Ensuring Earthquake-Proof Development in a Swiftly Developing Region through Neural Network Modeling of Earthquakes Using Nonlinear Spatial Variables

Mubeen ul Basharat 1, Junaid Ali Khan 1, Umer Khalil 2, Aqil Tariq 3, 4, *, Bilal Aslam 5 and Qingting Li 6, *

1 Department of Computer Science and Engineering, HITEC University, Taxila 47080, Punjab, Pakistan
2 ITC Faculty of Geo-Information Science and Earth Observation, University of Twente, 7522 NB Enschede, The Netherlands
3 Department of Wildlife, Fisheries and Aquaculture, College of Forest Resources, Mississippi State University, 775 Stone Boulevard, Starkville, MS 39762, USA
4 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China
5 School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA
6 Airborne Remote Sensing Center, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
* Correspondence: aqiltariq@whu.edu.cn (A.T.); liqt@radi.ac.cn (Q.L.)

Abstract: Northern Pakistan, the center of major construction projects due to the commencement of the China Pakistan Economic Corridor, is among the most earthquake-prone regions globally owing to its tectonic settings. The area has experienced several devastating earthquakes in the past, and these earthquakes pose a severe threat to infrastructure and life. Several researchers have previously utilized advanced tools such as Machine Learning (ML) and Deep Learning (DL) algorithms for earthquake predictions. This technological advancement helps with construction innovation, for instance, by designing earthquake-proof buildings. However, previous studies have focused mainly on temporal rather than spatial variables. The present study examines the impact of spatial variables to assess the performance of the different ML and DL algorithms for predicting the magnitude of short-term future earthquakes in North Pakistan. Two ML methods, namely Modular Neural Network (MNN) and Shallow Neural Network (SNN), and two DL methods, namely Recurrent Neural Network (RNN) and Deep Neural Network (DNN) algorithms, were used to meet the research objectives. The performance of the techniques was assessed using statistical measures, including accuracy, information gain analysis, sensitivity, specificity, and positive and negative predictive values. These metrics were used to evaluate the impact of including a new variable, Fault Density (FD), and the standard seismic variables in the predictions. The performance of the proposed models was examined for different patterns of variables and different classes of earthquakes. The accuracy of the models for the training data ranged from 73% to 89%, and the accuracy for the testing data ranged from 64% to 85%. The analysis outcomes demonstrated an improved performance when using an additional variable of FD for the earthquakes of low and high magnitudes, whereas the performance was less for moderate-magnitude earthquakes. DNN, and SNN models, performed relatively better than other models. The results provide valuable insights about the influence of the spatial variable. The outcome of the present study adds to the existing pool of knowledge about earthquake prediction, fostering a safer and more secure regional development plan involving innovative construction.

Keywords: North Pakistan; earthquake-proof development; earthquake prediction; neural network modeling; spatial variables
1. Introduction

Earthquakes are among the most perilous natural disasters, causing loss of life, serious calamities, and destruction of building infrastructures, resulting in both life and economic losses [1,2]. In the past two decades, earthquakes and tsunamis have naturally been the fatalist form of disasters, accounting for 58% of overall fatalities regardless of their comparatively low frequency [3]. In recent decades, a series of strong earthquakes have been responsible for inducing landslides in different global regions [4]. Flushing excess debris downstream of rivers, leading to bank erosion and floodplain accumulation, and channel eruptions that affect flood frequency, towns, ecosystems, and infrastructure are slower consequences of earthquakes [5].

Earthquakes also cause wide cracks and fractures on verges and mountain crests, thus increasing the occurrence of landslides that may last for years [6]. Due to the dreadful consequences of earthquakes, seismologists and geologists have a huge compel to make accurate predictions for earthquakes by considering the time, place, and magnitude [7–10]. Therefore, many researchers have adopted different methods for predicting an earthquake by considering the environmental and regional dynamics of different seismically prone areas.

Existing earthquake prediction research can be broadly split into four groups based on the methodologies used: (1) precursor signal investigation, (2) mathematical analysis, (3) Machine Learning (ML)-based algorithms, and (4) Deep Learning (DL)-based algorithms. For the first group, researchers have explored earthquake precursor signals. The researchers have observed different features, such as an increase in temperature [11,12], recording the behavior of animals [13,14], an emanation of radon gas [15], lithosphere–atmosphere ionosphere coupling [16], and variability of Aerosol Optical Depth (AOD) [17]. The air above the epicentral region of an impending significant earthquake frequently becomes overloaded with positive airborne ions due to the electromagnetic signals from the crustal rocks. These positive airborne ions have been shown to cause changes in stress hormone levels in animals and humans [13]. The water on the surface is also oxidized to hydrogen peroxide due to these charges, and this, plus oxidation, can cause behavioral changes among aquatic animals [13,14]. The concentration of Rn-222 in soil has been observed to be influenced before earthquakes [15].

Researchers have adopted statistical methods and techniques in the second type to accomplish earthquake prediction. The Fibonacci, Dual, and Lucas (FDL) approach [18], fuzzy mathematics [19], likelihood analysis of earthquake catalogs [20], stochastic models [21], statistical physics approach [22], Poisson distribution [23], etc., are some of the techniques that have been studied for predicting earthquakes.

The third type of research focused on using data-mining or ML methods such as Fuzzy Logic [24], AdaBoost [10], Decision Tree [25,26], Support Vector Machines (SVM) [27], K-nearest Neighbors [28], etc., to predict earthquakes based on previously recorded seismic data in the same region. DL algorithms have been used in the fourth type of work to anticipate the magnitude and timing of significant seismic events. DL is a rapidly expanding field of ML research that detects and classifies data patterns using multi-layered representations and is inspired by ANNs. The term ‘deep’ in this strategy refers to a series of levels, through which data from one level are converted into another level [29]. A suitable data transformation allows obtaining the most appropriate hierarchical representations of data as the number of layers (i.e., network depth) increases [30]. To date, various DL architectures (e.g., Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs)) have been presented [31–33], all of which have enhanced the state-of-the-art in ML research.

Although much work has been carried out on earthquake predictions, very few can accurately predict seismic events [2,34]. The insufficient earthquake predictions are due to the number of real-time factors that are difficult to analyze and the higher complexities of the prediction processes. Traditional and existing statistical methods cannot accurately predict the complex nonlinear correlations amongst earthquake occurrences [35]. However, due to their relatively better prediction performance, numerous ML and DL neural networks...
(NNs) have recently become the center of attention in the prediction of earthquakes for various situations and with different seismic parameters and have performed well [8,36–40]. Some of the most prominently used NN architectures in earthquake-related studies include the Multilayer Perceptron (MLP) [41], Backward Propagation Neural Network [42], Deep Neural Network (DNN) [43], and Feed-Forward Neural Network [44], RNN [45], Shallow Neural Network (SNN) [26], Pattern Recognition Neural Network (PRNN) [46], etc. Various NN architectures have also been used in other fields for prediction and classification purposes, and their exceptional performances have been documented. Some of the prominent architectures include the Modular Neural Network (MNN) [47], Long Short-Term Memory [48], Residual Network [49], etc.

