Prediction of earthquake magnitude and seismic vulnerability mapping using artificial intelligence techniques: a case study of Turkey

Saptadeep Biswas
NIT Agartala: National Institute of Technology Agartala
DHRUV KUMAR
NIT Agartala: National Institute of Technology Agartala
Uttam Kumar Bera (✉ bera_uttam@yahoo.co.in)
NIT Agartala: National Institute of Technology Agartala

Research Article

Keywords: Earthquake hazard, Geo-risk, AI, ML, ANN, GIS, Seismic vulnerability mapping, Turkey

Posted Date: May 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2863887/v1

Prediction of earthquake magnitude and seismic vulnerability mapping using artificial intelligence techniques: a case study of Turkey

Saptadeep Biswas[a], Dhruv Kumar[b], Uttam Kumar Bera[a]

[a]*Department of Mathematics, National Institute of Technology Agartala, Tripura, India.*
[b]*Department of Physics, National Institute of Technology Agartala, Tripura, India.*

Abstract

Earthquake threats can result in fatalities, property destruction, and other cascading effects. Since it is nearly impossible to prevent earthquakes, anticipating the location of future earthquakes and figuring out their likelihood could be very helpful in reducing the seismic threat. In this work, seismic hazard prediction is executed to forecast adverse results using a range of potential artificial intelligence (AI) techniques, including ML and ANN. In the case study, we have looked at Turkey, which was recently and badly damaged by two earthquakes in February 2023. To predict earthquake magnitude, this study used a variety of regression algorithms, including Decision Tree Regressor, Extra-Trees Regressor, Random Forest Regressor, Bayesian Ridge Regressor, and advanced gradient boosting decision tree (GBDT) algorithms such as XGBoost, LightGBM, and CatBoost, as well as three artificial neural networks (ANN). The predicted magnitude and risk zone of an earthquake are mapped using a geographic information system (GIS), and the maps performed well in terms of prediction. The generated maps is showing the expected earthquake risk based on historical data using the statistical computations. The ANN models perform exceptionally well, with R2 scores of 0.99 and 0.98 for training and case study data, respectively, and low values for MSE, MAE, and RMSE. ML models have demonstrated an exceptional ability to properly generalize from a single dataset, which implies they can accurately anticipates results for new and untested data. The results would be helpful to many local emergency preparedness and infrastructure planning organizations.

Keywords: Earthquake hazard, Geo-risk, AI, ML, ANN, GIS, Seismic vulnerability mapping, Turkey

---

1. Introduction

The type of incoming seismic waves, geography, rock-mass qualities, groundwater conditions, and other factors regulate how much the magnitude and depth of an earthquake affect how much the landscape is disturbed. Although the seismic motion is an extremely fast and instantaneous disturbance, its effects eventually fade, and over time, earthquakes can have an impact on how mountain landscapes change. A major mainland earthquake actually triggers a series of events. Earthquakes cause landslides, fissures, avalanches, fires, tsunamis, and earth-shaking, breaking, and liquefaction. Liquefaction beneath a structure's surface during an earthquake can cause substantial harm. These processes represent a risk when they interact with the human world because they each carry a different level of risk. Recent advancements in geo-technical domains and geographic decision support systems, along with an increase in the need for spatial data, have sparked the emergence of new opportunities with relation to seismic emergency planning and management concerns. Emergency managers increasingly need spatial data to make complicated decisions involving numerous stakeholders across multiple teams and criteria.

Turkey is located in a dangerous geographic location. Turkey is at the intersection of four tectonic plates: Anatolian,

---

Email address: `saptadeepbiswas7@gmail.com` (Saptadeep Biswas), `dhruvphy.stud@nita.ac.in` (Dhruv Kumar), `bera_uttam@yahoo.co.in` (Uttam Kumar Bera)

Arabian, African and Eurasian. Turkey's seismotectonic setting has long been characterised by the westward extension of Anatolia caused by the collision of the Arabian and Eurasian plates. At a rate of around 4.5 cm per year, the African plate is continuously pushing the Anatolian plate westward [35]. The North Anatolian fault and the East Anatolian fault, which are important transformations, are where the majority of the deformation related to the extrusion is accommodated. Moreover, a back-arc extensional regime in Western Anatolia is produced by the complex tectonic process of the African plate subducting along the Aegean arc and Cyprian arc. The North Anatolian Fault (NAF), East Anatolian Fault (EAF), North East Anatolian Fault (NEAF), and Aegean Arc regulate the majority of seismic events occurring in Turkey.

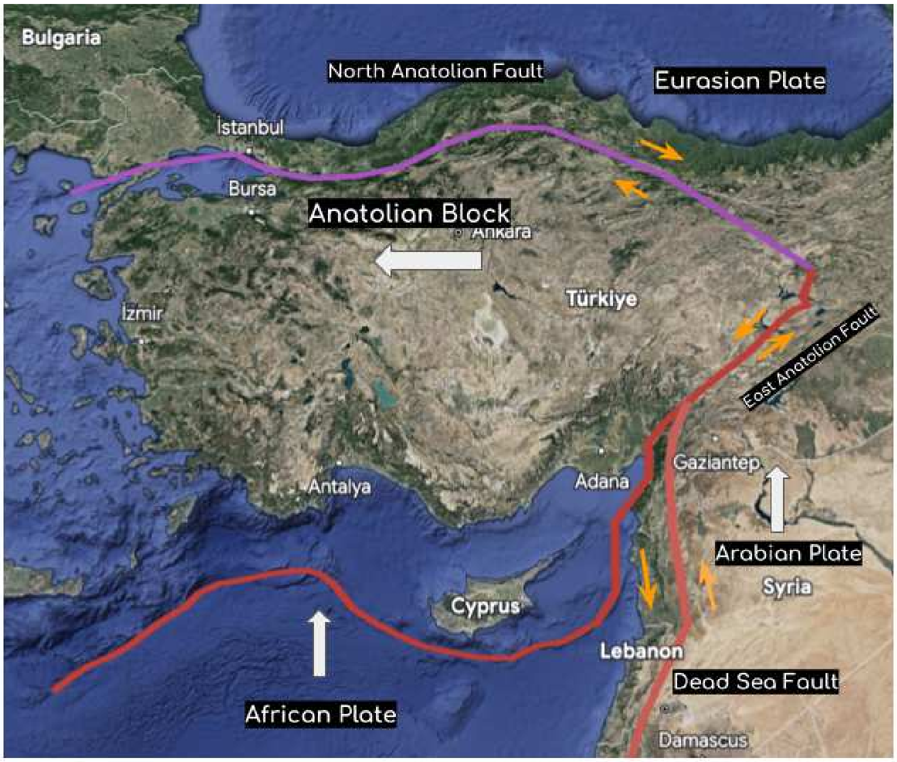


Figure 1: Geotectonic position of Turkey

At least 20,000 people died as a result of the magnitude 7.4 earthquake that rocked the Turkish city of Izmit on August 17, 1999 [18]. This terrible incident served as additional evidence in favor of a novel method for predicting earthquake locations in the future by examining the stress transfer along faults. An earthquake won't often occur on a section of a fault again until the stress rebuilds because each earthquake event relieves some of the stress that builds up on fault segments over time. Such reaccumulation often occurs over hundreds to thousands of years. The future earthquakes are likely to strike a new part of the fault, though, because an earthquake also transmits and increases stress further along the fault. The 1,400-kilometer-long (870-mile-long) east-west North Anatolia fault in Turkey has experienced a systematic development of large earthquakes from east to west. The USGS scientists estimated a 12% probability that an earthquake of magnitude 6.7 or bigger would occur in the Izmit area during the next 30 years, a forecast that was made just before the 1999 Izmit earthquake. The Izmit earthquake, according to USGS's preliminary assessments, increased strains on a fault segment in the Marmara region southeast of Istanbul. The frequency of minor earthquakes beneath the Sea of Marmara has significantly increased after the Izmit earthquake, supporting this finding. These results helped shape USGS predictions of a 62% likelihood of heavy shaking near Istanbul over the following 30 years.

A 7.8-magnitude earthquake with a 17.9-km depth struck Turkey and Syria at 4:17 a.m. on February 6, 2023, killing over 300 people immediately [1]. A powerful 7.6-magnitude earthquake rocked the Elbistan region in the southern Turkish province of Kahramanmaraş at 4:30 p.m., leaving 1,300 people dead. The earthquake was first felt at Pazarck, a village in southern Turkey, about an hour's drive to the north of Gaziantep, a significant industrial city. The epicenter of the second tremor was at Nurda, a city some 80 kilometers to the southwest. On February 17, 2023, news reports indicated that 10 Turkish provinces had experienced at least 40,642 fatalities, 108,068 injuries, and 41,791 damaged or destroyed homes [2]. At least 6,396 people died, 14,500 were hurt, 490 structures were destroyed, and thousands more were damaged in northwest Syria. Iskenderun, in Turkey, experienced land subsidence. Turkey was severely damaged by these earthquakes, which resulted in about 48,448 fatalities and 115,000 injuries across its 11 provinces[3]. A small tsunami was brought on by waves measuring 12 cm in Iskenderun, Turkey; 13 cm in Erdemli; and 17 cm in Famagusta, Cyprus. A maximum magnitude of 7.8 and an intensity level of IX were reported for the most powerful earthquake ever.

Arabian plate is moving into Eurasian plate, pushing the Anatolian plate—which Turkey lies on—outward and toward the north. The movement of the tectonic plates causes pressure to build up on the fault zones that are close to their boundaries. Cracks in the earth's crust that cause movement are known as faults. The sudden release of this pressure results in earthquakes and shaking of the ground. The East Anatolian fault or the Dead Sea Transform fault, one of the primary faults, is probably where the Turkey earthquake took place. The energy of an earthquake is actually produced by moving along a fault's area. The fault area that has shifted has grown in size in direct proportion to the earthquake's magnitude. An event with a magnitude of 7.8 most likely moved across an area that was about 190 km long and 25 km wide. The shaking will consequently be felt over a huge area. There will be a lot of aftershocks following large earthquakes as the crust recalibrates to the changes in stress. In the first twelve hours after the original quake, there were three other earthquakes in southeast Turkey that had a magnitude greater than 6.0. On the Sürgü Fault, a distinct but neighboring fault system, further to the north, an aftershock of magnitude 7.5 took place. All these heavy earthquakes occurred where the Dead Sea fault and the East Anatolian fault meet. This earthquake sequence caused serious damage to the northernmost portion of the Dead Sea fault and the majority of the southwest portion of the East Anatolian fault.

Hazard maps are created to highlight locations that are impacted by or sensitive to a specific hazard. By minimizing exposure to some risks and planning other development to offset or neutralize their potential negative impacts, planners, developers, and engineers can save lives and prevent financial losses by appropriately utilizing the resources provided by hazard maps. Many risks associated with earthquakes and their aftershocks can be identified and mapped individually in areas where doing so is a top priority. Landslides and soil liquefaction can result from ground shaking in areas that are vulnerable to these dangers. Understanding the type and size of local flood plains is crucial for managing earthquake risk because ground failure that damages dikes and dams can result in flooding.

---

[1] https://www.emsc-csem.org/Earthquake/earthquake.php?id=1218444

[2] U.S. Geological Survey, The M7.8 and M7.5 Kahramanmaraş Earthquake Sequence struck near Nurdağı, Turkey (Türkiye) on February 6, 2023, at URL https://www.usgs.gov/news/featured-story/m78-and-m75-kahramanmaras-earthquake-sequence-near-nurdagi-turkey-turkiye

[3] "Depremlerde can kaybı 48 bin 448 oldu" [Loss of life in earthquakes was 48 thousand 448], TRT Haber, 13 March 2023.

The East Anatolian Fault (EAF), a significant active left-lateral strike-slip fault in Turkey, experienced seismic activity in the 19th century [13, 10]. Seismically, it is currently active once more. As much as 400 kilometres of the fault's surface were ruptured during the earthquake of 2023. The EAF is among the most intense and damaging fault zones, producing several earthquakes of large size both historically and experimentally, according to Güvercin et al. (2022) [12]. Predicting earthquake magnitude is crucial for handling the situation. Under this circumstance, earthquake hazard maps must be produced using GIS and the estimated magnitude in order to identify the most dangerous areas, respond quickly to the right places, organize the mobilization of equipment, the removal or retrofit of assets, and assess damage and initiate recovery operations as a part of catastrophe mitigation measures. It is essential to determine the outputs of danger maps as quickly and precisely as feasible.

1.1. ***Related Work***

Seismic risk assessment plays a fundamental role in minimising potential losses due to upcoming earthquakes. It is a crucial instrument for establishing governmental policy about land use planning, building codes, insurance, and emergency preparedness. The impact of earthquakes on society and the economy can be reduced by accurate seismic hazard and risk estimation. In order to lessen prospective losses caused by future earthquakes in Nepal, Chaulagain et al. (2015) [5] evaluated structural vulnerability, seismic risk, and the consequent potential economic losses. In order to assess the seismic risk, recent exposure data, a newly designed structural vulnerability, and an existing probabilistic seismic hazard are all used. The Global Earthquake Model initiative's open-source platform for assessing seismic hazard and risk, known as OpenQuake-engine, was utilised to determine the seismic hazard and risk in Nepal. According to this study, the mid-western and eastern regions of Nepal have a larger seismic risk than the southern region of Nepal, which has the lowest seismic risk. Maio et al. (2016) [19] presented a case study of Faro city, with the goal of analyzing and evaluating the seismic susceptibility of its 354 old brick buildings. These goals were accomplished by using a vulnerability-index-based method to analyze the seismic risk of the old masonry buildings located in Faro's historical center (Ribeirinha region), which was done through thorough examination and survey. The generated results were analyzed using a geographic information system (GIS) tool. Using GIS techniques, the study of Bhagyaraj et al. (2023) [4] has mapped the earthquake-induced landslide susceptibility in Ladakh and Jammu and Kashmir, two recently constituted union regions of India. In order to prevent landslides from occurring, they determined the static safety factor, which is a crucial component of landslide hazard assessments. Due to valley obstruction and landslide dam breaking incidents, the Himalayan region also experienced landslide-induced damages in the downhill zones. Das et al. (2023) [8] demonstrated that experts conducted landslide susceptibility zonation (LSZ) assessment for several regions of this seismically active high mountainous terrain utilizing a variety of methodologies, mostly spanning from conventional subjective views to sophisticated machine learning models. Sangeeta et al. (2020) [32], Sangeeta et al. (2022) [31], and Samm-A et al. (2023) [30] all perform additional study on LSZ . The map combination strategy and the AHP method were consistently used with various area and mapping unit sizes, it has been observed.

Rahman et al. [28]'s study has assessed the risk of earthquake and fire in a particular location of Dhaka City while taking socioeconomic factors into consideration. FEMA-RVS visual screening method, ADPC (2004) fire hazard methodology, and World Bank (2014) social vulnerability methodology were all used to examine a sample of 350 structures for earthquake, fire hazard, and vulnerability to social factors. The research area's socioeconomic vulnerability state, as well

Table 1: Abbreviations

Names	Abbreviations
Artificial Neural Network	ANN
Adaptive Neuro-Fuzzy Inference System	ANFIS
Geographic Information System	GIS
Seismic Multi-channel Analysis of Surface Waves	MASW
Analytic Network Process	ANP
Rapid Visual Screening	RVS
Random Forest Regressor	RF
Precision Appraisal	PA
Frequency Ratio	FR
weights-of-evidence models	WEM
Radial Basis Function	RBF
Extreme Gradient Boosting Algorithm	XGBoost
Bayesian Ridge	BR
Deep Neural Network	DNN
Multivariate Adaptive Regression Splines	MARS
Boosted Regression Trees	BRT
Dam Break Model	DamBrk
Boosted-Tree model	BT
Decision Tree Regressor	DT
Category Gradient Boosting	CatBoost
Extra Trees Regressor	ET
Light Gradient Boosting Machine	LGBM
Mean Absolute Error	MAE
Mean Square Error	MSE
Root Mean Square Error	RMSE
R-squared score	R2
Normal Root Mean Square Log Error	NRMSLR
Federal emergency Management Agency	FEMA
Rapid Visual Screening	RVS
Multiple-criteria decision analysis	MCDA
Analytic Hierarchy Process	AHP

Table 2: Abbreviations

Names	Abbreviations
Technique for order of preference by similarity to ideal solution	TOPSIS
Peak Ground Acceleration	PGA
Delaware Environmental Navigator	DEN
Silhouette Clustering	SC
Hierarchical Clustering Analysis	HCA
Convolutional Neural Network	CNN
Fuzzy Analytic Hierarchy Process	FAHP
Multilayer Perceptron	MLP
Rapid Visual Survey by Institut Teknologi Sepuluh Nopember	RViSITS
Long short-term memory	LSTM
Fuzzy-Logistic Regression	fuzzy-LR
Multi-Criteria Evaluation–Logistic Regression	MCE–LR
Classification Tree Analysis	CTA

as the risk of earthquake and fire, have all been taken into account when creating the composite vulnerability score. It was displayed as a map created with ArcGIS 10 that showed structures in several vulnerable categories. The findings indicated that fire risk is relatively more prevalent in the research area than earthquake risk. Cox et al. (2011) [7] present a seismic site classification micro zonation for Port-au-Prince based on 35 shear wave velocity (VS) profiles gathered throughout the city and a revised geology map of the area. The multichannel analysis of surface waves (MASW) approach was used to acquire the VS profiles. The geologic map was created using field mapping and geomorphic interpretation of a digital elevation model (DEM). In order to classify seismic sites around the city using codes, relationships between the mean shear wave velocity over the top 30 m of the subsurface (VS30) and surficial geologic unit have been constructed. This article provides a site classification map for the National Earthquake Hazards Reduction Program.

The amount to which social, economic, and physical assets are vulnerable to natural disasters is referred to as urban vulnerability. In order to determine the index value of vulnerability while taking into account social, economic, environmental, and physical vulnerability, Alizadeh et al. (2018) [1] created a new hybrid framework utilising Analytic Network Process (ANP) and Artificial Neural Network (ANN) models. This score was then applied to Tabriz City, a seismically active province in Iran's northwest that frequently experiences destructive earthquakes with significant aftershocks and damage. In order to create an earthquake vulnerability map, a Geographical Information Systems (GIS) analysis was utilised to discover and assess quantitative vulnerability indicators. In order to create the training database, the classified and standardised indicators were then sorted and weighted using an ANP model. Standardized maps along with the training site maps were then supplied as input to Multilayer Perceptron (MLP) neural network generating an Earthquake Vulnerability Map (EVM). Ultimately, an EVM was created for Tabriz City, and the vulnerability levels in different zones were determined. The results show that the south and southeast areas of Tabriz City have low to moderate susceptibility, while some zones in the northeastern tract have critical vulnerability conditions. In their study, Malakar et al. (2022)

[20], they employed the hierarchy density-based clustering technique to automatically identify the Himalayan ranges' most active seismicity zones, which are situated along the Indo-Eurasian tectonic plate collision boundary. Effectively increasing earthquake susceptibility include geological, structural, and social factors. The study [21] examined seismic susceptibility in the Abbottabad District of Pakistan by incorporating different important supporting techniques to produce more precise results. To estimate the vulnerability zones of various hazard areas with the greatest degree of precision, hybrid models such as fuzzy-logistic regression (fuzzy-LR) and multi-criteria evaluation-logistic regression (MCE-LR) are trained using multiple criteria decision analysis-multiple criteria evaluation (MCDA-MCE) and fuzzy-multiple criteria analysis (fuzzy-MCDA).

For the risk assessment of seismic loss, a number of analytical techniques, including both qualitative and quantitative risk evaluations, are accessible since scenario-based seismic disaster risk assessment is an efficient method for enhancing earthquake relief. In order to estimate the potential social and economic damages from earthquakes, scenario-based seismic disaster risk assessment is frequently employed and has grown to be a highly significant tool for local residents and governments. In their study published in 2019, Zhuang et al. [38] suggested a strategy that takes into account the population as the element most at danger, as well as the peak ground acceleration (PGA) of an earthquake and the landslides that are caused by it. These three variables can be used to calculate the risk to social and economic elements as well as the possible repercussions of earthquakes. The PGA is a key component in calculating the magnitude of an earthquake and is a crucial factor in evaluating how much infrastructure is destroyed and moved as a result of an earthquake. The size and distance of the earthquake from the active fault have the biggest effects on the PGA. The population, the second aspect, has to do with the local social and economic context. Geological changes, monetary loss, and mortality are all effects of earthquake hazards. For the purpose of predicting and preventing earthquake damage, Earthquake Risk Mapping (ERM) is crucial knowledge. In order to construct hybrid models for ERM, Yariyan et al. [36] study employed the Classification Tree Analysis (CTA) learner model with three Gini, Entropy, Ratio split, and Fuzzy ART MAP (FAM) models. The Earthquake Risk Conditioning Factors, which encompass environmental, physical, and social elements, were chosen based on the opinions and experiences of experts.

In order to address geohazards and real-world issues, artificial intelligence (AI) approaches have gained a lot of traction and have been applied in a variety of applications, especially in the area of seismic risk and prediction. Jena et al. (2020) [15] work used clustering analysis to pinpoint earthquake-prone regions, created a CNN model to estimate probabilities, and calculated and compared the risk using two different equations. The CNN model evaluated probability due to its strong predictive capacity, and SC and PLC were used to comprehend spatial clustering, Euclidean distance between clusters, spatial relationships, and cross-correlation among the predicted Mw, PGA, and intensity including events depth. The vulnerability assessment was completed by using AHP. In order to do this, results from assessments of earthquake vulnerability, seismic amplification susceptibility, and earthquake probability were used to generate risk. Asim et al. (2017) [3] used machine learning classifiers in conjunction with the temporal sequence of previous seismic activity to forecast the magnitude of earthquakes for the Hindukush region. The study of Asim et al. (2018) [2] used a prediction model consisting of Support Vector Regressor (SVR), followed by Hybrid Neural Networks (HNN), and Improved Particle Swarm Optimization, to carry out earthquake prediction using sixty seismic features from the Hindukush, Chile, and Southern California regions. The study by Jena et al. (2020) [16] created an integrated model by utilizing the artificial neural network-analytic hierarchy process (ANN-AHP) model for creating the seismic risk assessment map for Aceh province, Indonesia. For assessment of earthquake risk in Sanandaj City, Iran, Yariyan et al. (2020) [37] proposed a hybrid FAHP-ANN model. The deep learning approach is an excellent choice for modeling and forecasting natural hazard risks. In numerous natural hazard predictions, the Long Short-Term Memory (LSTM) model has gained significant popularity. In Jena et al. (2021)[14], the earthquake susceptibility of the entire country of India is evaluated using the LSTM model and appropriate Geospatial Information Systems (GIS) methodologies. Their findings showed that the AI-based model outperformed conventional models and basic neural networks.

