

A new method to promptly evaluate spatial earthquake probability mapping using an explainable artificial intelligence (XAI) model

Ratiranjan Jena ^{a,b}, Biswajeet Pradhan ^{a,c,d,*}, Shilpa Gite ^{e,f}, Abdullah Alamri ^g, Hyuck-Jin Park ^c

^a Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

^b GIS & Remote Sensing Center, Research Institute of Sciences and Engineering, University of Sharjah, Sharjah 27272, United Arab Emirates

^c Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, South Korea

^d Earth Observation Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia

^e Artificial Intelligence and Machine Learning Department, Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune 412115, India

^f Symbiosis Centre of Applied AI (SCAAI), Symbiosis International (Deemed) University, Pune 412115, India

^g Department of Geology and Geophysics, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

ARTICLE INFO

Article history:

Received 22 March 2022

Revised 12 October 2022

Accepted 12 October 2022

Available online 17 October 2022

Keywords:

Earthquake probability

Explainable AI

Machine learning

GIS

ABSTRACT

Machine learning (ML) models have been extensively used in several geological applications. Owing to the increase in model complexity, interpreting the outputs becomes quite challenging. Shapley additive explanation (SHAP) measures the importance of each input attribute on the model's output. This study implemented SHAP to estimate earthquake probability using two different types of ML approaches, namely, artificial neural network (ANN) and random forest (RF). The two algorithms were first compared to evaluate the importance and effect of the factors. SHAP was then carried out to interpret the output of the models designed for the earthquake probability. This study aims not only to achieve high accuracy in probability estimation but also to rank the input parameters and select appropriate features for classification. SHAP was tested on earthquake probability assessment using eight factors for the Indian subcontinent. The models obtained an overall accuracy of 96 % for ANN and 98 % for RF. SHAP identified the high contributing factors as epicenter distance, depth density, intensity variation, and magnitude density in a sequential order for ANN. Finally, the authors argued that an explainable artificial intelligence (AI) model can help in earthquake probability estimation, which then open avenues to building a transferable AI model.

© 2022 International Association for Gondwana Research. Published by Elsevier B.V. All rights reserved.

1. Introduction

As a natural disaster, earthquake can result in financial losses and casualties. Earthquake probability is the spatial occurrence of earthquakes over a certain threshold of magnitude (4.5 Mw). Probability varies between 0 and 1 as a continuous value. The Indian subcontinent is one of the major seismic prone areas out of the three main active seismic belts in the world (Ring of Fire, Alpine–Himalayan Belt, and Oceanic Ridge Belt). In fact, half of the Indian subcontinent is prone to high-magnitude, damaging earthquakes. Indian plate colliding with Eurasian plate generated the Alpine–Himalayan belt and is the main source of earthquakes

in this region. The earthquakes experienced from the last few decades in this belt indicate the possibility of future earthquakes that may hit in this highly populated region. In the last decade, several machine learning techniques, such as self-organizing maps (SOM), artificial neural network (ANN), convolutional neural network (CNN), long- and short-term memory (LSTM), and generative adversarial network (GAN) (Wang et al., 2017; Wang et al., 2019; Berhich et al., 2020; Khan et al., 2020; Debnath et al., 2021), have been applied to probability as well as several other applications.

Successful utilization of ANNs have been conducted to solve the classification problems and pattern recognition in various fields of research, including image recognition (Adeli and Hung, 1993; Adeli and Hung, 1994; McIlraith and Card, 1997), optimal recognition algorithms (Bourbakis et al., 2007; Gopych, 2008; Hung and Adeli, 1993; Khashman and Sekeroglu, 2008), medicine (Lian and Lu, 2007; Wersing et al., 2007), earthquake hazard (Panakkat and Adeli, 2007), and geomorphology (Karunanithi et al., 1994). Previous research shows that probability is a measuring indicator for

* Corresponding author at: Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia.

E-mail address: Biswajeet.Pradhan@uts.edu.au (B. Pradhan).

earthquakes (Jena et al., 2020), which helps identify the earthquake potential zones (Wang et al., 2006). In a specific geographic position, the probability of an earthquake can be estimated within a period when the earthquake intensity crosses a particular threshold of VI (Jena et al., 2020). The literature features several studies on earthquake magnitude prediction (Panakkat and Adeli, 2007), probability, and hazard mapping (Jena et al., 2020; Sitharam and Vipin, 2011; Sitharam et al., 2015). Some studies conducted on earthquake probability modelling are based on popular machine learning methods, such as support vector machine (SVM), ANN, CNN (Jena et al., 2020; Xiong et al., 2021) and LSTM (Zhao and Takano, 1999). ANN and CNN models have been used for earthquake probability studies by utilizing seismicity indicators (Panakkat and Adeli, 2007; Jena et al., 2020). Several researchers proposed different neural network models such as a backpropagation algorithm (Priambodo et al., 2020), a recurrent neural network (Schäfer and Zimmermann, 2007; Shcherbakov et al., 2019), and a radial-basis function (Adeli and Karim, 2000; Liu et al., 2007; Mayorga and Arriaga, 2007). Panakkat and Adeli (2007) put forward the categorical earthquake magnitude prediction of a predefined (Mw 6) threshold of magnitude.

Shcherbakov et al. (2019) developed a methodology to estimate the probabilities of earthquakes above a specific magnitude. They implemented the Bayesian method, with the assumption of Epidemic Type Aftershock Sequence process. Their study is helpful to compute probabilities of a sequence of earthquakes for large events. In a recent work, Jena et al. (2020) proposed an integrated machine learning model for earthquake probability in the Aceh province in Indonesia. They achieved an accuracy of 84 % on earthquake probability mapping. Sitharam and Vipin (2011), Sitharam et al. (2015) conducted earthquake probabilistic assessment in India by adopting the topographic gradient technique. None of the outputs of the aforementioned models applied in earthquake research are interpreted to understand the impact of factors on obtained accuracy. Owing to the increasing complexity of these models, their results are quite challenging to interpret. This situation motivated the authors of the present study to apply a recently derived explainable artificial intelligence (XAI) technique that belongs to the machine learning community called Shapley additive explanation (SHAP) (Lundberg et al., 2018; Lubo-Robles et al., 2020). SHAP was employed to study the impact of several factors that are most useful for probability estimation in neural network and random forest (RF) architecture trained to differentiate earthquake from non-earthquake.

This research aims to demonstrate how to interpret the outputs of classification models based on an XAI method. Thus, an evaluation methodology was developed for feature selection based on SHAP (Lundberg and Lee, 2017) to interpret the output of ANN and RF to appropriately select factors for probability mapping. ANN and RF models were used because the aim was to make a comparison between two machine learning modelling approaches as well as measure the impacts of factors on models' accuracy (Tyagi et al., 2019). Machine learning models are widely used in several fields of application. There are several models used for the prediction or classification purposes. The SHAP explainability approach describes the internal function among nodes to estimate the importance, interaction, stability, local contribution, compactness, distance between explanations, and predict the information of every factor on a single prediction, respectively.

As a novelty, this study explains a comprehensive assessment of factors importance, threshold, stability, and interaction which is not yet been recorded in the earthquake studies. This is the first ever study on earthquake probability mapping and is explained through the XAI technique. To the best of the authors' knowledge, no work has been conducted on earthquake probability estimating models regarding their explainability to identify limitations and

enhance the model accuracy. As this study is the first XAI based earthquake probability study, we choose two basic models such as ANN and RF to understand the interpretability and applicability of XAI. This comparison was conducted because of an accuracy difference observed in probability mapping. Although ANN and RF are two well-known models, however, this study is explaining the factors interaction and factors stability for a global earthquake probability study through which the accuracy can be improved. The study estimated the threshold for the highly contributing factors that contribute towards the spatial probability of earthquakes. The SHAP allocated the values for every factor considered for the earthquake probability mapping. The factors' contribution was utilized to arrange and encode them based on their importance and SHAP values. In this scenario, a set of factors were chosen on the basis of the SHAP values. The factors were sorted, aiming for good accuracy. The initial features were then selected for further study.

The main contributions of this work are as follows: (1) ANN and RF were applied for earthquake probability estimation, and the outputs were compared on the basis of the SHAP values. (2) SHAP was implemented as a feature selection technique to select suitable contributing factors out of eight factors and estimate the quantitative results of the impact of every feature on earthquake probability mapping. (3) Probability assessment was conducted again without the high contributing factors to understand their impact on accuracy. The organization of this article is as follows. Section 1 provides a brief introduction on SHAP for feature selection and the structures of ANN and RF. Sections 2 and 3 are the data and methodology sections, respectively. Section 4 presents the experiment results, and they are discussed in Section 5. This work is concluded in Section 6.

2. Data

The input data used in this research were collected from different sources including the Geological Survey of India and Bhuvan portal in India. A digital elevation model with a spatial resolution of 30 m (USGS) and a scale of 1:250,000 was used for classification. The historical earthquake locations were sourced from the Geological Survey of India (Dasgupta et al., 2000).

