Volume 14, Issue 2, Page 37-51, 2025 https://doi.org/10.46810/tdfd.1605168 Research Article

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science
www.dergipark.gov.tr/tdfd

Multiple Regression-Based Prediction Method to Assess the Impact of PGA and Distance on Post-Earthquake Structural Damage Levels

Ömer Faruk NEMUTLU[1*]

[1] Bingol University, Engineering and Architecture Faculty, Civil Engineering Department, Bingöl, Türkiye
Ömer Faruk NEMUTLU ORCID No: 0000-0001-7841-3911

Corresponding author: ofnemutlu@bingol.edu.tr

(Received: 23.12.2024, Accepted: 08.04.2025, Online Publication: 27.06.2025)

Keywords
Multiple regression, Artificial intelligence, Damage distribution, Earthquake observation, Kahramanmaraş earthquakes

Abstract: This study was carried out to evaluate the accuracy of the damage assessments made after the 06 February 2023 Kahramanmaras earthquakes and to ensure that these data are a guide for future studies in the field of earthquake engineering. The relationship between damage levels, peak ground acceleration (PGA) values measured by Disaster and Emergency Management Affair (DEMA) stations and distances to earthquake-affected cities were analyzed. Unlike the studies in literature, evaluation was made on multiple input and multiple output parameters, and a separate regression model was used for each output data. As a result of regression analysis, a significant relationship was found between damage levels and PGA-distance parameters. The R^2 scores for the "No damage" and "Heavy damage" levels were found to be 0.75 and 0.71, respectively. In the analyzes made by reducing the damage levels to two main categories (damaged and undamaged), the R^2 scores were calculated as 0.63 and 0.6, respectively. These results show that there is a sufficient level of agreement between the input and output parameters, but they reveal that the dataset should be expanded, and the positional details of the structures should be obtained separately for higher accuracy. Within the scope of the study, linear regression, polynomial regression, random forest and gradient boosting models were used and their performances were compared. According to the results obtained, gradient boosting and random forest models were the models that exhibited the best compatibility according to damage levels (0.75 and 0.71 R^2 scores for No damage and Heavy damage, respectively). In particular, the fact that the random forest model gives the best results in 5 out of 6 damage levels shows that the model is a method that produces fast and reliable results in such complex analyses. As a result, it was determined that model performance at low conforming damage levels could be improved by expanding the data set and increasing the available data details. These findings make important contributions to the accuracy analysis of damage assessments after earthquakes and provide a scientific basis for similar studies.

PGA ve Mesafenin Deprem Sonrası Yapısal Hasar Seviyeleri Üzerindeki Etkisini Değerlendirmek için Çoklu Regresyon Tabanlı Tahmin Yöntemi

Anahtar Kelimeler
Çoklu regresyon, Yapay zeka, Hasar dağılımı, Deprem gözlemi, Kahramanmaraş depremleri

Öz: Bu çalışma, 06 Şubat Kahramanmaras depremleri sonrasında yapılan hasar tespitlerinin doğruluğunu değerlendirmek ve bu verilerin deprem mühendisliği alanındaki gelecekteki çalışmalar için rehber olmasını sağlamak amacıyla gerçekleştirilmiştir. Hasar seviyeleri, Afet ve Acil Durum Yönetimi Başkanlığı (AFAD) istasyonlarının ölçtüğü en büyük yer ivmesi (PGA) değerleri ve depremden etkilenen şehirlere olan mesafeleri arasındaki ilişki analiz edilmiştir. Literatürdeki çalışmalardan farklı olarak, çoklu giriş ve çoklu çıkış parametreleri üzerinden değerlendirme yapılmış ve her bir çıkış verisi için ayrı regresyon modeli kullanılmıştır. Regresyon analizleri sonucunda, hasar seviyeleri ile PGA-mesafe parametreleri arasında anlamlı bir ilişki tespit edilmiştir. "Hasar yok" ve "Ağır hasar" seviyeleri için R^2 skorları sırasıyla 0.75 ve 0.71 olarak bulunmuştur. Hasar seviyeleri iki ana kategoriye (hasarlı ve hasarsız) indirgenerek yapılan analizlerde ise R^2 skorları sırasıyla 0.63 ve 0.6 olarak hesaplanmıştır. Bu sonuçlar, giriş ve çıkış parametreleri arasında yeterli düzeyde uyum olduğunu göstermekle birlikte, daha yüksek doğruluk için veri setinin genişletilmesi ve yapıların konumsal detaylarının ayrı ayrı elde edilmesi gerektiğini ortaya koymaktadır. Çalışma kapsamında lineer regresyon, polinomal regresyon, random forest ve gradient boosting modelleri kullanılmış ve performansları karşılaştırılmıştır. Elde edilen sonuçlara göre gradient boosting ve random forest modelleri, hasar seviyelerine göre en iyi uyumu sergileyen modeller olmuştur. Bu modeller Hasarsız ve Ağır hasarlı durum için sırasıyla 0.75 ve 0.71 R^2 değerleri almıştır. Özellikle random forest modelinin 6 hasar seviyesinden 5'inde en iyi sonuçları vermesi, bu tür karmaşık analizlerde modelin hızlı ve güvenilir sonuçlar üreten bir yöntem olduğunu göstermektedir. Sonuç olarak, düşük uyum gösteren hasar seviyelerinde model performansının, veri setinin genişletilmesi ve mevcut veri detaylarının artırılmasıyla iyileştirilebileceği belirlenmiştir. Bu bulgular, depremler sonrası hasar tespitlerinin doğruluk analizine önemli katkılar sağlamakta ve benzer çalışmalar için bilimsel bir temel oluşturmaktadır.

1. INTRODUCTION

Earthquakes are one of the most important natural disasters within the borders of Turkiye in recent years. Because it contains the world's most active fault zones within its borders. The North Anatolian Fault Zone (NAFZ), the East Anatolian Fault Zone (EAFZ) and the West Anatolian Fault Zone (WAFZ) in the west of the country are faults that have the potential to produce significant earthquakes. In the past years, important earthquakes have occurred in these fault zones [1]. The 2003 Bingol Earthquake, the 2011 Van Earthquake, the 2020 Elazig Earthquake, the 2020 Izmir Earthquake, the 2023 Kahramanmaras Earthquakes are important destructive earthquakes that have occurred in Turkiye in the last quarter century [2]. These earthquakes caused significant loss of life and property. The Kahramanmaras earthquakes, which have passed for a very short time, are among the most important of these destructive earthquakes. The occurrence of two earthquakes with a magnitude of 7.7 M_w and 7.6 M_w only 9 hours apart has greatly increased the level of destruction and losses [3–6]. The first earthquake was an earthquake with an epicenter in Pazarcik. The earthquake, which occurred on February 06, 2023, at 04:17 local time, occurred on the EAFZ. When the surface fractures are examined, they are broken together with the EAFZ and the Oludeniz Fault Zone, which is the continuation of this fault zone. The surface deformation caused by this earthquake is about 300 km [7]. It continued the broken Amanos segment and proceeded to the city center of Hatay province. On the same day, after the first earthquake, the earthquake with the epicenter of Elbistan, which occurred at 13.24 local time, occurred on the Cardak Fault zone, one of the branches of the EAFZ [8].

The M_w 7.7 and M_w 7.6 magnitude earthquakes that occurred in Kahramanmaras on February 6, 2023 caused extensive loss of life and property. Field observations reveal that serious damage occurs due to design and construction errors, especially in reinforced concrete buildings. Among the main structural deficiencies, factors such as strong beam-weak column effect, short column formation, soft floor irregularities, errors in reinforcement placement and inadequate concrete quality stand out. In addition, the damage causes determined by field observations were also confirmed by nonlinear finite element analyses. This situation once again demonstrates the importance of evaluating the existing building stock in terms of compliance with earthquake regulations and carrying out the necessary retrofitting works [10]. The earthquakes of magnitude 7.8 and 7.6 and the Hatay earthquake of magnitude 6.4 that occurred on February 6, 2023 caused serious damage to various structures and critical infrastructures. Field observations and analyses evaluated the damage to residential, commercial and industrial structures, roads, bridges and energy systems [11]. Avğın et al. examines the acceleration records, spectral analyses and structural and geotechnical damage causes of earthquakes in their study. It was determined that 57% of the buildings in Kahramanmaras were damaged, and the most severe damage was concentrated in Dulkadiroğlu, Onikişubat and Göksun districts. The soft story effect, strong beam-weak column formation, inadequate shear wall use, low material quality and weak soil conditions are prominent among the damage causes [12]. Işık et al., in their study, examined the damages in 20 settlements located directly on the fault line and compared the PGA estimates in Turkey's current earthquake hazard maps with the actual measurements. In addition, reinforced concrete structures were evaluated in terms of earthquake engineering and pushover analyses were performed on a sample building model. The results showed that the target displacements were exceeded in some settlements and not in others. It was concluded that a more realistic representation of the earthquake hazard would increase the accuracy of building performance estimates [13].

