DEVELOPMENT OF A SEISMIC DAMAGE PREDICTION MODEL BY USING MACHINE LEARNING CLASSIFICATION ALGORITHMS WITH AN ARTIFICIAL DATASET

by

Ali Talha Atici

B.S., Civil Engineering, Yıldız Technical University, 2019

Submitted to the Kandilli Observatory and Earthquake
Research Institute in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Earthquake Engineering

Boğaziçi University

2023

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my advisor, Professor Ufuk Hancılar, for his support and guidance throughout my master's program. I am grateful to him for his feedback, guidance, and the considerable time he dedicated to our lengthy meetings during the thesis process.

I extend my thanks to my colleague, Onur Ülkü, who holds a special place in my education and professional journey. I appreciate him for our collaboration, invaluable contributions, his vision, and the ideas he shared.

Also, I want to thank my respected friends, Elif Yıldırım and Nurullah Açıkgöz, for their supportive comments, collaboration, and exchange of ideas.

I would like to express my deep gratitude to my family, who has always supported me and played a significant role in my journey to reach this point.

ABSTRACT

DEVELOPMENT OF A SEISMIC DAMAGE PREDICTION MODEL BY USING MACHINE LEARNING CLASSIFICATION ALGORITHMS WITH AN ARTIFICIAL DATASET

Assessing the potential damage to buildings due to a possible earthquake in a region and taking measures, such as strengthening or reconstruction of vulnerable structures, is critically important to minimize social and economic losses that are likely to occur. Evaluating the seismic performance of structures is a comprehensive and time-consuming process. However, using well-trained machine learning prediction models instead of traditional structural performance analyses can significantly reduce computation time. This thesis focuses on developing a damage prediction model using classification-based machine learning algorithms, utilizing a two-dimensional reinforced concrete frame system dataset that represents low to mid-rise, non-ductile buildings. The structural features forming the dataset are obtained from a comprehensive literature review on building stock characteristics in the Marmara region. Nonlinear time-history analyses are conducted using actual earthquake records with the OpenSeesPy framework. The maximum inter-story drift ratio is used as an engineering demand parameter to classify the damage state of buildings. Reliable machine learning models are developed with a balanced dataset. Twenty-four models are created using six variant ground motion intensity measures and four classification algorithms: k-Nearest Neighbors, Support Vector Machine, Decision Tree, and Random Forest. The best-performing model is determined by comparing performance metrics and the confusion matrix. In conclusion, the model developed with a dataset incorporating peak ground velocity and utilizing the Random Forest classification algorithm demonstrates the most effective performance with 92% prediction accuracy.

ÖZET

MAKİNE ÖĞRENMESİ SINIFLANDIRMA ALGORİTMALARI İLE YAPAY VERİ SETİ KULLANILARAK DEPREM HASARI TAHMİN MODELİ GELİŞTİRİLMESİ

Olası bir depremin binalarda meydana getirebileceği potansiyel hasarı belirlenmesi ve riskli yapıların güçlendirilmesi veya yeniden inşa edilmesi gibi önlemlerin alınması, sosyal ve finansal kayıpları en aza indirmek açısından kritik bir öneme sahiptir. Yapıların deprem performansının değerlendirilmesi kapsamlı ve zaman alıcı bir süreçtir. Ancak, geleneksel yapısal performans analizleri yerine iyi eğitilmiş makine öğrenmesi tahmin modellerinin kullanılması, hesaplama süresini önemli ölçüde azaltabilir. Bu tez, az ve orta katlı sünek olmayan yapıları ifade etmesi için iki boyutlu betonarme çerçeve sistemlerinden oluşan bir veri seti ve sınıflandırma tabanlı makine öğrenmesi algoritmaları kullanarak bir hasar tahmin modeli geliştirmeye odaklanmaktadır. Veri setini oluşturan yapısal özellikler, Marmara bölgesindeki bina stok özellikleri hakkında kapsamlı bir literatür taramasından elde edilmiştir. Doğrusal olmayan dinamik analizler, gerçek deprem kayıtları kullanılarak OpenSeesPy programı ile yapılmıştır. Katlar arası göreli en yüksek yer değiştirme oranı, binaların hasar durumunu sınıflandırmak için mühendislik talep parametresi olarak kullanılmıştır. Altı farklı yer hareketi şiddet ölçütü ve dört makine öğrenmesi sınıflandırma algoritması (k-Nearest Neighbors, Support Vector Machine, Decision Tree, and Random Forest olmak üzere) kullanılarak yirmi dört farklı model oluşturulmuştur. Modellerin performansı, performans metrikleri ve karışıklık matrisi karşılaştırılarak belirlenmiştir. Sonuç olarak, yirmi dört model arasında, pik yer hızını içeren veri seti ve Random Forest sınıflandırma algoritması kullanılan model, %92 oranında tahmin doğruluğu ile en etkili performansı sergilemektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . iii
ABSTRACT . . . iv
ÖZET . . . v
LIST OF FIGURES . . . viii
LIST OF TABLES . . . xi
LIST OF SYMBOLS . . . xvi
LIST OF ACRONYMS/ABBREVIATIONS . . . xvii
1. INTRODUCTION . . . 1
1.1. Objectives and Content of the Thesis . . . 2
1.2. Literature Review . . . 3
2. COMPILATION OF BUILDING INVENTORY AND STRUCTURAL ANALYSES . . . 7
2.1. Introduction . . . 7
2.2. Building Stock Characteristics in Marmara Region . . . 8
2.2.1. Number of Stories . . . 9
2.2.2. Story Height and Commercial Use of Ground Floor . . . 10
2.2.3. Span Length and Number of Spans . . . 11
2.2.4. Structural Elements and Dimensions . . . 12
2.2.5. Material Properties . . . 13
2.2.6. Reinforcement Ratios . . . 14
2.2.7. Local Soil Conditions . . . 14
2.3. Modeling of the Study Buildings . . . 14
2.3.1. Model Verification . . . 19
2.4. Selection of Ground Motion Records . . . 20
2.5. Intensity Measures Based on Earthquake Record . . . 26
2.6. Nonlinear Time History Analyses . . . 28
2.7. Assessment of the Analysis Results . . . 31
3. DEVELOPMENT OF MACHINE LEARNING MODELS . . . 35

3.1. Introduction 35
3.2. Data Pre-Processing 36
3.3. Preliminary Data Analysis 38
3.4. Machine Learning Algorithms 46
3.4.1. k-Nearest Neighbor 46
3.4.2. Support Vector Machine (SVM) 48
3.4.3. Decision Tree (DT) 50
3.4.4. Random Forest (RF) 52
3.5. Application of Machine Learning Algorithms 53
3.5.1. Evaluation of the Model Performance 54
3.6. The Results of Seismic Damage Prediction Models and Comparison of Model Performances 56
4. CONCLUSIONS 67
4.1. Summary of the Thesis 67
4.2. Observations and Findings 68
4.3. Future Studies 69
REFERENCES 70
APPENDIX A: SCRIPTS OF GENERATING ARTIFICIAL DATASET, OPENSEESPY, DATA PREPARATION, AND APPLICATION OF MACHINE LEARNING ALGORITHMS 77
APPENDIX B: THE SELECTED GROUND MOTIONS 78

LIST OF FIGURES

Figure 2.1. Distribution of reinforced concrete frame buildings in Istanbul based on the number of stories and the construction year (DEZIM, 2020). 10

Figure 2.2. Distribution of standard-story height for RC residential buildings (Bal et al., 2008). 11

Figure 2.3. Beam length distribution for all buildings (Bal et al., 2008). 12

Figure 2.4. Variation in the 28th-day characteristic compressive strength of buildings located in Marmara (Bal et al., 2008). 13

Figure 2.5. Dispersion of yield strength for reinforcing steel of S220 MPa grade (Bal et al., 2008). 13

Figure 2.6. Example building models in the OpenSeesPy framework: 3- and 5-story buildings. 16

Figure 2.7. The plan view of the 3-span structure displays the structural model's considered axis, serving as an example of loading tributary areas. . 16

Figure 2.8. Fiber discretization of a reinforced concrete section. 17

Figure 2.9. Confined and unconfined concrete stress-strain graphics. 18

Figure 2.10. Stress-strain graphics of reinforcing steel (OpenSees, Steel01). . . . 19

Figure 2.11. Response spectra of selected ground motions for period class 1 and 270 m/s V_{s30} (a) AFAD-TADAS database (b) CyberShake database [Baker and Lee's (2018) MATLAB Algorithm is used for plotting these figures.]. 25

Figure 2.12. Example time series for a selected ground motion (02.06.2023 Pazarcik Kahramanmaras Earthquake (M_w=7.8), Station 3145, East Component). 26

Figure 2.13. P-Delta effect: (a) undeformed configuration, (b) deformed configuration, (c) bending moment distribution (Burgos and Silva, 2023). 29

Figure 2.14. The variation of mode damping ratios with natural frequency for an analyzed example structure. 30

Figure 2.15. The outcomes of the NLTH analysis are presented as damage states corresponding to SF_1 and SF_2. 33

Figure 3.1. MIDR distributions for 3- and 5-story buildings. 38

Figure 3.2. MIDR distribution for 3- and 4-span buildings. 39

Figure 3.3. MIDR distributions conditioned on span length. 40

Figure 3.4. MIDR distributions conditioned on story height. 41

Figure 3.5. MIDR distributions conditioned on ground floor height. 41

Figure 3.6. MIDR distributions conditioned on soil condition. 42

Figure 3.7. The probability density functions of MIDR depending on column area for 3-story buildings. 43

Figure 3.8. The probability density functions of MIDR depending on column area for 5-story buildings. 43

Figure 3.9. The probability density functions of MIDR depending on concrete strength. 44

Figure 3.10. Scatter plot illustrating the relationship between MIDR and IMs, along with their respective distributions. 45

Figure 3.11. The schematic representation of kNN (Yang, 2019). 46

Figure 3.12. An illustration of a separable problem within a two-dimensional space. The support vectors, indicated by dark grey squares, delineate the margin, representing the maximum separation between the two classes (Cortes and Vapnik, 1995). 48

Figure 3.13. The boundary and margins identified by the Gaussian kernel with various spread values (Alpaydin, 2014). 50

Figure 3.14. Illustration of a dataset along with the corresponding decision tree. Oval nodes represent decision nodes, and rectangles represent leaf nodes (Alpaydin, 2014). 51

LIST OF TABLES

Table 2.1. Building characteristics, references, number of structures and variations. 9

Table 2.2. Information on reinforcing steel. 18

Table 2.3. Comparison of modal analysis results from OpenSeesPy and SAP2000 for 3-story buildings. 20

Table 2.4. List of period classes for recording selections. 21

Table 2.5. Selected ground motions for period class 1 and V_{s30} = 270 m/s. . . 24

Table 2.6. Limits of inter-story drift ratio (%) linked to damage states (Ghobarah, 2004). 32

Table 2.7. Distribution of damage levels. 32

Table 2.8. The NLTH analysis results as damage states for SF_1 and SF_2, overall results, and ratios. 34

Table 3.1. Randomly selected data points from the dataset; inputs and outputs. 36

Table 3.2. Processed data points with PGD as IM. 37

Table 3.3. Confusion Matrix and Calculation of Performance Metrics. 55

Table 3.4. Evaluation metrics of machine learning algorithms 58

Table 3.5. Confusion matrix of kNN algorithm, the dataset with PGA 59

Table 3.6. Confusion matrix of kNN algorithm, the dataset with PGV 59

Table 3.7. Confusion matrix of kNN algorithm, the dataset with PGD 59

Table 3.8. Confusion matrix of kNN algorithm, the dataset with $S_a(T_1)$ . . . 60

Table 3.9. Confusion matrix of kNN algorithm, the dataset with $S_d(T_1)$ . . . 60

Table 3.10. Confusion matrix of kNN algorithm, the dataset with CAV 60

Table 3.11. Confusion matrix of SVM algorithm, the dataset with PGA 61

Table 3.12. Confusion matrix of SVM algorithm, the dataset with PGV 61

Table 3.13. Confusion matrix of SVM algorithm, the dataset with PGD 61

Table 3.14. Confusion matrix of SVM algorithm, the dataset with $S_a(T_1)$ . . . 62

Table 3.15. Confusion matrix of SVM algorithm, the dataset with $S_d(T_1)$ . . . 62

Table 3.16. Confusion matrix of SVM algorithm, the dataset with CAV 62

Table 3.17. Confusion matrix of DT algorithm, the dataset with PGA 63

Table 3.18. Confusion matrix of DT algorithm, the dataset with PGV 63

Table 3.19. Confusion matrix of DT algorithm, the dataset with PGD 63

Table 3.20. Confusion matrix of DT algorithm, the dataset with $S_a(T_1)$ 64

Table 3.21. Confusion matrix of DT algorithm, the dataset with $S_d(T_1)$ 64

Table 3.22. Confusion matrix of DT algorithm, the dataset with CAV 64

Table 3.23. Confusion matrix of RF algorithm, the dataset with PGA 65

Table 3.24. Confusion matrix of RF algorithm, the dataset with PGV 65

Table 3.25. Confusion matrix of RF algorithm, the dataset with PGD 65

Table 3.26. Confusion matrix of RF algorithm, the dataset with $S_a(T_1)$ 66

Table 3.27. Confusion matrix of RF algorithm, the dataset with $S_d(T_1)$ 66

Table 3.28. Confusion matrix of RF algorithm, the dataset with CAV 66

Table B.1. Selected ground motions for period class 1 and V_{S30} 270 m/s. . . . 79

Table B.2. Selected ground motions for period class 1 and V_{S30} 560 m/s. . . . 80

Table B.3. Selected ground motions for period class 1 and V_{S30} 1130 m/s. . . 81

Table B.4. Selected ground motions for period class 2 and V_{S30} 270 m/s. . . . 82

Table B.5. Selected ground motions for period class 2 and V_{S30} 560 m/s. . . . 83

Table B.6. Selected ground motions for period class 2 and V_{S30} 1130 m/s. . . 84

Table B.7. Selected ground motions for period class 3 and V_{S30} 270 m/s. . . . 85

Table B.8. Selected ground motions for period class 3 and V_{S30} 560 m/s. . . . 86

Table B.9. Selected ground motions for period class 3 and V_{S30} 1130 m/s. . . 87

Table B.10. Selected ground motions for period class 4 and V_{S30} 270 m/s. . . . 88

Table B.11. Selected ground motions for period class 4 and V_{S30} 560 m/s. . . . 89

Table B.12. Selected ground motions for period class 4 and V_{S30} 1130 m/s. . . 90

Table B.13. Selected ground motions for period class 5 and V_{S30} 270 m/s. . . . 91

Table B.14. Selected ground motions for period class 5 and V_{S30} 560 m/s. . . . 92

Table B.15. Selected ground motions for period class 5 and V_{S30} 1130 m/s. . . 93

Table B.16. Selected ground motions for period class 6 and V_{S30} 270 m/s. . . . 94

Table B.17. Selected ground motions for period class 6 and V_{S30} 560 m/s. . . . 95

Table B.18. Selected ground motions for period class 6 and V_{S30} 1130 m/s. . . 96

Table B.19. Selected ground motions for period class 7 and V_{S30} 270 m/s. . . . 97

Table B.20. Selected ground motions for period class 7 and V_{S30} 560 m/s. . . . 98

Table B.21. Selected ground motions for period class 7 and V_{S30} 1130 m/s. . . 99

Table B.22. Selected ground motions for period class 8 and V_{S30} 270 m/s. . . . 100

Table B.23. Selected ground motions for period class 8 and V_{S30} 560 m/s. . . . 101

Table B.24. Selected ground motions for period class 8 and V_{S30} 1130 m/s. . . 102

Table B.25. Selected ground motions for period class 9 and V_{S30} 270 m/s. . . . 103

Table B.26. Selected ground motions for period class 9 and V_{S30} 560 m/s. . . . 104

Table B.27. Selected ground motions for period class 9 and V_{S30} 1130 m/s. . . 105

Table B.28. Selected ground motions for period class 10 and V_{S30} 270 m/s. . . 106

Table B.29. Selected ground motions for period class 10 and V_{S30} 560 m/s. . . 107

Table B.30. Selected ground motions for period class 10 and V_{S30} 1130 m/s. . . 108

LIST OF SYMBOLS

a_0	Mass-proportional damping coefficient
a_1	Stiffness-proportional damping coefficient
b	Strain-hardening ratio
E_s	Elasticity modulus of the reinforcing steel
E_c	Elasticity modulus of concrete
f_c	28^{th}-day concrete compressive strength
F_sy	Characteristic yield strength of the rebar steel
G	Dead load
Q	Live load
T_{sig}	Significant duration
ζ	Modal damping ratio
ζ_n	n^{th} mode modal damping ratio
ω_n	n^{th} mode natural frequency of the building

LIST OF ACRONYMS/ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
AFAD	Disaster and Emergency Management Authority
ATC	Applied Technology Council
CART	Classification and Regression Tree
CAV	Cumulative Absolute Velocity
CV	Cross Validation
DT	Decision Tree
EDP	Engineering Demand Parameter
FEM	Finite Element Method
IDA	Incremental Dynamic Analysis
IDR	Inter-story Drift Ratio
IM	Intensity Measure
JSIM	Japonesse Seismic Index Method
kNN	k-Nearest Neighbor
LSTM	Long Short-Term Memory
MIDR	Maximum Inter-story Drift Ratio
ML	Machine Learning
NAFZ	North Anatolian Fault Zone
NLTH	Non-Linear Time History
PGA	Peak Ground Acceleration
PDF	Probability Density Function
PGD	Peak Ground Displacement
PGV	Peak Ground Velocity
RBF	Radial-Basis Function
RELA	Rapid Earthquake Loss Assessment
RF	Random Forest
S_a	Spectral Acceleration

SAP2000	Structural Analysis Program 2000
S_d	Spectral Displacement
SDOF	Single Degree Of Freedom
SF	Scale Factor
SVM	Support Vector Machine
TBEC	Turkish Building Earthquake Code
V_{S30}	the time-averaged shear-wave velocity to a depth of 30 meters

1. INTRODUCTION

Especially in seismic-prone regions, it is necessary to take precautions, such as strengthening or renewing the structures before possible earthquakes, to maintain the use of the structures, minimize financial loss, and prevent loss of life. The performance of the structures should be determined to prioritize the measures to be taken.

The Marmara region is facing a potential earthquake threat, highlighting the growing significance of conducting thorough assessments of seismic hazards (Murru et al., 2016). Many studies (Atakan et al., 2002; Erdik et al., 2004; Kalkan et al., 2009; Gülerce and Ocak, 2013) have been conducted regarding the expected earthquake in the Marmara region and the associated risks. Taking immediate precautions can be considered a common conclusion of these studies. Hence, it is of paramount importance to assess whether the buildings can withstand such strong ground shaking, particularly in a city like Istanbul, with a dense population and an aging building stock.

Earthquake performance assessment of existing structures is one of the concerns in earthquake engineering that might directly affect human life. Various approaches have been developed for estimating potential damage levels in existing structures, and these can be categorized into three main groups: (i) methods for evaluating the seismic performance of specific buildings as outlined in modern seismic codes and portfolio-based seismic risk analyses, (ii) rapid damage assessment techniques that assign scores to buildings based on structural characteristics [such as ATC 21 (2002), PERA (Ilki et al., 2014), Japanese Seismic Index Method (JSIM, 2001), METU procedure (Sucuoğlu et al., 2007), P25 procedure (Bal et al., 2007)], and (iii) machine learning prediction models [referenced in studies like Zhang et al., 2018; Kiani et al., 2019; Mangalathu and Burton, 2019; Mangalathu et al., 2020; Roeslin et al., 2020; Stojadinović et al., 2021; Ulku et al., 2022; Wu and Sarno, 2022].

The first group of methods is intended to estimate the seismic performance of

individual buildings based on finite element analysis or to assess earthquake risk of classes of buildings through structural capacity and fragility relationships. Assessing buildings individually using this approach is crucial for minimizing losses. Regrettably, the exhaustive examination and structural analysis of millions of buildings using this method is both time-consuming and expensive. In contrast, seismic risk analyses based on an inventory approach necessitate pre-established fragility and vulnerability functions, developed either analytically or empirically, for each building class under consideration. The second approach predicts earthquake damage to buildings by assigning scores to structural features accessible through on-site surveys. The rapid assessment method offers a less precise estimate of structural performance compared to alternative methods. Machine Learning (ML) models for predicting damage states have become popular recently, although the ML models commonly use observed data collected after an earthquake. Nevertheless, observational data are often imbalanced, making it challenging to generate reliable models. In order to avoid the problem of an imbalanced dataset, the training and the test dataset for the machine learning model were constructed through the analysis of artificial structural assessments designed to accurately represent real-world scenarios.

1.1. Objectives and Content of the Thesis

This thesis describes the method of a prediction model for existing building damage states under possible earthquakes by using machine learning classification algorithms with analytically generated datasets. The development of the supervised ML models for predicting structure performance has two main steps. First, a balanced and unbiased dataset is created, and second, machine learning algorithms are developed.

Primarily, to generate a balanced and unbiased dataset, the central axes of three-dimensional residential-type buildings, designed to capture the features of structures in the Marmara region, were represented and modeled as two-dimensional frame systems in the Openseespy framework (McKenna et al., 2010). The structural characteristics of the buildings were compiled from studies that prepared statistical information on

the building inventory in the Marmara region, for instance, Bal et al. (2008), Azak et al. (2014), and DEZIM (2020). Nonlinear analyses were executed to determine the damage levels. For these analyses, eleven earthquake records were used for each structure, considering the fundamental vibration periods of the structures and the soil conditions. As a result of the analyses, the Maximum Inter-Story Drift Ratio (MIDR) values were recorded as the output of the dataset. MIDR values were grouped into intervals for the purpose of using classification algorithms in the machine learning model.

Secondly, the balanced dataset generated through the conducted analyses was utilized to develop reliable ML models. Twenty-four different models were created by employing four different classification algorithms (k-nearest neighbors, support vector machine, decision tree, random forest) and considering six different ground motion intensity levels, such as peak ground acceleration (PGA), peak ground velocity (PGV), peak ground displacement (PGD), spectral acceleration ($S_a(T_1)$), spectral displacement ($S_d(T_1)$), and cumulative absolute velocity (CAV). The results were examined by comparing the performance metrics of these models.

1.2. Literature Review

The utilization of ML has been increasing in recent years, owing to the significant role played by data science across various disciplines. The advantages of machine learning can primarily be categorized as computational efficiency, resolution of complex problems, making informed decisions through data analysis, and identifying and mitigating uncertainties (Xie et al., 2020). As a result, it has gained substantial popularity in scientific research and has found applications in engineering fields such as material science, bioengineering, earthquake engineering, and civil engineering.

Zhang et al. (2018) utilized machine learning algorithms to evaluate the structural safety (whether it is safe or unsafe) of a building after an earthquake, using a predefined threshold for residual collapse capacity. In this study, a two-dimensional model

of a four-story, three-bay reinforced concrete special moment frame was constructed in OpenSees. Classification and Regression Trees (CART) and Random Forests (RF) were utilized to develop an ML model for probabilistically determining the structural safety status of earthquake-damaged buildings. Prediction accuracies exceeding 90% were achieved when assessing the safety status. Zhang et al. suggested, based on the results of their study, that the proposed framework could be utilized for a rapid structural assessment of a damaged building's suitability for occupancy after an earthquake. Furthermore, they suggested that it could be integrated as a subroutine in the evaluation of community resilience or in the assessment and optimization of building lifecycle performance.

In 2019, Kiani et al. applied ten machine learning algorithms based on classification, including logistic regression, lasso regression, linear and quadratic discriminant analyses, Naïve Bayes, decision tree, random forest, support vector machine, neural networks, and k-Nearest Neighbors, to forecast structural reactions. This study used an 8-story construction with a steel moment-resisting frame, and it was represented in OpenSees as a two-dimensional model employing lumped plasticity. Structural responses were obtained from multiple strip analyses. Kiani et al. presented a detailed procedure for deriving fragility curves based on classification-based ML tools. They also investigated the sensitivity of the ML model with respect to the dataset size. The results indicated that logistic regression, lasso regression, and Naïve Bayes are not sensitive to dataset size, while the performance of discriminant analysis depends significantly on the size of the dataset employed.

Mangalathu and Burton (2019) used the Long Short-Term Memory (LSTM) deep learning technique to categorize structure damage states using ATC-20 color labels. The observational dataset, which consisted of 1552 green labels, 1674 yellow labels, and 197 red labels, was obtained from the 2014 South Napa earthquake in California and randomly split into training and test subsets. The LSTM model achieved an overall accuracy 86% in determining the ATC-20 color labels in the test set. Mangalathu and Burton emphasized the significant potential of utilizing the LSTM method for the rapid

assessment of building cluster damage states.