Table 3: **Review**

Author (Year)	Study Areas	Methods
Pal et al. (2008)[26]	Sikkim, Himalayan territory	GIS
Cox et al. (2011)[7]	Port-au-Prince, Haiti	MASW
Chaulagain et al. (2015)[5]	Nepal, Himalayan territory	Vulnerability functions, OpenQuake-engine
Rahman et al. (2015)[28]	Dhaka city, Bangladesh	FEMA-RVS, AHP, GIS
Maio et al. (2016)[19]	Faro, Portugal	Hybrid vulnerability-index-based method, GIS Tool.
Alizadeh et al. (2018)[1]	Tabriz city, Iran	ANP, ANN
Merciu et al. (2018)[22]	Bucharest, Romania	GIS
Nyimbili et al. (2018)[25]	Küçükçekmece Region, Istanbul, Turkey	MCDA, AHP, TOPSIS, GIS
Zhuang et al. (2019)[38]	Guyuan Region of China	PGA, DEM
Jena et al. (2020)[15]	Palu, Indonesia	HCA, SC, ML, CNN, AHP, GIS
Yariyan et al. (2020)[37]	Sanandaj City, Iran	AHP, FAHP, FAHP-ANN, MLP
Jena et al. (2020)[16]	Aceh province, Indonesia	ANN, AHP
Riyanto et al. (2020)[29]	Surabaya, Indonesia	RVS, RViSITS, FEMA
Jena et al. (2021)[14]	India	LSTM, GIS
Maqsoom et al. (2021)[21]	Abbottabad District, Pakistan	fuzzy-LR, MCE–LR
Yariyan et al. (2022)[36]	Sanandaj city, Iran	CTA, NN, GIS
Malakar et al. (2022)[20]	Himalayas	Fuzzy AHP, TOPSIS
Bhagyaraj et al. (2023)[4]	Ladakh, Jammu, Kashmir, India	Newmark's method, GIS
This paper	Turkey	ANN, XGBoost, CatBoost, RF, DT, ET, LGBM, BR

Earthquake hazard maps show the amount of shaking that can be reasonably anticipated at a given likelihood over a given period of time [34]. Despite the fact that these maps are frequently produced, experience has shown that they frequently fall short of predicting future earthquakes and what will happen during them. The location and size of upcoming earthquakes can be predicted with the use of information on active faults. Strong earthquake hazard assessments help us better grasp how frequently and violently the earth might tremble in the future at a certain site [11]. A seismic hazard assessment's goal is to identify the ground motions that should be taken into account when designing structures

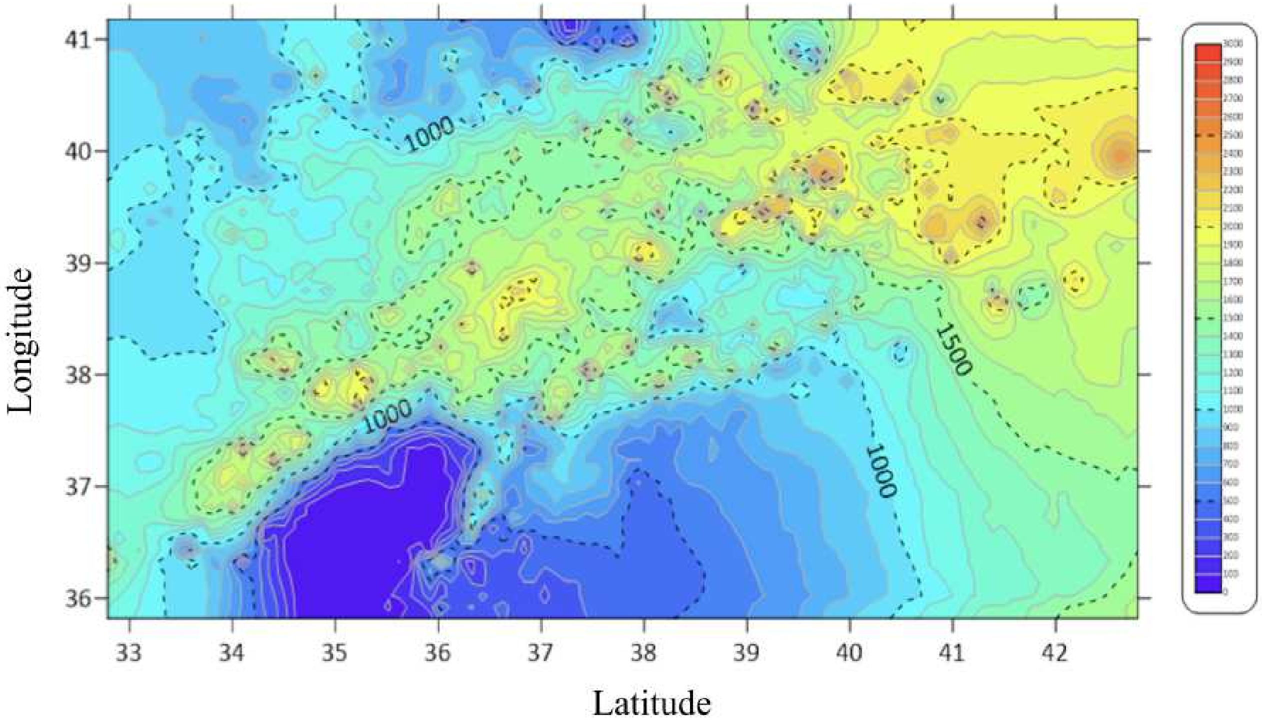


Figure 2: 2D Elevation Mapping for Turkey

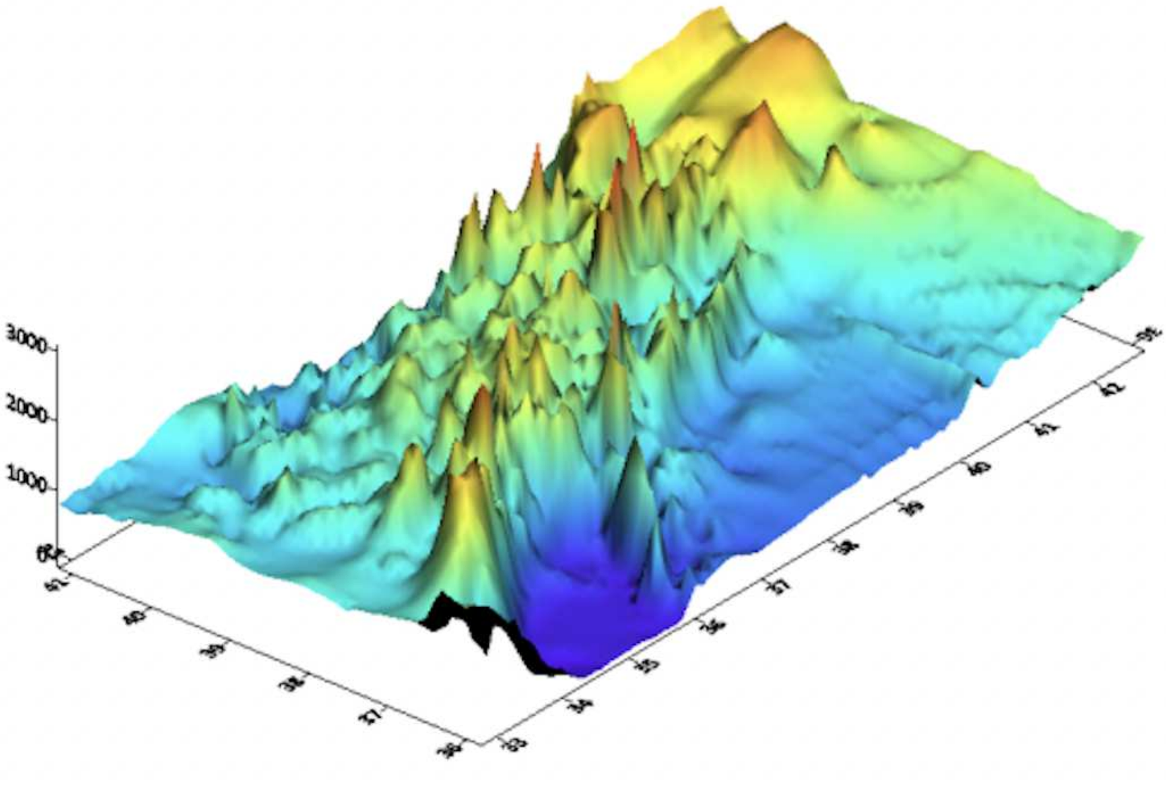


Figure 3: 3D Elevation Mapping for Turkey

or estimating risks. A model for the future earthquake source and a model to predict the movements at the site because of each earthquake scenario make up the two fundamental parts of any assessment of the earthquake hazard. Additionally, it lays the groundwork for effectively modifying mitigation strategies to increase a community's quake resistance. To lessen the harm, this study presents a Turkish Seismic Hazard Assessment and Mapping. In this study, earthquake prediction and the creation of seismic hazard maps are both accomplished using artificial intelligence (AI) techniques and geographic information systems (GIS). The central and southern regions of Turkey are the case study for this work, which focuses on the use of the GIS based on prior geophysical data using the ML and ANN technique.

1.2. ***Paper Structure***

This paper is structured as follows: Section 2 provides a geological explanation of the region mentioned earlier. In part 3, data analysis is covered. Sections 4 provide an outline of the research methods. Section 5 analyzes all of the findings from each proposed approach. In section 6, we've offered a case study to demonstrate the validity of our concepts. Section 7 includes a general analysis, benefits of our framework, and suggestions. The conclusions are mentioned in Section 8. Future research directions are presented in Section 9 to conclude our plans.

2. Geology and Tectonics

Turkey is primarily a mountainous nation, with significant lowland areas limited to the coastline borders. Less than two-fifths of the surface is below 1,500 feet in elevation, and around one-fourth is above 4,000 feet. In several locations, especially in the east, mountain crests are higher than 7,500 feet. The Demirkazk Peak (12,320 feet) and Mount Aydos (11,414 feet) are also noteworthy summits. In the southeast, Uludoruk Peak rises to a height of 15,563 feet. In contrast to the fact that only around one-sixth of the terrain is level or moderately sloping, steep slopes are typical throughout the nation. Several components of the physical environment are impacted by these mitigating factors, creating climates that are frequently far severe than would be predicted for a country with Turkey's latitude. With sedimentary rocks from the Paleozoic to the Quaternary, many intrusions, and vast amounts of volcanic material, Turkey's geology is complicated. The central massif, the southern folded zone, the northern folded zone, and the Arabian platform can be distinguished as the four primary geographical areas.

When tectonic plates move within the crust of the Earth, geological faults begin to rupture, causing earthquakes to happen. Turkey is in a particularly risky position since it is situated where three tectonic plates meet. Turkey experiences seismic activity as a result of the relative motions of four major tectonic plates: the Anatolian, African, Eurasian, and Arabian. As a result of many first-order plate boundary interactions between these plates, due to subduction, large-scale transform faulting, compressional mountain building, and crustal extension, the area's geology has developed. The Anatolian block is moving westward as a result of the Mediterranean Sea closing when the African, Arabian, and Eurasian plates collide. Along the North and East Anatolian faults, such motion is mostly absorbed. Between the Anatolian block to the north and the African plate in the south, Cyprus Arc serves as the convergent (subduction) border. A strike-slip system with a northward tendency that accommodates the differential motion of the African and Arabian plates is the Dead Sea fault, which is located further south. At the southern end of the Dead Sea fault, where speed decreases farther north, the Arabian plate slides through the African plate in a northerly direction at a rate of roughly 10 mm per year. The heavily populated Levant region has always faced serious risks from seismic events all along Dead Sea fault (eastern

Mediterranean). A complicated deformation zone in southern Turkey is the convergence point of the western edge of the East Anatolian fault, the north part of the Dead Sea fault, and also the eastern tip of the Cyprus Arc.

The regions of Turkey's central and south are of more interest to us. This region includes some significant locations like Gaziantep, Kahramanmaraş, and Hatay where there has been a great deal of destruction and loss of life. To understand the elevations of Turkey's surface, we utilize both 2D and 3D mapping [Figure 2, 3]. GIS technology was employed for this. Geographic Information System (GIS)-based mapping is the process of creating maps and researching geographic data using GIS technology. We have utilised Surfer software to perform 2D and 3D elevation profile mapping for the GIS-based mapping.

2.1. ***Gaziantep***

Gaziantep is located on Latitude: 37.0662° N and Longitude: 37.3833° E at an elevation of 856 meters (2,808 feet) above sea level with a total area of 7,194 square kilometers (2,778 square miles) and a population of around 2 million people according to the 2021 census, making it Turkey's sixth-largest city.

Gaziantep, Turkey is situated close to the Arabian Plate and the Eurasian Plate boundaries. Since the two regions are converging at a rate of about 2-4 centimeters per year, this region is noted for its intense seismic activity. These tectonic shifts may lead to stresses and tensions in the outermost layer of the Earth, which may end up in earthquakes. Seismic measurements of the region indicate several active faults, including the East Anatolian Fault, which runs along Turkey's eastern coast, and the North Anatolian Fault, which runs along the country's northern coast. These lines have the potential to cause major earthquakes with magnitudes of 6.5 or higher.

2.2. ***Kahramanmaraş***

Kahramanmaraş is a city located in the Mediterranean region of Turkey. with latitude: 37.5872° N and longitude: 36.9267° E with elevation: 526 meters (1,726 feet) above sea level, the area: Approximately 14,000 square kilometers (5,400 square miles). Southern Turkish city Kahramanmaraş lies 78 kilometers north of Gaziantep. East-northeast of Adana, it is located at the border of a productive plain beneath Ahr Mountain. Three significant routes across the Taurus Mountains—from Göksun, Elbistan, and Malatya—have their southern outlets close to this city.

The area of Kahramanmaraş, Turkey, lies adjacent to the East Anatolian Fault Zone, which separates the Eurasian and Arabian plates. This fault zone is known to be active and has caused multiple major earthquakes in the past, including a magnitude 6.0 earthquake in 2003 and a magnitude 6.1 earthquake in 2012. Earthquake risks are significant in Kahramanmaraş and other zones near the East Anatolian Fault Zone, as the region is prone to powerful earthquakes. Building collapse, landslides, and tsunamis, as well as fires and other secondary consequences, are all hazards caused by earthquakes.

2.3. ***Hatay***

Hatay, Turkey is a city located on latitude: 36.2083° N and longitude: 36.1622° E with elevation: ranges from sea level to 1,950 meters (6,398 feet) above sea level and area: 5,524 square kilometers (2,131 square miles), the population:

According to the 2021 census, the population of Hatay is around 1.6 million people, making it the eleventh-largest city in Turkey. The most southern province in Turkey is Hatay Province. Most of it is outside of Anatolia, on the eastern shore of the Levantine Sea. The province shares borders with Syria on its south and east, Adana on its northwest, Osmaniye on its north, and Gaziantep on its northeast. It is situated in portion of Ukurova, a sizable fertile plain near Cilicia. Antakya serves as its administrative hub. The north-easterly line of equal latitude and longitude passes through Hatay, where 46% of the area is made up of mountains, 33% is plain, and 20% is made up of plateaus and foothills.

The Hatay province of Turkey is situated near the East Mediterranean region, which is the border between the African and Eurasian plates. This area is renowned for its complicated tectonic setting, which includes several active faults that offer a major seismic risk. Because of the existence of several active faults, including the North Anatolian Fault, the East Anatolian Fault, and the Dead Sea Transform Fault, the seismic threat in the Hatay area is significant. These faults have the potential to cause earthquakes of varying magnitudes, spanning from minor tremors to major, devastating occurrences. There is a history of devastating earthquakes in the region, including a magnitude 6.2 earthquake in 1998 that caused considerable damage and loss of life.

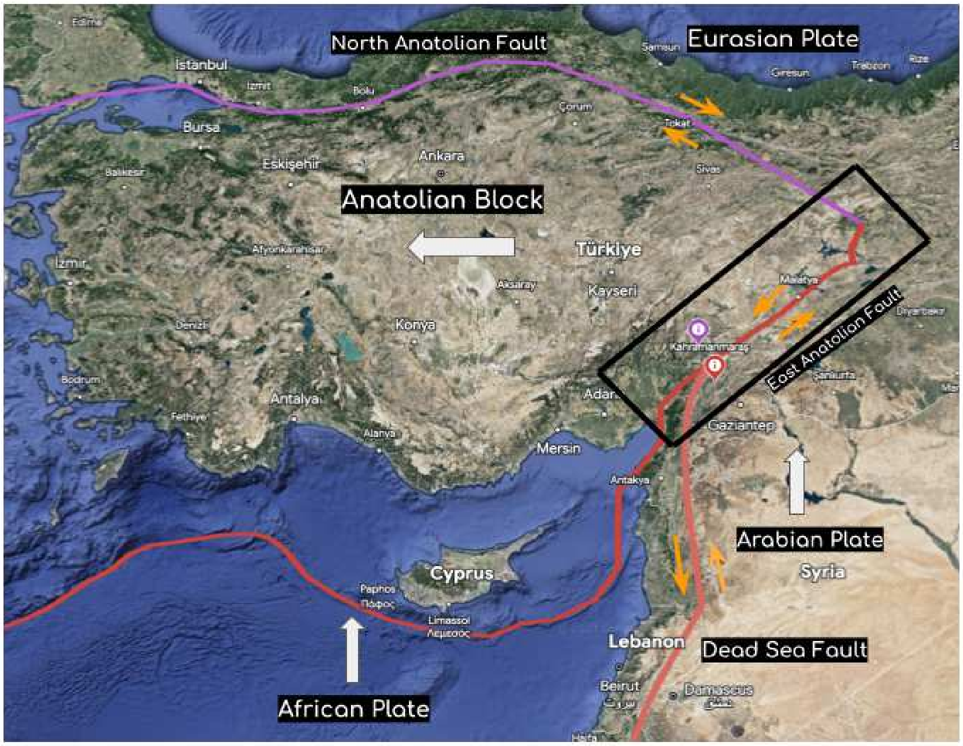


Figure 4: Mapping of Turkey

3. Data Analysis

3.1. Data Descriptions

Magnitude(mag): The magnitude of an earthquake is a measurement of its size at its epicenter. It's a logarithmic scale. The amplitude of the seismic waves used to estimate magnitude is roughly ten times greater during a magnitude 5 earthquake than during a magnitude 4 earthquake at the same distance from the event. On an Average earthquake total energy increases by a factor of around 32 for each unit rise in magnitude.

Depth(depth):

The depth indicates the point at which the earthquake begins to rupture. The WGS84 geoid, the mean sea level, or the average elevation of the seismic stations that provided arrival-time information for the earthquake location might all be used to measure this depth. The method utilized to locate the earthquake, which varies for each seismic network, has an impact on the reference depth selection. Because ComCat has data from several seismic networks, the technique for estimating depth varies depending on the event. The depth is the parameter that is least limited at the earthquake location, and because to the various depth measurement techniques, the error bars are frequently larger than the variation.

GAP: Between stations that are azimuthally adjacent, this is the biggest azimuthal gap. The more accurate the calculated horizontal position of the earthquake is, generally speaking, the lower this number. There are often significant location and depth uncertainty at earthquake locations where the azimuthal gap exceeds 180 degrees.

Minimum Distance (dmin): The horizontal distance between the nearest station and the epicenter is known as the Minimum Distance (dmin). A degree is roughly equal to 111.2 kilometers. In general, the accuracy of the estimated depth of the earthquake increases as this number decreases.

Root-Mean-Square (rms): The root-mean-square (RMS) travel residual with all weights is the parameter in a second. This value gives an indication of how closely the actual arrival times match those anticipated for this location. Smaller values indicate a better match of the data. The quality weights applied to the arrival time data, the method used to locate the earthquake, and the accuracy of the velocity model used to calculate the earthquake location all affect the value.

Table [4,5] lists the continuous and discrete parameters respectively that were employed in our analysis. The units, mean values, maximum values, minimum values, 75 percentile values, standard deviations (STD), typical values range, data type, and description of continuous variables are provided in Table[4]. Table[5] contains a list of discrete parameters along with their descriptions, sub-parameter counts, sub-parameter counts (count of how often they occur), and data types.

3.2. ***Date vs Magnitude***

Graph given in Figure[5], describes the magnitude of the earthquake over the course of January 2023 to March 2023. The high magnitude earthquakes are observed on 6^{th} February i.e., around 7.8 magnitudes, along with a number of earthquakes of high destructive magnitude.

3.3. ***Magnitude vs Depth***

The graph given in Figure[6], describes the relationship between an earthquake's magnitude with depth. For high-magnitude earthquakes, the depth is usually less. For instance, the depth for 7.8 magnitudes is just 10 meters.