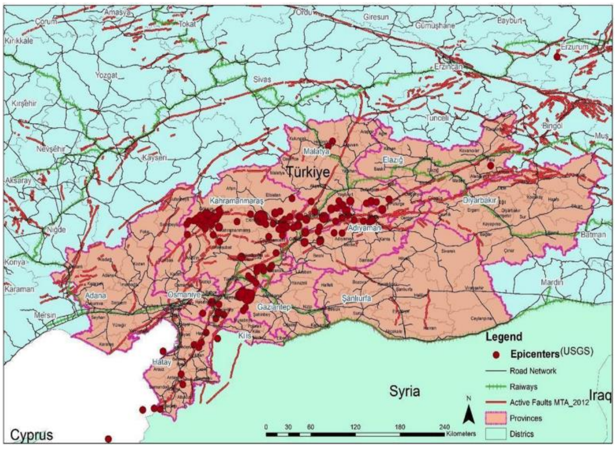


Figure 1. Cities affected by the Kahramanmaras earthquakes[9]

Research has shown that indicators such as ground motion parameters, PGA, play a critical role in determining the level of structural damage during earthquakes. For example, Tao & Cai investigated the relationship between ground motion parameters and simulated structural damage and emphasized the importance of PGA in terms of damage estimation [14]. In addition, Zhou & Sun stated that a number of factors should be considered in post-earthquake damage assessments and emphasized the importance of ground motion characteristics among these factors [15]. Similarly, Liang et al. examined the effect of epicenteral distance on structural damage and showed that increasing distances were generally associated with less damage [16]. In the context of the Kahramanmaras earthquakes, Karaşin emphasized the unique factors contributing to structural damage in these events, especially addressing the effects of the duration and intensity of shaking. The study found that local geological conditions and material quality caused the diversity of damage [17]. In addition, Zengin & Aydin study emphasized that the observed damage is particularly attributed to the inadequacy of construction practices and poor material quality in the region, which increases the vulnerability of buildings to earthquakes [18]. Research supports the development of a multi-dimensional approach to understand structural weaknesses and solve the problems we face. Such methods allow for the development of strategies for improving construction materials and construction practices. Therefore, careful examination of factors such as PGA, epicenter distance and construction quality enables engineers and policy makers to develop more effective building codes and disaster response strategies [19].

In this study, the effects of PGA and distance to the earthquake epicenter on the post-earthquake damage levels of structures were investigated using a multiple regression model. The February 6, 2023 Kahramanmaras earthquake was used to create and validate the model. These earthquakes, which caused great destruction in Kahramanmaras and its surroundings, affected a wide area with different ground properties and building types. The 7.7 and 7.6 magnitude main shocks and the aftershocks following these main shocks seriously tested the resistance capacities of structures in the region and caused extensive damage. Therefore, the Kahramanmaras earthquakes provide a comprehensive data set to analyze the effects of variables such as PGA and distance on building damage.

Although the attenuation relationships in the literature show a relationship between PGA and distance, this relationship is not combined with damage. This combination can be achieved with multiple regression-based approaches. Some studies in this field examine how high ground accelerations change the damage level [20]. Multiple regression-based approaches allow for fast and low-cost large-scale damage analyses in this field.

Artificial intelligence has become an important part of our lives today. Multiple regression analyses, which are needed in such studies, can now be easily performed with artificial intelligence tools. Linear regression is the simplest regression technique for determining the linear relationship between the dependent variable and the independent variables. The resulting model expresses the relationship between the variables with a linear equation. The analysis of this model is easy and its results can be interpreted clearly. However, linear regression only gives effective results in cases where linear relationships exist. Polynomial regression is used when the relationships between the dependent variable and the independent variables are not linear. Therefore, it makes it possible to capture nonlinear relationships by adding second or higher order terms to model the more complex structure of the data [21]. However, polynomial regression is more susceptible to the problem of overfitting, especially when high-degree polynomials are used [22]. Gradient boosting and random forest are powerful machine learning methods based on decision trees. Random forest combines the predictions of each tree by creating multiple decision trees. This model provides high accuracy and low variance, and is more resistant to overfitting [23]. Studies have shown that gradient boosting generally provides better prediction performance than random forest. For example, gradient boosting models have been shown to provide mean squared error (MSE) compared to random forest [23]. However, random forest may be a more easily implemented option due to the flexibility of the model [24]. In another study, comparisons between gradient boosting and random forest showed that although gradient boosting provides higher accuracy, random forest requires less processing time [25]. In such a case, the integration of researchers into the fields they work in has become inevitable[26–28]. As part of this study, it is aimed to be used in studies in the field of earthquakes with artificial intelligence tools. Studies between artificial intelligence tools and earthquakes are available in the literature. Artificial intelligence tools are used in the classification of damage that occurs after earthquakes and earthquake risk analysis studies[29,30]. It is also used in damage assessment studies from satellite images. In his study, Nemutlu 2024[31] made an assessment on the level of damage in the earthquake-affected regions using satellite images of the Kahramanmaras earthquakes. In the study, satellite images taken before February 6 and satellite images after February 6 were evaluated through image processing techniques and deep learning models, and examinations were made on the determination of areas where the number of damaged structures is intense. On the other hand, there are studies on the detection of damage with damaged building visuals[32,33].

In the context of this research, the relationship between earthquake parameters and damages will be examined by using artificial intelligence tools. In the study, the parameters of the earthquakes were obtained from the earthquake stations, and the relationship between the damaged levels of the damaged structures collected by field work was analyzed by regression models and machine learning methods. The maximum ground acceleration (PGA) caused by the earthquake and the distance of the earthquake stations to the cities affected by the earthquake were used as input data, and the level of damage to the buildings after the earthquake was evaluated by accepting the output data as a dependent variable. The analyzes and analysis processes are explained in detail in the following sections. The results obtained were examined with their justifications.

2. MATERIAL AND METHOD

2.1. Kahramanmaras Earthquakes and Seismicity of the Region

Two destructive earthquakes caused significant loss of life and property at the 06 February 2023. Over 250000 buildings collapsed or severely damaged. 11 cities were directly affected by the earthquake. These cities are Adiyaman, Malatya, Kahramanmaras, Hatay, Elazig, Sanliurfa, Kilis, Gaziantep, Diyarbakir, Adana and Osmaniye. Most of the destruction is concentrated in the provinces of Adiyaman, Kahramanmaras, Hatay and Malatya [34,35]. In total, more than 14 million people living in 11 provinces were directly affected by the earthquake, and more than 50000 people lost their lives because of the collapsed buildings caused by the earthquake[36]. Figure 1 shows the 11 cities affected by the earthquake on a map of Turkiye. Figure 2 shows the fault lines where the earthquake occurred and the fractures that occurred.

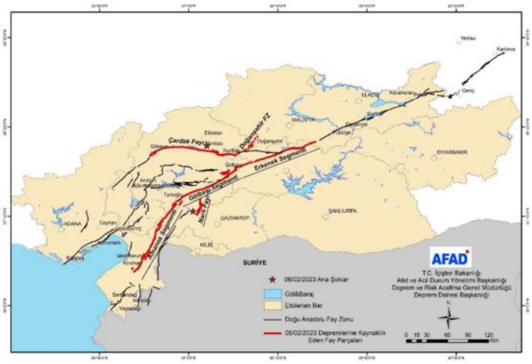
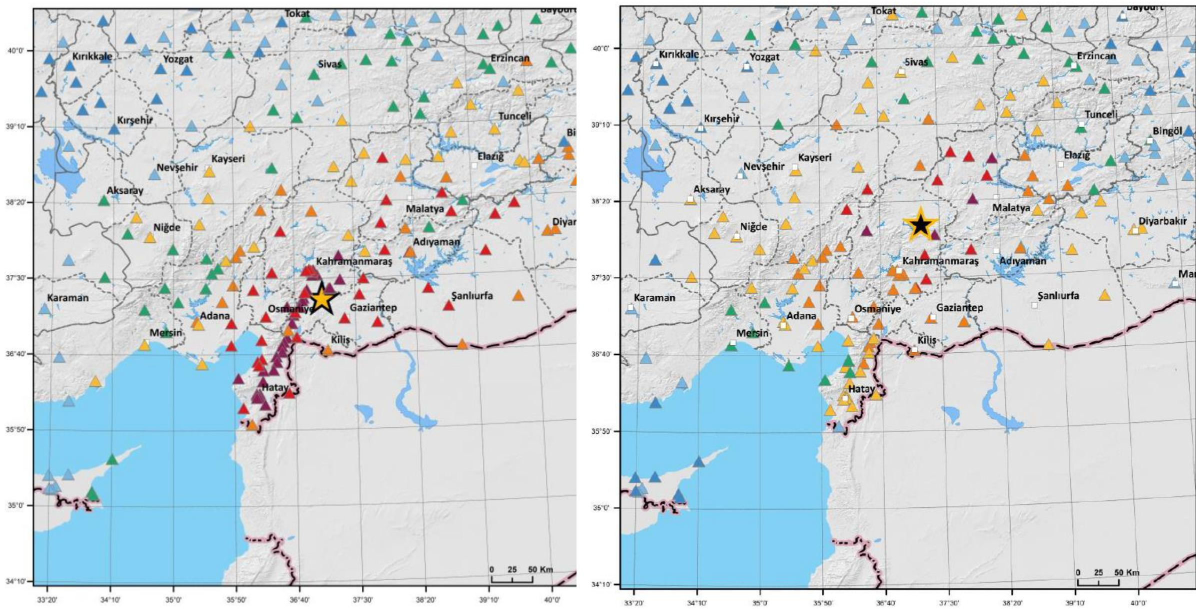


Figure 2. Map showing fault parts of the East Anatolian Fault[8]



(a) 7.7 M_w Magnitude earthquake (b) 7.6 M_w Magnitude earthquake

Figure 3. Distribution of accelerometer stations in the area that recorded the 7.7 M_w and 7.6 M_w Magnitude earthquake[37]

After the earthquakes, many institutions, especially the Disaster and Emergency Management Affairs (DEMA), took measures related to the earthquake. The distribution of DEMA stations, which have a widespread station network in Turkiye, is given in Figure 3. Figure 3 shows the epicenters of the two earthquakes and the stations and locations that took records because of the earthquakes[38]. Aftershocks occurred after the earthquakes. The most important aftershock of these was the 6.4 M_w magnitude earthquake centered in Hatay on February 20. Figure 4 shows aftershock activity from 6 February to 6 May.

