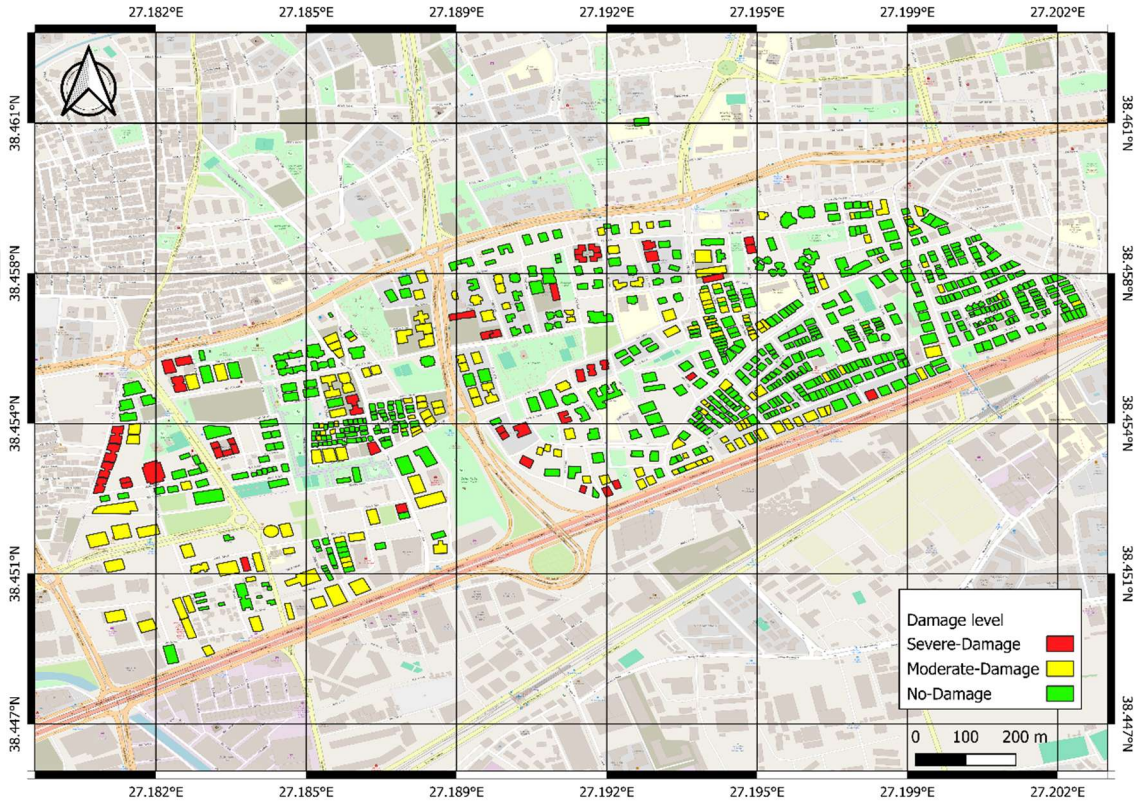

Figure 3.2: Izmir Study Case Building Distribution by Damage Class.

3.1.3.1 Collected data

Approximately 900 building records were collected through field surveys. These records include essential structural and spatial attributes such as building location, number of stories, year of construction, and observed damage severity. The damage assessments were carried out through visual inspections by highly qualified engineering teams from the Ministry of Environment, Urbanization, and Climate Change of Turkiye. The classification of damage states was based on professional engineering judgment following nationally adopted post-earthquake assessment procedures, ensuring a high level of reliability and consistency in the recorded observations.

3.1.3.2 Generated data

Seismic and ground-motion characteristics for each building were generated using REDAS (Rapid Earthquake Damage Assessment System) software. The parameters extracted include Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV), seismic intensity, and rupture distance. These features were linked to each building's location and incorporated into the dataset as essential input variables for model training and evaluation. Including these parameters enables the machine learning models to account for the spatial variability of earthquake effects and their impact on observed building damage.

3.1.3.3 Online available databases

Open-source geospatial datasets have become essential tools for addressing complex real-world problems, particularly in disaster risk assessment and urban analysis. One of the most prominent platforms is OpenStreetMap (OSM), a global, community-driven project that provides freely accessible and continuously updated cartographic data. OSM offers detailed information on buildings, roads, transportation networks, and public infrastructure, compiled from ground surveys, GPS traces, aerial imagery, and other open-access sources. Due to its richness and global coverage, OSM has been widely used in domains like urban planning, navigation, disaster response, and environmental monitoring.

In this study, the geographic coordinates of the surveyed buildings were used to extract their corresponding footprints from OSM. Geographic Information System (GIS) tools were then employed to compute the plan area of each building, which serves as an important structural descriptor in damage assessment and vulnerability modeling.

Additionally, elevation data for each building location was retrieved using the Open-Elevation API, which provides elevation values above mean sea level for any given coordinate. Elevation is a particularly relevant factor in seismic risk analysis, as it can influence local site effects, wave propagation characteristics, and overall hazard exposure. The integration of elevation data further enhances the spatial resolution and physical relevance of the dataset.

Collectively, the combination of field-based inspection data, seismic parameters, and open-source geospatial information provides a robust, scalable, and cost-effective dataset. This integrated data framework supports reliable machine learning model development and strengthens the overall credibility of the proposed approach for post-earthquake damage assessment and decision-making.

3.1.4 Data distribution

The dataset used in this case study comprises approximately 900 reinforced concrete (RC) buildings that were inspected and assessed following the earthquake. The distribution of observed damage states shows a clear imbalance: about 7% of the buildings were classified as severely damaged, 4% as moderately damaged, 17% as lightly damaged, and the remaining 72% were categorized as undamaged, which can be illstrated in (figure 3.3) This distribution reflects the overall damage pattern observed in the affected region and highlights the challenge of modeling imbalanced post-earthquake datasets.

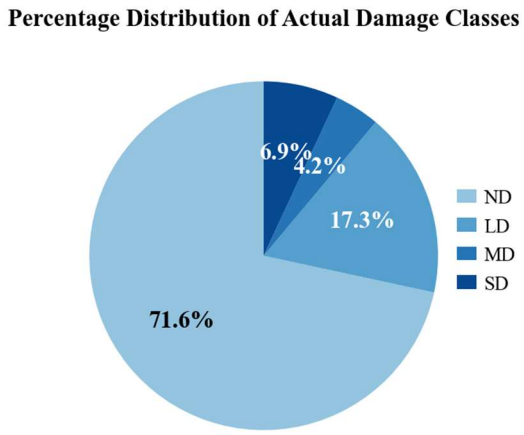

Figure 3.3: Izmir Study Case Actual Damage Distribution.

The surveyed buildings are located at distances ranging approximately from 50 to 70 km (figure 3.5) from the fault rupture. The elevation of the building sites varies from near sea level up to approximately 170 m above mean sea level, capturing moderate topographic variability across the study area.

In terms of seismic characteristics, the Peak Ground Acceleration (PGA) values associated with the buildings range from approximately 96 gal to 111 gal (Figure 3.4), while the Peak Ground Velocity (PGV) values vary between1 cm/s and 17.23 cm/s. The corresponding seismic intensity values span from 6.30 to 6.70 (Figure 3.6), indicating moderate levels of ground shaking across the dataset. Additionally, the site conditions, represented by the VS30 parameter, range from 261 m/s to 700 m/s, covering a broad spectrum of soil stiffness classes.

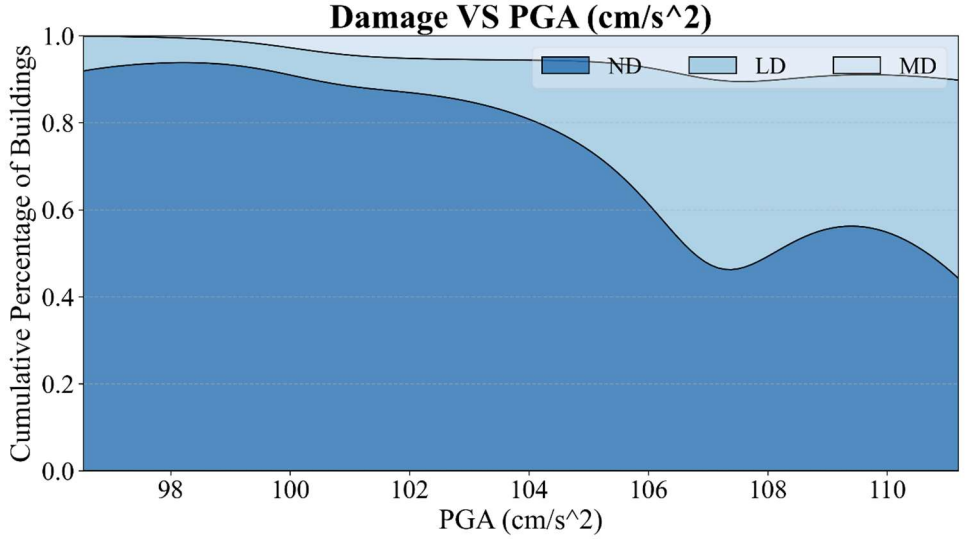

Figure 3.4: Statistical Distribution of Izmir Study Cases Structures Damage. States by PGA.

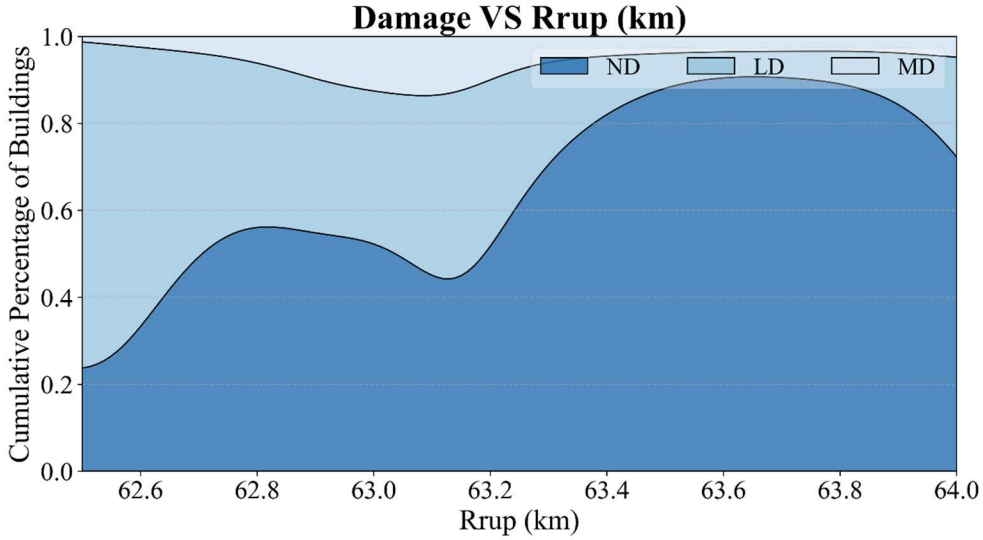

Figure 3.5: Statistical Distribution of Izmir Study Cases Structures Damage States by Rupture Distance.

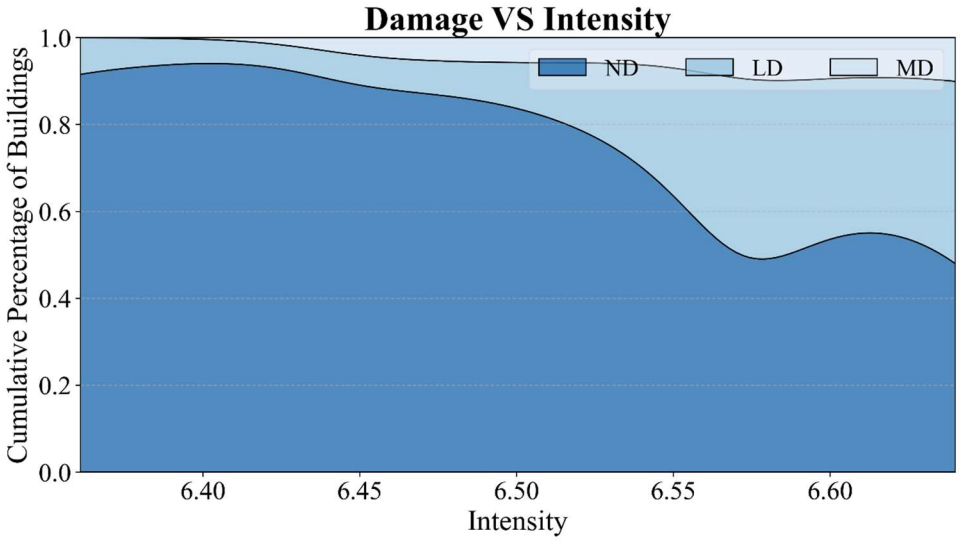

Figure 3.6: Statistical Distribution of Izmir Study Cases Structures Damage States by Intensity.

Overall, the dataset represents a diverse and realistic spectrum of structural damage states, seismic demand indicators, and site-specific conditions, forming a solid foundation for the development, training, and evaluation of machine learning models for post-earthquake damage classification.

3.1.5 Feature set configuration

To examine the influence of different feature combinations on model performance, the dataset used in this section was divided into two configurations. The first configuration, referred to as Model 1, includes a set of building-related attributes—namely footprint area, floor class, elevation, and VS30—together with key seismic intensity measures, including peak ground acceleration (PGA), peak ground velocity (PGV), macro seismic intensity, the spectral acceleration at 1 second (S01), and rupture distance.

The second configuration, Model 2, builds upon Model 1 by incorporating the age of the building construction as an additional feature. This extension allows for a direct evaluation of the role of aging and potential material degradation in influencing building damage and improving prediction accuracy.

To standardize structural information and reduce variability, building height was classified into three floor-based categories: Class A (1–4 floors), Class B (5–8 floors), and Class C (9–12 floors). This floor-based categorization represents the only preprocessing

step applied to the dataset, ensuring that the original characteristics of the data were largely preserved while maintaining consistency across the analyzed models.

3.1.6 Machine learning algorithms

This subsection describes the complete machine learning workflow adopted in this study, from data preparation to model evaluation. A structured and sequential approach was followed to ensure reliable training, unbiased validation, and meaningful performance comparison. Each step was designed to enhance model robustness and suitability for post-earthquake damage classification tasks.

3.1.6.1 Data cleaning and filtering

Before model implementation, the dataset underwent a rigorous cleaning process to remove null values, corrupted entries, and typographical errors. Feature encoding was applied where necessary to align with the research objectives. Filtering was subsequently performed on the entire dataset to eliminate outliers falling outside the 95% normal distribution, as well as any illogical data points. This stage was critical to mitigating noise, preventing algorithmic bias, and ensuring that anomalous values did not compromise the model's accuracy or generalization capabilities.

3.1.6.2 Data splitting

Following the validation of the dataset through cleaning and filtering, the data was partitioned into three distinct subsets: a training set (80%), a testing set (10%) for performance tuning and initial evaluation, and a final evaluation set (10%). This final subset consisted of entirely unseen data, which remained untouched during the training and testing phases. This approach is vital for providing an objective evaluation of the model's generalization potential and its readiness for deployment in real-world scenarios.

3.1.6.3 Tuning ML model hyperparameters

The behavior, sensitivity, and computational efficiency of machine learning algorithms are fundamentally governed by their hyperparameters. To achieve optimal performance,

a systematic grid-search method was utilized to identify configurations superior to default settings. This process involved evaluating all predefined hyperparameter combinations and selecting the optimal setup based on specific performance metrics. To prevent information leakage and minimize overfitting, this tuning was conducted using the validation subset, ensuring the model's robustness and accuracy on unseen data.

3.1.6.4 K-folds and cross-validation

To ensure the development of an unbiased model and determine the ideal data split, K-fold cross-validation was implemented. While the initial split reserved 10% for independent testing, a Stratified 20-fold cross-validation was applied specifically to the training set. This high number of folds was selected to address the dataset's significant imbalance, maximizing the representation of minority classes in each fold and reducing estimation bias. The final model was selected based on its ability to maintain high accuracy across the testing dataset.

3.1.6.5 Fitting of training data

Upon completion of hyperparameter tuning and cross-validation, the model was trained using the designated training split. During this phase, the algorithm identified underlying patterns within the data based on the optimized parameter configurations. This process resulted in a finalized model ready for comprehensive performance evaluation and subsequent analysis.

3.1.6.6 Fitting of validation data

A segment of the data was kept as a validation set to evaluate the model's performance independently of the training phase. This subset serves as a substitution for real-world applications, allowing for an unbiased evaluation of how the model handles unseen data. This step is essential for establishing confidence in the model’s generalizability and the reliability of the final results.

3.1.6.7 Comparison of utilized models

To compare the performance of the six models, accuracy and F1-score were selected as the main evaluation metrics. Accuracy indicates the proportion of correctly classified instances; however, it may provide an overly optimistic view when the dataset is imbalanced. For this reason, the F1-score was also adopted to ensure a more balanced evaluation by combining precision and recall into a single measure. Using both indicators allowed the models to be assessed not only in terms of overall correctness but also in their ability to correctly identify minority classes, leading to the selection of the model that demonstrates the best overall balance across all damage categories.

3.2 Kahramanmaras Earthquake

This section provides an overview of the February 6, 2023, Kahramanmaras earthquake sequence and its severe impact on the built environment in southern Turkiye. It describes the spatial distribution and underlying causes of observed building damage, with a particular focus on structural and site-related vulnerabilities. The section also introduces the datasets used in this study, detailing their sources, composition, and relevance for machine-learning-based damage prediction, where the workflow can be summarized in (Figure 3.7).

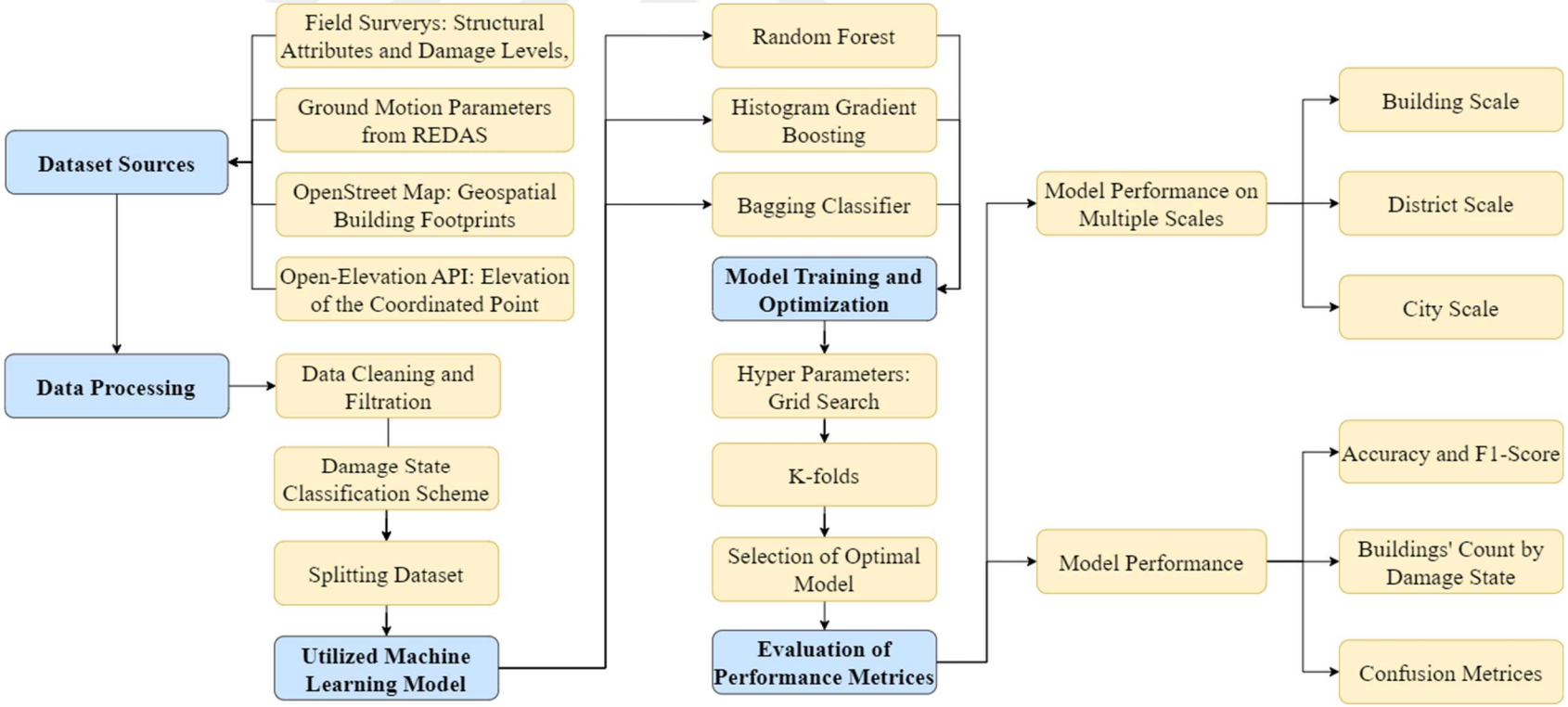

Figure 3.7: Kahramanmaras Study Case, Methodology Workflow.

3.2.1 Location and seismic event

A series of powerful earthquakes struck the southern region of Turkiye on 6 February 2023, with the main epicenter located in the Pazarcık district of Kahramanmaras Province. The first event, with a magnitude of Mw 7.8, was followed about nine hours later by a second major earthquake of approximately Mw 7.6 near Elbistan in the same province. These earthquakes ruptured several segments of the East Anatolian Fault Zone, producing a surface rupture extending for more than 500 km. The impacts were widespread, affecting an area of roughly 350,000 km² across 11 provinces, including Kahramanmaras, Hatay, Gaziantep, Adana, Adıyaman, Diyarbakır, Malatya, Kilis, Osmaniye, Elazığ, and Şanlıurfa. This region, home to nearly 14 million people, experienced severe structural damage, extensive infrastructure disruption, and various geotechnical hazards such as landslides and liquefaction.

3.2.2 The distribution of damage in Kahramanmaras

The two earthquakes struck near the city of Kahramanmaras, and due to the proximity of the epicenters to the urban core, the city suffered extensive damage. Satellite-based Synthetic Aperture Radar (SAR) change-detection analyses indicated that roughly 17.37% of the urban area was destroyed. The most severely affected districts were Dulkadiroğlu and Onikişubat, which include densely populated residential and commercial neighborhoods. In these areas, numerous buildings either collapsed or experienced severe, non-repairable damage, particularly older structures constructed prior to the enforcement of updated seismic design regulations. Although detailed building-specific casualty data are limited, the wider Kahramanmaras province reported more than 12,600 deaths, with a considerable portion likely concentrated in the city center [45,46].
From an engineering standpoint, the earthquakes revealed multiple critical structural weaknesses:

1. Insufficient Seismic Design Provisions: A significant number of the collapsed buildings did not incorporate essential earthquake-resistant design features, such as adequate ductile detailing, proper confinement of columns, and structural

systems like shear walls that are intended to enhance energy dissipation and improve seismic performance.

2. Soft-Story Vulnerability: A large portion of mid-rise structures contained soft-story configurations, often created by open ground floors used for retail or parking. These weak stories performed poorly under strong lateral forces, frequently leading to partial or total collapse.
3. Soil Amplification Effects: Much of central Kahramanmaras is situated on soft alluvial soil, which is known to amplify ground shaking. Ground deformation mapping revealed an average displacement of 0.46 meters, which contributes to the abundance of structural failures in the city.
4. A considerable portion of the damaged buildings had been constructed prior to the 2000s and therefore did not comply with contemporary seismic design regulations. Typical deficiencies observed in these structures included the use of low-strength concrete, insufficient transverse reinforcement, and improper reinforcement lap splicing.
5. Resonance with Ground Motions: A substantial number of the affected buildings were between 5 and 8 stories tall, which corresponded to the dominant frequency range of the recorded earthquakes. This resonance phenomenon increased lateral displacement demands and caused many structures to exceed their intended design limits.

Overall, the 2023 Kahramanmaras earthquakes highlighted the critical importance of enhancing seismic resilience in vulnerable regions. The extensive damage demonstrated the impacts of outdated construction practices and reinforced the necessity of integrating ductility, structural redundancy, and dynamic response considerations into both new building designs and the retrofitting of existing structures. Improving the resilience of the built environment is therefore essential to mitigate potential losses in future earthquakes.

3.2.3 Dataset

The dataset forms the backbone of any machine learning model, and the quality of the final predictions depends heavily on the quality and completeness of the data used. In this

study, the dataset was carefully assembled by integrating information from several different databases. This merging process ensured that the final dataset was both comprehensive and diverse, capturing a wide range of structural and environmental characteristics.

The compiled dataset represents buildings across the entire Kahramanmaras region, one of the areas most severely impacted by the devastating earthquakes of February 6, 2023. By bringing together data from multiple sources, the dataset provides a detailed and realistic representation of the building stock, enabling the machine learning models to learn meaningful patterns related to earthquake-induced damage.

Overall, this integrated dataset serves as a strong foundation for developing reliable and generalizable predictive models, ensuring that the analysis reflects real-world conditions and supports effective disaster management strategies.

3.2.3.1 Collected data

The field data collected from the affected region forms the foundation of the dataset used in this thesis. This information was gathered across the Kahramanmaras province in the southern region of Turkiye, one of the areas most severely impacted by the 2023 earthquakes. The dataset includes essential building characteristics such as geographic coordinates, structural type, construction age, and number of floors. Most importantly, it contains the damage states assigned by qualified engineers through systematic visual inspections and professional engineering judgment.