3.4. ***Correlation plots and matrix***

A clear picture of the co-relationships between various factors can be seen in the correlation graphs in Figure [7]. Here, we have selected five characteristics, and are Magnitude (mag), Depth (depth), Gap (gap), minimum distance (d-

Table 4: **Continuous features from Data-set**

Parameters	Description	Units	Max	Min	75 %	Mean	STD	Typical Values	data type
Magnitude (mag)	The magnitude for the earthquake	Richter scale	7.80	4.50	5.00	4.84	0.46	[-1.0, 10.0]	float64
Depth (depth)	Depth of the earthquake	Kilometers	108.14	0.00	10.16	12.86	12.28	[0, 1000]	float64
Gap (gap)	The largest azimuthal gap between azimuthally adjacent stations.	Degrees	191.0	13.0	71.0	57.26	28.96	[0.0, 180.0]	float64
Minimum Distance (dmin)	Horizontal distance from the epicentre to the nearest station.	Degrees	2.68	0.11	1.12	0.89	0.45	[0.4, 7.1]	float64
Root Mean Square (rms)	The root mean square (RMS) travel time residual, in a sec, using all weights.	Seconds	1.37	0.33	0.88	0.75	0.18	[0.13, 1.39]	float64

Table 5: **Discrete features from Data-set**

Parameters	Description	Sub-parameter count	Sub-parameters (Count of number of times appear)	data type
Magnitude Type (magType)	The method or algorithm used to calculate the preferred magnitude for the event.	3	mb(163),mww(46), mwr(39)	int64
Network (net)	The ID of a data contributor. Identifies the network considered to be the preferred source of information for this event.	1	us(248)	int64
The seismic station (location-Source)	The network that originally authored the reported location of this event.	2	us(246), isk(2)	int64

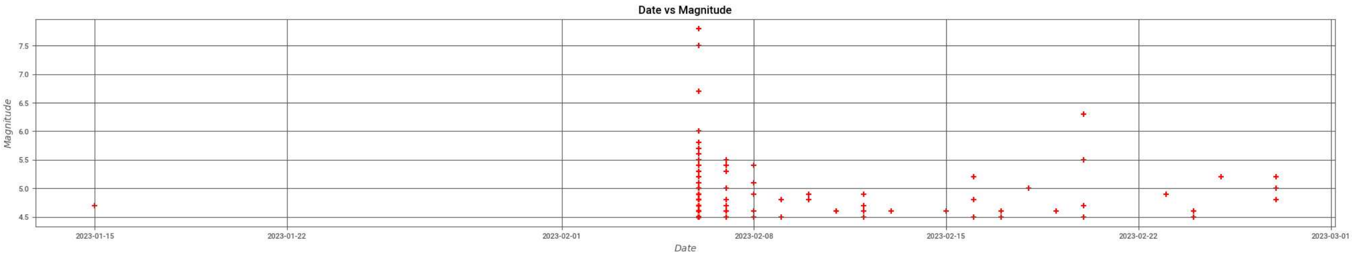


Figure 5: Date vs Magnitude

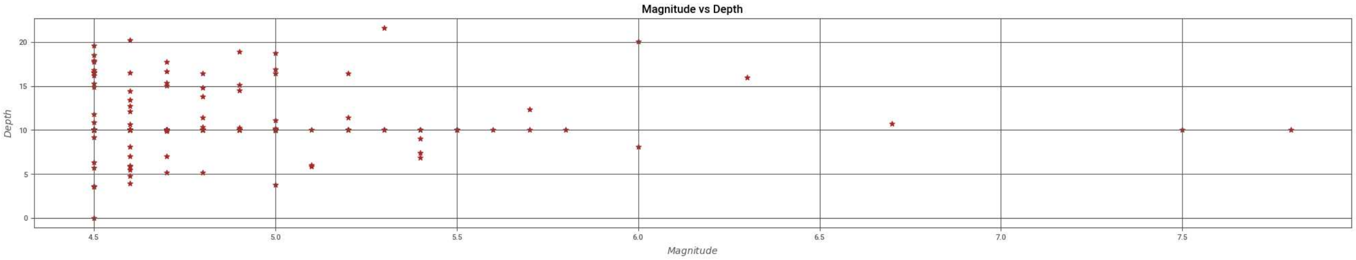


Figure 6: Magnitude vs Depth(m)

min), and root mean square error (RMS) all related to earthquakes. The values and counts of the corresponding features are displayed in the diagonal graphs. The correlation trends between various features are shown in additional figures. Figure[8] gives the numerical description of how much a parameter is dependent on others. Positive values indicated positive dependence and simultaneously negative values indicate negative dependence.

*3.5. **Danger-Zone Mapping***

Here, in this section we have done the required danger zone mapping depending upon the magnitude of earth quack observed till march 2023. We have classifier four Zones namely: Red zone, orange zone, yellow zone, and green zone.

Red Zone : It is the most dangerous zone with the highest earthquake magnitude. The range of magnitude lies between 5.9 to 8. The latitude and longitude positions are given in Figure [9]

Orange Zone : It is a moderately dangerous zone with a range of magnitude lying between 5.9 to 4.9. The latitude and longitude positions are given in Figure [9]

Yellow Zone : It is a mild dangerous zone with a range of magnitude lies between 4.9 to 4. The latitude and longitude positions are given in Figure [9]

Green Zone : It is a safe region where the magnitude lies less than 4. The earthquake causes no damage, the infrastructure can easily resist the destructive effects.

4. Methodology

Our study focuses on developing an effective method for determining earthquake vulnerability by utilizing the capability of ANN and ML algorithms. To do this, we acquired and processed data from the USGS spanning eight years (January 2015- February 2023). Latitude, longitude, earthquake magnitude (mag), depth (depth), gap (gap), minimum distance, root mean square (rms), depth error (depthError), magnitude error (magError), magnitude stations (magNst), magnitude types (magType-n), and location source (locationSource-n) are some of the important parameters included in the data

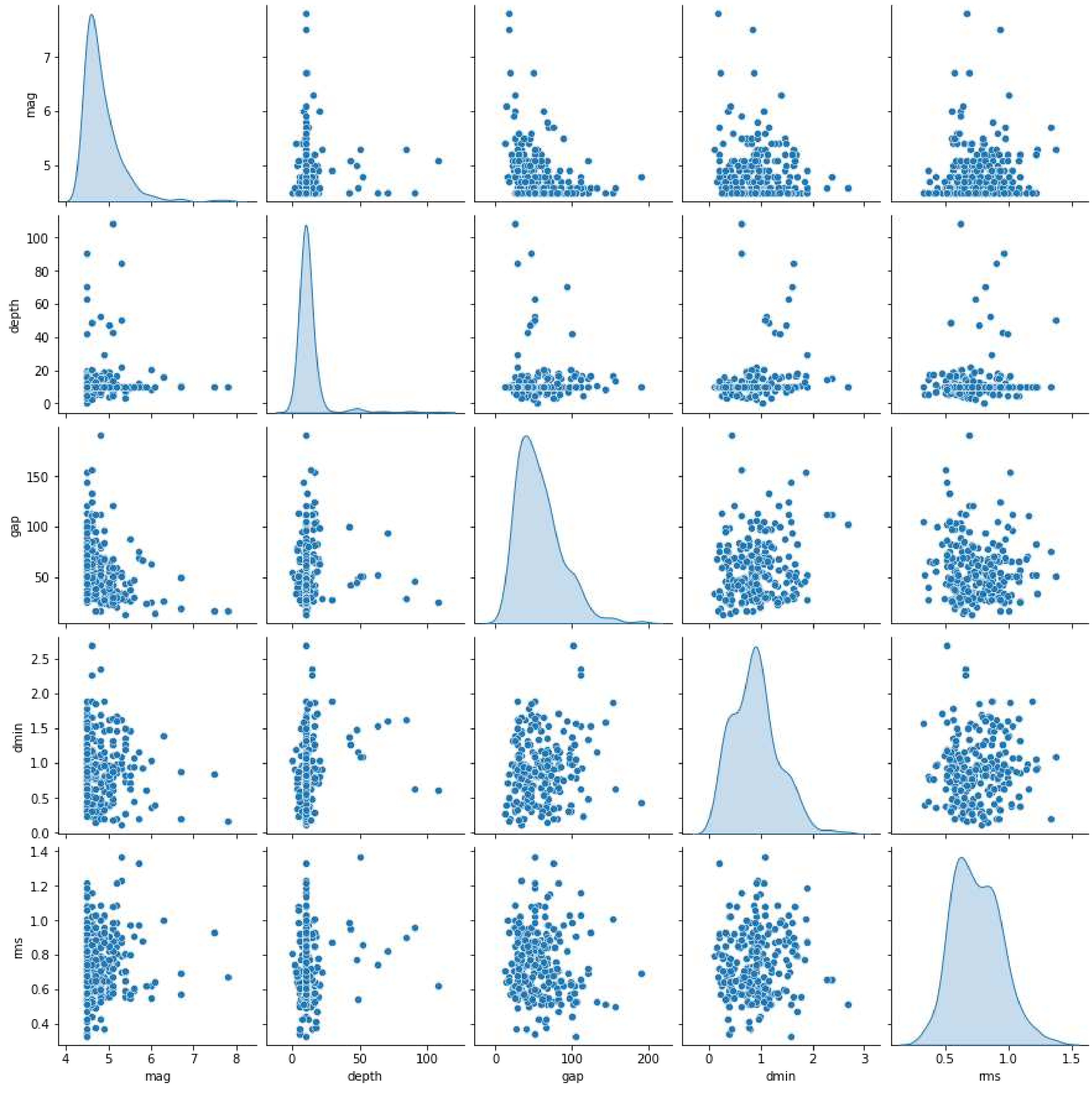


Figure 7: Correlation Plot

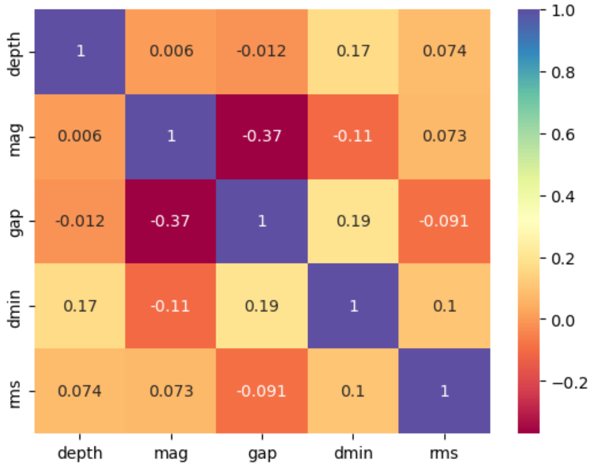


Figure 8: Correlation Matrix

that could have a direct impact on seismic mapping and earthquake magnitude prediction.

To ensure the quality and reliability of our results, we cleaned and filled in missing values in the raw data. We trained and examined three ANN and seven ML models using the cleaned data to determine the best technique for our investigation. We separated the dataset into two portions to train our models: a training set (85%) and a testing set (15%) for ML algorithms and simultaneously we separated the dataset into two portions to train our models: a training set (75%) and a testing set (25%) for ANN algorithms. The training set is then utilized in training our models on the link between various input parameters and earthquake risk. The models were then tested using the testing set, and their performance was evaluated using a variety of criteria. Our models are built with both continuous and discrete-valued parameters in mind, which are utilized to establish the weights and biases for all regression models. This method guarantees that our models can handle complicated interactions between input parameters and earthquake magnitude prediction.

We added an additional evaluation layer of a case study after performing all of this traditional analysis. We have obtained and analysed data from the USGS in the case study scenario, which spans over one month (01-March-2023 to 31-March-2023). Same are used in the data. Lastly, we produced seismic maps based on the estimated earthquake magnitude. Figure[10] is a flowchart that details the complete procedure, from data collection through model evaluation, to give an in-depth understanding of our technique. Our approach has the potential to enhance the evaluation of earthquake susceptibility and provide better disaster management solutions.

*4.1. **Machine Learning Algorithm***

For the prediction of the earthquake's 'magnitude,' we are using the ML regression algorithm. Here we have used seven ML models namely XGBoost, Random Forest, Extra Trees, Decision Tree, Bayesian Ridge, CatBoost, LightGBM.

*4.2. **Gradient boosting decision trees (GBDTs)***

Gradient-Boosted Decision Trees (GBDT) is a machine learning approach that combines the usage of decision trees and gradient boosting to enhance the performance of predictive models. It is a member of the family of ensemble learning techniques and includes merging the results of various models to arrive at a final prediction.

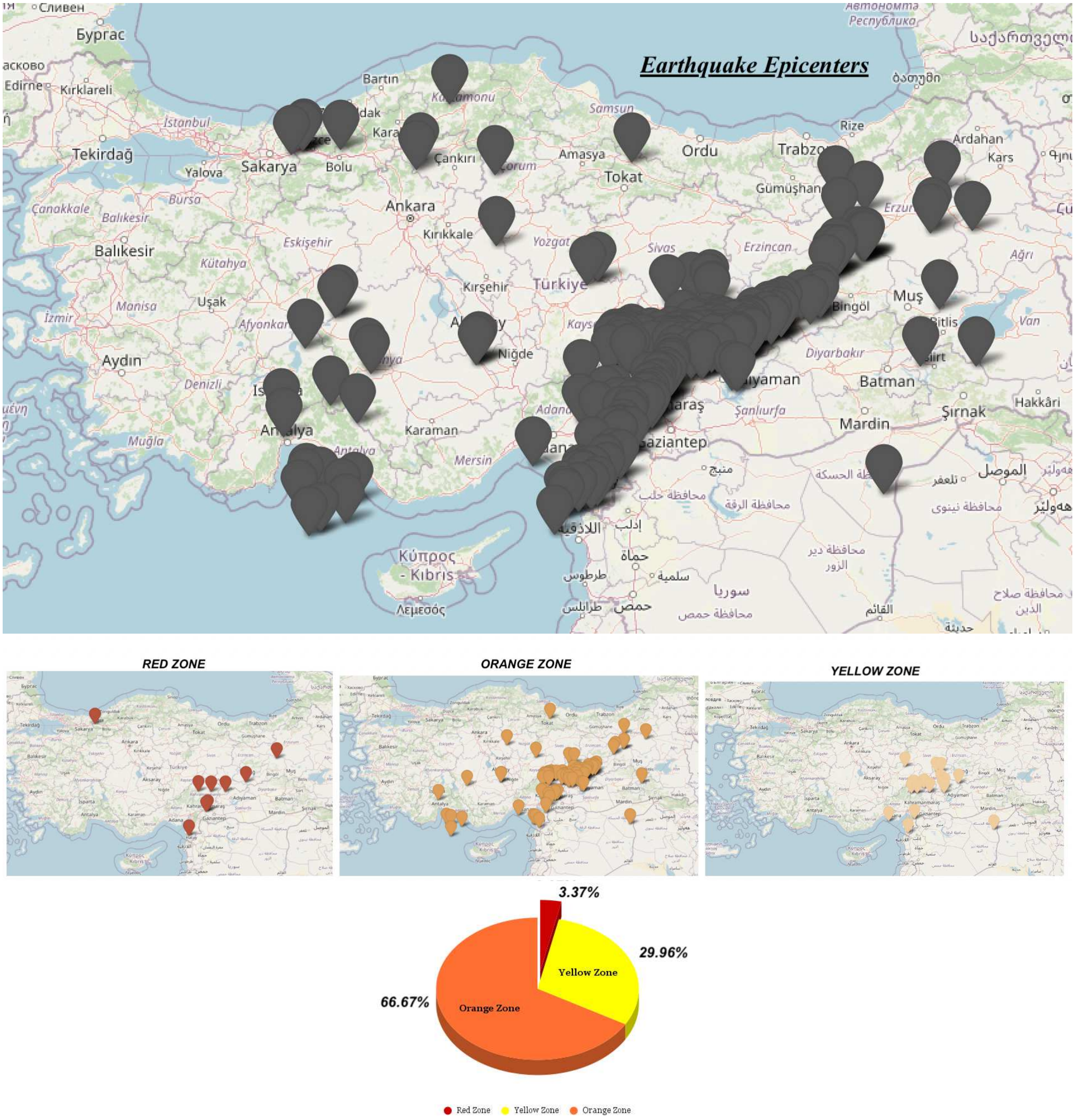


Figure 9: Earthquake Epicenters with Danger Levels

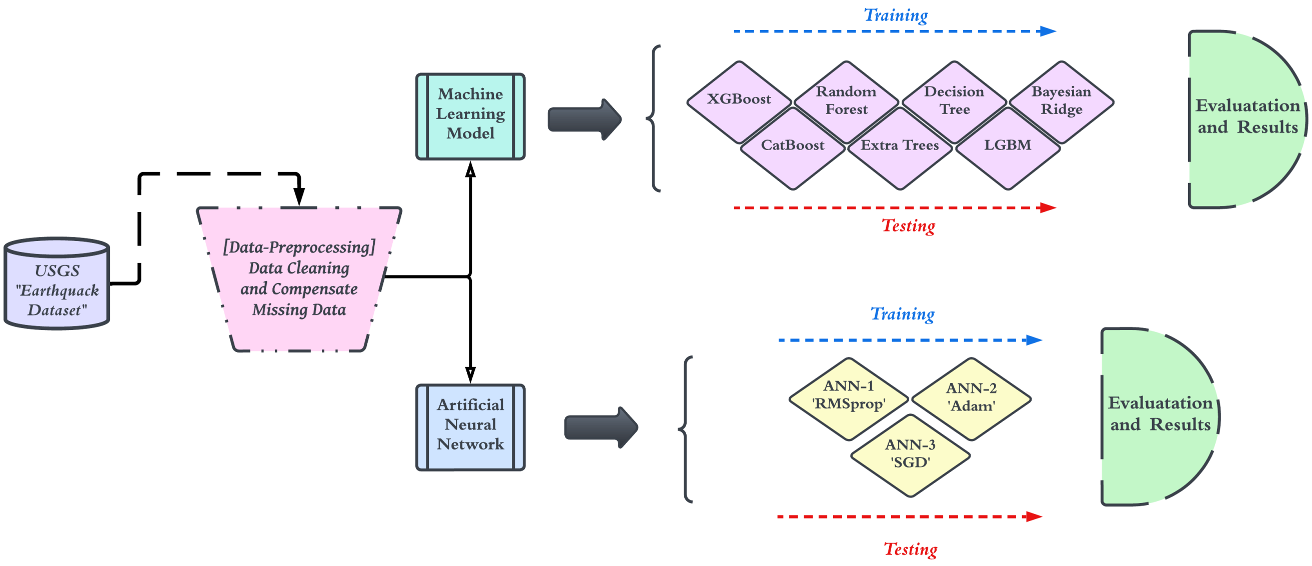


Figure 10: Flowchart of Methodology

In GBDT, decision trees are formed in a sequential manner, each one correcting the flaws of the one before it. In order for the procedure to function, the data are first fitted to a decision tree, and then further trees are fitted to the residual errors of the first tree. The final forecast is formed by adding the predictions from all the trees, with each tree being weighted according to how well it can lower the total error of the model.

4.2.1. ***eXtreme gradient boosting (XGBoost)***

An enhanced variant of the Gradient Boosting Algorithm is the XGBOOST algorithm. It is well-liked because of its strong prediction ability and straightforward implementation method. It creates the tree structure greedily rather than completely. It divides level-wise rather than leaf-wise in comparison to LightGBM. Negative gradients are considered in gradient boosting in order to optimise the loss function, but in this case Taylor's expansion is taken into consideration. This approach uses weighted quantiles to choose the best node split instead of evaluating every candidate split. Data is kept by XGBoost in the CPU's cache memory. One of XGBoost's distinguishing qualities is its proficiency in handling missing values and feature interactions, which makes it appropriate for use with complicated datasets. In order to avoid overfitting, it also enables regularised parameter adjustment. Both the tree booster and the linear booster, XGBoost, take sparse input and are built to handle it.

Nguyen et al. (2022) [24] developed ML models for the rapid seismic damage-state evaluation of steel moment frames during the post-earthquake period, which is an important and urgent activity. In order to accomplish this, eight ML models were investigated, including K-nearest neighbors, naive Bayes, decision trees, random forests (RF), adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), light gradient boosting, and category boosting. The performance of the XGBoost model (97%) was the best of the four boosting techniques taken into account.

4.2.2. ***Light gradient boosting machine (LightGBM)***

The Light Gradient Boosting Machine (LightGBM) is a gradient boosting architecture that accelerates training and uses less memory by employing a cutting-edge method called Gradient-based One-Side Sampling (GOSS). It was created by Microsoft with the goal of being effective and scalable for big data. LightGBM builds decision trees into the model iteratively, with each new tree trained to rectify the mistakes caused by the prior trees. It calculates the decision trees' splitting points using a gradient-based algorithm, which can dramatically speed up computation compared to more conventional techniques. During the training phase, the GOSS technique employed by LightGBM involves choosing sampling the examples with greater gradients, which lowers the number of instances that must be processed and may increase the model's accuracy. It also employs histogram-based methods to boost productivity and further decrease memory use. LightGBM improves the gradient boosting technique by incorporating an autonomous feature selection component and concentrating on cases with larger gradients. As a result, training may go much more quickly and predictability can be enhanced. For massive datasets, LightGBM is renowned for its quick training times and great accuracy. To speed up training even more, it supports parallel and GPU processing.

Due to the inapplicability of currently available locally applicable earthquake attenuation equations and the inapplicability of approaches based on observation data in areas without access to observation data, Chen et al. (2021) [6] presented a prediction model known as a light gradient boosting machine with a selection of features to address those problems and also to increase the accuracy of ground motion parameters. Based on a light gradient boosting machine (LightGBM) developed with historical strong motion data from the NGA-west2 database, it can quickly duplicate the distribution of strong motion close to the epicenter following an earthquake. A mine fire has the potential to quickly poison the mine's air, which could lead to the loss of life and, in some circumstances, the cessation of mining sector operations. In order to reduce the number of fatalities and material losses during the building of deep underground engineering projects, Kamran et al. (2023) [17] explored some intriguing elements of mine fire data utilizing Catboost and LightGBM methodologies.The results show that LightGBM algorithms outperformed Catboost in terms of performance, with accuracy levels of 92% and 89%, respectively.

4.2.3. ***Categorical boosting (CatBoost)***

Categorical Boosting (CatBoost) is a framework for gradient boosting that is made to handle categorical variables in data more efficiently. It was created by Yandex and is intended to be simple to use and extremely accurate. Similar to other gradient boosting systems, CatBoost incorporates decision trees into the model iteratively to remedy mistakes made by earlier trees. But, it has a few improvements that make it especially useful for categorical data. To make it simpler to locate the ideal splitting points, it employs an algorithm called Ordered Boosting that sorts the category characteristics prior to training the trees. L1 and L2 regularisation, as well as a cutting-edge method for handling missing data values, are additional approaches included in CatBoost that help to decrease overfitting and increase the model's accuracy. CatBoost's capability to handle huge datasets with mixed data types is one of its important characteristics. Also, it is extremely scalable and can be used with either a single machine or a distributed computing cluster.