Table 1 and Table 2 show the distances from the epicenters of the earthquake to the nearest settlements according to DEMA information. [38]

2.2. Artificial Intelligence Tools and Machine Learning Process

Today, the analysis of data with artificial intelligence has started to be included in the subjects of researchers. Artificial intelligence can be classified in general terms as methods and problems. When looking at the methods of artificial intelligence, one of the most widely used methods is machine learning[39,40]. Commonly used methods besides machine learning are given in Figure 5. Machine learning, on the other hand, is divided into two sub-headings as traditional methods and deep learning according to the evaluation made by researchers in the most general sense. Traditional methods are methods such as regression, support vector machines, decision trees, artificial neural networks[41,42]. Although deep learning is an artificial intelligence approach based on artificial neural networks, its usage area and the applied process distinguish it from other machine learning methods. On the other hand, the applicability of these methods is related to how the data will be evaluated. To evaluate the available data with artificial intelligence tools, it is necessary to determine the problem. Artificial intelligence evaluates data through two different problems. These problems are regression and classification problems. How the data will be analyzed and which problem definition it conforms to affects the success rate of artificial intelligence. In machine learning, the input data we have is trained by a model and obtained as output data. These models, which are located between the input and output data, vary according to the problem at hand and the data contained in this problem. Machine learning methods will be applied within the scope of the study. The main topic of study is regression problems. To evaluate regression problems, many regression models are within the scope of machine learning. Figure 6 provides the subheadings of machine learning and the details of these headings. The general concept of machine learning is given in Figure 7.

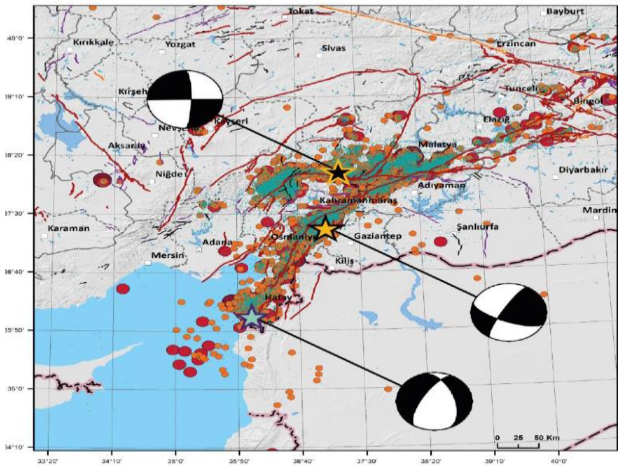


Figure 4. Distribution of aftershock activity from the 7.7 and 7.6 earthquakes (6 February to 6 May)[37]

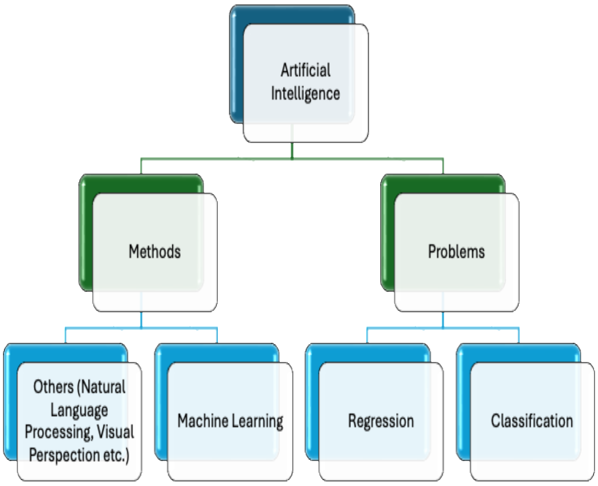


Figure 5. Methods and problems in the use of artificial intelligence

Table 1. Settlements near the epicenter affected by the M_w 7.7 earthquake

Country	City	County	Site	Distance (km)
Turkiye	Kahramanmaras	Pazarcik	Akdemir	2.72
Turkiye	Kahramanmaras	Pazarcik	Karahuyuk	2.84
Turkiye	Kahramanmaras	Turkoglu	Cennetpinari	3.75
Turkiye	Kahramanmaras	Pazarcik	Evri	4.48
Turkiye	Kahramanmaras	Pazarcik	Emiroglu	4.94

Table 2. Settlements near the epicenter affected by the M_w 7.6 earthquake[38]

Country	City	County	Site	Distance (km)
Turkiye	Kahramanmaras	Elbistan	Gumusdoven	1.70
Turkiye	Kahramanmaras	Ekinozu	Akpinar	2.09
Turkiye	Kahramanmaras	Elbistan	Ozcanli	4.90
Turkiye	Kahramanmaras	Ekinozu	Maarif	5.47
Turkiye	Kahramanmaras	Ekinozu	Ekinozu	5.72

Classification and regression are the two main problems in machine learning. What distinguishes these two problems from each other is the solution methods and the dataset in question. Should the data available is an uninterrupted continuous data set, the problem is considered as a regression problem. However, if the existing data is categorical, this problem is a classification problem. The output obtained in the classification problem is labels, while the outputs in the regression problem are numerical values. Classification algorithms; Logistic Regression, Support Vector Machines, Decision Trees, Bayesian, Random Forest, Gradient Boosting, Neural Networks. The algorithms used in the regression problem are linear regression, support vector regression, lasso, elastic net, random forest, decision trees, gradient boosting regression and neural networks. As can be seen, although the methods applied are similar, the versions differ according to the problems. The purpose of the classification problem is to increase the rate at which the model makes an accurate class prediction. The goal of the regression problem is to minimize the error between the actual values and the predicted value. For this reason, evaluation in regression problems is made with metrics such as R^2 score, mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE)[39].

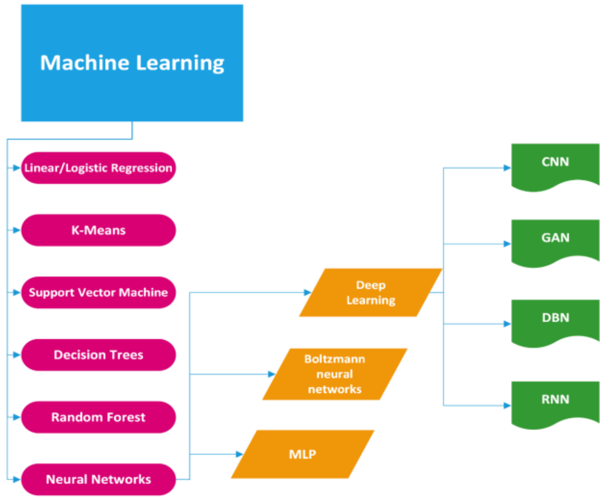


Figure 6. Methods of machine learning

The methods given in Figure 6 aim to model the relationship between an input variable and a target output variable. The purpose of regression models is to predict a continuous output after appropriate modeling. One of the most well-known models is linear regression. Looking at linear regression, it aims to directly reflect the relationship between the independent variable and the dependent variable. It is quick and easy to apply. But in general, it gives good results between variables that have a linear relationship. It is weak in modeling complex relationships. In polynomial regression [43], curves are obtained by adding polynomial terms of independent variables. These curves aim to reflect the relationship between dependent and independent variables. Support vector machines generalize by modeling data that lies between hyperplanes. In decision tree and random forest methods [44], it divides data into sections through simple rules. It aims to obtain stronger results by combining the results of the separated sections in the random forest method. Methods such as Gradient Boosting Regression [45], Neural Networks, K-Nearest Neighbors exhibit different approaches to minimize the distance between data. There are some things to consider when choosing one of these methods. The appropriate regression methods should be determined by considering the amount of data and the relevant dataset, the complexity of the model and the specific conditions of the features to be used in the study. Choices made without taking these situations into account will cause the results to be incompatible and incorrect. In contrast, models that seem to work in harmony do not give accurate results due to overfitting. Although the results obtained look good, overfitting is limited to the accuracy of the model and the dataset used.