The recorded building points are distributed across a wide area with an approximate radius of 28 km, covering both the urban core and surrounding settlements. The farthest surveyed point lies roughly 39 km from the fault line, capturing a broad range of shaking intensities and damage patterns. This spatial diversity strengthens the dataset, ensuring that the resulting analysis reflects real on-site conditions and the varying structural performance observed throughout the region.

3.2.3.2 Generated data

Ground-motion parameters used in this study were derived from scenario analyses designed to represent the seismic events that occurred on February 7, 2023. To model these scenarios, the REDAS software was employed to simulate the earthquake rupture process and estimate the corresponding magnitude. This approach allowed for the generation of realistic ground-motion characteristics that closely reflect the conditions experienced during the actual event.

The resulting dataset includes several key seismic parameters: Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV), seismic intensity measures, the shear-wave velocity of the upper 30 meters of the soil profile (VS30), and the rupture distance (Rrup). Together, these parameters provide a detailed description of the shaking environment and local site effects, offering valuable inputs for understanding building performance under the simulated earthquake scenario.

3.2.3.3 Online available databases

Publicly accessible online databases were also incorporated to enrich and expand the dataset. Platforms such as OpenStreetMap were used to extract building footprints, allowing the calculation of additional features—such as building area—based on precise geospatial coordinates. These geospatial attributes help create a more detailed representation of each structure and its surrounding context.

In addition, the Open Elevation API was utilized to obtain the ground elevation above sea level for every building location. By linking each structure's coordinates to its corresponding elevation value, the dataset captures important topographic information that may influence seismic response. Integrating these external data sources not only enhances the completeness of the dataset but also provides more meaningful inputs for the subsequent analysis and machine learning modeling.

3.2.4 Data distribution

The datasets used in this study capture a wide range of building characteristics and observed damage patterns following the earthquake. They include detailed records for

both masonry and reinforced concrete structures, covering diverse geographic, structural, and seismic conditions. This comprehensive distribution ensures a robust basis for analyzing building performance and developing reliable machine learning models.

3.2.4.1 Masonry buildings

The masonry dataset used in this study contains approximately 9,700 building records, representing a wide range of observed damage conditions following the earthquake. The distribution of damage states is as follows: about 25% of the buildings were classified as highly damaged, 1% as moderately damaged, 32% as lightly damaged, and 42% as undamaged, which is presented in Figure 3.8. This diversity provides a balanced foundation for understanding how different masonry structures responded under varying levels of seismic demand.

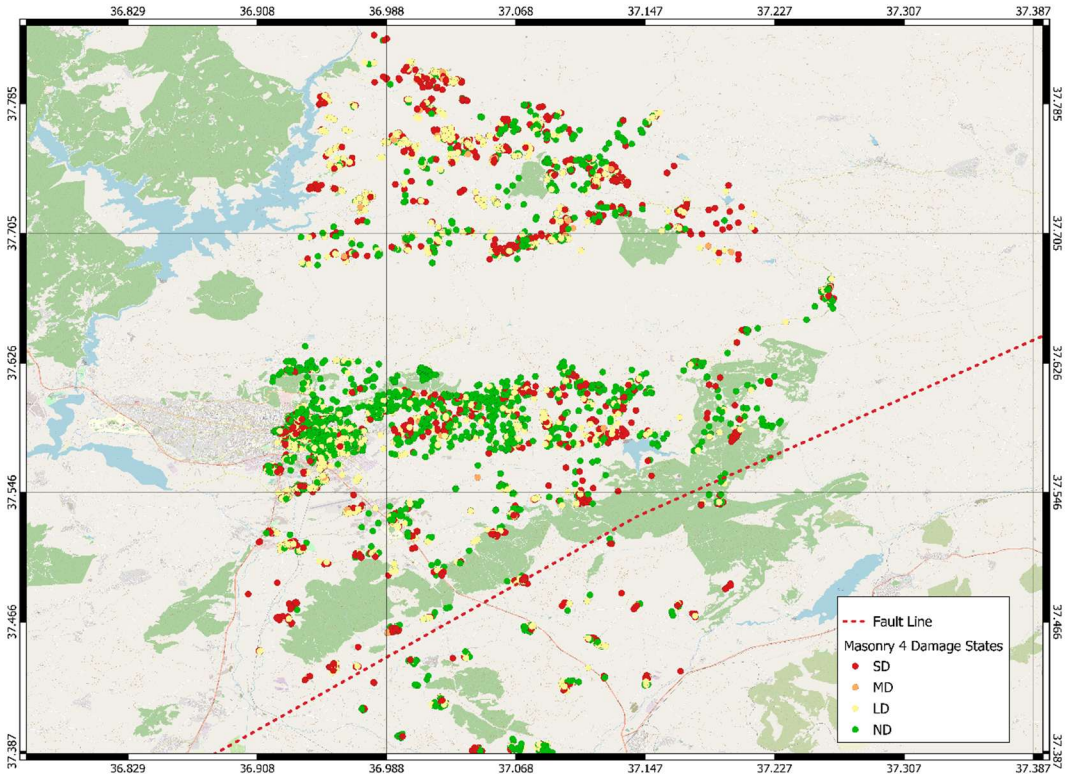

Figure 3.8: Masonry Building Distribution by Damage Class.

The data were collected from a near-fault region, with rupture distances (R_{rup}) averaging around 35 km. The surveyed areas span a broad topographic range, with ground elevations varying from 455 m to 1,764 m above sea level, as shown in (figure 3.9), and the damage

distribution by the elevation can be shown in (figure 3.10) as statistical distrubuted, reflecting the mountainous geography of the region. In total, building information was gathered from around 100 districts, ensuring comprehensive coverage of both densely populated urban areas and smaller rural settlements.

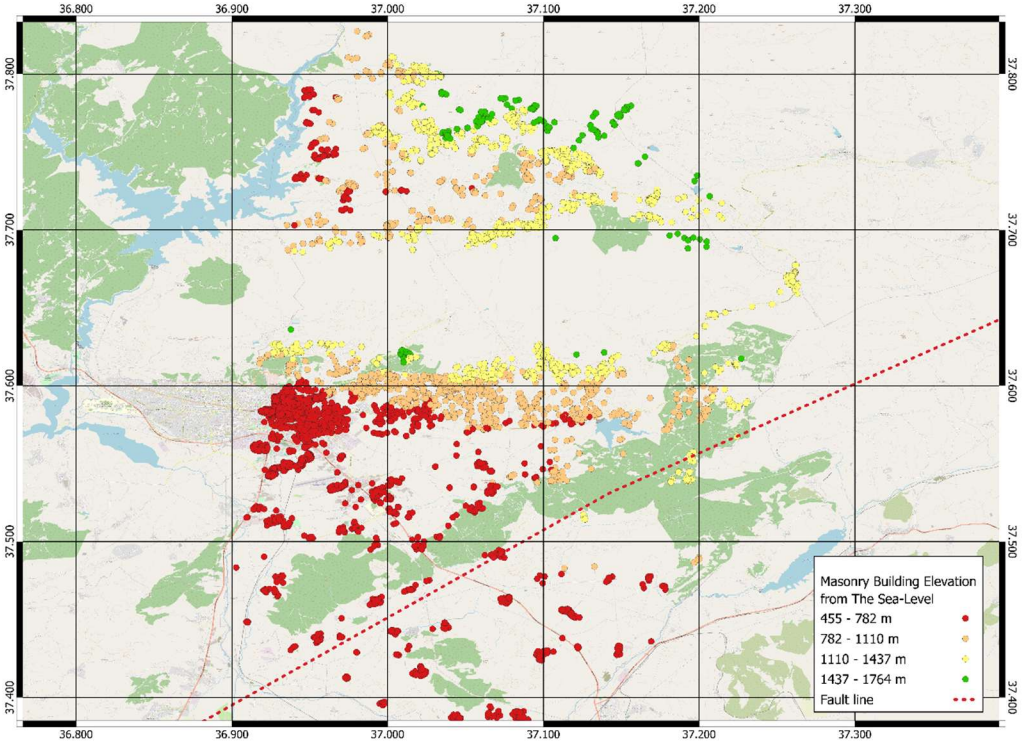

Figure 3.9: Masonry Building Distribution by Elevation.

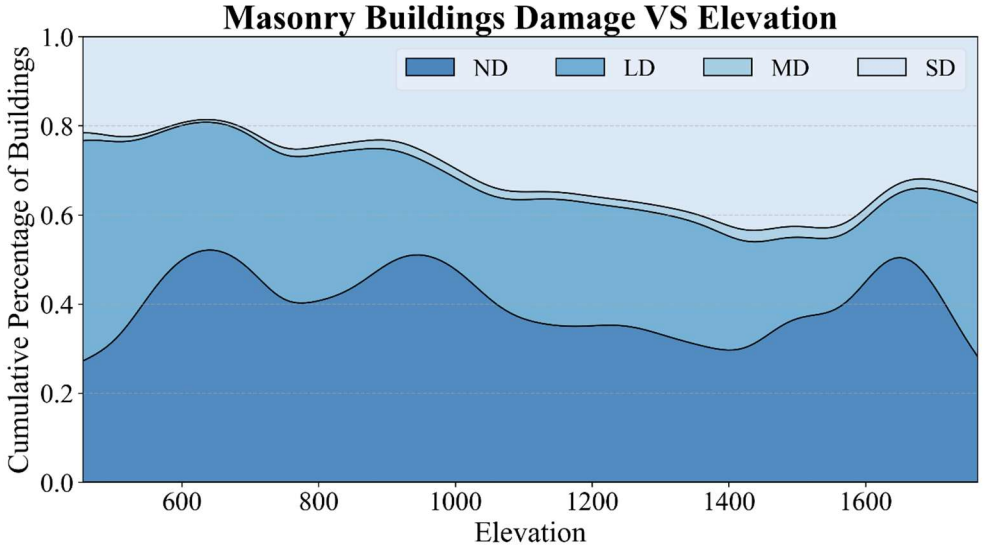

Figure 3.10: Statistical Distribution of Masonry Structures Damage States by Elevation.

Regarding the seismic input, the Peak Ground Acceleration (PGA) as geographically distributed is shown in Figure 3.11, where the recorded dataset ranges from 111 gal to 770 gal. The damage distribution by PGA is presented in Figure 3.12. The Peak Ground Velocity (PGV) varies between 11 cm/s and 129 cm/s. The corresponding seismic intensity values fall within the range of approximately 5.95 to 9.7, which can be presented on a map as in (figure 3.13), and the damage distribution by intensity shown in (figure 3.14), capturing both moderate and extreme shaking conditions. These variations in ground-motion parameters help provide a detailed and realistic representation of the seismic environment experienced by the masonry buildings.

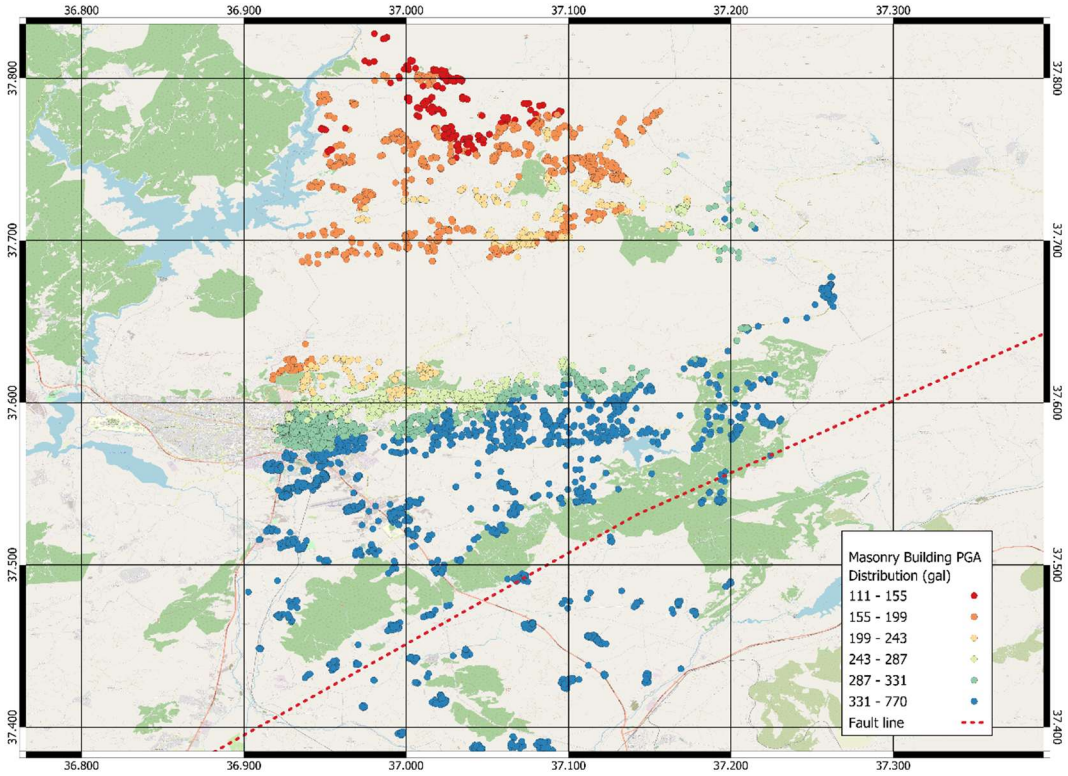

Figure 3.11: Masonry Building Distribution by PGA.

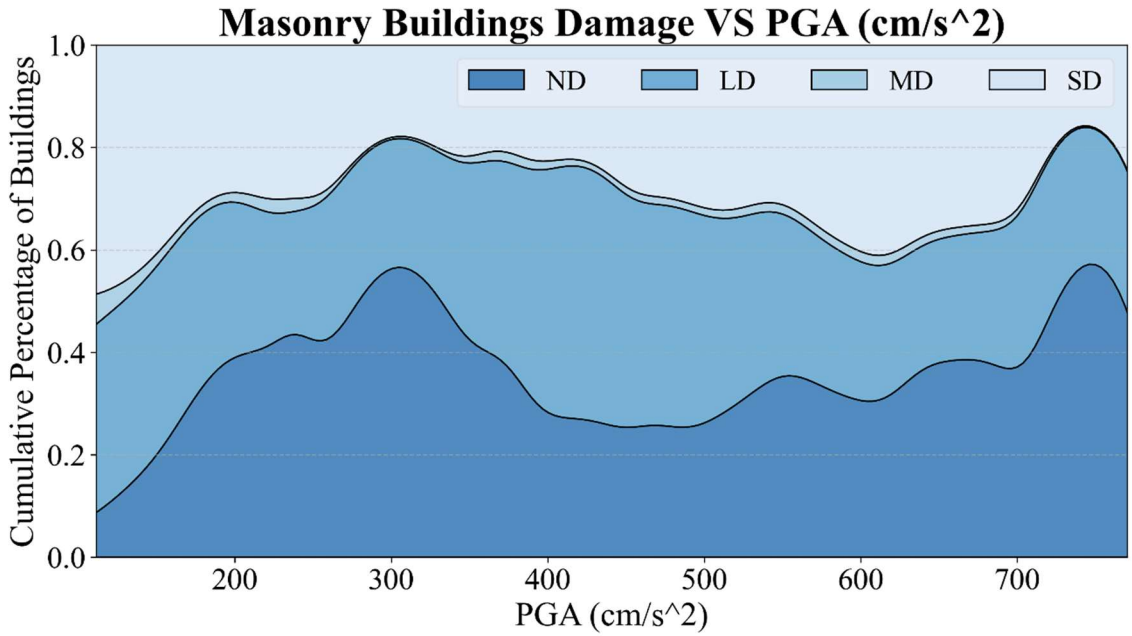

Figure 3.12: Statistical Distribution of Masonry Structures Damage States by PGA.

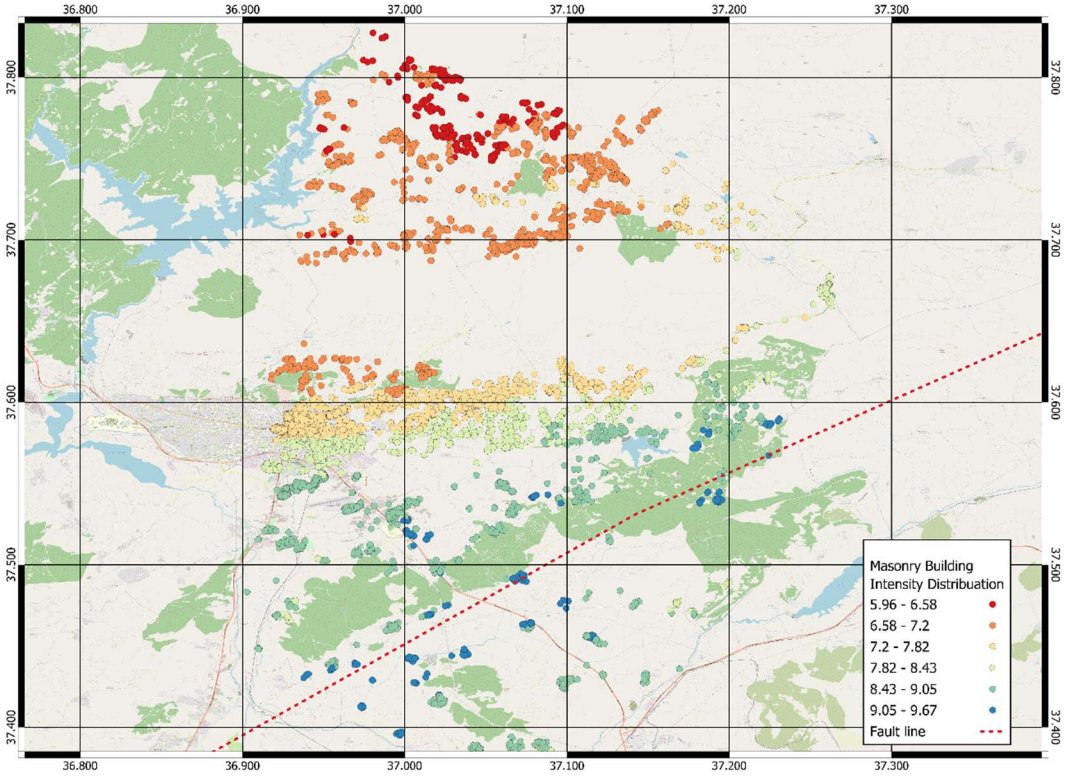

Figure 3.13: Masonry Building Distribution by Intensity.

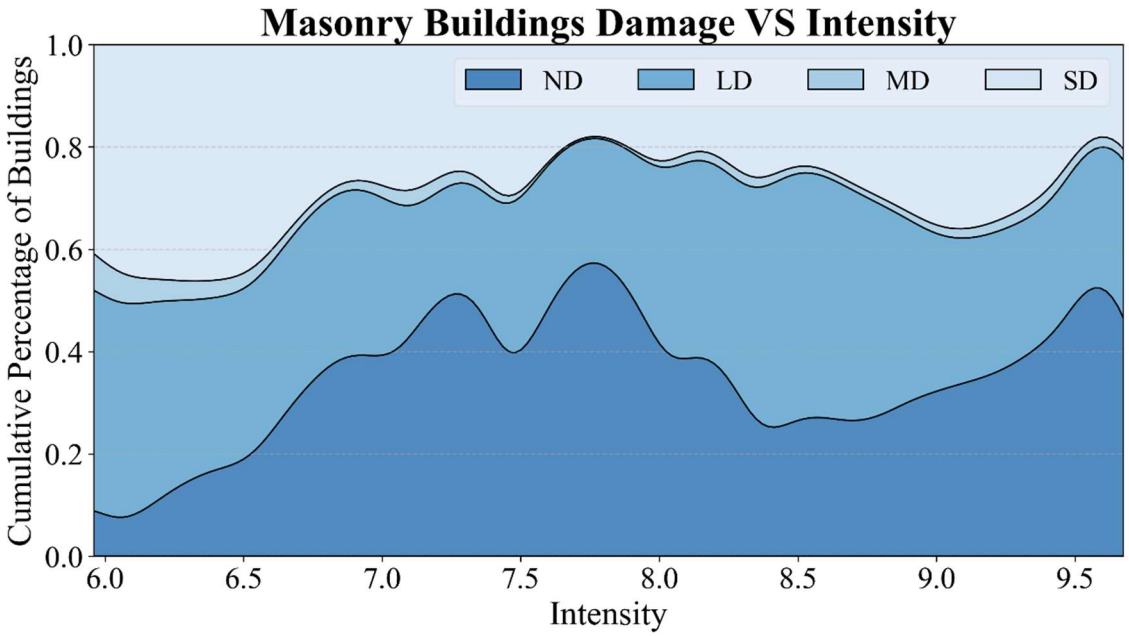

Figure 3.14: Statistical Distribution of Masonry Structures Damage States by Intensity.

Inspecting the distribution of damage by building age (as presented in Figure 3.15) shows an increase in the percentage of severely damaged buildings with increasing building age. The building ages span a wide range, from newly constructed structures to those up to 80 years old. (Figure 3.16) shows the distribution of damage by rupture distance, ranging from extremely close to the fault line up to 39 km away. This figure demonstrates that the damage varies with the distance from the fault line.

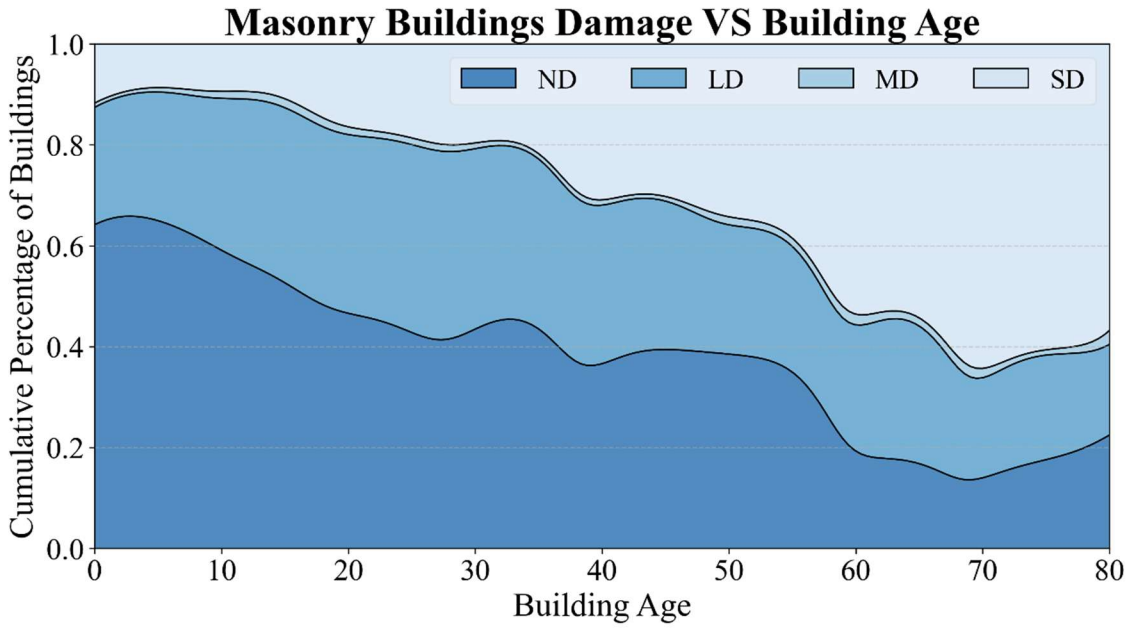

Figure 3.15: Statistical Distribution of Masonry Structures' Damage States by Age.

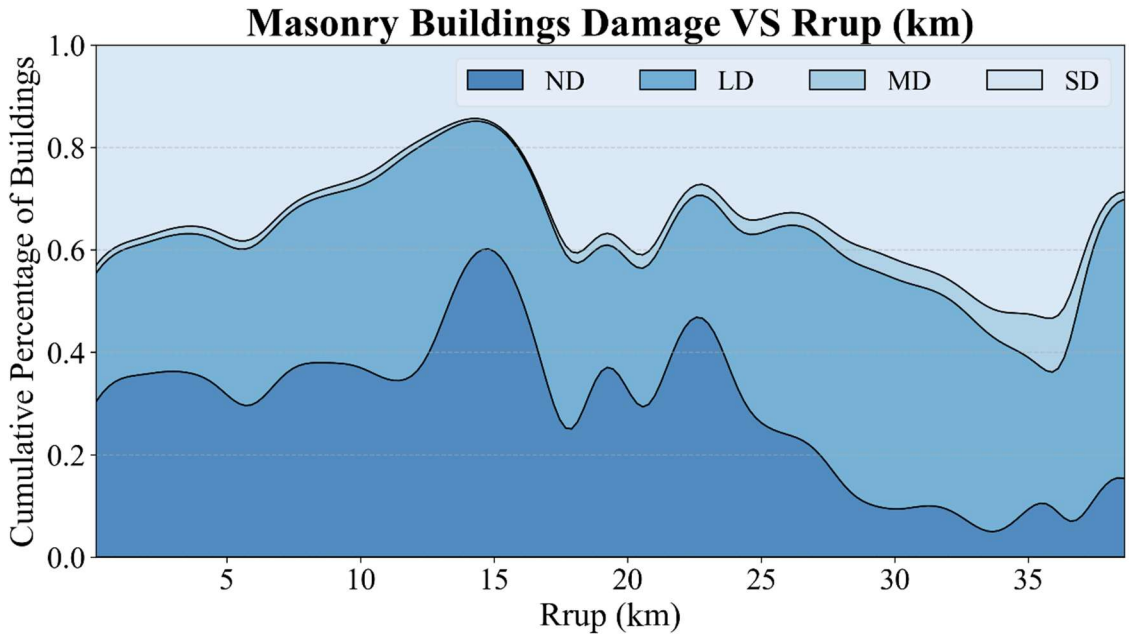

Figure 3.16: Statistical Distribution of Masonry Structures Damage States by Rrup.