Given the numerous works on earthquake hazard and probability (Iyengar and Ghosh, 2004; Nath et al., 2008; Boominathan et al., 2008; Kanth and Iyengar, 2006), several relevant factors were derived for the earthquake probability assessment. This study is conducted to estimate the earthquake spatial probability based on prediction techniques such as ANN and RF. The data used here are considered as conditioning factors to estimate the earthquake spatial probability. Earthquake probability values vary between 0 and 1. Because the probability needs to be presented in the stretched format through maps that represents the continuous values, where we assume 0 as low and 1 as high. The top 10 % values close to 1 and 0 can be considered as high and low and subsequently values in between can be stated as moderate. Detailed information about lineaments, faults, shear zones, and several features in India was documented by Dasgupta et al. (2000). An exhaustive lithology information was processed through the digitization and georeferencing of available maps and the clustered earthquake information using ArcGIS 10.6 software (Fig. 1). Finally, eight factors were considered for the earthquake probability mapping and classification. Table 1 summarizes the list of input parameters and their sources. To obtain samples for the training dataset, historical earthquakes were used, and some non-earthquake points were randomly created using ArcGIS 10.6 software. A total number of 504 samples were obtained in the region for training and testing of the models. A total of 504 samples were chosen, out of which

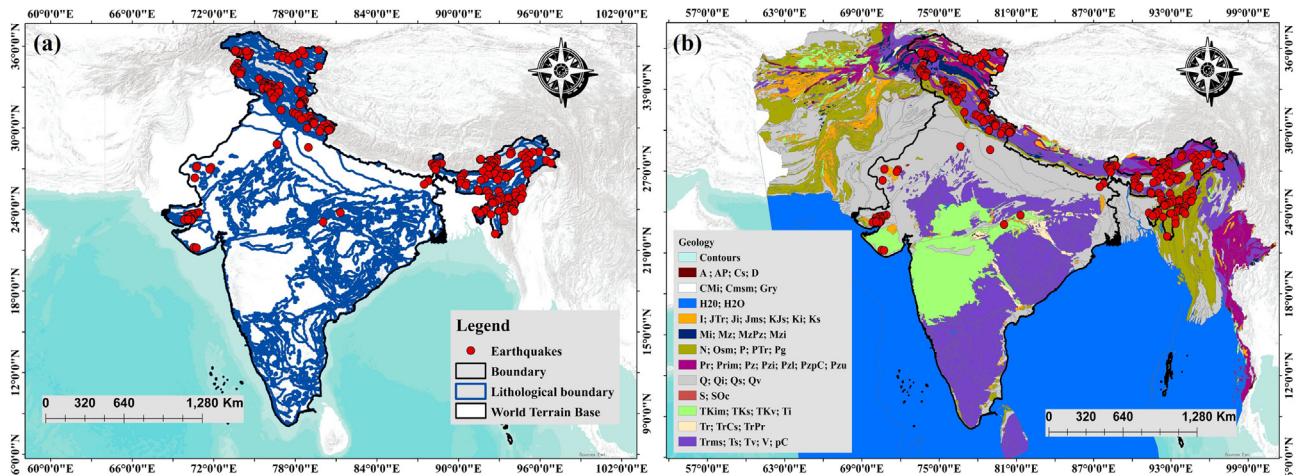


Fig. 1. a) Represents the study area and lithological boundaries in India, and b) shows the geological map of India, (A: Archean, Ap: Archean and proterozoic, CMi and Cmsm: Cambrian sedimentary and metamorphic rocks; Gry: Unidentified rocks; Cs: Carboniferous sedimentary rocks, D: Devonian rocks, H20: other regions, Ji, Ki and I: Jurassic and cretaceous igneous; Mi: Mesozoic igneous; JT: Triassic and Jurassic rocks JMs: Jurassic metamorphic and sedimentary, KJs: Jurassic and Cretaceous sedimentary, Ki and Ks: Cretaceous igneous and sedimentary, Mi: Mesozoic igneous; P and MzPz: Paleozoic and Mesozoic metamorphic, Mzi: Mesozoic intrusive, N: Neogene sedimentary, Osm: Ordovician metamorphic and sedimentary, Pg: Paleogene sedimentary, Prim: Permian igneous and metamorphic rocks; Ptr: Permian-Triassic rocks; Pr: Permian rocks, Pz: undifferentiated Paleozoic, Pzi: Paleozoic igneous rocks, Pzl: Lower Paleozoic, Pzu: Upper Palaeozoic metamorphic, PzPc: Palaeozoic undivided Precambrian, Q and Qi: Quaternary sediments and igneous rocks, Qs and Qv: Quaternary sand and volcanic, S and SOC: Silurian, TKim: Cretaceous and Tertiary igneous and metamorphic, TKs: Cretaceous and Tertiary sedimentary, TKv: Cretaceous and Tertiary volcanic, Ti: Tertiary igneous, TrCs, TrPr, TrMs: Upper Carboniferous - Lower Triassic sedimentary, Upper Permian - Lower Triassic sedimentary and Upper Mesozoic-Lower Triassic sedimentary; Tims: Triassic igneous and sedimentary, Ts, Tv and V: Tertiary sedimentary, Pz: Precambrian (Modified from Jena et al., 2021).

Table 1
Data used for the earthquake probability estimation and SHAP interpretation.

Factors	Code names	Sources	Resolution and scale	References	Importance
Slope Elevation	slope_elevation	United States Geological Survey (USGS)		The importance of all the factors is mentioned in. Shcherbakov et al. (2019) Sitharam et al. (2011, 2015) Jena et al., 2020a,b, c,d; Jena et al., 2021	The landscape processes are controlled by elevation and slope. These processes may reform the crustal faults with complicated tectonics.
Magnitude density	mag_dens	Geological map of India, GSI	30 m and 1:250,000		The occurrence probability of a specific magnitude can be understood through magnitude clusters.
Depth density	depth_dens	USGS Earthquake Catalogue			Shows the source of event and the depth information on the fault zone.
Epicenter distance	Epic_dis				Provides a zone of the clusters of events. This identifies the high probable zone.
Intensity variation	intens_ind				Current intensity measure shows the degree of consequences at high probable locations.
PGA density	pga_dens				PGA density provides ground acceleration information useful for hazard estimation.
Lithology	lithology_fault_den	Geological map of India, (https://www.gsi.gov.in), (bhuvan.nrsc.gov.in), (USGS World Geologic Map)			High magnitude events found in thrust faults, whereas earthquakes in strike-slip faults are more destructive.
					Very solid granites are mostly found in fault zones that transmit energy better than others.

252 non earthquake points were artificially generated while 252 are earthquakes. A ratio of 70 % and 30 % was randomly split into training and testing sets, respectively. Next, a data normalization method was employed to speed up the gradient descent optimization and activation function (Jena et al., 2020).

Table 1. Data used for the earthquake probability estimation and SHAP interpretation.

3. Methodology

3.1. Heat map generation for data summary

A heat map generally provides a visual summary of data and helps perceive complex information (Zhao et al., 2014). In simple

terms, this map represents the relationship between attributes that signifies the impact of attributes on others. The correlation values among factors lie between -1.0 and 1.0, as presented in Fig. 2. The positive correlation shows an effect, but the negative one shows the opposite effect. A zero correlation shows no effect on the other attributes. In supervised learning, attributes that are important are generally considered in determining the target (Guo et al., 2020). Therefore, the heat map is a plot that portrays the correlation of the attributes and plots based on the values (Zhao et al., 2014). The color strip shows the range of values of the correlation with colors (Fig. 2). The correlation among the data was established through a heat map, which ranges from -1 (negative correlation) to 1 (positive correlation) (Table 2).

Table 2. Correlation values for all 9 important factors.

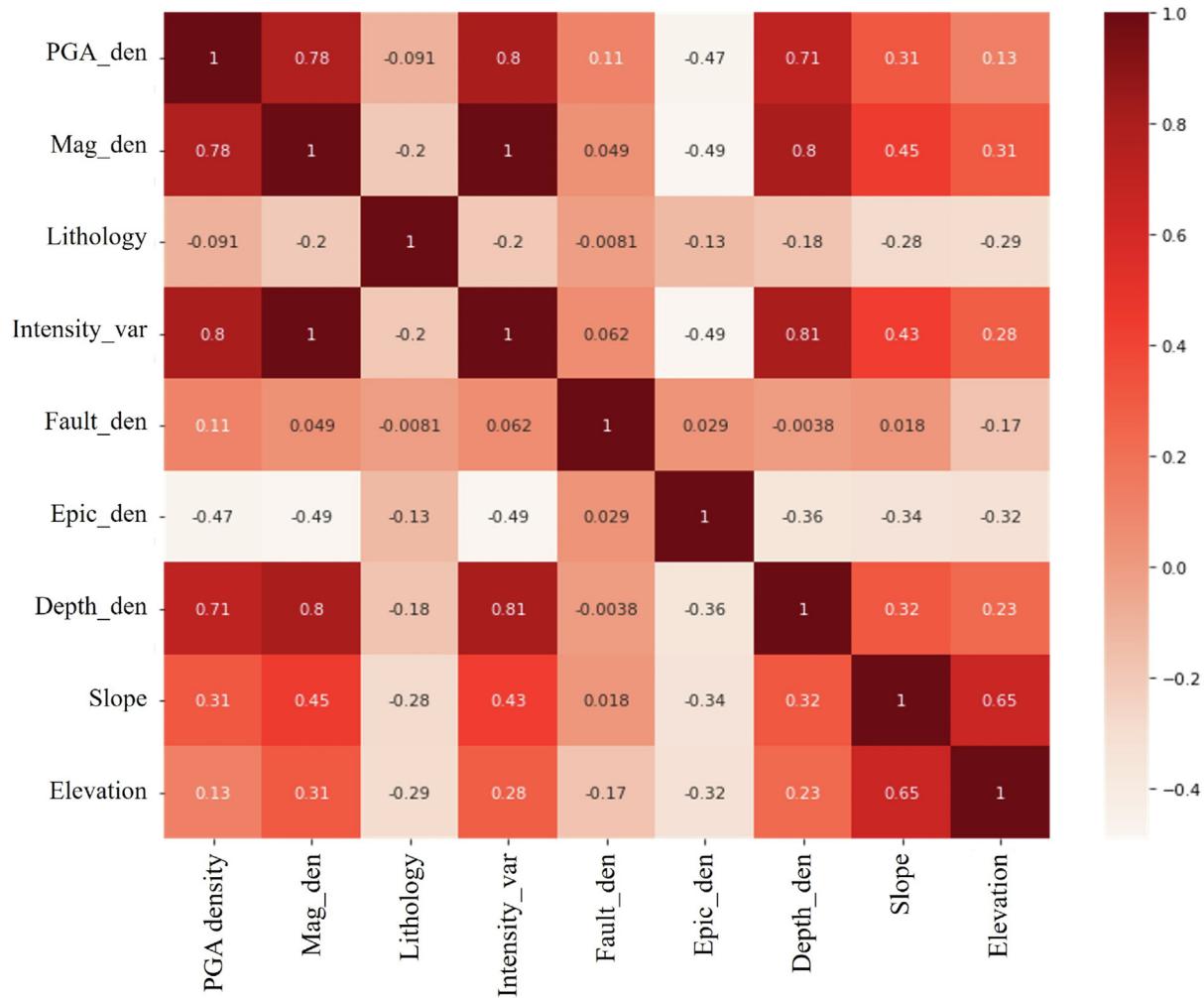


Fig. 2. Heat map presents the correlative factors and a comparison matrix to understand the importance of factors.