Recent earthquakes have been extensively reported to have liquefaction-induced lateral spreading that has caused terrible harm to lifelines and structures. In order to forecast the occurrence of liquefaction-induced lateral spreading, Demir et al. (2023) [9] designed ML models based on XGBoost, CatBoost, and LightGBM with 6704 lateral movement records during Christchurch earthquake in New Zealand, which occurred in 2011. Performance measurements showed that PSO-CatBoost performed better than other suggested models, demonstrating its efficacy.

4.3. ***Random Forest Regression Algorithm***

Random Forest Regression is an ensemble learning technique for regression problems that integrates numerous decision trees to produce a more precise and accurate model. In Random Forest Regression, a number of decision trees are trained using random feature subsets on various subsets of the training data. Based on some of the input characteristics, each decision tree is trained to predict the output variable. The average forecast made by each individual decision tree makes up the random forest's ultimate prediction. It can handle relations between input and output variables that are not linear.

4.4. ***Decision Tree Regression Algorithm***

Decision Tree Regression is a supervised learning algorithm. By recursively splitting the input data into smaller subsets based on the values of the input parameters, it generates a model that resembles a tree with the aim of minimising the variance of the target variable in each subset. The method terminates splitting when a minimum quantity of samples or a maximum tree depth, is satisfied. The method iterates through the tree from the root node to a leaf node to anticipate a new input by comparing the input feature values to threshold values at each node. The target variable's mean or median value for the training samples that have reached the leaf node is the prediction.

4.5. ***Extra Trees Regression Algorithm***

Extra Trees Regression (or Extremely Randomized Trees Regression) is a supervised learning algorithm similar to Random Forest Regression. It is also a kind of ensemble learning technique that brings together many decision trees to generate a model that is more precise and reliable. Extra Trees Regression establishes the splitting point at random from the range of values for each feature, contrasting Random Forest Regression, which determines the splitting point for each node based on the best split among a random group of features. This unpredictability adds more variation to the decision trees, which can help them be more accurate and more generalised.

4.6. ***Bayesian Ridge Regression Algorithm***

Bayesian Ridge Regression is a type of linear regression algorithm. The model parameters are computed through Bayesian inference using this probabilistic method for linear regression. To obtain a posterior distribution, it takes a prior distribution over the parameters and updates it on the basis of the observed data. For generating predictions and determining the degree of confidence in those forecasts, this posterior distribution encapsulates the uncertainty in the estimations of the model parameters.

4.7. ***Artificial Neural Networks (ANN)***

ANN is a type of machine-learning algorithm that is inspired by the structure and operation of the human brain. The fundamental goal of an artificial neural network is to mimic how the brain processes information by constructing a network of interconnected artificial neurons. An artificial neuron is a mathematical function that takes inputs, weighs them according to a set of criteria, adds a bias term, and then uses an activation function to produce an output. The following neuron in the network processes the output of the preceding neuron. The training technique modifies the weights and biasing of each neuron using a set of input(Features)-output(Labels) pairings.

Using the ANN technique, Sreejaya et al. (2021) [33] generated a prediction framework for ground-movement intensity measures for active shallow crustal earthquakes in India. The model was composed of 659 ground motion records gathered from 138 earthquakes recorded by different seismic networks with the aim of seismic hazard analysis. Pandit et al. (2021) [27] investigated the efficacy of the ANN and ANFIS in predicting earthquake magnitude using 47 earthquake data sets gathered from 1906 to 2019 in the United States. The outcome demonstrated that ANN is superior to ANFIS in terms of earthquake magnitude prediction. The third-largest contributor to the social losses brought on by an earthquake is thought to be landslides that are generated by the earthquake [23] The Newmark sliding displacement method-based regional seismic landslide hazard assessment is utilized in this case to determine the slopes that are most susceptible to a future seismic event. In the work of Nayek et al. (2022) [23], ANN-based prediction models for Newmark's sliding displacement were constructed.

As reflected in Figure[11], we built a simple ANN model with four layers: an input layer with three neurons, a first hidden layer with four neurons, a second hidden layer with two neurons, and an output layer with one neuron. In addition to the structural aspect, there are other aspects that have an impact on ANN performance:

4.7.1. ***Optimizers***

During the training phase, cost or loss functions are minimised using optimisation techniques. Numerous optimisation techniques, including Stochastic Gradient Descent (SGD), Adam, Adagrad, RMSprop, etc., can be employed. To reduce loss or mistakes, such methods alter the neural network's weights during training.

4.7.2. ***Matrics***

Metrics are used to evaluate the neural network's performance during both training and testing. Accuracy, precision, recall, F1 score, and mean squared error (MSE) are the most popular measures. These metrics enable evaluating how effectively a neural network performs a certain dataset.

4.7.3. ***Loss***

The loss function, which is a function in mathematics, is utilized to calculate the difference between the neural network's predicted output and its actual output. The optimisation method modifies the weights of the neural network to reduce the amount of error between the expected and actual outputs based on the loss function, which is an essential part of the neural network. Mean square error (MSE), cross-entropy loss, and binary cross-entropy loss are a few examples of frequent loss functions.

4.8. ***Experimental Setup***

All data and code implementations have been performed on a computer running macOS and equipped with an apple M1 chip, 8GB of unified memory, an 8-core CPU, a 7-core GPU, and a 16-core neural engine. Together with the Jupyter notebook, we are using Python 3.8.8 as our implementation language.

5. Results

5.1. ***Machine Learning Algorithm***

In table[6], we got the variable importance matrix of a machine learning algorithm which measures how much each input variable contributes to the model's accuracy or performance. It is used to discover which features are the most

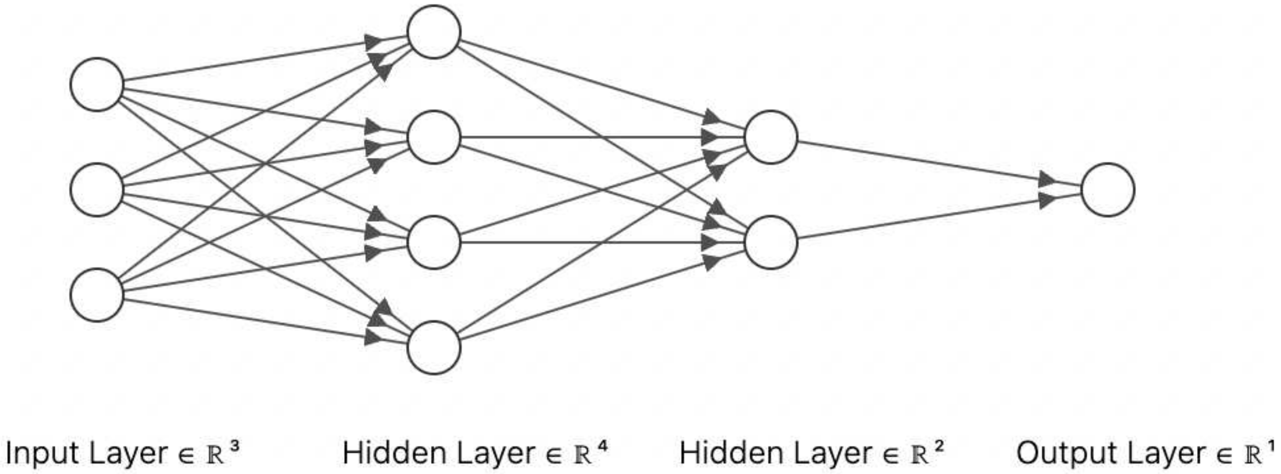


Figure 11: Sample Neural Network

pertinent in generating predictions.

Table 6: **Variable Importance for ML algorithms:**

Variable	Relative Importance	Scaled Importance	Percentage
magType	6106.4887695	1.0	0.2513940
magError	4787.5654297	0.7840128	0.1970962
Gap(gap)	4109.9545898	0.6730471	0.1692000
dmin	1569.2197266	0.2569758	0.0646022
latitude	1375.4831543	0.2252494	0.0566264
horizontal Error	1158.5717773	0.1897280	0.0476965
rms	1054.0560303	0.1726125	0.0433937
depth	729.2401123	0.1194205	0.0300216
longitude	621.0110474	0.10169698	0.0255660
depth error	472.6737061	0.0774052	0.0194592
locationSource	0.0490488	0.0000080	0.0000020

5.1.1. ***XGBoost Model***

The prediction performance of Extreme Gradient Boosting can be visualized in Figure[12] where the orange line represents the predicted values and the blue line represents the actual values. Quantitative evaluation of the model can be obtained from Table[7], which reports the following metrics: MAE of 0.2776, MSE of 0.1753, RMSE of 0.4018, R2 of -0.0923, RMSLE of 0.0629, MAPE of 0.0541, and TT of 0.1450 s. The accuracy of the model can also be assessed by examining Figure[13], where a higher number of data points lying along the x=y curve indicates better performance.

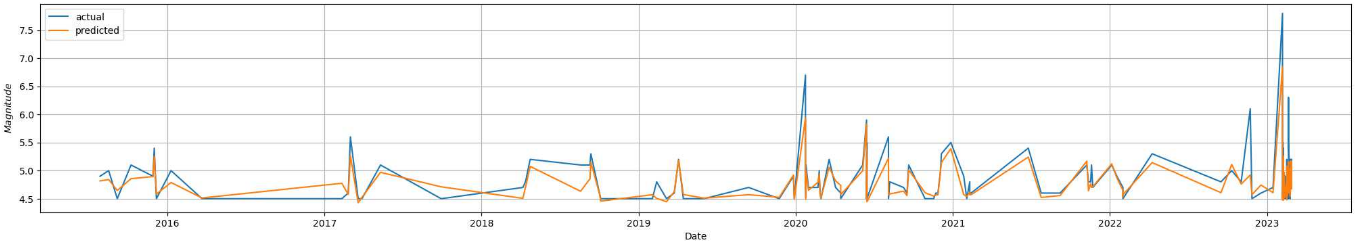


Figure 12: XGBoost Model Preciction

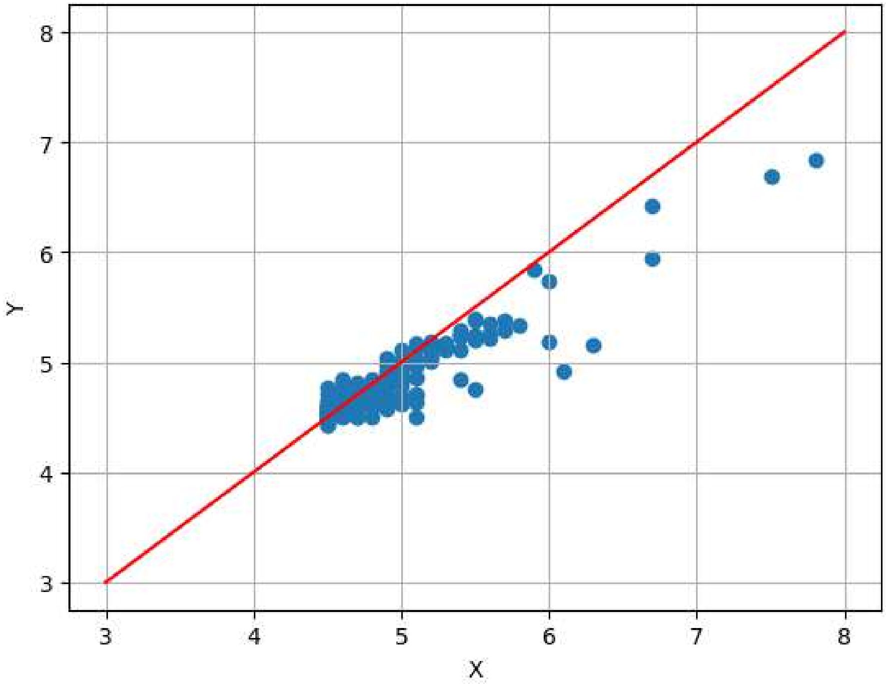


Figure 13: XGBoost model Performence

5.1.2. ***CatBoost Model***

The effectiveness of the Category Gradient Boosting model can be evaluated by comparing its predicted values, shown in orange in Figure[14], to the actual values, shown in blue. The model's performance metrics are summarized in Table[7], which reports an MAE of 0.2325, MSE of 0.1308, RMSE of 0.3483, R2 of 0.1775, RMSLE of 0.0544, MAPE of 0.0449, and TT of 1.0310 s. Figure[15] illustrates the model's accuracy, with a higher concentration of data points lying along the x=y curve indicating greater accuracy.

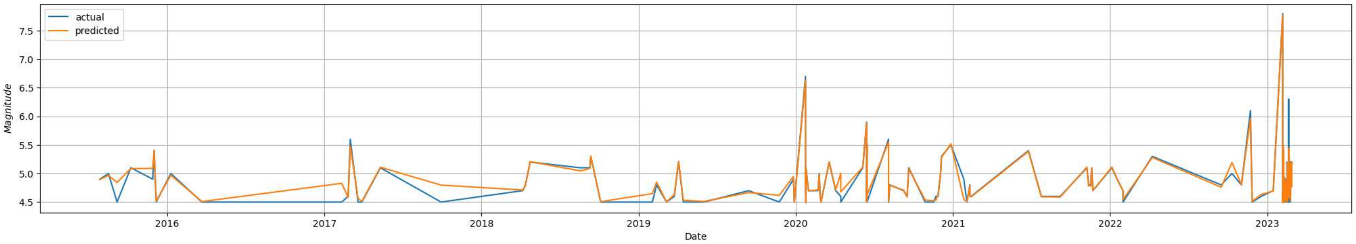


Figure 14: CatBoost Model Preciction

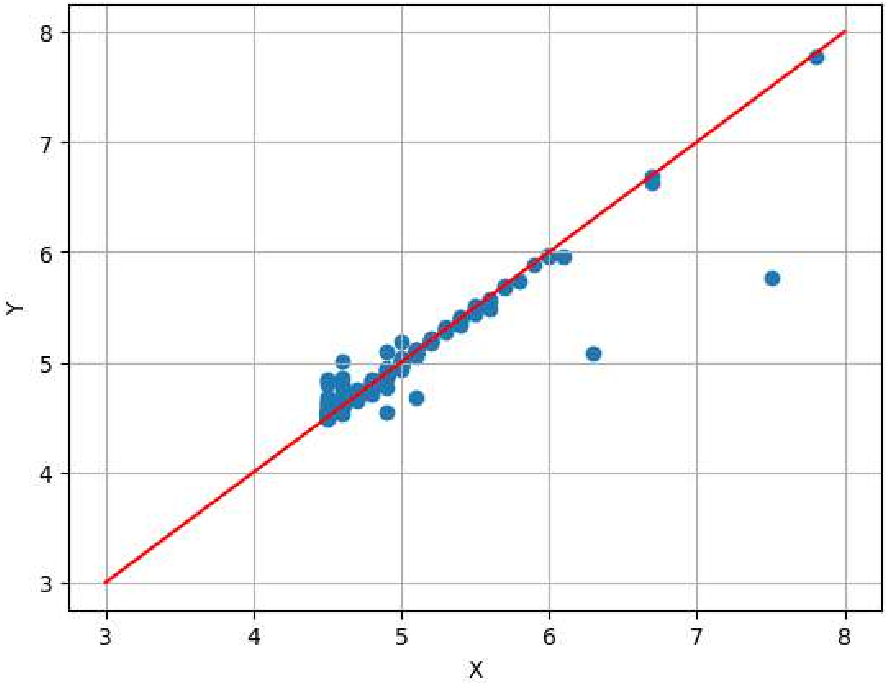


Figure 15: CatBoost Model Performance

5.1.3. ***Random Forest Model***

The performance of the Random Forest regression model can be observed in Figure[16], where the predicted values are represented by the orange line and the actual values by the blue line. The model's evaluation metrics can be obtained from Table[7], which shows an MAE of 0.2493, MSE of 0.1405, RMSE of 0.3539, R2 of 0.1803, RMSLE of 0.0553, MAPE of 0.0486, and TT of 0.1880 s. The model's accuracy can also be assessed by examining Figure[16], where a higher number of data points lying along the x=y curve indicates better performance.

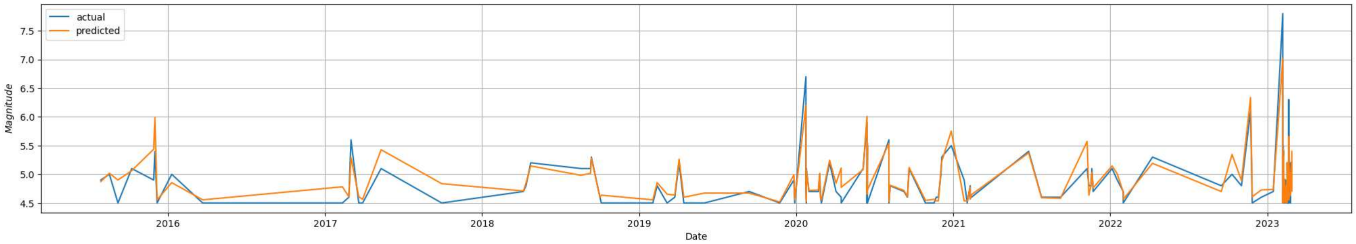


Figure 16: Random Forest Model prediction

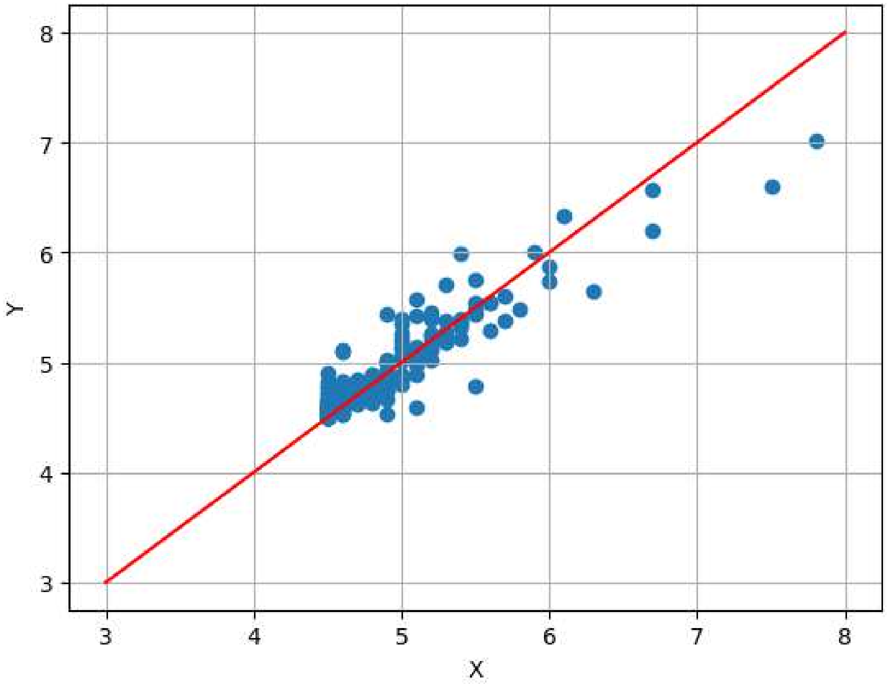


Figure 17: Random Forest Model Performance

5.1.4. ***Extra Trees Model***

The Extra Trees regression model's prediction performance can be observed in Figure [18], where the orange line represents the predicted values and the blue line represents the actual values. The model's evaluation metrics can be obtained from Table [7], which shows an MAE of 0.2252, MSE of 0.1264, RMSE of 0.3365, R2 of 0.1681, RMSLE of 0.0520, MAPE of 0.0434, and TT of 0.1750 s. The accuracy of the model can also be assessed by examining Figure [19], where a higher number of data points lying along the x=y curve indicates better performance.

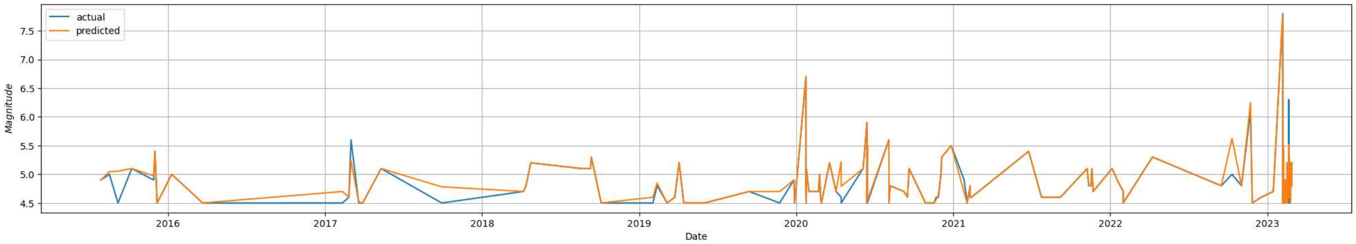


Figure 18: Extra Trees Model Prediction

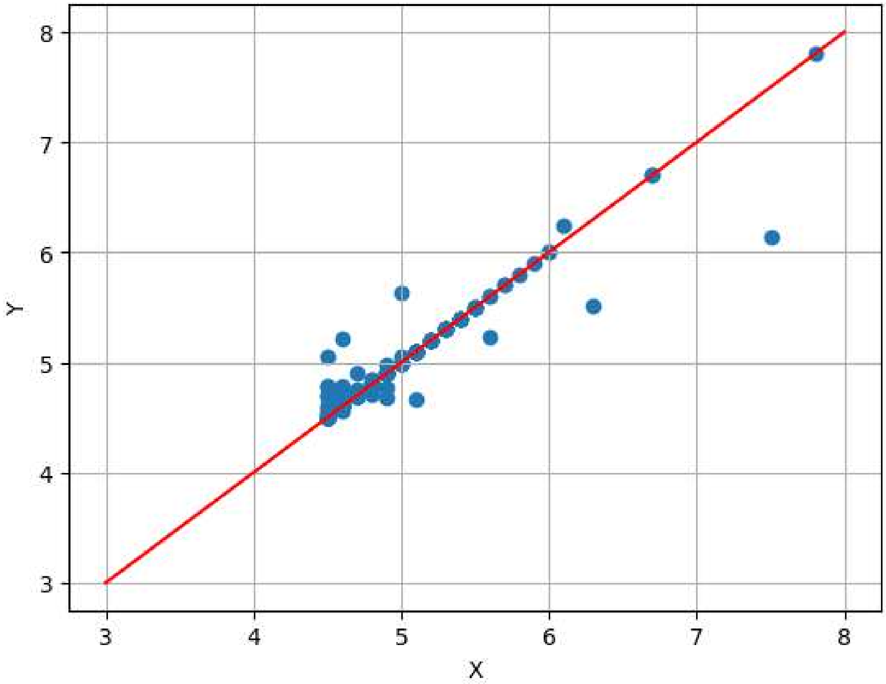


Figure 19: Extra Trees Model Performance

5.1.5. ***Decision Tree Model***

The Decision Tree regression model's prediction performance can be observed in Figure [22], where the orange line represents the predicted values and the blue line represents the actual values. The model's evaluation metrics can be obtained from Table [7], which shows an MAE of 0.3537, MSE of 0.3460, RMSE of 0.5350, R2 of -1.1420, RMSLE of 0.0815, MAPE of 0.0693, and TT of 0.0210 s. The accuracy of the model can also be assessed by examining Figure [23], where a higher number of data points lying along the x=y curve indicates better performance.