3.2.4.2 Reinforced concrete buildings

The reinforced concrete (RC) dataset includes around 31,000 buildings, providing a broad and detailed picture of how RC structures were affected by the earthquake. The damage

levels vary widely: about 17% of the buildings were heavily damaged, 3% showed moderate damage, 37% experienced only minor damage, and 43% were found to be undamaged, where the geographical distribution can be shown in Figure 3.17. All of this information was collected from the same region as the masonry dataset, ensuring that the buildings were exposed to similar seismic conditions.

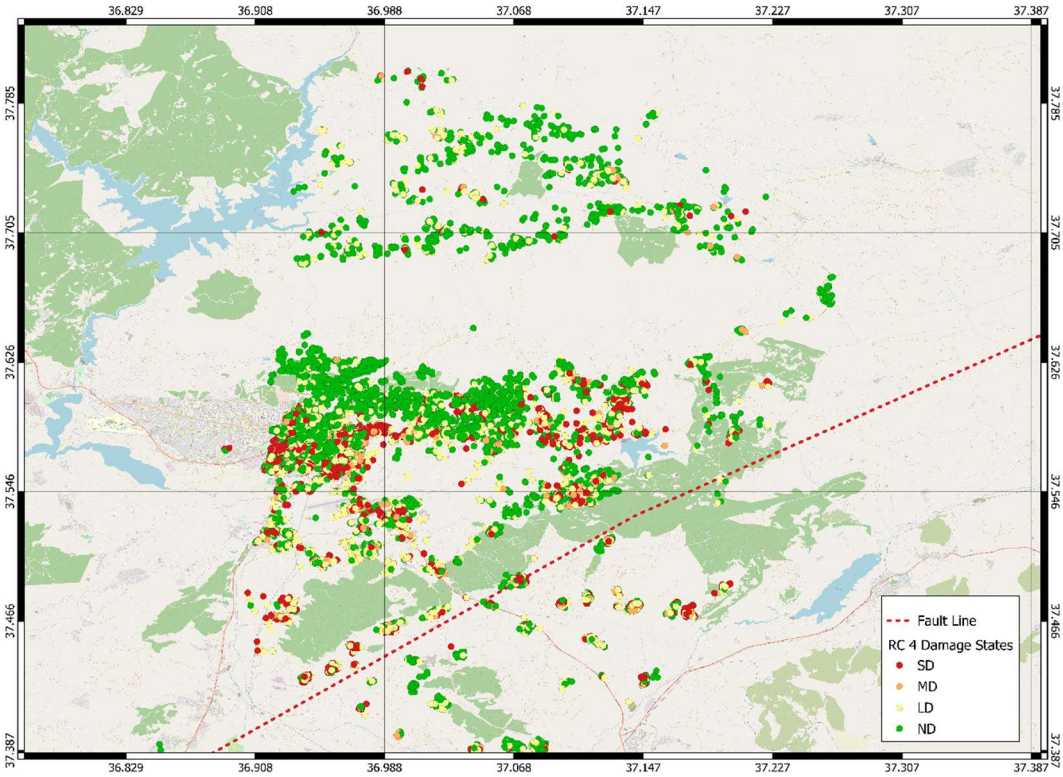

Figure 3.17: RC Building Distribution by Damage Class.

The RC buildings themselves represent a diverse range of characteristics. Their ages span from very new constructions—only about a year old—to older buildings built roughly 70 years ago, where the age distribution by damage can be shown in Figure 3.18. The number of stories also varies, with some structures reaching up to 15 floors. Data were gathered from 107 districts, giving a wide geographic coverage that includes dense urban centers as well as smaller surrounding areas.

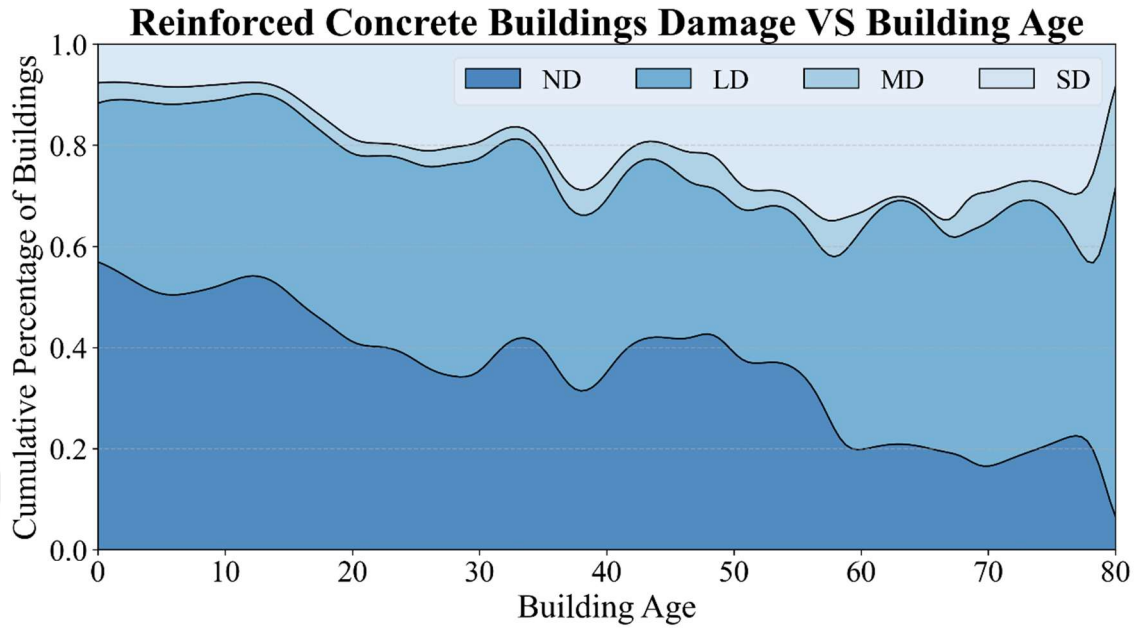

Figure 3.18: Statistical Distribution of RC Structures Damage States by Age.

(Figure 3.19) shows the geographical distribution of elevations of RC structures. The results indicate that the percentage of severely damaged buildings decreases as elevation above sea level increases (Figure 3.20). This trend is attributed to the increasing topographic distance as structures are located farther from the fault line.

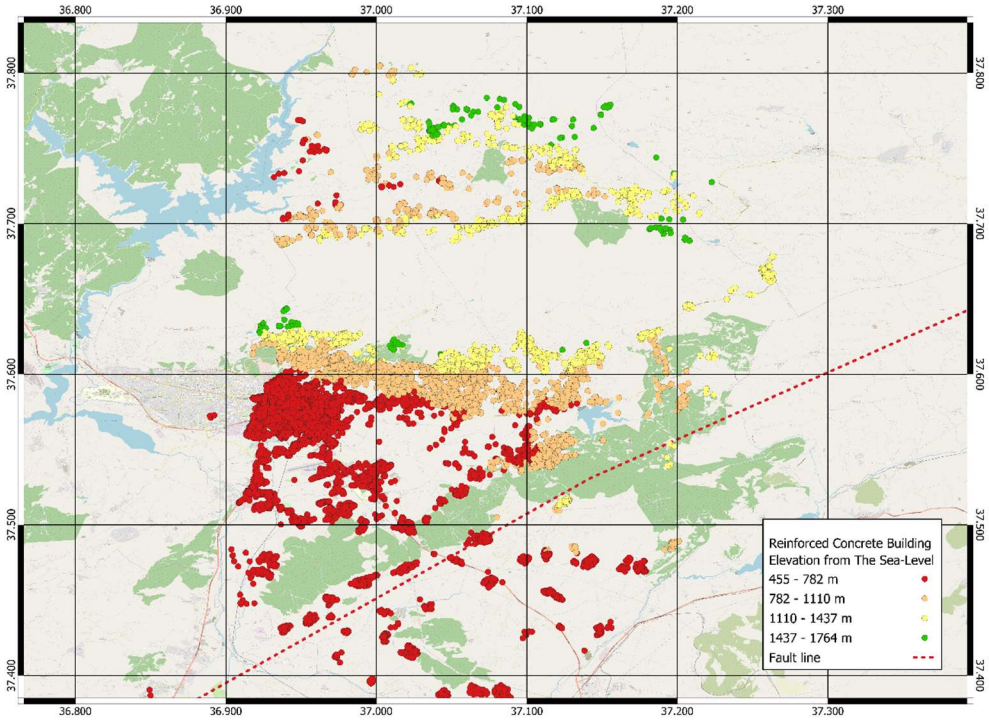

Figure 3.19: RC Building Distribution by Damage Elevations.

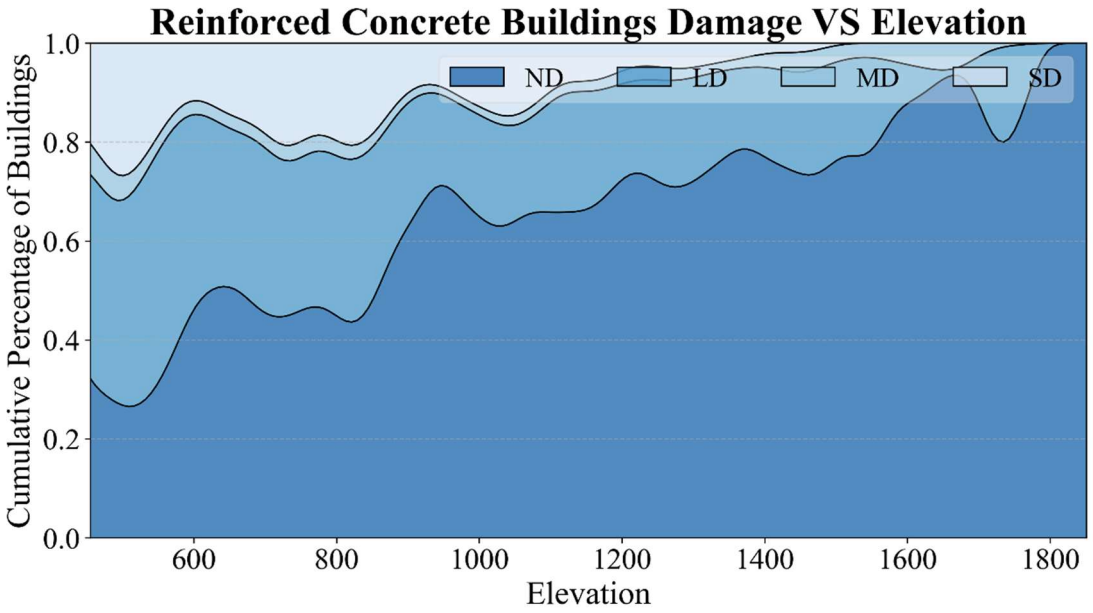

Figure 3.20: Statistical Distribution of RC Structures Damage States by Elevation.

The seismic shaking experienced by these buildings covered a wide range of intensities. The geographical distribution of Peak Ground Acceleration (PGA) is shown in Figure 3.21, where values range from 89 gal to 770 gal, and the damage state distribution by PGA is presented in Figure 3.22. Peak Ground Velocity (PGV) values varied between 11 cm/s and 129 cm/s. (Figure 3.23) shows the geographical distribution of intensity levels, ranging from approximately 5.85 to 9.7, while their distribution by damage level is illustrated in (Figure 3.24), capturing conditions from moderate shaking to extremely strong ground motion. The damage state distribution by rupture distance is presented in Figure 3.25, covering locations from the fault line up to 35 km away, and shows a decrease in the percentage of severely damaged structures with increasing rupture distance.

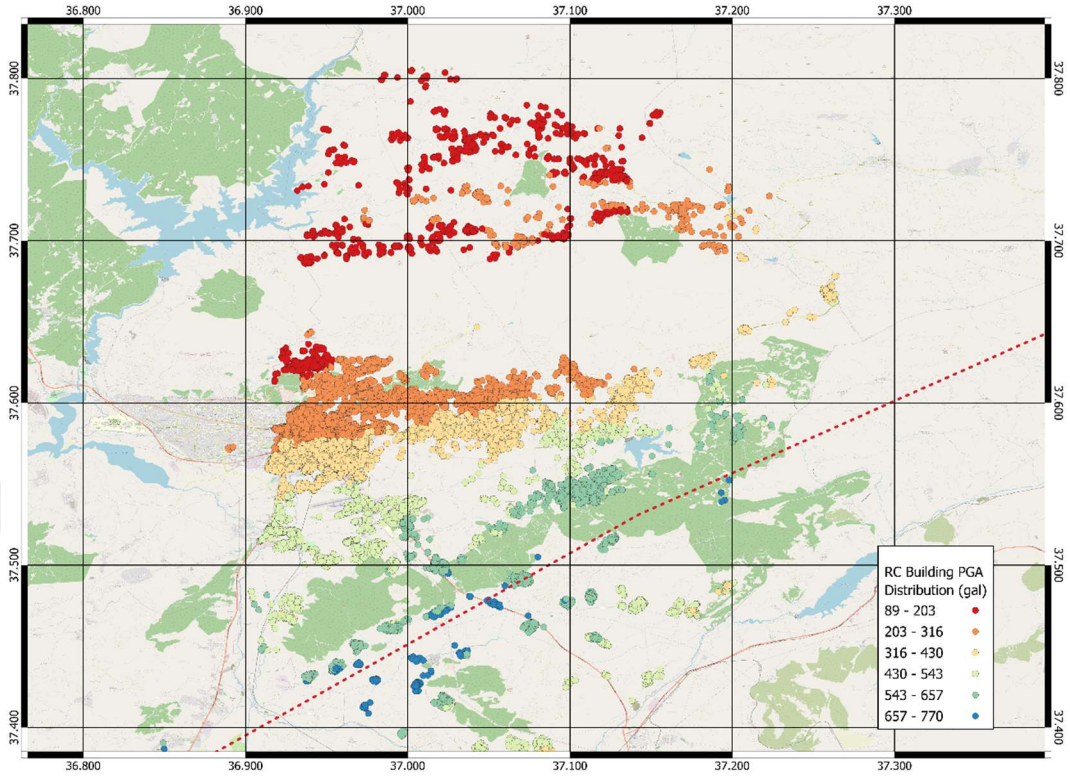

Figure 3.21: RC Building Distribution by PGA.

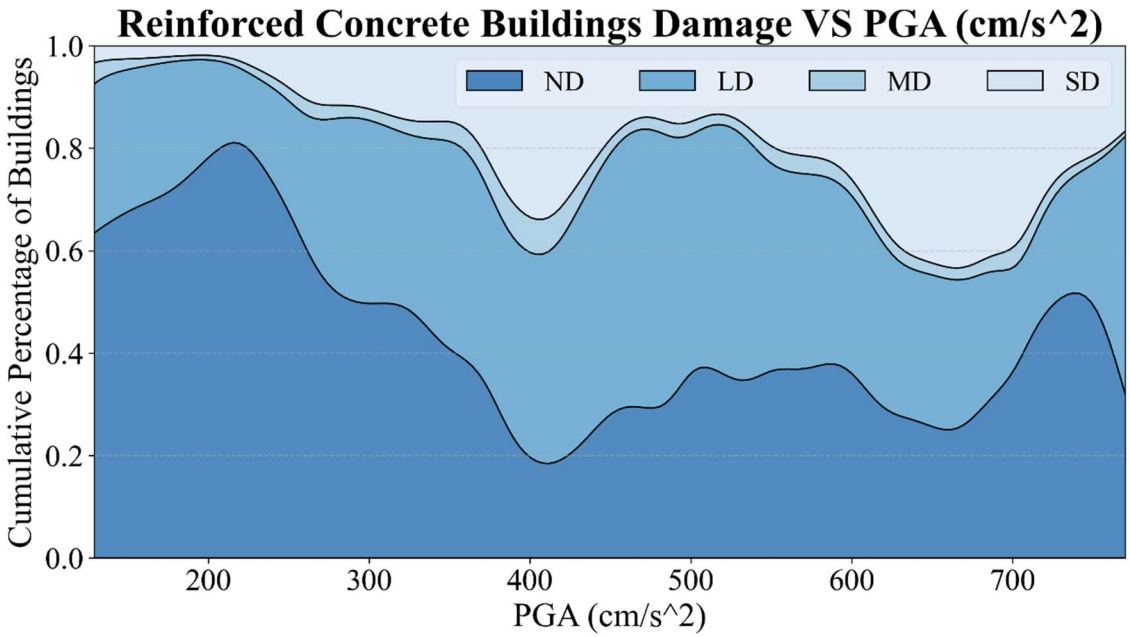

Figure 3.22: Statistical Distribution of RC Structures Damage States by PGA.

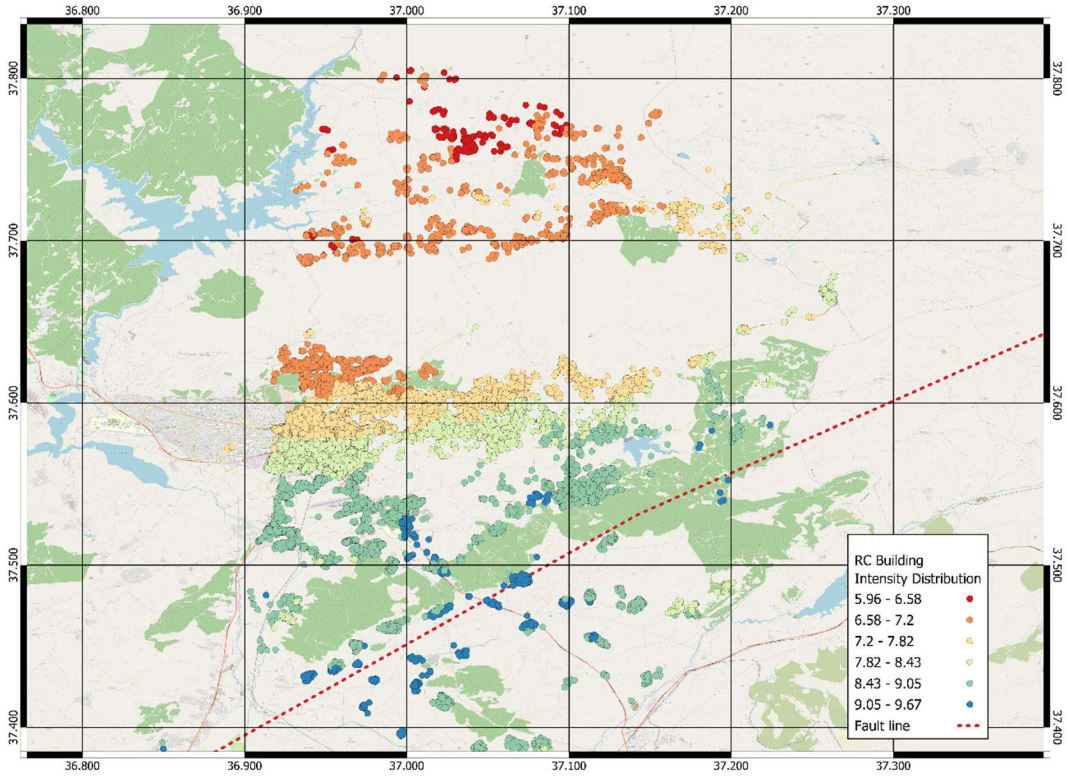

Figure 3.23: RC Building Distribution by Intensity.

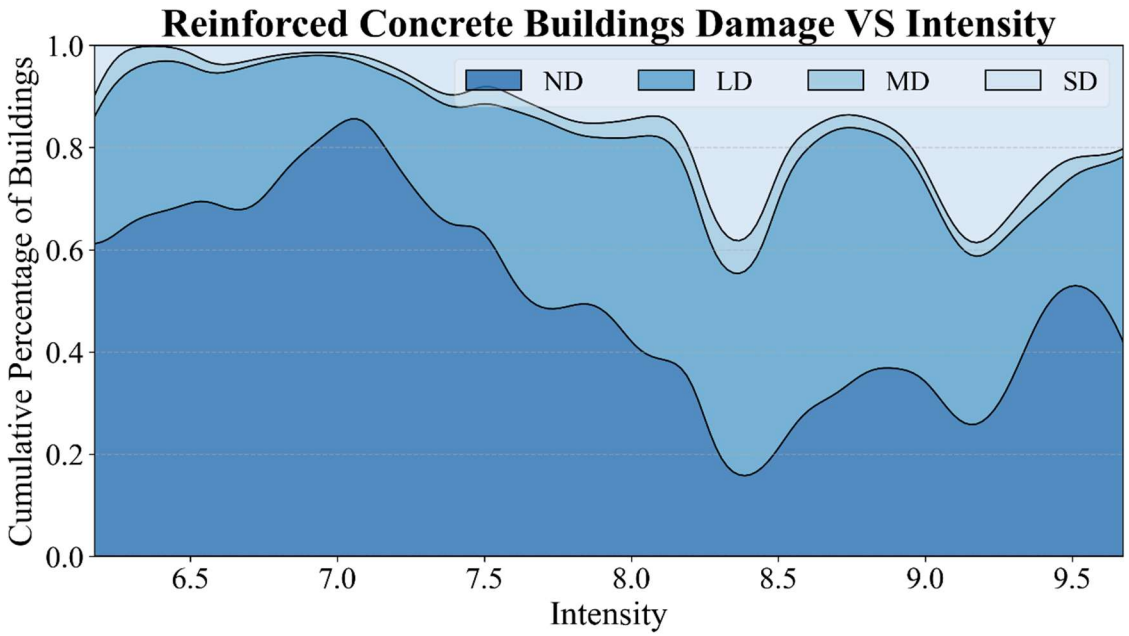

Figure 3.24: Statistical Distribution of RC Structures Damage States by Intensity.

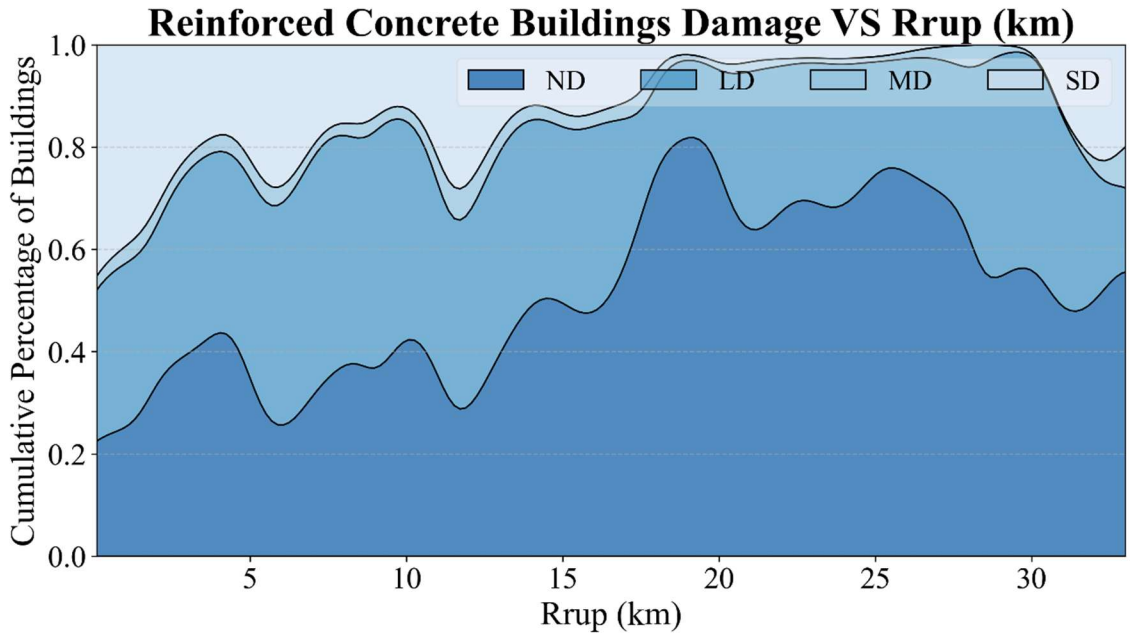

Figure 3.25: Statistical Distribution of RC Structures Damage States by Rrup.

Together, these variations—across building age, height, location, and shaking intensity—make the RC dataset a rich and realistic foundation for understanding how reinforced concrete buildings respond to major earthquakes and for developing reliable machine learning models.

3.2.5 Dataset used in the study case

In this study, the machine learning models were applied by considering both the type of structure and the damage-state classification scheme, reflecting the procedures used in post-earthquake visual screening inspections. Two primary structural types were examined: masonry buildings and reinforced concrete (RC) frame structures. This separation is motivated by the fundamental differences in damage mechanisms and inspection criteria between the two systems. For masonry buildings, damage assessment primarily focuses on crack patterns and failure modes in walls, whereas for RC buildings, evaluation is centered on the condition of structural members such as columns, beams, and joints. Treating these building types separately allows the models to better capture structure-specific damage characteristics.