Mangalathu et al. (2020) assessed the practicality of employing ML techniques to rapidly estimate earthquake-induced construction damage using four classification-based algorithms, including discriminant analysis, k-nearest neighbors, decision trees, and RF. The study utilized a dataset of 2276 damaged buildings resulting from the 2014 Napa earthquake. These buildings were classified based on their ATC-20 tags. The findings of the research indicated that the RF algorithm achieved an overall accuracy of 66% in label assignment, correctly identifying 79% of the actual yellow tags in the test set. As a result, researchers suggested that further efforts are required to facilitate the widespread implementation of ML for earthquake-induced structure damage forecasts.

Roeslin et al. (2020) present a case study on the development of a damage prediction model using four different classification-based ML algorithms (Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest). They analyzed post-earthquake observational data, which includes building damage information and structural characteristics for 237 buildings in Mexico City after the devastating 2017 earthquake in Puebla, Mexico. When comparing classification models, the RF model achieved the highest prediction accuracy. Consequently, this article introduces a method for utilizing machine learning in the assessment of post-earthquake damage data. An analysis of feature importance reveals that the building's location, seismic demand, and building height have the most significant impact.

Stojadinović et al. (2021) utilized a random forest to predict the damage probability distribution for the development of a rapid earthquake loss assessment (RELA) framework. The observational dataset contains 1979 buildings, which were collected after the 2010 Kraljevo earthquake. The proposed RELA framework consists of two stages: the pre-earthquake and co-earthquake stages. The pre-earthquake phase involves obtaining the characteristics of the portfolio in the designated area. The co-earthquake phase entails determining the level of damage of a representative set of structures in the portfolio to train the ML algorithm. Subsequently, the ML model is

used to predict the losses of other constructions in the portfolio. This study demonstrated that estimating losses with 10% of the portfolio yielded sufficiently accurate results, with less than a 20% relative error.

Ulku et al. (2022) classified damage states by applying six alternative classification-based machine learning algorithms: logistic regression, decision tree, random forest, k-nearest neighbor, support vector machine, and Naïve Bayes. In this study, collected data were utilized to develop a prediction model. The dataset includes building characteristics and damage states for 10451 damaged buildings in Elazığ province after the devastating 2020 Elazığ (Sivrice), Turkiye earthquake. The extensive damage state is relatively higher than the other target classes, leading to an imbalanced dataset. Under-sampling and over-sampling methods were used to overcome the issue of imbalanced datasets. The study also presented the importance of features that influence the classification of damaged buildings, with the number of stories showing the highest correlation. When evaluating the outcomes of classification algorithms, it was observed that the RF algorithm outperformed the others, achieving an average accuracy of 85% and an F1-score exceeding 68%.

Wu and Sarno (2022) used ML methods to create fragility curves for existing steel moment frames with masonry infills. They used a feedforward neural network to predict the collapse states of buildings, aiming to reduce computational costs significantly compared to traditional incremental dynamic analysis (IDA). The research focused on four steel moment frame buildings representing low- and mid-rise structures designed for low and high seismic activity. Two-dimensional finite element models of the study structures were executed in OpenSees. The study compared the fragility curves obtained through the proposed method with those derived from the time-consuming conventional IDA, which utilized 49 ground motions at small incremental steps but offered high accuracy in fragility curve estimation. The fragility curves produced using their proposed method closely mirrored the outcomes from the more complex IDA, especially when considering the influence of masonry infills on the overall structural response.

2. COMPILATION OF BUILDING INVENTORY AND STRUCTURAL ANALYSES

2.1. Introduction

This section details how the artificial dataset was created. Within the scope of this section, information is provided about the selection of structural characteristics, finite element modeling, nonlinear structural analysis method, selection of ground motion records, and evaluation of analysis results.

In creating the datasets for developing machine learning algorithms predicting earthquake damage, two methods can be employed: (i) observing damage in the field after earthquakes and (ii) using analytical models to predict consequences. Both approaches, as outlined by Baker et al. (2021), come with their own strengths and weaknesses. The empirical method, involving field observations, offers direct validation through evidence but demands meticulous data collection to ensure unbiased estimates of damage states across different levels of ground motion. Addressing imbalances in the observed dataset arising from uneven inspections at each damage level can be managed through techniques like under-sampling or over-sampling (Kaur et al., 2020; Ulku et al., 2022). Conversely, the analytical approach, while overcoming imbalanced datasets, requires a precise representation of real-world consequences in its predictions—a challenge highlighted by Baker et al. (2021).

This thesis opts for the analytical approach to create a dependable and well-balanced dataset for the development of a machine learning-based model for predicting earthquake damage. The dataset is prepared from simulations of NLTH analyses of representative buildings to mimic reality. The steps followed during the creation of the synthetic dataset are explained in subsections in this section of the thesis.

2.2. Building Stock Characteristics in Marmara Region

A strong earthquake is expected to occur in the near future in the Sea of Marmara, which is expected to affect the surrounding provinces (Murru et al., 2016; Atakan et al., 2002; Erdik et al., 2004; Kalkan et al., 2009; Gülerce and Ocak, 2013). In order to draw attention to this region, a comprehensive dataset has been created in this study by conducting tens of thousands of analyses covering the characteristics of the building stock in the Marmara region.

The characteristics of representative buildings were obtained through a literature review of the building inventory in the Marmara Region and Istanbul (Bal et al., 2007; Azak et al., 2014; DEZIM, 2020). The available resources and the analysis time required to consider all structural features and their variations limit the scope of the study. Hence, selected attributes are taken into account to represent the building stock of the region, as explained in the forthcoming sections.

Table 2.1 presents a summary of the structural features considered within the scope of the study. This table provides information about which reference was used to select each parameter, the range of values for each parameter, and the overall variations considered. The bottom row of the table indicates how many different building models were created based on these considered features.

Table 2.1. Building characteristics, references, number of structures and variations.

Building Characteristics	**References**	**Parameters**	**Number of Variations**
Number of Stories	DEZIM, 2020	3 and 5	2
Story Height	Bal et al., 2008	2.6, 2.8, and 3 m	3
Commercial Use of Base Floor	Bal et al., 2008 and Azak et al., 2014	'Yes' or 'No' If 'Yes', Ground Floor Height = 3.5 m	2
Number of Spans	Azak et al., 2014	3 and 4	2
Span Length	Bal et al., 2008 and Azak et al., 2014	2.5, 3, and 4 m	3
Column Dimensions	Azak et al., 2014	3 story: 25, 30, 35 cm 5 story: 30, 35, 40 cm	3
Beam Width		25 cm	1
Beam Depth	Bal et al., 2008	60 cm	1
Concrete Compr. Strength	Bal et al., 2008	5, 12, 20, 28, and 35 MPa	5
Steel Yield Str.	Bal et al., 2008	370 MPa	1
Reinforcement Ratio for Column	ABYYHY-1975 (Min. Requirement)	0.01	1
Reinforcement Ratio for Beam	ABYYHY-1975 (Min. Requirement)	0.004	1
Soil Condition (V_{S30})		1130, 560, 270 cm/s	3
Total Number of Variations (Number of structures in this study)			3240 (1080)

2.2.1. Number of Stories

Figure 2.1 represents the joint probability distribution based on the number of stories and the construction year of reinforced concrete buildings in Istanbul, according

to DEZIM (2020).

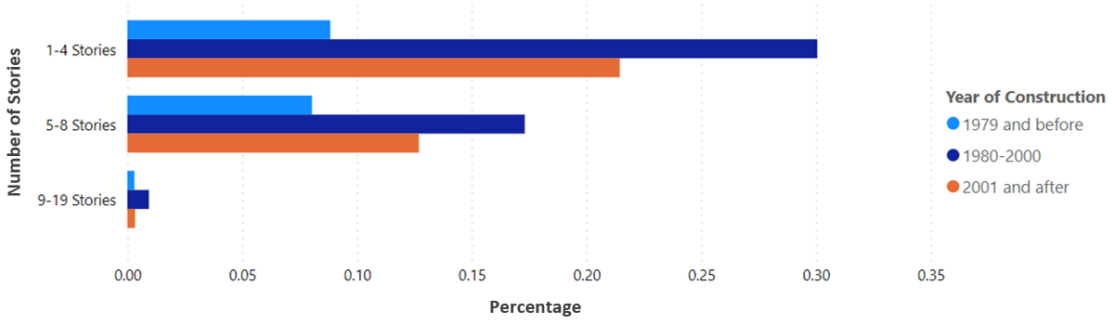


Figure 2.1. Distribution of reinforced concrete frame buildings in Istanbul based on the number of stories and the construction year (DEZIM, 2020).

As observed, the predominant reinforced concrete structures in Istanbul, the most populated city in the Marmara region, consist mainly of low-rise (1 to 4 floors) and medium-rise (5 to 8 floors) buildings erected between 1980 and 2000. To encompass a substantial portion of the buildings in the Marmara region for this study, 3-story buildings constructed between 1980 and 2000 were specifically chosen to represent low-rise structures, while 5-story buildings from the same time period were selected for the examination of mid-rise buildings.

2.2.2. Story Height and Commercial Use of Ground Floor

In order to encompass the features of the building inventory in the Marmara territory, the floor heights were determined by referring to the studies of Bal et al. (2008) and Azak et al. (2014). Both research studies highlighted the common practice of commercial use in Turkish construction. In the case of commercial use, the height of the ground floor is taller than that of the other floors. In addition to floor height irregularity, the lack of infill walls on the base floor, which are typically present on standard floors, results in lower floor stiffness on the ground floor. In this study, in the presence of commercial use, the height of the ground floor is considered to be 3.5 meters to implicitly represent the likely occurrence of soft-story behavior.

The normal floor heights are used in the analysis in three different variations, namely 2.6, 2.8, and 3.0 meters, based on the distribution of 938 sample buildings statistically evaluated in Bal et al. (2008). The distribution created by Bal et al. (2008) can be seen in Figure 2.2. According to this distribution, the average floor height is 2.84 meters.

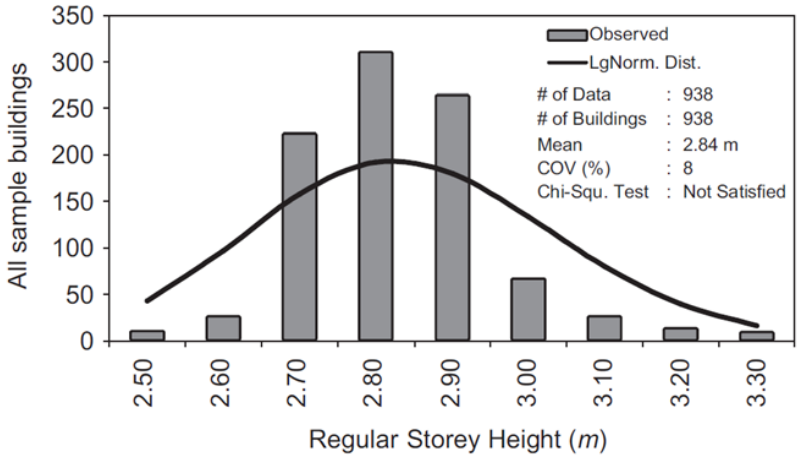


Figure 2.2. Distribution of standard-story height for RC residential buildings (Bal et al., 2008).

2.2.3. Span Length and Number of Spans

Figure 2.3 depicts the beam length distribution of 6150 observed buildings from the Marmara region, as detailed in Bal et al. (2008). The average beam length in this distribution is 3.37 meters, with a coefficient of variation value of 0.38. In addition, the discoveries of Azak et al. (2014) reveal an average span length calculated as 3.55 m with a standard deviation of 0.68 m. For the purposes of this study, beam lengths are chosen as 2.5, 3, and 4 meters to encompass the predominant portion of the interested building stock.

Additionally, based on the information from Azak et al. (2014), the mean plan dimensions of Marmara region buildings are 9.58 meters for the short dimension and 13.73 meters for the long dimension. The standard deviation for the short dimension is

3.64 meters, and for the long dimension, it is 7.84 meters. Based on these statistics and considering beam lengths of 2.5, 3, and 4 meters, it can be assumed that the number of spans is either 3 or 4.

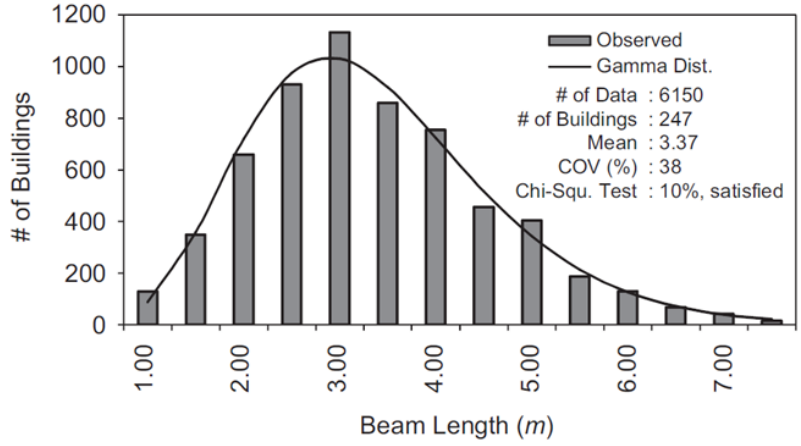


Figure 2.3. Beam length distribution for all buildings (Bal et al., 2008).

2.2.4. Structural Elements and Dimensions

In this thesis, all building plans and columns were supposed to be square for simplicity. Column dimensions are conditioned on the number of stories. Column dimensions are considered 25, 30, and 35 cm for 3-story buildings and 30, 35, and 40 cm for 5-story buildings. When determining the dimensions of columns, reference was made to the column dimensions based on the number of stories provided by Azak et al. (2014).

As per the findings of Bal et al. (2008), the average beam depth was recorded as 0.60 meters for constructions erected before the year 2000. Despite the minimum beam dimensions being specified as 20x30 cm in the ABYYHY (1975), this study has adopted a beam width of 25 cm.

2.2.5. Material Properties

The strength of structural materials was determined using statistical studies and distributions shared in the article by Bal et al. (2008). To encompass the concrete compressive strength of the building inventory in the Marmara region, values of 5, 12, 20, 28, and 35 MPa were considered based on the distribution seen in Figure 2.4 provided by Bal et al. (2008). For structures employing S220 steel, the average yield strength of steel was 371.13 MPa in the relevant article. A steel yield strength of 370 MPa was included in the calculations.

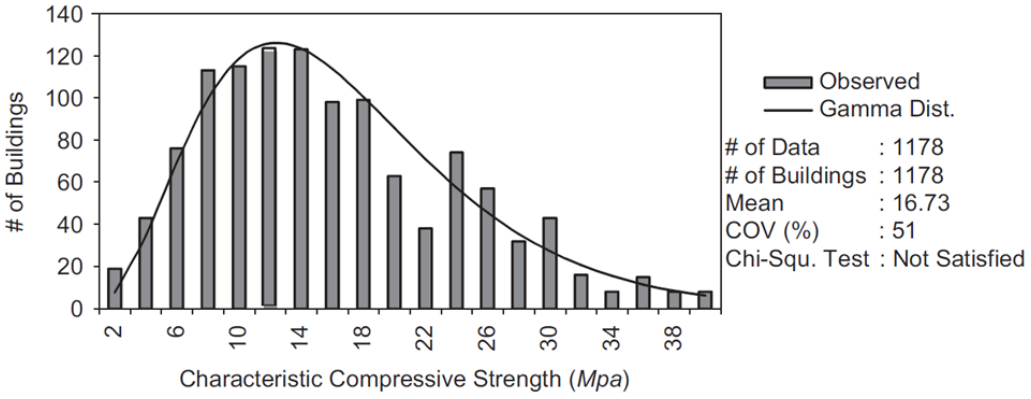


Figure 2.4. Variation in the 28th-day characteristic compressive strength of buildings located in Marmara (Bal et al., 2008).

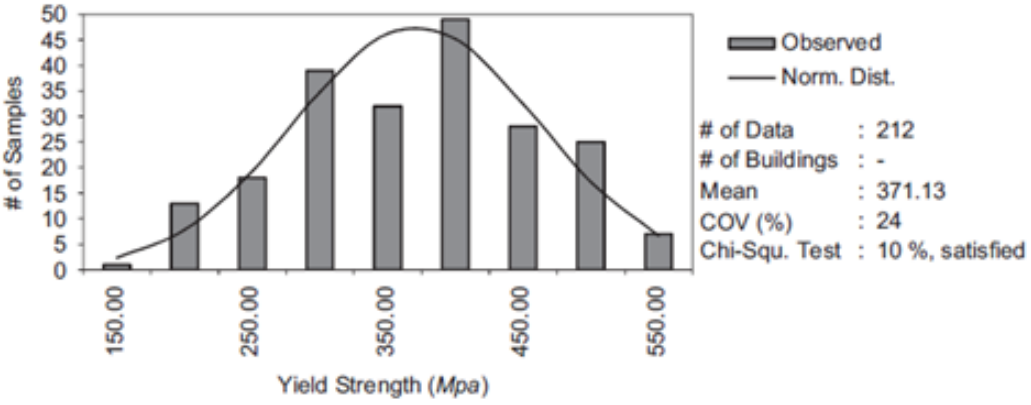


Figure 2.5. Dispersion of yield strength for reinforcing steel of S220 MPa grade (Bal et al., 2008).

2.2.6. Reinforcement Ratios

During the execution of nonlinear analyses, multiple methodologies are employed to elucidate the behavior of a given section, encompassing hinge definition, fiber section definition, and the finite element method. In this study, the fiber section definition method is utilized. The reinforcement ratio and layout are important in the fiber section definition of the structural analysis model. The reinforcement ratio in beams and columns is not determined through statistical studies; however, it is established to meet the minimum requirements specified in ABYYHY (1975) for reinforcement ratio and layout. An equal distribution of steel material fibers has been assigned to the four corners of the rectangular column and beam sections. For column sections, a ratio of 1%, and for beams, a ratio of 0.4% of reinforcement has been used. The fiber definition and reinforcement layout description are provided in detail in Section 2.3 of the thesis.

2.2.7. Local Soil Conditions

In this thesis, the variability of soil conditions is addressed through the incorporation of three distinct V_{S30} values, specifically 1130, 560, and 270 m/s, which respectively characterize good, moderate, and poor soil classes.

2.3. Modeling of the Study Buildings

Structural analysis models were systematically created using the OpenSessPy framework (McKenna et al., 2010) through a Python-based program (Python Software Foundation, 2020) specifically developed for this purpose. The simulations of the buildings were performed by modeling two-dimensional (2D) finite element models. The program and associated OpenSeesPy scripts can be found in Appendix A.

For simplicity, all building plans and columns were presumed to be square. The central axes of three-dimensional residential-type buildings, designed to represent the features of structures in the Marmara region, were simplified and modeled as two-

dimensional frame systems. Loads were applied based on the dimensions of each structure to maintain consistency in the mass of the structures. The loads applied to both the interior and exterior walls of the partition, assumed to be 2.5 kN/m^3, were distributed onto the beams. The slabs were regarded as reinforced concrete with a density of 25 kN/m^3 and a thickness of 15 cm. Superimposed dead and live loads were estimated at 2 kN/m^2. In seismic analysis, the mass is commonly calculated as a combination of dead and live loads (G + 0.3Q).

The buildings were simulated as 2D bare frame systems with reinforced concrete moment-resisting characteristics. Figure 2.6 is included to visualize the structural models and two sample buildings. Figure 2.7 displays the plan of the 3-span structure, indicating the considered axis of the building and the loading tributary areas. The A-A axis serves as an example of the 2D structural model. As seen, axis beams are loaded from both sides of the slab.

In the OpenSeesPy framework, it was assumed that beam-column links exhibit rigidity, and columns were fixed at the base level for all degrees of freedom. Every element was modeled using a 'nonlinearBeamColumn' object to ensure nonlinear behavior in the structure. This object relies on the non-iterative (or iterative) force formulation and accounts for the spread of plasticity along the element. The definition of this element is employed to calculate element deformations by integrating the internal forces of the section. This study designated five integration points along the frame members to introduce distributed plasticity.

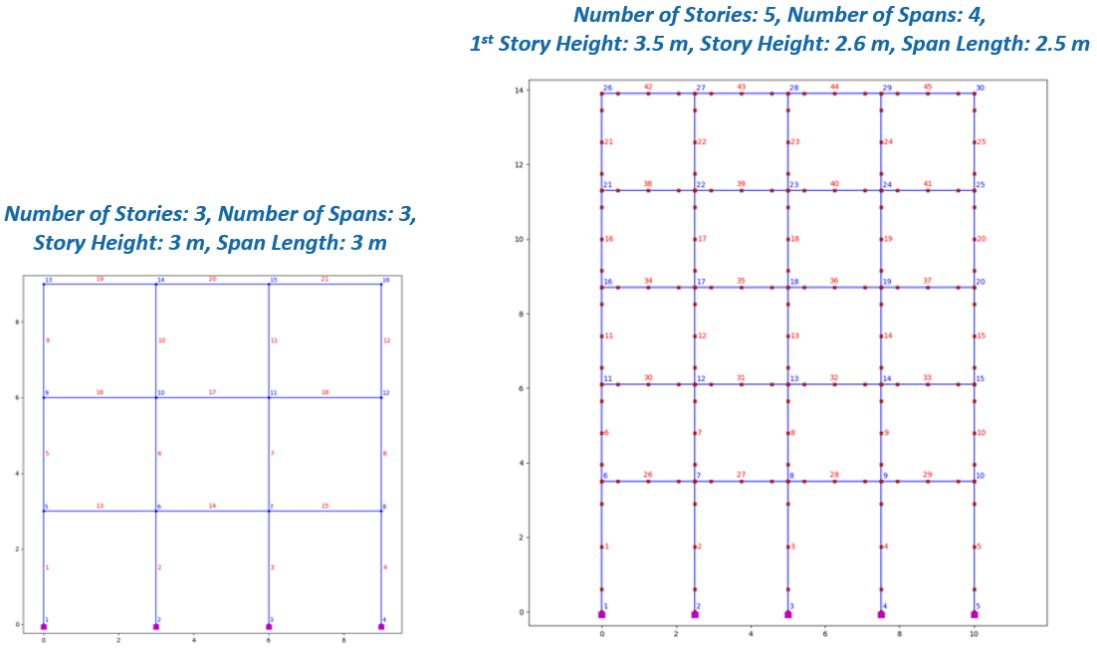


Figure 2.6. Example building models in the OpenSeesPy framework: 3- and 5-story buildings.

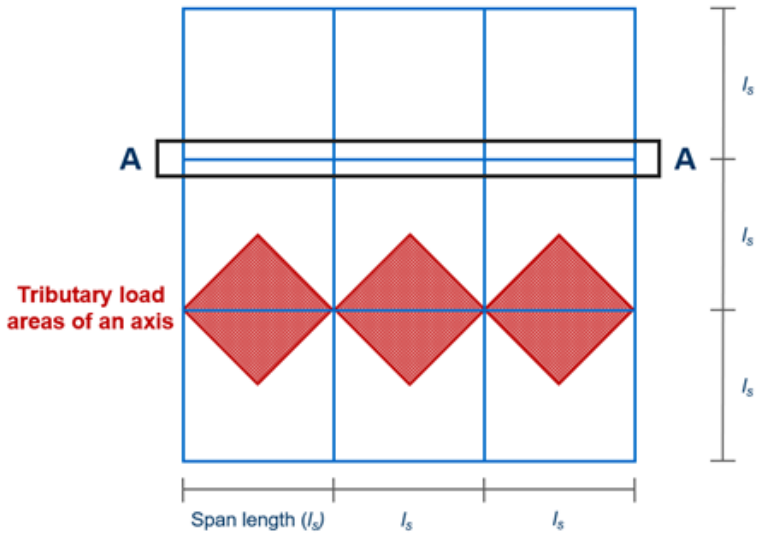


Figure 2.7. The plan view of the 3-span structure displays the structural model's considered axis, serving as an example of loading tributary areas.

In OpenSeesPy, distributed plasticity is implemented through the 'RCSection2d'

object, encapsulating the fiber model of a rectangular reinforced concrete section. Material properties for the fiber section are specified using 'Steel01' for steel and 'Concrete04' for concrete fibers within the OpenSeesPy framework. Section 2.2.5 defines the yield strength of reinforcing steel and the compressive strength of concrete based on the Marmara Region building inventory. While the 'RCSection2d' object enables the modeling of confined and unconfined concrete fibers (see Figure 2.8), the cover and core regions of column fibers are assigned as unconfined. This choice is made because, in pre-code buildings, column stirrups were not adequately tucked and were sparsely placed. As shown in Figure 2.8, the reinforcing steel fibers are evenly distributed at the four corners of the section for both columns and beam elements.