Table 2
Correlation values for all 9 important factors.

	PGA density	Magnitude density	Lithology	Intensity variation	Fault density	Epicenter distance	Depth density	Slope	Elevation
PGA density	1.000000	0.778130	-0.090914	0.804408	0.106673	-0.468737	0.707423	0.311059	0.127127
Magnitude density	0.778130	1.000000	-0.202541	0.997613	0.049425	-0.489645	0.804095	0.447108	0.306434
Lithology	-0.090914	-0.202541	1.000000	-0.196818	-0.008145	-0.130715	-0.180729	-0.282689	-0.287173
Intensity variation	0.804408	0.997613	-0.196818	1.000000	0.061889	-0.487680	0.810060	0.429931	0.284404
Fault density	0.106673	0.049425	-0.008145	0.061889	1.000000	0.028615	-0.003810	0.017941	-0.171081
Epicenter distance	-0.468737	-0.489645	-0.130715	-0.487680	0.028615	1.000000	-0.355250	-0.338303	-0.320368
Depth density	0.707423	0.804095	-0.180729	0.810060	-0.003810	-0.355250	1.000000	0.318249	0.232511
Slope	0.311059	0.447108	-0.282689	0.429931	0.017941	-0.338303	0.318249	1.000000	0.651396
Elevation	0.127127	0.306434	-0.287173	0.284404	-0.171081	-0.320368	0.232511	0.651396	1.000000

3.2. Implementation of SHAP

SHAP was used for identifying and interpreting the influence of the factors for earthquake probability models (see Fig. 3). Data were randomly divided into two sections for training and testing purposes to build the networks. The training (70 %) and test (30 %) datasets were utilized to develop the classification model and assess the performance of the models. SHAP was applied to the classification networks of ANN and RF to create additive attri-

butes, which were then utilized to ascertain relevant factors and determine the effects on classification outputs. In the following sections, the technique for earthquake probability using ANN and RF were compared using SHAP in detail. SHAP adopts the game theory concept, thereby assigning importance value to factors on the basis of the impact on the prediction by considering the presence and absence of factors during estimation (Lundberg and Lee, 2017). SHAP uses a linear additive feature attribute method to explain complex models, which is an XAI model. It is expressed as:

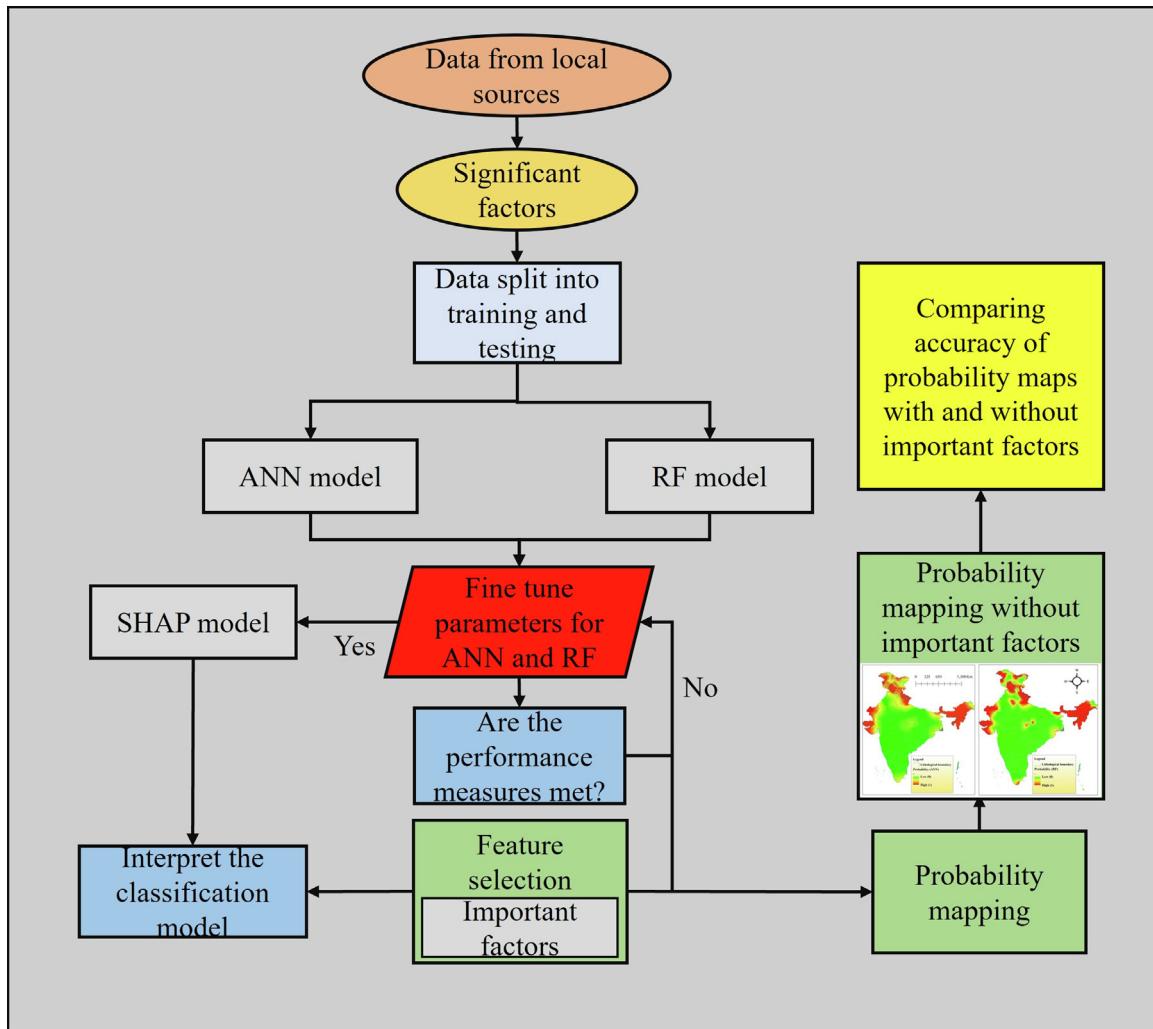


Fig. 3. Overall methodology shows the prediction models and interpretation using SHAP.

$$f(a) = g(a') = \emptyset_0 + \sum_{j=1}^J \emptyset_j a'_j \quad (1)$$

where, $f(a)$ is the prediction model, $g(a')$ is the explanation model, J represent the earthquake attributes, \emptyset_0 is the model prediction when no attributes are considered, \emptyset_j denote the SHAP values, and a'_j is the input vector indicating the presence of a particular attribute. SHAP is the model that can describe local and global models practically. The interpretable model generally considers the closeness to the instance. Several approaches are available, such as deep SHAP, tree SHAP, and kernel SHAP, to estimate the SHAP values for various applications (Lundberg and Lee, 2017). In the model agnostic estimation approach, kernel SHAP applies linear LIME and SHAP values to construct a local explanatory model (Ribeiro et al., 2016). Thus, in the present study, a kernel SHAP was implemented for accurate estimation of the SHAP values, which could improve sample efficiency.

The summary plot denotes the ranking of factor's importance and the factor effects. The factor's effect shows the contribution of each factor in the classification, while the factor's importance demonstrates the contribution of factors to the classification performance. Each point shows SHAP value in the summary for a factor and an observation. SHAP estimates the local interpretability by using the impact of each factor on classifying the individual class. Apart from the global interpretability, this capability allows to

analyse the detail about correctly classified and misclassified classes. This led to a deep comprehension regarding the model and the dataset.

3.3. ANN model

The data used in this study were characterized by N_v training patterns (X_p, T_p) , where p represents the pattern number. To reduce the difficulty on notation analysis, an augmented vector was assigned $X_p(N+1)$, which can control the hidden and output units. However, X_p is consistent with the input dimension of the Pth training pattern, whereas the output dimension Y_p is consistent with the Pth training pattern (Haykin, 2009). The details on ANN modeling using MLP classifier can be seen in previous studies (Nazzal et al., 2008; Park et al., 2013). The model utilized a rectified linear unit (ReLU) as an activation function in both output and input units. The input of (j th) hidden units, $net_p(j)$, can be mathematically expressed as:

$$net_p(j) = \sum_{k=1}^{n+1} W_{hi}(j, k) X_p(k) \quad 1 \leq j \leq N_h \quad (2)$$

Here, the output activation function $Op(j)$ for the P th pattern can be expressed as.

$$O_p(j) = f(net_p(j)). \quad (3)$$

For this model, the authors chose a nonlinear activation function, which is a sigmoid function that can be presented as:

$$f(\text{net}_p(j)) = \frac{1}{1 + e^{-\text{net}_p(j)}} \quad (4)$$

where, N and K represent the input and index, respectively, and $W_{hi}(J, K)$ portrays the connecting weight between the input (K th) and hidden units (J th). The sklearn metrics were implemented to check the performance of the classifier based on root-mean-square-error (RMSE) and accuracy.

MSE can be represented as:

$$E = 1/N \sum_{p=1}^{N_v} E_p = 1/N \sum_{p=1}^{N_v} \sum_{i=1}^M [T_p(i) - Y_p(i)]^2 \quad (5)$$

and,

$$E_p = \sum_{i=0}^M [T_p(i) - Y_p(i)]^2 \quad (6)$$

where, E_p and T_p are considered as the P th pattern error and requisite output, respectively. The classification error for the i th unit can be expressed as:

$$E_i = 1/N_v \sum_{p=1}^M [T_p(i) - Y_p(i)]^2 \quad (7)$$

For the i th output and P th training pattern, the expression can be written as:

$$Y_p(i) = \sum_{k=1}^{N+1} W_{oi}(i, k) X_p(k) + \sum_{j=1}^{N_h} W_{oi}(i, j) O_p(j) \quad (8)$$

where, $W_{oi}(i, k)$ represents the weight of input-to-output nodes, and $W_{oh}(i, j)$ denotes the weights of the hidden-to-output nodes.