In addition, multiple damage-state classification schemes were adopted to support different operational objectives. The binary classification scheme is intended for the rapid identification of safe and unsafe structures, enabling immediate post-earthquake decisions. The three-class damage scheme, commonly referred to as the traffic-light classification, introduces an intermediate category for buildings that require further assessment, supporting authorities in prioritizing inspections and allocating emergency resources during the early response phase. Finally, the four-class damage classification provides a more detailed representation of damage severity, closely aligning with observed post-earthquake damage states and offering deeper insight for comprehensive damage evaluation and recovery planning.

3.2.6 Machine learning algorithm

This section presents the structured workflow followed to develop, train, and assess the machine learning models applied in this study. The methodology includes data preprocessing, model tuning, validation procedures, and comparative performance analysis to achieve reliable and unbiased damage classification results. Each phase of the process is carefully designed to improve prediction accuracy and to support effective decision-making in post-earthquake scenarios.

3.2.6.1 Data cleaning and filtering

The raw data was subjected to a cleaning phase to rectify non-existent points, corrupted data, and misspellings, alongside feature encoding tailored to the specific research focus. To further refine the dataset, filtering was applied to exclude any illogical entries and outliers residing outside the 95% normal distribution. This systematic preparation was essential to minimize noise and prevent anomalous values from dominating the model's learning process, thereby enhancing generalization.

3.2.6.2 Data splitting

When the dataset was validated, it was divided into three functional segments: 80% for training, 10% for testing and performance tuning, and 10% for final evaluation. The

evaluation set served as a strictly independent data source, never exposed to the model during the development phase. This separation is fundamental to obtaining an unbiased reflection of the model's true ability to solve real-world problems when faced with entirely new information.

3.2.6.3 Tuning ML model hyperparameters

Hyperparameters significantly influence an algorithm's sensitivity and data-handling characteristics. To move beyond default configurations, a grid-search approach was employed to explore the hyperparameter space in a structured manner. By evaluating combinations through a predefined metric, the most effective configuration was identified. This procedure was executed via the validation dataset to avoid information leakage, thereby improving the model's accuracy and reducing the risk of overfitting.

3.2.6.4 K-folds and cross-validation

K-fold cross-validation was utilized to ensure the model remained well-generalized and unbiased. Following the 80/10 split for training and testing, a Stratified 10-fold cross-validation was applied to the training data. This ensured that class distributions remained consistent across all folds. The final model selection was determined by analyzing the confusion matrix, specifically prioritizing the configuration that minimized the confusion between the "worst damaged" and "safe" cases, thereby enhancing the model's practical safety.

3.2.6.5 Fittings of training data

Following the conclusion of the cross-validation and tuning phases, the finalized training split was used to fit the model. During this stage, the model internalized data patterns according to the optimized parameters. The result of this process was a stable, finalized model prepared for rigorous performance testing and final comparative analysis.

3.2.6.6 Fitting of validation data

In the development workflow, a portion of the data was held as a validation set to serve as an objective performance check. This data is not involved in the direct training of the model; rather, it provides a means to evaluate the model's performance on unseen inputs. This stage ensures that the resulting model is generalizable and that its performance metrics reflect real-world applicability with a high degree of confidence.

3.2.6.7 Comparison of utilized models

The performance of the six models was assessed primarily using the confusion matrix, with accuracy and F1-score applied as additional evaluation metrics. The confusion matrix was emphasized because it clearly illustrates classification errors, which is particularly important in safety-critical applications. Accuracy provided an overall measure of correct predictions, whereas the F1-score ensured balanced evaluation in the presence of class imbalance by accounting for both precision and recall. By combining these metrics, the models were systematically compared to determine the one that achieved the most reliable overall performance, especially in identifying important minority damage classes.

4. RESULTS EVALUATION AND DISCUSSION

This section presents a comprehensive evaluation of machine learning–based frameworks for post-earthquake building damage assessment using real data from the 2020 Izmir and 2023 Kahramanmaras earthquakes. Model performance is systematically analyzed across multiple damage classification schemes and spatial scales, ranging from individual buildings to district and city levels. In addition, feature importance analysis is employed to identify the key seismic, structural, and site-specific parameters governing damage prediction and model reliability.

4.1 Izmir Earthquake

This section evaluates six machine learning models developed to predict building damage using real field data collected after the 2020 Izmir earthquake. The models were applied to the two datasets described in Section 4.1.5 to assess their ability to classify buildings into three damage-state categories. Furthermore, a feature importance analysis was conducted to identify the structural, seismic, and site-related variables that most strongly influence the model predictions. The results of this analysis offer practical insights that can contribute to more accurate and effective post-earthquake risk assessment.

4.1.1 Accuracy and F1 score

Figures 4.1 and 4.2 illustrate the comparative performance of six machine learning algorithms in terms of classification accuracy and F1-score, respectively, for Model 1 and Model 2. The key distinction between the two modeling frameworks is that Model 2 incorporates building age as an additional input feature, whereas Model 1 relies on the baseline feature set.

Figure 4.1 shows that Model 2 consistently outperforms Model 1 across all algorithms. While Model 1 achieves accuracies in the range of approximately 66% (DTR) to 77% (HGB), the inclusion of building age in Model 2 increases predictive performance, with accuracies rising to approximately 75%–87%. The most notable improvements are observed for ensemble-based methods, particularly Random Forest (RF) and Histogram Gradient Boosting (HGB), reaching 86% and 87% accuracy under Model 2, respectively. This highlights the strong contribution of building age in enhancing the models' ability to differentiate between damage states.

A similar pattern is evident in Figure 4.2. Model 2 yields consistently higher F1-scores across all algorithms, with values ranging from approximately 74% (GB) to 86% (HGB), compared to 46% (DTR) to 76% (HGB) for Model 1. The improvement in F1-score indicates that the inclusion of building age not only increases overall classification accuracy but also improves the balance between precision and recall across damage classes. This is particularly important in post-earthquake damage assessment, where older buildings are generally more vulnerable, and misclassification of damaged structures may lead to critical safety implications.

Overall, these results demonstrate that building age is a highly informative feature for post-earthquake damage classification. Its inclusion in Model 2 leads to a systematic performance improvement across all machine learning algorithms, with ensemble models—especially Random Forest, XGBoost, and Histogram Gradient Boosting—showing the greatest benefit. This finding reinforces the importance of incorporating structural deterioration and construction-era effects when developing data-driven frameworks for rapid seismic damage assessment.

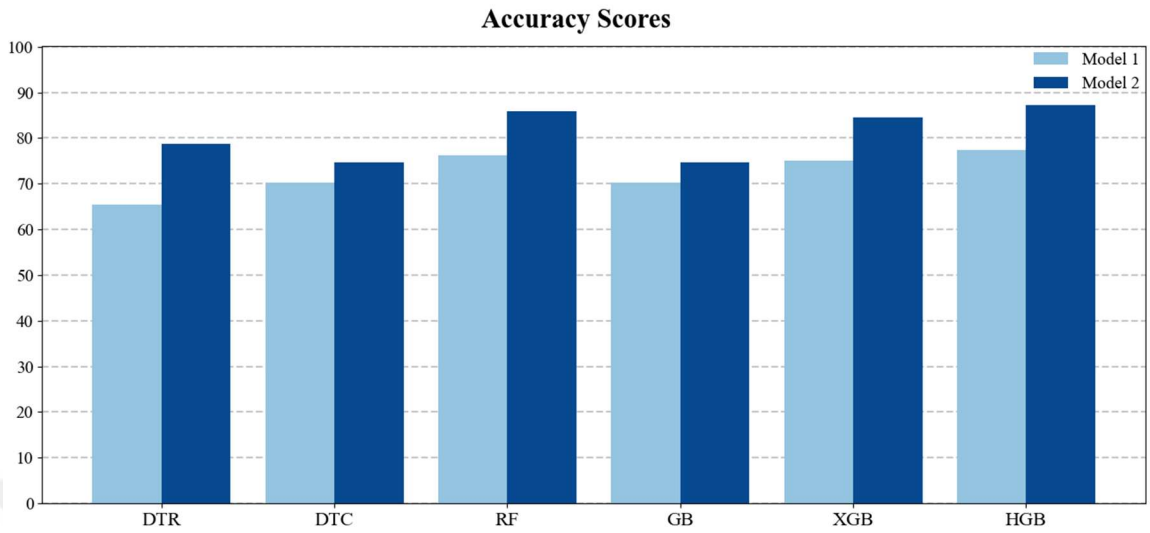

Figure 4.1: Accuracy Score of Izmir Case for Model 1 and Model 2.

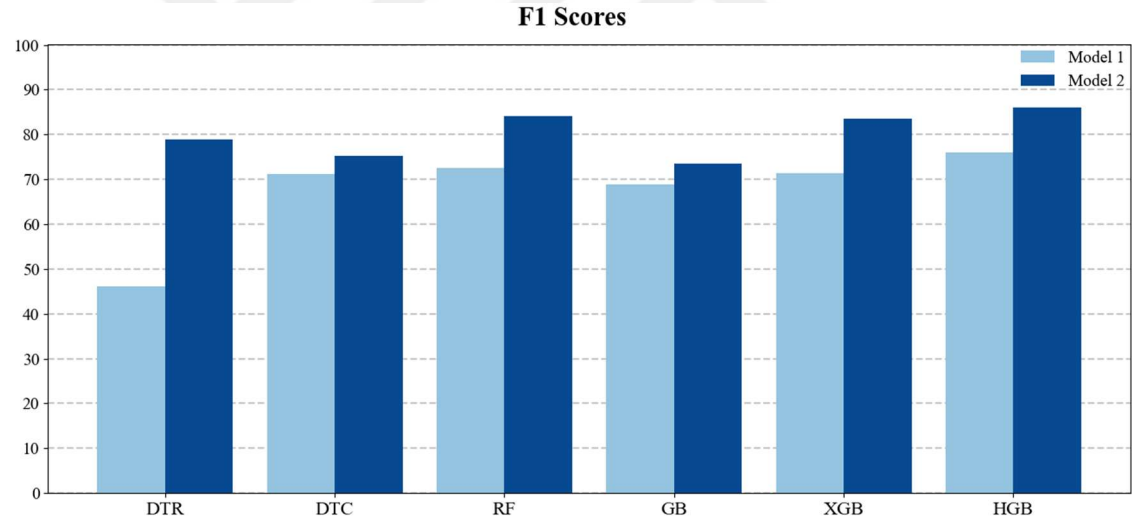

Figure 4.2: F1-Score of Izmir Case for Model 1 and Model 2.

4.1.2 Model 1 performance evaluation

Model 1 evaluates the performance of the machine learning algorithms when building age is not included in the input features. The confusion matrices in Figure 4.3 show that most models tend to classify the “Safe” class correctly, but struggle to distinguish between “Not Safe” and “To Be Checked” buildings.

The Random Forest (RF) model shows strong performance for the “Safe” class, correctly predicting about 93% of safe buildings. However, it performs very poorly for the “Not Safe” class, where 83% of not safe buildings are incorrectly classified as “Safe.” For the “To Be Checked” class, only 44% are correctly identified, while 50% are misclassified as “Safe.” This indicates a clear tendency to underestimate damage severity.

The Decision Tree Classifier (DTC) shows similar behavior. Although 80% of safe buildings are correctly classified, only 50% of “Not Safe” buildings are correctly identified, and the remaining 50% are predicted as “To Be Checked.” For the “To Be Checked” class, half of the cases are misclassified as “Safe,” again showing difficulty in separating adjacent damage levels.

The Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Histogram Gradient Boosting (HGB) models also show a strong bias toward predicting the “Safe” class. For example, 67% of “Not Safe” buildings are classified as “Safe” in both GB and XGB, and 66% in HGB. Although these models correctly identify between 88% and 91% of safe buildings, they frequently misclassify damaged buildings as safe. The “To Be Checked” class remains particularly challenging, with a large portion predicted as “Safe.”

The Decision Tree Regressor (DTR) shows slightly better balance for the “Not Safe” class compared to some other models, correctly identifying 50% of these cases. However, it still misclassifies the remaining 50% as either “To Be Checked” or “Safe.” Similar to the other models, it also struggles to clearly separate “To Be Checked” from “Safe.”

Overall, the results of Model 1 indicate that removing building age leads to systematic underestimation of damage. Most models show high accuracy for the “Safe” class but poor performance for the more severe damage categories. This explains the observed misclassification patterns and highlights the importance of including vulnerability-related features in the model.

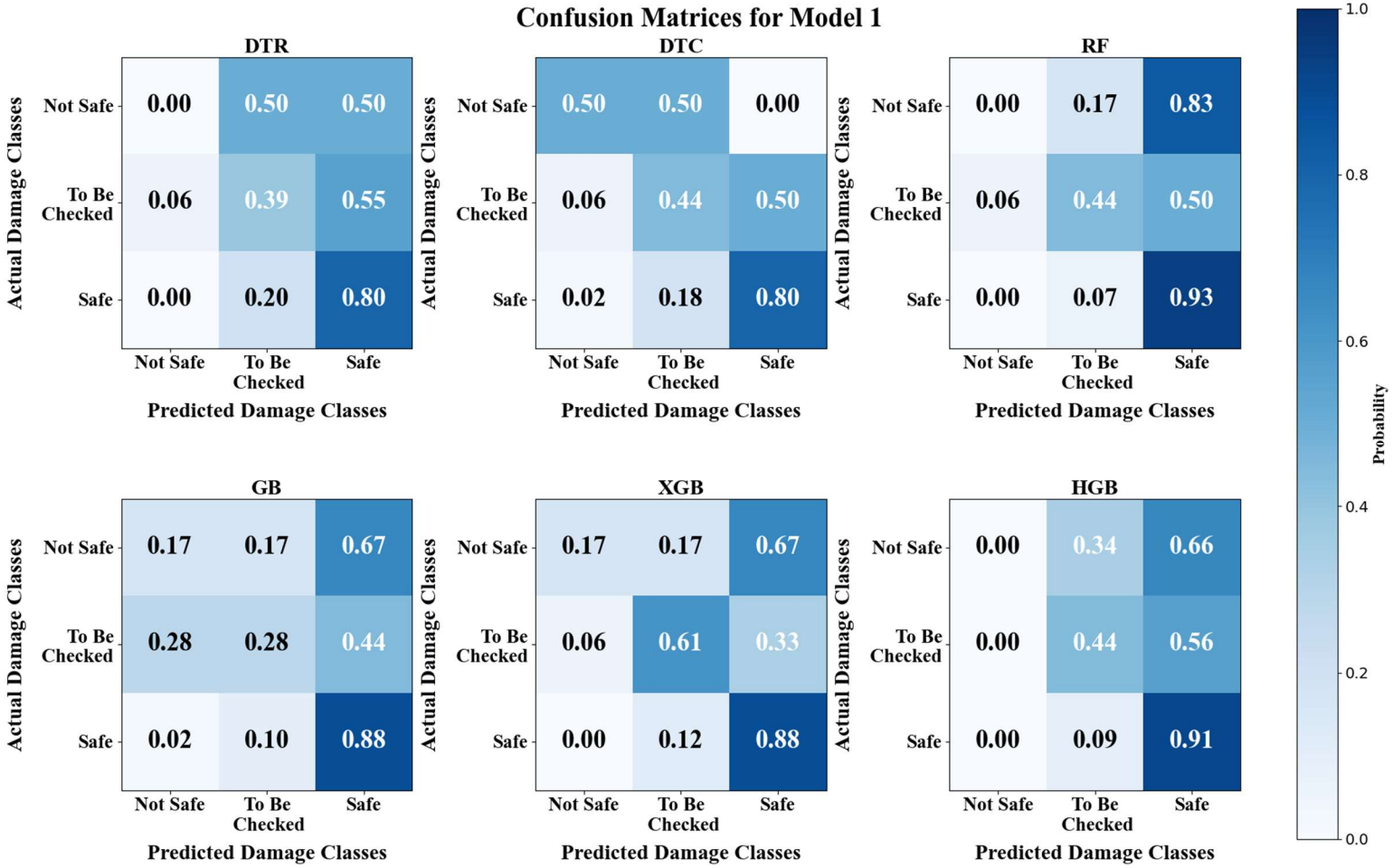

Figure 4.3: Confusion Matrices of Izmir Case for Model 1.

4.1.3 Model 2 performance evaluation

The Model 2 evaluates the performance of the machine learning algorithms when building age is included in the input features. The confusion matrices in Figure 4.4 show that the models now correctly classify both the “Not Safe” and “Safe” classes with high accuracy in most cases, thanks to the addition of this vulnerability-related feature, while the “To Be Checked” class remains the most challenging but is handled substantially better than in Model 1.

The Random Forest (RF) model shows excellent performance overall. It correctly predicts 100% of “Not Safe” buildings and 96% of “Safe” buildings. For the “To Be Checked” class, 38% are correctly identified, while 62% are misclassified as “Safe.” This indicates a clear reduction in the tendency to underestimate damage severity compared to Model 1.

The Decision Tree Classifier (DTC) shows notably poorer performance than the other models. It fails to correctly identify any “Not Safe” buildings (0% correct), with 80% misclassified as “Safe” and 20% as “To Be Checked.” For the “To Be Checked” class, 54% are correctly identified. However, no “Safe” buildings are correctly predicted as “Safe”; instead, 83% are classified as “To Be Checked” and 17% as “Not Safe,” revealing significant confusion across adjacent damage levels.

The Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Histogram Gradient Boosting (HGB) models demonstrate strong performance, particularly for the “Not Safe” and “Safe” classes. For example, XGB correctly identifies 100% of “Not Safe” buildings and 93% of “Safe” buildings, HGB achieves 75% for “Not Safe” and 98% for “Safe,” and GB reaches 50% for “Not Safe” and 87% for “Safe.” Although the “To Be Checked” class is still partially misclassified as “Safe” (69% in GB, 54% in both XGB and HGB), the overall separation of damage levels is markedly improved.

The Decision Tree Regressor (DTR) shows strong balance across classes, correctly identifying 100% of “Not Safe” buildings, 46% of “To Be Checked” (with the remaining 54% misclassified as “Safe”), and 85% of “Safe” buildings. It performs particularly well in distinguishing severe damage while maintaining good accuracy for the “Safe” class.

Overall, the results of Model 2 indicate that including building age greatly enhances the model’s ability to identify “Not Safe” buildings and reduces the systematic underestimation of damage severity observed in Model 1. Most models now achieve high

accuracy for both “Safe” and “Not Safe” classes, although the “To Be Checked” class continues to pose some difficulty. This confirms the critical importance of including vulnerability-related features, such as building age, for reliable damage assessment.

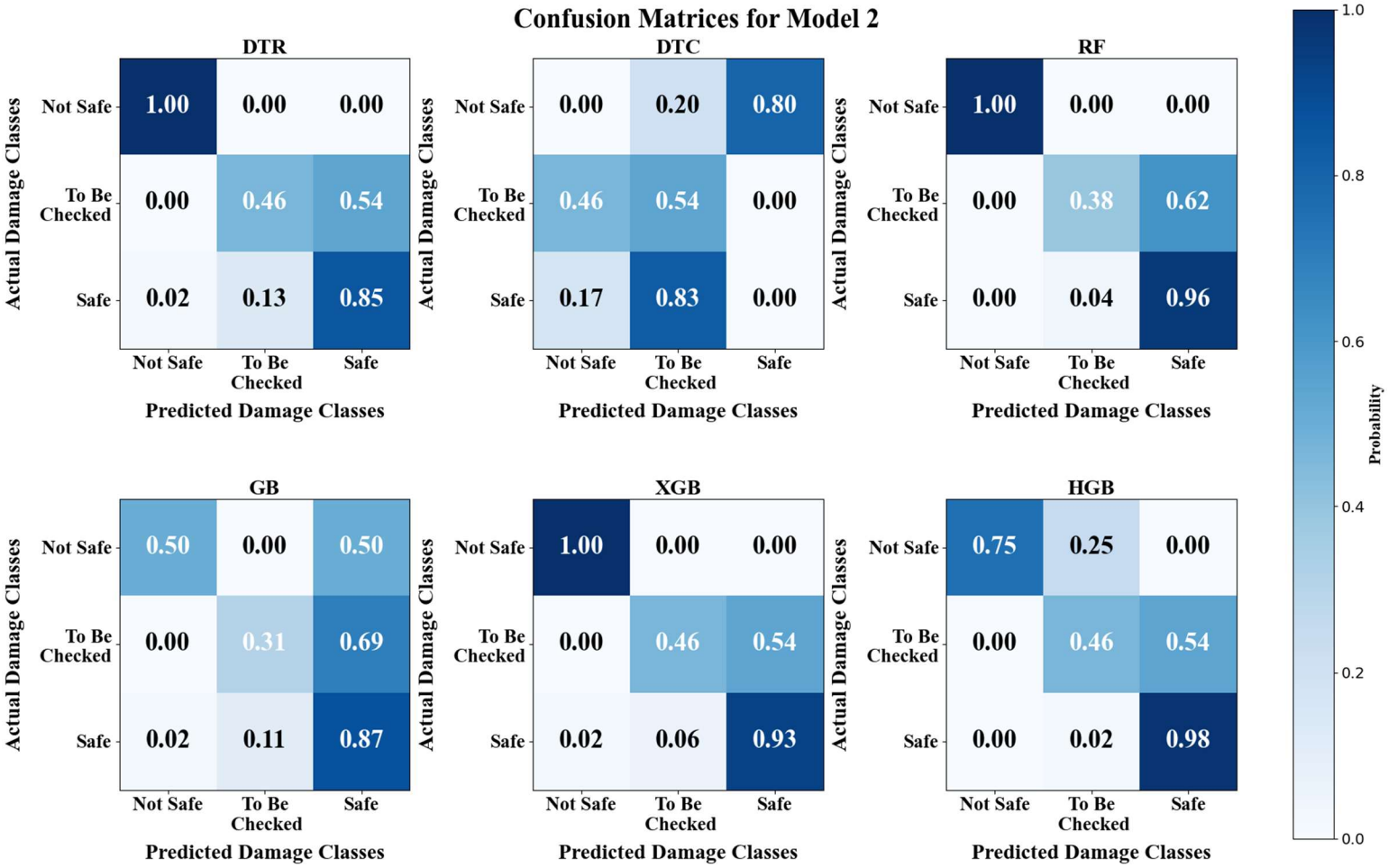

Figure 4.4: Confusion Matrices of Izmir Case for Model 2.

4.1.4 Model 1 features importance analysis

Figure 4.5 illustrates the relative importance of the input variables included in Model 1, as determined using the Histogram Gradient Boosting (HGB) classifier. The reported importance scores reflect the contribution of each feature to the model's ability to predict building damage states.

The results show that building area is the most important feature, contributing about 23% to the model. Elevation comes next at around 22%, followed by the number of floors at approximately 18.6%. This means that the size of the building, its height, and its location in terms of elevation play a major role in damage prediction. These structural and geometric characteristics have a stronger effect on the model than most ground-motion parameters.

Peak Ground Acceleration (PGA) contributes about 15.5%, making it the most important earthquake-related parameter in the model. This confirms that ground-motion intensity is still a key factor in predicting damage, although its effect is lower than that of some structural features. The S1 parameter has a moderate contribution (about 6.7%), while VS30 and PGV contribute around 5.6% and 4.8%, respectively. These parameters provide additional information, but their influence is smaller compared to area, elevation, and number of floors.

Rupture distance (Rrup) has a small contribution (around 2.8%), and macroseismic intensity has almost no effect on the model (less than 1%). This suggests that these two variables do not significantly improve the model's ability to classify damage levels.

Overall, Model 1 mainly depends on building size, height, elevation, and PGA for predicting damage states. Site parameters and macroseismic intensity have a secondary role. These results help explain how the HGB model makes its decisions and why structural characteristics are more influential when other vulnerability-related features are not included.

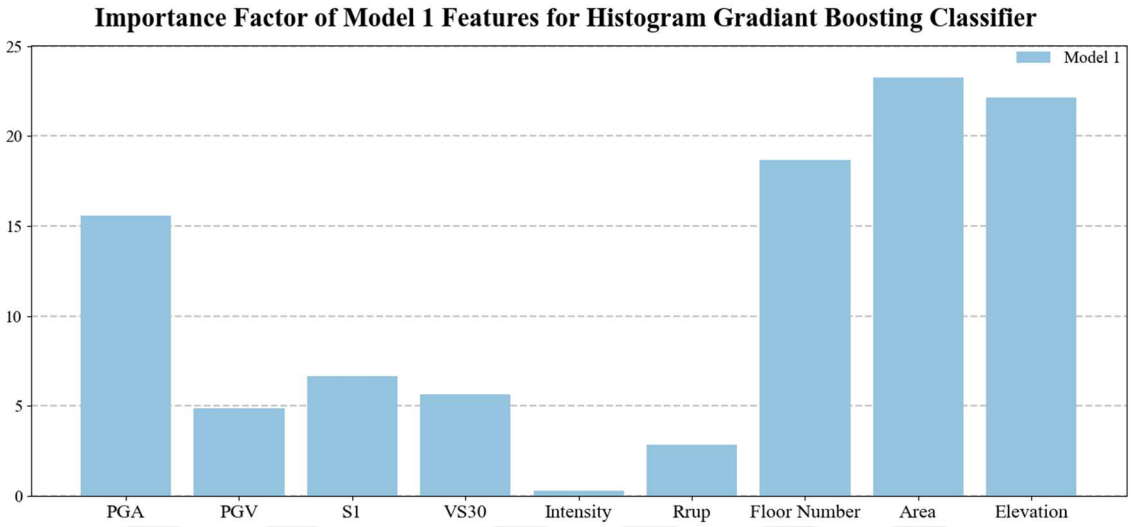

Figure 4.5: Feature Importance Factor of Izmir Case for Model 1.