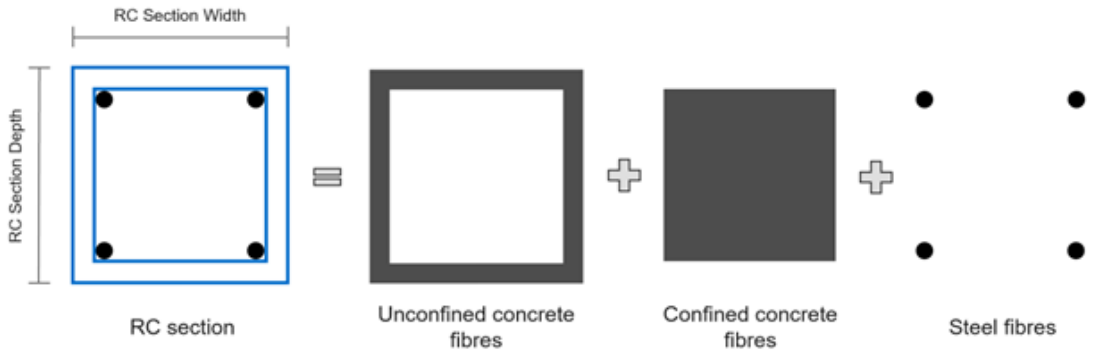


Figure 2.8. Fiber discretization of a reinforced concrete section.

Concrete material is defined in OpenSessPy using 'Concrete04' to represent the identical envelope curve developed by Mander et al. (1988). For each concrete compressive strength, the unconfined concrete strain at maximum stress and ultimate stress is considered to be 0.002 and 0.005, respectively. The elasticity modulus (E_c) is calculated using Equation 2.1.

$$E_c = 5000 \times \sqrt{f_c} \quad MPa \tag{2.1}$$

where f_c is the 28^{th}-day compressive strength of concrete.

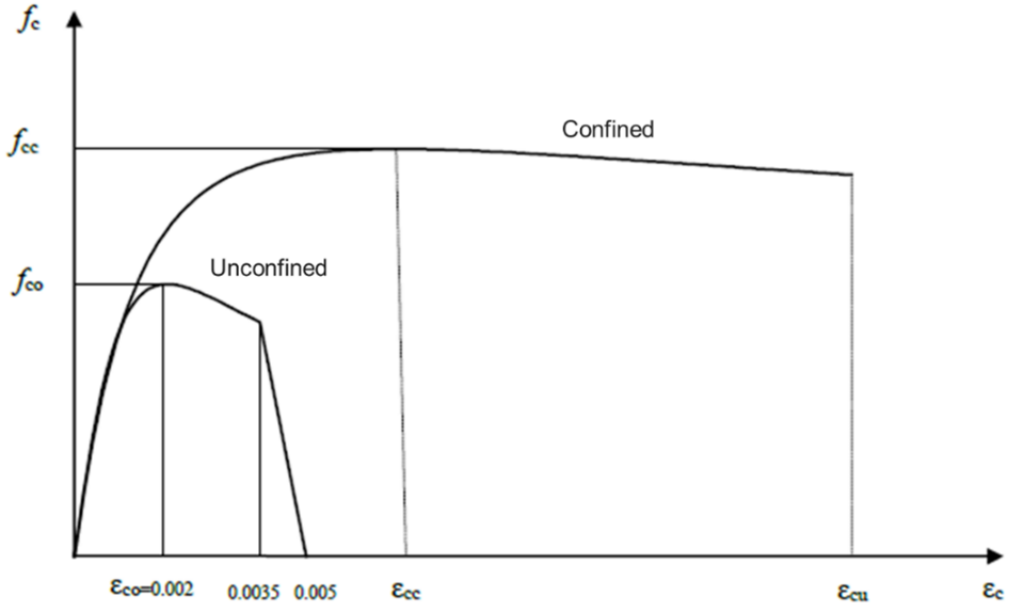


Figure 2.9. Confined and unconfined concrete stress-strain graphics.

The characteristic yield strength of the rebar steel is 370 MPa, and the ultimate strength is 444 MPa, with a strain-hardening ratio of 3.4×10^{-3} (the ratio between the post-yield tangent and the initial elastic tangent). The elasticity modulus of the reinforcing steel is used as 2×10^5 MPa.

Table 2.2. Information on reinforcing steel.

$\mathbf{f_{sy}}$ **(MPa)**	$\mathbf{E_s}$ **(MPa)**	**b**
370	2×10^5	3.4×10^{-3}

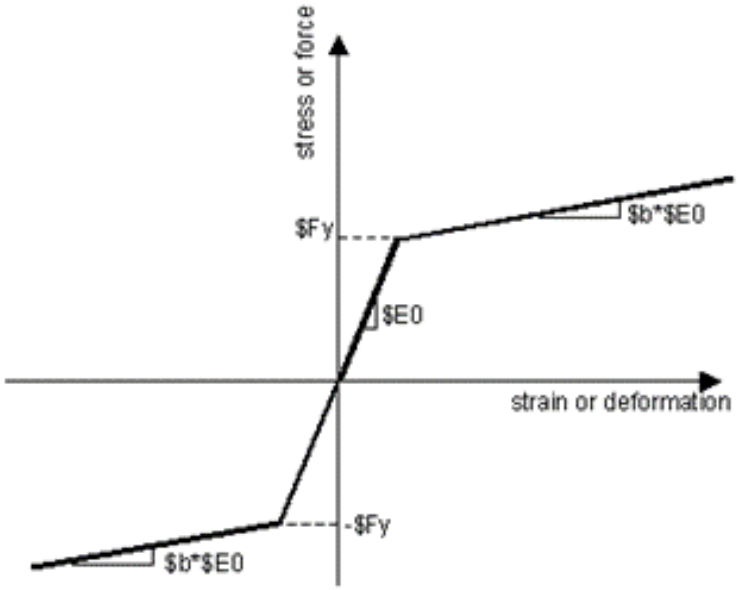


Figure 2.10. Stress-strain graphics of reinforcing steel (OpenSees, Steel01).

2.3.1. Model Verification

In order to validate the structural models, static analyses were conducted on these two-dimensional models using SAP2000 (Computers and Structures, Inc., 2000) and then compared while maintaining the same assumptions and load conditions.

The modeling and analysis in SAP2000 followed a linear-elastic procedure and were employed to validate the model generated with OpenSeesPy. The fundamental vibration periods and modal participating mass ratios obtained from free vibration analyses of the building's finite element models in SAP2000 and OpenSeesPy were found to fit. Table 2.3 compares vibration periods and participating mass ratios for

3-story buildings with identical features.

Table 2.3. Comparison of modal analysis results from OpenSeesPy and SAP2000 for 3-story buildings.

Mode	**Period (OpenSeesPy)**	**Period (SAP2000)**	**Mass Participation Percent (OpenSeesPy)**	**Mass Participation Percent (SAP2000)**
1^{st}	0.436	0.420	0.90 (UX)	0.91 (UX)
2^{nd}	0.152	0.148	0.39 (RY)	0.38 (RY)
3^{rd}	0.101	0.100	0.01 (UX)	0.01 (UX)
4^{th}	0.042	0.041	0.71 (UZ)	0.73 (UZ)
5^{th}	0.036	0.034	0.31 (RY)	0.43 (RY)

2.4. Selection of Ground Motion Records

Earthquake records were utilized in the context of NLTH analyses. The ground acceleration-time records of earthquakes were sourced from the CyberShake dataset and the AFAD-TADAS data repository.

A dataset was created by selecting earthquake records from AFAD-TADAS, originating within the territory of Turkiye and displaying fault types and characteristics similar to those found in the Marmara region. These records were merged with the CyberShake earthquake dataset shared in the research accomplished by Baker et al. in 2021. The earthquake dataset enables us to access well-documented ground shakes worldwide from different stations and earthquakes in Turkiye. Selected ground motions are appropriate for seismological specifications for the Marmara region, and they match with the conditional mean spectrum to create an optimal list of ground motion records.

When selecting earthquake records for structural analysis models, the structure's period and site conditions were taken into consideration. For the 1080 models, earth-

quake records were chosen while categorizing the structures into ten groups based on their first natural vibration periods to reduce the computational load. The shortest and longest first vibration periods were divided into ten equal intervals, and the period classes for the structures were determined based on these intervals. When choosing earthquake records, the period corresponding to the midpoint of the period class intervals was considered.

In the selection of earthquake records, it was assumed that variations in the structure's first vibration period within ±0.09 seconds would not significantly affect the analysis results. The classification included in Table 2.4 provides the boundary values of the period classes, the period value considered when selecting earthquake records, and the number of structures in each period class.

Table 2.4. List of period classes for recording selections.

Fundamental Vibration Period Class	**Period Intervals**	**Center of Period Intervals**	**Number of Building Models**
1	0.19 – 0.28	0.24	70
2	0.28 – 0.37	0.33	189
3	0.37 – 0.46	0.41	255
4	0.46 – 0.54	0.50	234
5	0.54 – 0.63	0.59	161
6	0.63 – 0.72	0.67	90
7	0.72 – 0.80	0.76	47
8	0.80 – 0.89	1.85	21
9	0.89 – 0.98	1.94	9
10	0.98 – 1.07	1.02	4
Total Number of Buildings			1,080

In 1766, earthquakes caused significant ruptures along the North Anatolian Fault Zone (NAFZ) beneath the Sea of Marmara. However, a 150 km stretch remained unruptured near the Istanbul metropolitan area, referred to as the 'Marmara seismic gap.' It is anticipated that this region will likely experience a major earthquake in the

foreseeable future. Considering time dependence and stress transfer, there is a 42% probability of a significant seismic event occurring in Istanbul within the next 30 years (Murru et al., 2016). Many researchers drew attention to the problem of earthquakes larger than moment magnitude 7 in the Marmara region (Parsons et al., 2000; Parsons et al., 2004; Hubert-Ferrari et al., 2000; Atakan et al., 2002; Erdik et al., 2004; Kalkan et al., 2009; Gülerce and Ocak, 2013).

Murru et al. (2016) introduced a multi-segment earthquake with simultaneous rupture of the West Marmara, Central Marmara, and Cinarcik segments, characterized by a right-lateral strike-slip mechanism. In this thesis, this earthquake is assumed to be the dominant event with a magnitude of 7.56. The location of the analyzed buildings is presumed to be 10 km from the rupture projection. However, the selection of ground motion records is based on a source-to-site distance of less than 50 km. Additionally, this study incorporates the variability of soil conditions by considering three values of V_{S30}: 1130, 560, and 270.

Based on this information, the selection of eleven ground motion records and corresponding scale factors is achieved through the application of Baker and Lee's (2018) MATLAB algorithm. This selection is made for each V_{S30} value, aligning with the target distribution conditioned on the first vibration period class. Among the 11 earthquakes selected for each period class and V_{S30} value, six are drawn from the global CyberShake dataset, while the remaining five are derived from the AFAD-TADAS dataset, featuring earthquakes that transpired within the geographical confines of Turkiye.

The target response spectrum, which is conditioned on the first vibration period, is calculated by considering the dominant event as the sole seismic source for the hazard analysis. The S_a values have been computed for the site's V_{S30} value. Ground motion is determined through a distribution that accounts for the inherent uncertainties in ground motion calculations. 2.5 and 97.5 percentile values are utilized to represent the distribution of S_a. Figure 2.11 shows that the S_a(T) values of selected strong ground motion records remain within the distribution. In this figure, the solid blue

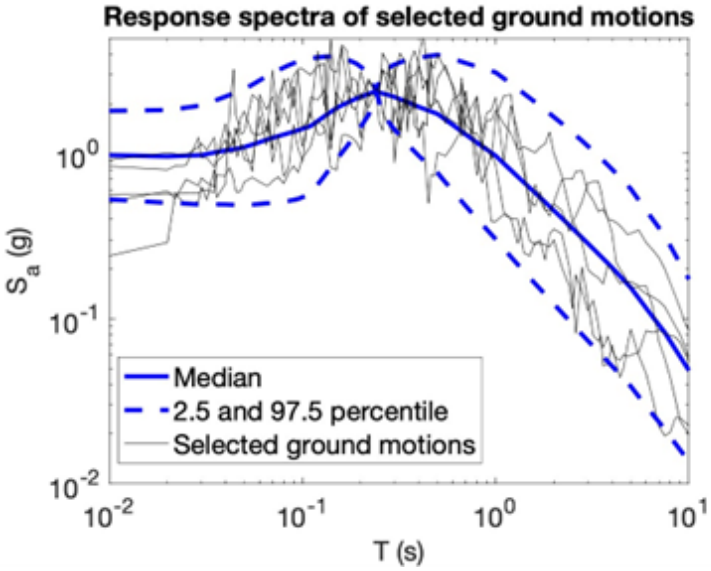
line indicates the median value, and the dashed lines stand for 2.5 and 97.5 percentiles of the conditional mean spectrum.

For example, selected earthquake records and the records' scale factor, magnitude, station, and intensity measures (PGA, PGV, and PGD) are given in Table 2.5 for fundamental vibration period class 1 and 270 m/s V_{S30}. Additionally, the response spectra of scaled recordings from the table are visualized in Figure 2.11. The earthquakes from the AFAD-TADAS and CyberShake databases are presented in separate graphs.

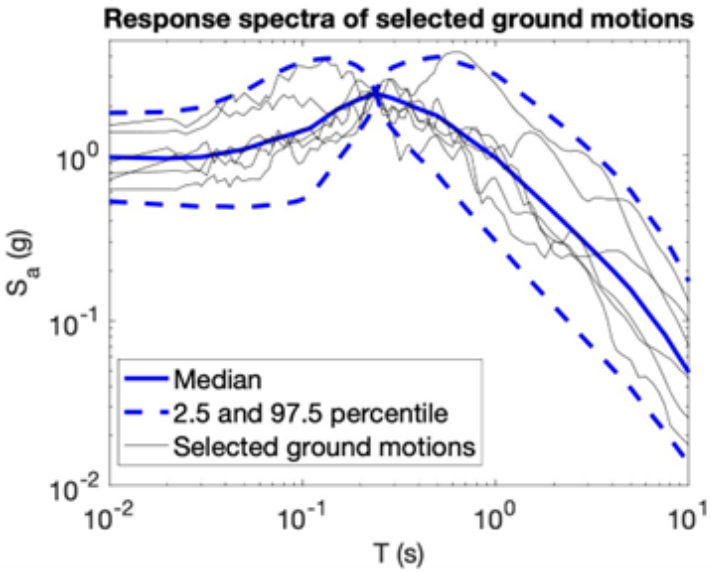
Figure 2.12 depicts the acceleration-time graph of the seismic activity recorded on February 2, 2023, in Gaziantep (M_w=7.8). This recording is the east component of the earthquake at AFAD-TADAS station number 3145. Furthermore, the figure also includes the velocity-time, displacement-time, and cumulative absolute velocity-time graphs corresponding to the same seismic event. These graphs are based on the data presented in the 3rd row of Table 2.5, but these graphics represent the earthquake record without applying a scale factor.

Table 2.5. Selected ground motions for period class 1 and V_{s30} = 270 m/s.

Selected Earthquake	**M_w**	**Horiz. Comp.**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
Pazarcik Kahramanmaras (02.06.2023)	7.8	East	3135	0.67	0.90	43.84	32.43
Pazarcik Kahramanmaras (02.06.2023)	7.8	East	3144	1.05	0.83	140.0	139.3
Pazarcik Kahramanmaras (02.06.2023)	7.8	East	3145	0.79	0.56	121.4	84.84
Izmit Earthquake (17.08.1999)	7.4	East	8101	1.59	0.59	89.64	36.84
Pazarcik Kahramanmaras (02.06.2023)	7.8	East	8002	1.16	0.24	44.40	86.47
S. San Andreas; PK+ CH+CC+BB+NM+ SM+NSB+SSB	8.0	2nd	s351	2.16	0.57	45.41	78.99
S. San Andreas	7.3	1st	Silverwood Lake	3.79	1.40	219.6	185.6
Anacapa-Dume, alt 2	7.3	2nd	p24	1.37	0.80	121.8	103.6
Clamshell-Sawpit	6.5	1st	s480	2.65	0.70	71.25	76.51
S. San Andreas	8.0	1st	s165	0.63	0.91	81.90	65.86
S. San Andreas;BB+ NM+SM+NSB+SSB	8.0	1st	s566	1.98	1.53	374.7	294.5



(a)



(b)

Figure 2.11. Response spectra of selected ground motions for period class 1 and 270 m/s V_{s30} (a) AFAD-TADAS database (b) CyberShake database [Baker and Lee's (2018) MATLAB Algorithm is used for plotting these figures.].

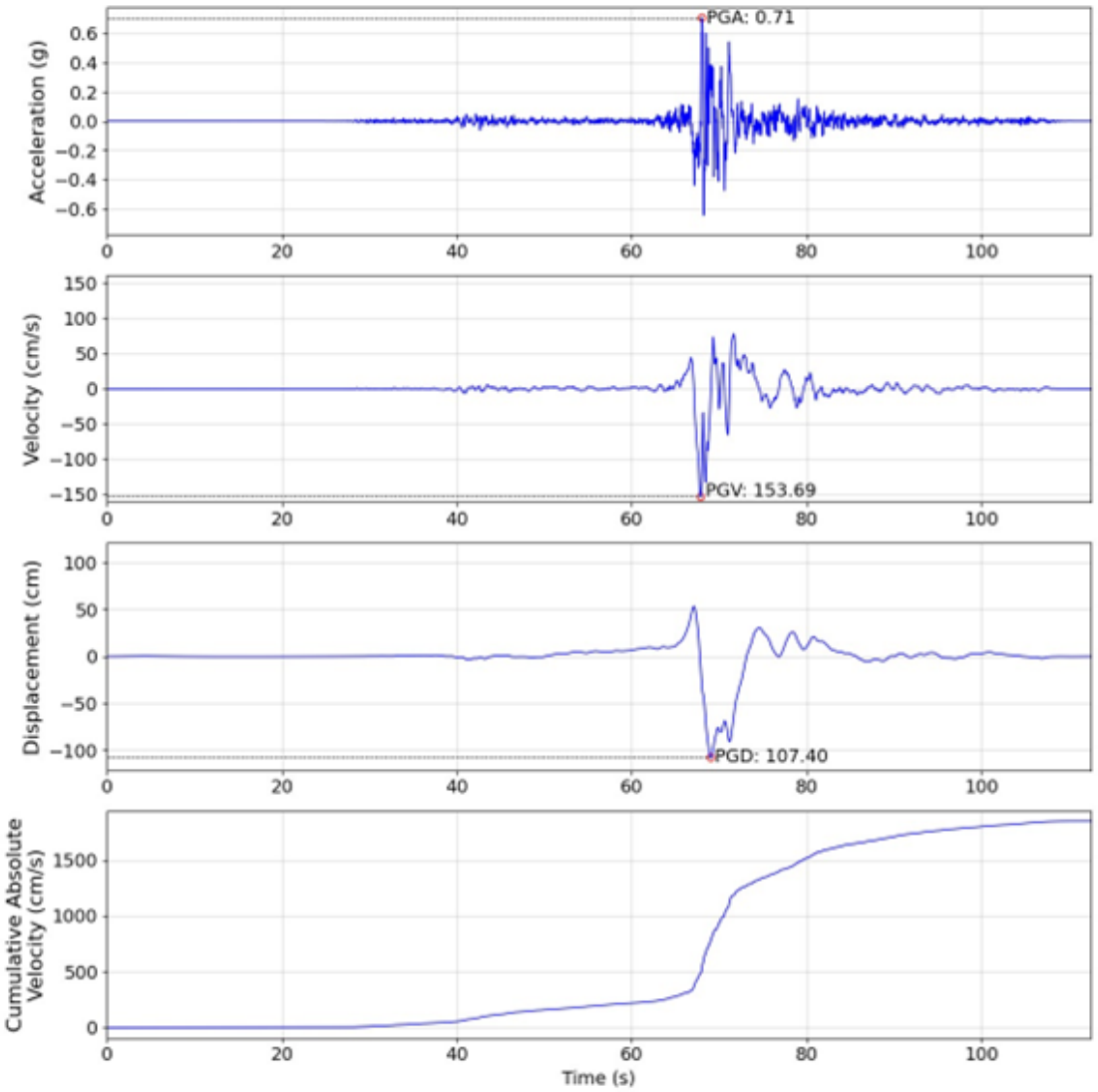


Figure 2.12. Example time series for a selected ground motion (02.06.2023 Pazarcik Kahramanmaras Earthquake (M_w=7.8), Station 3145, East Component).

2.5. Intensity Measures Based on Earthquake Record

The Intensity Measure (IM) signals the degree of shaking resulting from intense ground motion at a specific location. Single-parameter intensity definitions have been developed to understand the damage caused by earthquakes. These parameters include the maximum acceleration of ground motion, the frequency content of motion, and the duration of ground motion, depending on earthquake acceleration records or spectra.

Since these parameters are based on the acceleration records of earthquakes, they can be considered a more realistic absolute measure of earthquakes compared to the Mercalli Intensity or Richter Local Magnitude (Celep, 2019). When preparing the artificial dataset to represent the intensity of an earthquake, IMs corresponding to the ground motion records were added to the dataset.

For chosen records, six widely used IMs are computed, categorized as peak ground motion parameters [Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV), Peak Ground Displacement (PGD)], spectral parameters [Spectral Acceleration (S_a), Spectral Displacement (S_d)], and energy parameter [Cumulative Absolute Velocity (CAV)] (Aquib et al., 2022).

PGA is derived from the absolute maximum values in the acceleration-time series. PGV is calculated from the absolute maximum values of the velocity-time series, obtained by taking the derivative of the acceleration-time series. Similarly, PGD is obtained from the absolute maximum values of the displacement-time series, derived by taking the derivative of the velocity-time series.

The meaning of peak values is the absolute maximum values of the time series of records. Acceleration is the derivative of velocity, while velocity is the derivative of displacement.

$$PGA = \max(|\ddot{u}_g|) \tag{2.2}$$

$$PGV = \max\left(\left|\int \ddot{u}_g dt\right|\right) = \max(|\dot{u}_g|) \tag{2.3}$$

$$PGD = \max\left(\left|\iint \ddot{u}_g dt dt\right|\right) = \max(|u_g|) \tag{2.4}$$

S_a and S_d represent earthquake ground motion intensity across various frequencies, commonly applied in structural engineering and seismology to assess ground shaking at specific locations. They are expressed as frequency functions, denoted as $S_a(T)$ and $S_d(T)$, with T representing the period of vibration. In this study, the vibration period

is assumed to be the first mode period of the buildings. The Newmark-Beta method is used to calculate the displacement of a single degree of freedom (SDOF) system for each time interval of ground motion (u_t). In Equation 2.6, ω_0 is the natural frequency of the SDOF system.

$$S_d(T) = max|u_t| \tag{2.5}$$

$$S_a(T) = max|\ddot{u}_t| \approx {\omega_0}^2 S_d = \omega_0 S_v \tag{2.6}$$

The IMs, as mentioned earlier, rely on structural configuration and are calculated from spectral values. In contrast, certain IMs utilize accelerograms directly, called energy IMs. Energy denotes the total energy per unit weight absorbed by an infinite set of undamped single-degree-of-freedom oscillators with a uniform distribution of periods on (0 ∞). CAV is determined by calculating the area under the absolute accelerogram, signifying the continuous accumulation of acceleration in a ground motion record. CAV serves as a valuable indicator of the initiation of structural damage. In Equation 2.7, T_{sig} is the significant duration, and *a(t)* is the acceleration time series (Aquib et al., 2022).

$$CAV = \int_0^{T_{sig}} |a(t)|\, dt \tag{2.7}$$

2.6. Nonlinear Time History Analyses

Nonlinear analyses with ground motion records were conducted using the Openseespy framework. The methods employed for these dynamic analyses are explained in this section.

P-Delta effects factored into the calculations. The P-delta effect, regarded as a geometrically induced nonlinearity, emerges as a second-order (non-linear) phenomenon when an axial load interacts with lateral displacement, resulting in the creation of an

extra moment as illustrated in Figure 2.13. If the displacement or the axial load (or both) are large enough, the P-Delta component of the bending moment M cannot be neglected (Burgos and Silva, 2023).