3.4. RF architecture

RF is a supervised classification algorithm (Provost et al., 2017). It generates decision trees randomly and merges them to form a forest. A decision model is used to improve accuracy. RF is also considered as an ensemble model used for classification. RF produces a root node feature through random splitting that can distinguish from the decision tree. To improve accuracy, RF selects random features. RF is faster than the bagging and boosting methods (Pal, 2005). In some cases, RF provides better results when using SVM, which is a neural network.

The core aim of RF is to generate a huge number of decision tree models for training $\{h(X, \theta_k); k = 1, \dots\}$, which are unrelated. All the decision trees predict the target sample individually using a classification algorithm. Finally, the output shows the mode of sample classification. RF can be improved by generating unrelated training sets that can decrease the model variance. Several training sets, $\{h_1(X), h_2(X), \dots, h_k(X)\}$, are acquired by sample training and then generates RF model. A voting process takes place to generate the output of RF, as shown in Eq. (9).

$$H(x) = \arg_z^{\max} \sum_{i=1}^k I(h_i(x) = Z). \quad (9)$$

$H(x)$ indicates the RF model, and h_i is a decision tree model. The indicative function is expressed as $I(h_i(x) = Z)$, and the output variable is Z .

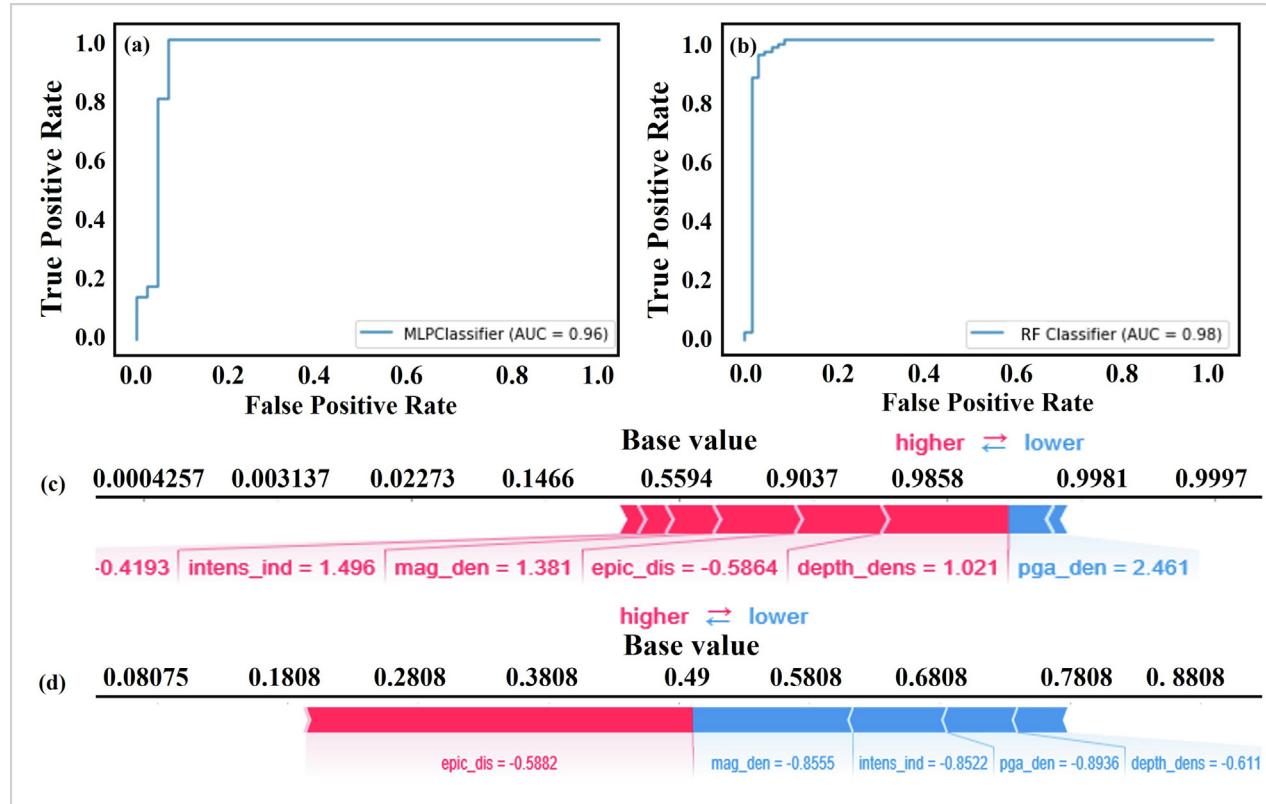


Fig. 4. (a) Model accuracy for earthquake probability for ANN, (b) RF, (c) SHAP force plot used for the local interpretation of initial prediction for ANN, and (d) RF model. “The blue feature attributions pushed to lower outcome, whereas the red feature attributions outcomes pushed to higher than the “base value”. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.5. Performance metrics

To assess the proposed models, evaluation metrics were used (Jena et al., 2020). Models diagnostic ability was represented through ROC. ROC portrays the true positive rate against false positive. The area under the curve (AUC) depicts the prediction values. The prediction was better when the AUC was larger (Jena et al., 2020). The prediction ability of ANN and RF models were demonstrated based on ACC (overall accuracy), recall, F1-score, and precision. They can be expressed as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F1\text{ score} = 2TP / (2TP + FP + FN) \quad (12)$$

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

4. Results

In this section, earthquake probability assessment was performed using ANN and RF, and the ROC graph for both was plotted (Fig. 4a). Approximately, 96 % accuracy was observed in ANN; 98 % accuracy in RF (Fig. 4b). Next, the feature explanations achieved by

SHAP were graphically portrayed. A bar plot illustrated the global relevance of the factors. To understand the factors' influence in a model, the shifting of an individual characteristic was plotted using a partial dependence plot (García and Aznarte, 2020). Force plot was generated to depict the local interpretation for the initial prediction. Several other visualizations are also possible because the SHAP values are specific to every prediction classification. A good alternative to the partial dependence plot is SHAP dependence graphs, which reflects a logical relationship among the features. The authors highlighted the global significance using force plots, summary graphs, dependence plots, and heat map (García and Aznarte, 2020). Fig. 4c and 4d present force plots of earthquake probability obtained by ANN and RF models, respectively. These figures show that multiple factors work together that help push the output from the “base” to the “predicted” value during the initial prediction. Factors that have high classification value are shown in red, and blue indicates a lower value. For the earthquake class, the sample has a prediction of 0.96 for ANN, and the baseline is 0.55. Similarly, a prediction accuracy of 0.98 and a baseline of 0.49 are observed for RF. The depth density, magnitude density, and epicenter distance can improve the final result of ANN, whereas epicenter distance is the only one for RF. All other factors diminish the prediction accuracy.

The authors utilized another visualization technique to illustrate a summary plot that integrates the features' impacts with significance. The SHAP values of a factor and a sample that impact the model output is plotted on the summary plot. Fig. 5a and 5b demonstrate the summary plots for the earthquake classes for

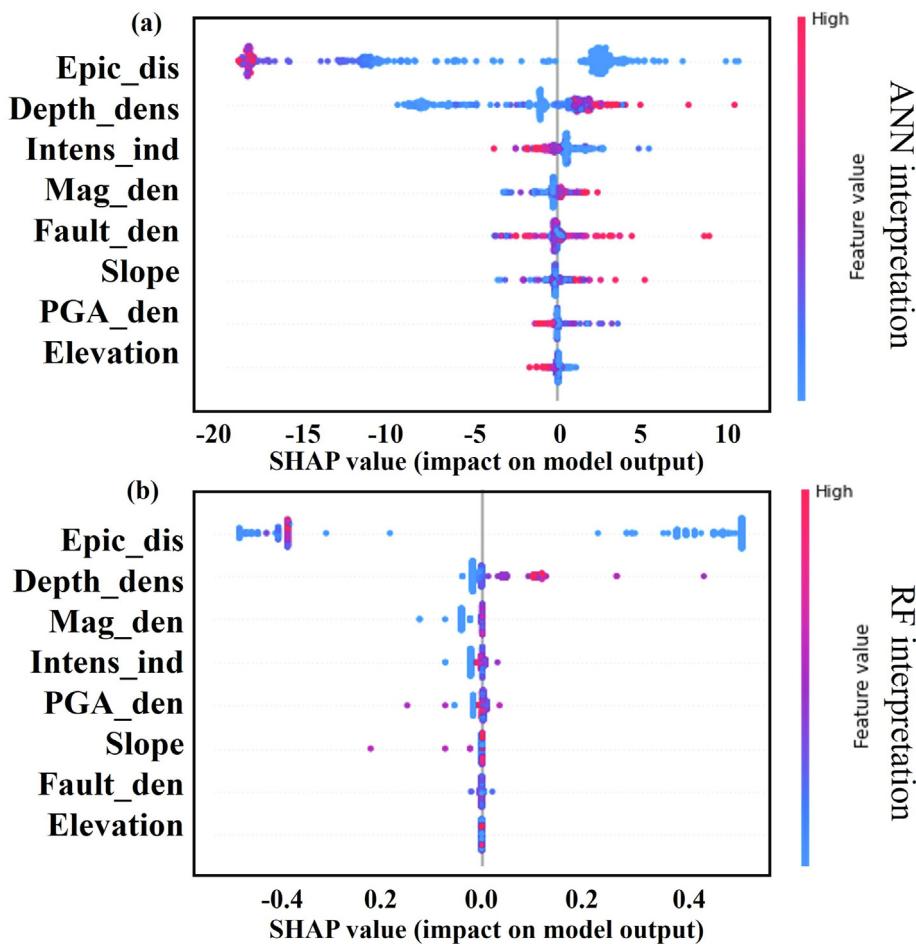


Fig. 5. Summary plot based on: (a) SHAP values for all features for prediction using ANN; and (b) SHAP values for all features for prediction using RF shows the importance of factors on model's output.

different labels of ANN and RF models, respectively. The x-axis represents the SHAP value, and the y-axis shows the factors. The color ranges from low to high, denoting the magnitude of a feature's value. SHAP values can be achieved per feature using overlapping points jittered in y-axis direction based on the features' significance values ranked in order (García and Aznarte, 2020). According to Fig. 5a, epicenter distance, depth density, intensity variation, and magnitude density contents significantly impact the classification. Similarly, in Fig. 5b, epicenter distance, depth density, magnitude density, and intensity variation are important. However, if the sequence of both figures are compared, the RF model works differently with magnitude density and intensity variation as the priority factors. The x-axis with positive values in red color represents high importance. The feature saturation is negatively associated with the target factor.