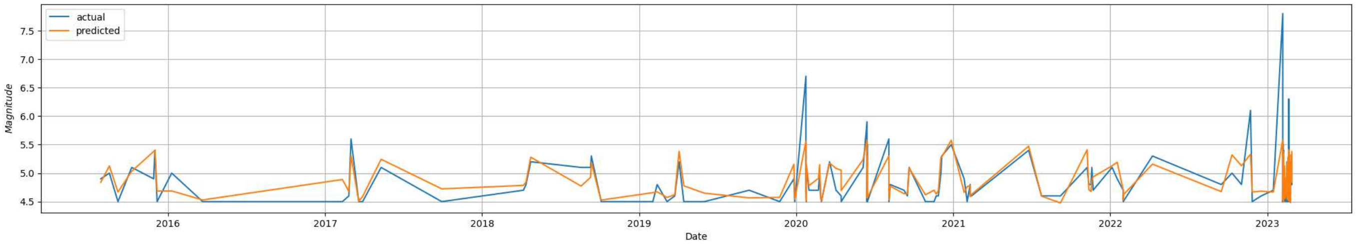


Figure 20: Decision Tree Model Prediction

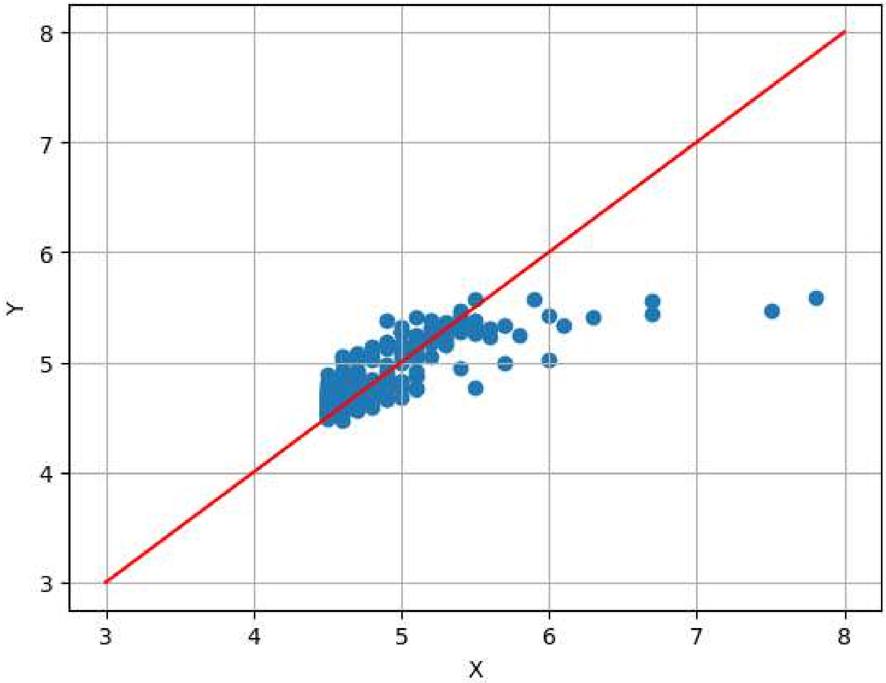


Figure 21: Decision Tree Model Performance

5.1.6. ***LGBM Model***

The Light Gradient Boosting (LGBM) model's prediction performance can be observed in Figure [22], where the orange line represents the predicted values and the blue line represents the actual values. The model's evaluation metrics can be obtained from Table [7], which shows an MAE of 0.2466, MSE of 0.1343, RMSE of 0.3513, R2 of 0.1491, RMSLE of 0.0555, MAPE of 0.0483, and TT of 0.1330 s. The accuracy of the model can also be assessed by examining Figure [23], where a higher number of data points lying along the x=y curve indicates better performance.

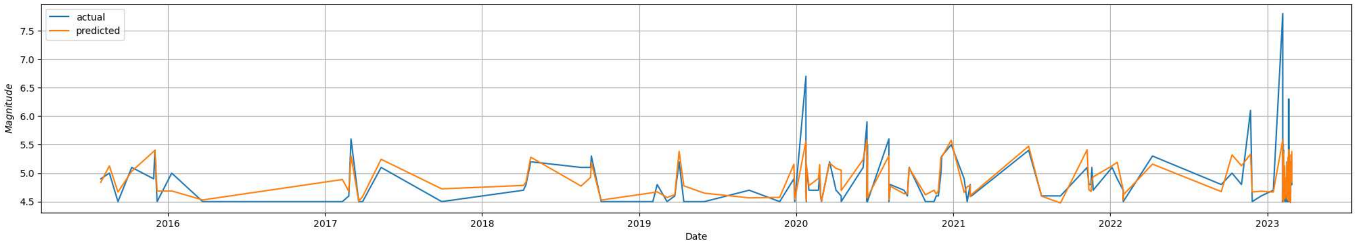


Figure 22: LGBM Model Prediction

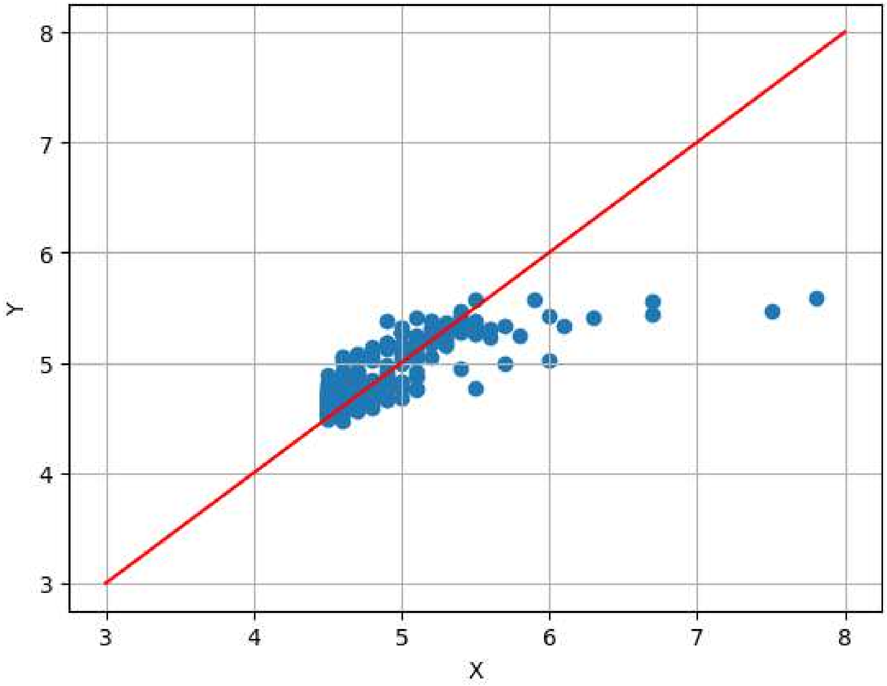


Figure 23: LGBM Model Performance

5.1.7. ***Bayesian Ridge Model***

The prediction made by the Bayesian Ridge regression model is visualized in Figure [24] with the predicted values plotted in orange and the actual values shown in blue. The evaluation metrics for the model, as shown in Table [7], include an MAE of 0.2367, MSE of 0.1461, RMSE of 0.3607, R2 of 0.2096, RMSLE of 0.0559, MAPE of 0.0458, and a TT of 0.0230 s.

The model's accuracy can also be assessed by observing the scatter plot in Figure [25], where the ideal scenario would be to have most of the data points lying along the x=y line. Therefore, a larger number of data points closer to the line indicates better performance of the model.

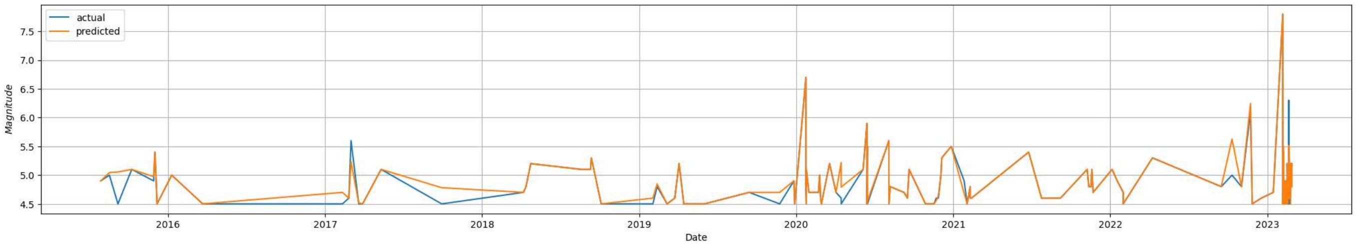


Figure 24: Bayesian Ridge Model Prediction

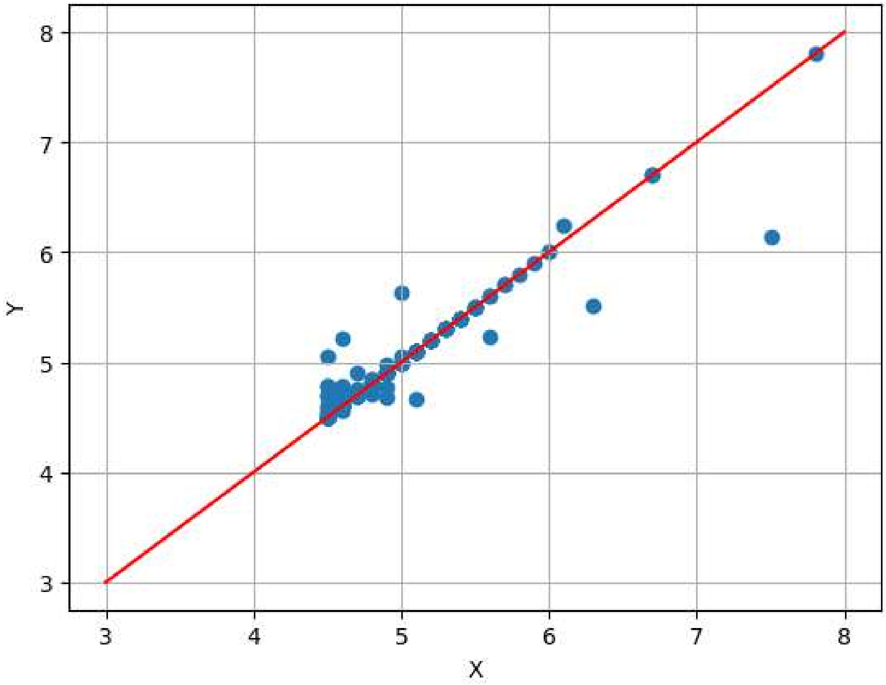


Figure 25: Bayesian Ridge Model Performance

Table 7: **Machine Learning Models Result**

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Seconds)
Extra Trees Regressor	0.2252	0.1264	0.3365	0.1681	0.0520	0.0434	0.1750
CatBoost Regressor	0.2325	0.1308	0.3483	0.1775	0.0544	0.0449	1.0310
Light Gradient Boosting	0.2466	0.1343	0.3513	0.1491	0.0555	0.0483	0.1330
Random Forest Regressor	0.2493	0.1405	0.3539	0.1803	0.0553	0.0486	0.1880
Bayesian Ridge	0.2367	0.1461	0.3607	0.2096	0.0559	0.0458	0.0230
(XGBoost) Extreme Gradient Boosting	0.2776	0.1753	0.4018	-0.0923	0.0629	0.0541	0.1450
Decision Tree Regressor	0.3537	0.3460	0.5350	-1.1420	0.0815	0.0693	0.0210

Table[8,9] lists the details of the ML model that we employed in our research, with its default parameters to the best parameters that we have used.

Table 8: **Parameter Settings for ML Models**

Model	Parameters	Best parameters
DecisionTreeRegressor	criterion = ’squared-error’, splitter = ’best’, max-depth = None, min-samples-split = 2, min-samples-leaf = 1, random_state=None	criterion = ’squared-error’, splitter = ’best’, max_depth = None, min_samples_split = 2, min_samples_leaf = 1, random_state = 42.
RandomForestRegressor	n_estimators = 100, criterion = ’squared-error’, max_depth = None, min_samples_split = 2, min_samples_leaf = 1, max_features = ’auto’, bootstrap = True, oob_score = False, n_jobs = None, random_state = None, verbose = 0, warm_start = False	n_estimators = 11, criterion=’squared-error’, max_depth = None, min_samples_split = 2, min_samples_leaf = 1, max_features = ’auto’, bootstrap = True, oob_score = False, n_jobs = None, random_state = 42, verbose = 0, warm_start=False.
XGBoost	objective = None, n_estimators = 1000, seed = None	objective = ’reg:linear’, n_estimators = 12, seed = 42

5.2. *Artificial Neural Networks (ANN)*

ANNs are a subset of sophisticated machine learning algorithms made up of numerous layers of interconnected neurons that cooperate to carry out intricate calculations. We illustrate a small NN in Figure [26] as an example. As illustrated in Figure [27], we created three different neural network topologies for our investigation, each with five layers. Twelve neurons are present in the dense input layer, which is the top layer. Each neuron receives a single parameter (features) from the dataset. The next layer is a Dense layer with 32 neurons, which are extensively connected with the neurons from the input layer. We use the ’relu’ activation function in this layer to introduce non-linearity in the network. The subsequent layer is another Dense layer with 16 neurons, which are extensively connected with the previous layer neurons. Again, we use the ’relu’ activation function to introduce non-linearity. The third layer is another Dense layer with 8 neurons, which are extensively connected with the previous layer neurons. We use ’relu’ as the activation function. The fourth layer is a Dense layer with 4 neurons, which are extensively connected with the previous layer neurons. We use

Table 9: **Parameter Settings for ML Models**

Model	Parameters	Best parameters
LGBM	boosting_type: str = 'gbdt', num_leaves: int = 31, max_depth: int = -1, learning_rate: float = 0.1, n_estimators: int = 100, subsample_for_bin: int = 200000, objective: Union[str, Callable, NoneType] = None, class_weight: Union[Dict, str, NoneType] = None, min_split_gain: float = 0.0, min_child_weight: float = 0.001, min_child_samples: int = 20, subsample: float = 1.0, subsample_freq: int = 0, colsample_bytree: float = 1.0, reg_alpha: float = 0.0, reg_lambda: float = 0.0, random_state: Union [int, numpy.random.mtrand.RandomState, NoneType] = None, n_jobs: int = -1, silent: Union[bool, str] = 'warn', importance_type: str = 'split'	boosting_type: str = 'gbdt', num_leaves: int = 31, max_depth: int = -1, learning_rate: float = 0.3, n_estimators: int = 11, subsample_for_bin: int = 200000, objective: Union[str, Callable, NoneType] = None, class_weight: Union[Dict, str, NoneType] = None, min_split_gain: float = 0.0, min_child_weight: float = 0.001, min_child_samples: int = 20, subsample: float = 1.0, subsample_freq: int = 0, colsample_bytree: float = 1.0, reg_alpha: float = 0.0, reg_lambda: float = 0.0, random_state: 42, n_jobs: int = -1, silent: Union[bool, str] = 'warn', importance_type: str = 'split'
CatBoost	iterations=None, learning_rate=None, loss_function='RMSE'	iterations=1000, learning_rate=0.031996, loss_function='RMSE'(0.387631)

'relu' as the activation function in this layer as well. Finally, the output layer is a Dense layer with 1 neuron, which is extensively connected with the previous layer neurons. The output layer provides the final result of the computation performed by the neural network. The optimizer used in each of the three neural network models we developed is different.

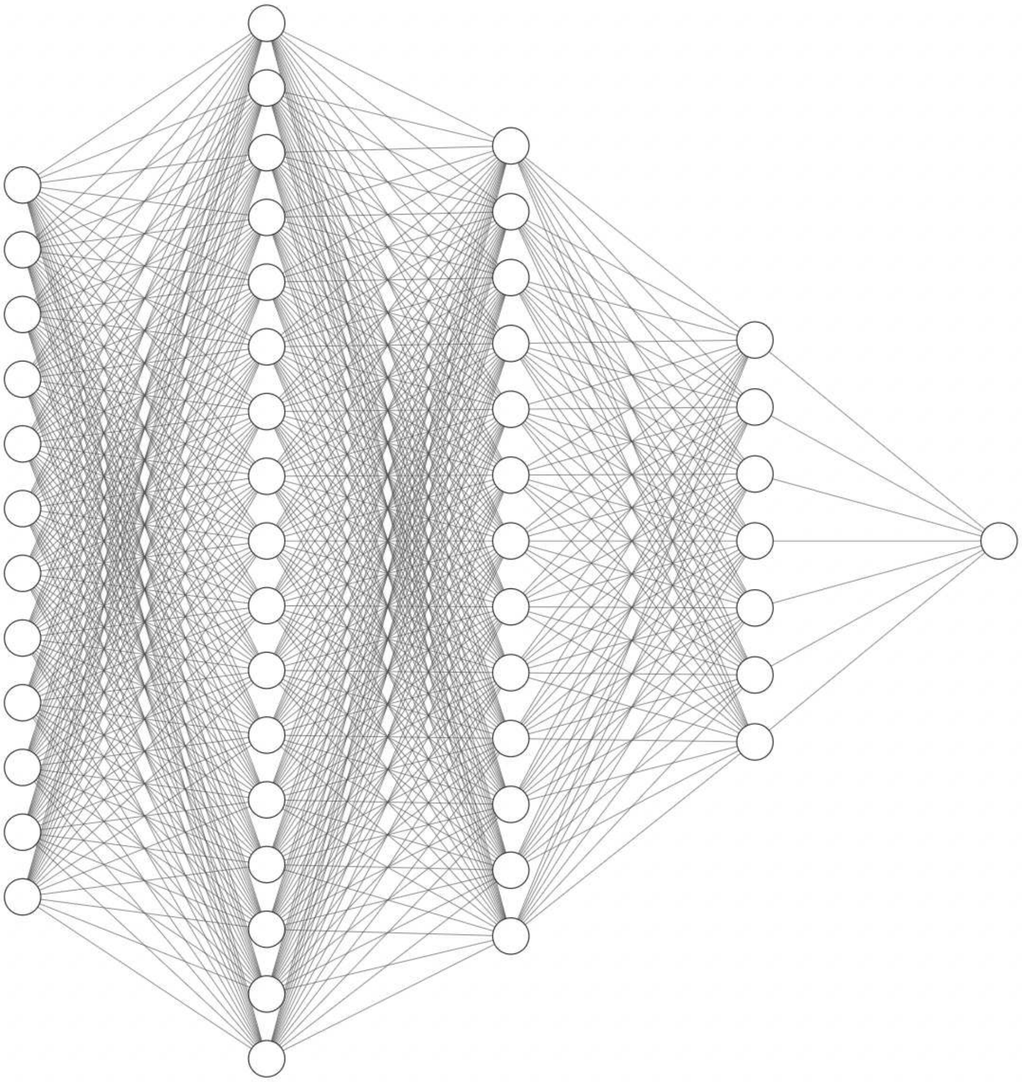


Figure 26: Neural Network

5.3. ***ANN-1***

The neural network architecture for ANN-1 is the same as that in Figure [27]. After defining all five layers of the neural network, we constructed it using the 'RMSprop' optimizer, 'Mean Absolute Error' as the loss function, and 'Mean Squared Error' as the evaluation metric. Our neural network architecture and optimizer are in place, and we are prepared to train the model and make predictions.

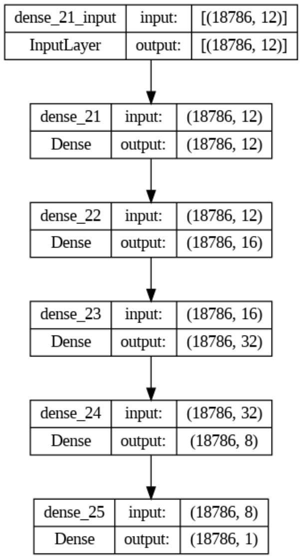


Figure 27: Neural Network Layers for ANN Models

5.3.1. ***Evaluation of ANN-1***

We trained ANN-1 using 100 iterations and monitored the loss function (MAE) and evaluation metric (MSE) with each iteration. Figure [28] shows the trend of MAE and MSE over the 100 iterations. After 100 iterations, we achieved a loss (MAE) of 0.02880, MSE of 0.00183, RMSE of 0.02918, R2 of 0.99999, MAE of 0.02918, and NRMSLE of 0.07120. The fact that our ANN-1 model's loss (MAE) and MSE for both training and validation are nearly identical suggests that it is not overfitting or underfitting the data points. Thus, we may conclude that ANN-1 performs very well with the provided data.

In Figure [29], we can see the performance of the ANN-1 model for the given data. According to the investigations, the more points that lie along the x=y (red line), the more accurate the model is (in Figure [29]). Figure [30] displays the error function profile (Predicted value vs Actual value) for ANN-3. This graph is useful in determining the profile of the error function.