Figure 7. Basic concept demonstration in machine learning

2.3. Relationship Between Damaged Structures and Earthquake Records

As outlined in this study, regression methods, random forest and gradient boosting methods were applied to the models. The random forest method is based on the decision tree method. The segmented data is trained on different subsets of the dataset. Estimates are made by averaging all tree estimates. As the name suggests, in this method, sub-datasets are created based on random samples in the training data set. Because the segmented data is trained with different data sets, diversity increases, and this randomness makes the model resistant to overfitting. In a nutshell, each decision tree makes an independent prediction, and these predictions are averaged for regression. Therefore, it is suitable for use in complex datasets due to its ability to overcome complexity in large data sets more easily and its overfitting resistance. Conversely, when looking at the gradient boosting regression method, it focuses on each tree correcting the errors of the previous model by creating decision trees, that is, weak estimators. In contrast to the random forest method, the final estimate is considered as the weighted sum of all decision trees. Due to its gradual approach to minimizing errors, it has high applicability in complex data sets. As part of this research, the damage data obtained from the field studies carried out after the Kahramanmaras earthquakes, the acceleration values related to the earthquake from DEMA's station network and the distances to the city centers will be evaluated. As it is known in the Kahramanmaras earthquakes, 11 cities in Turkiye were directly affected and structures were damaged. Table 3 gives the distribution of damage levels according to cities obtained from the damage assessment reports made by the Ministry of Environment, Urbanization and Climate Change in the field [46]. Buildings are classified according to 6 different damage levels: No Damage, Low Damaged, Medium Damaged, Heavy Damaged, Requiring Urgent Demolition and Collapsed. Until the date of obtaining this data, a total of 38330 buildings collapsed due to the earthquake. Moreover, Table 4 gives the PGA values taken from DEMA's data stations after the Kahramanmaras earthquakes and the distance of the stations to the city center. Table 4 also shows the city where the stations are located and the station code. Together with the information given in Table 3 and Table 4, the relationship between damage levels and PGA values produced by the earthquake will be evaluated over distance with artificial intelligence methods. In the study, PGA values obtained from the stations and the distance to the city center will be used as independent variables. Looking at the station data, since there are PGA values for two different directions, east-west and north-south, the PGA value, which is larger than these two directions, was used as the PGA value. The other independent variable, the distance parameter, is the distance of the stations to the city center. To be used in the study, the distance between the coordinates of the stations and the coordinates of 11 city centers affected by the earthquake was calculated. This variable is given as the calculated distance in Table 4. The dependent variables are the damage levels obtained from the damage assessment results of the cities. The damaged building data given in Table 3 was used as output data in models where PGA and calculated distance expressions obtained from earthquake stations were used as input data. In summary, how does the PGA value produced by the earthquake and the distance to the study area change the damage level after the earthquake? This is the general concept of the study. The change in damage levels was examined by evaluating the increase or decrease of PGA and the approach or decrease of the distance together.

When the data from the study were assessed, it was determined that the problem was a regression problem. Therefore, regression methods were used. However, since there is no study to examine the relationship between these direct dependent and independent variables, multiple regression models will be used. In this study, input and output data are multiple variables. Maximum PGA and calculated distance data to be used as input data and damage levels will be evaluated through multiple regression models. Figure 8 shows the stages of applying the multiple regression model with different variables.

Table 3. Distribution of building damage in earthquake-affected cities according to damage levels [46]

Cities/Damage States	Number of Building	No Damage	Low Damage	Medium Damage	Heavy Damage	Urgently Demolished	Collapsed
Adana	324345	276691	39541	5118	2923	37	35
Adiyaman	110354	38666	38576	4629	20201	2329	5953
Diyarbakir	183730	129986	45602	3355	4708	59	20
Elazig	27760	11767	7945	506	7441	48	53
Gaziantep	282693	188639	68429	5524	14047	1994	4060
Hatay	342531	140337	103549	12874	64283	8038	13450
Kahramanmaras	225230	93168	79027	5987	35229	4423	7396
Kilis	34346	20188	11191	486	1867	151	463
Malatya	155204	60825	48690	2783	36046	1810	5050
Osmaniye	133992	87674	35006	1094	9010	530	678
Sanliurfa	321065	195565	112690	3192	7706	740	1172
Total	2141250	1243506	590246	45548	203461	20159	38330

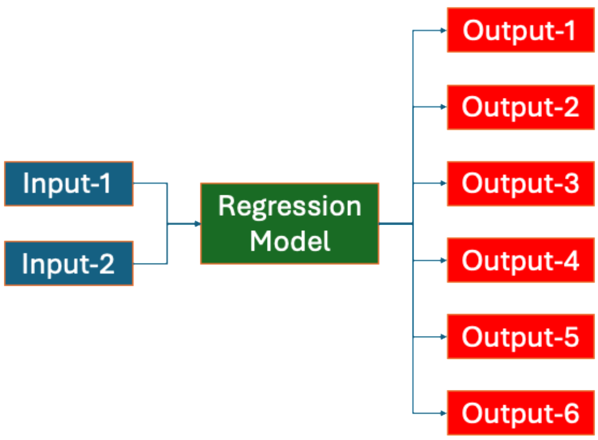


Figure 8. Machine learning process of multiple dependent and independent variables

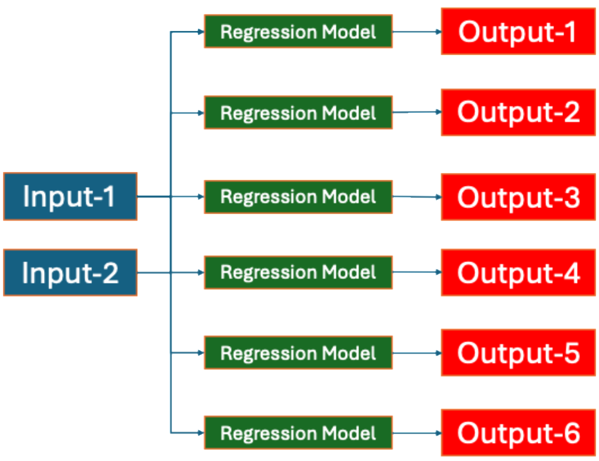


Figure 9. Machine learning process using multiple regression model within the scope of the study

Figure 8 illustrates the stages of machine learning for multiple dependent and independent variables. However, the multiple regression model to be applied within the scope of the study has a different approach from this process. Since a single regression model and multivariate status will not give appropriate results in the study, while multiple regression is applied in this study, model approaches also diversify. As can be seen in Figure 9, in the machine learning model used within the scope of the study, more than one independent variable aims to predict different dependent variables by training with different regression models. Even though the damage levels directly reflect the degree of damage, since the building entering each damage level does not enter the other damage levels, it reveals the necessity of evaluating the dependent variables through a separate regression model with the independent variables. This situation can be given as an example for a clearer understanding. When the distance calculated with PGA is evaluated directly through a single regression model, it will not be able to classify between damage levels. In other words, the PGA value and the calculated distance variables cannot determine the degree of damage. Whether it is a slightly damaged structure, or a heavily damaged structure is independent of PGA and distance parameters. However, it is thought that there is a relationship between the obtained PGA value, and the distance value calculated with the number of damaged structures obtained because of damage detection. Therefore, it is not intended to estimate the level of damage to the damaged structure. The aim of this study is to evaluate the relationship between the number of structures belonging to the damage levels and the PGA value of the earthquake and the distance to the area where the damage occurred. This is the purpose of using different regression models for different levels of damage. Multiple variation, which is generally accepted, is differentiated in this study. The regression analyses and results obtained according to this approach are given in the following sections of the study. In the study, more than one regression method was tried, and the most appropriate regression model was determined for the relevant damage level. The regression analyses performed within the scope of the study were carried out using the code prepared on Phyton. The libraries and models used in the preparation of the code are as follows [47]:

- import pandas as pd
- import numpy as np
- import matplotlib.pyplot as plt
- import seaborn as sns
- from sklearn.model_selection import train_test_split
- from sklearn.preprocessing import PolynomialFeatures
- from sklearn.linear_model import LinearRegression
- from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
- from sklearn.tree import DecisionTreeRegressor
- from sklearn.metrics import mean_squared_error, r2_score
- from sklearn.preprocessing import StandardScaler

The stations used within the scope of the study include stations in cities affected by the earthquake that took records of 7.6 and 7.7 M_w magnitudes. The data were taken for two earthquakes and are given together in Table 4. Therefore, Table 4 contains values for two different stations with the same name. This situation is broken down according to the earthquakes given in the last column.

Table 4. Acceleration values from the stations and the distance of the stations to the city center