Furthermore, in previous studies, only history-based time-series data were used to predict earthquakes in a specific area; hence, excellent or accurate results could not be produced [50]. The study of earthquakes’ spatial and temporal properties is an important branch of seismic science divided into two general forecast categories: long-term forecasting based on months and years and short-term forecasting based on days and hours [51,52]. Short-term earthquake prediction is more challenging due to the complexity of earthquake phenomena [53], crustal blocks-and-faults structure [54], the intricacies of the Earth’s lithosphere, and the lack of a credible approach for such forecasts [55]. Moreover, short-term seismic events are directly associated with social infrastructure and human lives [56]. Some studies [55,57] have focused on the variables directly correlated with the expected output variable. However, these studies ignored the less correlated variables when making predictions. Moreover, most existing studies are based on the temporal correlation between independent and dependent variables rather than spatial correlations. Furthermore, it is also considered important to change the parameters of the DL or ML models, which include the neurons, hidden layers, learning rate, epochs, etc., to obtain the best architecture for the model [58].

This provides the novelty of the present study, where both the spatial and temporal characteristics of a large area and real-time data, along with history-based data, are considered to make earthquake predictions using multiple state-of-the-art NNs. The overall outcomes of the study are the primary contributions of this study. Generally, this study had the following objectives:

1. To utilize two ML methods, MNN and SNN, and two DL methods, RNN and DNN, and check their applicability and performance for earthquake prediction. The choice of methods was based on the literature review and the application and performance of the techniques in different research areas.
2. To explore the spatiotemporal viewpoint of earthquake prediction, as the results of prediction models for certain input variables are different, as each method extracts the information from the input variables differently.
3. To examine the prediction accuracy of the models by introducing and analyzing the role of an additional spatial variable, namely Fault Density (FD), involved in the earthquake-prediction process.
4. To use accuracy, information gain analysis (IGA), and sensitivity metrics for evaluating the descriptive power of each input variable, including the suggested FD, to evaluate the overall performance of the considered models.

The rest of the study is organized such that the study area is described in Section 2. Data preparation, the considered seismic variables, and the description of the methods used are given in Section 3. The obtained results of this study are presented in Section 4, and the discussion about the general outcomes of the study and their importance are presented in Section 5. Eventually, Section 6 is the conclusion, offering a brief description of the outcomes of the present study.

2. Study Area

The targeted area for this study is the northern part of Pakistan. Despite being the most spectacular and fascinating region of Pakistan, the northern part of Pakistan is a geohazard
notspot. It is among the most seismically active regions globally [27,35]. Three of the world’s prominent mountain ranges, the Hindukush, the Karakorum, and the Himalayas, meet here. These mountain ranges were formed because of the conformation of the Indian and Eurasian tectonic plates during Eocene. The region’s seismic activity is caused by its location as a subduction zone (compression) between the colliding Indian and Eurasian plates [34]. The region is known for its intermediate-depth seismic events (70–300 km). However, some deep seismic events (>300 km) have also been reported [46]. Numerous opposing reasons for the intriguing deep seismicity of the region have been presented by researchers. Still, the halting point is that no suitable and persuasive explanations are presented [27].

3. Materials and Methods

The proposed process of predicting short-term earthquakes is demonstrated in Figure 1. The estimation of the dependent variable (earthquake magnitude) was the final goal of this study. The dependent variable categorizes the magnitude of the largest earthquake occurring in the subsequent week. The initial step in the process was acquiring the data associated with the three chosen pixels that met C1 and C2, which are mentioned as criteria in Figure 1. Subsequently, for every record of the data, independent variables were computed. Then, the splitting of data into three portions, namely, training (50%), validation (25%), and testing (25%), was carried out. The Natural Breaks categorization procedure assisted in determining class limits and record rearrangement, resulting in a consistent class distribution across all three datasets (training, validation, and testing). In other words, all classes were equally characterized in the training, testing, and validation of datasets. The models were trained, tested, and validated, and the accuracy assessment was performed. All the involved steps are explained in the subsequent subsections.

Figure 1. The methodological framework of the study.
3.1. Data Collection and Processing

The data for this study were collected from the United States Geological Survey (USGS) and the International Institute of Earthquake Engineering and Seismicity (IIEES). The analysis period ranged from 1 January 1970 to 31 December 2021. The frequency of the earthquakes per year for the considered period in the region is shown in Figure 2. The highest number of earthquakes is in the year 2005. This is the same year that the catastrophic earthquake of $M_w$ 7.6 hit Northern Pakistan. Figure 2 shows that there has been a notable surge in the number of occurrences in recent years.

![Figure 2. Frequency of earthquake occurrence per year from 1970 until 2021 in Northern Pakistan.](image)

After acquiring and saving the raw catalog data from the two specified sources, the data were put together, and the identified identical rows were eliminated. Only the latitude, longitude, and depth columns from catalog data were taken as input variables for prediction purposes. Next, the catalog data were sorted to eliminate events with moment magnitudes ($M_w$) less than 3 (Richter Scale). This step resulted in data containing more critical earthquakes. Numerous studies [59,60] have also adopted the same sort of methodology. They removed earthquakes with $M_w$ 3 or less because they are not regarded as significant and can introduce errors in the modeling. Moreover, the used earthquake catalog had a magnitude of completeness as $M_w$ 3, so events with $M_w$ 3 were considered in the analysis.

After filtering the data, the locations of the events with an $M_w$ greater than or equal to 3 are shown in Figure 3. The total number of events recorded during the considered period in North Pakistan is 12,646. The most critical events with significant magnitudes are displayed in red. Seismic events with magnitudes ranging from 3 to 7.7 were spread across the whole region, as shown in Figure 3.
Figure 3. Distribution of earthquakes based on the magnitude of occurrence (greater than 3 Richter) and location from 1970 until 2021 in Pakistan and its surroundings.

Following the data collection and filtering, a grid size of $1 \times 1$ degree was built for the study area. To study the more earthquake-prone areas in Pakistan’s Northern region, only pixels with a minimum of 300 seismic occurrences (C1) and at least one $M_w$ 5 (C2) event were included. As a result, just three pixels met the C1 and C2 requirements. It was assumed that setting the threshold of 300 seismic events and $M_w$ 5 will result in more exposed areas or hotspots that require the utmost attention, as future earthquakes somehow align with the previous earthquake trends. Consequently, the three pixels were selected as the input pixels for future investigation. One of the pixels that met the criteria is shown in a window in Figure 3.