The average absolute SHAP value of features can be considered in the bar graph, as shown in Fig. 6. The x-axis values in the bar graph depict the magnitude of the difference in log-odds. As observed in Fig. 5, the SHAP values are globally allocated to the features. In the current scenario, all features are continual, and their influence is arranged vertically on classification. Epicenter distance clearly has the most significant feature on earthquake classification (class 1), whereas depth density is crucial on class 1. The high values of other features show a similar effect in the classification of earthquake and non-earthquake. Fig. 6a and 6b show the magnitude differences and their impacts for ANN and RF, respectively.

A SHAP dependence plot was presented through a scatter plot, which portrays the effect of a single feature on the model predic-

tions. The interaction between depth density and epicenter distance was plotted for both ANN (Fig. 7a) and RF models (Fig. 7b) to better understand the relationships based on SHAP values. If the feature values are constant, the dependence plot denotes the expected output. This plot can reveal the dependency of the model on a feature by illustrating the shifting of outputs with the variation of features. This is presented for epicenter distance and depth density in Fig. 7, where the change of colors reflects the change of values. The magnitude of the features is presented in the x and y-axes. In fact, the effect of depth density is depicted for variation of epicenter distance from -1 to 0. The red color denotes the high values of epicenter distance, while the blue represents low values. In Fig. 7a, values for depth density are negative when depth density is less than 0. The low possibility of the background class can be observed because of the low SHAP values for depth density and epicenter distance.

The idea behind the SHAP feature importance estimation is to understand the impact of factors on the model's output. Because the SHAP interpretation reveals the factors with high importance, and large Shapley values can contribute to the model's earthquake probability mapping. In this study, the average impact of each factor was estimated with respect to the impact of magnitude. According to the results, epicenter distance, depth density, intensity variation, and magnitude density have high impacts, respectively. To estimate the global importance, this study summed up all the absolute Shapley values for a particular factor across the data. In general, SHAP feature importance is based on the magnitude of factor attributions, whereas permutation feature impor-

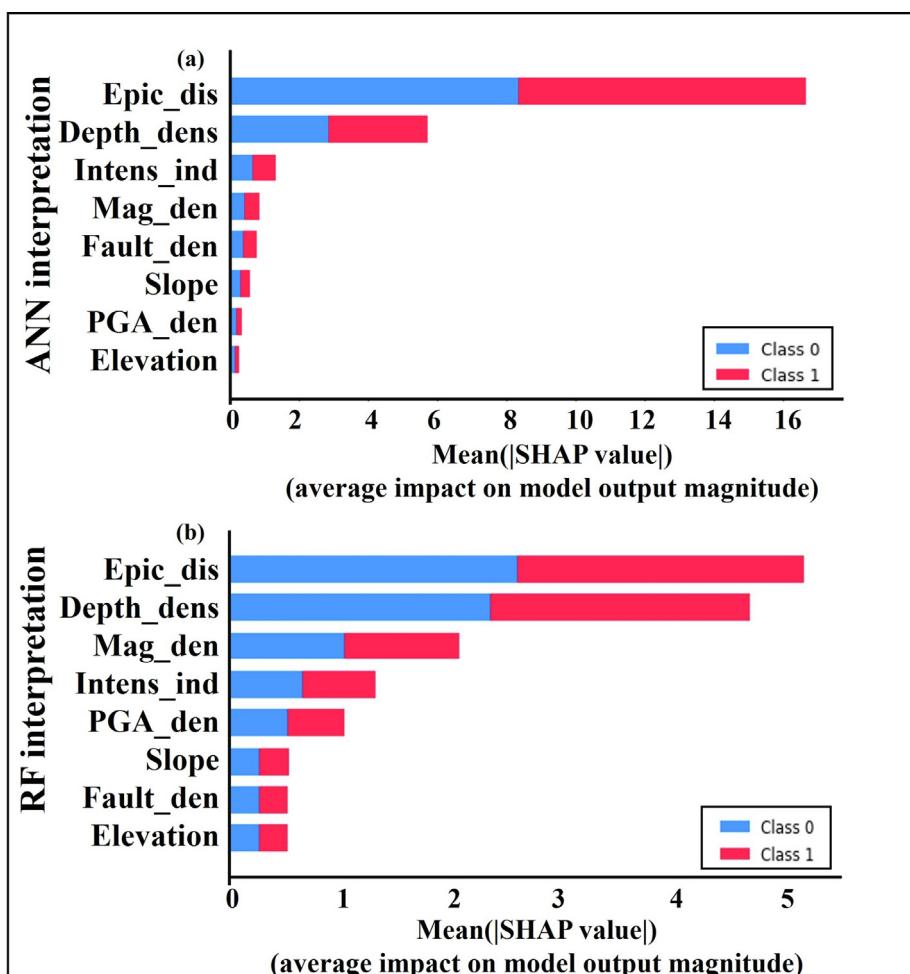


Fig. 6. (a) Feature importance for ANN, and (b) RF model based on classes (Binary classification).

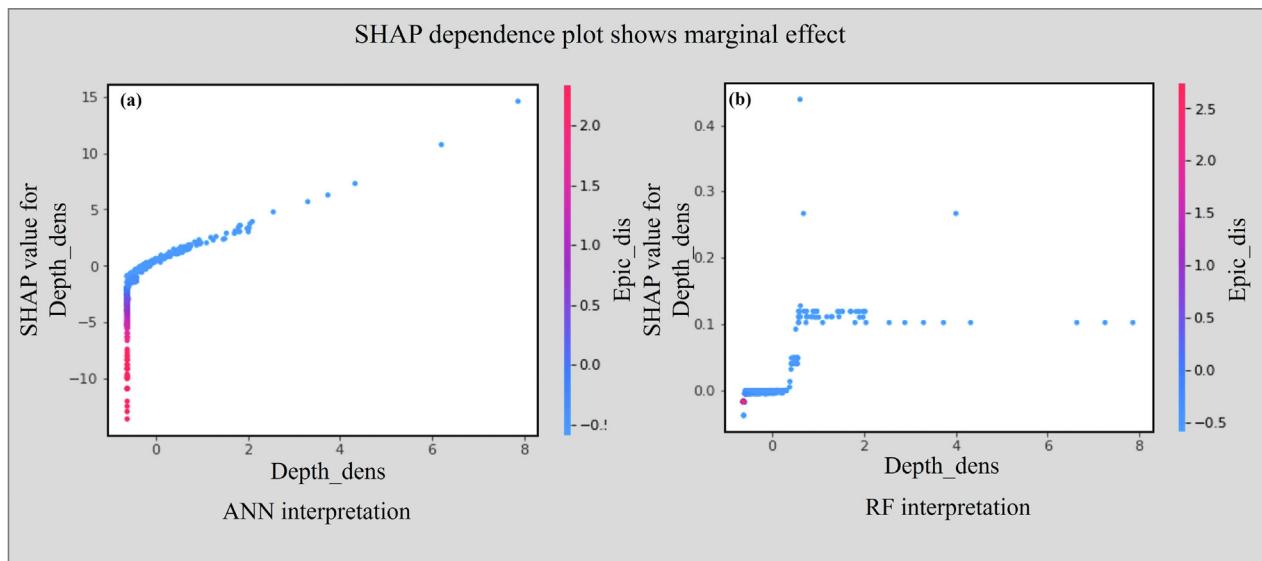


Fig. 7. (a) SHAP dependence plot shows the marginal effect of epic_dis on depth_dens for ANN, and (b) RF model.

tance depends on model performance. The feature importance plot only contains importance, which is not informative. On the other hand, the summary plot shows the Shapley value for each point on a factor and an instance. The factors are ordered in the graph based on their importance. To the end, an accuracy comparison was conducted with and without the important features for earthquake probability estimation, which shows the low accuracy in case of probability results without important features.

5. Discussion

Eight significant factors were used to implement and verify the proposed ANN and RF models for earthquake probability assessment, i.e., earthquake and non-earthquake classification. RF shows better result than ANN in case of NW and Northern areas, while similar results could be observed in NE region. However, Allah Bund Fault and east-trending Kachch Mainland Gedi fault are responsible for the high probability while in the Himalayan collision zone accountable for high probability. SHAP is confirmed to be a suitable method for selecting the important features to understand its contribution to a model's classification based on the allocated SHAP values for interpreting the effect on the probability map. Additionally, the features such as epicenter distance, depth

density, intensity variation, and magnitude density show high contribution to the model outputs for both ANN and RF. By contrast, other features have low contribution. However, the analysis reveals that the sequence of importance is different for the two models, which helps them achieve good accuracy (Fig. 6). All the features were fed to the ANN and RF models to check the influence on classification. Subsequently, the low contributing features and high contributing features were fed individually into the model and achieved the outputs. At the end, the accuracy of both models was compared based on SHAP values by using various visualization techniques.

The accuracy of ANN (96 %) is less than that of RF (98 %) based on the SHAP values with all the factors considered for classification. Similarly, after identifying the important factors, the same prediction classification was performed by inputting only low contributing features. The output accuracies are 83 % (ANN) and 87 % (RF), and the quantitative results are presented in Tables 3 and 4. The developed models perform well for earthquake probability based on high-ranked features. Similarly, using low-ranked features for the models can lead to low accuracy. The resulted maps are presented without high-ranked features. Fig. 8a and 8b show the ROC plot for both the models without high-ranked features. The accuracy is reduced by 13 % and 11 % accuracy for the ANN

Table 3

Quantitative results attained the ANN model without the high-contributed features.