4.1.5 Model 2 features importance analysis

Figure 4.6 presents the relative importance of the input variables used in Model 2, as derived from the Histogram Gradient Boosting (HGB) classifier. The importance scores represent the extent to which each feature contributes to the prediction of post-earthquake building damage states.

The results show that rupture distance (Rrup) is the most important feature (about 21%). Building age (around 21%) and building area (about 20%) follow closely. This means that in Model 2, both vulnerability-related factors (such as age and size of the building) and proximity to the earthquake source play a major role in damage prediction. The high importance of building age highlights its strong influence in representing structural vulnerability.

The number of floors also has a significant contribution (about 17%), showing that building height affects seismic performance. Peak Ground Acceleration (PGA) contributes around 8%, confirming that ground-motion intensity is still important, although less dominant compared to vulnerability and geometric features in this model.

Elevation has a moderate contribution (around 5%), while VS30 contributes about 3%. Peak Ground Velocity (PGV), S1, and macroseismic intensity each have very small importance values (around 1–2%), indicating that they add limited additional information to the model.

Overall, Model 2 relies mainly on rupture distance, building age, building area, and number of floors for damage classification. Compared to Model 1, the inclusion of building age makes the model more focused on structural vulnerability, which helps improve the accuracy and reduce confusion between similar damage states.

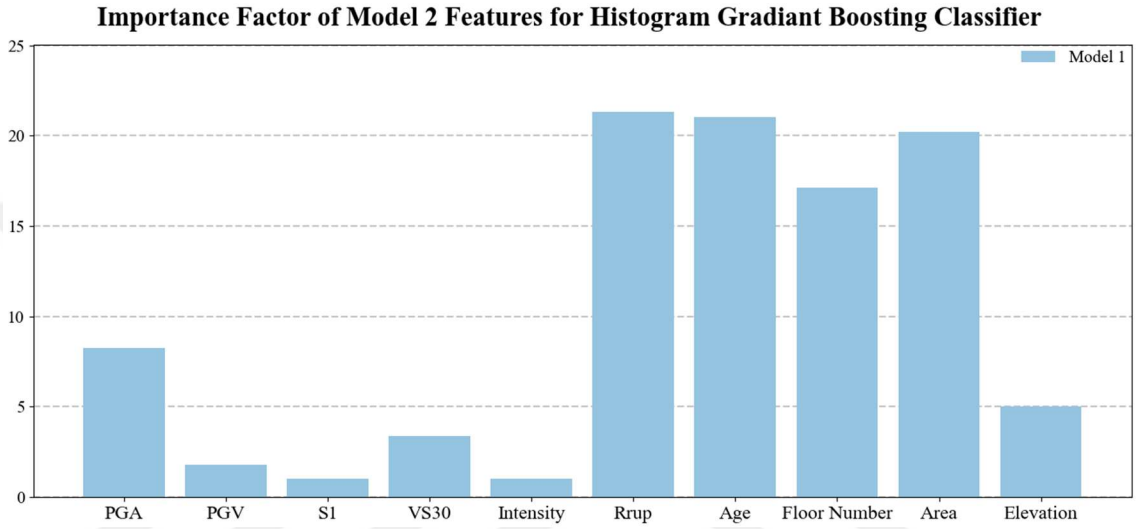

Figure 4.6: Feature Importance Factor of Izmir Case for Model 2.

4.2 Kahramanmaras Earthquake

This section provides a detailed assessment of the machine learning models developed to predict earthquake-related damage in masonry and reinforced concrete buildings, using field observations from the 2023 Kahramanmaras earthquake. The analysis is performed at multiple spatial scales—building, district, and city—to examine both detailed and aggregated prediction performance. Damage classification is explored using two-, three-, and four-class schemes, highlighting the models' ability to distinguish between different severity levels and their limitations in intermediate cases. In addition, feature importance analysis is conducted to identify the key structural, seismic, and site-specific factors that most strongly influence model predictions, providing practical insights for post-earthquake risk assessment.

4.2.1 Accuracy and F1 score

This section compares the performance of the models in terms of accuracy and F1-score for both masonry and reinforced concrete (RC) buildings across different damage classification schemes. A consistent pattern is observed for both structural types: as the number of damage classes increases, overall model performance tends to decline. In the binary classification case, the models achieve their best results, with accuracy values above 70% and F1-scores close to or exceeding 80%. These outcomes demonstrate a strong ability to distinguish between safe and unsafe buildings, which is especially important for rapid post-earthquake decision-making, where the prompt identification of potentially dangerous structures is essential.

When the classification is expanded to three damage classes, a noticeable reduction in both accuracy and F1-score is observed for both masonry and RC buildings. Performance levels converge around 65%, suggesting that the introduction of an intermediate damage category increases ambiguity between classes. While the models remain capable of distinguishing clearly safe and unsafe buildings, confusion becomes more frequent for structures that fall near damage boundaries, such as those requiring further assessment.

The most significant performance degradation occurs under the four-class damage classification scheme. For masonry buildings, accuracy falls below 50%, with a corresponding decline in F1-score, indicating substantial difficulty in differentiating between closely related damage states such as low and moderate damage. RC buildings demonstrate relatively better performance under this scheme, maintaining accuracies slightly above 50%; however, the associated F1-scores still decline, reflecting increased misclassification among intermediate damage levels.

Overall, the results illustrated in (figures 4.7, 4.8, and 4.9) highlight an important trade-off between damage classification granularity and predictive reliability. While multi-class schemes provide more detailed insight into structural damage, they also introduce higher uncertainty and reduced classification confidence. These findings suggest that simpler classification schemes may be more suitable for rapid post-earthquake screening, whereas more detailed classifications may require additional data or advanced modeling strategies to achieve reliable performance.

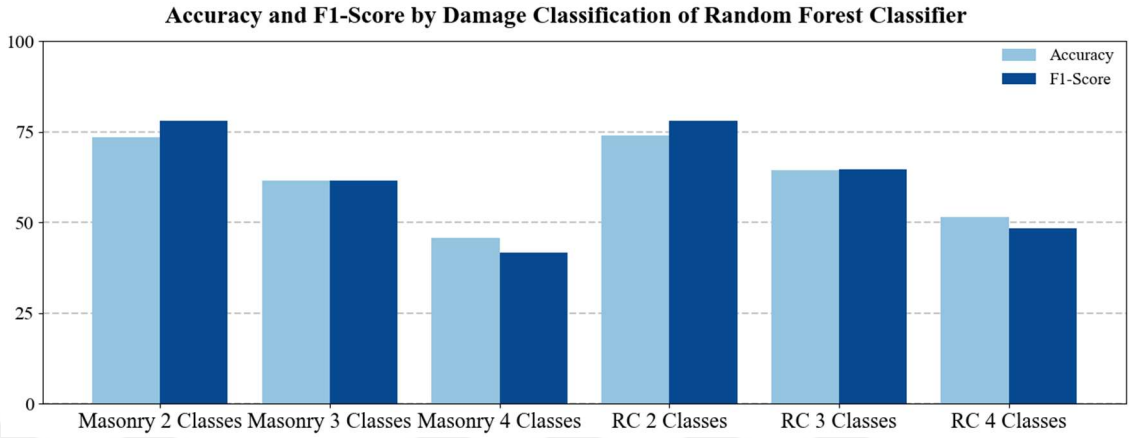

Figure 4.7: Accuracy and F1-Score by Damage Class of Random Forest Classifier.

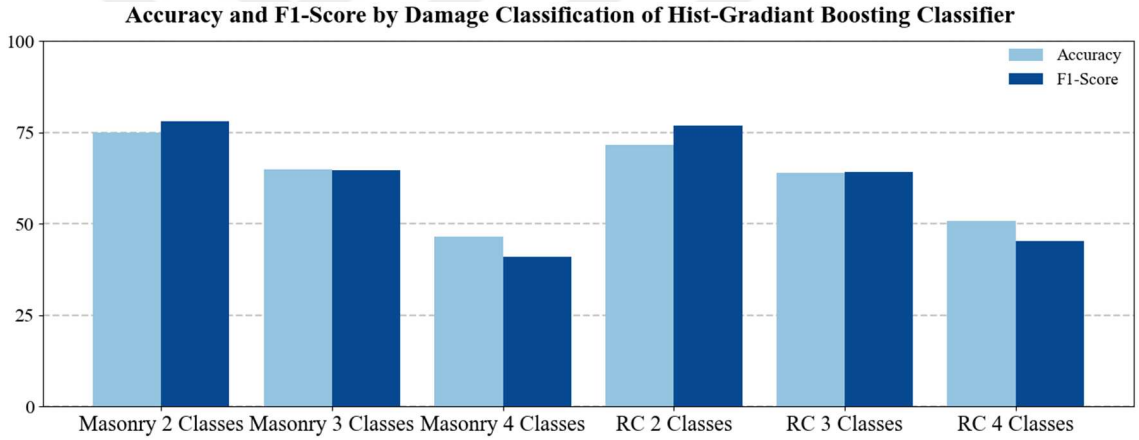

Figure 4.8: Accuracy and F1-Score by Damage Class of Hist-Gradiant Boosting Classifier.

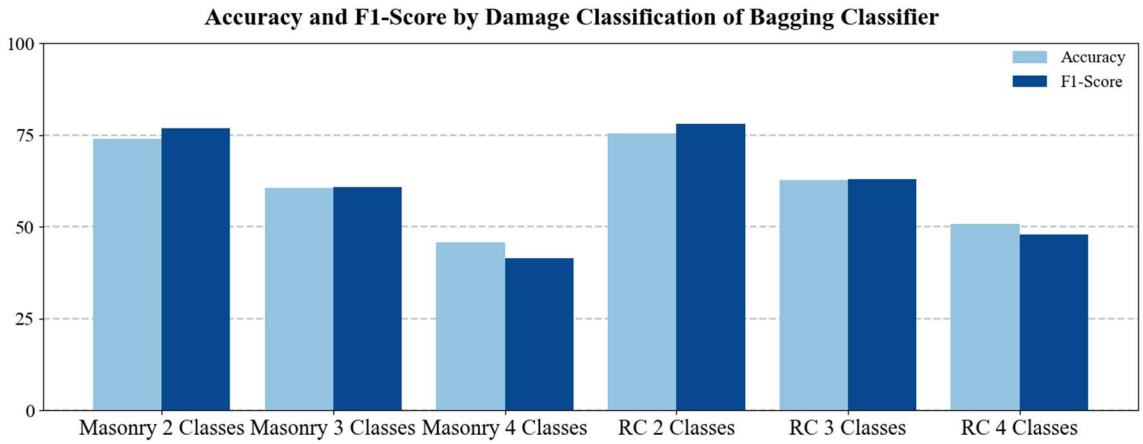

Figure 4.9: Accuracy and F1-Score by Damage Class of Bagging Classifier.

4.2.2 Building scale

The building-scale evaluation was performed by generating a confusion matrix for each case and for every machine learning algorithm. In addition, accuracy, precision, recall, and F1 scores were calculated to provide deeper insight into the model's behavior and to better interpret the overall performance.

4.2.2.1 Masonry buildings' two damage states classification

For the binary damage classification of masonry buildings, the confusion matrix shows a considerable challenge in accurately identifying unsafe structures. Nearly 40% of the buildings labeled as “Unsafe” were mistakenly classified as “Safe,” revealing a critical weakness in the model’s ability to distinguish between severely affected and stable buildings. This level of misclassification is especially concerning in a post-earthquake context, where underestimating unsafe structures could lead to serious safety risks and ineffective allocation of emergency resources. Given this substantial error rate, the binary classification approach proves unreliable for practical decision-making, as it significantly underestimates the true extent of damage and may result in misleading or overly optimistic assessments. Across the three machine learning models, the classification performance for masonry buildings in two damage states is generally strong, with all models showing high precision but moderate recall when identifying unsafe structures. Random Forest achieves an overall accuracy of 0.74 and a macro-F1 of 0.73. Histogram Gradient Boosting performs similarly, and the Bagging Classifier also follows a comparable trend. (Figure 4.10) shows that Random Forest and Histogram Gradient Boosting provide similar recall for unsafe buildings (0.57–0.58), while the Bagging Classifier performs slightly better with a recall of 0.61, indicating superior sensitivity to hazardous cases. Precision for unsafe predictions remains high across all models (0.81–0.86), showing consistent reliability when labeling buildings as unsafe. For the safe class, recall is highest for Histogram Gradient Boosting (0.91), followed by Random Forest (0.90) and Bagging (0.86). Overall, while all models confidently detect safe structures, they miss a notable portion of unsafe ones, with Bagging offering the strongest unsafe detection capability.

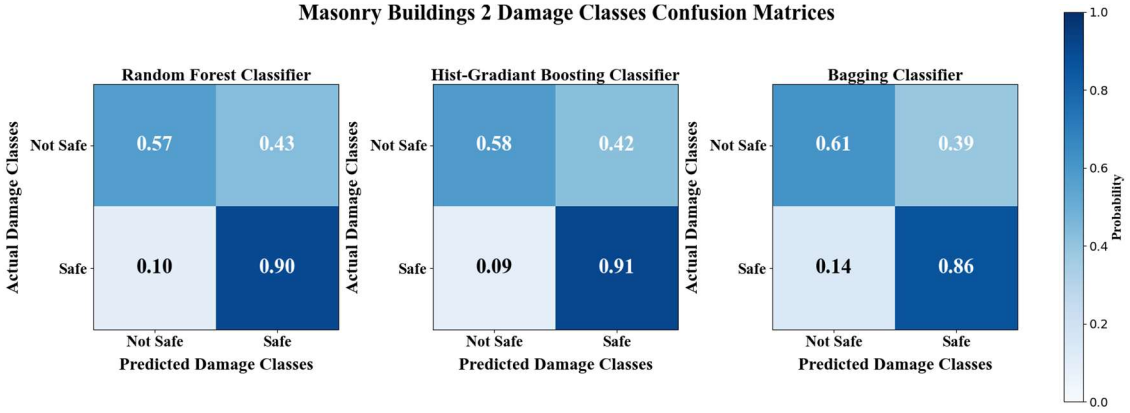

Figure 4.10: Confusion Matrices of Different Models for Masonry Buildings 2 Damage Classes.

4.2.2.2 Masonry buildings' three damage states classification

In the three-class damage classification scenario, the confusion matrix offers a more detailed and balanced understanding of the models' ability to differentiate between varying levels of seismic risk. The findings indicate that approximately 10% of the buildings that were actually classified as "Unsafe" were incorrectly predicted as "Safe," representing a significant improvement over the results obtained from the binary classification approach. A larger portion -around 30 - of unsafe buildings was classified as requiring "Further Assessment," meaning that although these structures were not directly identified as unsafe, they were still flagged for closer inspection rather than overlooked entirely.

For buildings that were actually "Safe," the models performed consistently well, with more than 75% correctly identified, especially when using the Histogram-Based Gradient Boosting classifier. This reliability in recognizing non-critical structures is important because it helps reduce unnecessary inspections and allows emergency teams to focus their efforts where they are needed most.

Overall, the three-state classification approach proves to be a practical and effective way to support post-earthquake decision-making. By keeping misclassifications between "Safe" and "Unsafe" buildings relatively low, the method offers a more trustworthy basis for prioritizing field inspections, allocating resources, and guiding rapid response operations in the aftermath of a major seismic event.

(Figure 4.11) demonstrates that all models are strong in performance in identifying unsafe and safe buildings compared to the intermediate "Further Assessment" class. For Random Forest, the overall accuracy is 0.62 with a macro-F1 score of 0.62, reflecting the increased confusion introduced by the third damage category. Random Forest yields recall values of 0.63 for unsafe and 0.72 for safe buildings, while the intermediate class remains more difficult to detect (0.50). Histogram Gradient Boosting improves recall for all classes—0.68 for unsafe, 0.51 for further assessment, and 0.76 for safe—and also achieves stronger precision, particularly for unsafe and safe categories. The Bagging Classifier performs similarly but with slightly lower recall for safe buildings (0.65). Overall, Histogram Gradient Boosting provides the most balanced precision–recall performance, while all models show considerable confusion around the intermediate damage state.

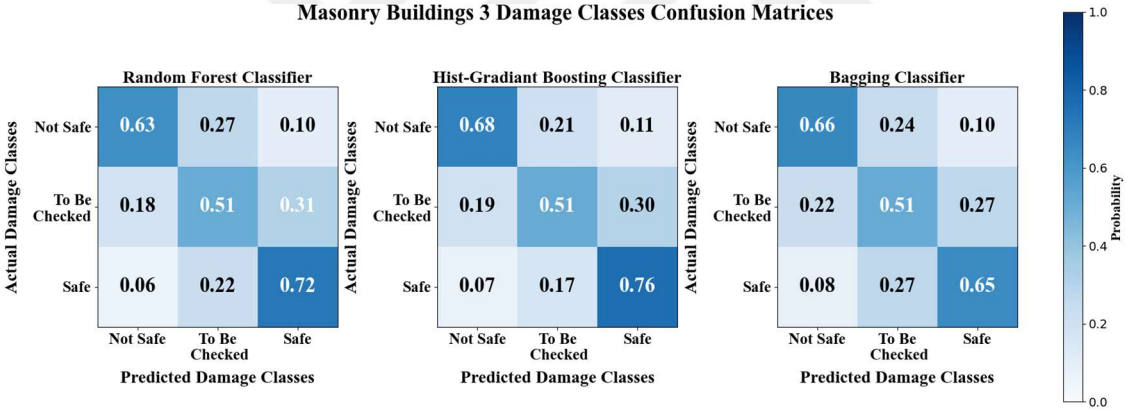

Figure 4.11: Confusion Matrices of Different Models for Masonry Buildings 3 Damage Classes.

4.2.2.3 Masonry buildings' four damage states classification

For the four-class damage classification of masonry structures, the confusion matrix reveals that most misclassifications tend to occur between damage states that are adjacent in severity. For example, roughly one-quarter of buildings categorized as "Severe-Damaged" were mistakenly predicted as "Low-Damaged," highlighting the challenge of distinguishing between the higher levels of structural damage. Similarly, a notable portion of buildings that were actually "Non-Damaged" were misclassified as "Low-Damaged," suggesting that the model sometimes struggles to recognize structures that are truly intact versus those experiencing minor damage. On a positive note, instances where the model

confused the most extreme categories - The “Severely Damaged” and “Non-Damaged” categories were relatively rare, each representing less than 10% of the total cases. This distribution suggests that although the model may experience some difficulty in identifying intermediate damage states, it performs effectively in separating the least damaged buildings from those that are most severely affected, particularly for masonry structures.

The three models shown in Figure 4.12 show clear strengths in identifying the extreme classes—Severe Damage (SD) and No Damage (ND)—but struggle with the intermediate Moderate Damage (MD) and Light Damage (LD) categories. Random Forest achieves an accuracy of 0.45 and a macro-F1 score of 0.41, indicating that classification difficulty increases considerably when moving from two or three to four categories. All models achieve strong recall for SD (0.63–0.65) and ND (0.66–0.77), indicating stable detection of fully damaged or undamaged buildings. However, recall for MD is critically low across all models (0.03–0.06), demonstrating substantial difficulty distinguishing this class. Light damage performs moderately better, with a recall of around 0.43–0.46. Histogram Gradient Boosting provides marginal improvements in ND recall (0.77) and LD precision, but overall, all three models exhibit significant overlaps among middle damage states while maintaining robust performance for the most and least damaged buildings.

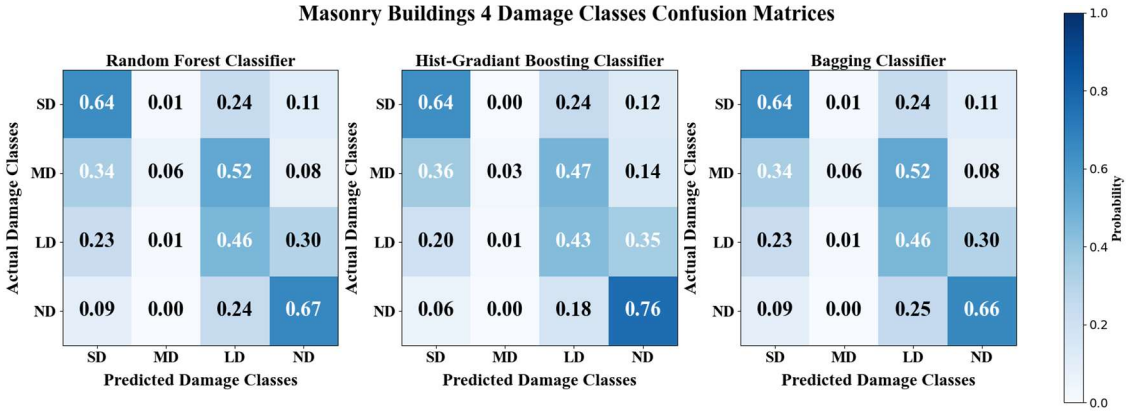

Figure 4.12: Confusion Matrices of Different Models for Masonry Buildings 4 Damage Classes.

4.2.2.4 Reinforced concrete frame structures: two damage states classification

For reinforced concrete (RC) frame structures categorized into just two damage states, the confusion matrix highlights a concerning pattern. Between 40% and 50% of buildings that were actually "Unsafe" were incorrectly predicted as "Safe," indicating that the model struggles to reliably detect structures at serious risk of collapse. This high rate of misclassification raises questions about the model's sensitivity to severe structural damage and suggests that relying on this simplified two-class system could be risky. In practical terms, using such a binary classification could lead to significant underestimation of unsafe buildings, which in turn may compromise post-earthquake risk assessments and decision-making. Essentially, this approach could provide a false sense of security, potentially putting people and resources in danger by overlooking buildings that need urgent attention. All models perform strongly overall, as demonstrated in (Figure 4.13). Random Forest achieves an accuracy of 0.74 and a macro-F1 of 0.73, consistent with its balanced behavior across the two categories. All models show consistently high precision for unsafe predictions (0.87–0.91). Bagging achieves the highest recall for unsafe buildings (0.60), outperforming Random Forest (0.54) and Histogram Gradient Boosting (0.48), making it the most effective at capturing hazardous cases. At the same time, Histogram Gradient Boosting achieves the strongest recall for safe buildings (0.95), followed closely by Random Forest (0.94) and Bagging (0.91). While all models are highly reliable when predicting that a building is unsafe, their lower unsafe recall values indicate that some unsafe RC buildings remain misidentified as safe—an issue reduced but not eliminated by the Bagging Classifier.

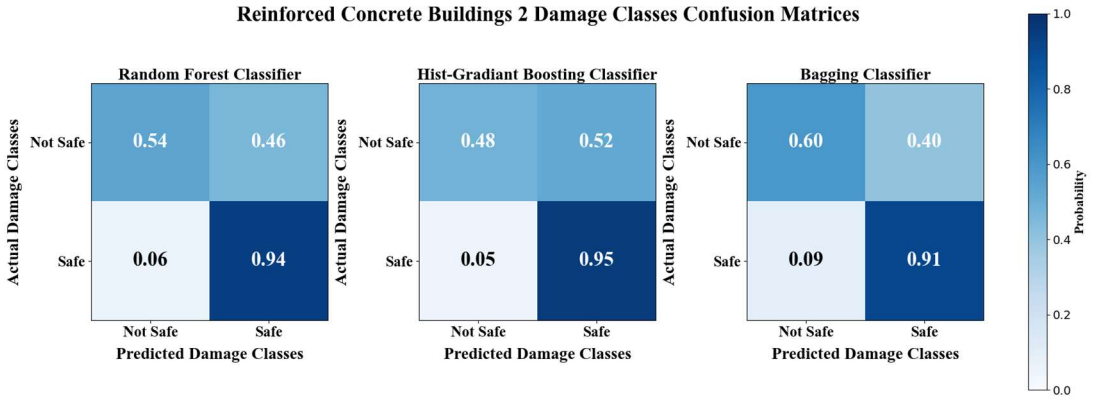

Figure 4.13: Confusion Matrices of Different Models for RC Buildings 2 Damage Classes.