Moreover, statistics about the earthquake events for each pixel are listed in Table 1. The standard deviation (std) of the events in the three pixels shows that the $M_w$ of the events is relatively less dispersed in relation to the average $M_w$. Moreover, the total number of events in the selected pixels is 1035.

<table>
<thead>
<tr>
<th>Pixel (Row-Column)</th>
<th>No. of Events</th>
<th>Avg of Magnitude</th>
<th>Variance of Magnitude</th>
<th>Std of Magnitude</th>
<th>Magnitude Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4–9)</td>
<td>361</td>
<td>4.763</td>
<td>0.297</td>
<td>0.54</td>
<td>6.8–3</td>
</tr>
<tr>
<td>(7–11)</td>
<td>373</td>
<td>4.236</td>
<td>0.348</td>
<td>0.43</td>
<td>7.6–3</td>
</tr>
<tr>
<td>(11–5)</td>
<td>301</td>
<td>4.387</td>
<td>0.423</td>
<td>0.48</td>
<td>7.1–3</td>
</tr>
</tbody>
</table>
3.2. Dependent and Independent Variables

In the present work, the earthquake prediction problem was deemed as a classification problem. The input data were converted into well-structured records before feeding into the proposed models in this study. Every data record contains numerous independent variables (which are basically the used seismic variables in this study) and a single dependent variable (earthquake magnitude). The dependent (output) variable represents the maximum magnitude of the seismic event in the next seven days. The magnitude of the next week’s extreme earthquake was projected to be one of the four magnitude classes listed in Table 2.

Table 2. The boundaries of output classes.

<table>
<thead>
<tr>
<th>Range (Dependent Variable)</th>
<th>Class</th>
<th>No. of Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>3–4</td>
<td>1</td>
<td>276</td>
</tr>
<tr>
<td>4–5</td>
<td>2</td>
<td>376</td>
</tr>
<tr>
<td>5–6</td>
<td>3</td>
<td>256</td>
</tr>
<tr>
<td>6–7</td>
<td>4</td>
<td>127</td>
</tr>
</tbody>
</table>

Former research has demonstrated that an imbalanced dataset resulting from the classification of the dependent variable might considerably jeopardize the performance of ML models for earthquakes [61]. As a result, in this investigation, the frequency distribution of the dependent variable was used to designate the periods, and the Natural Breaks classification approach [4,6,35] was used to create the class bounds.

A total of 19 independent variables referenced from previous studies were considered. These include longitude, latitude, 16 commonly used seismic variables, depth, and the suggested $FD$ variable. They comprise 20 input variables. These variables were normalized (between 0 and 1) before being processed by the models. The commonly used 16 seismic input variables and their description are listed in Table 3.

Table 3. Description of the used seismic variables (adopted from [26,37,62]).

<table>
<thead>
<tr>
<th>ID</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b-value</td>
<td>b-value from the prominent Gutenberg–Richter (GR) law</td>
</tr>
<tr>
<td>2</td>
<td>X1</td>
<td>Increment of b between i and i-4 events</td>
</tr>
<tr>
<td>3</td>
<td>X2</td>
<td>Rise amongst i-4 and i-8 events</td>
</tr>
<tr>
<td>4</td>
<td>X3</td>
<td>Rise amongst i-8 and i-12 events</td>
</tr>
<tr>
<td>5</td>
<td>X4</td>
<td>Rise amongst i-12 and i-16 events</td>
</tr>
<tr>
<td>6</td>
<td>X5</td>
<td>Rise amongst i-16 and i-20 events</td>
</tr>
<tr>
<td>7</td>
<td>X6</td>
<td>Maximum magnitude from the events documented through the last week using OU’s law</td>
</tr>
<tr>
<td>8</td>
<td>X7</td>
<td>Likelihood of events with magnitude equal or greater than 6.0, computed as $P (Ms \geq 6) = e^{-3b/\log 6} = e^{-0.3b}$</td>
</tr>
<tr>
<td>9</td>
<td>a-value</td>
<td>a-value of GR law</td>
</tr>
<tr>
<td>10</td>
<td>$\eta$</td>
<td>Resultant of the mean square deviation (MSD) from the regression line centered on GR law</td>
</tr>
<tr>
<td>11</td>
<td>$\Delta M$</td>
<td>Remainder of the maximum noted magnitude and the anticipated extreme one centered on GR law</td>
</tr>
<tr>
<td>12</td>
<td>$T$</td>
<td>Elapsed time (the duration amongst the last n events), computed from $T = t_n - t_1$</td>
</tr>
<tr>
<td>13</td>
<td>$\mu$</td>
<td>Mean time between key seismic events (characteristic events) among the last n events</td>
</tr>
<tr>
<td>14</td>
<td>dE1/2</td>
<td>Rate of the square root of seismic energy</td>
</tr>
<tr>
<td>15</td>
<td>C</td>
<td>Coefficient of variation</td>
</tr>
<tr>
<td>16</td>
<td>Mmean</td>
<td>Mean magnitude of the last n events</td>
</tr>
</tbody>
</table>

The variable b-value is from the prominent Gutenberg Richter (GR) law [63]. Panakkat and Adeli [37] suggested this variable and computed it using the least-squares method. However, given the lack of sturdiness in processing the infrequent earthquakes, Reyes,
Morales-Esteban and Martínez-Álvarez [26] proposed that the b-value can be computed using the maximum likelihood approach provided in Equation (1).

\[ b = \frac{\log(e_i)}{\left(\frac{1}{n}\right) \sum_{j=0}^{n-1} M_{i-j} - M_0} \]  

where \( e \) is the Euler’s number (around 2.718), \( n \) is the sum of events deemed prior to the event \( e_i \), \( M_{i-j} \) is the magnitude of \( e_i \), and \( M_0 \) indicates the cutoff magnitude. Based on the previously conducted research [26,39,64,65], the \( n \) value was set to 50 in this study. The other variables listed in Table 3 were computed based on their mentioned description.

Along with the most frequently previously utilized variables, this study used a different variable, named FD, in short-term earthquake prediction techniques. The preliminary hypothesis for this study is that the probability of a large earthquake can increase in the region, given the short distances to the active faults [66]. The kernel density estimation (KDE) analysis was applied to the faults data layer to compute the FD for transmitting the effect of the adjacent faults [67,68]. The search radius was set as a cardinal parameter of KDE analysis. The distance that magnifies the association between the neighborhood defects and the dependent variable is the absolute radius of the KDE analysis. This distance was determined by employing bivariate Moran’s I [69], as Yousefzadeh et al. [55] suggested. The distance that amplifies Moran’s I index between the magnitude of the biggest earthquake in the subsequent week (the dependent variable) and the distance from the faults (independent variable) was regarded as the absolute distance of the KDE analysis. The KDE was calculated for the entire research area, and the FD variable was determined by its value in each cell.