Code*	Province	PGA_NS** (cm/s^2)	PGA_EW** (cm/s^2)	MaksPGA** (cm/s^2)	Calculated Distance(km)	Earthquake***
0122	Adana	57.34304	52.33138	57.34304	258.024439	Pazarcik
0125	Adana	128.551	83.12265	128.551	217.6078245	Pazarcik
0127	Adana	54.98741	50.81184	54.98741	294.1709142	Pazarcik
0130	Adana	81.09731	68.23101	81.09731	245.943058	Pazarcik
0120	Adana	112.4618	115.9806	115.9806	194.8052095	Pazarcik
0129	Adana	49.91854	42.1567	49.91854	334.647704	Pazarcik
0123	Adana	41.42038	39.65318	41.42038	239.3306136	Pazarcik
0118	Adana	50.09951	38.23767	50.09951	243.5799995	Pazarcik
0119	Adana	43.46264	47.31035	47.31035	200.9083487	Pazarcik
0128	Adana	11.68445	14.22345	14.22345	276.3505367	Pazarcik
0124	Adana	8.571593	8.76492	8.76492	300.2768086	Pazarcik
0129	Adana	154.462	172.1792	172.1792	334.647704	Pazarcik
0127	Adana	56.0935	62.72347	62.72347	294.1709142	Pazarcik
0122	Adana	48.44631	67.45694	67.45694	258.024439	Elbistan
0130	Adana	79.31747	79.89938	79.89938	245.943058	Elbistan
0125	Adana	70.09405	50.6768	70.09405	217.6078245	Elbistan
0120	Adana	20.67546	25.01525	25.01525	194.8052095	Elbistan
0118	Adana	27.48687	24.4715	27.48687	243.5799995	Elbistan
0123	Adana	17.93101	27.66621	27.66621	239.3306136	Elbistan
0128	Adana	19.07407	19.72894	19.72894	276.3505367	Elbistan
0124	Adana	15.07867	20.11232	20.11232	300.2768086	Elbistan
0119	Adana	10.10975	11.75999	11.75999	200.9083487	Elbistan
0208	Adiyaman	30.19949	14.00124	30.19949	55.44248446	Pazarcik
0213	Adiyaman	242.2791	171.6946	242.2791	55.43172643	Pazarcik
0201	Adiyaman	474.1206	879.9495	879.9495	72.42297102	Pazarcik
0210	Adiyaman	65.90985	61.3746	65.90985	72.93748504	Pazarcik
0214	Adiyaman	61.67553	54.38109	61.67553	49.00909816	Pazarcik
0213	Adiyaman	121.297	126.6186	126.6186	55.43172643	Elbistan
0205	Adiyaman	44.87774	54.6579	54.6579	92.13885028	Elbistan
2107	Diyarbakir	74.75684	112.2655	112.2655	269.106092	Pazarcik
2104	Diyarbakir	72.83684	116.4655	116.4655	270.9526975	Pazarcik
2101	Diyarbakir	77.07944	71.42427	77.07944	323.8146848	Pazarcik
2103	Diyarbakir	53.7313	43.12138	53.7313	328.5451819	Pazarcik
2106	Diyarbakir	72.30118	61.69453	72.30118	307.0785457	Pazarcik
2108	Diyarbakir	20.65492	20.26613	20.65492	329.2730899	Pazarcik
2107	Diyarbakir	28.64484	47.6136	47.6136	269.106092	Pazarcik
2104	Diyarbakir	27.52314	21.22324	27.52314	270.9526975	Elbistan
2101	Diyarbakir	25.76558	21.59068	25.76558	323.8146848	Elbistan
2103	Diyarbakir	19.86305	23.82505	23.82505	328.5451819	Elbistan
2106	Diyarbakir	9.178418	8.270238	9.178418	307.0785457	Elbistan
2108	Diyarbakir	7.947267	5.894371	7.947267	329.2730899	Elbistan
2310	Elazig	60.45616	51.19656	60.45616	72.68669944	Pazarcik
2309	Elazig	38.26139	35.33888	38.26139	46.62227564	Pazarcik
2304	Elazig	32.68223	49.40283	49.40283	115.9637753	Pazarcik
2307	Elazig	33.62553	38.10754	38.10754	122.5618978	Pazarcik
2305	Elazig	58.1771	53.67463	58.1771	136.4866903	Pazarcik
2310	Elazig	41.466	55.23522	55.23522	72.68669944	Elbistan

2309	Elazig	62.65357	38.65393	62.65357	46.62227564	Elbistan
2308	Elazig	69.7981	48.52414	69.7981	101.0596715	Elbistan
2302	Elazig	12.35801	16.33764	16.33764	125.7133286	Elbistan
2307	Elazig	12.95982	15.56112	15.56112	122.5618978	Elbistan
2305	Elazig	5.356134	5.302636	5.356134	136.4866903	Elbistan
2712	Gaziantep	555.5879	592.3544	592.3544	36.92711957	Pazarcik
2703	Gaziantep	156.6342	165.0642	165.0642	44.01870845	Pazarcik
2709	Gaziantep	154.0308	127.0069	154.0308	45.0806162	Pazarcik
2711	Gaziantep	142.6439	119.6102	142.6439	44.24795116	Pazarcik
2718	Gaziantep	654.4308	630.312	654.4308	57.1246742	Pazarcik
2707	Gaziantep	98.64927	89.27893	98.64927	66.62504841	Pazarcik
2704	Gaziantep	102.235	160.6396	160.6396	77.31523048	Pazarcik
2703	Gaziantep	93.68232	63.4492	93.68232	44.01870845	Elbistan
2704	Gaziantep	34.59009	63.62961	63.62961	77.31523048	Elbistan
2718	Gaziantep	34.47282	50.5759	50.5759	57.1246742	Elbistan
3143	Hatay	381.0936	351.3809	381.0936	84.01575389	Pazarcik
3138	Hatay	888.7299	746.6645	888.7299	77.71924355	Pazarcik
3144	Hatay	611.2695	763.3625	763.3625	72.16215953	Pazarcik
3137	Hatay	428.373	670.1654	670.1654	65.65887763	Pazarcik
3134	Hatay	246.1068	203.9094	246.1068	75.35026556	Pazarcik
3145	Hatay	591.8801	692.2899	692.2899	58.09676004	Pazarcik
3139	Hatay	577.1307	504.8208	577.1307	51.91871923	Pazarcik
3116	Hatay	164.2769	168.8629	168.8629	51.84022805	Pazarcik
3142	Hatay	651.6892	739.2937	739.2937	41.45588121	Pazarcik
3112	Hatay	171.8594	83.63697	171.8594	48.92942351	Pazarcik
3115	Hatay	286.7226	241.5	286.7226	44.18495112	Pazarcik
3146	Hatay	483.8456	346.9315	483.8456	37.96742306	Pazarcik
3133	Hatay	221.4053	147.2227	221.4053	35.08977603	Pazarcik
3141	Hatay	961.1165	868.8192	961.1165	24.8151231	Pazarcik
3124	Hatay	572.6323	638.3214	638.3214	10.17359091	Pazarcik
3125	Hatay	822.616	1121.948	1121.948	11.50918597	Pazarcik
3135	Hatay	740.9707	1372.071	1372.071	40.44057724	Pazarcik
3123	Hatay	655.5713	593.9404	655.5713	8.00447787	Pazarcik
3132	Hatay	515.3094	514.6342	515.3094	6.803978332	Pazarcik
3126	Hatay	1178.116	999.3831	1178.116	9.612220001	Pazarcik
3131	Hatay	363.0329	366.0505	366.0505	5.644666684	Pazarcik
3129	Hatay	1351.5	1198.743	1351.5	7.465738456	Pazarcik
3136	Hatay	534.2245	401.9692	534.2245	5.686417693	Pazarcik
3140	Hatay	194.6867	218.7093	218.7093	23.72669254	Pazarcik
3147	Hatay	56.44854	47.51172	56.44854	30.11693011	Pazarcik
3143	Hatay	42.89935	39.84327	42.89935	84.01575389	Elbistan
3138	Hatay	49.2678	68.71472	68.71472	77.71924355	Elbistan
3144	Hatay	59.115	78.10838	78.10838	72.16215953	Elbistan
3134	Hatay	30.57111	40.03998	40.03998	75.35026556	Elbistan
3137	Hatay	23.03817	25.60015	25.60015	65.65887763	Elbistan
3139	Hatay	43.35793	57.54239	57.54239	51.91871923	Elbistan
3116	Hatay	17.08	19.196	19.196	51.84022805	Elbistan
3142	Hatay	10.38088	21.28696	21.28696	41.45588121	Elbistan
3115	Hatay	25.77447	27.45891	27.45891	44.18495112	Elbistan
3146	Hatay	17.67443	18.28782	18.28782	37.96742306	Elbistan
3141	Hatay	25.71274	23.11699	25.71274	24.8151231	Elbistan
3133	Hatay	19.90386	18.1046	19.90386	35.08977603	Elbistan
3135	Hatay	18.14687	15.50154	18.14687	40.44057724	Elbistan
3124	Hatay	21.75652	32.18029	32.18029	10.17359091	Elbistan
3125	Hatay	25.62747	21.04758	25.62747	11.50918597	Elbistan
3123	Hatay	23.0113	24.32187	24.32187	8.00447787	Elbistan
3132	Hatay	17.45727	22.97341	22.97341	6.803978332	Elbistan
3129	Hatay	22.78477	26.62058	26.62058	7.465738456	Elbistan
3136	Hatay	18.60377	22.79383	22.79383	5.686417693	Elbistan
3140	Hatay	29.10271	30.20007	30.20007	23.72669254	Elbistan
3147	Hatay	5.370752	7.258578	7.258578	30.11693011	Elbistan
4615	Kahramanmaras	584.6534	556.6476	584.6534	64.0362702	Pazarcik
NAR	Kahramanmaras	784.5689	619.7074	784.5689	63.45176373	Pazarcik
4616	Kahramanmaras	610.3447	428.5635	610.3447	82.31432921	Pazarcik
4630	Kahramanmaras	178.5622	124.0367	178.5622	82.71060379	Pazarcik
4629	Kahramanmaras	338.9347	248.1954	338.9347	80.70715678	Pazarcik
4632	Kahramanmaras	359.4571	299.2487	359.4571	80.26537483	Pazarcik
4625	Kahramanmaras	447.0017	466.3326	466.3326	85.83923916	Pazarcik
4624	Kahramanmaras	357.252	310.8946	357.252	89.25660956	Pazarcik
4614	Kahramanmaras	2165.615	2178.72	2178.72	66.38797756	Pazarcik
4626	Kahramanmaras	108.8081	223.0931	223.0931	92.71471793	Pazarcik
4621	Kahramanmaras	363.8016	295.5592	363.8016	93.49970259	Pazarcik
4620	Kahramanmaras	300.4047	320.9304	320.9304	94.53707107	Pazarcik
4619	Kahramanmaras	302.0343	194.7355	302.0343	96.50513536	Pazarcik
4618	Kahramanmaras	125.6644	159.4222	159.4222	97.25026352	Pazarcik