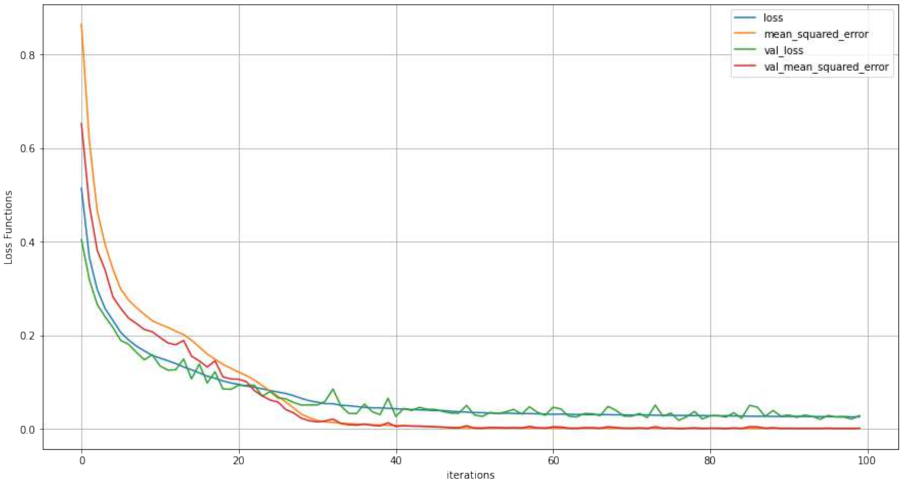


Figure 28: Loss function vs iterations for ANN-1

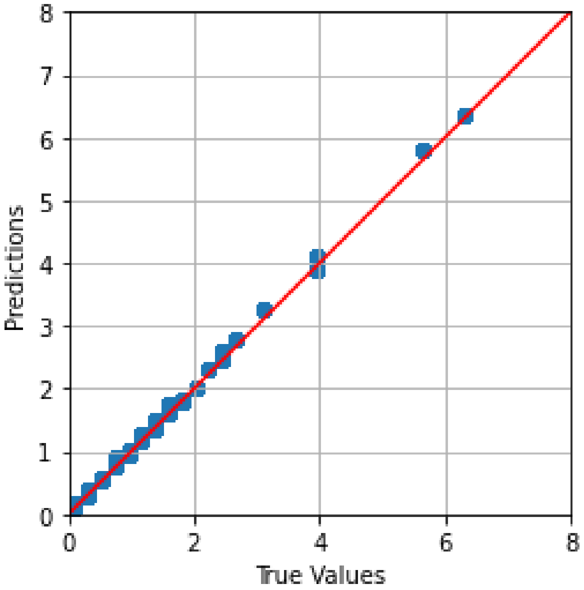


Figure 29: Performance of ANN-1 Model

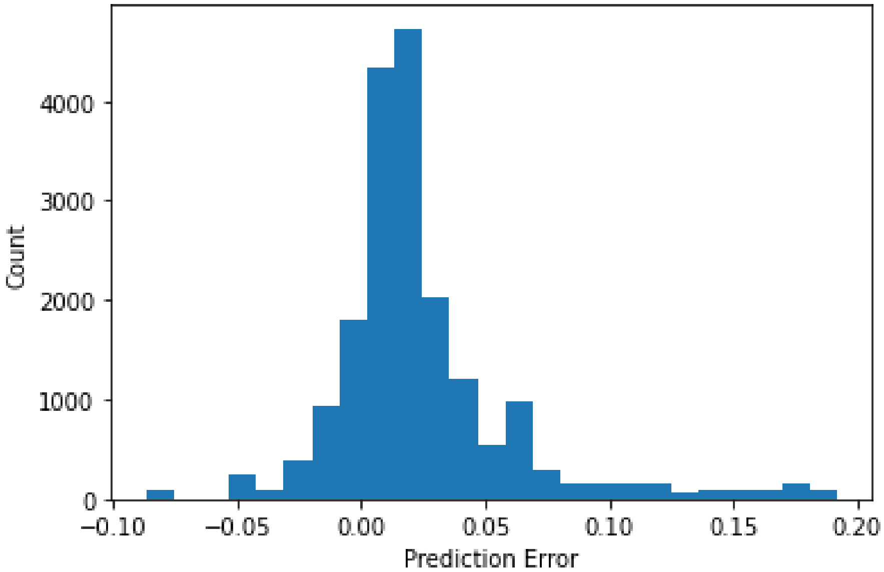


Figure 30: Error(Predicted Value - True Value) for ANN-1 Model

5.4. ***ANN-2***

We employed the same neural network architecture for ANN-2 as in Figure [27]. Using the "Adam" optimizer, "Mean Absolute Error" as the loss function, and "Mean Squared Error" as the evaluation metric, we constructed the neural network after specifying each of its five layers. Our ANN-2 model is now prepared for prediction and training.

5.4.1. ***Evaluation of ANN-2***

For the training of ANN-2, we used 100 iterations. The graph in Figure [31] shows the trend of the loss function (MAE) and metrics (MSE) with each iteration. At the end of training, we achieved a loss (MAE) of 0.02596 and metrics (MSE) of 0.01284. Additionally, we obtained MSE = 0.01369, RMSE = 0.02659, R2 = 0.99999, MAE = 0.02659, and NRMSLE = 0.05619.

The loss (MAE) and 'MSE' values for both training and validation are almost equal, indicating that the model is not underfitting or overfitting the data points. Therefore, ANN-2 is showing good performance. Figure [32] displays the performance of the ANN-2 model for the given data. The more points that lie along the x=y line (red line), the more accurate the model. Note that the values in the graph are not scaled. Figure [33] shows the error function profile (predicted value - actual value) of the ANN-2 model. We can see that there is a uniform distribution with a precision of 0.01, indicating that the model's predictions are accurate. Figure [33] displays the error function profile (Predicted value - Actual value) for ANN-3. This graph is useful in determining the profile of the error function.

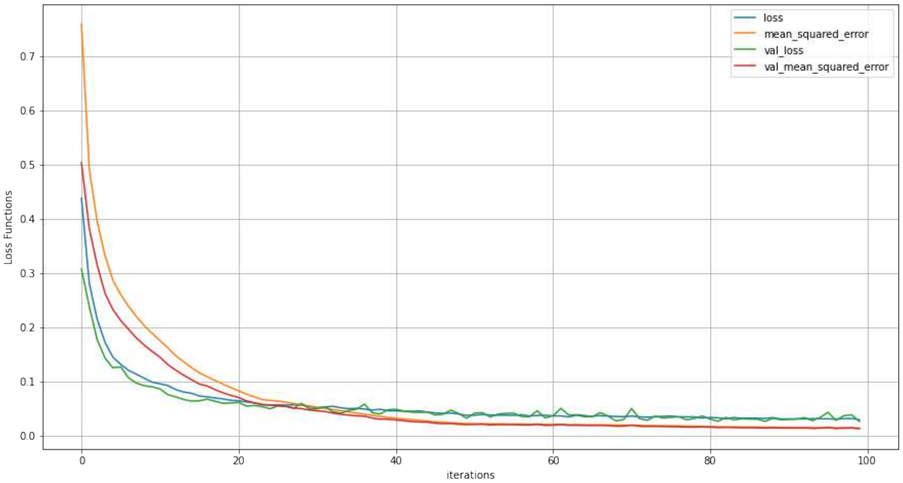


Figure 31: Loss function vs iterations for ANN-2

5.5. ***ANN-3***

For ANN-3, we used the same neural network architecture as shown in Figure [27]. After defining all five layers of the neural network, we compiled it using the 'SGD' optimizer, 'Mean Absolute Error' as the loss function, and 'Mean Squared Error' as the evaluation metric. Our ANN-3 model is now ready for training and prediction.

5.5.1. ***Evaluation of ANN-3***

To train ANN-3, you have used 100 iterations, and the resulting loss (MAE) and metrics (MSE) values for both the training and validation sets are shown in Figure [34]. At the end of 100 iterations, the final loss (MAE) value is 0.17620, and the final metrics (MSE) value is 0.12055. Additionally, the RMSE, R2, MAE, and NRMSLE values are 0.18148,

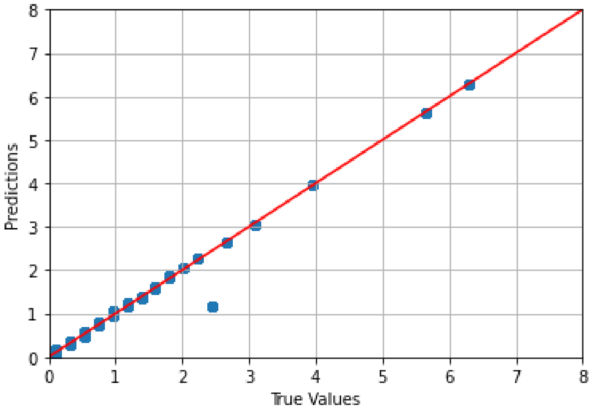


Figure 32: Performance of ANN-2 Model

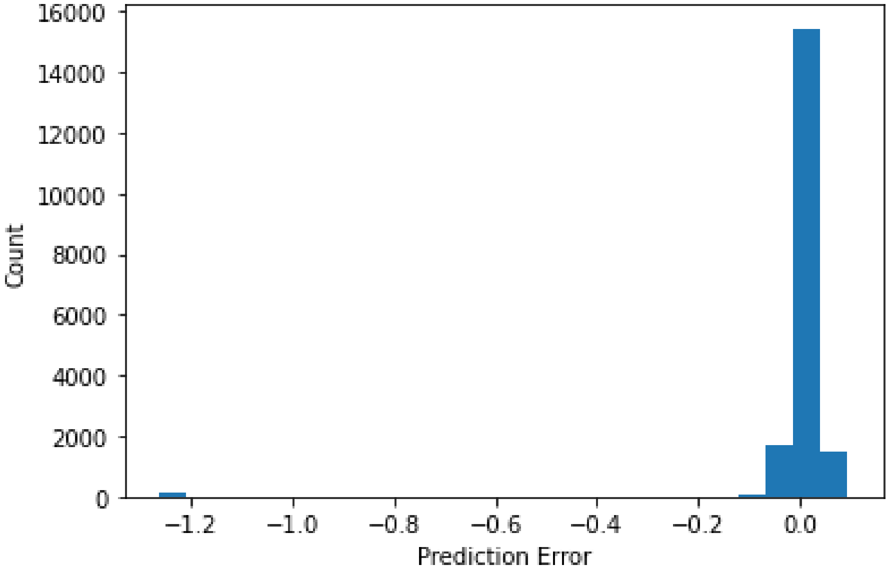


Figure 33: Error(Predicted Value - True Value) for ANN-2 Model

0.99999, 0.18148, and 0.44657, respectively. These results indicate that ANN-3 is performing well on the given dataset, as the loss and metrics values for both the training and validation sets are similar, suggesting that the model is not overfitting or underfitting the data.

Figure [35] shows the performance of the ANN-3 model on the given data, where the more points lie along the x=y (red line), the more accurate the model. It is important to note that the values in the graph [35] are not scaled. Overall, the results suggest that ANN-3 is a good model for the given dataset and can be used for further analysis or decision-making. Figure [36] displays the error function profile (Predicted value - Actual value) for ANN-3. This graph is useful in determining the profile of the error function.

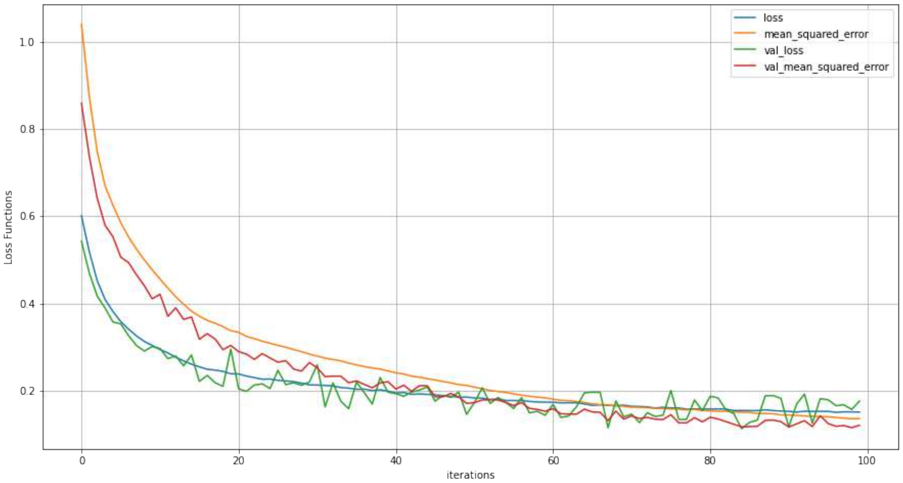


Figure 34: Loss function vs iterations for ANN-3

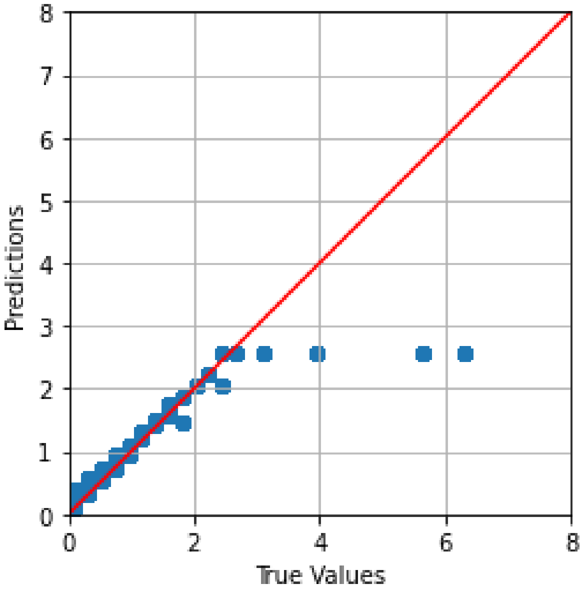


Figure 35: Performance of ANN-3 Model

5.6. *Mapping*

An earthquake's magnitude is a measurement of the energy released during the occurrence. The Richter scale or the ”moment magnitude scale” is typically used to express it. While the moment magnitude scale considers the area of the

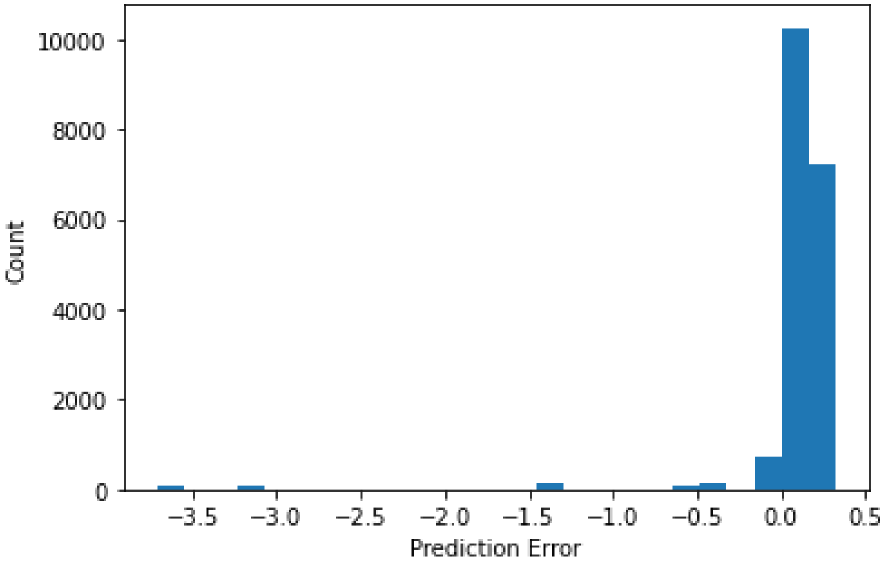


Figure 36: Error(Predicted Value - True Value) for ANN-3 Model

Table 10: **Artificial Neural Network Models Result**

ANN Model	Optimizer	loss = MAE	matrics = MSE	MSE	RMSE	R2	MAE	NRMSLE
ANN-1	RMSprop	0.02880	0.00183	0.00188	0.02918	0.99999	0.02918	0.07120
ANN-2	Adam	0.02596	0.01284	0.01369	0.02659	0.99999	0.02659	0.05619
ANN-3	SGD	0.17620	0.12055	0.14055	0.18148	0.99999	0.18148	0.44657

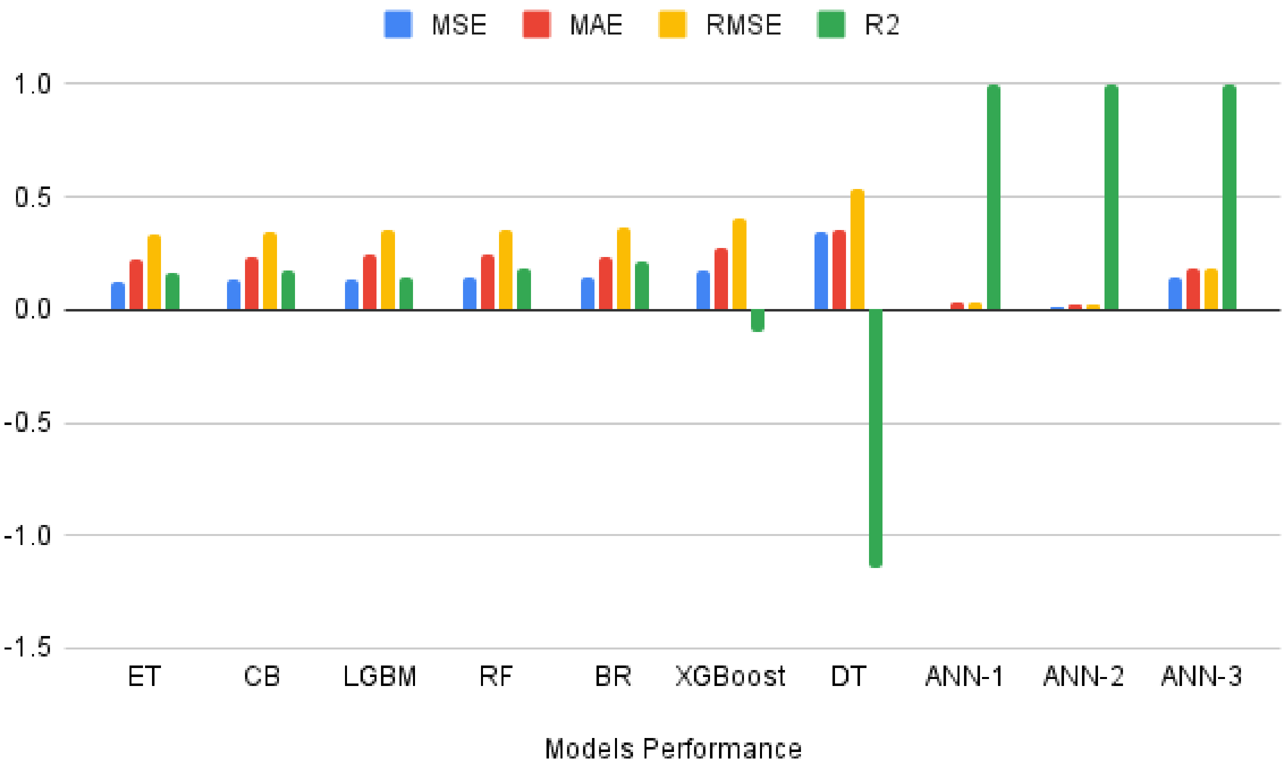


Figure 37

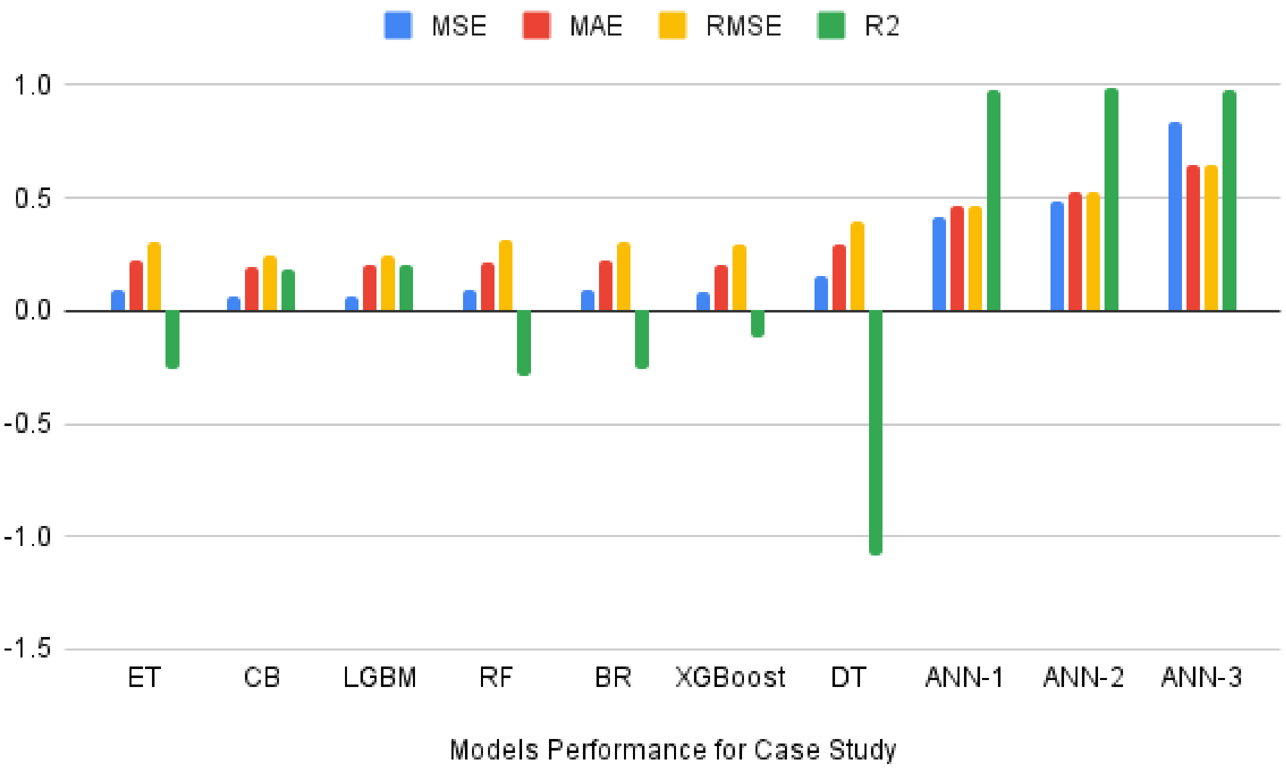


Figure 38

ruptured fault as well as the amount of slip and the strength of the rock that broke during the earthquake, the Richter scale is a logarithmic scale that assigns a single number to quantify the seismic energy released by an earthquake. Based on geological information, past earthquake activity, and other pertinent characteristics, seismic maps are used to pinpoint regions at high risk of earthquakes. The severity levels used in the mapping of earthquake hazards normally vary from 1 (extremely low) to 10 (extremely high). The magnitude of the earthquake, the distance from the epicentre, the depth of the earthquake, the local geology and topography, and the vulnerability of the built environment and population are some of the elements used to determine the severity levels. In order to create maps with high precision and adaptability, cartography, a Python library, has been used. Apart from that, we interpolated seismic epicentres using the Scipy Python programme. The interpolation tool allows method = cubic, which is responsible for all of the contouring structures on the map.