3.3. Models

As previously discussed, four NN models were utilized in this study. These are explained in the following.

3.3.1. Shallow Neural Network (SNN)

The present study utilized an MLP with three input, three hidden, and three output layers referred to as SNN. The MLP, often known as an ANN, is a commonly used ML technique for classifying and predicting problems in various sectors of science. Several studies have shown the superior prediction performance of MLP over other traditional methods for various purposes [41,62,70,71]. That is why it was selected as one of the NNs to be used in the present studies. The main components of MLP are the input layer, also known as the hidden layer, and the output layer. Input layers are responsible for feeding the network with the input variables, and output layers give the result. All the intermediate layers are considered hidden layers [72]. The number of layers is dependent on the number of input and output variables. The required hidden layers are obtained through a trial-and-error process. Backpropagation is the most commonly known algorithm used for training the ANN model to determine the weights. Backpropagation continues the network’s training until the least error is reached, at which point the successfully trained MLP model is utilized to classify the data [25,73].

3.3.2. Recurrent Neural Network (RNN)

To capture nonlinear correlations among data, DL methods such as RNN are used [74]. RNNs are considered the most effective in considering time-series data to make predictions, so it was selected in this study. Numerous prediction-based studies targeting different subject areas have documented an adequate performance of RNN [31,75–77]. Among all NNs, RNNs are the deepest and can produce and handle reminiscences of arbitrary-length sequences of input configurations [31]. RNN is capable of building relationships among units from a targeted cycle. RNN can also map from the history of earlier inputs to focus vectors on tenet and permit a collection of earlier inputs to be maintained in the network’s
3.3.3. Modular Neural Network (MNN)

MNN is a type of NN centered on the divide-and-conquer notion [47]. MNN comprises several NNs that are internally connected. MNN can enhance the reliability and generalization ability compared with single NNs. Due to this advantage of MNN, several researchers have successfully applied it for modeling purposes [78–80]. This was the reason for the selection of MNN for the present study. More complex systems are handled through MNNs. In this case, NNs are considered modules, with each solving a separate portion of the problem [58]. The integrator is responsible for dividing the problems into sub-problems and allocating those subproblems to intermediate modules, then gathering results from submodules to extract and combine the final result [81].

3.3.4. Deep Neural Network (DNN)

This study used a DNN feed-forward architecture to make earthquake predictions. DNN has been used in several previous studies and demonstrates good prediction performance [62,71,82–84]. DNN is a deep structure possessed by several hidden layers. This makes DNN a particular type of ANN; thus, it was selected for this study. DNN attempts to present a hierarchical depiction beneath data and comprehend the arrangements by constructing numerous layers of information-processing modules in hierarchical styles. The increased number of hidden layers of DNN results in satisfactory data transformations in its structures, thus extracting the most suitable hierarchical depiction of the data [85]. Speech recognition, object detection, and image classification are some domains in which DNN has shown significant improvements [86]. Moreover, the generality of DNNs, accessibility of computer hardware for augmenting their process, and open-source code, primarily when the task at hand handles abundant data, are some reasons for the augmented prominence of DNNs [87].

3.4. Prediction Methodology

The MNN, SNN, DNN, and RNN algorithms were trained and calibrated using the training and validation sections. The trained models were then used to determine the earthquake’s class for the testing data for the next seven days. Finally, the testing data were used to create a confusion matrix that accounted for the expected and projected classes.

All the NNs were calibrated to accomplish superior generalization while diminishing the overfitting problem. In this study, the purpose of the calibration process was to determine the finest pattern of the hyperparameters of every model. Previously, researchers have utilized various metaheuristics methods, for instance, artificial bee colony [88], coronavirus optimization [89], genetic algorithm [90], and particle swarm optimization [91], to deal with the problem of hyperparameter tuning. However, in this study, the weight decay parameter [92] and the dropout rate [93] were used to diminish the impact of overfitting for the NNs. The four models were tuned for the dropout rate, weight decay activation function, and the number of layers and nodes. Repeated modification, training, and validation of the models were carried out to accomplish the best models exhibiting superior performance (free of over and underfitting). Various hyperparameters of the models were iteratively modified, including the number of layers, the number of units per layer, dropouts, learning rate, and regularization. The optimal hyperparameters were chosen as the set of hyperparameters that resulted in the best model performance.

Four-fold cross-validation was used to conduct the calibration process, a commonly adopted practice for earthquake-prediction-related studies [26,94]. In particular, once the data was reserved for testing purposes (25%); the rest of the data were distributed into four equivalent parts. The training and validation were accomplished in four iterations. In each iteration, training was carried out using three parts, while the remaining part was used for
validation. The ultimate validation count was attained and computed from the mean of the four validation counts. Eventually, after the successful training and determination of the optimal hyperparameters using the validation count, the four trained models, RNN, MNN, SNN, and DNN, were used to make predictions based on the testing data. These were not fed to the models during training and validation.

3.5. Evaluation

In this study, the outcomes resulting from the application of the testing dataset were evaluated using the IGA, accuracy, and sensitivity measures. Primarily, IGA was utilized to gauge the descriptive ability of individual input variables and the extent to which each ML and DL algorithm can benefit from these variables. According to the IGA, the feature that significantly reduces entropy is considered a highly significant attribute for classification [95]. The IGA of an attribute A over the dataset S can be defined using Equation (2) [96].

\[
Gain(S, A) = Entropy(S) - \sum_{v \in values(A)} \frac{S_v}{S} Entropy(S_v)
\]

where \(Entropy(S)\) represents the entropy of the entire dataset, \(S_v\) represents the subset of \(S\) for which the attribute \((A)\) has a value \((v)\). More specifically, the \(Entropy(S)\) was calculated as a degree of impurity using Equation (3) [96].

\[
Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 p_i
\]

where \(p_i\) is the possibility that a specific occurrence goes to class \(i\). The number of classes is represented by \(c\).