4.2.2.5 Reinforced concrete frame structures: three damage states classification

When considering reinforced concrete (RC) buildings with a three-class damage scheme, the confusion matrix reveals that the model performs much better at distinguishing between safe and unsafe structures. Fewer than 10% of buildings that were actually "Unsafe" were mistakenly classified as "Safe," while about 35% were flagged as needing "Further Assessment," reflecting some caution in borderline cases. The model was particularly strong in identifying "Safe" buildings, correctly classifying over three-quarters of them. Among the tested approaches, the Histogram-Based Gradient Boosting model achieved the most stable and dependable performance among the evaluated methods. This three-class system strikes a good balance: it minimizes the risk of overlooking truly unsafe buildings while still providing a practical way to identify structures that may require closer inspection. Overall, this classification approach is well-suited for post-earthquake risk assessments, as demonstrated in (Figure 4.14), helping authorities make informed and safer decisions by clearly distinguishing between buildings that are safe, potentially unsafe, or in need of further evaluation. It shows balanced but imperfect performance. Random Forest achieves an overall accuracy of 0.64 and a macro-F1 of 0.65, reflecting moderate but stable predictive capability across all classes. Random Forest provides moderate recall for unsafe (0.58), further assessment (0.62), and safe (0.73) categories, though precision for the intermediate class remains limited. Histogram Gradient Boosting improves recall for safe buildings (0.77) and maintains strong precision for unsafe and safe predictions, overall delivering the most consistent accuracy across classes. Bagging shows similar recall for unsafe (0.59) and further assessment (0.61) categories, though with slightly lower precision compared to HGB. In general, all models perform reliably for extreme classes, while the intermediate "Further Assessment" state continues to introduce classification uncertainty.

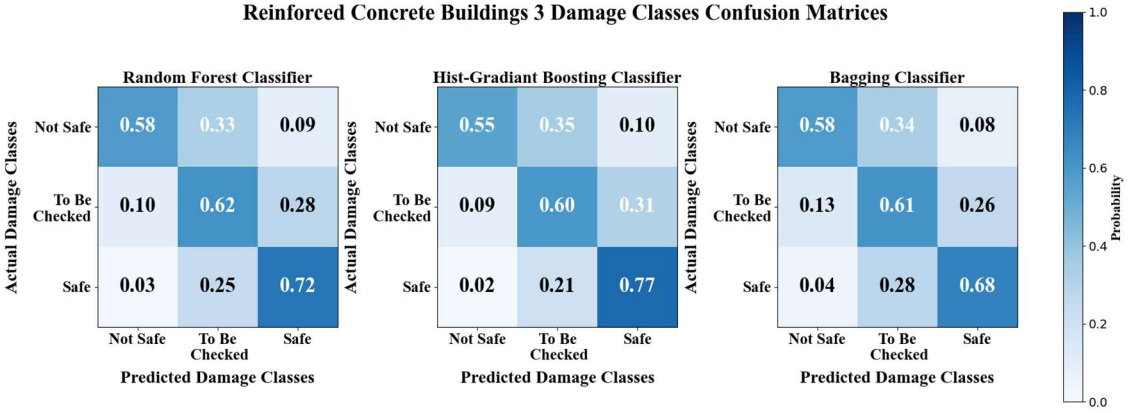

Figure 4.14: Confusion Matrices of Different Models for RC Buildings 3 Damage Classes.

4.2.2.6 Reinforced concrete frame structures: four damage states classification

When the classification scheme for reinforced concrete (RC) buildings was extended to four damage categories, the model showed a noticeable increase in misclassifications between neighboring damage levels. According to the confusion matrix, roughly 30% of buildings that were actually "Severely Damaged" were mistakenly predicted as only "Low-Damaged," highlighting the challenge of accurately identifying the most critical structural conditions. Similarly, about 20% of buildings that were truly "Non-Damaged" were incorrectly classified as "Low-Damaged," suggesting that the model sometimes interprets minor imperfections as more significant damage. On the positive side, extreme misclassifications—where completely intact buildings are confused with severely damaged ones—were relatively rare, occurring in fewer than 10% of cases. This pattern indicates that while the model occasionally struggles with intermediate damage levels, it remains generally reliable at identifying buildings that are either fully safe or at serious risk, which is crucial for prioritizing inspections and post-earthquake interventions.

The models consistently demonstrate strong recall for severe damage (0.59–0.61), light damage (0.59–0.60), and no damage (0.69–0.80). Random Forest achieves an accuracy of 0.49 and a macro-F1 score of 0.43, reflecting the increasing complexity of four-category prediction. Moderate damage remains the most challenging category, with very low recall across all models (0.05–0.14), indicating frequent misclassification into neighboring classes. Histogram Gradient Boosting achieves the highest no-damage recall (0.80), while Bagging and Random Forest yield more balanced but slightly lower values. Precision

values show similar trends, with high confidence in SD and ND predictions but lower discriminative ability for MD and LD cases. Overall, while the models accurately identify the most and least damaged RC buildings, they show reduced reliability for intermediate structural states as presented in Figure 4.15).

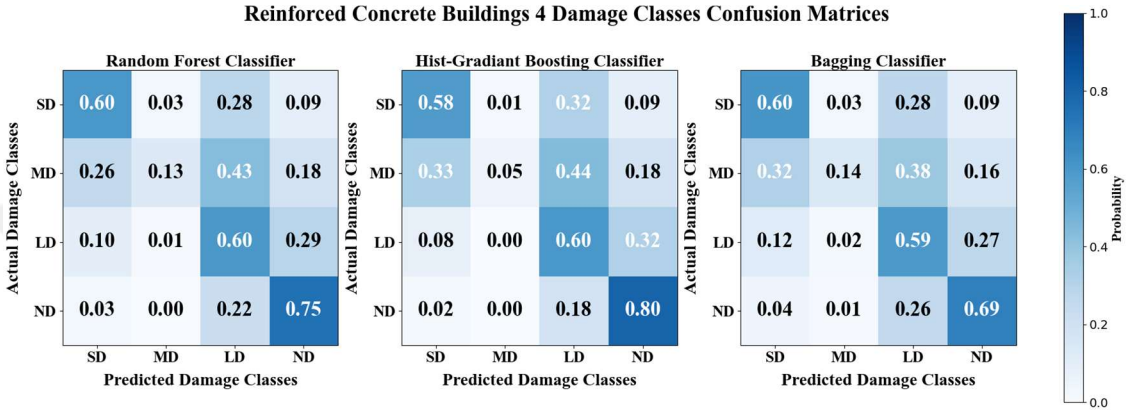

Figure 4.15: Confusion Matrices of Different Models for RC Buildings 4 Damage Classes.

4.2.3 District scale

The The district-level evaluation is performed to examine the effectiveness of the proposed machine learning framework at an intermediate spatial scale, serving as a bridge between detailed building-based predictions and large-scale city-level analysis. At this scale, individual building damage predictions were aggregated to characterize damage patterns at the district level. To ensure a representative evaluation, ten districts were randomly selected from the approximately one hundred districts within the study area. The results indicate that the model is capable of effectively reproducing the distribution of damage classes across different districts, demonstrating consistency between predicted and observed damage patterns. This analysis highlights the model's robustness in capturing spatial variability in earthquake-induced damage and underscores its potential for supporting district-level rapid damage assessment and post-earthquake decision-making.

4.2.3.1 Masonry buildings: two damage states classification

For masonry buildings classified into two damage levels, the model demonstrated reasonably strong predictive performance at the district scale. The predicted number of buildings within each damage category showed close agreement with the observed damage counts across the randomly selected districts, as shown in Figure 4.16. This consistency indicates that the simplified two-class damage scheme allows for clearer class separation and more reliable representation of district-level damage patterns for masonry structures.

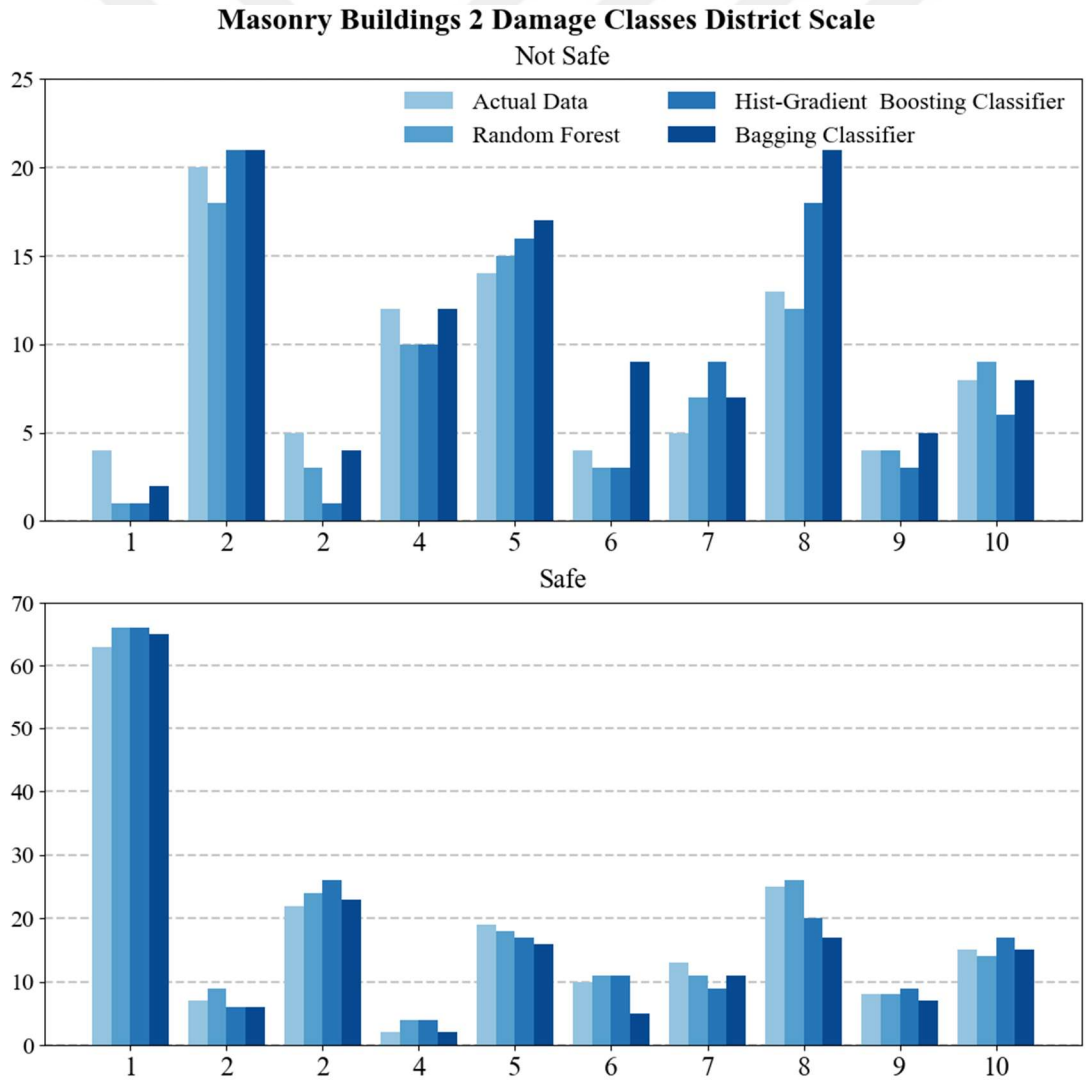

Figure 4.16: Comparison between Different Models' Predictions for Masonry Structures 2 Damage Classes in District Scale.

4.2.3.2 Masonry buildings' three damage states classification

In the three-category damage classification setting, the model showed a strong overall capability to estimate the actual distribution of masonry buildings across the different damage classes at the district scale. The predicted distributions closely followed the observed damage patterns, indicating satisfactory overall performance. Nevertheless, in some districts, the “Need Further Assessment” category was slightly overestimated, suggesting moderate challenges in clearly distinguishing this intermediate damage state from the adjacent classes, as shown in Figure 4.17. These discrepancies highlight the inherent difficulty of separating borderline damage levels in masonry structures, particularly when damage characteristics overlap.

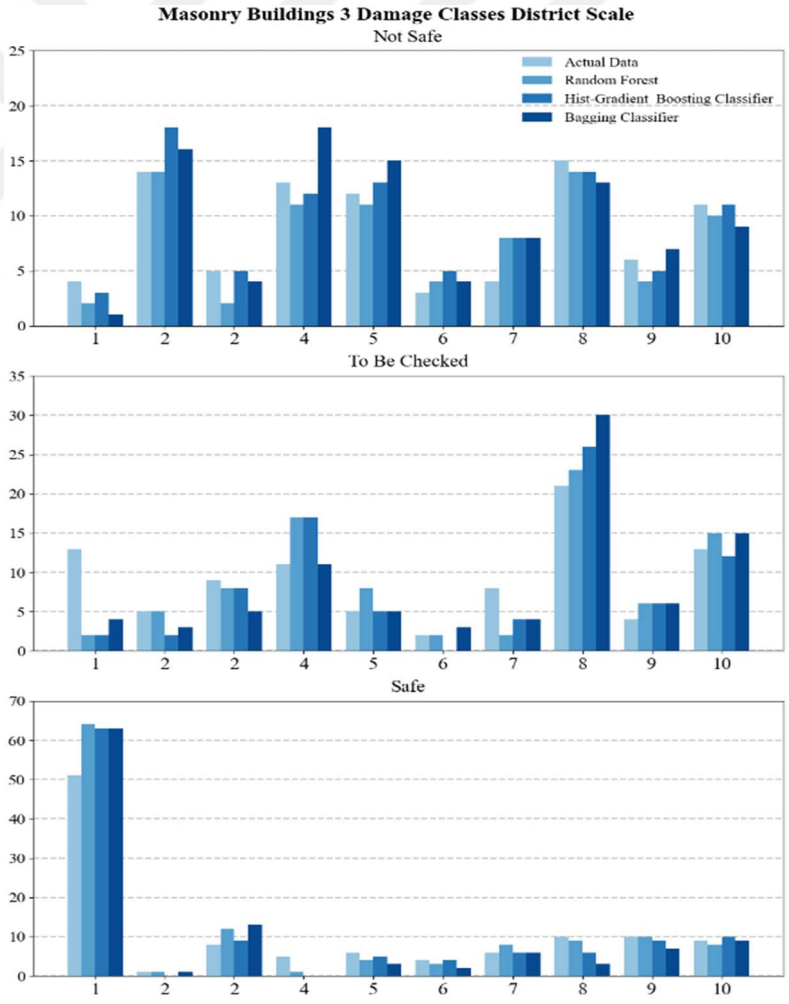

Figure 4.17: Comparison between Different Models Predictions for Masonry Structures 3 Damage Classes in District Scale.

4.2.3.3 Masonry buildings' four damage states classification

In the four-class damage classification case for masonry buildings, the model achieved reasonable accuracy in estimating the number of structures identified as “Severely Damaged” and “No-Damage” at the district level, demonstrating dependable detection of the most critical damage category. However, some discrepancies were observed between the “Moderate-Damaged” and “Low-Damaged” classes, where occasional misclassifications occurred, as shown in Figure 4.18. This overlap indicates similarities in the underlying damage features and highlights the greater difficulty associated with more detailed damage classifications, especially when separating slight damage from undamaged conditions.

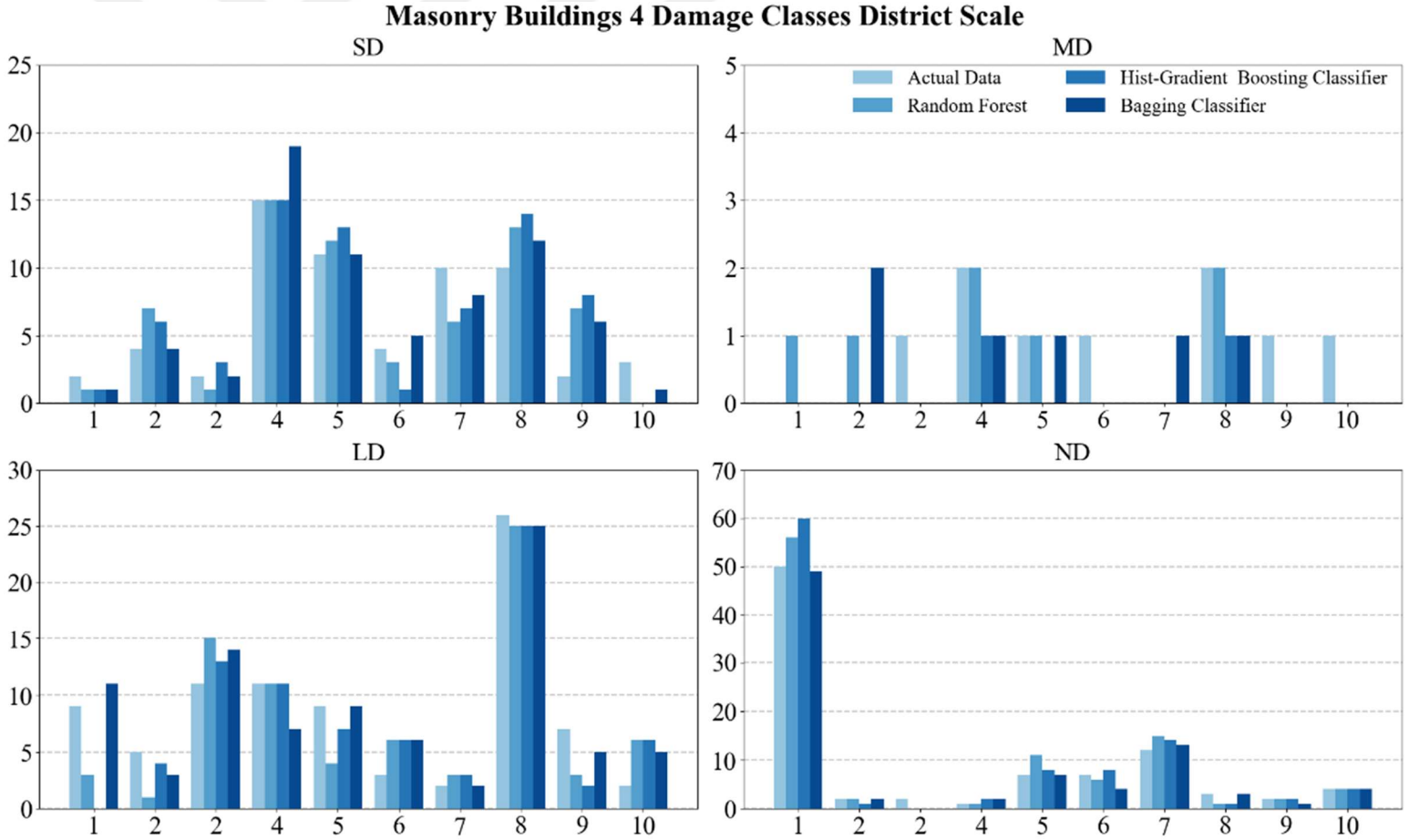

Figure 4.18: Comparison between Different Models Predictions for Masonry Structures 4 Damage Classes in District Scale.

4.2.3.4 Reinforced concrete frame structures: two damage states classification

In the binary damage classification of reinforced concrete (RC) buildings, the model produced accurate predictions in most districts, with predicted damage counts closely matching the observed data, as shown in Figure 4.19. However, notable discrepancies were identified in a few districts, reflecting variability in model performance across the study area. These differences suggest that local site conditions and structural characteristics influence the model's ability to generalize, even within a simplified two-class damage framework.

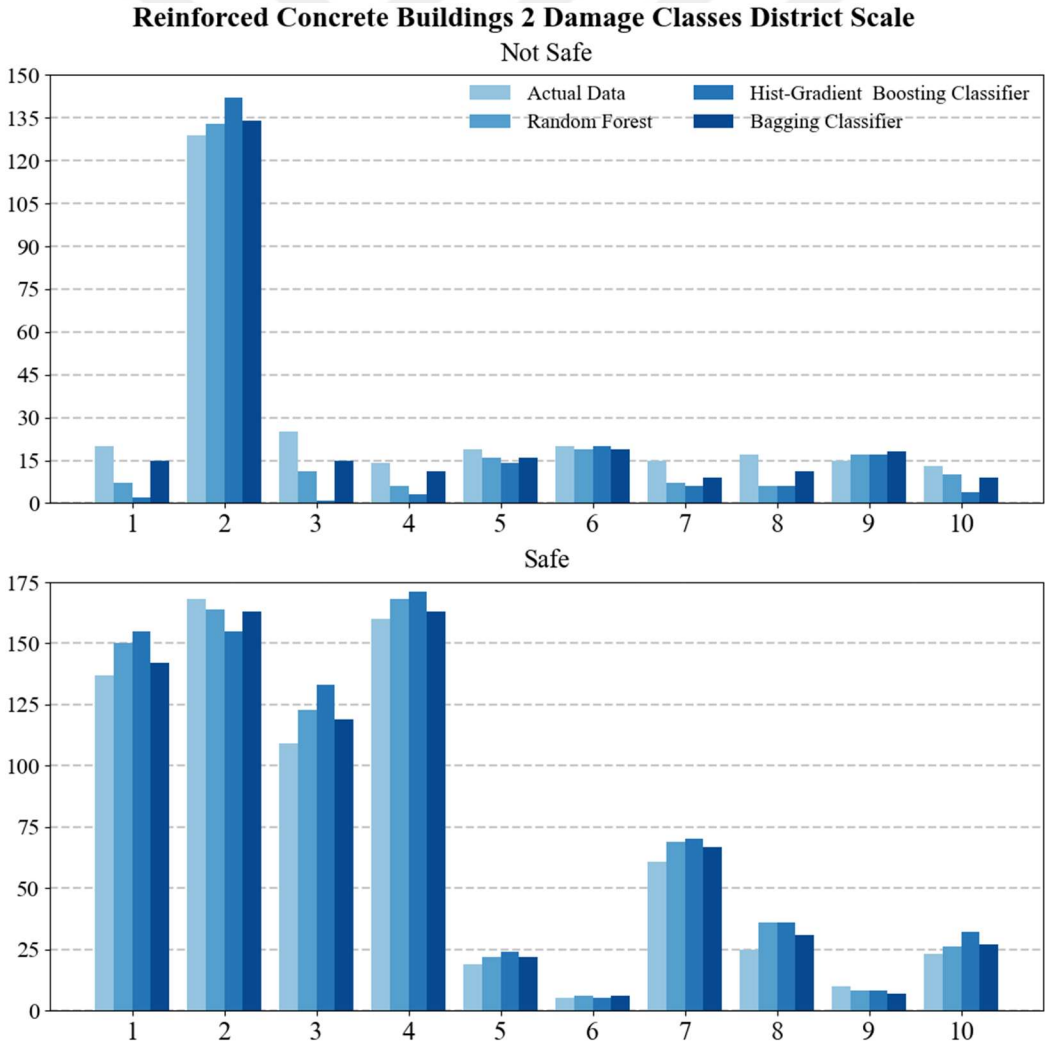

Figure 4.19: Comparison between Different Models' Predictions for RC Structures 2 Damage Classes in District Scale.

4.2.3.5 Reinforced concrete frame structures: three damage states classification

(Figure 4.20) The three-class damage classification results for reinforced concrete (RC) buildings demonstrate reasonable accuracy at the district scale, especially by the Bagging Classifier. The model showed some difficulty in clearly separating the three damage categories, leading to noticeable overlap and confusion among damage states across several districts. This behavior suggests that the damage characteristics of RC buildings exhibit greater complexity and similarity between adjacent classes, making intermediate damage levels more challenging to distinguish using the available input features.

4.2.3.6 Reinforced concrete frame structures: four damage states classification

In the four-class damage classification scenario for reinforced concrete (RC) buildings, the model encountered notable difficulties in accurately distinguishing between damage levels at the district scale. In particular, the “Low-Damaged” category was frequently overestimated, resulting in a distorted distribution of predicted building counts when compared to the observed data, as shown in Figure 4.21. This tendency indicates increased confusion among adjacent damage states and highlights the limitations of finer damage stratification for RC structures under the given feature set and data availability.

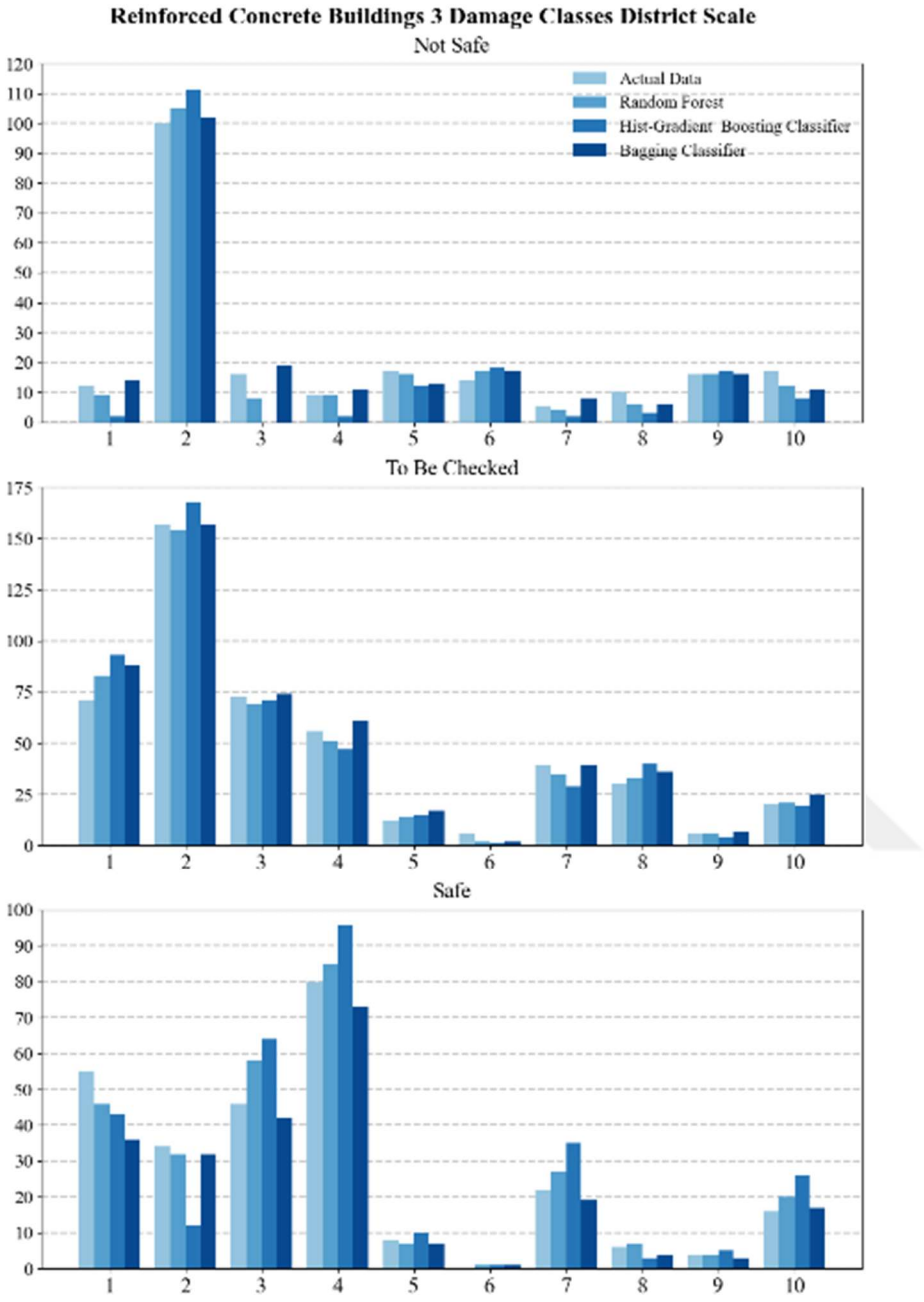

Figure 4.20: Comparison between Different Models Predictions for RC Structures 3 Damage Classes in District Scale.