The visual representation based on the magnitude of the earthquake is shown in Figure [39]. The levels of severity used in the mapping operated from 4 (mild and innocuous) to 8 (dangerous and destructive). The most damaging earthquakes have occurred in the areas highlighted in dark red. The majority of these areas are located along the East Anatolian Fault and extend up to the junction of the East Anatolian Fault and the North Anatolian Fault. The East Anatolian Fault is a major fault system in Turkey that runs from the Armenian border in the east to the Aegean Sea in the west. The fault is a result of the collision between the Arabian and Eurasian plates, which has been ongoing for millions of years. The fault is capable of producing large and damaging earthquakes, as evidenced by the historical record. The North Anatolian Fault is another major fault system in Turkey that runs parallel to the East Anatolian Fault, but further to the north. The fault is also capable of producing large and damaging earthquakes, and has been responsible for several catastrophic earthquakes in the past.

On the earthquake magnitude map, the areas highlighted in dark red represent the locations of the most damaging earthquakes. These areas are frequently characterised by strong strain accumulation along the fault zones and high levels of seismic activity. As a result of the interaction between the tectonic forces causing the faulting and the local geology and terrain, powerful and destructive earthquakes frequently occur in these areas. It's critical to take into account both the earthquake's magnitude and the level of risk it poses to infrastructure and humans when determining where to place earthquake hazards. A number of variables, including the distance from the epicenter, the strength of the ground shaking, and the vulnerability of the built environment to damage, are commonly used to determine the danger zones in earthquake hazard mapping. Figure [40] provides a visual representation of the danger zones for earthquakes, which is an expansion of Figure [39]. The mapping is based on the levels of severity used in the mapping, which operate from 1 (red danger zone) to 4 (low intensity). The red danger zone represents areas where the earthquake poses the greatest risk to life and infrastructure, while the green zone represents areas that are safer and less likely to experience significant damage.

In earthquake hazard mapping, the danger zones are typically established based on historical data and scientific analysis. The aim is to identify areas that are at higher risk of experiencing strong ground shaking and other earthquake-related hazards, such as liquefaction and landslides. By identifying these areas, policymakers, urban planners, and emergency managers can take steps to mitigate the potential impact of earthquakes on people and infrastructure. It is worth noting that earthquake hazard mapping is an ongoing process that requires regular updates based on new data and analysis. As our understanding of the seismic hazards and the built environment improves, the danger zones may need to be adjusted to reflect the latest information. It is, therefore, essential to maintain a continuous and comprehensive monitoring system to keep the hazard maps up-to-date and to inform the decision-making process. Seismic maps can be a valuable resource for identifying earthquake-prone regions, and it's crucial to take the appropriate precautions to make sure that these regions have the infrastructure, food supplies, and disaster relief services needed to minimize the potential effects of seismic activity.

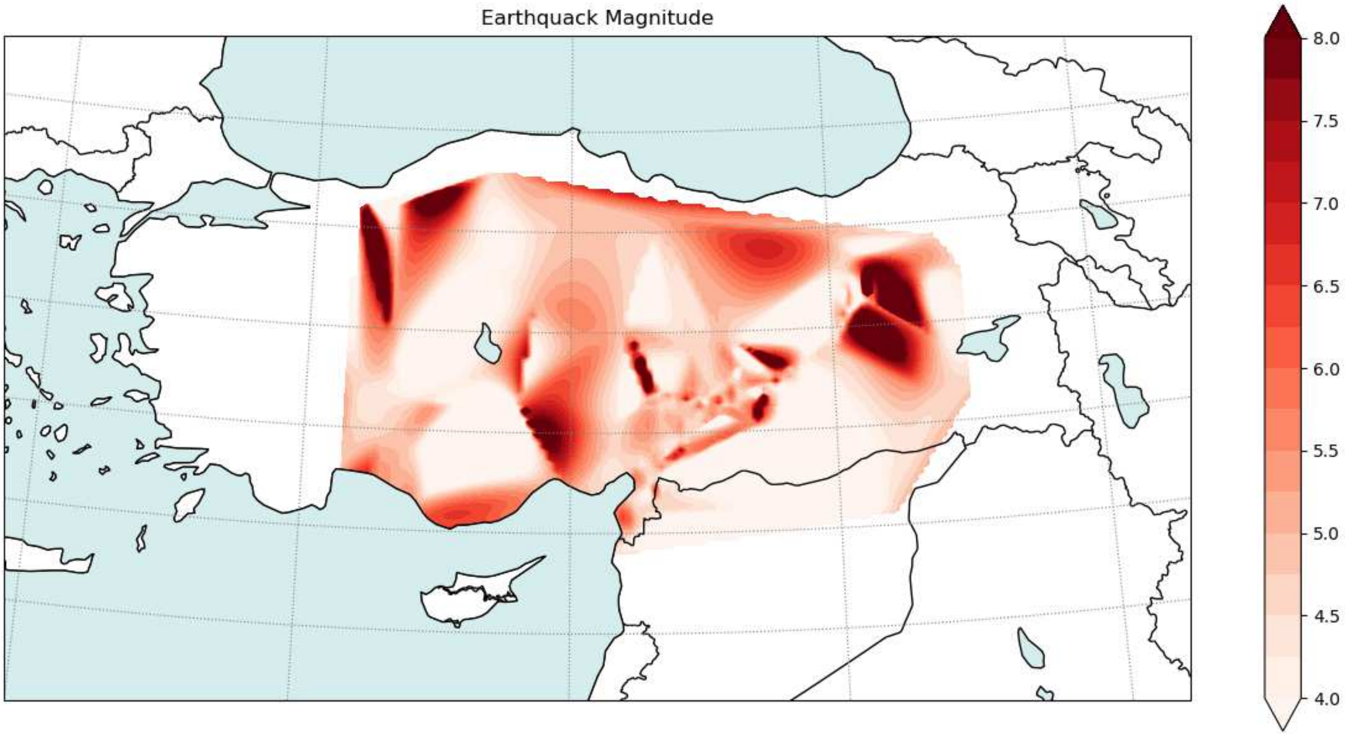


Figure 39: Earthquake Magnitude mapping

6. Case Study

We used information collected from 01-03-2023 to 31-03-2023 for the case study. The data contains the same latitude, longitude, earthquake magnitude (mag), depth (depth), gap (gap), minimum distance, root mean square (rms), horizontal error (horizontalError), depth error (depthError), magnitude error (magError), magnitude stations (magNst), magnitude types (magType-n), and location source (locationSource-n) parameters that we used to train all three ANN and seven ML models. We then analysed the performance of each of the following ML and ANN models before generating a map based on the forecasts provided by the models.

6.1. Machine Learning Models

The performance of the ML models for the case study data is given in Table [11]. The ML models are performing really well on the case study data, which is a great indication that our model is very accurately predicting the magnitude of earthquakes and giving valuable data for Seismic mapping. The visualization in Figure[41] illustrates the performance of the ML models applied to the case study data. A higher concentration of points along the diagonal x=y line indicates better performance of the models. Figure [42] provides a comprehensive display of both actual and predicted earthquake

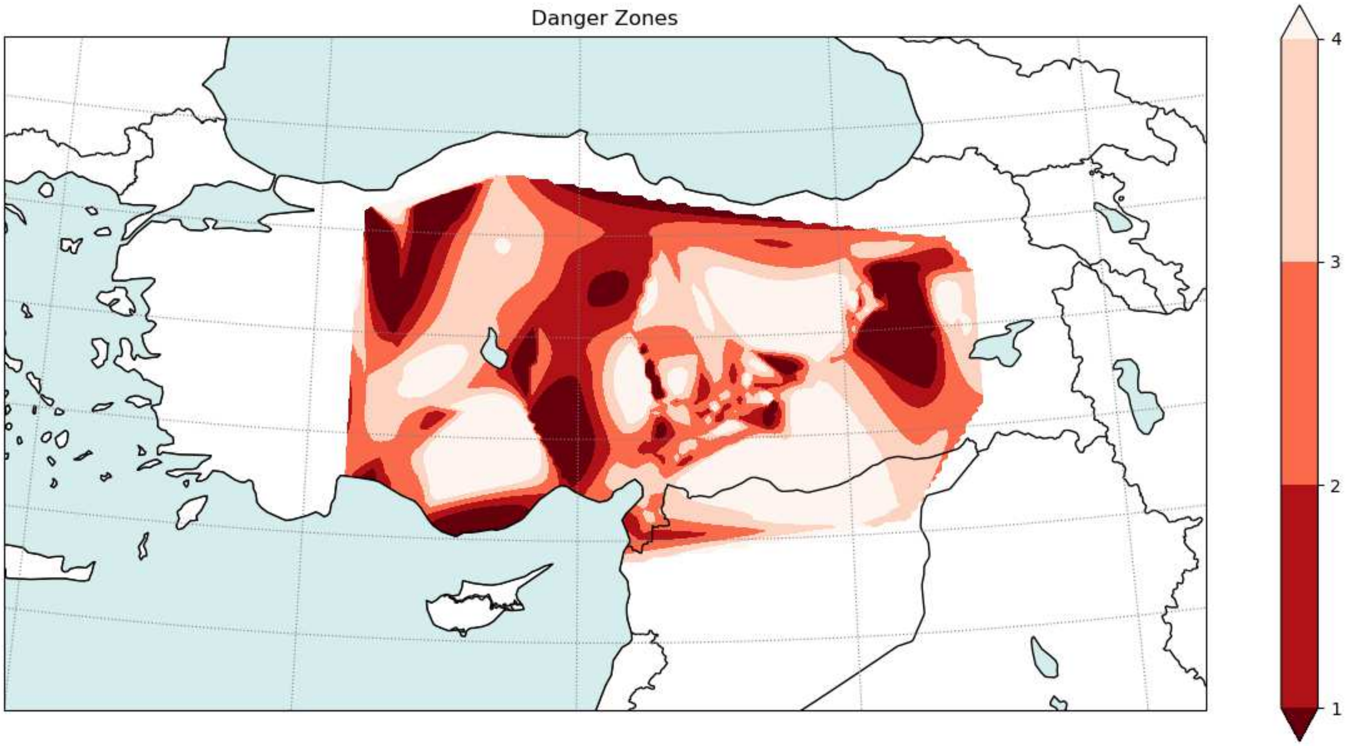


Figure 40: Earthquake danger zone mapping

magnitudes, along with their corresponding dates. This information can aid in the evaluation of the predictive performance of earthquake forecasting models.

In Figure[43], we utilized the predicted magnitude data obtained from the case study to create an earthquake magnitude map. The original data map is displayed in the first mapping, while the corresponding maps for various ML models are also shown. It is noteworthy that XGBoost, CatBoost, and LGBM ML models exhibit exceptional performance in this regard.

Table 11: **Machine Learning Models Result for Case Study**

Model	MSE	RMSE	R2	MAE
Extra Trees Regressor	0.0957	0.3094	-0.2601	0.2224
CatBoost Regressor	0.0620	0.2491	0.1834	0.1984
Light Gradient Boosting	0.0607	0.2464	0.2010	0.1997
Random Forest Regressor	0.0978	0.3128	-0.2877	0.2127
Bayesian Ridge	0.0957	0.3094	-0.2601	0.2224
(XGBoost) Extreme Gradient Boosting	0.0850	0.2910	-0.1195	0.2055
Decision Tree Regressor	0.1586	0.3983	-1.0877	0.2933

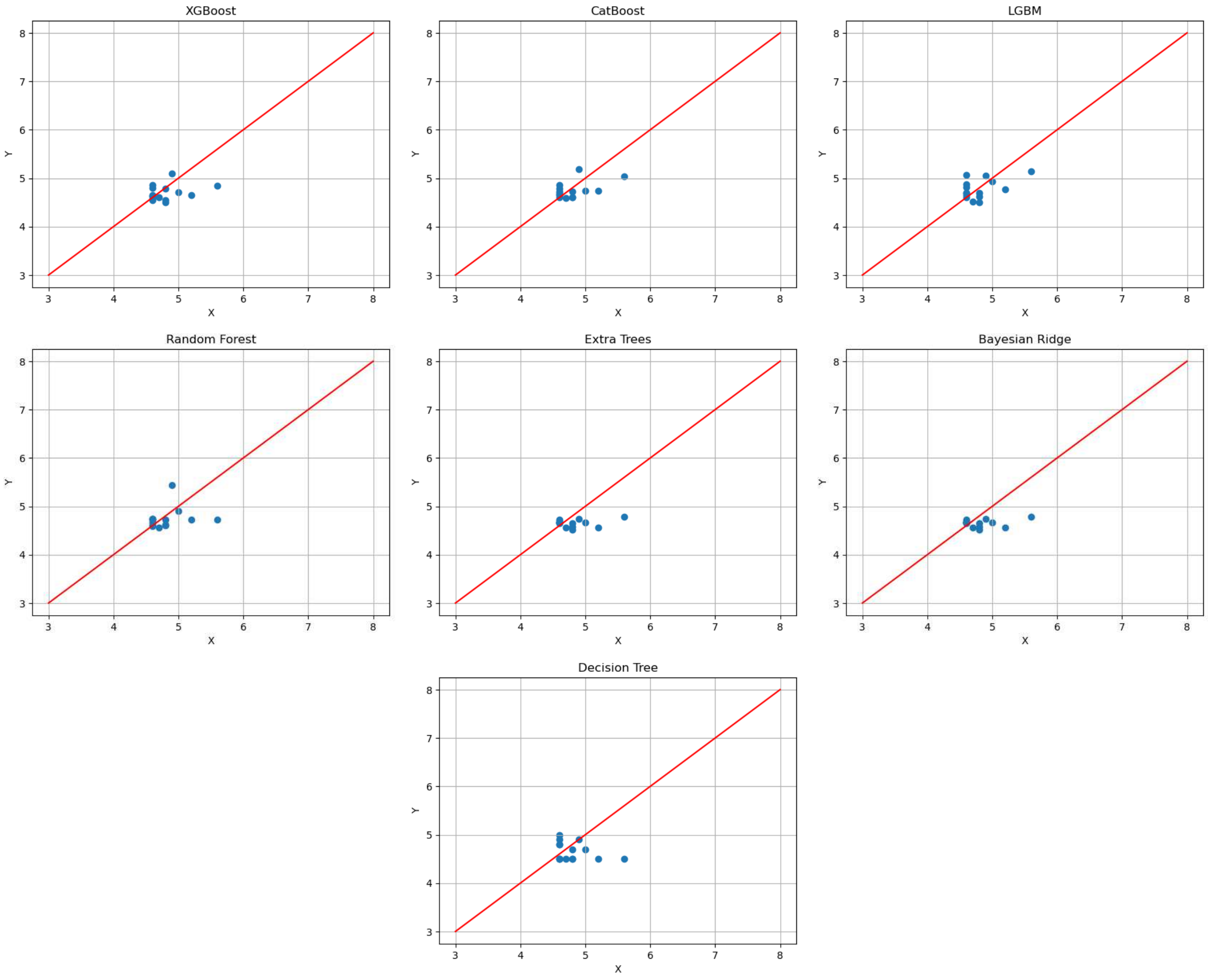


Figure 41: ML models performance for the Case Study

6.2. *Artificial Neural Networks*

The ANN models are performing really well on the case study data, which is a great indication that our models are very accurately predicting the magnitude of earthquakes and giving great data for Seismic mapping. The performance of the ANN models for the case study data is given in Table [12].

The coefficients of determination (R2) for all of the ANN models are quite high, with values of 0.98012 for ANN-1, 0.98710 for ANN-2, 0.97847 for ANN-3. A high R2 value indicates that the model is capable of describing an important component of the variance in the target variable using the input variables. It means that the model conveys the relationships between the variables accurately and is a decent match for the data. Furthermore, we observe that the error (MAE, MSE, RMSE) is greater for the case study data than for the training data. However, the prevalence of error is very low, demonstrating that the ANN models are performing very well. From Table[12] NRMSE is calculated to be infinity, meaning that the predicted data range is zero or close to zero. In this case, the denominator of the NRMSE equation becomes very small or approaches zero, causing the NRMSE value to become infinitely large. This can happen when the predicted data has no variability or is constant, and the model attempts to predict different values for it.

Table 12: **Artificial Neural Network Models Result for Case Study Data**

ANN Model	Optimizer	MSE	RMSE	R2	MAE	NRMSLE
ANN-1	RMSprop	0.41930	0.46178	0.98012	0.46178	infinite
ANN-2	Adam	0.48860	0.52882	0.98710	0.52882	infinite
ANN-3	SGD	0.83792	0.64974	0.97847	0.64974	infinite

7. Overall Analysis and Advantages of our Framework with various suggestions

In this research, we addressed several AI approaches, such as machine learning, deep learning, and neural networks, that could potentially be utilized to improve and make the process for seismic hazard maps and predicting earthquake's magnitude efficient. We have included an in-depth discussion of each technique, including its benefits and drawbacks, and emphasized recent study initiatives that have successfully applied these techniques to seismic hazard mapping and making predictions. AI techniques that we have used in this paper could be helpful to government authorities and disaster coordinators to reach more informed choices about disaster preparation and response through the generation of more accurate seismic hazard maps. For example, they can assist in identifying areas at higher risk of earthquakes, allowing precautions such as more durable infrastructure and escape planning to be put into effect, which can save lives and reduce damage in the event of an earthquake. They can also help provide the necessary disaster relief facilities and resources that are readily available within the crucial time window.

This paper presents the development and execution of three ANN models and seven ML models. These models were extensively compared in our study. The input variables which were used in the study are categorized into continuous and discrete variables. The class of continuous variables contains latitude, longitude, Earthquake's magnitude(mag), Depth(depth), Gap(gap), Minimum Distance, Root mean Square(rms), Horizontal Error (horizontalError), Depth Error(depthError), Magnitude Error(magError), Magnitude Stations (magNst). Now, the class of discrete variables contains Magnitude Types (magType-n), and Location Source (locationSource-n).

The investigations of the seismic mapping and Earthquake's magnitude predictions by the ANN and ML models, and the relative significances of the input variables on the response prediction yield the following conclusions:
1. Artificial Intelligence (AI) algorithms are able to analyze vast amounts of data very quickly and efficiently, enabling them to identify patterns and connections that human analysts could have overlooked. This can result in more precise and reliable seismic hazard maps, reducing the probability of earthquake-related tragic events. In accordance with our findings, ANNs outperform typical ML algorithms by a wide margin. The ANN-2 model with the 'Adam' optimizer is the best model for Earthquake magnitude prediction since it is the most accurate and effective fast.

2. The developed ANN and ML models are capable of reliably estimating seismic mapping and Earthquake magnitude prediction. The coefficients of determination (R2) in the maximum for the entire ANN models with each having value of 0.999 for the complete data set and having values of 0.98012 for ANN-1, 0.98710 for ANN-2, and 0.97847 for ANN-3. A high R2 score value indicates that the model is able to explain a large portion of the variance in the target variable based on the input variables. This is generally considered to be a good sign, as it suggests that the model is able to capture the

underlying relationships between the input and output variables.

3. The error function for all the ANN models is very low. Hence, A low value of MAE, MSE, RMSE, NRMSE indicates that the model's predictions are close to the actual values of the target variable. In other words, a low error value suggests that the model is performing well and is able to accurately predict the target variable based on the input variables. But in our case study, we get an interesting result of having NRMSE infinite. Which indicates that the predicted data range is zero or close to zero. In this case, the denominator of the NRMSE equation becomes very small or approaches zero, causing the NRMSE value to become infinitely large. This can happen when the predicted data has no variability or is constant, and the model attempts to predict different values for it.

4. Based on the results shown in Figures [37, 38], we can see that the error values (MSE, MAE, RMSE) for both the training and case study data are similar, with the case study data sometimes even showing lower errors. To clarify, the plot shows how a model performs on both training and case study data. The error values reflect how well the model predictions corresponded to the actual values, with smaller values showing superior performance. According to the Figures [37, 38], the model works well on both datasets, with error values that are near each other.

5. Furthermore, in some instances, the error values for case study data are smaller than those for training data. This demonstrates that the model can generalize well to new data that it has not been taught on, which is a desirable trait for any forecasting model. Overall, the findings show that the model is working well and may be helpful in practical applications. The coefficients of determination (R2) for all of the ML models, on the other hand, are quite low, with values of -0.2601 for ET, 0.1834 for CatBoost, 0.2010 for LGBM, -0.2877 for RF, -0.2601 for BR, -0.1195 for XGBoost, and 1.0877 for DT. A poor R2 score indicates that the model is unable to clarify an important portion of the variance in the target variable using the input variables. It means the model is not correctly capturing the relationships between variables.