In addition to IGA, after making the predictions using the trained models, the initial observations and expected values from the testing data were utilized to structure the confusion matrix. Firstly, the accuracy of the models was computed, utilizing the confusion matrix that indicates the number of events successfully predicted by the model. Then, Equation (4) was used to compute the accuracy.

\[
Accuracy = \frac{TP + TN}{TP + FP + TN + FN}
\]

where \(TP\) (true positive) represents the earthquake that occurred and was predicted correctly by the model. \(TN\) (true negative) indicates that there was no earthquake and that the model did not predict this. \(FP\) (false positive) means no earthquake happened, but the model predicted one, while \(FN\) (false negative) means an earthquake happened, but the model could not predict one. The confusion matrix [10] was used to describe these measures.

Subsequently, the sensitivity of the models was computed. This indicates how precisely a model predicted the earthquakes that occurred (positive class). Equation (5) was used to calculate the sensitivity.

\[
Sensitivity = \frac{TP}{TP + FN}
\]

Moreover, to comprehend the behavior of the DNN model, its Positive Predictive Value (PPV) and Negative Predictive Value (NPV) were calculated. Specificity denoted the degree of definite negative predictions of models and was computed using Equation (6).

\[
Specificity = \frac{TN}{TN + FP}
\]
PPV indicates the ratio of definite positive predictions amongst all the positive predictions, calculated using Equation (7).

\[
PPV = \frac{TP}{TP + FP}
\]  

(7)

NPV signifies the ratio of actual definite predictions out of all the negative predictions. Equation (8) was used to compute the NPV.

\[
NPV = \frac{TN}{TN + FN}
\]  

(8)

4. Results

4.1. Importance Analysis of the Input Variables

The outcomes of IGA are listed in Table 4. The higher the IGV, the better the information provided by the variable to the used models. The outcomes showed that, when predicting earthquakes, the additionally introduced FD (a nonlinear variable) in this study has a greater IGV than some of the other variables, such as X1, X2, X3, X4, X5, and depth. The elapsed time (T) is the variable with the highest IGV, demonstrating that it offers the most information to the used models.

Table 4. Information Gain Value (IGV) for the input variables.

<table>
<thead>
<tr>
<th>ID</th>
<th>Variables</th>
<th>IGV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X6</td>
<td>0.31</td>
</tr>
<tr>
<td>2</td>
<td>T</td>
<td>0.121</td>
</tr>
<tr>
<td>3</td>
<td>Latitude</td>
<td>0.071</td>
</tr>
<tr>
<td>4</td>
<td>b-value</td>
<td>0.062</td>
</tr>
<tr>
<td>5</td>
<td>( \chi_7 )</td>
<td>0.062</td>
</tr>
<tr>
<td>6</td>
<td>a-value</td>
<td>0.062</td>
</tr>
<tr>
<td>7</td>
<td>Mmean</td>
<td>0.061</td>
</tr>
<tr>
<td>8</td>
<td>( \eta )</td>
<td>0.06</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>0.055</td>
</tr>
<tr>
<td>10</td>
<td>Longitude</td>
<td>0.051</td>
</tr>
<tr>
<td>11</td>
<td>dE1/2</td>
<td>0.031</td>
</tr>
<tr>
<td>12</td>
<td>( \Delta M )</td>
<td>0.03</td>
</tr>
<tr>
<td>13</td>
<td>FD</td>
<td>0.025</td>
</tr>
<tr>
<td>14</td>
<td>( \mu )</td>
<td>0.02</td>
</tr>
<tr>
<td>15</td>
<td>Depth</td>
<td>0.00</td>
</tr>
<tr>
<td>16–20</td>
<td>X1, X2, X3, X4, X5</td>
<td>0</td>
</tr>
</tbody>
</table>

4.2. Used Different Variable Combinations

Table 5 shows the different variable combinations used to operate the four models with the input variables. For three different variable sets, the used variables are different. The first set did not contain the FD variable, whereas the second set did not contain the depth variable. However, all the variables were included for the third variable set.

Table 5. Examined variables sets.

<table>
<thead>
<tr>
<th>Variables Set</th>
<th>16 Seismicity Variables (Table 3), Longitude and Latitude</th>
<th>FD</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>*</td>
<td>-</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

* signifies that the variable(s) is included in a particular set.
4.3. Training and Testing Results of the Models

The ideal hyperparameters for the ML (SNN and MNN) and DL (RNN) techniques are shown in in Table 6. For all the models listed in Table 6, the optimal activation function was Logistic, while all of them had 1 hidden layer and 14 neurons. To obtain the best architecture for every model, the parameters of the models were altered, including the activation function, neurons, hidden layers, learning rate, epochs, etc.

<table>
<thead>
<tr>
<th>Variables Set</th>
<th>Model</th>
<th>Parameter and Structure Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Activation</td>
</tr>
<tr>
<td>1</td>
<td>MNN</td>
<td>Logistic</td>
</tr>
<tr>
<td></td>
<td>RNN</td>
<td>Logistic</td>
</tr>
<tr>
<td></td>
<td>SNN</td>
<td>Logistic</td>
</tr>
<tr>
<td>2</td>
<td>MNN</td>
<td>Logistic</td>
</tr>
<tr>
<td></td>
<td>RNN</td>
<td>Logistic</td>
</tr>
<tr>
<td></td>
<td>SNN</td>
<td>Logistic</td>
</tr>
<tr>
<td>3</td>
<td>RNN</td>
<td>Logistic</td>
</tr>
<tr>
<td></td>
<td>SNN</td>
<td>Logistic</td>
</tr>
</tbody>
</table>

The models were operated using different activation functions, but the Logistic activation function was found to be effective in terms of performance. As this study had a single input dataset, there was only a single hidden layer for the models. There were 14 datasets inside the single input, so the number of neurons was 14. The decay rate shows the extent of loss function for the output layer. The decay rate of MNN and RNN was different for the first and third set, respectively. However, for most of the models for different variables set, this was 0.

Table 7 contains the details of the structure of the ideal DNN, along with the shape of the output and the number of parameters. The selected ideal DNN structure was the one with 1 input layer, 6 hidden layers, and 1 output layer. In addition, there were four nodes in the output layer, together with an activation function (SoftMax) to forecast the four classes. There were two different kinds of layer types for DNN, namely, dense and dropout. There were different numbers of units inside the dense layer (512, 256, 4). Moreover, a constant value of 0.4 was selected for the dropout layer. The output shape depends on the number of units, so it was the same as the number of units. Finally, the model generated the number of parameters.