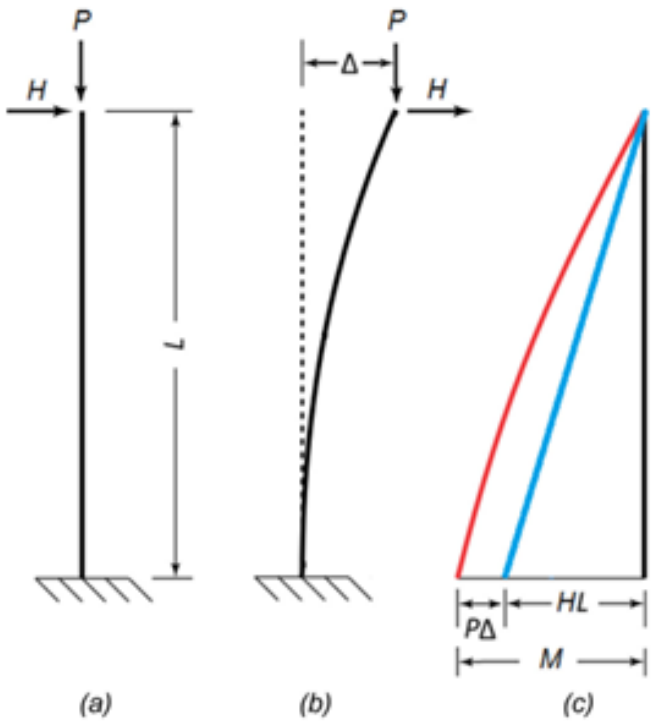


Figure 2.13. P-Delta effect: (a) undeformed configuration, (b) deformed configuration, (c) bending moment distribution (Burgos and Silva, 2023).

The Rayleigh damping ratio method was employed for dynamic analyses. When applying the Rayleigh damping ratio, it is crucial to select modes where all modes' damping ratios significantly contribute to the behavior, as emphasized by Chopra (2013). A damping ratio of 5 percent was adopted for the reinforced concrete frame system, and this damping ratio was enforced for the 1st and 4th modes. Consequently, the damping ratios for the 2nd and 3rd modes were slightly lower, and the 5th mode had a higher damping ratio. For the subsequent modes, their damping ratios would be considerably higher, rendering their contributions insignificant.

The damping ratio is correlated with stiffness and mass. The coefficient a_0, associ-

ated with mass-proportional damping, is related to the stiffness-proportional damping coefficient a1. The Rayleigh damping ratio is formulated as in Equation 2.8, and graphs depicting mass-proportional and stiffness-proportional damping, along with Rayleigh damping, are presented for an analyzed example structure in Figure 2.14.

$$\zeta_n = \frac{a_0}{2\omega_n} + \frac{a_1\omega_n}{2} \tag{2.8}$$

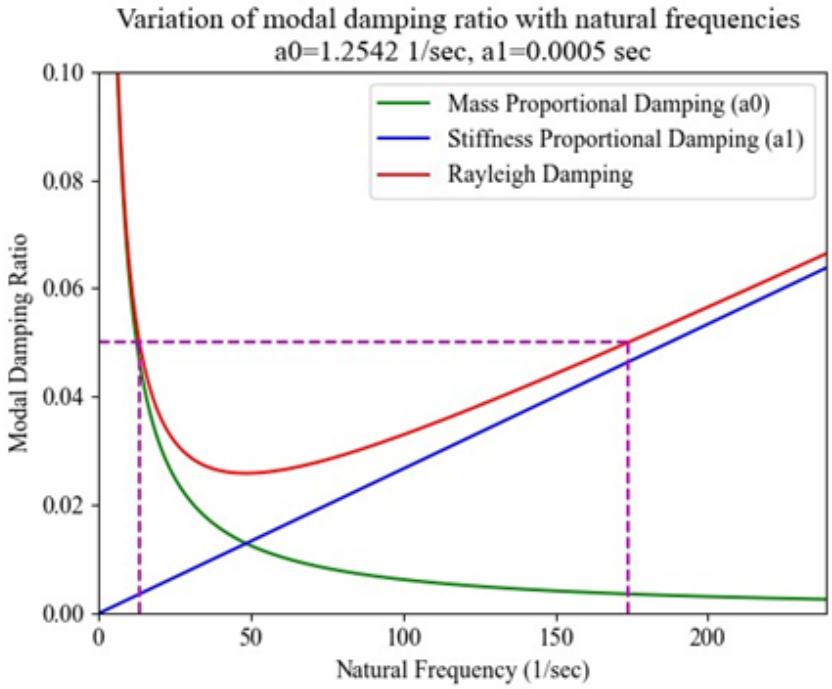


Figure 2.14. The variation of mode damping ratios with natural frequency for an analyzed example structure.

Numerical integration of the equations of motion was performed using Newmark's time integration method with the average constant acceleration method (β=0.25 and γ =0.5). The equilibrium equations were solved at each time step of the seismic record using the Newton-Raphson algorithm, which employs the tangent stiffness during each equilibrium iteration within the time step.

Numerous factors can lead to convergence challenges in nonlinear analysis, in-

cluding excessive displacements, inadequate algorithms for solving the nonlinear equilibrium equations, inappropriate material definitions, or insufficient tolerance. In addressing these convergence errors within OpenSees, the script was modified with diverse solvers and adjustments to tolerances to successfully conclude the analysis. Additional methods and procedures are detailed in the provided OpenSeesPy scripts, accessible in Appendix A.

2.7. Assessment of the Analysis Results

Nonlinear time-history analyses were performed for 1080 artificial buildings, using three different soil conditions and 11 actual earthquake records. The results of each earthquake record applied to an artificial building were added to the dataset as separate data. Artificial buildings with the aim of representing structures in the Marmara region, diversified in terms of structural characteristics and soil conditions, were differentiated in terms of earthquake loads. As a result, analyses were added to the dataset, totaling 35640 analysis data points. The inputs of the dataset include structural features, ground conditions, and intensity measures of earthquake records. The output of the dataset is the level of damage to the structure. A portion of the dataset is provided in Table 3.1.

In this thesis study, the performance of structures was expressed in terms of damage levels. The Maximum Interstory Drift Ratio (MIDR) was used as an engineering demand parameter (EDP). To calculate the MIDR, the horizontal displacement values at each floor of the structure were recorded for each step of the earthquake record applied during each dynamic analysis. As a result of the time-history analysis, relative inter-story drift values were calculated for each step. The calculated maximum interstory drift value was divided by the floor height to compute the MIDR. This value was utilized for structural analysis results and damage level assessment.

The categorization of the obtained MIDR for each structure from the analysis results was determined based on the "drift ratio limits" calculated for non-ductile

moment-resisting frame systems in the Ghobarah (2004) study.

Table 2.6. Limits of inter-story drift ratio (%) linked to damage states (Ghobarah, 2004).

Damage States	**Limits of Inter-story Drift Ratio (%)**
No Damage	0.00 – 0.10
Slight Damage	0.10 – 0.20
Moderate Damage	0.20 – 0.50
Extensive Damage	0.50 – 1.00
Complete Damage	>1.00

Regarding the considered inter-story drift ratio limits, 57.2% (20395 buildings) of the target data in the dataset are categorized as complete damage level and 26.7% (9514 buildings) are in the "extensive" damage category. The distribution of damage levels resulting from the analysis results is depicted in Table 2.7.

Table 2.7. Distribution of damage levels.

Damage State	**Number of analyses**	**Ratio**
Complete Damage	20395	57.2%
Extensive Damage	9514	26.7%
Moderate Damage	4914	13.8%
Slight Damage	702	2.0%
No Damage	115	0.3%
Total	**35640**	**100%**

As a result of this damage distribution, it is observed that a significant portion of the conducted analyses have high damage levels, indicating that the dataset is imbalanced. This outcome has been interpreted as being due to the earthquake recordings

selected with scale factors, which produce excessive ground shaking intensity measures. To overcome the problem of an imbalanced dataset, a second scale factor, as applied in the Wu and Sarno (2023) study, was calculated for the earthquake records. The calculation for the second scale factor, as shown in Equation 2.9, depends on the ratio of the complete damage level limit to the average MIDRs obtained in the analysis set. Since the average MIDR value in the analysis set is greater than the complete damage level limit, the second scale factor is lower compared to the first scale factor.

$$SF_2 = SF_1 \times \frac{IDR_{CompleteDamageLimit}}{MIDR_{Mean}} \tag{2.9}$$

In Equation 2.9, $IDR_{\text{Complete Damage Limit}}$ and $MIDR_{\text{Mean}}$ indicate 0.01 and 0.0461, respectively. Consequently, the calculation of the second scale factor involves reducing the first scale factor by multiplying it by 0.22. By using the SF_2, earthquake records with lower amplitudes were obtained, and NLTH analyses were repeated. When the results of the two sets of analyses were combined, a balanced dataset was prepared with 71280 completed analyses. The results as damage state distributions for both scale factors are illustrated in Figure 2.15, and the details of these distributions can be seen in Table 2.8.

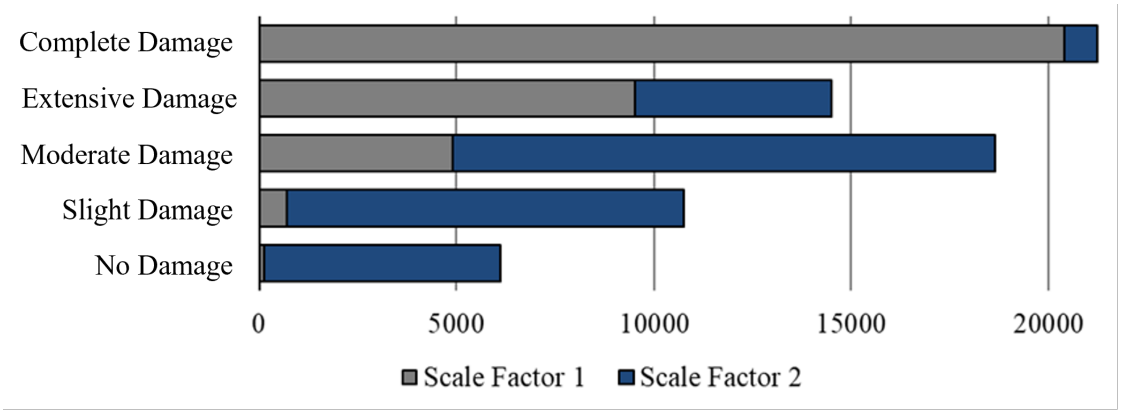


Figure 2.15. The outcomes of the NLTH analysis are presented as damage states corresponding to SF_1 and SF_2.

Table 2.8. The NLTH analysis results as damage states for SF_1 and SF_2, overall results, and ratios.

Damage State	**SF1**	**SF2**	**Overall**	**Ratio**
No Damage	115	5996	6111	8.6%
Slight Damage	702	10042	10744	15.1%
Moderate Damage	4914	13743	18657	26.2%
Extensive Damage	9514	4996	14510	20.4%
Complete Damage	20395	863	21258	29.8%
Total	**35640**	**35640**	**71280**	**100%**

3. DEVELOPMENT OF MACHINE LEARNING MODELS

3.1. Introduction

Machine learning in civil and earthquake engineering has been limited compared to other disciplines; nevertheless, it has inflated in recent years. Some noteworthy studies in this area include research on the categorization of structural damage caused by earthquakes (Zhang et al., 2018; Mangalathu and Burton, 2019; Mangalathu et al., 2020; Ulku et al., 2022; Roeslin et al., 2020); the rapid loss assessment (Stojadinović et al., 2021) as well as the development of fragility curves (Kiani et al., 2019; Wu and Sarno, 2022).

Machine learning involves instructing computers to enhance performance based on sample data or prior knowledge. In this process, a model characterized by specific parameters and learning entails running a computer program to refine these parameters by leveraging training data or prior experiences. The model can make future predictions, extract insights from data, or perform both tasks simultaneously. ML can be classified into three main types: supervised, unsupervised, and reinforcement learning. In the first one, the aim is to develop a mapping from input to output, guided by a supervisor who provides the correct values. Unsupervised learning exclusively involves input data without associated output values. The goal is to discern inherent patterns and regularities within the input dataset. Reinforcement learning is employed in cases where a system's output consists of action sequences. In such instances, individual actions matter less than the overall policy, which is the sequence of correct actions needed to reach a goal. The focus is on assessing policy quality and learning from past successful action sequences to create effective procedures (Alpaydin, 2014).

Randomly selected four data from the artificial dataset are presented in Table 3.1. The dataset contains input and output information, so supervised learning algorithms have been employed. Supervised learning can be categorized into two main types:

classification and regression algorithms. Since the target values in the dataset represent the level of structural damage with specific classes (No Damage, Slight, Moderate, Extensive, and Complete), and in other words, the outputs are not numerical, the machine learning model has been trained using classification algorithms.

Table 3.1. Randomly selected data points from the dataset; inputs and outputs.

Data Label	**Input/Output**	**Data - 1**	**Data - 3799**	**Data - 71142**	**Data - 71280**
Number of Stories	Input	3	3	5	5
Number of Spans	Input	3	3	4	4
Span Length (m)	Input	2.5	2.5	4	4
Story Height (m)	Input	2.6	2.8	3	3
Column Dimension (m)	Input	0.25	0.25	0.4	0.4
Concrete Strength (MPa)	Input	5	12	35	35
Commercial Use	Input	Yes	No	No	No
Soil Condition (m/s)	Input	1130	1130	1130	270
PGA (g)	Input	0.362	0.083	0.036	0.177
PGV (cm/s)	Input	17.668	8.425	8.887	43.298
PGD (cm)	Input	13.069	8.202	10.449	34.031
Sa(T1) (g)	Input	0.516	1.047	0.169	0.673
Sd(T1) (cm)	Input	0.004	0.005	0.001	0.005
CAV (m/s)	Input	7.767	8.576	1.482	6.318
MIDR	Output	0.0108	0.0014	0.0008	0.0053
Damage State	Output	Complete	Slight	No Damage	Extensive

3.2. Data Pre-Processing

After generating the dataset, a series of pre-processing steps are applied to create the final dataset used in developing the machine learning prediction model. Pre-processing techniques apply to assure optimal performance. Normalization and transforming categorical data into numerical values are the expected techniques in data pre-processing, as they contribute to refining the dataset and fostering improved performance and interpretability in the subsequent stages of model development.

In this process, columns are assumed to be square, as outlined in Section 2.2.4, and their dimensions are transformed into column area to better represent column characteristics. Consequently, the original column dimensions are removed from the final dataset while the calculated column area is added. Certain stable features, such as beam dimensions and steel strength, are excluded from the dataset. Additionally, the only categorical feature, "Commercial Use," is converted into a binary system. Therefore, the categorical feature is transformed into a numerical one. Apart from stable ones, numerical features are normalized to prevent any single numerical feature from dominating the data due to its scale. In essence, normalization ensures a more balanced contribution of features to the model. These transformation aims to enhance the classification accuracy of the model.

Furthermore, five out of six intensity measures (IM) are deleted from the final dataset for each intensity measure. At the same time, the remaining undeleted IM is utilized as input for the model training process. Finally, MIDR values, the target values in the dataset, are categorized into damage levels based on the limits provided in Table 2.6. The damage levels given in the table are represented with integers ranging from 0 to 4.

Table 3.2. Processed data points with PGD as IM.

Data Label	**Input/Output**	**Data - 1**	**Data - 3799**	**Data - 71142**	**Data - 71280**
Number of Stories - Normalized	Input	0	0	1	1
Number of Spans - Normalized	Input	0	0	1	1
Span Length - Normalized	Input	0	0	1	1
Story Height - Normalized	Input	0	0.5	1	1
Column Area - Normalized	Input	0	0	1	1
Concrete Strength - Normalized	Input	0	0.2333	1	1
Commercial Use	Input	1	0	0	0
Soil Condition - Normalized	Input	1	1	1	0
PGD - Normalized	Input	0.01347	0.00765	0.01034	0.03854
Damage State	Output	4	1	0	3

3.3. Preliminary Data Analysis

In this section, the distributions of MIDR values relative to the input have been visualized in detail to examine the generated dataset. This preliminary analysis observes the impact of input variations and the relationships between the inputs and the outcome.

MIDR values exceeding 2% have been set equal to 0.02 to ease the interpretation of the figures. When utilizing machine learning algorithms, the highest damage level, complete, has a MIDR threshold of 0.01. In other words, analysis results with MIDR values exceeding 2% were already labeled as a complete damage state.

MIDR distributions for 3- and 5-story buildings are shown in Figure 3.1. While the MIDR distribution for both building types does not exhibit a significant change, for the moderate damage state, structures with 5-story buildings have a higher proportion compared to structures with 3-story buildings. However, the ratio of 3-story structures is slightly higher for other damage levels.

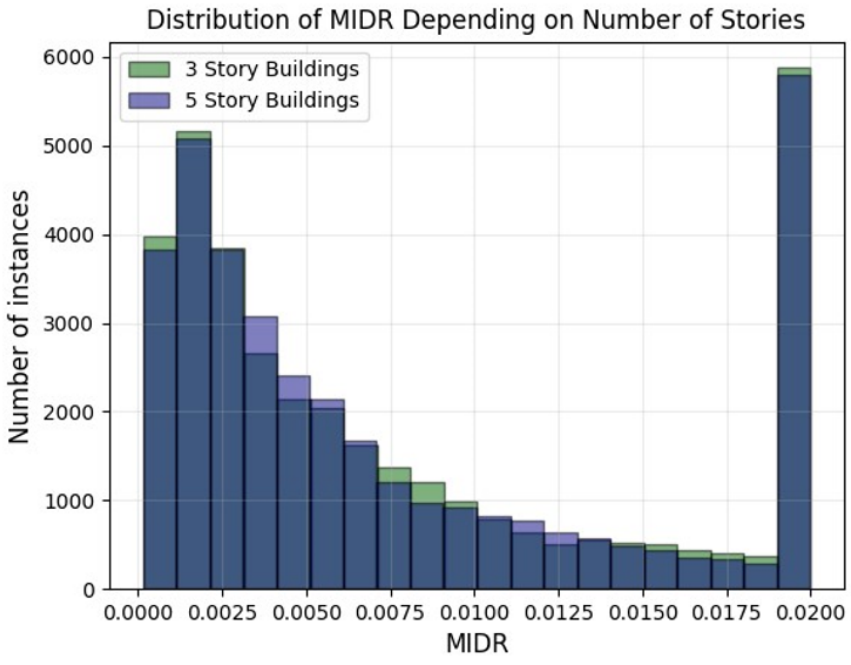


Figure 3.1. MIDR distributions for 3- and 5-story buildings.

Figure 3.2 illustrates MIDR distributions for 3- and 4-span buildings. Similar to the MIDR distribution observed for 3- and 5-story buildings, the number of spans does not considerably impact the MIDR distribution. There are only two distinctions in Figure 3.2: the number of instances is slightly higher in the "no damage" category for 4-span buildings, while 3-span buildings have approximately 400 more instances in the "complete damage" category.

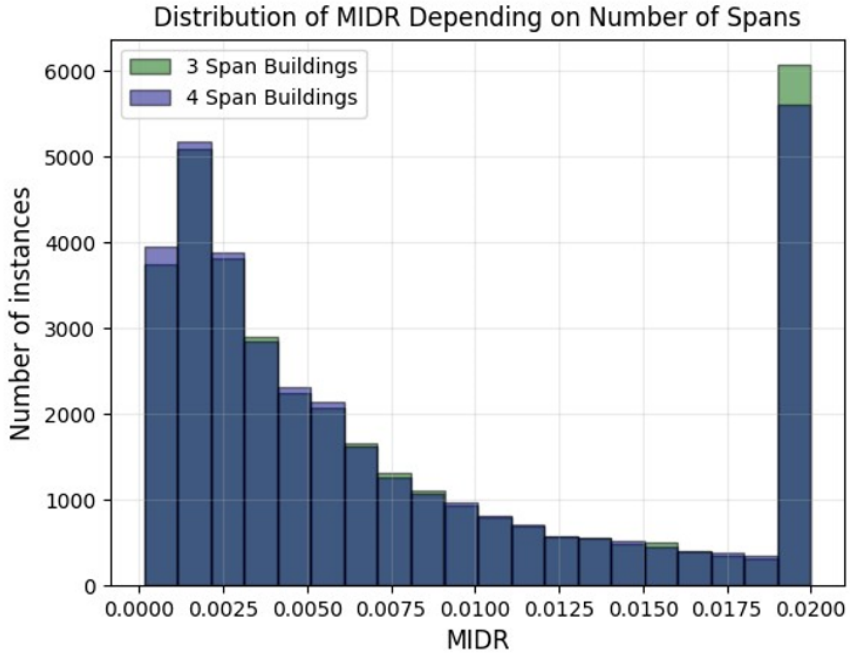


Figure 3.2. MIDR distribution for 3- and 4-span buildings.

In contrast to the previous graphs, the impact of span length on MIDR values is apparent in Figure 3.3. There is an observed correlation where an increase in the span length corresponds to an increase in structural damage.

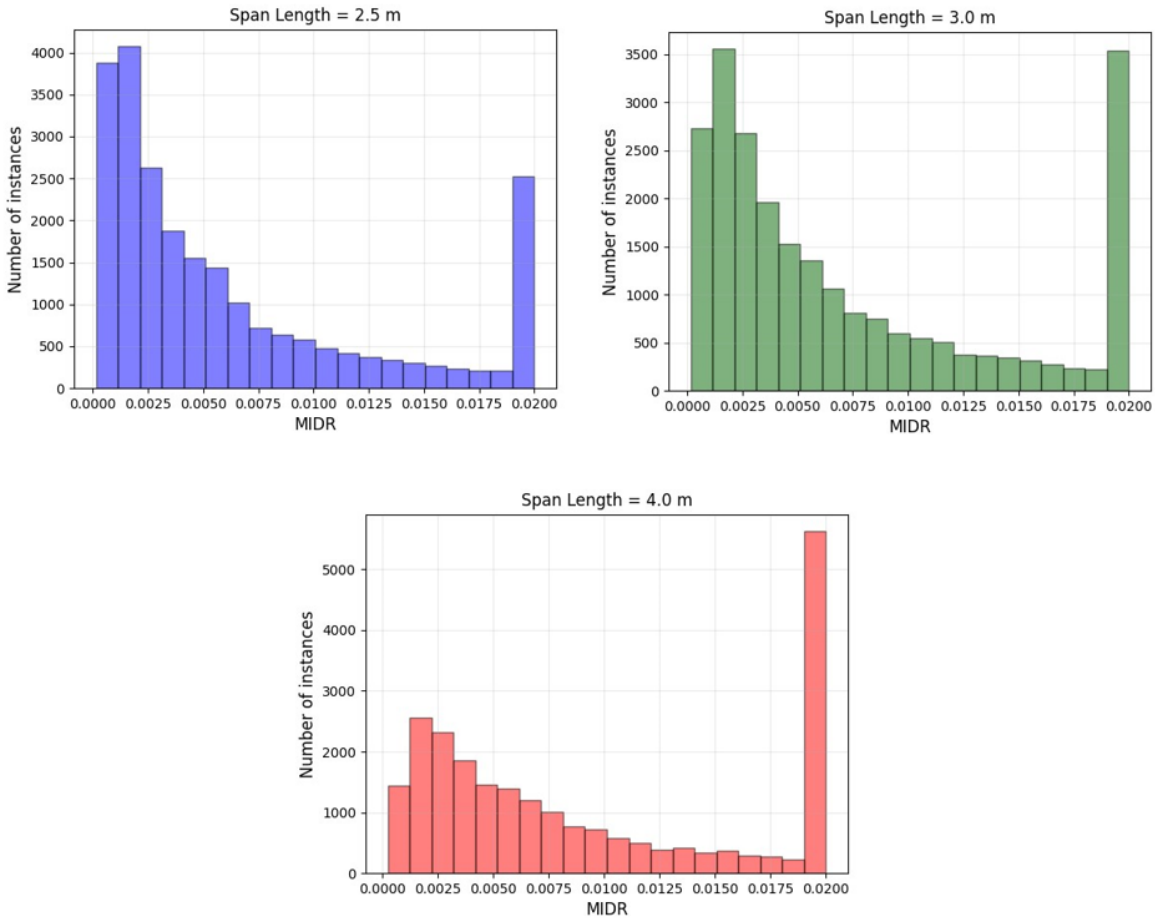


Figure 3.3. MIDR distributions conditioned on span length.

The impact of floor height on MIDR distribution has been examined in Figure 3.4 and Figure 3.5. In Figure 3.4, situations where the floor height is 2.6, 2.8, and 3 meters are shown in different graphs. When examining the provided MIDR distributions in these graphs, it is observed that there is a proportional relationship between floor height and damage levels. In Figure 3.5, the case where the ground floor is operated for commercial purposes is compared to the standard case. In this study, where the base floor is utilized for commercial use, the ground floor height is modeled as 3.5 meters, and in standard cases, the ground floor height is equal to the specified floor height without creating a soft story condition. It is clearly observed here that the ground floor height being different from the normal floor height, at 3.5 meters, is a parameter directly affecting the structural damage.