4617	Kahramanmaras	145.3257	115.1562	145.3257	98.49700167	Pazarcik
4611	Kahramanmaras	349.7206	321.1143	349.7206	94.30820604	Pazarcik
4613	Kahramanmaras	146.9338	153.5201	153.5201	129.5609468	Pazarcik
4631	Kahramanmaras	22.20652	19.33552	22.20652	115.7158986	Pazarcik
4612	Kahramanmaras	140.97	122.222	140.97	155.4221265	Pazarcik
4628	Kahramanmaras	91.09563	82.54909	91.09563	156.6641583	Pazarcik
4631	Kahramanmaras	337.3846	388.6079	388.6079	115.7158986	Pazarcik
4611	Kahramanmaras	194.4007	139.037	194.4007	94.30820604	Pazarcik
4620	Kahramanmaras	66.82375	81.33087	81.33087	94.53707107	Pazarcik
4625	Kahramanmaras	73.45961	50.68318	73.45961	85.83923916	Pazarcik
4617	Kahramanmaras	55.97399	82.69461	82.69461	98.49700167	Pazarcik
4612	Kahramanmaras	635.4467	523.2124	635.4467	155.4221265	Pazarcik
4614	Kahramanmaras	160.8168	206.0473	206.0473	66.38797756	Pazarcik
4624	Kahramanmaras	65.00184	79.7458	79.7458	89.25660956	Pazarcik
NAR	Kahramanmaras	126.5214	110.4217	126.5214	63.45176373	Elbistan
4615	Kahramanmaras	44.47337	73.751	73.751	64.0362702	Elbistan
4616	Kahramanmaras	57.54619	53.50307	57.54619	82.31432921	Elbistan
4613	Kahramanmaras	80.61347	78.2478	80.61347	129.5609468	Elbistan
7901	Kilis	53.11445	16.55168	53.11445	57.27891365	Pazarcik
7901	Kilis	50.9099	49.81428	50.9099	57.27891365	Elbistan
4408	Malatya	100.0891	137.1811	137.1811	47.30044778	Pazarcik
4406	Malatya	108.7379	131.3439	131.3439	32.95217427	Pazarcik
4409	Malatya	38.00886	28.49102	38.00886	79.5883046	Pazarcik
4412	Malatya	63.57863	68.89719	68.89719	34.31656851	Pazarcik
4410	Malatya	33.70423	45.49616	45.49616	84.48601049	Pazarcik
4405	Malatya	91.11829	126.4967	126.4967	65.47416203	Pazarcik
4404	Malatya	136.2437	137.4162	137.4162	47.68172015	Pazarcik
4414	Malatya	106.6179	163.844	163.844	36.62772173	Pazarcik
4407	Malatya	43.36136	33.08422	43.36136	52.09549835	Pazarcik
4413	Malatya	13.15253	10.18892	13.15253	60.14259075	Pazarcik
4409	Malatya	287.0381	218.0397	287.0381	79.5883046	Elbistan
4406	Malatya	467.2015	409.3123	467.2015	32.95217427	Elbistan
4410	Malatya	112.0973	127.2469	127.2469	84.48601049	Elbistan
4412	Malatya	159.0325	126.3764	159.0325	34.31656851	Elbistan
4405	Malatya	155.4112	158.0522	158.0522	65.47416203	Elbistan
4414	Malatya	81.40928	63.00617	81.40928	36.62772173	Elbistan
4404	Malatya	45.36233	48.54014	48.54014	47.68172015	Elbistan
4413	Malatya	36.78609	50.93417	50.93417	60.14259075	Elbistan
8002	Osmaniye	242.9514	202.8933	242.9514	101.8297877	Pazarcik
8003	Osmaniye	141.5669	185.7379	185.7379	111.5543084	Pazarcik
8004	Osmaniye	168.4261	181.8594	181.8594	144.7633524	Pazarcik
8002	Osmaniye	65.87371	45.50682	65.87371	101.8297877	Elbistan
8003	Osmaniye	48.69694	66.60214	66.60214	111.5543084	Elbistan
6304	Sanliurfa	210.8972	238.2282	238.2282	198.0535956	Pazarcik
6305	Sanliurfa	126.6591	104.0897	126.6591	230.8136543	Pazarcik
6306	Sanliurfa	65.89738	55.98942	65.89738	278.1915873	Pazarcik
6303	Sanliurfa	117.4226	114.4394	117.4226	221.9514892	Pazarcik
6302	Sanliurfa	59.94751	51.16346	59.94751	285.6036786	Pazarcik
6303	Sanliurfa	29.4311	21.68857	29.4311	221.9514892	Elbistan
6306	Sanliurfa	35.99724	27.16897	35.99724	278.1915873	Elbistan
6302	Sanliurfa	27.00732	19.33995	27.00732	285.6036786	Elbistan

*: DEMA Station Number for the cities.
**: Direction of recorded station.
***: Epicenter of earthquake.

3. RESULTS

As part of this research, the data were evaluated by multi-regression method. Linear regression, polynomial regression, gradient boosting and random forest regression models were used as regression methods. The models used are shown in Table 5. These regression models were analyzed with the code prepared on Phyton software[47].

Table 5. Model used in the study

Models
Linear Regression
Polynomial Regression
Gradient Boosting
Random Forest

All the regression models used were used for each damage level. Among the models, the model that is the most compatible, that is, the one that gives the highest R^2 value, was determined. When the models with the highest agreement for damage levels are examined, it is seen that the highest agreement is achieved in 5 out of 6 damage levels in the random forest regression model. The most compatible model for the heavy damage level was gradient boosting regression. The distribution of the selected regression models according to their damage levels is given in Table 6.

Table 6. Regression models selected for damage levels

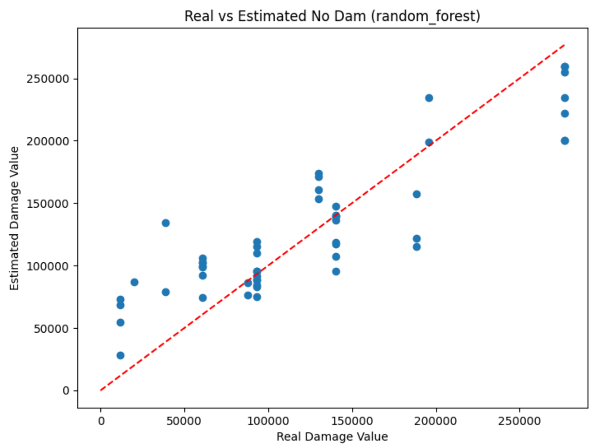
Damage Level	Best Fitting Model
No Damage	Random Forest
Low Damage	Random Forest
Medium Damage	Random Forest
Heavy Damage	Gradient Boosting
Urgently Demolished	Random Forest
Collapsed	Random Forest

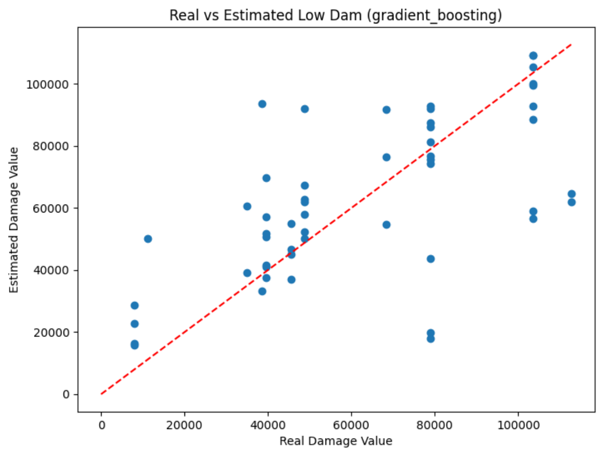
Table 7 shows the distribution of R^2 scores for damage levels and the predictive power results based on interpretation. R^2 values reflect the relationship between damage levels and PGA values of stations and station distance. When the data are examined, it is seen that there is a good level of harmony between the earthquake parameters of the undamaged and heavily damaged structures. The R^2 scores of no damage and heavy damage structures are 0.75 and 0.71, respectively. Damage levels, including low damage, medium damage, urgent demolition and collapsed buildings, are 50% compatible. It is expected that the R^2 scores of moderately damaged structures will be low. The determination of moderately damaged structures in the damage assessments made in the field is a subjective situation based on interpretation. The level of knowledge of the technical staff that determines the level of damage in the field, the approach of the damage assessment forms, and the inability to determine the damage of the structure make the concept of a moderately damaged building variable. In field observations, one expert's low damage can be detected as moderate damage by another expert. Likewise, this situation exists between medium damage and heavy damage. Therefore, high compliance in the prediction of moderate damage is not technically expected. Predictive power values based on interpretation were determined by the researcher. This evaluation shows the level of compliance of the number of structures belonging to the damage levels and the PGA and distance parameters of the earthquake. Even if the values between 0.4-0.5 are considered low in regression models, it is thought that higher predictive power will be achieved by increasing the number of data and the number of stations in this study.

Table 7. R^2 scores of damage levels and interpretive predictive power

Damage Levels	R^2 Score	Predictive Power
No Damage	0.75	Good
Low Damage	0.45	Medium
Medium Damage	0.46	Medium
Heavy Damage	0.71	Good
Urgently Demolished	0.56	Medium
Collapsed	0.61	Medium

With this study, it is seen that the number and quality of data will increase the consistency between them. The levels of damaged buildings examined within the scope of the study are the total numbers in the cities where the buildings damaged by the earthquake are located. Therefore, the calculated distances are a single value calculated based on the coordinates of the city center. If the distribution of the damaged structure is coordinate-oriented, the relationship between the damage level and the station data will be seen more clearly, as the distance of the evaluated station to the structure entering the relevant damage level can be better determined. Studies show that the effects of the structures affected by earthquakes, the damage mechanism and the causes of damage can be better reflected with the data at the stations close to the building. This shows that there is a relationship between damage levels and the location of the station and the earthquake parameters it takes measurements. This study shows that the relationship described can be revealed by regression models. The fact that the number of less damaged and heavily damaged structures is higher in number compared to other damage levels increases the success rate of compliance. When we separated the R^2 scores obtained in the study as damaged and undamaged, the R^2 scores were calculated as 0.63 and 0.6, respectively. These ratios show that if the number of data is increased and the coordinates of the damaged structure data used in the study are determined, they will reflect the relationship very well.