<table>
<thead>
<tr>
<th>Layer Type</th>
<th>Units</th>
<th>Activation</th>
<th>Value</th>
<th>Output Shape</th>
<th>No. of Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dense</td>
<td>512</td>
<td>Tanh</td>
<td>-</td>
<td>(None, 512)</td>
<td>127,543</td>
</tr>
<tr>
<td>Dropout</td>
<td>512</td>
<td>ReLU</td>
<td>0.4</td>
<td>(None, 512)</td>
<td>119,843</td>
</tr>
<tr>
<td>Dropout</td>
<td>256</td>
<td>ReLU</td>
<td>0.4</td>
<td>(None, 256)</td>
<td>85,843</td>
</tr>
<tr>
<td>Dropout</td>
<td>256</td>
<td>Tanh</td>
<td>0.4</td>
<td>(None, 256)</td>
<td>79,453</td>
</tr>
<tr>
<td>Dropout</td>
<td>512</td>
<td>ReLU</td>
<td>0.4</td>
<td>(None, 512)</td>
<td>116,321</td>
</tr>
<tr>
<td>Dropout</td>
<td>256</td>
<td>ReLU</td>
<td>0.4</td>
<td>(None, 256)</td>
<td>84,387</td>
</tr>
<tr>
<td>Dense</td>
<td>4</td>
<td>SoftMax</td>
<td>0.4</td>
<td>(None, 4)</td>
<td>516</td>
</tr>
</tbody>
</table>

Total parameters: 613,906; Trainable parameters: 613,906; Optimizer: RMSprop; Loss function: Categorical Crossentropy; Metric: Accuracy
The accuracies obtained by different ML and DL models for the training and testing data on the three different variable sets are listed in Tables 8 and 9, respectively. The SNN model obtained the best overall training accuracy, followed by MNN, RNN, and DNN for the three sets of variables, as shown in Table 8. In contrast, the MNN model outperformed other models in terms of testing accuracy. After the MNN model, RNN and SNN models had the highest testing accuracy values, respectively. The DNN model achieved the least testing accuracy among all the models, as listed in Table 9.

Table 8. Accuracy of the used models for training data.

<table>
<thead>
<tr>
<th>Variables Set</th>
<th>MNN</th>
<th>DNN</th>
<th>RNN</th>
<th>SNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>84.1%</td>
<td>74.1%</td>
<td>73.7%</td>
<td>83.40%</td>
</tr>
<tr>
<td>2</td>
<td>83.3%</td>
<td>81.7%</td>
<td>82.1%</td>
<td>89.30%</td>
</tr>
<tr>
<td>3</td>
<td>78.5%</td>
<td>81.7%</td>
<td>81.7%</td>
<td>88.20%</td>
</tr>
</tbody>
</table>

Table 9. Accuracy of the used models for testing data.

<table>
<thead>
<tr>
<th>Variables Set</th>
<th>MNN</th>
<th>DNN</th>
<th>RNN</th>
<th>SNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>84.9%</td>
<td>64.9%</td>
<td>83.3%</td>
<td>78.5%</td>
</tr>
<tr>
<td>2</td>
<td>85.7%</td>
<td>74.1%</td>
<td>81.7%</td>
<td>81.7%</td>
</tr>
<tr>
<td>3</td>
<td>83.7%</td>
<td>73.7%</td>
<td>82.1%</td>
<td>81.7%</td>
</tr>
</tbody>
</table>

4.4. Validation of the Models

The sensitivity obtained for the four classes for MNN, DNN, RNN, and SNN is exhibited in Table 10. SNN was the most accurate model for forecasting high-magnitude earthquakes, whereas DNN was the most accurate model for low-magnitude earthquakes. High sensitivity values of models for class 1 (earthquakes between Mw 3 and 3.7) mean that all models performed efficiently in estimating the class 1 earthquakes, whereas low sensitivity values of models for class 2 (earthquakes between Mw 3.7 and 4.5) mean that almost all models showed comparatively poor performance in estimating the class 2 earthquakes. However, for classes 3 and 4, the performances of the models were relatively moderate, as can be witnessed in Table 10.

Table 10. Outcomes of the sensitivity analysis of the used models.

<table>
<thead>
<tr>
<th>Variables Set</th>
<th>Model</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MNN</td>
<td>86.20%</td>
<td>52.30%</td>
<td>72.50%</td>
<td>68.90%</td>
</tr>
<tr>
<td>2</td>
<td>MNN</td>
<td>82.70%</td>
<td>77.30%</td>
<td>81.40%</td>
<td>61.40%</td>
</tr>
<tr>
<td>3</td>
<td>MNN</td>
<td>79.20%</td>
<td>76.30%</td>
<td>42.80%</td>
<td>68.90%</td>
</tr>
<tr>
<td>1</td>
<td>DNN</td>
<td>94.00%</td>
<td>66.90%</td>
<td>82.00%</td>
<td>68.40%</td>
</tr>
<tr>
<td>2</td>
<td>DNN</td>
<td>98.00%</td>
<td>66.90%</td>
<td>76.70%</td>
<td>73.10%</td>
</tr>
<tr>
<td>3</td>
<td>DNN</td>
<td>88.70%</td>
<td>59.30%</td>
<td>78%</td>
<td>67.30%</td>
</tr>
<tr>
<td>1</td>
<td>RNN</td>
<td>91.50%</td>
<td>60.30%</td>
<td>79%</td>
<td>88.30%</td>
</tr>
<tr>
<td>2</td>
<td>RNN</td>
<td>88.00%</td>
<td>60.30%</td>
<td>76.50%</td>
<td>86.80%</td>
</tr>
<tr>
<td>3</td>
<td>RNN</td>
<td>91.50%</td>
<td>52.30%</td>
<td>79.50%</td>
<td>86.80%</td>
</tr>
<tr>
<td>1</td>
<td>SNN</td>
<td>95.00%</td>
<td>56.30%</td>
<td>68%</td>
<td>88.30%</td>
</tr>
<tr>
<td>2</td>
<td>SNN</td>
<td>88.00%</td>
<td>48.30%</td>
<td>69.60%</td>
<td>95.80%</td>
</tr>
<tr>
<td>3</td>
<td>SNN</td>
<td>89.70%</td>
<td>56.30%</td>
<td>75%</td>
<td>88.30%</td>
</tr>
</tbody>
</table>

Moreover, as the performance of the DNN model was relatively better than the other models, the PPV, NPV, and specificity for variables set 3 for the DNN model were also calculated and are presented in Table 11. The PPV value of 84.10% for class 4, predicted by the DNN model, is promising. However, as shown in Table 11, there is a trade-off between NPV and specificity, signifying that the chances of false-positive predictions are more likely for high specificity.
Table 11. Different measures for the DNN model on variables set 3.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity</td>
<td>87.30%</td>
<td>77.30%</td>
<td>85.70%</td>
<td>80.90%</td>
</tr>
<tr>
<td>PPV</td>
<td>88.10%</td>
<td>76.50%</td>
<td>84.10%</td>
<td>84.10%</td>
</tr>
<tr>
<td>NPV</td>
<td>86.10%</td>
<td>76.10%</td>
<td>84.50%</td>
<td>84.10%</td>
</tr>
</tbody>
</table>

5. Discussion

Earthquake is among the major catastrophes that cause significant casualties and damage to infrastructures [9,32,97]. Hindu Kush, Himalayas, and the Karakoram Mountains range in the North of Pakistan are amongst the most seismically lively regions globally. Unfortunately, the North of Pakistan has a record of suffering mild to rigorous and, at times, overwhelming earthquakes, primarily due to its presence at the boundary of the Indian and Eurasian plates [27,35]. Although naturally occurring earthquakes cannot be stopped, their prediction and adequate protection measures can prevent the loss of human life and several valuables.