Classes	Precision	Recall	F1-Score	Support	OA
Class 0	0.84	0.73	0.78	71	
Class 1	0.79	0.88	0.83	80	
accuracy			0.81	151	
macro avg	0.81	0.80	0.81	151	0.80
weighted avg	0.81	0.81	0.81	151	

Table 4

Quantitative results attained the RF model without the high-contributed features.

Classes	Precision	Recall	F1-Score	Support	OA
Class 0	0.96	0.82	0.88	88	
Class 1	0.79	0.95	0.86	63	
accuracy			0.87	151	0.87
macro avg	0.87	0.89	0.87	151	
weighted avg	0.89	0.87	0.88	151	

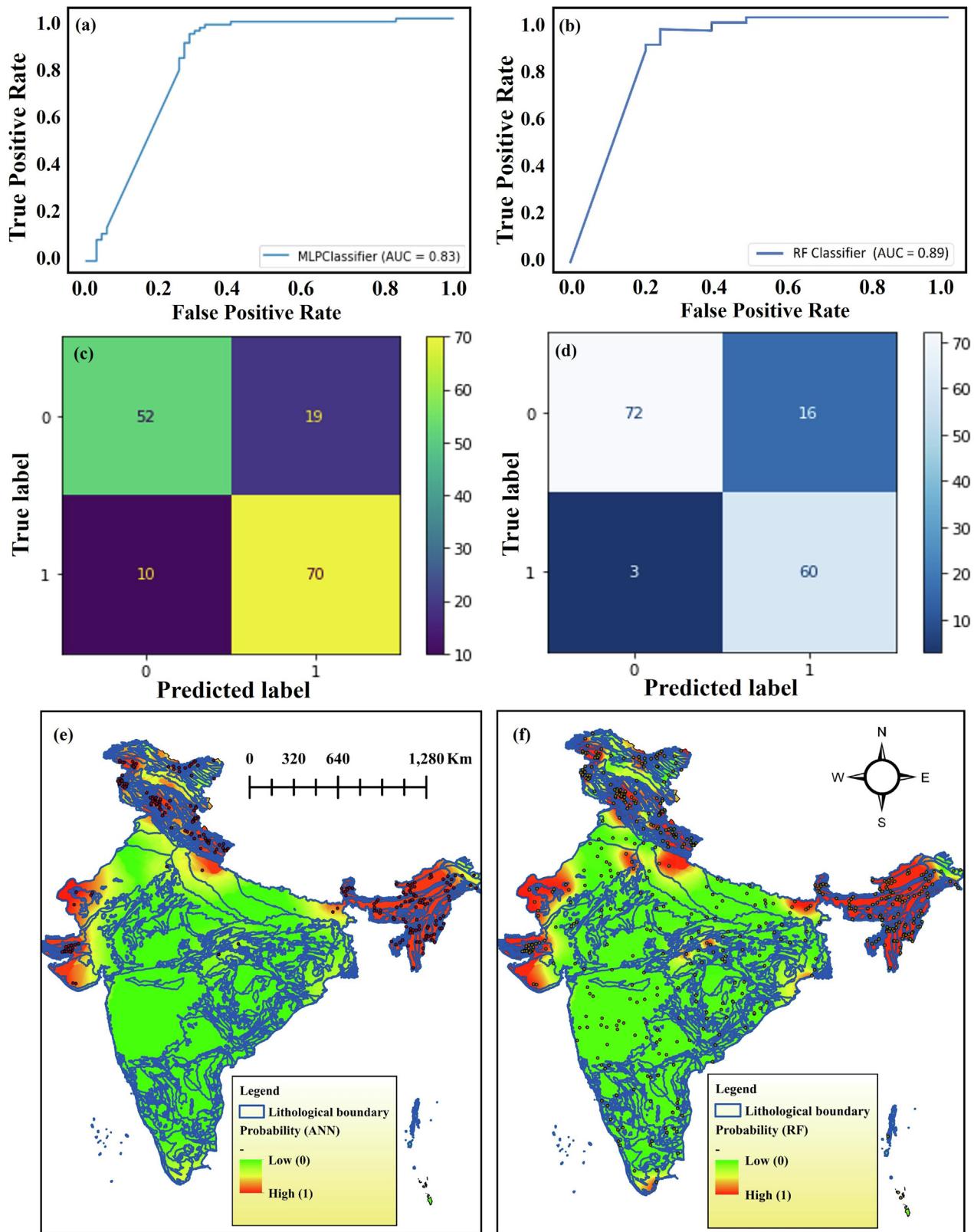


Fig. 8. (a) Prediction accuracy for ANN, and (b) RF; (c) confusion matrix for ANN, and (d) RF; and prediction classification for (e) ANN, and (f) RF without high contributing factors.

and RF models, respectively (Table 4). Fig. 8c and d present the confusion matrix for both the models where the true positives are 70 and 60 for the ANN and RF models. Fig. 8e and 8f show the earthquake probability results in the form of geospatial maps.

As seen in Fig. 8, the models cannot produce good probability maps because the important factors are removed from the analysis. Nevertheless, the results show that the RF model predicts better than ANN. Moreover, the most probable locations are found in the

Himalayan areas. Features with low contribution cannot cooperate properly with the model parameters and result in misclassified pixels (predicted more FPs) that subsequently lead to low accuracy.

Both models predict and misclassify earthquake classes as non-earthquake. By contrast, when using all the input features, especially the high contributing ones, the suggested RF technique

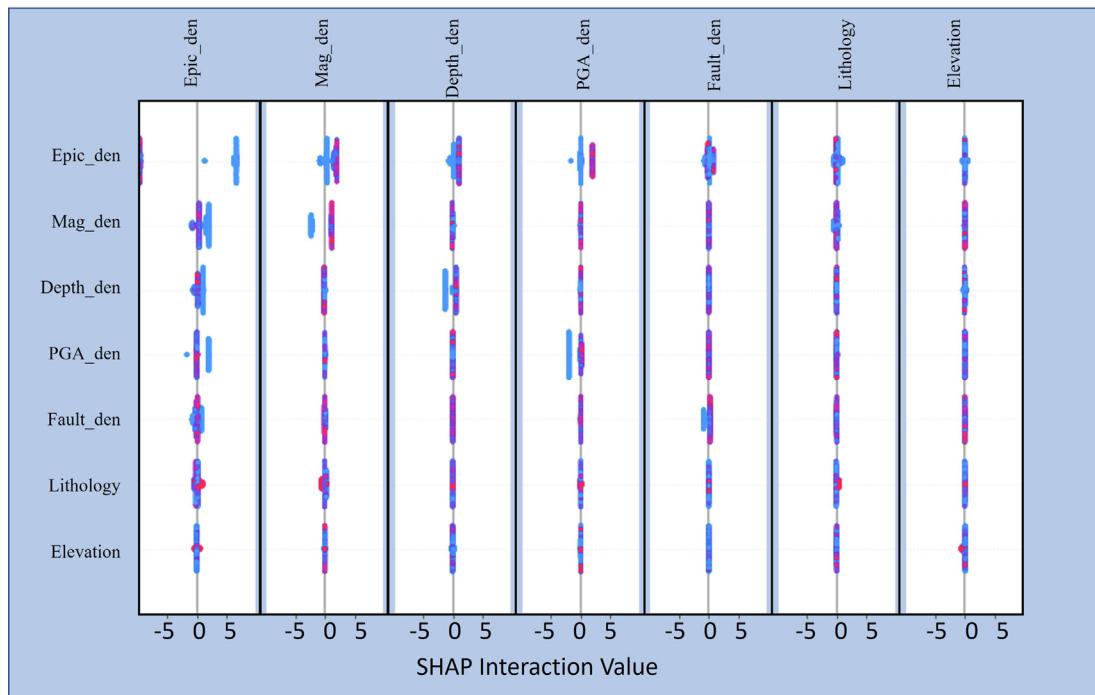


Fig. 9. SHAP interaction plot for the RF model (accuracy 97 %).