.

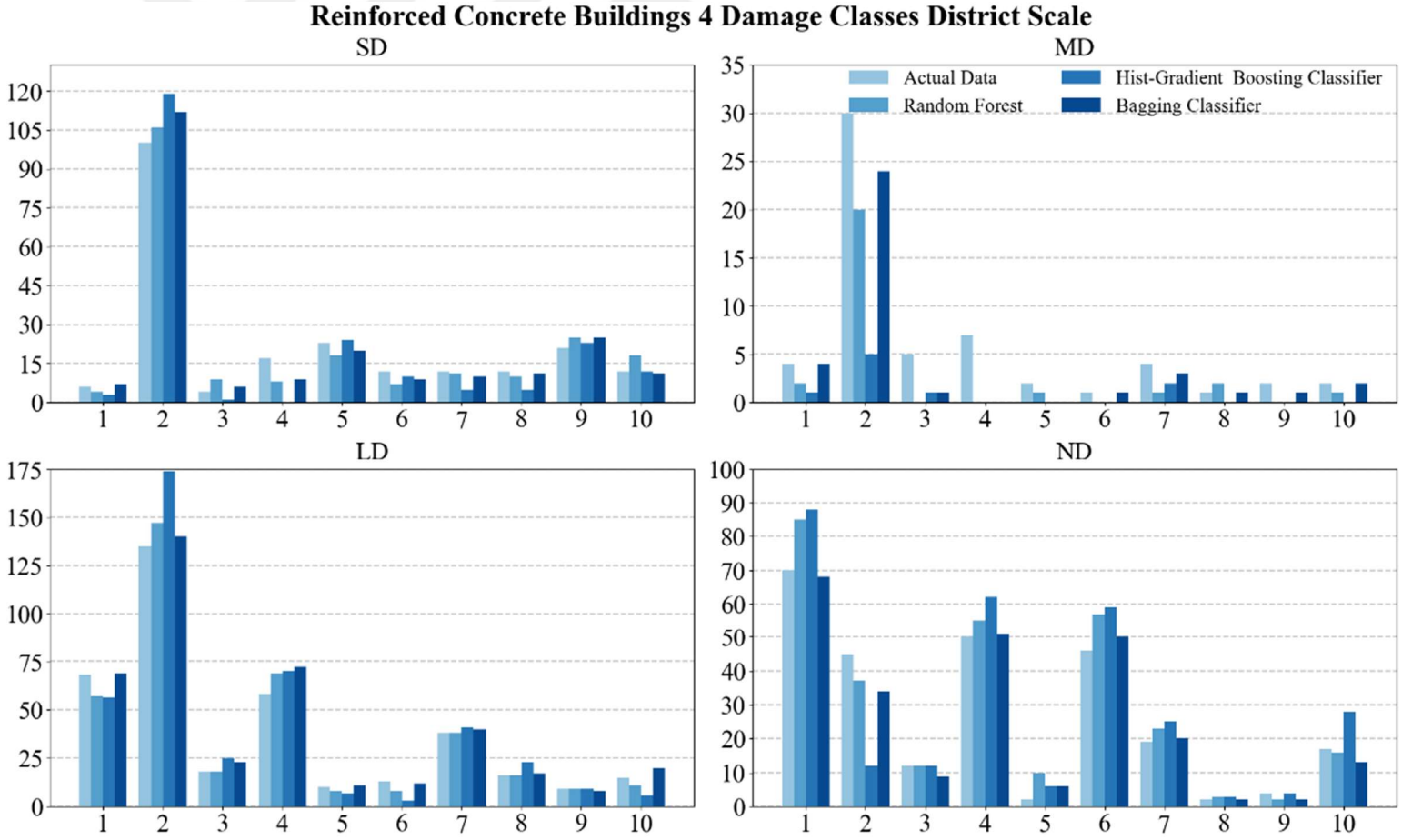

Figure 4.21: Comparison between Different Models Predictions for RC Structures 4 Damage Classes in District Scale

4.2.4 City-wide scale

The city-level evaluation broadens the performance assessment by consolidating damage predictions across the entire validation dataset, allowing for a thorough analysis at the urban scale. Like the district-level approach, this method compares the predicted and actual distributions of damage classes; however, it does so for the complete study area instead of limiting the analysis to specific districts or subsets. This large-scale aggregation provides insight into the model's overall ability to capture city-wide damage trends and dominant damage patterns. By evaluating the predicted damage distribution against the actual post-earthquake observations, the analysis highlights the consistency, robustness, and potential limitations of the proposed framework when applied to urban-level rapid damage assessment.

4.2.4.1 Masonry buildings: two damage states classification

At the city scale, the model demonstrated high predictive accuracy for masonry buildings under the binary damage classification scheme. The predicted number of buildings in each damage category closely matched the actual distribution observed in the validation dataset (as shown in Figure 4.22), indicating that the simplified two-class framework provides a reliable representation of overall urban-level damage patterns.

4.2.4.2 Masonry buildings' three damage states classification

In the three-category damage classification case for masonry buildings, the model demonstrated high predictive capability at the city scale. As illustrated in Figure 4.23, the discrepancies between the predicted and observed numbers of buildings in each damage class were small, confirming the model's reliability in representing the overall urban damage distribution.

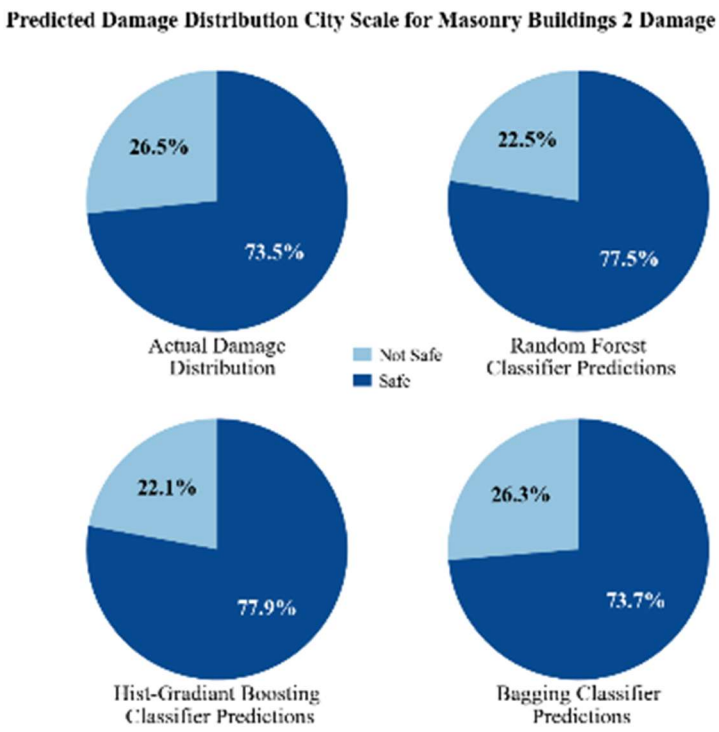

Figure 4.22: Accuracy Comparison Between Different Models for Masonry Structures 2 Damage Classes.

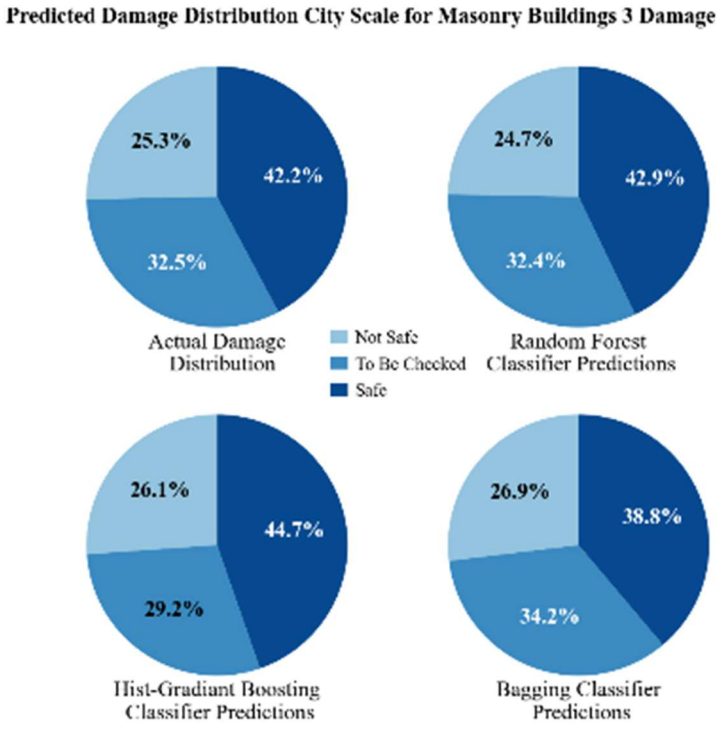

Figure 4.23: Accuracy Comparison Between Different Models for Masonry Structures 3 Damage Classes.

4.2.4.3 Masonry buildings' four damage states classification

At the urban scale, the predicted results for masonry buildings in the four-class damage scheme showed strong consistency with the actual observations, especially for the "High-Damaged" category. The model, as shown in Figure 4.24, performed well, where it captured the overall trends in damage distribution across the urban area, highlighting its potential utility for large-scale post-earthquake damage assessment and decision-making.

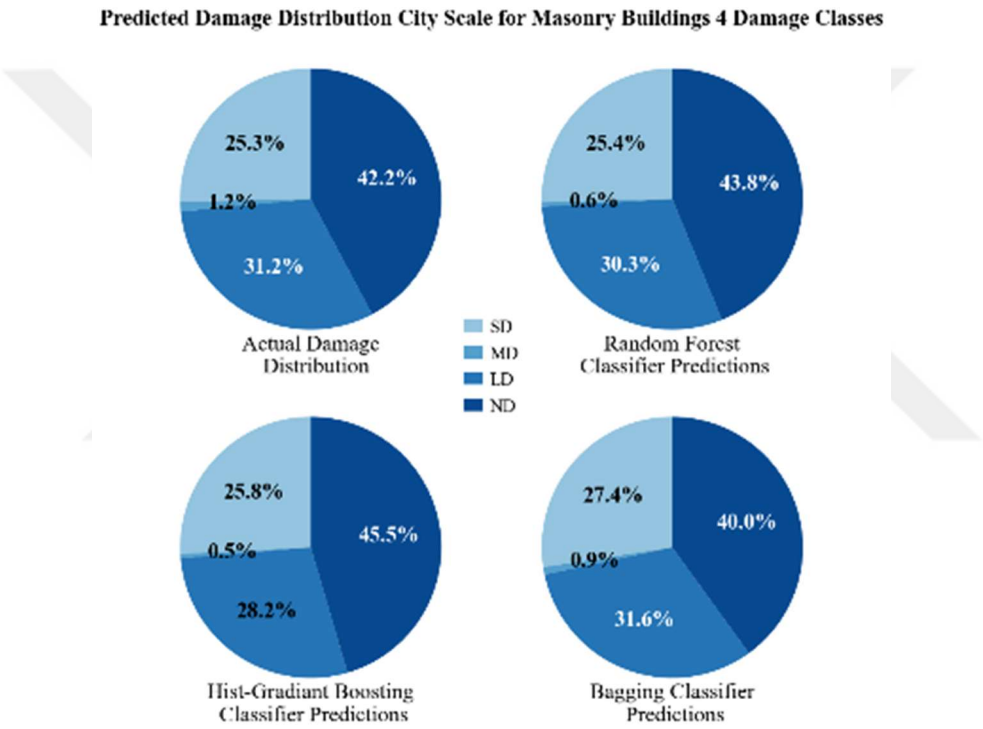

Figure 4.24: Accuracy Comparison Between Different Models for Masonry Structures 4 Damage Classes.

4.2.4.4 Reinforced concrete frame structures: two damage states classification

For reinforced concrete (RC) buildings classified into two damage categories, the model produced accurate predictions at the city-wide level. Specifically, 590 buildings were predicted as "Not-Safe," closely matching the 622 "Not-Safe" buildings observed in the validation dataset. This close agreement demonstrates the model's ability to reliably represent overall urban damage patterns for RC structures under a simplified two-class framework as presented in (Figure 4.25).

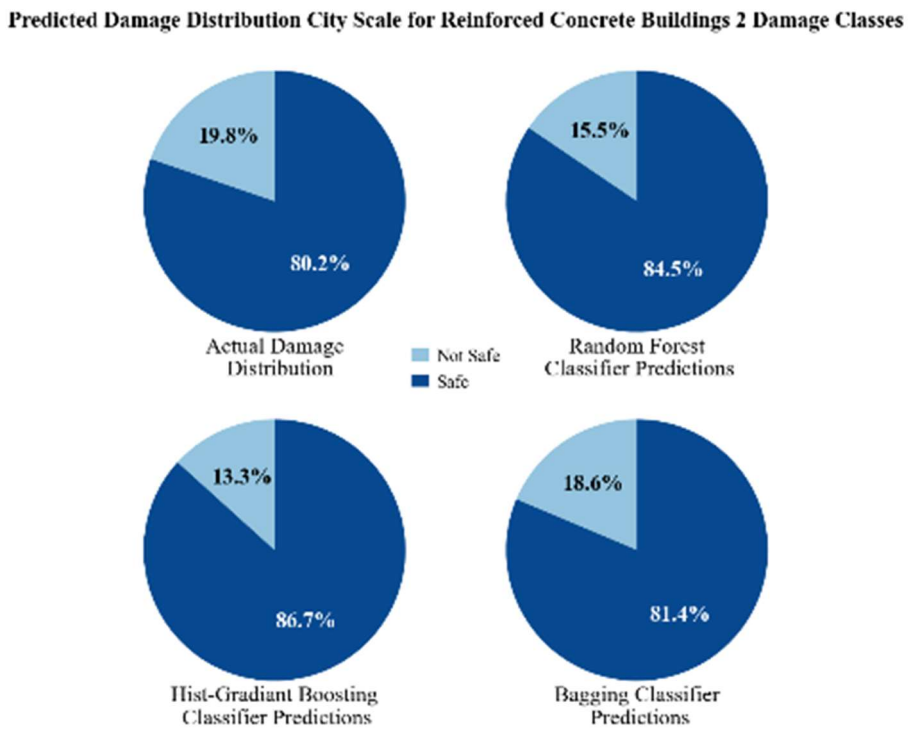

Figure 4.25: Accuracy Comparison Between Different Models for RC Structures 2 Damage Classes.

4.2.4.5 Reinforced concrete frame structures: three damage states classification

In the three-class scenario, the model demonstrated high predictive accuracy at the city-wide level, particularly for the "Not-Safe" category (Figure 4.26), which shows that the predicted building count closely matched the observed number in the validation dataset. This outcome highlights the model's effectiveness for large-scale operational damage assessment of RC structures.

4.2.4.6 Reinforced concrete frame structures: four damage states classification

In the four-class damage classification scenario for RC buildings, the model demonstrated strong performance, particularly in accurately estimating the number of buildings in the "High-Damaged" category. (Figure 4.27) The results indicate the model's ability to capture detailed structural damage distributions across the city, highlighting its potential for supporting comprehensive urban-scale post-earthquake assessments.

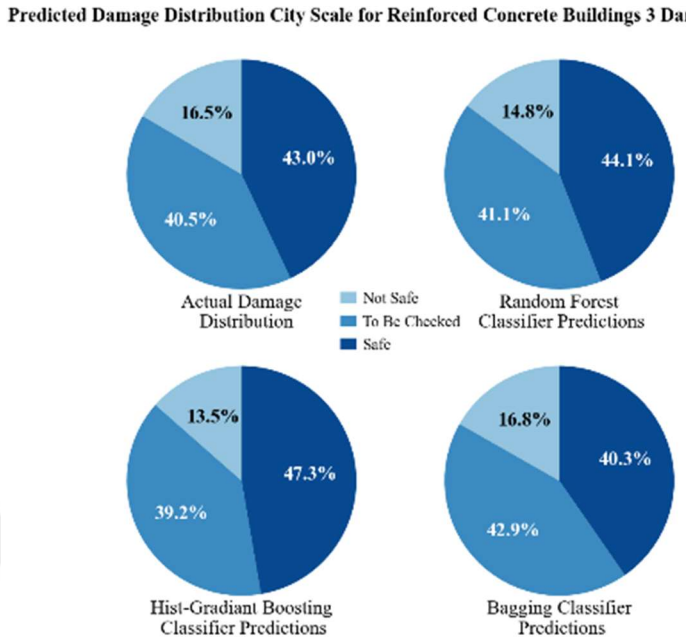

Figure 4.26: Accuracy Comparison Between Different Models for RC Structures 3 Damage Classes.

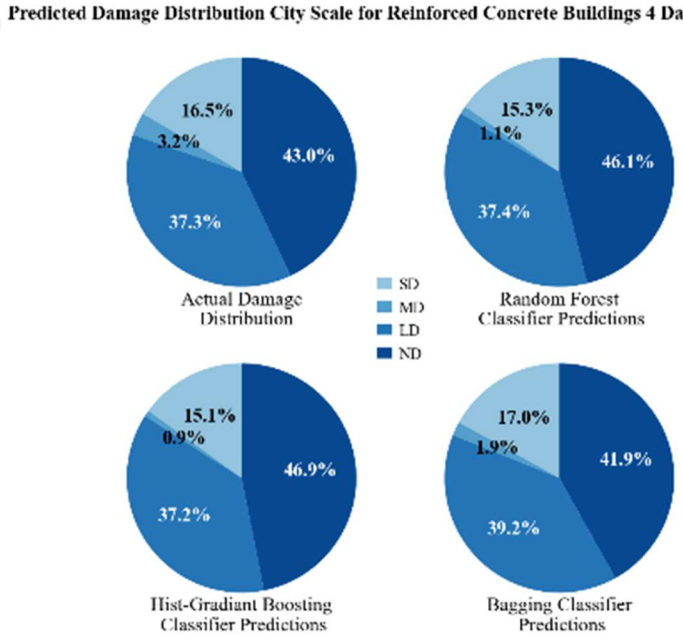

Figure 4.27: Accuracy Comparison Between Different Models for RC Structures 4 Damage Classes.

4.2.5 Features importance analysis:

Feature importance analysis was performed for the three damage-class scenarios for both masonry and reinforced concrete (RC) buildings. The purpose was to determine the key variables influencing model predictions and to better understand which structural, site-related, and ground-motion factors play the most significant role in accurately classifying building damage.

4.2.5.1 Masonry buildings

Feature importance analysis was carried out independently for the three damage-classification schemes for both masonry and reinforced concrete buildings in order to determine the most influential predictors for each model. In the Random Forest approach, ground-motion variables were the primary contributors, followed by building age, elevation, and rupture distance, emphasizing the joint effect of seismic demand and structural attributes. For the Histogram-based model, rupture distance was the leading feature, representing over 36% of the total importance, with building age ranking second. In the Bagging classifier, Peak Ground Acceleration (PGA), building age, elevation, and rupture distance each contributed at similar levels (approximately 15%), indicating that several factors collectively shaped the model's predictions.

Overall, these results emphasize that earthquake-induced damage is governed by a combination of seismic intensity, building vulnerability, and site-specific conditions, and that the relative importance of each factor can vary depending on the model's structure. This insight not only confirms the robustness of the models but also provides practical guidance for identifying buildings and areas at higher risk in post-earthquake scenarios.

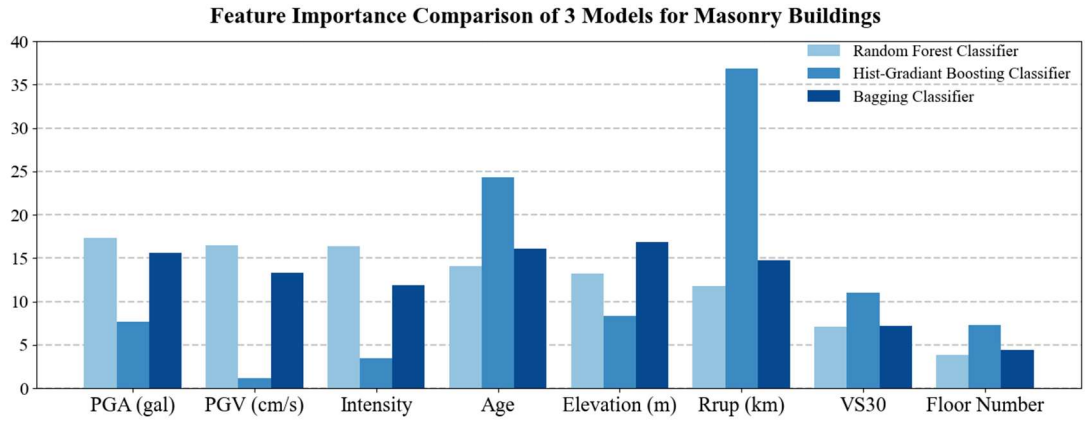

Figure 4.28: Feature Importance Analysis of Masonry Buildings.

4.2.5.2 Reinforced concrete frame structures

For reinforced concrete buildings, the feature importance results showed trends comparable to those observed for masonry structures, while also revealing some model-specific variations. In the Random Forest model, ground-motion variables were the strongest predictors, followed by building age, elevation, and rupture distance, highlighting the interaction between seismic demand and structural susceptibility. In the Histogram-based model, rupture distance was the most influential feature, with building age, VS30, Peak Ground Acceleration (PGA), and elevation also making meaningful contributions to prediction accuracy. For the Bagging classifier, Peak Ground Velocity (PGV) ranked as the top predictor, closely followed by PGA, building age, and elevation, demonstrating that both seismic intensity measures and structural attributes significantly affect damage classification results. These findings demonstrate that, while all models capture the fundamental influence of seismic and structural factors, the relative importance of specific predictors can vary across modeling approaches, providing nuanced insights for urban earthquake risk assessment.

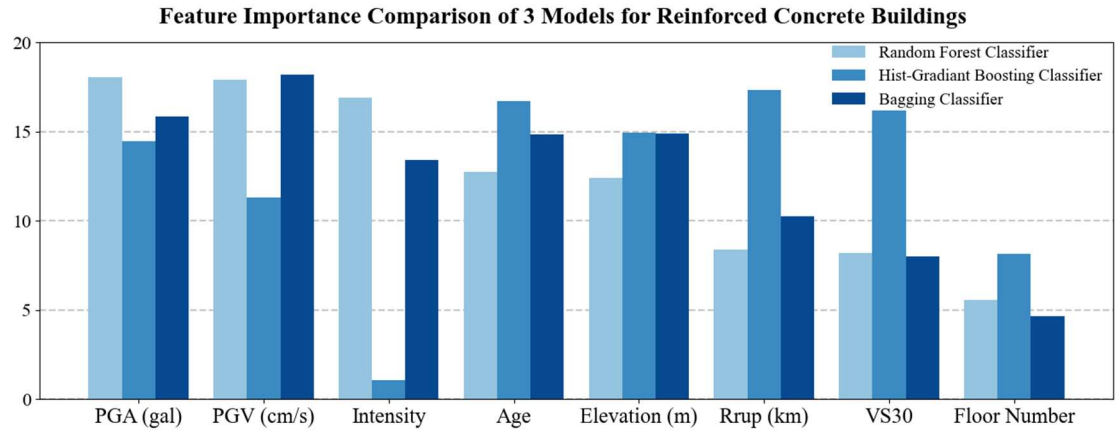

Figure 4.29: Feature Importance Analysis of Reinforced Concrete Buildings.