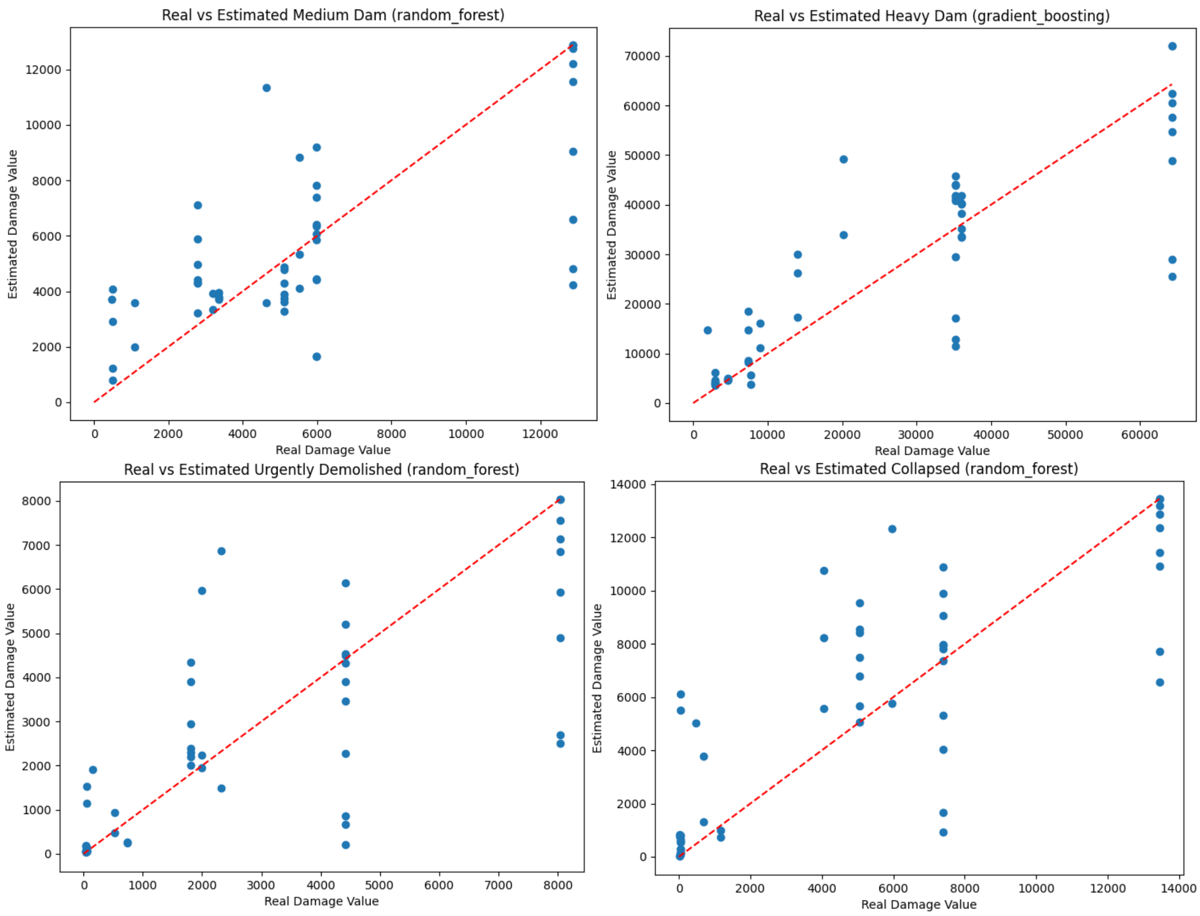


Figure 10. Damaged building values and predicted damaged building values based on damage levels

Figure 10 shows the distribution of actual and estimated values for damage levels. As can be seen at damage levels with high R^2 scores, scattering is lower. Damage levels of No Damage and Heavy Damaged structures are concentrated close to the trend line. In addition, when the values were examined, the actual and estimated values were collected in some value ranges due to the evaluation of the city-based damage distribution. Figure 10 supports the R^2 scores in Table 7. In such studies, the presence of coordinated information of damaged structures shows that it will increase the prediction power.

Table 8. Comparison of Random Forests and Gradient Boosting Regression Models [24,25]

Characteristics	**Random Forest**	**Gradient Boosting**
Model Structure	Trees trained in parallel	Sequentially trained trees
Performance	It's faster	Provides higher accuracy
Overfitting Risks	Less	Higher
Forecast Merge	Average of the forecasts of the trees	Focused on error correction
Hyperparameter Setting	Less sensitive	Requires more careful tuning

Table 8 shows the comparison between the two regression models that showed the highest agreement with the damage levels in the study. When the characteristics of random forest and gradient boosting regression models are examined, the model structure, model performance, overfitting risk, and hyperparameter settings used to combine predictions differ. As used in the study, random forest is more suitable for solving regression problems of complex data[48,49]. As can be seen in Table 6, 5 of the 6 damage levels used in the study give the best fit in the random forest model. The fact that the predictions in the decision trees are averaged, fast, have a low risk of overfitting, and have low precision in hyperparameter settings provides an advantage in the evaluation of complex data sets. Within the scope of this study, the random forest regression model showed high compatibility.

4. DISCUSSION AND CONCLUSION

Verification and evaluation of the damage assessment made after earthquakes will guide the evaluations to be made in future earthquakes. As outlined in this study, the relationship between the damage levels obtained because of the damage assessments made after the February 6, 2023 Kahramanmaras earthquakes and the maximum ground acceleration (PGA) of DEMA stations recorded in the Kahramanmaras earthquakes and their distances to the cities affected by the earthquake were examined. The results obtained after the evaluations are given below.

- In the study, unlike the literature, multiple input and multiple output parameters were evaluated in the multi regression model.
- A separate regression model was used for each output data in the regression problem in the study. In this way, the multiple regression approach is completely different compared to other studies.

- After the regression analysis, R^2 scores between damage levels and PGA-distance parameters and adaptive strength values based on interpretation were obtained. When the results are examined, there is a high agreement between the number of buildings that enter the heavy damage and no damage levels, and the values measured by the stations. The R^2 scores for the "No damage" and "Heavy damage" damage levels are 0.75 and 0.71, respectively. When the evaluation was reduced from 6 damage levels to 2 damage levels, damaged and undamaged, the R^2 scores were obtained as 0.63 and 0.6, respectively. This indicates that there is a sufficient level of harmony between the input and output parameters. However, it has been concluded that in case of higher success, the data set should be expanded, and the coordinate data of the damaged structures should be obtained separately.
- In regression analysis, more than one regression model (linear regression, polynomial regression, random forest and gradient boosting) was used to determine the best fit models among these models. Considering the results obtained according to the damage levels, the regression models that showed the best fit were the gradient boosting and random forest models.
- It is seen that the random forest regression model is a preferable model in terms of being the best compatible model in 5 out of 6 damage levels, and in complex analyzes such as damage-earthquake parameter relationship, hyperparameter sensitivity is low, fast and minimizes the risk of overfitting.
- It has been seen by the results of the study that low compliance levels of damage can be increased by increasing the number of data and detailing the level of information about the data.
- In such multiple regression analysis studies, all input data are analyzed and evaluated with a single regression model. In this study, the model that fits best was used for each input data. While this can be used in cases with independent input data as in this study, it would not be correct to use it on input data that is directly related to each other. Therefore, in the continuation of these studies, researchers can examine the relationship between damage and earthquake characteristics with a single regression model with high data sets.

REFERENCES

[1] Nemutlu ÖF, Balun B, Sari A. Damage assessment of buildings after 24 January 2020 Elazig-Sivrice earthquake. Earthquakes and Structures. 25 Mart 2021;20(3):325-35.

[2] Balun B, Nemutlu OF, Benli A, Sari A. Estimation of probabilistic hazard for Bingol province, Turkey. Earthquakes and Structures. 25 Şubat 2020;18(2):223-31.

[3] Avcil F, Işık E, İzol R, Büyüksaraç A, Arkan E, Arslan MH, vd. Effects of the February 6, 2023, Kahramanmaraş earthquake on structures in Kahramanmaraş city. Nat Hazards. Şubat 2024;120(3):2953-91.

[4] Altunsu E, Güneş O, Öztürk S, Sorosh S, Sarı A, Beeson ST. Investigating the structural damage in Hatay province after Kahramanmaraş-Türkiye earthquake sequences. Engineering Failure Analysis. 2024;157:107857.

[5] Öztürk S, Altunsu E, Güneş O, Sarı A. Investigation of industrial structure performances in the Hatay and Gaziantep provinces during the Türkiye earthquakes on February 6, 2023. Soil Dynamics and Earthquake Engineering. Nisan 2024;179:108569.

[6] Nemutlu ÖF, Balun B, Sarı A. 06 Şubat 2023 Kahramanmaraş Depremleri Kaynaklı Yapısal Hasarların Adıyaman İli Özelinde İncelenmesi. İçinde Konya; 2023.