6. The seismic maps we obtain allow us to identify specific areas with a high risk of earthquakes. It is crucial that these red zone areas are equipped with adequate disaster relief facilities, ample food resources, and strong infrastructure to ensure the safety and well-being of the local population during a potential earthquake. When an area is designated as a red zone, it indicates that a seismic event is more likely to occur in that location. It is crucial to have the proper disaster relief infrastructure and resources close at hand in these places in order to lessen the possible effects of earthquakes. This entails having prepared personnel, medical equipment, and emergency shelters ready to act swiftly and effectively in the case of an earthquake. Additionally, keeping an abundance of food supplies is crucial after an earthquake because access to food and clean water may be restricted or eliminated. It is critical to ensure that these places have enough food and other basic supplies to maintain the local people until outside relief arrives. And finally, strong infrastructure is necessary to survive an earthquake's effects. This includes structures like bridges, buildings, and other essential infrastructure constructed to resist earthquakes. To guarantee that a region is adequately prepared for the possible impact of an earthquake, it is critical to invest in the infrastructure upgrades required.

8. Conclusions

In this work, AI methods and GIS are used to predict earthquakes and produce seismic hazard maps. This work focuses on the earlier geological information using the ML and ANN technique, and the case study includes the center and southern areas of Turkey. With the determination of expected locations, greater focus is placed on design, seismic restoration, and evaluating the dependability of the existing structures in these places. No specific method has yet been approved for figuring out the specifics of forthcoming earthquakes.

1. Our findings indicate that ANNs outperform traditional machine learning algorithms significantly. The ANN-2 model, optimized with the 'Adam' algorithm, is the most accurate and efficient model for predicting earthquake magnitudes.

2. The developed ANN and ML models are reliable in estimating seismic mapping and predicting earthquake magnitudes. The ANN models have a maximum coefficient of determination (R2) of 0.999 for the complete dataset, with values of 0.98012 for ANN-1, 0.98710 for ANN-2, and 0.97847 for ANN-3. A high R2 score indicates that the model can explain a large portion of the variance in the target variable based on the input variables. This suggests that the models can capture the underlying relationships between the input and output variables, which is a good sign.

3. The error function for all ANN models is low, indicating accurate predictions of the target variable based on input variables. Low MAE, MSE, RMSE, and NRMSE values suggest good model performance. However, in our case study, we found an interesting result of having an infinite NRMSE. This indicates that the predicted data range is close to or zero, resulting in a very small denominator in the NRMSE equation, leading to an infinitely large value. This occurs when the predicted data has no variability or is constant, and the model attempts to predict different values for it.

4. One significant observation is the ability to generalize to new data. This is a desirable trait for any forecasting model. Overall, the results show that the model is reliable and has potential for usage in real-world scenarios.

5. Although the ML models are obtaining low MSE, MAE, and RMSE values, the accompanying R2 values are rather low, highlighting the limitations of these ML models.

6. The seismic maps generated through AI analysis can help identify areas with a high risk of earthquakes. When these areas are designated as red zones, it is essential to have adequate disaster relief facilities, food resources, and strong infrastructure in place to ensure the safety and well-being of the local population during an earthquake. This means having trained personnel, medical equipment, and emergency shelters readily available.

9. Plans for the Future

In our research, we have complemented seismic mapping with earthquake magnitude prediction to get a more complete understanding of the seismic risks in a given area. By combining these two methods, we are able to determine where earthquakes are most likely to occur as well as estimate the possible magnitude of such earthquakes, which may help with disaster planning and response. We may better identify regions that may be at risk of secondary hazards, such as landslides and liquefaction, which may exacerbate the damage caused by earthquakes, by integrating several data sources and modeling tools. We can produce more precise danger maps and better guide land-use planning and infrastructure development by adding information on soil type, groundwater levels, and other geospatial elements.

Future studies will concentrate on optimization methods to improve the accuracy of earthquake magnitude estimation and seismic mapping in the targeted locations.

Acknowledgements We acknowledge the technical and scientific support from Department of Mathematics, National Institute of Technology Agartala. Turkey earthquake data has been collected from the U.S. Geological Survey (USGS). As a result, we are grateful to the USGS.

Authors contribution All authors contributed to conceptualize and design the study. Saptadeep Biswas and Dhruv Kumar worked together to collect data, analyze datasets, prepare them for analysis, and do those tasks. The results, assessments, and visualisations were validated under the supervision of Uttam Kumar Bera by Saptadeep Biswas and Dhruv Kumar. Saptadeep Biswas and Dhruv Kumar wrote the first draft of the manuscript, and all authors provided feedback. The final manuscript was read and approved by all the authors.

Funding The authors declare that they got no funding, grants, or other forms of support while preparing this paper.

Data availability The manuscript comprises the majority of the data used in and/or analyzed during the current study. In all other cases, data will be made available upon request from the corresponding author.

Declarations

Competing interests There are no competing interests of the authors to declare that are pertinent to the subject matter of this study. Each author indicates that they are not connected to or a part of any group or organization that has a financial or non-financial interest in the topics or materials covered in this work.

Ethical approval There are no studies by any of the authors in this article that used humans or animals as subjects.

References

[1] M. Alizadeh, I. Ngah, M. Hashim, B. Pradhan, and A. Pour. A Hybrid Analytic Network Process and Artificial Neural Network (ANP-ANN) Model for Urban Earthquake Vulnerability Assessment. *Remote Sensing*, 10(6):975, 6 2018.

[2] K. M. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez. Earthquake prediction model using support vector regressor and hybrid neural networks. *PLoS ONE*, 13(7), 2018.

[3] K. M. Asim, F. Martínez-Álvarez, A. Basit, and T. Iqbal. Earthquake magnitude prediction in Hindukush region using machine learning techniques. *Natural Hazards*, 85(1):471–486, 1 2017.

[4] U. Bhagyaraj and S. Kolathayar. Seismic Hazard Assessment and Landslide Vulnerability Mapping for Ladakh, and Jammu & Kashmir Using GIS Technique. *Journal of the Geological Society of India*, 99(3):377–382, 3 2023.

[5] H. Chaulagain, H. Rodrigues, V. Silva, E. Spacone, and H. Varum. Seismic risk assessment and hazard mapping in Nepal. *Natural Hazards*, 78(1), 2015.

[6] J. Chen, H. Tang, W. Chen, and N. Yang. A Prediction Method of Ground Motion for Regions without Available Observation Data (LGB-FS) and Its Application to both Yangbi and Maduo Earthquakes in 2021. *Journal of Earth Science*, 33(4), 2022.

[7] B. R. Cox, J. Bachhuber, E. Rathje, C. M. Wood, R. Dulberg, A. Kottke, R. A. Green, and S. M. Olson. Shear wave velocity- and geology-based seismic microzonation of port-au-prince, Haiti. *Earthquake Spectra*, 27(SUPPL. 1), 2011.

[8] S. Das, S. Sarkar, and D. P. Kanungo. A critical review on landslide susceptibility zonation: recent trends, techniques, and practices in Indian Himalaya. *Natural Hazards*, 115(1):23–72, 1 2023.

[9] S. Demir and E. K. Sahin. Predicting occurrence of liquefaction-induced lateral spreading using gradient boosting algorithms integrated with particle swarm optimization: PSO-XGBoost, PSO-LightGBM, and PSO-CatBoost. *Acta Geotechnica*, 2023.

[10] A. Doğangün, B. Yön, O. Onat, M. Emin Öncü, and S. Sağıroğlu. Seismicity of East Anatolian of Turkey and Failures of Infill Walls Induced by Major Earthquakes. *Journal of Earthquake and Tsunami*, 15(04), 8 2021.

[11] J. R. Elliott. Earth Observation for the Assessment of Earthquake Hazard, Risk and Disaster Management. *Surveys in Geophysics*, 41(6):1323–1354, 11 2020.

[12] S. E. Güvercin, H. Karabulut, A. Konca, U. Doğan, and S. Ergintav. Active seismotectonics of the East Anatolian Fault. *Geophysical Journal International*, 230(1):50–69, 3 2022.

[13] A. Hubert-Ferrari, L. Lamair, S. Hage, S. Schmidt, M. N. Çağatay, and U. Avşar. A 3800 yr paleoseismic record (Lake Hazar sediments, eastern Turkey): Implications for the East Anatolian Fault seismic cycle. *Earth and Planetary Science Letters*, 538:116152, 5 2020.

[14] R. Jena, S. P. Naik, B. Pradhan, G. Beydoun, H.-J. Park, and A. Alamri. Earthquake vulnerability assessment for the Indian subcontinent using the Long Short-Term Memory model (LSTM). *International Journal of Disaster Risk Reduction*, 66:102642, 12 2021.

[15] R. Jena, B. Pradhan, G. Beydoun, A. M. Alamri, Ardiansyah, Nizamuddin, and H. Sofyan. Earthquake hazard and risk assessment using machine learning approaches at Palu, Indonesia. *Science of the Total Environment*, 749, 2020.

[16] R. Jena, B. Pradhan, G. Beydoun, Nizamuddin, Ardiansyah, H. Sofyan, and M. Affan. Integrated model for earthquake risk assessment using neural network and analytic hierarchy process: Aceh province, Indonesia. *Geoscience Frontiers*, 11(2):613–634, 3 2020.

[17] M. Kamran, R. K. Wattimena, D. J. Armaghani, P. G. Asteris, I. M. Jiskani, and E. T. Mohamad. Intelligent based decision-making strategy to predict fire intensity in subsurface engineering environments. *Process Safety and Environmental Protection*, 171, 2023.

[18] K. A. Korkmaz, A. Irfanoglu, and A. H. Kayhan. Seismic risk assessment of buildings in Izmir, Turkey. *Natural Hazards*, 54(1):97–119, 7 2010.

[19] R. Maio, T. M. Ferreira, R. Vicente, and J. Estêvão. Seismic vulnerability assessment of historical urban centres: Case study of the old city centre of Faro, Portugal. *Journal of Risk Research*, 19(5), 2016.

[20] S. Malakar and A. K. Rai. Seismicity clusters and vulnerability in the Himalayas by machine learning and integrated MCDM models. *Arabian Journal of Geosciences*, 15(22):1674, 11 2022.

[21] A. Maqsoom, B. Aslam, M. Awais, U. Hassan, W. S. Alaloul, M. A. Musarat, and M. I. Qureshi. Efficiency of multiple hybrid techniques for the earthquake physical susceptibility mapping: the case of Abbottabad District, Pakistan. *Environmental Earth Sciences*, 80(19):678, 10 2021.

[22] C. Merciu, I. Ianos, G.-L. Merciu, R. Jones, and G. Pomeroy. Mapping accessibility for earthquake hazard response in the historic urban centre of Bucharest. *Natural Hazards and Earth System Sciences*, 18(7):2011–2026, 7 2018.

[23] P. S. Nayek and M. Gade. Artificial neural network-based fully data-driven models for prediction of newmark sliding displacement of slopes. *Neural Computing and Applications*, 34(11), 2022.

[24] H. D. Nguyen, J. M. LaFave, Y. J. Lee, and M. Shin. Rapid seismic damage-state assessment of steel moment frames using machine learning. *Engineering Structures*, 252, 2022.

[25] P. H. Nyimbili, T. Erden, and H. Karaman. Integration of GIS, AHP and TOPSIS for earthquake hazard analysis. *Natural Hazards*, 92(3), 2018.

[26] I. Pal, S. K. Nath, K. Shukla, D. K. Pal, A. Raj, K. K. Thingbaijam, and B. K. Bansal. Earthquake hazard zonation of Sikkim Himalaya using a GIS platform. *Natural Hazards*, 45(3), 2008.

[27] A. Pandit and S. Panda. Prediction of earthquake magnitude using soft computing techniques: ANN and ANFIS. *Arabian Journal of Geosciences*, 14(13), 2021.

[28] N. Rahman, M. A. Ansary, and I. Islam. GIS based mapping of vulnerability to earthquake and fire hazard in Dhaka city, Bangladesh. *International Journal of Disaster Risk Reduction*, 13, 2015.

[29] W. Riyanto, D. Irawan, T. Joko Wahyu Adi, D. Iranata, and A. Rizki Amalia. Earthquake Vulnerability Assessment of High-Rise Buildings in Surabaya using RViSITS Android Application. In *IOP Conference Series: Materials Science and Engineering*, volume 739, 2020.

[30] A. Samm-A, A. S. M. M. Kamal, and M. Z. Rahman. Earthquake and rainfall-induced landslide hazard assessment of Kutupalong Rohingya camp using meteorological and geological information. *Stochastic Environmental Research and Risk Assessment*, 3 2023.

[31] Sangeeta and B. K. Maheshwari. Spatial predictive modelling of rainfall- and earthquake-induced landslide susceptibility in the Himalaya region of Uttarakhand, India. *Environmental Earth Sciences*, 81(8):237, 4 2022.

[32] Sangeeta, B. K. Maheshwari, and D. P. Kanungo. GIS-based pre- and post-earthquake landslide susceptibility zonation with reference to 1999 Chamoli earthquake. *Journal of Earth System Science*, 129(1):55, 12 2020.

[33] K. P. Sreejaya, J. Basu, S. T. Raghukanth, and D. Srinagesh. Prediction of Ground Motion Intensity Measures Using an Artificial Neural Network. *Pure and Applied Geophysics*, 178(6), 2021.

[34] S. Stein, B. D. Spencer, and E. M. Brooks. Metrics for Assessing Earthquake-Hazard Map Performance. *Bulletin of the Seismological Society of America*, 105(4):2160–2173, 8 2015.

[35] S. S. Tezcan and M. Ipek. A reconnaissance report: 1995 Dinar, Turkey, earthquake. *Engineering Structures*, 18(12), 1996.

[36] P. Yariyan, R. Ali Abbaspour, A. Chehreghan, M. Karami, and A. Cerdà. GIS-based seismic vulnerability mapping: a comparison of artificial neural networks hybrid models. *Geocarto International*, 37(15):4312–4335, 8 2022.

[37] P. Yariyan, H. Zabihi, I. D. Wolf, M. Karami, and S. Amiriyan. Earthquake risk assessment using an integrated Fuzzy Analytic Hierarchy Process with Artificial Neural Networks based on GIS: A case study of Sanandaj in Iran. *International Journal of Disaster Risk Reduction*, 50:101705, 11 2020.

[38] J. Zhuang, J. Peng, X. Zhu, and W. Huang. Scenario-based risk assessment of earthquake disaster using slope displacement, PGA, and population density in the Guyuan region, China. *ISPRS International Journal of Geo-Information*, 8(2), 2019.

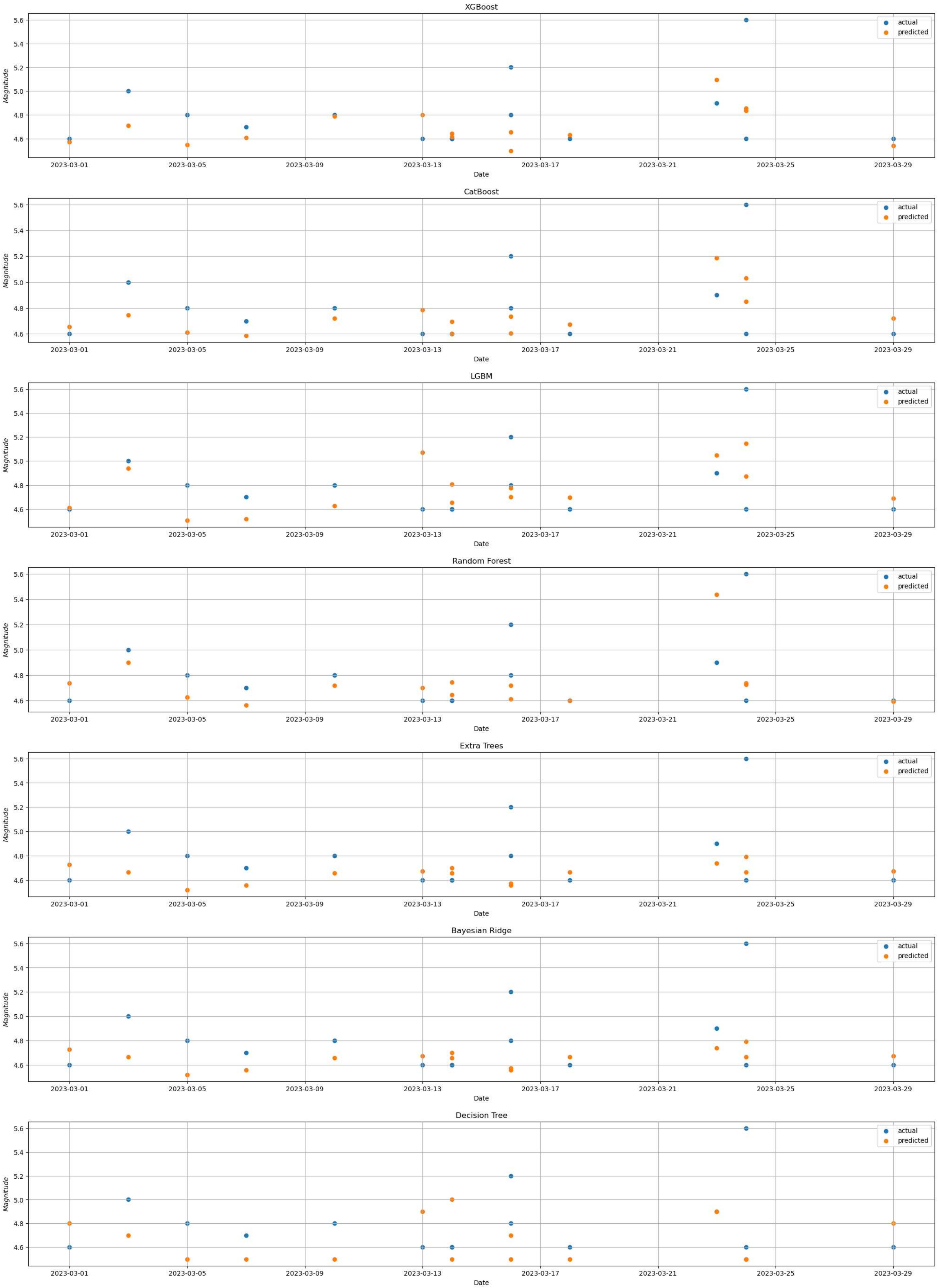
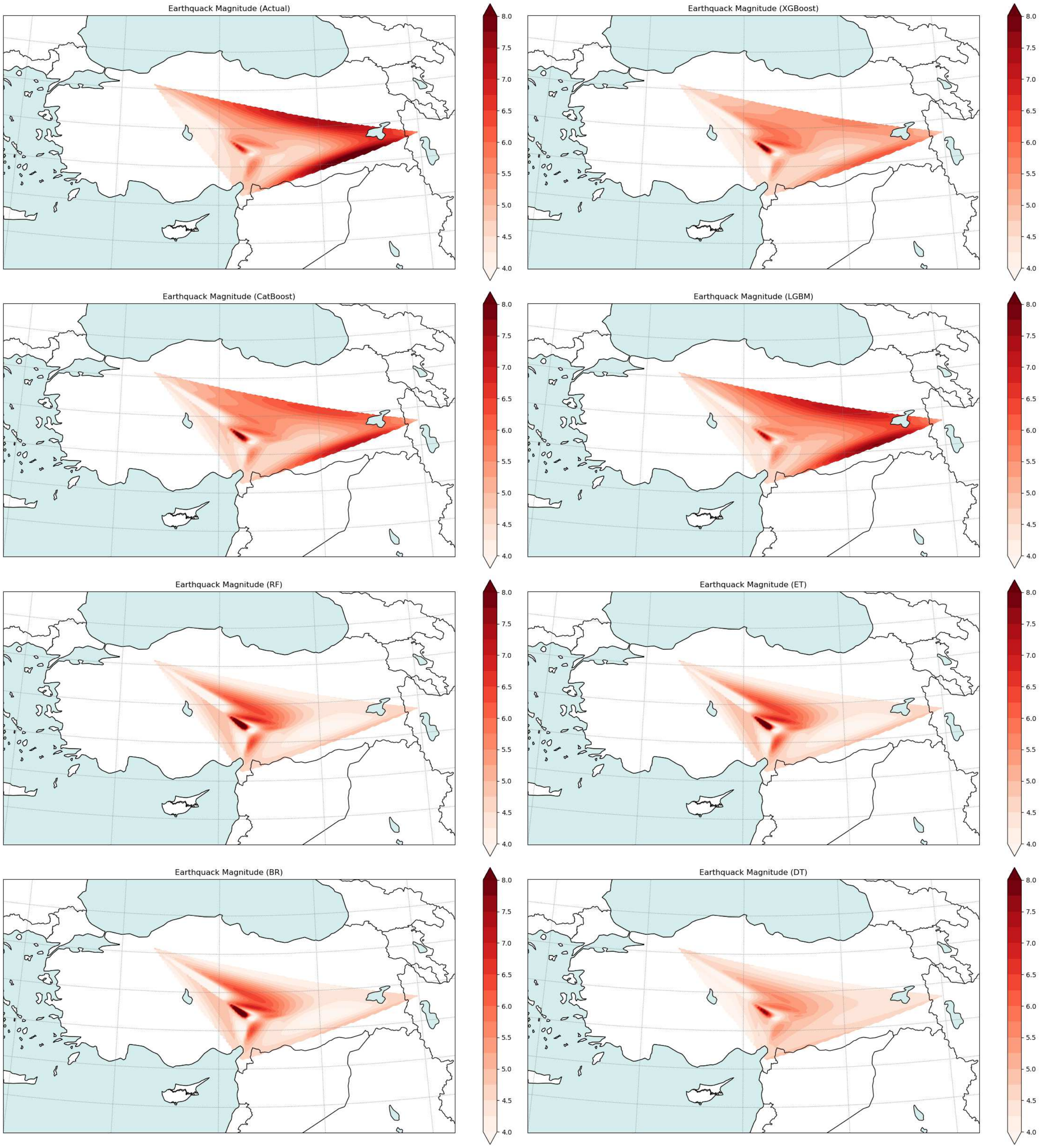


Figure 42: ML models prediction performance for the Case Study


Earthquack Magnitude (Actual)

Earthquack Magnitude (XGBoost)

Earthquack Magnitude (CatBoost)

Earthquack Magnitude (LGBM)

Earthquack Magnitude (RF)

Earthquack Magnitude (ET)

Earthquack Magnitude (BR)

Earthquack Magnitude (DT)

Figure 43: ML models Mapping for the Case Study