4.2.6 Comparison between the proposed model and the fragility curve procedure

The performance of the proposed machine learning model was evaluated by comparing its results with those obtained from the conventional fragility curve methodology, a commonly applied approach in seismic damage estimation. The fragility curves were generated using REDAS software, and the damage limit states were defined according to the HAZUS Manual, as previously described in Table 2.2.

The fragility curves were developed based on three seismic demand parameters: Peak Ground Acceleration (PGA), Spectral Displacement (Sd), modified in accordance with the Kandilli Observatory, and macroseismic intensity.

Table 4.1 compares the actual damage distribution of reinforced concrete buildings in Kahramanmaraş with predictions from the Bagging model and HAZUS-based fragility curves for PGA, Sd, and intensity. The Bagging model closely matches the actual distribution, particularly for the ED+CD and SD classes, showing strong agreement with observed damage patterns. In contrast, the HAZUS–PGA and HAZUS–Sd approaches significantly underestimate the most severe damage state (ED+CD) and overestimate moderate damage (MD). The HAZUS–Intensity method provides a better estimate of severe damage than the other fragility-based parameters. However, it still shows noticeable deviation from the actual distribution compared to the machine learning model.

Table 4.1: Comparison between Bagging Machine learning method and Fragility Curve Method for Reinforced Concrete Structures in Kahramanmaras City.

Reinforced Concrete				
Kahramanmaras City	ND	SD	MD	ED + CD
Actual	17%	37%	3%	43%
Bagging	17%	39%	2%	42%
HAZUS PGA	17%	20%	34%	29%
HAZUS Sd	38%	21%	30%	11%
HAZUS Intensity	34%	14%	17%	35%

Table 4.2 presents a comparison between the actual damage distribution of masonry buildings in Kahramanmaraş and the predictions obtained from the Bagging model and HAZUS-based methods. The Bagging model shows very close agreement with the observed data, especially for the SD and ED+CD classes, indicating good predictive performance. In contrast, the HAZUS–PGA approach strongly overestimates the most severe damage state (ED+CD) and underestimates ND and SD classes. The HAZUS–Intensity method provides a more balanced distribution than HAZUS–PGA, but it still deviates noticeably from the actual damage pattern compared to the machine learning model.

Table 4.2: Comparison between Bagging Machine learning method and Fragility Curve Method for Masonry Structures in Kahramanmaras City.

MASONRY				
Kahramanmaras City	ND	SD	MD	ED + CD
Actual	25%	32%	1%	42%
Bagging	27%	32%	1%	40%
HAZUS PGA	12%	9%	20%	59%
HAZUS Intensity	44%	15%	16%	26%

Table 4.3 shows that the Bagging model generally provides a closer match to the actual damage distribution compared to the HAZUS-based approaches, although some deviations are observed in certain damage states. In several cases, Bagging captures the dominant damage class more accurately, particularly for buildings experiencing extensive

or moderate damage. In contrast, the HAZUS–PGA and HAZUS–Sd methods tend to overestimate moderate damage while underestimating the most severe damage states. The HAZUS–Intensity approach sometimes improves the estimation of severe damage, but overall, it still shows larger discrepancies from the observed data compared to the machine learning model.

Table 4.3: Comparison between Bagging Machine learning method and Fragility Curve Method for Reinforced Concrete Structures in Kahramanmaras City's districts.

Reinforced Concrete				
Kahramanmaras City	ND	SD	MD	ED + CD
District 1				
Actual	34%	9%	12%	45%
Bagging	36%	3%	17%	44%
HAZUS PGA	11%	16%	34%	39%
HAZUS Sd	33%	20%	35%	12%
HAZUS Intensity	27%	14%	19%	40%
District 2				
Actual	8%	2%	51%	39%
Bagging	10%	1%	49%	41%
HAZUS PGA	19%	25%	36%	19%
HAZUS Sd	39%	23%	30%	8%
HAZUS Intensity	34%	16%	19%	30%
District 3				
Actual	11%	1%	37%	50%
Bagging	38%	0%	18%	44%
HAZUS PGA	7%	16%	37%	39%
HAZUS Sd	20%	20%	39%	20%
HAZUS Intensity	13%	9%	14%	65%
District 4				
Actual	14%	3%	42%	42%
Bagging	3%	1%	30%	66%
HAZUS PGA	17%	24%	36%	23%
HAZUS Sd	36%	23%	32%	9%
HAZUS Intensity	30%	15%	19%	36%
District 5				
Actual	1%	25%	74%	1%
Bagging	0%	67%	31%	1%
HAZUS PGA	13%	16%	34%	36%
HAZUS Sd	41%	20%	30%	8%
HAZUS Intensity	37%	17%	19%	28%

Table 4.4 shows that the Bagging model generally follows the actual damage distribution of masonry buildings more closely than the HAZUS-based approaches, especially in areas with very high extensive and complete damage (ED+CD). Although some differences appear in the SD class, the machine learning model still captures the dominant damage pattern better than the fragility-based methods. The HAZUS–PGA approach frequently overestimates moderate damage (MD) and, in some cases, underestimates the most severe damage. The HAZUS–Intensity method shows large deviations from the observed data, particularly by overpredicting the non-damaged (ND) class in several cases.

Table 4.4: Comparison between Bagging Machine learning method and Fragility Curve Method for Masonry Structures in Kahramanmaras City's districts

Masonry				
Kahramanmaras City	ND	SD	MD	ED + CD
District 1				
Actual	4%	15%	0%	81%
Bagging	5%	38%	0%	57%
HAZUS PGA	10%	9%	23%	58%
HAZUS Intensity	44%	17%	18%	20%
District 2				
Actual	32%	52%	3%	13%
Bagging	42%	38%	1%	18%
HAZUS PGA	34%	16%	24%	25%
HAZUS Intensity	79%	12%	7%	3%
District 3				
Actual	45%	28%	1%	25%
Bagging	18%	59%	0%	24%
HAZUS PGA	10%	10%	23%	57%
HAZUS Intensity	45%	17%	18%	20%
District 4				
Actual	6%	32%	0%	61%
Bagging	3%	30%	0%	67%
HAZUS PGA	9%	9%	22%	60%
HAZUS Intensity	43%	17%	18%	22%
District 5				
Actual	1%	14%	0%	85%
Bagging	2%	10%	0%	88%
HAZUS PGA	9%	9%	22%	60%
HAZUS Intensity	42%	17%	18%	22%

The fragility curve results at both the city scale and the district scale show noticeable difficulties in accurately estimating the damage distribution. One of the main reasons for this limitation is the difference in damage limit state definitions between the HAZUS methodology and the classification system adopted by the Ministry of Environment, Urbanization, and Climate Change. Although both approaches aim to describe similar physical damage conditions, the thresholds and descriptions of adjacent damage states are not fully consistent. This mismatch leads to overlapping classifications and increases the likelihood of misrepresentation, particularly between neighboring damage levels such as slight and moderate or moderate and extensive damage.

As a result, the fragility-based predictions often shift damage proportions from one class to another, even when the overall damage trend appears reasonable. This issue becomes more evident at the district scale, where local variability further amplifies the sensitivity to classification differences.

In contrast, the machine learning model demonstrates higher accuracy and better alignment with the observed damage data. This improvement can be attributed to the fact that the model was trained directly using real field data labeled according to the official damage categories provided by the Ministry. By learning from the same classification framework used in practice, the machine learning approach is better able to capture the actual distribution of damage states and reduce confusion between adjacent categories.

5. CONCLUSION AND RECOMMENDATIONS

Earthquakes represent one of the most sudden and devastating natural hazards, causing significant risks to both human safety and socio-economic structures. Although earthquakes themselves cannot be prevented, the findings of this thesis confirm that effective disaster management—supported by rapid and reliable damage estimation—can substantially reduce casualties and mitigate social and economic losses. Timely post-earthquake damage assessment is therefore essential for enabling authorities to initiate emergency response actions, prioritize inspections, and allocate resources efficiently. In this context, machine learning has emerged over the past decade as a powerful paradigm for managing large and complex datasets, offering enhanced predictive capabilities by capturing nonlinear, highly interdependent relationships among seismic, structural, and geospatial variables.

This thesis proposed and evaluated a machine-learning framework for post-earthquake building damage prediction, utilizing real field inspection data that reflect authentic damage conditions and observed structural performance. The field-collected data were systematically integrated with ground-motion parameters generated through scenario-based analyses using the REDAS software, providing spatially consistent seismic demand estimates at each building location. In addition, geospatial features derived from multiple open-source databases were incorporated to further enrich the dataset. The proposed framework was evaluated using two major earthquake case studies in Turkiye: the 2020 Izmir earthquake and the 2023 Kahramanmaras earthquake sequence.

The Izmir earthquake provided a well-documented real-world benchmark for assessing the performance of the proposed models. Structural characteristics and observed damage states were obtained through post-earthquake field inspections and combined with seismic and geospatial data, forming a robust dataset for training and evaluating six machine learning algorithms. To examine the impact of feature selection, two dataset

configurations were considered. Model 1 excluded building age, whereas Model 2 incorporated building age as an additional explanatory variable. The results demonstrate that Model 2 consistently outperformed Model 1, achieving prediction accuracies close to 90% when using Random Forest and Histogram Gradient Boosting classifiers, compared to approximately 80% for Model 1.

Confusion matrix analysis further underscores the importance of building age in damage prediction. Model 2 exhibited a strong capability to correctly identify severely damaged buildings with minimal confusion with the undamaged class. In contrast, Model 1 showed limited reliability in identifying severe damage, with nearly half of severely damaged buildings misclassified as undamaged. These findings clearly highlight the critical role of building age in capturing structural vulnerability. Feature importance analysis supports this conclusion, indicating that building area (22.0%) and building age (21.2%) are the most influential features in Model 2, while Peak Ground Acceleration (PGA) remains the dominant seismic parameter, contributing approximately 31.7% of total importance. In Model 1, structural attributes played a more dominant role, with building area (25.8%) and number of floors (23.5%) being the primary contributors to model performance.

The Kahramanmaras earthquake sequence represents an extreme seismic scenario, characterized by two large-magnitude events that affected a vast region and multiple provinces in southern Turkiye. The dataset assembled for this case study covers the city of Kahramanmaras and represents the spatial distribution of damage for two predominant structural systems: masonry buildings and reinforced concrete buildings. Three machine learning algorithms were employed to identify the most suitable predictive models under three damage classification schemes: binary classification, three-class (traffic-light) classification, and four-class damage classification. Model performance was evaluated across three spatial scales: building, district, and city.

At the building scale, the models achieved approximately 80% accuracy for binary classification; however, performance decreased to around 65% and 50% for the three-class and four-class schemes, respectively. Confusion matrix analysis revealed that binary classification resulted in a high rate of unsafe buildings being misclassified as safe, reaching nearly 50% for both masonry and reinforced concrete structures. In contrast, the traffic-light classification provided a more balanced and reliable performance, with less

than 10% confusion between safe and unsafe classes. Among the tested algorithms, the Bagging classifier showed the most robust performance for the three-class damage scheme, whereas the four-class classification exhibited substantial confusion between adjacent damage states.

At the district scale, binary classification maintained relatively high accuracy, while the three-class classification achieved reasonable performance with moderate confusion involving the "To Be Checked" category for both structural types. The four-class classification showed improved identification of severely damaged and undamaged buildings, although overlap persisted between moderate and low damage states. At the city scale, the models demonstrated stable and high accuracy across all classification schemes for both masonry and reinforced concrete buildings, particularly when Bagging-based ensemble methods were employed.

Overall, this thesis presents a comprehensive and transferable workflow for rapid post-earthquake damage estimation using machine learning. The developed models can be directly applied to areas with comparable building practices and seismic characteristics, such as the Eastern Anatolian region. For application in other regions, retraining with locally representative data is essential to accurately capture regional structural behavior and vulnerability. The findings further indicate that binary damage classification at the building scale is not suitable for operational use due to its tendency to underestimate unsafe structures. While four-class damage classification provides greater detail, its practical reliability is limited by significant confusion between adjacent damage states, likely arising from uncertainties in post-earthquake damage labeling. Among the evaluated approaches, the three-class traffic-light classification emerges as the most effective and operationally reliable scheme for both masonry and reinforced concrete buildings, offering an optimal balance between accuracy, interpretability, and decision-making relevance. This classification scheme is therefore particularly well-suited for rapid damage estimation and emergency response applications in post-earthquake disaster management.

REFERENCES

[1] **Fidan, H.** (2024, October 29). İzmir'de 117 kişinin yaşamını yitirdiği depremin üzerinden 4 yıl geçti. *Anadolu Ajansı.* https://www.aa.com.tr/tr/gundem/Izmirde-117-kisinin-yasamini-yitirdigi-depremin-uzerinden-4-yil-gecti/3377806

[2] **Wertheimer, B. T.** (2023, February 6). Turkey earthquake: Death toll could increase eightfold, WHO says. *BBC.* https://www.bbc.com/news/world-europe-64533851

[3] **FEMA**, Applied Technology Council, Hamburger, R. O., Holmes, W. T., Roeder, C., Scawthorn, C., Ellingwood, B., Kennedy, R. P., Mahin, S., Anagnos, T., Bendimerad, F. M., Baker, J., Bonneville, D., & Seligson, H. (2012). *Seismic performance assessment of buildings Volume 1 – Methodology.*

[4] **Kiani, J., Camp, C., & Pezeshk, S**. (2019). On the application of machine learning techniques to derive seismic fragility curves. *Computers & Structures, 218,* 108–122. https://doi.org/10.1016/j.compstruc.2019.03.004

[5] **Kim, T., Kwon, O., & Song, J.** (2018). Response prediction of nonlinear hysteretic systems by deep neural networks. *Neural Networks, 111,* 1–10. https://doi.org/10.1016/j.neunet.2018.12.005

[6**] Wang, W., Li, L., & Qu, Z.** (2023). Machine learning-based collapse prediction for post-earthquake damaged RC columns under subsequent earthquakes. *Soil Dynamics and Earthquake Engineering, 172,* 108036. https://doi.org/10.1016/j.soildyn.2023.108036

[7] **Zhou, Z., Wang, M., Han, M., Yu, X., & Lu, D.** (2024). Prediction of damage potential in mainshock–aftershock sequences using machine learning algorithms. *Earthquake Engineering and Engineering Vibration, 23*(4), 919–938. https://doi.org/10.1007/s11803-024-2280-6

[8] **Zhang, Y., Burton, H. V., Sun, H., & Shokrabadi, M.** (2017). A machine learning framework for assessing post-earthquake structural safety. *Structural Safety, 72,* 1–16. https://doi.org/10.1016/j.strusafe.2017.12.001

[9] **Alcantara, E. A. M., & Saito, T.** (2023). Machine learning-based rapid post-earthquake damage detection of RC resisting-moment frame buildings. *Sensors, 23*(10), 4694. https://doi.org/10.3390/s23104694

[10] **Tocchi, G., Misra, S., Padgett, J. E., Polese, M., & Di Ludovico, M.** (2023). The use of machine-learning methods for post-earthquake building usability assessment: A predictive model for seismic-risk impact analyses. *International Journal of Disaster Risk Reduction, 97*, 104033. https://doi.org/10.1016/j.ijdrr.2023.104033

[11] **Cosgun, C.** (2023). Machine learning for the prediction of evaluation of existing reinforced concrete structures performance against earthquakes. *Structures, 50,* 1994–2003. https://doi.org/10.1016/j.istruc.2023.02.127

[12] **Zhang, H., Cheng, X., Li, Y., He, D., & Du, X.** (2022). Rapid seismic damage state assessment of RC frames using machine learning methods. *Journal of Building Engineering, 65,* 105797. https://doi.org/10.1016/j.jobe.2022.105797

[13] **Tang, Q., Dang, J., Cui, Y., Wang, X., & Jia, J.** (2021). Machine learning-based fast seismic risk assessment of building structures. *Journal of Earthquake Engineering, 26*(15), 8041–8062. https://doi.org/10.1080/13632469.2021.1987354

[14] **Bhatta, S., & Dang, J.** (2023). Seismic damage prediction of RC buildings using machine learning. *Earthquake Engineering & Structural Dynamics, 52*(11), 3504–3527. https://doi.org/10.1002/eqe.3907

[15] **Bhatta, S., Kang, X., & Dang, J.** (2024). Machine learning prediction models for ground motion parameters and seismic damage assessment of buildings at a regional scale. *Resilient Cities and Structures, 3*(1), 84–102. https://doi.org/10.1016/j.rcns.2024.03.001

[16] **Ghimire, S., Guéguen, P., Giffard-Roisin, S., & Schorlemmer, D.** (2022). Testing machine learning models for seismic damage prediction at a regional scale using building-damage dataset compiled after the 2015 Gorkha Nepal earthquake. *Earthquake Spectra, 38*(4), 2970–2993. https://doi.org/10.1177/87552930221106495

[17] **Wang, Y., Chew, A. W. Z., & Zhang, L.** (2022). Building damage detection from satellite images after natural disasters on extremely imbalanced datasets. *Automation in Construction, 140,* 104328. https://doi.org/10.1016/j.autcon.2022.104328

[18] **Braik, A. M., & Koliou, M.** (2024). Automated building damage assessment and large-scale mapping by integrating satellite imagery, GIS, and deep learning. *Computer-Aided Civil and Infrastructure Engineering, 39*(15), 2389–2404. https://doi.org/10.1111/mice.13197

[19] **Demircioğlu, M. B., Erdik, M., Hancılar, U., Harmandar, E., Kamer, Y., Şeşetyan, K., Tuzun, C., Yenidoğan, C., & Zülfikar, A. C.** (2010). *Earthquake loss estimation routine (ELER©) (Version 3.0): Technical manual.* Boğaziçi University, Department of Earthquake Engineering.

[20] **FEMA.** (2003). *Multi-hazard loss estimation methodology: Earthquake model (HAZUS-MH MR1).* U.S. Department of Homeland Security.

[21] **Erdik, M., Aydinoglu, N., Fahjan, Y., Sesetyan, K., Demircioglu, M., Siyahi, B., Durukal, E., Ozbey, C., Biro, Y., Akman, H., & Yuzugullu, O.** (2003). Earthquake risk assessment for Istanbul metropolitan area. *Earthquake Engineering and Engineering Vibration, 2*(1), 1–23. https://doi.org/10.1007/bf02857534

[22] **Quinlan, J. R.** (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106. https://doi.org/10.1007/bf00116251

[23] **GeeksforGeeks**. (2025a, June 30). Decision tree. *GeeksforGeeks.* https://www.geeksforgeeks.org/machine-learning/decision-tree/

[24] **Breiman, L.** (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

[25] **Genuer, R., Poggi, J., & Tuleau-Malot, C.** (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*(14), 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

[26] **Reis, I., Baron, D., & Shahaf, S.** (2018). Probabilistic random forest: A machine learning algorithm for noisy data sets. *The Astronomical Journal, 157*(1), 16. https://doi.org/10.3847/1538-3881/aaf101

[27] **Random forests for complete beginners.** (2020, May 14). *DataSource.ai.* https://www.datasource.ai/en/data-science-articles/random-forests-for-complete-beginners

[28] **Shi, Y., Ke, G., Chen, Z., Zheng, S., & Liu, T.** (2022). Quantized training of gradient boosting decision trees. *arXiv.* https://doi.org/10.48550/arxiv.2207.09682

[29] **Chen, T., & Guestrin, C.** (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference*

on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

[30] **Clark, B., & Lee, F.** (2025, November 17). What is gradient boosting? *IBM.* https://www.ibm.com/think/topics/gradient-boosting

[31] **GeeksforGeeks.** (2025b, September 5). Implementation of XGBOOST (eXtreme Gradient Boosting). *GeeksforGeeks.* https://www.geeksforgeeks.org/machine-learning/implementation-of-xgboost-extreme-gradient-boosting/

[32] **Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.** (2017). LightGBM: A highly efficient gradient boosting decision tree. *HAL.* https://hal.science/hal-03953007

[33] **Nhat-Duc, H., & Van-Duc, T.** (2023). Comparison of histogram-based gradient boosting classification machine, random forest, and deep convolutional neural network for pavement raveling severity classification. *Automation in Construction, 148,* 104767. https://doi.org/10.1016/j.autcon.2023.104767

[34] **Breiman, L.** (1996). Bagging predictors. *Machine Learning, 24,* 123–140. https://doi.org/10.1023/A:1018054314350

[35] **González, S., García, S., Del Ser, J., Rokach, L., & Herrera, F.** (2020). A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Information Fusion, 64,* 205–237. https://doi.org/10.1016/j.inffus.2020.07.007

[36] **Papatheodorou, K., Theodoulidis, N., Klimis, N., Zulfikar, C., Vintila, D., Cardanet, V., Kirtas, E., Toma-Danila, D., Margaris, B., Fahjan, Y., Panagopoulos, G., Karakostas, C., Papathanassiou, G., & Valkaniotis, S.** (2023). Rapid earthquake damage assessment and education to improve earthquake response efficiency and community resilience. *Sustainability, 15*(24), 16603. https://doi.org/10.3390/su152416603

[37] **OpenStreetMap Contributors.** (2021). OpenStreetMap database [PostgreSQL via API]. OpenStreetMap Foundation. Retrieved December 22, 2021, from https://www.openstreetmap.org

[38] **Open-Elevation.** (n.d.). Free and open-source elevation API. *Open-Elevation.* Retrieved April 22, 2025, from https://open-elevation.com/

[39] **scikit-learn developers**. (n.d.). Cross-validation: Evaluating estimator performance. In *scikit-learn documentation.* Retrieved January 22,

2026, from https://scikit-learn.org/stable/modules/cross_validation.html

[40] **Aktuğ, B., TiRyakiOğlu, İ., SözbiLiR, H., Özener, H., Özkaymak, Ç., YiĞiT, C. Ö., Solak, H. İ., EyübagiL, E. E., GeliN, B., Tatar, O., & Softa, M.** (2021). GPS-derived finite source mechanism of the 30 October 2020 Samos earthquake, Mw = 6.9, in the Aegean extensional region. *Turkish Journal of Earth Sciences, 30*(SI-1), 718–737. https://doi.org/10.3906/yer-2101-18

[41] **Dogan, G. G., Yalciner, A. C., Yuksel, Y., Ulutaş, E., Polat, O., Güler, I., Şahin, C., Tarih, A., & Kânoğlu, U.** (2021). The 30 October 2020 Aegean Sea tsunami: Post-event field survey along Turkish coast. *Pure and Applied Geophysics, 178*(3), 785–812. https://doi.org/10.1007/s00024-021-02693-3

[42] **Yakut, A., Sucuoğlu, H., Binici, B., Canbay, E., Donmez, C., İlki, A., Caner, A., Celik, O. C., & Ay, B. Ö.** (2021). Performance of structures in İzmir after the Samos island earthquake. *Bulletin of Earthquake Engineering, 20*(14), 7793–7818. https://doi.org/10.1007/s10518-021-01226-6

[43] **Cetin, K. O., Zarzour, M., Cakir, E., Tuna, S. C., & Altun, S.** (2023). 2-D and 3-D basin site effects in Izmir-Bayrakli during the October 30, 2020 Mw7.0 Samos earthquake. *Bulletin of Earthquake Engineering, 21*(12), 5419–5442. https://doi.org/10.1007/s10518-023-01738-3

[44] **Demirel, I. O., Yakut, A., & Binici, B.** (2022). Seismic performance of mid-rise reinforced concrete buildings in Izmir Bayrakli after the 2020 Samos earthquake. *Engineering Failure Analysis, 137,* 106277. https://doi.org/10.1016/j.engfailanal.2022.106277

[45] **Wang, X., Feng, G., He, L., et al.** (2023). Evaluating urban building damage of 2023 Kahramanmaras, Turkey earthquake sequence using SAR change detection. *Sensors, 23,* 6342. https://doi.org/10.3390/s23146342

[46] **Avğın, S., Köse, M. M., & Özbek, A.** (2024). Damage assessment of structural and geotechnical damages in Kahramanmaraş during the February 6, 2023 earthquakes. *Engineering Science and Technology, International Journal, 57,* 101811. https://doi.org/10.1016/j.jestch.2024.101811

CURRICULUM VITAE

Name Surname **:** Bilal Ein Larouzi

EDUCATION **:**

- **B.Sc.** **:** 2017, Prince Sattam bin Abdulaziz University, Faculty of Engineering, Civil Engineering Department
- **M.Sc.** **:** 2020, Istanbul Kültür University, Faculty of Engineering, Structural Engineering Program

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2017 - 2018 Rakaez Almodon Al-Motakdma Contracting Company – Riyadh City – KSA – Site Civil Engineer
- 2023 – present Abak Engineering Consult – Riyadh City – KSA – Structural Engineer

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Fahjan, Y., & Larouzi, B. E. (2025).** *Ground motion Parameters importance in machine learning-based damage prediction*. https://doi.org/10.31462/icearc2025_ce_eqe_392
- **Larouzi, B. E., & Fahjan, Y. (2026).** From Localized Collapse to City-Wide impact: ensemble Machine Learning for Post-Earthquake Damage Classification. *Infrastructures*, *11*(1), 25. https://doi.org/10.3390/infrastructures11010025

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- **Yazici, G., & Larouzi, B. E. (2025)**. An OpenSEES graphical user interface for structural dynamics and earthquake engineering instruction. *Scientific Reports*, *15*(1), 34456. https://doi.org/10.1038/s41598-025-17632-8