[7] Hussain E, Kalaycıoğlu S, Milliner CW, Çakir Z. Preconditioning the 2023 Kahramanmaraş (Türkiye) earthquake disaster. Nature Reviews Earth & Environment. 2023;4(5):287-9.

[8] AFAD. 06 Şubat Kahramanmaraş(Pazarcık ve Elbistan) Depremleri Saha Çalışmaları Ön Değerlendirme Raporu. AFAD; 2023.

[9] Şenol Balaban M, Doğulu C, Akdede N, Akoğlu H, Karakayalı O, Yılmaz S, vd. Emergency response, and community impact after February 6, 2023 Kahramanmaraş Pazarcık and Elbistan Earthquakes: reconnaissance findings and observations on affected region in Türkiye. Bulletin of Earthquake Engineering. 01 Şubat 2025;23(3):1053-81.

[10] Altunişik AC, Arslan ME, Kahya V, Aslan B, Sezdirmez T, Dok G, vd. Field Observations and Damage Evaluation in Reinforced Concrete Buildings After the February 6th, 2023, Kahramanmaraş–Türkiye Earthquakes. J Earthquake and Tsunami. 01 Aralık 2023;17(06):2350024.

[11] Yuzbasi J. Post-Earthquake Damage Assessment: Field Observations and Recent Developments with Recommendations from the Kahramanmaraş Earthquakes in Türkiye on February 6th, 2023 (Pazarcık M7.8 and Elbistan M7.6). Journal of Earthquake Engineering. :1-26.

[12] Avğın S, Köse MM, Özbek A. Damage assessment of structural and geotechnical damages in Kahramanmaraş during the February 6, 2023 earthquakes. Engineering Science and Technology, an International Journal. 01 Eylül 2024;57:101811.

[13] Işık E, Hadzima-Nyarko M, Avcil F, Büyüksaraç A, Arkan E, Alkan H, vd. Comparison of Seismic and Structural Parameters of Settlements in the East Anatolian Fault Zone in Light of the 6 February Kahramanmaraş Earthquakes. Infrastructures. 2024;9(12).

[14] Tao D, Cai Y. Study on the Relation Between Ground Motion Parameters and Simulated Earthquake Damage of Simplified Masonry Structures. 2018.

[15] Zhou Q, Sun B. Study on Earthquake Damage Distribution of Multistory Masonry Buildings. The Open Civil Engineering Journal. 2015;9(1):435-41.

[16] Liang H, Li J, Yang Z, Li YQ. Effects of Epicentral Distance and Seismogenic Fault Distance on Seismic Damage of Dams in Wenchuan Earthquake.

Applied Mechanics and Materials. 2013;353-356:2187-90.
[17] Karaşin İB. Comparative Analysis of the 2023 Pazarcık and Elbistan Earthquakes in Diyarbakır. Buildings. 2023;13(10):2474.
[18] Zengin B, Aydin F. The Effect of Material Quality on Buildings Moderately and Heavily Damaged by the Kahramanmaraş Earthquakes. Applied Sciences. 2023;13(19):10668.
[19] Choi H, Sanada Y, KASHIWA H, Watanabe Y, Tanjung J, Jiang H. Seismic Response Estimation Method for Earthquake-damaged RC buildings. Earthquake Engineering & Structural Dynamics. 2016;45(6):999-1018.
[20] Huang HC. Characteristics of earthquake ground motions and the H/V of microtremors in the southwestern part of Taiwan. Earthquake Engineering and Structural Dynamics. 2002;31(10):1815-29.
[21] Bilen A, Özer AB. Regresyon Yöntemlerine Dayalı Suç Tespit Analizi Karşılaştırması Elazığ İli Örneği. Fırat Üniversitesi Mühendislik Bilimleri Dergisi. 2022;34(1):115-21.
[22] Shen Y, Wang L, Jian W, Shang J, Wang X, Ju L, vd. Big-Data and Artificial-Intelligence-Assisted Vault Prediction and EVO-ICL Size Selection for Myopia Correction. British Journal of Ophthalmology. 2021;107(2):201-6.
[23] Suparwito H, Polina AM. Prediction of Tobacco Leave Grades With Ensemble Machine Learning Methods. 2019;1-6.
[24] Liou L, Mostofsky E, Lehman LL, Salia S, Barrera FJ, Ying W, vd. Survival Machine Learning Methods for Mortality Prediction After Heart Transplantation in the Contemporary Era. Plos One. 2025;20(1):e0313600.
[25] Pahno S, Yang J, Kim SS. Use of Machine Learning Algorithms to Predict Subgrade Resilient Modulus. Infrastructures. 2021;6(6):78.
[26] Özçelik STA, Üzen H, Şengür A, Çelebi A. Derin öğrenme ile panoramik radyografi görüntülerinden diş segmentasyonu: UNet, FPN ve PSPNet karşılaştırması. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi. 2024;13(4):1347-54.
[27] Yusufoğlu E, Fırat H, Üzen H, Özçelik STA, Çiçek İB, Şengür A, vd. A Comprehensive CNN Model for Age-Related Macular Degeneration Classification Using OCT: Integrating Inception Modules, SE Blocks, and ConvMixer. Diagnostics. 2024;14(24):2836.
[28] Üzen H. İmalat Sistemlerinde Derin Öğrenme Tabanlı Doku Hata Tespiti [Doktora]. [Fen Bilimleri Enstitüsü]: İnönü Üniversitesi; 2022.
[29] Mangalathu S, Sun H, Nweke CC, Yi Z, Burton HV. Classifying earthquake damage to buildings using machine learning. Earthquake Spectra. 01 Şubat 2020;36(1):183-208.
[30] Xia H, Wu J, Yao J, Zhu H, Gong A, Yang J, vd. A Deep Learning Application for Building Damage Assessment Using Ultra-High-Resolution Remote Sensing Imagery in Turkey Earthquake. International Journal of Disaster Risk Science. 01 Aralık 2023;14(6):947-62.
[31] Nemutlu ÖF. Detection of earthquake damage using pre and past-earthquake satellite data. İçinde Tashkent, Uzbekistan; 2024. s. 256-66.
[32] Aloisio A, Rosso MM, De Leo AM, Fragiacomo M, Basi M. Damage classification after the 2009 L'Aquila earthquake using multinomial logistic regression and neural networks. International Journal of Disaster Risk Reduction. Ekim 2023;96:103959.
[33] Sheibani M, Ou G. The development of Gaussian process regression for effective regional post-earthquake building damage inference. Computer-Aided Civil and Infrastructure Engineering. 01 Mart 2021;36(3):264-88.
[34] Adanur S, Altunişik AC, Bayraktar A, Akköse M. Comparison of near-fault and far-fault ground motion effects on geometrically nonlinear earthquake behavior of suspension bridges. Natural Hazards. 2012;64(1):593-614.
[35] Işık E, Avcil F, Büyüksaraç A, İzol R, Arslan MH, Aksoylu C, vd. Structural damages in masonry buildings in Adıyaman during the Kahramanmaraş (Turkiye) earthquakes (Mw 7.7 and Mw 7.6) on 06 February 2023. Engineering Failure Analysis. 2023;151(May).
[36] Nemutlu ÖF, Sari A, Balun B. 06 Şubat 2023 Kahramanmaraş Depremlerinde (Mw 7.7-Mw 7.6) Meydana Gelen Gerçek Can Kayıpları Ve Yapısal Hasar Değerlerinin Tahmin Edilen Değerler İle Karşılaştırılması. Afyon Kocatepe University Journal of Sciences and Engineering. 27 Ekim 2023;23(5):1222-34.
[37] AFAD. 06 Şubat 2023 Pazarcık-Elibstan Kahramanmaraş(Mw:7.7-Mw:7.6) Depremleri Raporu. AFAD: AFAD; 2023 Haz s. 140.
[38] AFAD. tadas.afad.gov.tr. 2023.
[39] Mahesh B. Machine Learning Algorithms -A Review. C. 9, International Journal of Science and Research (IJSR). 2019.
[40] Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science. 22 Mart 2021;2(3):160.
[41] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, vd. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data. 31 Mart 2021;8(1):53.
[42] Mishra C, Gupta D. Deep Machine Learning and Neural Networks: An Overview. IAES International Journal of Artificial Intelligence (IJ-AI). 01 Haziran 2017;6:66.
[43] Ostertagova E. Modelling Using Polynomial Regression. Procedia Engineering. 31 Aralık 2012;48:500-6.
[44] Breiman L. Random Forests. Machine Learning. 01 Ekim 2001;45(1):5-32.
[45] Zemel RS, Pitassi T. A Gradient-Based Boosting Algorithm for Regression Problems.
[46] Çevre Şehircilik ve İklim Değişikliği Bakanlığı. csb.gov.tr. Basın Bülteni. 2023.
[47] Python [Internet]. Python.org. 2024 [a.yer 21 Aralık 2024]. Erişim adresi: https://www.python.org/

[48] Nadkarni SB, Vijay GS, Kamath RC. Comparative Study of Random Forest and Gradient Boosting Algorithms to Predict Airfoil Self-Noise. Engineering Proceedings. 2023;59(1).
[49] Sousa M, Sant'Ana R, Fernandes R, Duarte J, Aploinário J, Thomä R. Comparison of Random Forest and Gradient Boosting Fingerprints to Enhance an Outdoor Radio-frequency Localization System. 2020.