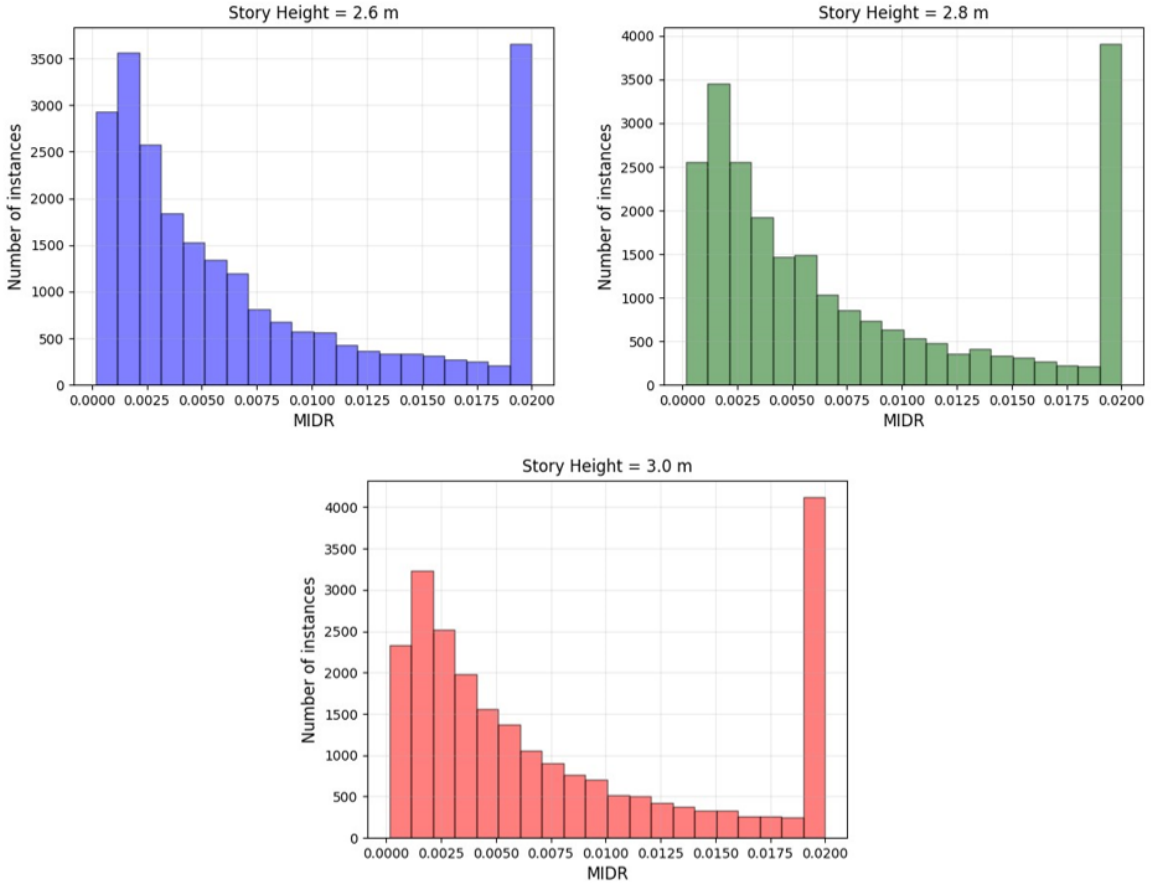


Figure 3.4. MIDR distributions conditioned on story height.

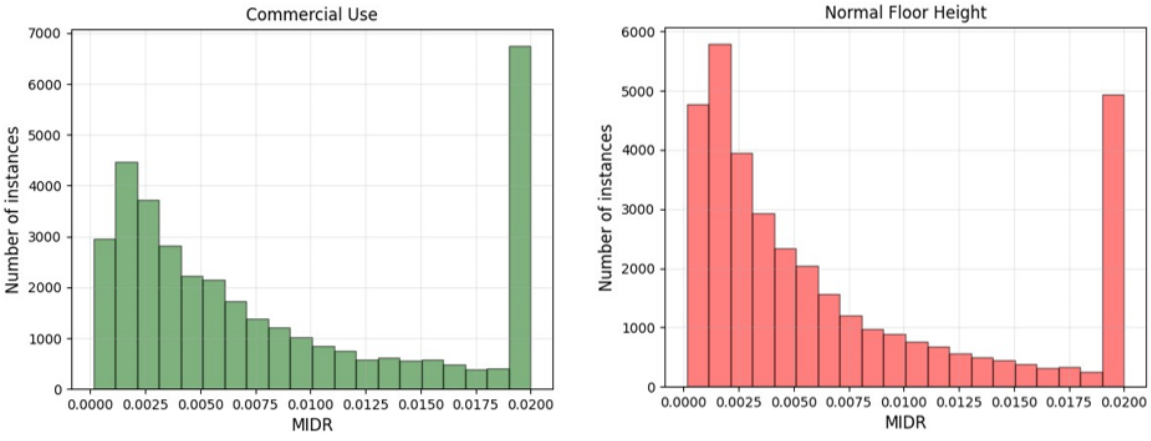


Figure 3.5. MIDR distributions conditioned on ground floor height.

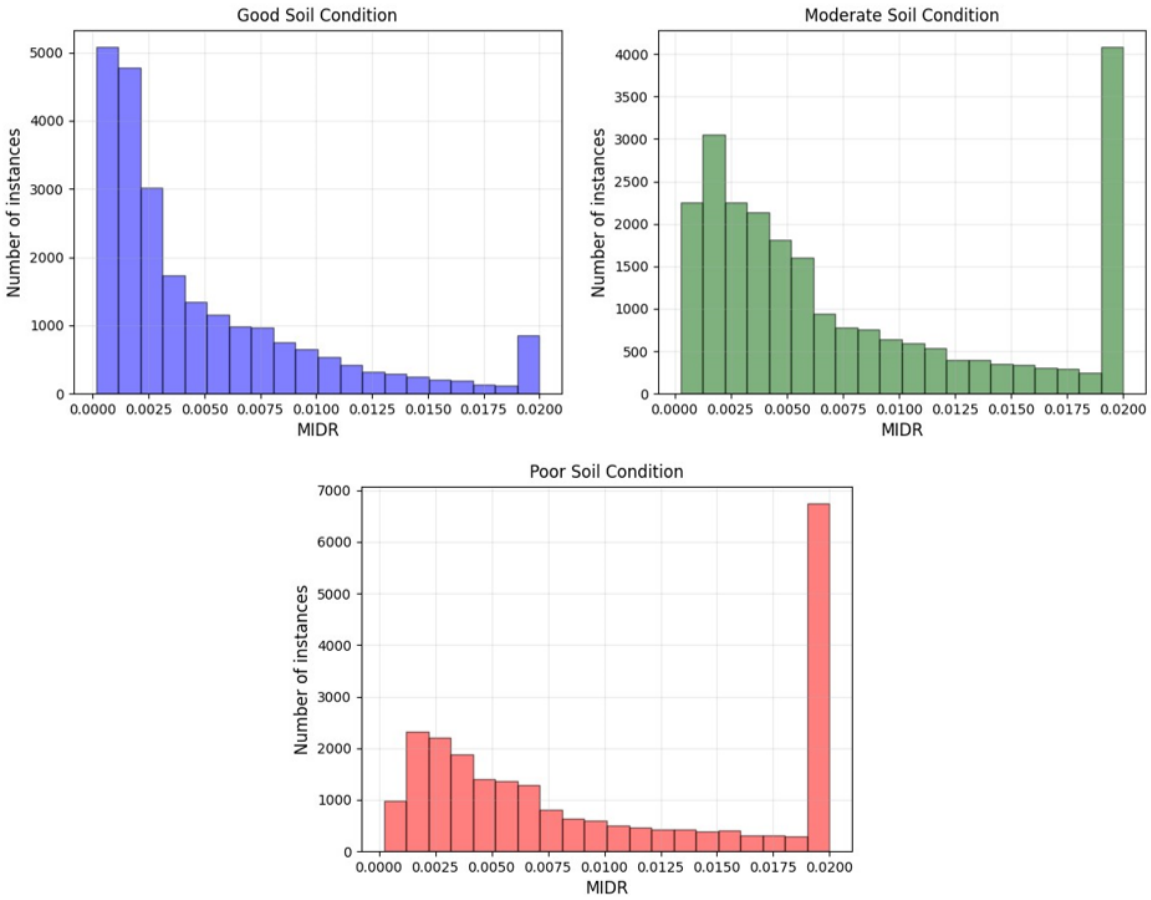


Figure 3.6. MIDR distributions conditioned on soil condition.

For examining the impact of another input, the soil condition, on the damage state of structures, MIDR distributions for each soil condition are provided in Figure 3.6. It is visible in the figure that as the soil condition improves, the damage rate significantly decreases.

Figure 3.7 and Figure 3.8 compare MIDR probability density functions (PDFs) for 3-story and 5-story buildings based on column dimensions. In Figure 3.9, the MIDR PDFs are drawn based on concrete strength. As expected, it is observed that as the column area and concrete strength increase, the damage rates decrease. Black dashed lines indicate the damage level boundaries in these three figures.

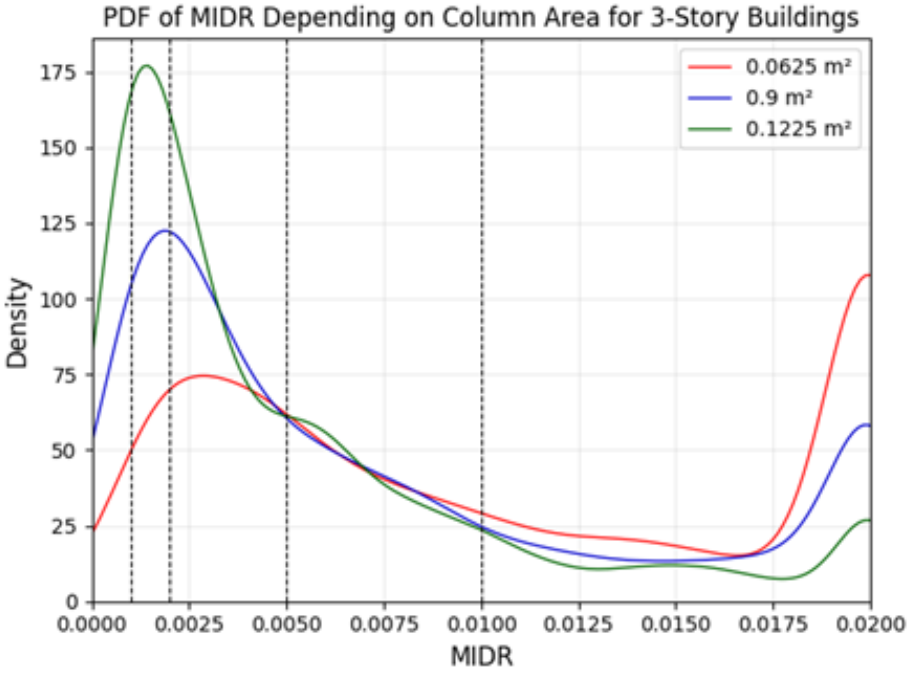


Figure 3.7. The probability density functions of MIDR depending on column area for 3-story buildings.

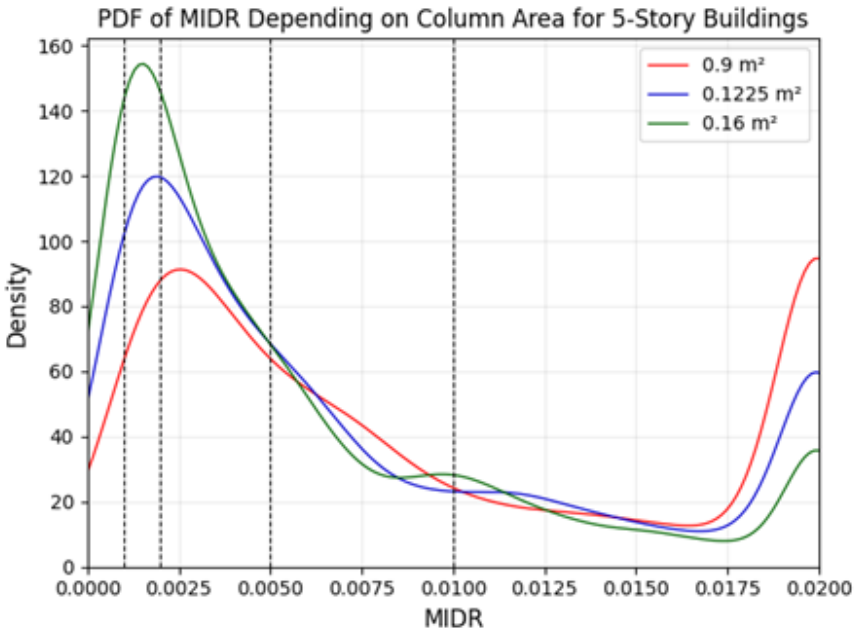


Figure 3.8. The probability density functions of MIDR depending on column area for 5-story buildings.

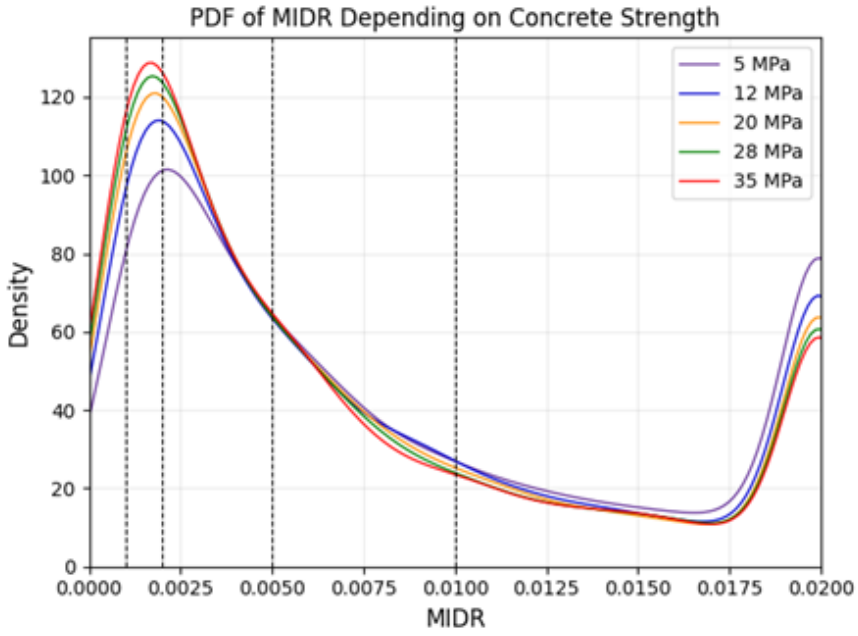


Figure 3.9. The probability density functions of MIDR depending on concrete strength.

Figure 3.10 shows the relationship between MIDR and IM values as a scatter plot. Additionally, the distribution of each IM is shown above the scatter plots. The IM type that provides the most distinct trend with MIDR is observed to be CAV. However, the figure shows a proportional relationship between MIDR and all IM types.

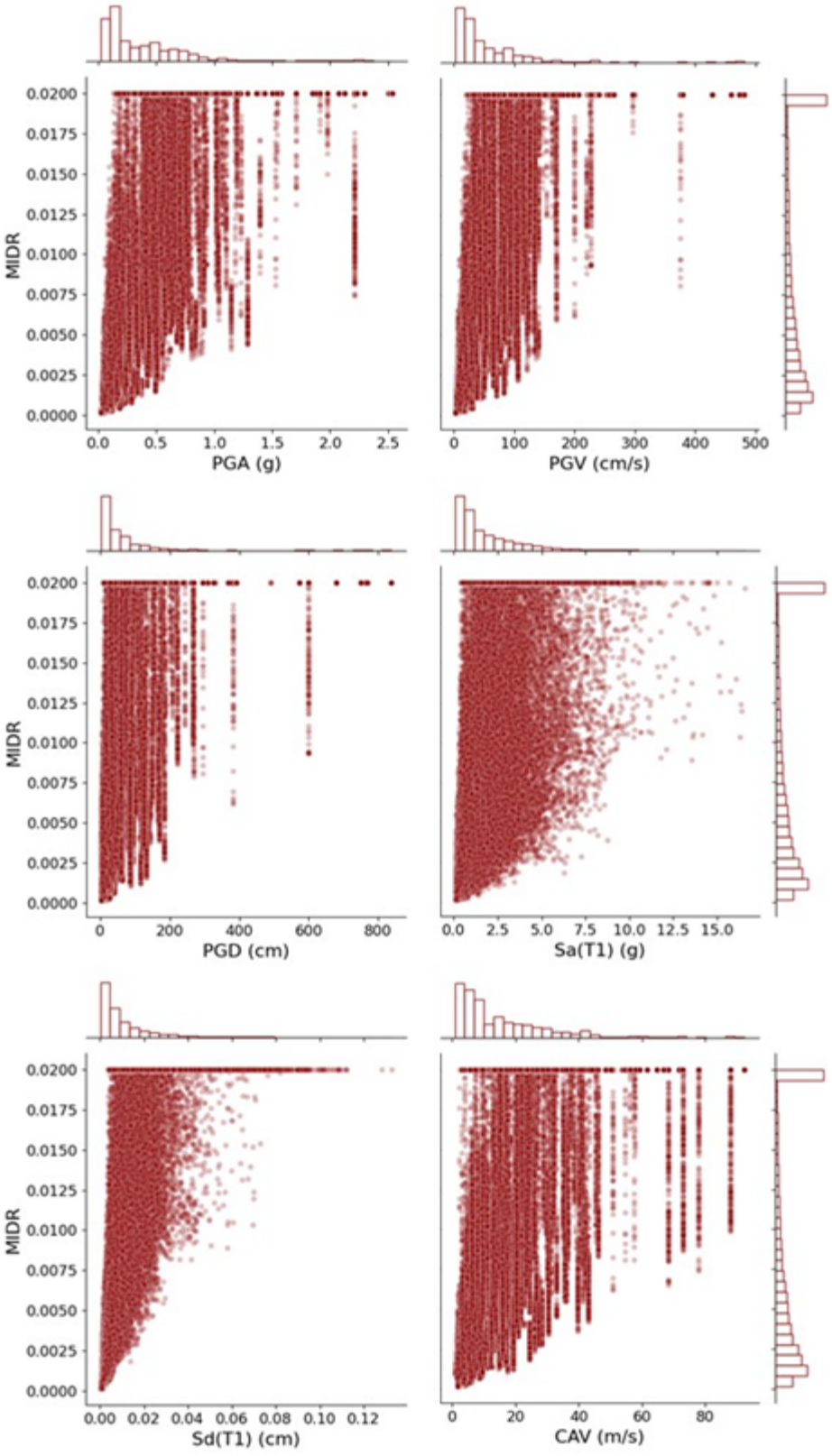


Figure 3.10. Scatter plot illustrating the relationship between MIDR and IMs, along with their respective distributions.

3.4. Machine Learning Algorithms

Machine learning consists of a wide array of training systems, and the selection of the algorithm type is contingent on the model's purpose and the dataset at hand (Alpaydin, 2014). In this research, the aim of the machine learning model is to forecast the state of building damage using input data comprising building features, soil conditions, and measures of ground motion intensity. Specifically, the study concentrates on classification-oriented machine learning algorithms, such as k-nearest neighbors, support vector machines, decision trees, and random forests.

3.4.1. k-Nearest Neighbor

The k-Nearest Neighbor (kNN) algorithm is a classification method that operates as a voting algorithm, utilizing a specified distance measure and a set of k sample points to classify a data point. It does not make any assumptions regarding the distribution of the underlying data and depends on the similarity of features. In kNN, the classification of a data point is established by a majority vote from its K nearest neighbors, assigning it to the class that is most prevalent among them (Kiani et al., 2019).

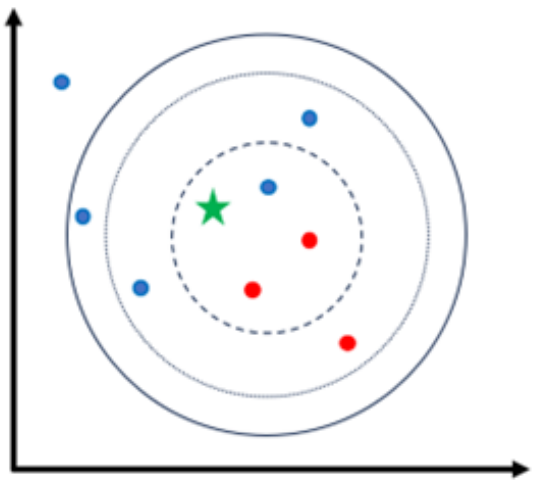


Figure 3.11. The schematic representation of kNN (Yang, 2019).

Figure 3.11 illustrates a classification problem featuring two distinct classes: red dots and blue dots. The goal is to determine the green star's class. When k is set to 3 (as indicated by the smaller dashed circle), the green star is classified as a red dot. On the other hand, with k set to 7 (as shown by the larger solid circle), the green star is classified as a blue dot. However, if k is set to 6 (as depicted by the dotted circle), it becomes challenging to classify the green star due to a tie. This reveals the significance of choosing the appropriate k parameter and its sensitivity. It is advisable for k to be an odd number. If k is too small, it may result in overfitting and increased sensitivity to data noise, whereas larger k values may introduce higher bias and reduced accuracy since they may include non-neighboring samples. Determining the ideal k value is not straightforward, often requiring iterative cross-validation (Yang, 2019).

In the context of this thesis, three hyperparameters are employed to enhance prediction performance. Firstly, ten odd numbers ranging from one to nineteen were experimented with for k values. Moreover, the weights hyperparameter is designated as "uniform" and "distance." "Uniform" weight assigns equal importance to all neighbors, and the "distance" weight makes closer neighbors more substantially influence the prediction. In addition, the distance hyperparameter is assigned as "Euclidean" and "Manhattan" distance metrics to measure the distance between data points. The "Euclidean" distance calculates the shortest distance between two points in Euclidean space, while the "Manhattan" or "Taxicab" distance is determined by adding up the absolute differences between their coordinates (Krause, 1987). The equation of both distance metrics is given in Equation 3.1 and Equation 3.2.

$$d_{Euclidean}\left(A, B\right) = \sqrt{\left(x_a - x_b\right)^2 + \left(y_a - y_b\right)^2} \tag{3.1}$$

$$d_{Manhattan}\left(A, B\right) = |x_a - x_b| + |y_a - y_b| \tag{3.2}$$

3.4.2. Support Vector Machine (SVM)

The support vector machine (SVM) distinguishes between classes using a boundary known as the margin. The objective is to maximize the distance between the margin and instances on both sides to identify the optimal boundary (optimal hyperplane in Figure 3.12) for class separation (Cortes and Vapnik, 1995).

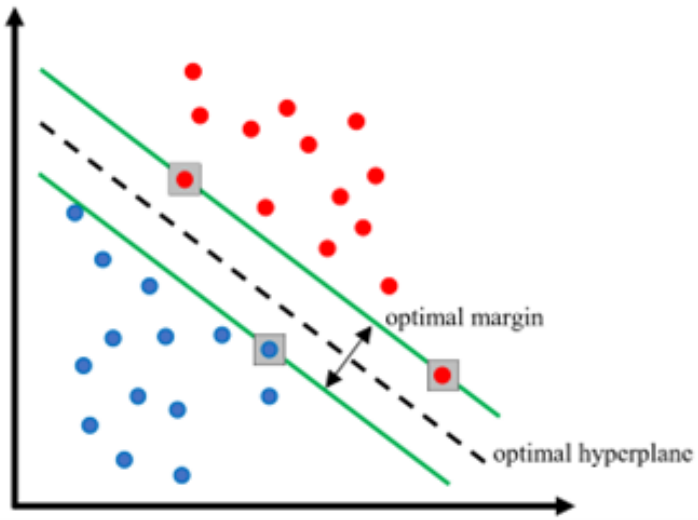


Figure 3.12. An illustration of a separable problem within a two-dimensional space. The support vectors, indicated by dark grey squares, delineate the margin, representing the maximum separation between the two classes (Cortes and Vapnik, 1995).

This method is suitable only for datasets that permit linear separation. When dealing with nonlinear problems, instead of attempting to accommodate a nonlinear model, nonlinear transformation should be performed using appropriately selected basis functions to map the problem to a new space. Subsequently, a linear model is applied in this transformed space. The linear model in the transformed space is equivalent to a nonlinear model in the original space. This method is suitable for addressing both classification and regression problems (Alpaydin, 2014).

This thesis employs the radial-basis function (RBF) kernel, one of the widely

adopted kernel functions. Also recognized as the Gaussian kernel, the RBF kernel gauges the similarity between two data points across infinite dimensions and subsequently facilitates classification through a majority vote. The radial-basis function with Euclidean distance (Alpaydin, 2014) is given below:

$$K(x^t, x) = \exp\left(-\frac{\|x^t - x\|^2}{2s^2}\right) \tag{3.3}$$

The kernel function approaches zero as the Euclidean distance between two points increases. Consequently, points that are far apart are more likely to be dissimilar.