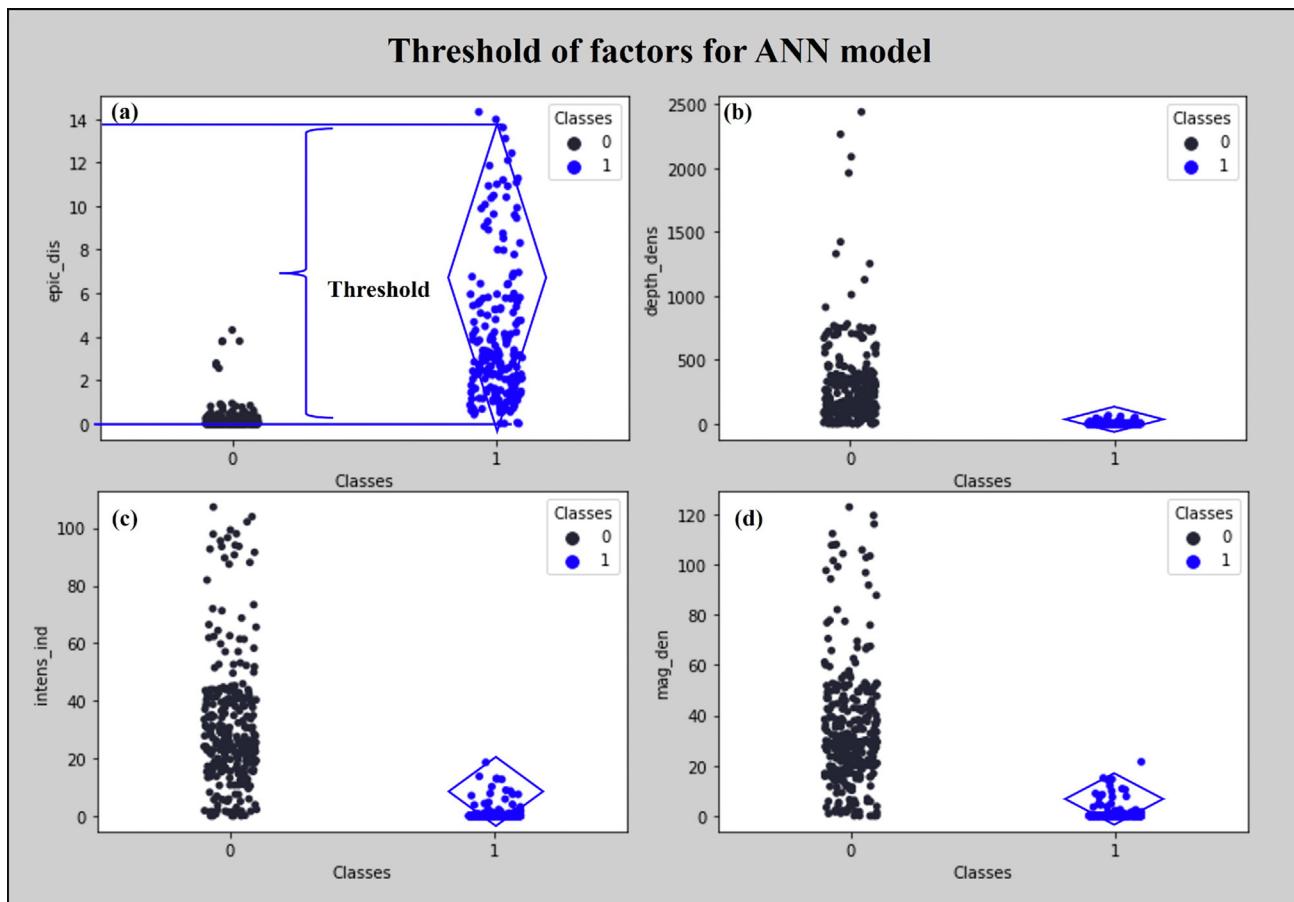


Fig. 10. Predicted classes were plotted and thresholds of top factors were estimated for ANN model.

gets better results than ANN and generates good prediction classification maps (see Fig. 8). The number of training data was very low for the training and testing, which might be another reason for the low accuracy in ANN compared with that in RF. The figures demonstrate the effect of the high contributing features on the performance of the ANN and RF models. These features can be the game changer for the models to identify a smaller number of FP pixels for the classes and classify the images more accurately.

Moreover, to understand the effectiveness of SHAP for probability mapping, the results were compared with other studies. According to the authors' knowledge and the literature review presented in Section 1, SHAP was applied for the first time for earthquake probability mapping in this research. Therefore, the results achieved here were compared with several probability mapping results. In a recent work, Jena et al. (2020) implemented an integrated approach of ANN-TOPSIS to estimate earthquake probability in Aceh province in Indonesia with an accuracy of 88 %. Jena et al. (2020b) applied ANN in the classification procedure for probability mapping in Aceh with an accuracy of 84 %. Han et al. (2020) implemented RF method for vulnerability mapping using in Gyeongju, South Korea with an overall accuracy of 94 %. By comparing the results with other methods, the RF technique demonstrates reasonably good efficiency in earthquake probability mapping, which can be understood through the SHAP model.

SHAP interaction plot shows a matrix of summary plots where the main effects can be seen on the diagonal side whereas the interaction effects are off diagonal (Fig. 9). The SHAP values are similar to the main effects for a linear model however, the higher order interaction effects are captured by all the interactions based on the pairwise interaction terms.

Spatial probability of earthquakes with highly significant factors such as; epicenter distance, depth density, intensity variation, and magnitude density were plotted as presented in Fig. 10 for ANN and in Fig. 11 for RF models. It can be seen that the epicenter distance varies from 0 to 14 km. However, the possibility for most of the earthquakes can be expected at a depth of 0–50 km (Fig. 10a-b). Similarly, a high possibility of earthquakes can be observed with weak to fair intensity and the areas not much hit by high magnitude earthquakes is considered as a seismic gap (Fig. 10c-d). According to the RF model, the epicenter distance varies from 0 to 2 km. However, the possibility for most of the earthquakes can be expected at a depth of 0–700 km (Fig. 11a-b). Similarly, a high possibility of earthquakes can be observed with high intensity and the areas that can hit by high magnitude earthquakes vary from 0 to 60 number of events which can be the source for future possible events (Fig. 11c-d).

These plots show a comparative assessment of a possible threshold of factors that is responsible for obtaining an accuracy of 96 % and 98 % for ANN and RF models, respectively (Figs. 10 & 11). However, the interactions among four major factors are different for both the models that generated little difference in accuracy and the final output.

6. Conclusions

This research aims to (1) apprehend model decisions, (2) comprehend complex intrinsic non-linear relations, and (3) ascertain the models' suitability for earthquake probability mapping. The critical point is to grasp the reason behind the probability accuracy performed by data-driven methods based on specific input data.

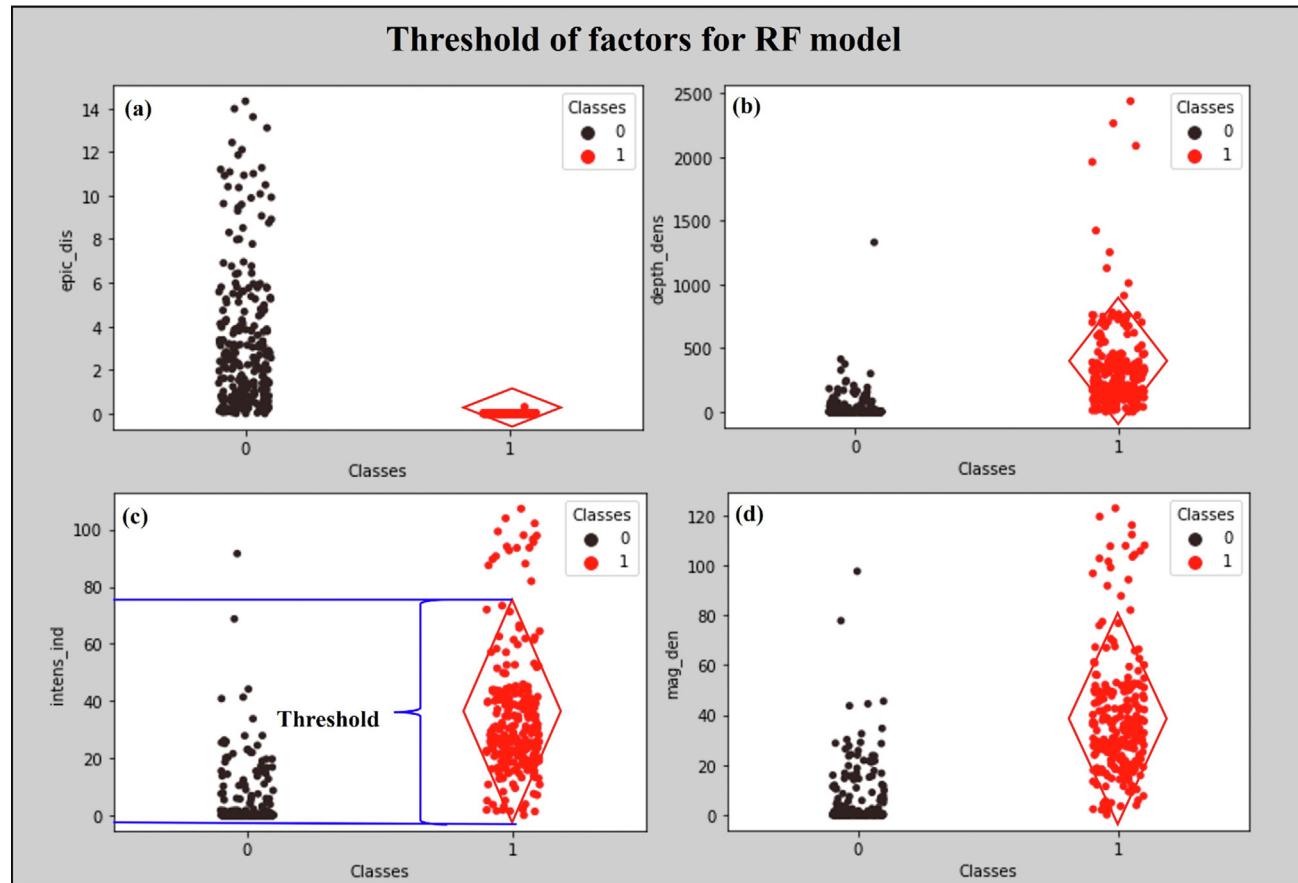


Fig. 11. Predicted classes were plotted and thresholds of top factors were estimated for RF model.

The study implemented an XAI model named SHAP to interpret the classification models' output and analyze the importance of features. The study found that SHAP is a suitable technique to identify appropriate features for prediction classification by allocating SHAP values for each data point that influences the model's output. Eight significant features were considered as model inputs to estimate earthquake probability locations. Based on the SHAP values, epicenter distance, depth density, magnitude density, and intensity variation contributed significantly to the output of the proposed models. Other features had low contribution. The two models were compared, and they gave different results. This study achieved both visualization and quantitative outcomes, which confirmed that the proposed ANN and RF models can enhance the overall accuracy and generate good probability results using the high contributing features for earthquakes. According to ANN's result, epicenter distance, depth density, intensity variation, and magnitude density are important with an accuracy of 96 %. However, the RF model shows that epicenter distance, depth density, magnitude density, and intensity variation are important with an accuracy of 98 %. From both the models, it can be concluded that epicenter distance and depth density are globally stable factors, while magnitude density and intensity variation are unstable and found to be fluctuating. The threshold range for the globally stable factors such as the distance from the epicenter are ANN (0–14 km), RF (0–2 km) and depth variation ANN (0–50 km), RF (0–700 km), respectively. This study also confirms that the slope and elevation don't affect much in earthquake probability assessment for the ANN and RF model. SHAP yields the features' importance, which are then used to improve predictions. The method is applicable and suitable for interpreting the machine learning models. SHAP allows to deeply analyze the data, thus leading to accurate future earthquake probability mapping. Future research can deal with the application of other AI based explainability models such as LIME (Local Interpretable Model-Agnostic Explanations), Randomized Input Sampling for Explanation (RISE) and Class Activation Mapping (CAM) and rationale techniques in earthquake spatial probability assessment.