A mechanism suitable for earthquake prediction that can offer positive predictions is a pressing need at present. A system that is competent in earthquake prediction should predict the accurate location, exact magnitude range, a specific incidence timespan, and the likelihood of incidence. However, no such comprehensive earthquake prediction system has existed to date [8,27]. Several researchers have tackled this problem individually, and investigations have been conducted to predict some of the aforementioned characteristics. However, a holistic study is missing.

The present study attempts to predict earthquakes of Richter magnitude greater than or equal to 3 in Northern Pakistan using four models (two ML and two DL models): DNN, RNN, SNN, and MNN, combined with 20 seismicity indicators from previously conducted investigations. The arithmetically computed seismicity indicators from the former earthquakes signify the seismic trend of the region. These indicators were used as input to the approaches for earthquake prediction. The seismic indicators were mathematically computed using the records from the earthquake catalog of IIEES and USGS.

Interestingly, some former investigations [8,10] suggested that the cutoff magnitude centered on the GR law must be computed in advance. Moreover, following the calculation of the cutoff magnitude, all events below the computed cutoff magnitude must be excluded. The reason for doing this is to safeguard that inadequate and deceptive information is not fed into the model [10]. Nevertheless, this approach of computing and utilizing the cutoff magnitude results in dropping a big chunk of the data, which is not suitable for operating the models. Therefore, the present research did not apply the cutoff magnitude to the data but excluded all the events with a Richter $M_{W}$ less than 3.

The seismic variables used in this study are listed in Table 3. These include 16 traditionally used variables and the variables of longitude, latitude, and depth, and a new variable FD. The variables of depth and FD were identified by IGA as being spatial variables of intermediate significance. Therefore, to further investigate the role of the FD variable along with depth, the four used algorithms were operated with distinct sequences of input variables presented in Table 5.

It is important to point out that the variables (X1, X2, X3, X4, and X5) with low IGA values were not removed from the input vector, which is the opposite of the widespread practice of excluding such variables. The reasoning for not excluding these variables stems from the notion that a variable’s effectiveness correspondingly hinges on the capability of the fundamental model. A practical model would benefit from the valuable information coming from less important variables and could offer improved predictions. It is also important to mention that several architectures with distinct hyperparameters were tested to locate the perfect DNN model. The model with the most excellent validation accuracy was chosen as the ideal model. The details of the perfect structure of the DNN model for variable set 3 are listed in Table 7.
Contemplating the two variables of depth and FD, it appears that MNN and SNN were able to utilize the underlying information held by these variables. However, the DNN and RNN could not successfully exploit these two variables. Meanwhile, MNN was the most successful ML algorithm in terms of utilizing the data stored by the depth and FD variables. The deep neural structure of MNN could be a reason for such efficacy by MNN. Furthermore, this structure can obtain valuable information from minimally associated independent input variables.

To analyze the performance of the used ML and DL models in this study, the accuracy and the sensitivity measures were computed. Accuracy was selected as a generic metric for examining the general implementation of the models. The purpose of using sensitivity was to understand how each model performed for each of the four classes. More specifically, the overall performance of the used models is represented by accuracy, whereas the ability of the models to accurately perceive the earthquakes that occurred is indicated by sensitivity.

Sensitivity values for classes 1 and 4 are higher than those for classes 2 and 3. According to Table 10, the highest sensitivity values for classes 1 and 4 occurred when the models were using variables set 1. The most straightforward explanation for this is that the magnitudes for these classes are more closely linked to the FD variable. The low sensitivity values for classes 2 and 3 relatives to classes 1 and 4 for all models are most likely due to significant noise in the magnitude data.

Even though MNN performs much better than the other considered models in terms of accuracy, the sensitivity values were not the highest for MNN, as can be seen from Tables 8–10. The best models in terms of sensitivity for predicting classes 1, 2, 3, and 4 were SNN on the variables set 1, DNN on the variables set 1, RNN on the variables set 2, and MNN on the variables set 1. A closer look at the listed sensitivity results in Table 10 reveals that optimal predictions of classes 1, 3, and 4 occurred when the models were exploiting the additional FD variable. This specifies the appropriateness and practicality of the variable, specifically for predicting large-magnitude earthquakes. Classes 3 and 4 can be predicted with the probability of good sensitivity values using the FD variables.

Moreover, the accuracy for both training and testing data for SNN and MNN, which are ML techniques, is surprisingly more than the DL techniques of DNN and RNN, as can be witnessed from Tables 8 and 9. Technically, this is impossible because ML employs fewer layers for analysis, whereas DL uses more layers, which leads to more accurate results [86,98]. The constructed ML models demonstrated a markedly higher prediction capability than the DL models. This may mean that the DL models considered the classification problem to be linear while it was nonlinear in its nature. Additionally, the different hyperparameter settings affect the performance of ML and DL models [99]. Even though the hyperparameters were changed, and optimal models were constructed during the training process, the choice of the hyperparameters used for the models might not be compatible with the data considered for the analysis.

The DNN model outperformed other models regarding the sensitivity values for class 1. In some cases, such as for class 2, other models outperformed DNN. Amongst the used models, DNN had the most extraordinary complexity. Therefore, in certain situations, this lower sensitivity might be due to its greater parametrization, as was previously observed in an investigation conducted by Mignan and Broccardo [75]. The authors suggested that, due to the structured and tabular nature of catalog data and the insufficient number of calculated features, NN models with shallow structures might compete with DNNs in earthquake prediction. A few other investigations also observed such an adherence regarding the predictive ability of SNNs and deep learning [100,101]. However, the involvement of several hidden layers in DNN allows for one to understand features at distinct levels of perception [98]. This was a key reason for the better performance of the DNN model as compared to other models.