In this basis function, two hyperparameters can be optimized using cross-validation, such as regularization parameter (C) and spread value (s^2). In Figure 3.13, it can be observed that larger spread values (s^2) smooth the boundaries. The regularization parameter C is finely adjusted through cross-validation. It determines the balance between maximizing the margin and minimizing errors. If C is excessively large, non-separable points incur a substantial penalty, potentially resulting in the storage of numerous support vectors and leading to overfitting. If it is too small, solutions that are too simple and prone to underfitting may be encountered. The best spread value and regularization parameter can be determined through cross-validation, and in this study, a broad range of hyperparameters is used to improve prediction performance.

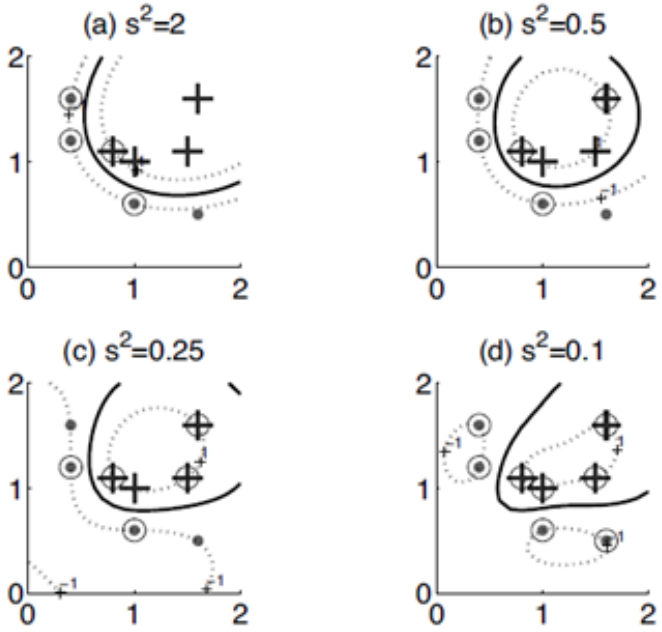


Figure 3.13. The boundary and margins identified by the Gaussian kernel with various spread values (Alpaydin, 2014).

3.4.3. Decision Tree (DT)

The decision tree is an effective nonparametric technique applicable to both classification and regression tasks. It consists of internal decision nodes and terminal leaves, as seen in Figure 3.14. When an input is provided, a test is conducted at each node, and the branch taken depends on the outcome. This process initiates at the root and iterates recursively until a leaf node is reached, at which juncture the output is determined by the value inscribed in the leaf. A decision tree is also considered a nonparametric model since no parametric form is assumed for class densities. The tree structure is not predetermined, but rather, it evolves during the learning process. Branches and leaves are added as needed, contingent on the complexity inherent in the data's underlying problem.

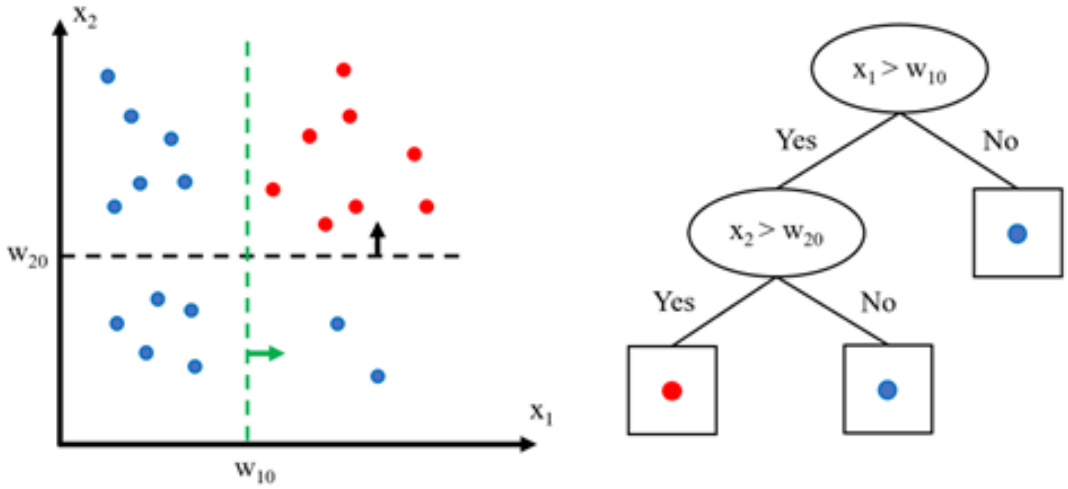


Figure 3.14. Illustration of a dataset along with the corresponding decision tree. Oval nodes represent decision nodes, and rectangles represent leaf nodes (Alpaydin, 2014).

One of the main reasons that decision trees are popular is the interpretability of the method. The decision rules are usually in the form of "if-then-else" conditions that are easily understandable. Numerous trees can encode a training set without error, and our goal is to find the smallest one in terms of the number of nodes and decision node complexity. Tree learning algorithms look for the best split at each step, starting at the root with the complete training data. This process continues recursively, splitting the data until no further splits are needed, creating labeled leaf nodes (Alpaydin, 2014).

This study uses two hyperparameters to find the best split, such as minimum sample split and maximum depth. The former indicates the minimum number of data points required to split an internal node, and the latter indicates the maximum depth of the tree. To reduce the computational burden, the complexity and size of the trees should be limited by setting the hyperparameters.

3.4.4. Random Forest (RF)

The random forest is a widely used ensemble learning algorithm aggregating multiple decision tree models. While a dataset is typically divided into training and test sets in machine learning applications, the random forest algorithm takes the training set and divides it into numerous sub-training sets. Decision trees are then constructed for each of these subsets. The final prediction model is received by averaging the outcomes of all the individual decision trees. This approach helps avoid overfitting and significantly enhances the model's overall accuracy (Ho, 1998).

Various criteria can be defined for determining the optimal split at each note in a decision tree. In this study, the Gini index proposed by Breiman et al. (1984) is utilized, which is a widely used method for evaluating the impurity or homogeneity of a set of data points. Among impurity measures such as entropy or log loss, the Gini impurity is preferred because it tends to be computationally more efficient. The calculation of the Gini impurity for a node in a decision tree is performed as follows:

$$i\,(t) = 1 - \sum_{j}^{k} p^2\,(\ j \mid t\ ) \tag{3.4}$$

Where i is the Gini index, t is the node, k is the number of classes, and the probability of randomly picking an element of class j in the node t is $p(j \mid t)$.

In constructing random forests for classification, the objective is to minimize Gini impurity at each node. Lower Gini impurity values indicate better splits in separating the classes in the dataset. Gini impurity ranges from 0 to 0.5, respectively, indicating a perfectly pure node and a completely impure node with equally distributed classes. This process is repeated recursively until the tree is fully grown, and the random forest is an ensemble of such trees.

The model performance could be significantly increased by altering the hyperparameters. While using random forest classification algorithms to develop a prediction

model, four hyperparameters are tuned. The first hyperparameter, max features, determines the number of features considered when looking for the best split. Smaller values foster more diverse trees, whereas larger values may yield more correlated trees. The second hyperparameter, minimum sample split, is borrowed from decision trees and denotes the minimum number of data points needed to split an internal node. Lower values may lead to more splits, potentially causing overfitting, while higher values can result in a simpler, less complex model. The third hyperparameter, minimum samples at a leaf node, adjusts the minimum number of samples required to form a leaf node. This impacts the granularity of the leaves in decision trees, with smaller values potentially leading to more detailed but noisy leaves and larger values producing a smoother, more generalized model. The final hyperparameter involves the number of trees in the forest. Increasing this number enhances model robustness but also raises computational costs. Fine-tuning these hyperparameters allows for a balance between model complexity, diversity, and computational efficiency.

3.5. Application of Machine Learning Algorithms

Within the scope of this thesis, twenty-four machine learning models were created using four different classification algorithms (k-nearest neighbors, support vector machine, decision tree, and random forest) and six different intensity measures (PGA, $S_a(T_1)$, PGV, PGD, $S_d(T_1)$, CAV). These models' performance and confusion metrics were examined to select the best-performing model.

Scikit-learn (Pedregosa et al., 2011) is employed for implementing machine learning algorithms. Scikit-learn is a Python module that leverages a robust environment to offer cutting-edge implementations of numerous well-known ML algorithms. It maintains an easy-to-use interface tightly integrated with the Python language. Scikit-learn facilitates evaluating an estimator's performance or selecting parameters through cross-validation with the GridSearchCV object, where "CV" denotes "cross-validated" (Pedregosa et al., 2011).

To assess the efficiency of machine learning methods in forecasting damage states, the complete dataset is split into two subsets: a training set and a testing set. The training set is employed to develop the prediction model, and the testing set is employed to assess the model's performance. In this research, 80% of the data is assigned to the training set, and the remaining data is designated for the testing set.

Moreover, the GridSearchCV object is employed to fine-tune the most appropriate hyperparameters, enhancing model performance through the utilization of k-fold cross-validation techniques. The hyperparameters are adjusted within the specified range outlined in the preceding section for each algorithm employed in this thesis. The training set is divided into ten cross-validation sets, employing the Stratified K-Folds method to evaluate the model's performance across diverse subsets of the data.

3.5.1. Evaluation of the Model Performance

Evaluating the performance of a classification-related machine learning model is significant to assess its effectiveness in making predictions. The confusion matrix is one of the most common techniques to interpret and assess the effectiveness of the models. The confusion matrix can calculate several performance metrics: precision, recall, and accuracy. In ML applications, a combination of performance metrics could provide a more exhaustive assessment of the model's performance, depending on the specific goals and characteristics of the problem. This thesis interprets the confusion matrix, these metrics, and the F1-score as part of model performance evaluation.

Confusion Matrix

The confusion matrix provides a tabular representation of actual class outcomes versus model predictions, illustrated in Table 3.3. Diagonal elements in the matrix signify correct predictions, while off-diagonal elements represent false predictions.

Table 3.3. Confusion Matrix and Calculation of Performance Metrics.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Predicted Class</td><td rowspan="2">Total</td><td rowspan="2">Recall</td></tr>
<tr><td>Positive</td><td>Negative</td></tr>
<tr><td rowspan="2">Actual Class</td><td>Positive</td><td>True Positive</td><td>False Negative</td><td>Total of Actual Positive</td><td>True Positive / Total of Actual Positive</td></tr>
<tr><td>Negative</td><td>False Positive</td><td>True Negative</td><td>Total of Actual Negative</td><td>True Negative / Total of Actual Negative</td></tr>
<tr><td colspan="2">Total</td><td>Total of Predicted Positive</td><td>Total of Predicted Negative</td><td colspan="2" rowspan="2">Accuracy = Sum of True Predictions / Total number of predictions</td></tr>
<tr><td colspan="2">Precision</td><td>True Positive / Total of Predicted Positive</td><td>True Negative / Total of Predicted Negative</td></tr>
</table>

Accuracy

Accuracy offers a straightforward measure of overall correctness and is easily understandable. It is well-suited for datasets with uniformly distributed classes. However, accuracy may cause misinterpretation of imbalanced datasets. A high accuracy score may not precisely indicate the model's effectiveness, particularly if it demonstrates strong performance in predicting the majority class but performs inadequately on the minority class.

Precision and Recall

Precision and recall serve as valuable metrics for evaluating the performance of models developed with imbalanced datasets, offering a comprehensive assessment. Precision reveals the ratio of correctly predicted instances to the total instances predicted as a particular class. On the other hand, recall highlights the ratio of correctly predicted samples to the total instances of the actual class for each target feature. By considering both precision and recall, the evaluation becomes more comprehensive, particularly in scenarios where imbalances in the dataset may impact traditional metrics like accuracy.

F1-score

The F1-score performance metric is commonly used in ML to evaluate the classification model performance. It is advantageous when dealing with imbalanced datasets or an uneven class distribution. The F1-score presents the harmonic mean of precision and recall, with the optimal value being one and the worst score zero. Precision and recall contribute equally to the F1-score, and its formula is expressed as follows:

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3.5}$$

3.6. The Results of Seismic Damage Prediction Models and Comparison of Model Performances

The data set produced in Section 2 was developed and tested using classification-based machine learning models described in Section 3. While 80% of the artificially generated data set was used to train the machine learning models, the remaining 20% was used for testing. The test data set consisted of 14,256 data points, whereas 57,024 data points were used for training the model.

Machine learning models created using different classification algorithms and intensity measures were evaluated using performance metrics such as precision, recall, accuracy, and F1-score, as well as confusion matrices. Performance metric comparisons are provided in Table 3.4. However, solely relying on this table may make it challenging to decide which intensity measure can be used to develop a better model, even if it can be ensured that the machine learning model using the RF algorithm provides the highest performance. This is because models developed with PGA, PGV, PGD, and CAV intensity measures show performance greater than 92% for each performance metric.

For a more detailed model comparison, confusion matrices were used. The con-

fusion matrices are presented in tables between 3.5 and 3.28. For determination of the best prediction model, the test results of machine learning models developed with different data sets, using the RF classification algorithm and intensity measures such as PGA, PGV, PGD, and CAV, were compared using the confusion matrices provided in Tables 30, 31, 32, and 35, respectively. When these tables are examined, each model demonstrates a high ability to make predictions. However, a detailed examination of the confusion matrices reveals a distinctive pattern for the PGV-based model did not predict any instances in the "no damage" or "slight damage state" as having a significantly different damage level, such as "extensive damage state" or "complete damage state," unlike models developed with the other three intensity measures using the RF algorithm. This characteristic enhances the reliability of the model developed using the RF algorithm with PGV.

As a result, a comparison was conducted among the 24 generated machine learning models. The model developed with a dataset incorporating PGV and utilizing the RF classification algorithm was the most proficient in predicting seismic damage.

Table 3.4. Evaluation metrics of machine learning algorithms

Classification Algorithm	**IM**	**Precision**	**Recall**	**Accuracy**	**F1-Score**
k-Nearest Neighbor (kNN)	PGA	0.59	0.58	0.61	0.58
	PGV	0.46	0.45	0.49	0.45
	PGD	0.47	0.46	0.51	0.46
	$S_a(T_1)$	0.57	0.57	0.59	0.57
	$S_d(T_1)$	0.57	0.57	0.59	0.57
	CAV	0.47	0.46	0.51	0.46
Support Vector Machine (SVM)	PGA	0.70	0.69	0.72	0.70
	PGV	0.59	0.57	0.61	0.57
	PGD	0.53	0.51	0.54	0.51
	$S_a(T_1)$	0.69	0.68	0.69	0.68
	$S_d(T_1)$	0.69	0.68	0.69	0.68
	CAV	0.62	0.59	0.64	0.60
Decision Tree (DT)	PGA	0.89	0.90	0.90	0.89
	PGV	0.90	0.90	0.90	0.90
	PGD	0.90	0.90	0.90	0.90
	$S_a(T_1)$	0.67	0.66	0.68	0.66
	$S_d(T_1)$	0.67	0.66	0.68	0.66
	CAV	0.90	0.90	0.90	0.90
Random Forest (RF)	PGA	0.92	0.92	0.93	0.92
	PGV	0.92	0.92	0.92	0.92
	PGD	0.92	0.92	0.93	0.92
	$S_a(T_1)$	0.71	0.68	0.70	0.69
	$S_d(T_1)$	0.70	0.68	0.70	0.68
	CAV	0.92	0.92	0.92	0.92

Table 3.5. Confusion matrix of kNN algorithm, the dataset with PGA

kNN - PGA		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	696	434	78	0	0	1208	0.58
	Slight D.	383	992	728	50	7	2160	0.46
	Mod. D.	38	619	2132	759	156	3704	0.58
	Ext. D.	0	55	850	1377	654	2936	0.47
	Comp. D.	0	8	145	666	3429	4248	0.81
Total		1117	2108	3933	2852	4246	0.61	
Precision		0.62	0.47	0.54	0.48	0.81		

Table 3.6. Confusion matrix of kNN algorithm, the dataset with PGV

kNN - PGV		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	502	487	219	0	0	1208	0.42
	Slight D.	474	671	907	107	1	2160	0.31
	Mod. D.	188	712	1670	940	194	3704	0.45
	Ext. D.	3	106	941	997	889	2936	0.34
	Comp. D.	0	6	243	852	3147	4248	0.74
Total		1167	1982	3980	2896	4231	0.49	
Precision		0.43	0.34	0.42	0.34	0.74		

Table 3.7. Confusion matrix of kNN algorithm, the dataset with PGD

kNN - PGD		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	500	383	291	31	3	1208	0.41
	Slight D.	349	635	902	199	75	2160	0.29
	Mod. D.	179	741	1468	816	500	3704	0.40
	Ext. D.	49	218	780	863	1026	2936	0.29
	Comp. D.	2	82	550	986	2628	4248	0.62
Total		1079	2059	3991	2895	4232	0.43	
Precision		0.46	0.31	0.37	0.30	0.62		

Table 3.8. Confusion matrix of kNN algorithm, the dataset with $S_a(T_1)$

kNN - $S_a(T_1)$		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	705	385	116	2	0	1208	0.58
	Slight D.	341	982	780	55	2	2160	0.45
	Mod. D.	81	691	2070	684	178	3704	0.56
	Ext. D.	1	71	852	1336	676	2936	0.46
	Comp. D.	0	4	233	703	3308	4248	0.78
Total		1128	2133	4051	2780	4164	0.59	
Precision		0.63	0.46	0.51	0.48	0.79		

Table 3.9. Confusion matrix of kNN algorithm, the dataset with $S_d(T_1)$

kNN - $S_d(T_1)$		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	705	385	116	2	0	1208	0.58
	Slight D.	341	982	780	55	2	2160	0.45
	Mod. D.	81	691	2070	684	178	3704	0.56
	Ext. D.	1	71	854	1332	678	2936	0.45
	Comp. D.	0	4	234	705	3305	4248	0.78
Total		1128	2133	4054	2778	4163	0.59	
Precision		0.63	0.46	0.51	0.48	0.79		

Table 3.10. Confusion matrix of kNN algorithm, the dataset with CAV

kNN - CAV		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	501	391	309	5	2	1208	0.41
	Slight D.	379	641	1010	103	27	2160	0.30
	Mod. D.	243	553	2122	485	301	3704	0.57
	Ext. D.	128	286	1013	825	684	2936	0.28
	Comp. D.	16	71	660	352	3149	4248	0.74
Total		1267	1942	5114	1770	4163	0.51	
Precision		0.40	0.33	0.41	0.47	0.76		

Table 3.11. Confusion matrix of SVM algorithm, the dataset with PGA

SVM - PGA		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	877	304	27	0	0	1208	0.73
	Slight D.	287	1271	587	15	0	2160	0.59
	Mod. D.	14	404	2593	628	65	3704	0.70
	Ext. D.	0	0	702	1638	596	2936	0.56
	Comp. D.	0	0	54	367	3827	4248	0.90
Total		1178	1979	3963	2648	4488	0.72	
Precision		0.74	0.64	0.65	0.62	0.85		

Table 3.12. Confusion matrix of SVM algorithm, the dataset with PGV

SVM - PGV		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	677	406	125	0	0	1208	0.56
	Slight D.	386	810	930	34	0	2160	0.38
	Mod. D.	75	359	2455	695	120	3704	0.66
	Ext. D.	0	17	883	1213	823	2936	0.41
	Comp. D.	0	0	172	509	3567	4248	0.84
Total		1138	1592	4565	2451	4510	0.61	
Precision		0.59	0.51	0.54	0.49	0.79		

Table 3.13. Confusion matrix of SVM algorithm, the dataset with PGD

SVM - PGD		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	668	320	219	1	0	1208	0.55
	Slight D.	350	695	1045	67	3	2160	0.32
	Mod. D.	120	409	2414	460	301	3704	0.65
	Ext. D.	23	115	1030	792	976	2936	0.27
	Comp. D.	5	20	585	444	3194	4248	0.75
Total		1166	1559	5293	1764	4474	0.54	
Precision		0.57	0.45	0.46	0.45	0.71		

Table 3.14. Confusion matrix of SVM algorithm, the dataset with $S_a(T_1)$

SVM - $S_a(T_1)$		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	848	314	46	0	0	1208	0.70
	Slight D.	249	1335	573	3	0	2160	0.62
	Mod. D.	19	534	2612	474	65	3704	0.71
	Ext. D.	0	17	865	1523	531	2936	0.52
	Comp. D.	0	0	131	559	3558	4248	0.84
Total		1116	2200	4227	2559	4154	0.69	
Precision		0.76	0.61	0.62	0.60	0.86		

Table 3.15. Confusion matrix of SVM algorithm, the dataset with $S_d(T_1)$

SVM - $S_d(T_1)$		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	858	305	45	0	0	1208	0.71
	Slight D.	267	1333	558	2	0	2160	0.62
	Mod. D.	23	551	2633	432	65	3704	0.71
	Ext. D.	0	16	891	1480	549	2936	0.50
	Comp. D.	0	0	132	544	3572	4248	0.84
Total		1148	2205	4259	2458	4186	0.69	
Precision		0.75	0.60	0.62	0.60	0.85		

Table 3.16. Confusion matrix of SVM algorithm, the dataset with CAV

SVM - CAV		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	659	411	138	0	0	1208	0.55
	Slight D.	333	941	845	41	0	2160	0.44
	Mod. D.	27	391	2538	573	175	3704	0.69
	Ext. D.	0	14	967	1316	639	2936	0.45
	Comp. D.	0	0	90	518	3640	4248	0.86
Total		1019	1757	4578	2448	4454	0.64	
Precision		0.65	0.54	0.55	0.54	0.82		

Table 3.17. Confusion matrix of DT algorithm, the dataset with PGA

DT - PGA		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	1107	95	6	0	0	1208	0.92
	Slight D.	99	1899	160	2	0	2160	0.88
	Mod. D.	4	223	3293	175	9	3704	0.89
	Ext. D.	0	7	232	2504	193	2936	0.85
	Comp. D.	0	5	23	231	3989	4248	0.94
Total		1210	2229	3714	2912	4191	0.90	
Precision		0.91	0.85	0.89	0.86	0.95		

Table 3.18. Confusion matrix of DT algorithm, the dataset with PGV

DT - PGV		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	1107	97	4	0	0	1208	0.92
	Slight D.	92	1890	176	2	0	2160	0.88
	Mod. D.	1	217	3304	173	9	3704	0.89
	Ext. D.	1	4	217	2509	205	2936	0.85
	Comp. D.	0	0	9	221	4018	4248	0.95
Total		1201	2208	3710	2905	4232	0.90	
Precision		0.92	0.86	0.89	0.86	0.95		

Table 3.19. Confusion matrix of DT algorithm, the dataset with PGD

DT - PGD		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	1124	82	2	0	0	1208	0.93
	Slight D.	90	1915	150	5	0	2160	0.89
	Mod. D.	1	205	3298	184	16	3704	0.89
	Ext. D.	0	8	185	2537	206	2936	0.86
	Comp. D.	0	7	8	255	3978	4248	0.94
Total		1215	2217	3643	2981	4200	0.90	
Precision		0.93	0.86	0.91	0.85	0.95		

Table 3.20. Confusion matrix of DT algorithm, the dataset with $S_a(T_1)$

DT - $S_a(T_1)$		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	790	332	84	2	0	1208	0.65
	Slight D.	213	1339	576	32	0	2160	0.62
	Mod. D.	46	520	2448	578	112	3704	0.66
	Ext. D.	0	38	752	1547	599	2936	0.53
	Comp. D.	0	2	134	573	3539	4248	0.83
Total		1049	2231	3994	2732	4250	0.68	
Precision		0.75	0.60	0.61	0.57	0.83		

Table 3.21. Confusion matrix of DT algorithm, the dataset with $S_d(T_1)$

DT - $S_d(T_1)$		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	783	364	61	0	0	1208	0.65
	Slight D.	222	1266	668	4	0	2160	0.59
	Mod. D.	17	467	2557	585	78	3704	0.69
	Ext. D.	0	21	786	1484	645	2936	0.51
	Comp. D.	0	1	87	567	3593	4248	0.85
Total		1022	2119	4159	2640	4316	0.68	
Precision		0.77	0.60	0.61	0.56	0.83		

Table 3.22. Confusion matrix of DT algorithm, the dataset with CAV

DT - CAV		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	1125	83	0	0	0	1208	0.93
	Slight D.	95	1914	150	1	0	2160	0.89
	Mod. D.	4	212	3318	159	11	3704	0.90
	Ext. D.	2	10	205	2524	195	2936	0.86
	Comp. D.	0	2	9	223	4014	4248	0.94
Total		1226	2221	3682	2907	4220	0.90	
Precision		0.92	0.86	0.90	0.87	0.95		

Table 3.23. Confusion matrix of RF algorithm, the dataset with PGA

RF - PGA		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	1137	70	1	0	0	1208	0.94
	Slight D.	67	1973	119	1	0	2160	0.91
	Mod. D.	0	140	3436	121	7	3704	0.93
	Ext. D.	0	3	158	2597	178	2936	0.88
	Comp. D.	0	2	5	188	4053	4248	0.95
Total		1204	2188	3719	2907	4238	0.93	
Precision		0.94	0.90	0.92	0.89	0.96		

Table 3.24. Confusion matrix of RF algorithm, the dataset with PGV

RF - PGV		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	1137	70	1	0	0	1208	0.94
	Slight D.	74	1961	125	0	0	2160	0.91
	Mod. D.	0	145	3440	115	4	3704	0.93
	Ext. D.	0	0	146	2568	222	2936	0.87
	Comp. D.	0	0	3	167	4078	4248	0.96
Total		1211	2176	3715	2850	4304	0.92	
Precision		0.94	0.90	0.93	0.90	0.95		

Table 3.25. Confusion matrix of RF algorithm, the dataset with PGD

RF - PGD		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	1134	74	0	0	0	1208	0.94
	Slight D.	74	1952	130	4	0	2160	0.90
	Mod. D.	3	144	3434	121	2	3704	0.93
	Ext. D.	0	2	134	2609	191	2936	0.89
	Comp. D.	0	0	6	184	4058	4248	0.96
Total		1211	2172	3704	2918	4251	0.93	
Precision		0.94	0.90	0.93	0.89	0.95		

Table 3.26. Confusion matrix of RF algorithm, the dataset with $S_a(T_1)$

RF - $S_a(T_1)$		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	802	335	71	0	0	1208	0.66
	Slight D.	168	1307	671	14	0	2160	0.61
	Mod. D.	24	379	2708	484	109	3704	0.73
	Ext. D.	0	20	717	1569	630	2936	0.53
	Comp. D.	0	0	126	461	3661	4248	0.86
Total		994	2041	4293	2528	4400	0.70	
Precision		0.81	0.64	0.63	0.62	0.83		

Table 3.27. Confusion matrix of RF algorithm, the dataset with $S_d(T_1)$

RF - $S_d(T_1)$		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	812	330	66	0	0	1208	0.67
	Slight D.	182	1329	643	6	0	2160	0.62
	Mod. D.	20	454	2637	512	81	3704	0.71
	Ext. D.	0	14	768	1544	610	2936	0.53
	Comp. D.	0	0	120	503	3625	4248	0.85
Total		1014	2127	4234	2565	4316	0.70	
Precision		0.80	0.62	0.62	0.60	0.84		

Table 3.28. Confusion matrix of RF algorithm, the dataset with CAV

RF - CAV		Predicted Class					Total	Recall
		No D.	Slight D.	Mod. D.	Ext. D.	Comp. D.		
Actual Class	No D.	1128	80	0	0	0	1208	0.93
	Slight D.	69	1955	136	0	0	2160	0.91
	Mod. D.	3	152	3429	114	6	3704	0.93
	Ext. D.	1	0	140	2589	206	2936	0.88
	Comp. D.	0	0	4	180	4064	4248	0.96
Total		1201	2187	3709	2883	4276	0.92	
Precision		0.94	0.89	0.92	0.90	0.95		

4. CONCLUSIONS

4.1. Summary of the Thesis

The data-driven methods, including machine learning, deep learning, and big data analytics, have surged in recent years due to enhancements in computer technology. This thesis aims to get the benefits of evolving technology in earthquake engineering by exploring the viability of employing machine-learning classification algorithms. The objective is to develop a seismic damage prediction model capable of estimating the building damage state without conducting traditional structural analyses.