Funding

The Centre for Advanced Modelling and Geospatial Information Systems, Faculty of Engineering and IT, University of Technology Sydney, funded this research. This work is also in part supported by the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia, under Project RSP-2021/14.

CRediT authorship contribution statement

Ratiranjan Jena: Investigation, Data curation, Formal analysis, Writing – original draft. **Biswajeet Pradhan:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Visualization, Writing – review & editing. **Shilpa Gite:** Writing – review & editing. **Abdullah Alamri:** Writing – review & editing. **Hyuck-Jin Park:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Adeli, H., Hung, S.L., 1993. Fuzzy neural network learning model for image recognition. *Integr. Comput-Aid E.* 1, 43–55.
- Adeli, H., Hung, S.L., 1994. Machine learning: neural networks, genetic algorithms, and fuzzy systems. John Wiley & Sons, Inc.. ISBN (Print) 9780471016335, 1 (211).
- Adeli, H., Karim, A., 2000. Fuzzy-wavelet RBFNN model for freeway incident detection. *J. Transp. Eng.* 6, 464–471.
- Berhich, A., Belouadha, F.Z., Kabbaj, M.I., 2020. March. LSTM-based models for earthquake prediction. In Proceedings of the 3rd International Conference on Networking, Information Systems & Security, 1–7.
- Boominathan, A., Dodagoudar, G.R., Suganthi, A., Maheswari, R.U., 2008. Seismic hazard assessment of Chennai city considering local site effects. *J. Earth Syst. Sci.* 2, 853–863.
- Bourbakis, N., Kakumanu, P., Makrogiannis, S., Bryll, R., Panchanathan, S., 2007. Neural network approach for image chromatic adaptation for skin color detection. *Int. J. Neural Syst.* 1, 1–12.
- Debnath, P., Chittora, P., Chakrabarti, T., Chakrabarti, P., Leonowicz, Z., Jasinski, M., Gono, R., Jasińska, E., 2021. Analysis of Earthquake Forecasting in India Using Supervised Machine Learning Classifiers. *Sustainability* 13 (2), 971.
- García, M.V., Aznarte, J.L., 2020. Shapley additive explanations for NO₂ forecasting. *Ecol. Inform.* 56, 101039.
- Dasgupta, S., Narula, P.L., Acharyya, S.K., Banerjee, J., 2000. Seismotectonic Atlas of India and its Environs. Geological Survey of India.
- Gopych, P., 2008. Biologically plausible BSDT recognition of complex images: the case of human faces. *Int. J. Neural Syst.* 18 (06), 527–545.
- Guo, H., Zhang, W., Ni, C., Cai, Z., Chen, S., Huang, X., 2020. Heat map visualization for electrocardiogram data analysis. *BMC Cardiovasc. Disor.* 20 (1), 1–8.
- Han, J., Kim, J., Park, S., Son, S., Ryu, M., 2020. Seismic vulnerability assessment and mapping of Gyeongju, South Korea using frequency ratio, decision tree, and random forest. *Sustainability* 12 (18), 7787.
- Haykin, S.S., 2009. Neural networks and learning machines. Pearson Education India, 3/E.
- Hung, S.L., Adeli, H., 1993. Parallel backpropagation learning algorithms on cray Y-MP8/864 supercomputer. *Neurocomputing* 5 (6), 287–302.
- Iyengar, R.N., Ghosh, S., 2004. Microzonation of earthquake hazard in greater Delhi area. *Curr. Sci. India* 87 (9), 1193–1202.
- Jena, R., Pradhan, B., Beydoun, G., Alamri, A.M., Sofyan, H., 2020. Earthquake hazard and risk assessment using machine learning approaches at Palu, Indonesia. *Sci. Total Environ.* 749, 141582.
- Jena, R., Pradhan, B., Naik, S.P., Alamri, A.M., 2021. Earthquake risk assessment in NE India using deep learning and geospatial analysis. *Geosci. Front.* 12, (3) 101110.
- Kanth, S.R., Iyengar, R.N., 2006. Seismic hazard estimation for Mumbai city. *Curr. Sci. India*, 1486–1494.
- Karunanithi, N., Grenney, W.J., Whitley, D., Bovee, K., 1994. Neural networks for river flow prediction. *J. Comput. Civil Eng.* 8 (2), 201–220.
- Khan, I., Choi, S., Kwon, Y.W., 2020. Earthquake detection in a static and dynamic environment using supervised machine learning and a novel feature extraction method. *Sensors* 20 (3), 800.
- Khashman, A., Sekeroglu, B., 2008. Document image binarisation using a supervised neural network. *Int. J. Neural Syst.* 18 (05), 405–418.
- Lian, H.C., Lu, B.L., 2007. Multi-view gender classification using multi-resolution local binary patterns and support vector machines. *Int. J. Neural Syst.* 17 (06), 479–487.
- Liu, X., Chen, X., Wu, W., Peng, G., 2007. A neural network for predicting moisture content of grain drying process using genetic algorithm. *Food Control* 18 (8), 928–933.
- Lubo-Robles, D., Devegowda, D., Jayaram, V., Bedle, H., Marfurt, K.J., Pranter, M.J., 2020. Machine learning model interpretability using SHAP values: application to a seismic facies classification task. In SEG International Exposition and Annual Meeting. Houston, USA. DOI:10.1190/segam2020-3428275.1.
- Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.W., Newman, S.F., Kim, J., Lee, S.I., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* 2 (10), 749–760.
- Lundberg, S.M., Lee, S.I., 2017, December. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on NeurIPS 4768–4777. Long Beach, CA, USA.
- Mayorga, R.V., Arriaga, M., 2007. Non-linear global optimization via parameterization and inverse function approximation: An artificial neural networks approach. *Int. J. Neural Syst.* 17 (05), 353–368.
- McIlraith, A.L., Card, H.C., 1997. Birdsong recognition using backpropagation and multivariate statistics. *IEEE T. Signal Proces.* 45 (11), 2740–2748.
- Nath, S.K., Thingbaijam, K.K.S., Raj, A., 2008. Earthquake hazard in Northeast India—A seismic microzonation approach with typical case studies from Sikkim Himalaya and Guwahati city. *J. Earth Syst. Sci.* 117 (2), 809–831.
- Nazzal, J.M., El-Emary, I.M., Najim, S.A., 2008. Multilayer perceptron neural network (MLPs) for analyzing the properties of Jordan Oil Shale 1. *World Appl. Sci. J.* 5 (5), 546–552.
- Panakkat, A., Adeli, H., 2007. Neural network models for earthquake magnitude prediction using multiple seismicity indicators. *Int. J. Neural Syst.* 17 (01), 13–33.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *Int. J. Rem. Sens.* 26 (1), 217–222.
- Park, S., Choi, C., Kim, B., Kim, J., 2013. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ. Earth Sci.* 68 (5), 1443–1464.

- Priambodo, B., Mahmudy, W.F., Rahman, M.A., 2020. Earthquake Magnitude and Grid-Based Location Prediction using Backpropagation Neural Network. *Knowl. Eng. Data Sci.* 3 (1), 28–39.
- Provost, F., Hibert, C., Malet, J.P., 2017. Automatic classification of endogenous landslide seismicity using the Random Forest supervised classifier. *Geophys. Res. Lett.* 44 (1), 113–120.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. August. "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA. 1135–1144.
- Schäfer, A.M., Zimmermann, H.G., 2007. Recurrent neural networks are universal approximators. *Int. J. Neural Syst.* 17 (04), 253–263.
- Shcherbakov, R., Zhuang, J., Zöller, G., Ogata, Y., 2019. Forecasting the magnitude of the largest expected earthquake. *Nat. Commun.* 10 (1), 1–11.
- Sitharam, T.G., Vipin, K.S., 2011. Evaluation of spatial variation of peak horizontal acceleration and spectral acceleration for south India: a probabilistic approach. *Nat. Hazards* 59 (2), 639–653.
- Sitharam, T.G., Kolathayar, S., James, N., 2015. Probabilistic assessment of surface level seismic hazard in India using topographic gradient as a proxy for site condition. *Geosci. Front.* 6 (6), 847–859.
- Tyagi, D., Haq, M.A., Rahaman, G., Baral, P., Datta, J., 2020. Comparison of Performance of Artificial Neural Network (ANN) and Random Forest (RF) in the Classification of Land. *Proceedings of UASG 2019: Unmanned Aerial System in Geomatics*. Roorkee, India, 225–236.
- Zhao, Y., Takano, K., 1999. An artificial neural network approach for broadband seismic phase picking. *B. Seismol. Soc. Am.* 89 (3), 670–680.
- Wang, K., Chen, Q.F., Sun, S., Wang, A., 2006. Predicting the 1975 Haicheng earthquake. *B. Seismol. Soc. Am.* 96 (3), 757–795.
- Wang, Q., Guo, Y., Yu, L., Li, P., 2017. Earthquake prediction based on spatio-temporal data mining: an LSTM network approach. *IEEE Trans. Emerg. Topics Comput.* 8 (1), 148–158.
- Wang, T., Zhang, Z., Li, Y., 2019. EarthquakeGen: Earthquake generator using generative adversarial networks. In *SEG Technical Program Expanded Abstracts 2019*. Society of Exploration Geophysicists, 2674–2678.
- Wersing, H., Kirstein, S., Göttling, M., Brandl, H., Dunn, M., Mikhailova, I., Goerick, C., Steil, J., Ritter, H., Körner, E., 2007. Online learning of objects in a biologically motivated visual architecture. *Int. J. Neural Syst.* 17 (04), 219–230.
- Xiong, P., Tong, L., Zhang, K., Shen, X., Battiston, R., Ouzounov, D., Iuppa, R., Crookes, D., Long, C., Zhou, H., 2021. Towards advancing the earthquake forecasting by machine learning of satellite data. *Sci. Total Environ.* 771, 145256.
- Zhao, S., Guo, Y., Sheng, Q., Shyr, Y., 2014. Advanced heat map and clustering analysis using heatmap3. *Biomed. Res. Int.* 986048. <https://doi.org/10.1155/2014/986048>.