After learning the superiority of the four models, the prediction abilities of the models per class were assessed. The outcomes demonstrated that when the objective was to utilize a general model to predict both high- and low-magnitude earthquakes, RNN or SNN
could be a better choice. Nevertheless, as per the sensitivity analysis of classes 1 and 3, MNN or DNN can be a suitable option, as they can better detect and sense low- and moderate-magnitude earthquakes than other models. Regardless of the size of the network and the substantial number of parameters required for the training of the DNN model, the outcomes showed that this complicated model was considerably effective in employing the information from the FD and depth variables. Additionally, the SNN model outperformed other models in predicting high-magnitude earthquakes. This performance of SNN was unexpected, as its structure is comparatively straightforward, and the associations between the input variables and the higher-magnitude earthquakes are quite complicated.

From the viewpoint of a disaster management organization, any model attempting to predict earthquakes should not produce false alarms, as they can cause massive panic and economic loss [46]. Centered on this notion, specificity, \(NPV\), and \(PPV\) were computed for all four classes of the DNN model, and the results are shown in Table 11. The specificity, \(PPV\), and \(NPV\) values for class 1 are relatively higher than those in other classes. For class 4, the \(PPV\) value of 84.10% for the DNN model is very promising.

The advantages of the proposed methods for earthquake magnitude prediction become clear after a comparison with similar studies published previously in the same region and other parts of the world. Asim, Martínez-Álvarez, Basit and Iqbal [46] utilized four ML techniques, RNN, PRNN, LPBoost Ensemble, and Random Forest, to predict earthquakes in the Hindukush region of Northern Pakistan using eight seismic indicators. These indicators were centered on the eminent geophysical facts of GR’s inverse law, dissemination of distinctive earthquake magnitudes and seismic quiescence, and mathematically computed from the earthquake catalog of the region. The authors observed that the LPBoost Ensemble (79%) and PRNN (79%) have the highest accuracy for validation data, while the LPBoost Ensemble (65%) and RNN (64%) have the highest accuracy for the testing data. Compared to this previously conducted study, the present study used more seismic variables, and the accuracy of most of the models for validation and testing data was relatively higher. For example, the testing accuracy of SNN for variable set 2 was 89.30%, while the testing accuracy of MNN for variable set 1 was 84.9%. Moreover, the DNN model also outperforms all the used models in the previous study in terms of specificity.

Moreover, Aslam, Zafar, Khalil and Azam [34] used eight seismic features based on seismological notions, such as the dissemination of typical earthquake extents, the eminent geophysical specifics of Gutenberg–Richter’s inverse law, and seismic quiescence for earthquake prediction in Northern Pakistan. The authors developed a hybrid classification system based on a Support Vector Regressor (SVR) and Hybrid Neural Network (HNN) for earthquakes greater than or equal to 5.5. The accuracy assessment of the SVR-HNN model revealed that the model has a specificity of 86% and a sensitivity of 61%. However, the model attained an accuracy of 79% for the testing data. Compared to the results of the present study, this previously conducted study not only used fewer variables but also attained relatively less accuracy for the developed model compared to the sensitivity results of most of the models of the present study.

Furthermore, Yousefzadeh, Hosseini and Farnaghi [62] introduced and investigated the impact of spatial parameters on the performance of four ML models, specifically, SNN, DNN, DT, and SVM, for predicting future earthquakes’ magnitude in Iran. The results showed that the SVM and DT model achieved the highest accuracy for training data while DT and DNN achieved the highest accuracy for testing data. The authors also used different variable sets for different models and found that DNN and SVM models better detect intermediate- and high-magnitude earthquakes than other models. As the present study introduced different models than the mentioned study, the outcomes of this study also concluded that the DNN model is better than the other practiced models. This establishes the superiority of DNN compared to other conventional techniques.

Based on the outcomes of the present study, it can be stated that the adopted methods can be assessed for their accuracy in other locations by considering spatial variables. Then,
a comparison can be carried out to highlight the limitations or adequacies of the methods for the two cases.

6. Conclusions

In the present study, two ML algorithms of SNN, MNN, and two DL algorithms of RNN, and DNN, were utilized to predict short-term earthquakes in the Northern Part of Pakistan, a seismically active region. An additional seismic variable named FD was used, along with the most frequently utilized seismic variables defined in the earlier research. Three different variable sets containing different variables were separately utilized to check the performance of the individual model in response to different variables. When used in different variable sets, the variables of depth and FD facilitated the accuracy of the used ML and DL earthquake prediction models to a certain extent.

The outcomes demonstrated adequate performances of DNN and SNN in predicting the low-magnitude earthquakes, respectively. The DNN model demonstrated an accuracy of 98% for variables set 2 and class 1, whereas the accuracy was 94% for variables set 1. The SNN model showed an accuracy of 95% for variables set 1 and class 1. However, the performance of both RNN and SNN was more encouraging in dealing with high-magnitude events. The SNN model exhibited an accuracy of 95.80% for variables set 2 and class 4, while the RNN model demonstrated an accuracy of 88.30% for variables set 1 and class 4.

6.1. Limitations

The present study only used historical earthquake records from two open-access sources. For future studies, it is important to acquire data from multiple sources to better represent the documented earthquake events in the area. Moreover, it could also be beneficial to acquire some field data from the relevant authorities in the area. Furthermore, different input variables can also be used along with the variables used in this study to check for their influence on the model performance.

6.2. Implications and Future Research

The model presented in this study can be used to predict earthquakes with low and high magnitudes, as advocated by the accuracy assessment results. Moreover, earthquake prediction systems can be a great deal of help for the concerned authorities. An alert triggered by such a system can allow for controls to activate supplies and halt critical damage-causing systems such as electricity and nuclear power plants to avoid fatalities.

Future research can assess the usability of further DNN architectures, such as CNNs, for earthquake prediction and contrast their functioning with other techniques to establish the most superior model. Additionally, the effect of the FD variable on the performance of other contemporary and conventional methods can also be assessed in the context of other seismic regions.


Funding: This research was funded by the National Key Research and Development Program of China, grant number 2019YFE0127700; the National Natural Science Foundation of China, grant number 42071321.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Data Availability Statement: The data used in this study can be obtained from the corresponding author on a reasonable request.

Conflicts of Interest: There is no conflict of interest to report relevant to this work.

References


32. Tariq, A.; Yan, J.; Gagnon, A.S.; Riaz Khan, M.; Mumtaz, F. Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest. *Geo-Spat. Inf. Sci.* **2022**, 1–19. [CrossRef]


35. Last, M.; Rabinowitz, N.; Leonard, G. Predicting the maximum earthquake magnitude from seismic data in Israel and its neighboring countries. *PLoS ONE* **2016**, *11*, e0146101. [CrossRef]


70. Hu, Z.; Rao, K.R. Particulate air pollution and chronic ischemic heart disease in the eastern United States: A county level ecological study using satellite aerosol data. *Environ. Health* 2009, 8, 26. [CrossRef]