The scope of this thesis has been delimited to facilitate the completion of the study. Consequently, the focus of this research has been directed towards the building stock of the Marmara Region, which stands out due to its high seismicity. Additionally, due to the impracticality of representing the entire diversity of structures in this region, the variables were limited, and a dataset was created by considering 1080 different structure fractions and three different soil types. NLTH analyses were conducted for two-dimensional models. The analysis results formed an imbalanced dataset primarily exhibiting "extensive damage" and "complete damage" levels in the majority of structures. The scaling factors of the selected records were reduced by 78% to overcome the imbalanced dataset issue, and the analyses were repeated with lowered scale factors. A data warehouse containing 71,280 entries was utilized to enhance machine learning algorithms.

Six different intensity parameters, such as peak ground acceleration (PGA), peak ground velocity (PGV), peak ground displacement (PGD), spectral acceleration ($S_a(T_1)$), spectral displacement ($S_d(T_1)$), and cumulative absolute velocity (CAV), were employed to express the severity of the selected earthquake records. For each intensity measure, 24 separate earthquake damage prediction models were developed using k-nearest neighbors (kNN), support vector machines (SVM), decision trees (DT), and

random forests (RF) classification algorithms. The one developed using the RF classification algorithm trained with the PGV dataset demonstrated the best performance among these models.

In summary, the model developed in this study predicts the damage state for non-ductile reinforced concrete structures based on a given ground motion intensity level, soil conditions, and structural characteristics. While all mentioned intensity measures can be used, it is recommended to use PGA, PGV, PGD, and CAV since the model's performance is lower for $S_a(T_1)$ and $S_d(T_1)$. Structural features that need to be provided as input include the number of stories, story height, number of spans, span length, column area, and concrete strength. Beam dimensions, steel yield strength, and reinforcement ratio in structural elements are considered single values and are not included in the inputs. The V_{S30} value should be entered as the soil condition. The model can estimate structural damage with greater than 92% accuracy with the requested inputs and suggested intensity measures.

4.2. Observations and Findings

In this thesis, the MIDR distributions were examined based on selected parameters to investigate their impact on MIDR, which was used as the EDP in this thesis. The observations here suggest that parameters such as span length, story height, the presence of a soft story (commercial use of the ground floor), column dimensions, and concrete strength values significantly alter the MIDR distribution, indicating a direct impact.

The findings of the thesis revealed that predictions utilizing the Random Forest (RF) algorithm for PGA, PGV, PGD, and CAV as IM achieved an overall accuracy of around 92% in categorizing damage states. Moreover, the RF algorithm for each IM accurately assigned more than 87% to each actual damage state in the test set.

The DT algorithm follows RF in this study and provides the second-best results.

Similar to RF, models developed with IM values other than $S_a(T_1)$ and $S_d(T_1)$ achieved an overall accuracy rate of approximately 89% in the tests. Additionally, for each IM, the DT algorithm exhibited a recall score exceeding 85% for every actual damage state in the test set.

4.3. Future Studies

A more comprehensive and extensive study is necessary to enhance the capability of predicting earthquake-induced building damage through machine learning algorithms. Further investigations may include conducting analyses in three dimensions, covering all types of structures, expanding the scope of structural characteristics, and considering various regions. The machine learning model developed through this approach has the potential to actively replace traditional structural analysis methods and find applications in the construction sector. For example, it can expedite decision-making to determine which structures or areas should be prioritized when implementing precautions in disaster-prone regions. Furthermore, considering the fastness of this method, iterative decision-making for structural design alternatives becomes a feasible option.

Additionally, ML algorithms can be expanded to address various hazards (such as floods, windstorms, and fires) and assess building impact measures, including financial losses and business interruption.

REFERENCES

ABYYHY, 1975, "Regulation on Buildings to be Built in Disaster Areas, Ministry of Construction and Settlement". (In Turkish: Afet Bolgelerinde Yapılacak Yapılar Hakkında Yonetmelik, İmar ve İskan Bakanlığı) Ankara, Turkiye.

AFAD-TADAS, 2020, Earthquake Department of the Disaster and Emergency Management Presidency, Ankara, Türkiye, https://tadas.afad.gov.tr, accessed on Jan 11, 2023.

Alpaydin, E., 2014, *Introduction to Machine Learning*, Third Edition, MIT Press, Massachusetts.

Atakan, K., A. Ojeda, M. Meghraoui, A. Barka, M. Erdik, and A. Bodare, 2002, "Seismic hazard in Istanbul following the August 17, 1999 Izmit and November 12, 1999 Düzce earthquakes", *Bulletin of the Seismological Society of America*, Vol. 92, No. 1, 466–482.

ATC 21, 2002. "Rapid Visual Screening of Buildings for Potential Seismic Hazards - A Handbook (FEMA 154 Report)", Second Edition, ATC-21, Applied Technology Council, California, USA.

Aquib, T. A., J. Sivasubramonian, and P. M. Mai, 2022, "Analysis of Ground Motion Intensity Measures and Selection Techniques for Estimating Building Response", *Applied Sciences*, Vol. 12, No. 23, 2022.

Azak, T. E., B. Ö. Ay, and S. Akkar, 2014, "A statistical study on geometrical properties of Turkish reinforced concrete building stock", *2nd European Conference on Earthquake Engineering and Seismology*, Istanbul, Turkiye.

Baker, J. W., B. Bradley, and P. Stafford, 2021, *Seismic Hazard and Risk Analysis*, Cambridge University Press, Cambridge.

Baker, J. W., C. Goulet, N. Luco, S. Rezaeian, and G. Teng, 2021, "A Subset of CyberShake Ground Motion Time Series for Response History Analysis." *Earthquake Spectra*, Vol. 37, No. 2, pp. 1162-1176

Baker, J. W., and C. Lee, 2018, "An Improved Algorithm for Selecting Ground Motions to Match a Conditional Spectrum", *Journal of Earthquake Engineering*, Vol. 22, No. 4, pp. 708–723.

Bal, E. I., S. S. Tezcan, and F. G. Gülay, 2007, "P25 Rapid Screening Method to Determine the Collapse Vulnerability of Reinforced Concrete Buildings", *Sixth National Conference on Earthquake Engineering*, Istanbul, Turkiye.

Bal, E.I., H. Crowley, R. Pinho, and F.G. Gülay, 2008, "Detailed assessment of structural characteristics of Turkish RC building stock for loss assessment models", *Soil Dynamics and Earthquake Engineering*, Vol. 28, No. 10–11, pp. 914-932.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, 1984, "Classification and Regression Trees", *Taylor and Francis*, Belmont, Calif.: Wadsworth.

Burgos, R. B., and L. E. Silva, 2023, "Evaluation of the P- (P-Delta) effect in columns and frames using the two-cycle method based on the solution of the beam-column differential equation", *MethodsX*, Vol. 11, pp. 102248.

Celep, Z., 2019, *Deprem Mühendisliğine Giriş ve Depreme Dayanıklı Yapı Tasarımı*, Beta Printing Publishing Distribution Inc., Istanbul, Turkiye.

Chopra, A. K., 2013, *Dynamic of Structures: Theory and Applications to Earthquake Engineering*, Forth Edition, Pearson Education, NJ.

Computers and Structures, Inc. (SAP2000). Static and dynamic finite element analysis of structures, 2000.

Cortes, C., and V. Vapnik, 1995, "Support-vector networks", *Machine learning*, Vol. 20, pp. 273–297.

DEZIM, 2020, "The project of updating probable earthquake losses estimation for Istanbul province", Istanbul Metropolitan Municipality, Directorate of Earthquake and Ground Research.

Erdik, M., M. Demircioglu, K. Sesetyan, E. Durukal, and B. Siyahi, 2004, "Earthquake hazard in Marmara Region, Turkey", *Soil Dynamics and Earthquake Engineering*, Vol. 24, No.8, pp. 605–631.

Ghobarah, A., 2004, "On drift limits associated with different damage levels", *International Workshop on Performance-Based Seismic Design: concepts and implementations*, Department of Civil Engineering, McMaster University.

Gulerce, Z. and S. Ocak, 2013, "Probabilistic seismic hazard assessment of Eastern Marmara Region", *Bulletin of Earthquake Engineering*, Vol. 11 No.5, pp. 1259–1277.

Ho, T. K., 1998, "The random subspace method for constructing decision forests", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 8, pp. 832-844.

Hubert-Ferrari, A., A. Barka, E. Jacques, S. Nalbant, B. Meyer, R. Armijo, P. Tapponnier and G. C. P. King, 2000, "Seismic hazard in the Marmara Sea following the 17 August 1999 Izmit earthquake", *Nature*, Vol. 404, pp. 269-272.

Ilki, A., M. Comert, C. Demir, K. Orakcal, D. Ulugtekin, M. Tapan, and N. Kumbasar, 2014, "Performance Based Rapid Seismic Assessment Method (PERA) for Reinforced Concrete Frame Buildings", *Advances in Structural Engineering*, Vol. 17, No. 3, pp. 439-459.

Kalkan, E., P. Gülkan, N. Yilmaz, and M. Çelebi, 2009, "Reassessment of probabilistic seismic hazard in the Marmara Region", *Bulletin of the Seismological Society of America*, Vol. 99, No. 4, pp. 2127–2146.

Kaur, H., H. S. Pannu, and A. K. Malhi, 2020. "A systematic review on imbalanced data challenges in machine learning", *ACM Computing Surveys*, Vol. 52, No.4, pp. 1–36.

Krause, E. F., 1987, *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*, 1975 edition, Addison-Wesley Publishing Company, Menlo Park, California.

JSIM, 2005, *Technical Manual for Seismic Evaluation and Seismic Retrofit of Existing Reinforced Concrete Buildings*, English Version First Edition, The Japan Building Disaster Prevention Association, Tokyo, Japan.

Kiani J., C. Camp, and S. Pezeshk, 2019, "On the application of machine learning techniques to derive seismic fragility curves". *Computers and Structures*, Vol. 218, pp. 108–122.

Mander, J. B., M. J. N. Priestley, and R. Park, 1988, "Theoretical Stress-Strain Model for Confined Concrete", *Journal of Structural Engineering*, Vol. 114, No. 8, pp. 1804-1826.

Mangalathu, S., and H. V., Burton, 2019, "Deep learning-based classification of earthquake-impacted buildings using textual damage descriptions", *International Journal of Disaster Risk Reduction*, Vol. 36, p. 101111.

Mangalathu, S., H. Sun, C. C. Nweke, Z. Yi, and H. V. Burton, 2020, "Classifying earthquake damage to buildings using machine learning", *Earthquake Spectra*, Vol. 36, No. 1, pp. 183–208.

McKenna, F., M. H. Scott, and G. L. Fenves, 2010, "Nonlinear finite-element analysis software architecture using object composition", *Journal of Computing in Civil Engineering*, Vol. 24, No. 1, pp. 95–107.

Murru, M., A. Akinci, G. Falcone, S. Pucci, R. Console, and T. Parsons, 2016, "M 7 earthquake rupture forecast and time-dependent probability for the sea of Marmara region, Turkiye", *Journal of Geophysical Research: Solid Earth*, Vol. 121, pp. 2679–2707.

Parsons, T., S. Toda, R. S. Stein, A. Barka, and J. H. Dieterich, 2000, "Heightened odds of large earthquakes near Istanbul: An interaction-based probability calculation", *Science*, Vol. 288, No. 5466, pp. 661–665.

Parsons, T., 2004, "Recalculated probability of M ¿= 7 earthquakes beneath the Sea of Marmara, Turkey". *Journal of Geophysical Research*, Vol. 109, No. 5.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, 2011, "Scikit-learn: Machine learning in Python", *the Journal of machine Learning research*, Vol. 12, pp. 2825-2830.

Python Software Foundation. (2020). "Python Language Reference", version 3.9.

Roeslin, S., Q. Ma, H. Juárez-Garcia, A. Gómez-Bernal, J. Wicker, and L. Wotherspoon, 2020, "A machine learning damage prediction model for the 2017 Puebla-Morelos, Mexico, earthquake", *Earthquake Spectra*, Vol. 36, pp. 314–339.

Sucuoğlu, H., M. Eeri, U. Yazgan, and A. Yakut, 2007, "A Screening Procedure for Seismic Risk Assessment in Urban Building Stocks", *Earthquake Spectra*, Vol. 23, pp. 441-458.

Stojadinović, Z., M. Kovačević, D. Marinković, and B. Stojadinović, 2021, "Rapid earthquake loss assessment based on machine learning and representative sampling", *Earthquake Spectra*, Vol. 38, No. 1, pp. 152-177.

The MathWorks Inc., 2022, MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts.

Ulku, O., A. T. Atici, F. Yesilirmak, and U. Hancilar, 2022, "Classifying and predicting earthquake damage by using machine learning after the 2020 Elazığ, Turkiye, earthquake", *3rd European Conference on Earthquake Engineering and Seismology*, Bucharest, Romania.

Wu, J., and L. Sarno, 2023, "A machine-learning method for deriving state-dependent fragility curves of existing steel moment frames with masonry infills", *Engineering Structures*, Vol. 276, p. 115345.

Xie, Y., M. Ebad Sichani, J. E. Padgett, and R. DesRoches, 2020, "The promise of implementing machine learning in earthquake engineering: A state-of-the-art review". *Earthquake Spectra*. Vol. 36, No. 4, pp. 1769-1801.

Yang, X., 2019, "6 - Data mining techniques, Introduction to Algorithms for Data Mining and Machine Learning", *Academic Press*, pp. 109-128.

Zhang Y., H.V. Burton, H. Sun, and M., Shokrabadi, 2018, "A machine learning framework for assessing post-earthquake structural safety". *Structural Safety*, Vol. 72, pp. 1–16.

APPENDIX A: SCRIPTS OF GENERATING ARTIFICIAL DATASET, OPENSEESPY, DATA PREPARATION, AND APPLICATION OF MACHINE LEARNING ALGORITHMS

The source code has been made publicly available through the GitHub repository, accessible via the provided link: https://github.com/alitatici/SeismicDamagePrediction.

APPENDIX B: THE SELECTED GROUND MOTIONS

Table B.1. Selected ground motions for period class 1 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
1	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.67	0.90	43.84	32.43
1	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	1.05	0.83	140.04	139.31
1	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3145	0.79	0.56	121.41	84.84
1	270	Izmit Earthquake (17.08.1999)	7.4	East Comp.	8101	1.59	0.59	89.64	36.84
1	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	8002	1.16	0.24	44.40	86.47
1	270	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB	8	2nd Comp.	s351	2.16	0.57	45.41	78.99
1	270	S. San Andreas	7.3	1st Comp.	Silverwood Lake	3.79	1.40	219.61	185.60
1	270	Anacapa-Dume, alt 2	7.3	2nd Comp.	p24	1.37	0.80	121.82	103.58
1	270	Clamshell-Sawpit	6.5	1st Comp.	s480	2.65	0.70	71.25	76.51
1	270	S. San Andreas	8	1st Comp.	s165	0.63	0.91	81.90	65.86
1	270	S. San Andreas;BB+NM+SM+NSB+SSB	8	1st Comp.	s566	1.98	1.53	374.69	294.50

Table B.2. Selected ground motions for period class 1 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
1	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.65	0.87	42.54	31.46
1	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	4404	3.27	0.45	45.48	93.64
1	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	1.02	0.80	136.04	135.33
1	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	2707	4.59	0.42	23.13	15.87
1	560	Ekinozu - Kahramanmaras (02.06.2023)	7.7	East Comp.	129	2.13	0.37	30.41	45.22
1	560	Chino, alt 2	6.5	1st Comp.	s541	1.92	0.71	44.15	18.99
1	560	Puente Hills	7.3	2nd Comp.	s351	3.84	1.18	111.61	81.89
1	560	San Jacinto (SB to C)	7.3	1st Comp.	Moreno valley	1.13	0.84	65.04	73.38
1	560	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB	8	2nd Comp.	s351	2.09	0.55	43.94	76.43
1	560	Santa Ynez Connected	6.5	1st Comp.	s076	3.26	1.04	85.01	40.66
1	560	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG+CO	8	2nd Comp.	s566	2	1.23	199.08	382.07

Table B.3. Selected ground motions for period class 1 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
1	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	4630	1.42	0.25	21.31	11.95
1	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	4404	2.05	0.29	28.51	58.70
1	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.64	0.50	85.36	84.91
1	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.41	0.55	26.83	19.85
1	1130	Ekinozu - Kahramanmaras (02.06.2023)	7.7	East Comp.	129	1.34	0.23	19.13	28.45
1	1130	San Jacinto (SB to C)	7.3	2nd Comp.	s564	1.42	0.45	36.70	63.02
1	1130	S. San Andreas	8	2nd Comp.	s778	1.16	0.45	58.89	124.50
1	1130	S. San Andreas;SSB	6.5	1st Comp.	s648	2.7	0.50	30.06	17.36
1	1130	Santa Monica Connected alt 2	7.3	2nd Comp.	Pasadena	2.59	0.55	37.77	13.56
1	1130	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG+CO	8	2nd Comp.	Antelope	0.9	0.47	62.75	185.24
1	1130	S. San Andreas	6.5	2nd Comp.	s778	4.9	1.15	81.88	119.38

Table B.4. Selected ground motions for period class 2 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
2	270	Izmit Earthquake (17.08.1999)	7.4	East Comp.	4106	3.44	0.50	122.97	144.59
2	270	Izmit Earthquake (17.08.1999)	7.4	East Comp.	8101	2.05	0.76	115.57	47.50
2	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	1.01	0.62	132.76	112.63
2	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	2703	1.46	0.24	24.12	23.48
2	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3139	0.5	0.29	77.72	60.16
2	270	S. San Andreas	6.5	1st Comp.	s359	4.62	1.91	295.55	99.62
2	270	Clamshell-Sawpit	6.5	1st Comp.	Rio Hondo	2.13	0.78	108.44	72.20
2	270	Puente Hills	7.3	1st Comp.	s351	1.21	0.65	67.82	61.46
2	270	S. San Andreas;PK+CH+CC+BB+NM+SM	8	1st Comp.	Pedley	3.44	0.93	226.44	598.59
2	270	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG	8	1st Comp.	Rio Hondo	2.72	0.76	119.38	107.65
2	270	Oak Ridge Connected	6.5	1st Comp.	s066	1.89	0.91	76.60	38.23

Table B.5. Selected ground motions for period class 2 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
2	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3145	0.82	0.49	92.17	93.58
2	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3115	1.11	0.32	45.64	33.65
2	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3125	0.46	0.50	47.21	43.65
2	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.35	0.47	22.90	16.94
2	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	4404	2.95	0.40	58.72	120.05
2	560	S. San Andreas	6.5	1st Comp.	s248	4.4	2.21	99.80	113.17
2	560	Puente Hills	7.3	1st Comp.	s351	1.06	0.57	59.41	53.84
2	560	S. San Andreas;NM+SM	7.3	1st Comp.	s167	2.07	0.73	39.89	20.79
2	560	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG	8	1st Comp.	Rio Hondo	2.39	0.67	104.89	94.59
2	560	San Gabriel	7.3	2nd Comp.	s313	2.28	0.64	75.93	55.09
2	560	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG	8	2nd Comp.	s080	2.39	0.69	52.60	81.95

Table B.6. Selected ground motions for period class 2 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
2	1130	Ekinozu - Kahramanmaras (02.06.2023)	7.7	East Comp.	4620	3.01	0.25	70.54	115.00
2	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3142	0.48	0.31	41.29	29.98
2	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3142	0.35	0.26	25.41	29.48
2	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3141	0.2	0.20	16.11	12.19
2	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3131	0.66	0.24	31.70	34.47
2	1130	S. San Andreas	6.5	1st Comp.	s248	2.55	1.28	57.84	65.58
2	1130	S. San Andreas	8	1st Comp.	Perris	2.51	0.33	37.95	88.51
2	1130	Santa Monica Connected alt 1	6.5	2nd Comp.	p24	1.43	0.49	29.61	14.07
2	1130	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG	8	1st Comp.	Rio Hondo	1.39	0.39	61.01	55.01
2	1130	S. San Andreas	7.3	1st Comp.	s380	1.98	0.68	63.53	63.00
2	1130	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG	8	2nd Comp.	s080	1.38	0.40	30.37	47.32

Table B.7. Selected ground motions for period class 3 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
3	270	Izmit Earthquake (17.08.1999)	7.4	North Comp.	8101	1.18	0.38	63.33	48.05
3	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3123	0.91	0.60	169.97	57.93
3	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.65	0.51	86.69	86.24
3	270	Duzce Earthquake (12.11.1999)	7.1	East Comp.	1401	1.3	1.07	85.58	14.56
3	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3145	1.01	0.61	113.52	115.26
3	270	Puente Hills (LA)	7.3	1st Comp.	s351	1.01	0.75	95.86	44.62
3	270	San Jacinto (SB to C)	7.3	2nd Comp.	Moreno valley	3.9	2.30	479.43	391.51
3	270	Simi-Santa Rosa	6.5	1st Comp.	s066	2.29	0.66	126.22	63.97
3	270	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB+BG	8	2nd Comp.	s351	4.99	0.65	79.05	265.79
3	270	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB	8	2nd Comp.	Chino	1.55	0.69	95.51	106.18
3	270	S. San Andreas	8	1st Comp.	s603	4.46	0.76	168.02	207.95

Table B.8. Selected ground motions for period class 3 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
3	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3141	0.34	0.34	27.40	20.72
3	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.54	0.43	72.02	71.65
3	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3135	0.58	0.44	29.11	33.30
3	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3115	1.19	0.35	48.93	36.07
3	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	4616	0.98	0.49	82.46	69.81
3	560	S. San Andreas	8	1st Comp.	s603	3.7	0.63	139.39	172.52
3	560	S. San Andreas	6.5	1st Comp.	s465	4.93	1.97	141.62	79.87
3	560	Palos Verdes Connected	8	2nd Comp.	s313	2.48	0.57	61.14	61.44
3	560	Sierra Madre Connected	7.3	1st Comp.	s560	1.6	0.59	49.01	29.72
3	560	San Gabriel	7.3	1st Comp.	s313	2.89	0.87	86.97	129.04
3	560	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB+BG	8	2nd Comp.	s351	4.15	0.54	65.74	221.04

Table B.9. Selected ground motions for period class 3 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
3	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.3	0.24	40.01	39.80
3	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3145	0.47	0.28	52.83	53.63
3	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3115	0.66	0.19	27.14	20.01
3	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	2703	0.51	0.08	8.43	8.20
3	1130	Ekinozu - Kahramanmaras (02.06.2023)	7.7	North Comp.	129	2.16	0.33	40.38	55.42
3	1130	Elsinore;T+J+CM	7.3	2nd Comp.	s764	1.61	0.72	45.97	130.12
3	1130	North Frontal (West)	7.3	1st Comp.	Mt. Marie Louise N Grass Valley 5	1.52	0.75	91.73	75.21
3	1130	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB+BG	8	1st Comp.	s351	2.63	0.26	73.09	124.40
3	1130	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB	8	1st Comp.	Nuevo 2	1.91	0.63	33.35	81.46
3	1130	S. San Andreas;CH+CC+BB+NM+SM+NSB	8	2nd Comp.	s605	0.7	0.33	59.54	119.21
3	1130	Chino, alt 2	6.5	1st Comp.	Pleasants Peak	1.84	0.52	37.77	43.38

Table B.10. Selected ground motions for period class 4 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
4	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3123	0.72	0.48	134.48	45.83
4	270	Izmit Earthquake (17.08.1999)	7.4	North Comp.	8101	1.66	0.53	89.09	67.60
4	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.76	1.02	49.73	36.79
4	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.81	0.51	78.57	72.40
4	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.63	0.50	84.03	83.59
4	270	S. San Andreas;NSB+SSB+BG+CO	8	2nd Comp.	Seven Oaks Dam	1.28	0.64	134.51	238.51
4	270	S. San Andreas;BB+NM+SM+NSB	8	1st Comp.	s123	2.17	2.23	459.92	573.72
4	270	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB	8	2nd Comp.	Chino	1.42	0.64	87.50	97.28
4	270	S. San Andreas	8	1st Comp.	Seven Oaks Dam	1.55	0.82	121.30	158.30
4	270	Simi-Santa Rosa	6.5	1st Comp.	s066	2.3	0.66	126.77	64.25
4	270	S. San Andreas	8	1st Comp.	s603	4.33	0.73	163.13	201.89

Table B.11. Selected ground motions for period class 4 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
4	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.67	0.41	88.07	74.71
4	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3139	0.75	0.39	108.95	89.24
4	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.48	0.38	64.02	63.68
4	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.58	0.78	37.95	28.07
4	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	2703	1.63	0.26	26.93	26.21
4	560	S. San Andreas	6.5	1st Comp.	s167	4.76	1.54	225.67	118.99
4	560	Cleghorn	6.5	2nd Comp.	s648	3.84	1.19	90.22	60.76
4	560	S. San Andreas;NSB+SSB+BG+CO	8	2nd Comp.	Seven Oaks Dam	0.98	0.49	102.99	182.61
4	560	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB	8	2nd Comp.	s351	1.77	0.47	37.21	64.73
4	560	Puente Hills (LA)	7.3	2nd Comp.	s351	1.26	0.84	82.57	28.44
4	560	S. San Andreas	8	1st Comp.	s603	3.33	0.56	125.45	155.27

Table B.12. Selected ground motions for period class 4 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
4	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.31	0.42	20.29	15.01
4	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.33	0.21	32.01	29.50
4	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	4404	2.24	0.31	31.15	64.14
4	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.35	0.22	46.01	39.03
4	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3125	0.57	0.45	42.55	37.62
4	1130	Elsinore;W+GI+T+J	7.3	2nd Comp.	s682	2.48	0.66	37.38	38.09
4	1130	S. San Andreas;SM+NSB	7.3	1st Comp.	s167	3.98	1.03	133.51	177.65
4	1130	S. San Andreas	8	1st Comp.	s603	1.76	0.30	66.30	82.06
4	1130	Santa Monica Connected alt 2	7.3	1st Comp.	s313	0.59	0.38	27.03	24.69
4	1130	Chino, alt 2	6.5	1st Comp.	Pleasants Peak	2.56	0.73	52.56	60.35
4	1130	S. San Andreas	8	2nd Comp.	s313	1.32	0.29	31.32	53.92

Table B.13. Selected ground motions for period class 5 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
5	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3140	1.22	0.27	96.48	102.48
5	270	Izmit Earthquake (17.08.1999)	7.4	East Comp.	8101	2.25	0.84	126.85	52.14
5	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.89	0.70	118.70	118.08
5	270	Izmit Earthquake (17.08.1999)	7.4	North Comp.	8101	1.74	0.56	93.38	70.85
5	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3142	0.73	0.47	62.79	45.59
5	270	S. San Andreas;NSB+SSB+BG+CO	8	2nd Comp.	Seven Oaks Dam	1.44	0.71	151.33	268.32
5	270	S. San Andreas;NM+SM+NSB+SSB+BG	8	1st Comp.	s248	2.52	2.13	378.78	838.03
5	270	S. San Andreas;CC+BB+NM+SM	8	2nd Comp.	s526	2.13	0.92	472.55	750.18
5	270	Sierra Madre	6.5	2nd Comp.	Rio Hondo	2.46	0.60	75.74	34.20
5	270	S. San Andreas	8	1st Comp.	Seven Oaks Dam	1.58	0.84	123.65	161.36
5	270	S. San Andreas;BB+NM+SM+NSB+SSB	8	1st Comp.	s566	1.04	0.80	196.81	154.69

Table B.14. Selected ground motions for period class 5 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
5	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3139	0.36	0.18	52.30	42.83
5	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3132	1.4	0.71	94.41	44.46
5	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.66	0.52	88.03	87.57
5	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.52	0.70	34.03	25.17
5	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.64	0.39	84.12	71.37
5	560	Sierra Madre	6.5	2nd Comp.	Rio Hondo	1.83	0.44	56.34	25.44
5	560	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB+BG	8	2nd Comp.	s443	4.49	1.50	240.86	769.08
5	560	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG	8	1st Comp.	Rio Hondo	1.96	0.55	86.02	77.57
5	560	Raymond	6.5	1st Comp.	s351	1.96	0.59	43.67	32.07
5	560	S. San Andreas	8	1st Comp.	Seven Oaks Dam	1.18	0.63	92.35	120.51
5	560	Garlock;GC+GW	8	1st Comp.	Antelope	4.39	0.45	78.01	216.27

Table B.15. Selected ground motions for period class 5 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
5	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.27	0.36	17.67	13.07
5	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3125	0.6	0.47	44.78	39.60
5	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.34	0.27	45.35	45.11
5	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3131	0.72	0.26	34.58	37.60
5	1130	Izmit Earthquake (17.08.1999)	7.4	East Comp.	4106	1.13	0.16	40.40	47.50
5	1130	S. San Andreas	8	1st Comp.	s399	2.36	0.63	63.22	150.11
5	1130	S. San Andreas	7.3	2nd Comp.	s380	2.79	1.10	114.78	269.08
5	1130	S. San Andreas;CH+CC+BB	8	1st Comp.	s167	2.68	0.25	67.06	73.52
5	1130	Malibu Coast, alt 2	6.5	1st Comp.	s195	1.05	0.35	30.53	21.96
5	1130	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB	8	1st Comp.	s351	1.22	0.41	31.36	38.53
5	1130	San Gabriel	7.3	1st Comp.	s313	1.66	0.50	49.95	74.12

Table B.16. Selected ground motions for period class 6 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
6	270	Izmit Earthquake (17.08.1999)	7.4	North Comp.	8101	1.27	0.41	68.16	51.71
6	270	Izmit Earthquake (17.08.1999)	7.4	East Comp.	8101	1.18	0.44	66.52	27.34
6	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3145	1.3	0.78	146.12	148.35
6	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3131	1.09	0.39	48.95	28.57
6	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3140	2.29	0.51	181.09	192.36
6	270	S. San Andreas;NSB+SSB+BG+CO	8	1st Comp.	Seven Oaks Dam	0.86	0.62	83.77	72.58
6	270	North Frontal (West)	7.3	1st Comp.	Mt. Marie Louise N Grass Valley 5	3.78	1.87	228.12	187.03
6	270	S. San Andreas;BB+NM+SM+NSB+SSB	8	1st Comp.	s566	1	0.77	189.24	148.74
6	270	Sierra Madre	6.5	2nd Comp.	Rio Hondo	2.49	0.60	76.67	34.62
6	270	S. San Andreas;CC+BB+NM+SM	8	2nd Comp.	s526	1.93	0.84	428.18	679.74
6	270	Anacapa-Dume, alt 2	7.3	1st Comp.	p24	1.18	0.72	103.30	103.45

Table B.17. Selected ground motions for period class 6 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
6	560	Izmit Earthquake (17.08.1999)	7.4	East Comp.	8101	0.86	0.32	48.48	19.93
6	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3115	1.02	0.30	41.94	30.92
6	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	1.15	0.91	153.38	152.58
6	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3134	0.79	0.16	31.23	52.39
6	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.68	0.42	89.38	75.83
6	560	S. San Andreas;CH+CC+BB	8	1st Comp.	s167	3.99	0.37	99.83	109.46
6	560	Cucamonga	6.5	1st Comp.	s586	2.22	1.71	77.60	25.58
6	560	S. San Andreas;BB+NM+SM+NSB+SSB	8	1st Comp.	s566	0.73	0.56	138.14	108.58
6	560	Verdugo	7.3	2nd Comp.	s351	0.52	0.50	65.02	16.17
6	560	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG	8	1st Comp.	Rio Hondo	2.54	0.71	111.48	100.52
6	560	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB	8	1st Comp.	s351	2.99	1.01	76.85	94.42

Table B.18. Selected ground motions for period class 6 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
6	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3131	0.4	0.14	17.96	10.48
6	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3125	0.34	0.37	34.90	32.26
6	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3145	0.48	0.29	53.95	54.78
6	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.35	0.22	46.01	39.03
6	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.59	0.47	78.69	78.28
6	1130	S. San Andreas;NSB+SSB+BG+CO	8	1st Comp.	Seven Oaks Dam	0.32	0.23	31.17	27.01
6	1130	S. San Andreas	8	1st Comp.	Seven Oaks Dam	0.74	0.39	57.91	75.57
6	1130	S. San Andreas	8	2nd Comp.	Pasadena	2.4	0.37	49.16	60.39
6	1130	S. San Andreas;SM	7.3	1st Comp.	s165	1.19	0.30	37.30	28.22
6	1130	S. San Andreas	8	2nd Comp.	s690	0.72	0.66	113.83	242.46
6	1130	S. San Andreas	7.3	2nd Comp.	s380	2.76	1.09	113.55	266.18

Table B.19. Selected ground motions for period class 7 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
7	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.71	0.45	68.87	63.46
7	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.89	1.19	58.24	43.08
7	270	Izmit Earthquake (17.08.1999)	7.4	North Comp.	8101	2.44	0.78	130.95	99.36
7	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3124	0.75	0.43	84.24	35.45
7	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3139	0.96	0.49	139.45	114.23
7	270	S. San Andreas	7.3	1st Comp.	s586	2.76	1.85	265.55	262.14
7	270	S. San Andreas;CC+BB+NM+SM	8	2nd Comp.	s526	2.17	0.94	481.43	764.26
7	270	S. San Andreas;NSB+SSB+BG+CO	8	1st Comp.	Seven Oaks Dam	0.67	0.48	65.26	56.54
7	270	S. San Andreas;NSB+SSB+BG+CO	8	2nd Comp.	Seven Oaks Dam	1.16	0.58	121.90	216.15
7	270	Chino, alt 1	6.5	1st Comp.	Chino	1.26	0.58	164.27	82.34
7	270	S. San Andreas;BB+NM+SM+NSB+SSB	8	1st Comp.	s566	1.05	0.81	198.70	156.17

Table B.20. Selected ground motions for period class 7 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
7	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.5	0.32	48.50	44.69
7	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.63	0.84	41.23	30.49
7	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.78	0.48	102.53	86.98
7	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3142	0.58	0.38	49.89	36.22
7	560	Ekinozu - Kahramanmaras (02.06.2023)	7.7	East Comp.	4617	3.69	0.31	100.94	136.18
7	560	S. San Andreas;CH+CC+BB	8	1st Comp.	s167	3.92	0.36	98.08	107.54
7	560	North Frontal (West)	7.3	1st Comp.	Mt. Marie Louise N Grass Valley 5	2.9	1.44	175.01	143.49
7	560	S. San Andreas;NSB+SSB+BG+CO	8	2nd Comp.	Seven Oaks Dam	0.83	0.41	87.22	154.66
7	560	S. San Andreas	6.5	2nd Comp.	Seven Oaks Dam	2.34	0.82	100.26	23.51
7	560	S. San Andreas;BB+NM+SM+NSB	8	1st Comp.	s313	3.28	0.69	84.80	188.01
7	560	Anacapa-Dume, alt 2	7.3	1st Comp.	p24	0.84	0.51	73.54	73.64

Table B.21. Selected ground motions for period class 7 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
7	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3125	0.34	0.37	34.90	32.26
7	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3131	0.35	0.13	16.81	18.28
7	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	8003	1.09	0.16	32.29	54.39
7	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.39	0.24	51.26	43.49
7	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.58	0.46	77.36	76.95
7	1130	S. San Andreas;CH+CC+BB+NM+SM	8	2nd Comp.	s165	0.74	0.44	43.37	127.47
7	1130	S. San Andreas;SM+NSB	7.3	1st Comp.	s167	4.18	1.08	140.22	186.58
7	1130	S. San Andreas;CH+CC+BB	8	1st Comp.	s167	1.98	0.18	49.54	54.32
7	1130	S. San Andreas;SM	7.3	1st Comp.	s165	1.16	0.30	36.36	27.51
7	1130	Clamshell-Sawpit	6.5	2nd Comp.	Rio Hondo	1.6	0.50	51.43	30.21
7	1130	S. San Andreas;BB+NM+SM+NSB+SSB+BG+CO	8	2nd Comp.	s732	1.18	0.43	60.21	134.33

Table B.22. Selected ground motions for period class 8 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
8	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	1.55	1.22	206.73	205.65
8	270	Izmit Earthquake (17.08.1999)	7.4	North Comp.	8101	1.33	0.43	71.38	54.16
8	270	Ekinozu - Kahramanmaras (02.06.2023)	7.7	North Comp.	4617	4.27	0.24	96.58	92.43
8	270	Golyaka - Duzce (23.11.2023)	5.9	East Comp.	8102	1.95	0.80	143.55	45.59
8	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.67	0.42	64.99	59.89
8	270	S. San Andreas;NSB+SSB+BG+CO	8	1st Comp.	Seven Oaks Dam	0.82	0.59	79.88	69.20
8	270	S. San Andreas	7.3	1st Comp.	s586	3.1	2.08	298.27	294.43
8	270	Simi-Santa Rosa	6.5	1st Comp.	s066	2.46	0.70	135.59	68.72
8	270	S. San Andreas	7.3	1st Comp.	s167	2.01	0.82	157.40	137.76
8	270	Sierra Madre	7.3	1st Comp.	s560	1.88	0.83	147.38	105.74
8	270	S. San Andreas	8	2nd Comp.	Chino	2.26	0.82	157.06	156.94

Table B.23. Selected ground motions for period class 8 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
8	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3125	0.55	0.60	56.45	52.18
8	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3139	0.65	0.33	94.42	77.34
8	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.97	0.60	127.50	108.17
8	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3131	0.57	0.21	25.60	14.94
8	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.45	0.28	43.65	40.22
8	560	S. San Andreas	6.5	2nd Comp.	s125	4.98	1.13	102.90	88.42
8	560	San Gabriel	7.3	1st Comp.	s313	4.56	1.37	137.22	203.60
8	560	S. San Andreas;CH+CC+BB	8	1st Comp.	s167	4.11	0.38	102.84	112.75
8	560	Puente Hills (LA)	7.3	1st Comp.	s351	0.64	0.48	60.75	28.27
8	560	Clamshell-Sawpit	6.5	1st Comp.	Rio Hondo	1.4	0.57	88.48	45.05
8	560	Simi-Santa Rosa	6.5	1st Comp.	s066	1.66	0.47	91.50	46.37

Table B.24. Selected ground motions for period class 8 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
8	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3125	0.27	0.30	27.71	25.62
8	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.23	0.15	22.31	20.56
8	1130	Ekinozu - Kahramanmaras (02.06.2023)	7.7	East Comp.	4617	2.12	0.18	57.99	78.24
8	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3115	0.49	0.14	20.15	14.85
8	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.49	0.30	64.41	54.64
8	1130	San Jacinto;SBV+SJV+A	7.3	1st Comp.	s524	1.94	0.54	34.92	13.83
8	1130	S. San Andreas	7.3	2nd Comp.	s380	2.97	0.84	100.50	134.95
8	1130	S. San Andreas;PK+CH+CC+BB+NM+SM	8	2nd Comp.	Pedley	1.01	0.25	58.05	75.05
8	1130	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB+SSB	8	2nd Comp.	Chino	0.61	0.27	37.59	41.79
8	1130	S. San Andreas;CC+BB+NM+SM+NSB+SSB+BG	8	1st Comp.	Rio Hondo	1.15	0.32	50.47	45.51
8	1130	S. San Andreas;CC+BB+NM+SM	8	2nd Comp.	s526	0.62	0.36	92.36	133.48

Table B.25. Selected ground motions for period class 9 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
9	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3131	2.08	0.75	93.41	54.51
9	270	Golyaka - Duzce (23.11.2023)	5.9	East Comp.	8102	1.45	0.59	106.74	33.90
9	270	Izmit Earthquake (17.08.1999)	7.4	North Comp.	8101	2.42	0.78	129.88	98.54
9	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	1.09	0.86	145.38	144.62
9	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.43	0.27	41.71	38.43
9	270	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB	8	1st Comp.	s526	1.66	0.71	229.85	325.78
9	270	Santa Monica Connected alt 2	6.5	2nd Comp.	p25	3.78	2.54	222.10	194.99
9	270	S. San Andreas	7.3	1st Comp.	s167	1.99	0.81	155.83	136.39
9	270	S. San Andreas	8	1st Comp.	Seven Oaks Dam	1.28	0.68	100.17	130.72
9	270	S. San Andreas;CC+BB+NM+SM	8	1st Comp.	s526	1.24	1.00	258.58	134.16
9	270	Chino, alt 1	6.5	2nd Comp.	Chino	1.91	0.80	147.73	59.81

Table B.26. Selected ground motions for period class 9 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
9	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.82	1.10	53.66	39.69
9	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.28	0.18	27.16	25.03
9	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.69	0.54	92.03	91.55
9	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3131	1.33	0.48	59.73	34.86
9	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3139	0.61	0.31	88.61	72.58
9	560	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB	8	1st Comp.	s526	1.06	0.45	146.77	208.03
9	560	Oak Ridge Connected	6.5	2nd Comp.	s066	2.04	0.99	88.92	48.89
9	560	Sierra Madre	7.3	1st Comp.	s560	1.24	0.55	97.21	69.74
9	560	S. San Andreas	8	1st Comp.	Seven Oaks Dam	0.82	0.44	64.17	83.74
9	560	S. San Andreas	7.3	1st Comp.	s545	2.73	0.48	119.63	116.43
9	560	S. San Andreas;CC+BB+NM+SM+NSB+SSB	8	2nd Comp.	s668	1.88	1.58	111.18	309.59

Table B.27. Selected ground motions for period class 9 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
9	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	2703	1.09	0.18	18.01	17.53
9	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3131	0.67	0.24	30.09	17.56
9	1130	Ekinozu - Kahramanmaras (02.06.2023)	7.7	East Comp.	4617	2.12	0.18	57.99	78.24
9	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	4616	0.48	0.24	40.39	34.19
9	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.35	0.28	46.68	46.44
9	1130	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB	8	1st Comp.	s526	0.53	0.23	73.39	104.01
9	1130	S. San Andreas;BB+NM+SM+NSB	7.3	2nd Comp.	s361	2.29	0.78	47.63	96.47
9	1130	San Jacinto;SJV+A+C	8	1st Comp.	s720	2.87	0.30	67.52	60.78
9	1130	San Gabriel	6.5	1st Comp.	s161	0.93	0.28	42.67	11.42
9	1130	S. San Andreas;PK+CH+CC+BB+NM+SM	8	2nd Comp.	s486	1.28	0.37	127.34	368.55
9	1130	Elsinore;GI	6.5	2nd Comp.	Pleasants Peak	2.96	0.81	71.06	72.74

Table B.28. Selected ground motions for period class 10 and V_{S30} 270 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
10	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3124	0.36	0.23	34.92	32.18
10	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3131	1.75	0.63	78.59	45.86
10	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3140	1.88	0.41	148.67	157.92
10	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.89	0.70	118.70	118.08
10	270	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3139	0.82	0.42	119.12	97.57
10	270	S. San Andreas;CC+BB+NM+SM+NSB+SSB	8	2nd Comp.	s668	2.98	2.50	176.23	490.74
10	270	S. San Andreas;CH+CC+BB+NM+SM+NSB	8	1st Comp.	s566	1.26	0.89	235.12	374.27
10	270	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB	8	1st Comp.	s526	1.68	0.72	232.62	329.71
10	270	S. San Andreas	7.3	1st Comp.	s167	2.01	0.82	157.40	137.76
10	270	S. San Andreas;CC+BB+NM+SM	8	1st Comp.	s526	1.21	0.97	252.33	130.91
10	270	S. San Andreas	8	1st Comp.	Seven Oaks Dam	1.33	0.71	104.08	135.83

Table B.29. Selected ground motions for period class 10 and V_{S30} 560 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
10	560	Izmit Earthquake (17.08.1999)	7.4	East Comp.	8101	1.01	0.38	56.94	23.40
10	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3135	0.91	1.22	59.55	44.05
10	560	Ekinozu - Kahramanmaras (02.06.2023)	7.7	East Comp.	4617	3.18	0.27	86.99	117.36
10	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.56	0.44	74.69	74.30
10	560	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3131	1.09	0.39	48.95	28.57
10	560	S. San Andreas	8	1st Comp.	s228	1.87	1.28	109.28	140.88
10	560	Anacapa-Dume, alt 2	7.3	1st Comp.	p24	1.62	0.99	141.82	142.03
10	560	S. San Andreas;PK+CH+CC+BB+NM+SM+NSB	8	1st Comp.	s526	1.05	0.45	145.39	206.07
10	560	S. San Andreas	8	1st Comp.	Seven Oaks Dam	0.83	0.44	64.96	84.77
10	560	Clamshell-Sawpit	6.5	1st Comp.	Rio Hondo	1.51	0.61	95.43	48.59
10	560	S. San Andreas	7.3	1st Comp.	s545	2.72	0.47	119.20	116.00

Table B.30. Selected ground motions for period class 10 and V_{S30} 1130 m/s.

Period Class	**V_{S30}**	**Selected Earthquake**	**M_w**	**Horizontal Component**	**Station**	**Scale Factor**	**PGA (g)**	**PGV (cm/s)**	**PGD (cm)**
10	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3144	0.49	0.30	64.41	54.64
10	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	North Comp.	3125	0.45	0.35	33.59	29.70
10	1130	Ekinozu - Kahramanmaras (02.06.2023)	7.7	East Comp.	4617	1.58	0.13	43.22	58.31
10	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3144	0.28	0.22	37.34	37.15
10	1130	Pazarcik Kahramanmaras (02.06.2023)	7.8	East Comp.	3131	0.54	0.20	24.25	14.15
10	1130	Anacapa-Dume, alt 2	7.3	1st Comp.	p24	0.81	0.49	70.91	71.01
10	1130	S. San Andreas	7.3	2nd Comp.	s380	2.44	0.96	100.38	235.32
10	1130	S. San Andreas;CH+CC+BB	8	1st Comp.	s167	1.94	0.18	48.54	53.22
10	1130	Puente Hills	7.3	2nd Comp.	s351	1.23	0.38	35.75	26.23
10	1130	Clamshell-Sawpit	6.5	1st Comp.	s480	1.46	0.38	39.25	42.16
10	1130	S. San Andreas;NSB+SSB	7.3	1st Comp.	s526	0.55	0.25	85.88	71.94