An Ensemble Random Forest Model for Seismic Energy Forecast

Sukh Sagar Shukla [1], Jaya Dhanya [2], Praveen Kumar [3], Priyanka Sihag [3], and Varun Dutt [4]

[1]Research Scholar, School of Civil and Environmental Engineering, Indian Institute of Technology Mandi, Himachal Pradesh, India

[2]Assistant professor, School of Civil and Environmental Engineering, Indian Institute of Technology Mandi, Himachal Pradesh, India

[3]Research Scholar, School of Computing and Electrical Engineering, Indian Institute of Technology Mandi, Himachal Pradesh, India

[4]Professor, School of Computing and Electrical Engineering, Indian Institute of Technology Mandi, Himachal Pradesh, India

Correspondence: Sukh Sagar Shukla (D22178@students.iitmandi.ac.in)

Abstract. Seismic energy forecasting is critical for hazard preparedness, but current models have limits in accurately predicting seismic energy changes. This paper fills that gap by introducing a ~~new ensemble random forest model designed specifically~~ novel ensemble-based Random Forest framework for seismic energy forecasting. Building on ~~an existing paradigm, provided by Raghukanth et al. (2017)~~a previously established methodology, the global energy time series is decomposed into intrinsic mode functions (IMFs) using ensemble empirical mode decomposition for better representation. Following this approach, we split the data into stationary (IMF_1) and non-stationary (sum of IMF_2-IMF_6) components for modeling. We acknowledge the inadequacy of intrinsic mode functions (IMFs) in capturing seismic energy dynamics, notably in anticipating the final values of the time series. To ~~address this restriction~~overcome this limitation, the yearly seismic energy time series is also fed along with the stationary and non-stationary parts as inputs to the developed models. ~~Here, we employed~~ In this study, we employ Support Vector Machine (SVM), Random Forest (RF), ~~Instance-Bases~~ Instance-Based learning (IBk), ~~Linear Regression (LR~~Ridge Regression (RR), and Multi-Layer Perceptron (MLP) algorithms for the modelling. Furthermore, the five models discussed above were suitably employed in a ~~novel regression-based ensemble random forest algorithm~~ stacked regression ensemble using Random Forest as the meta-learner to arrive at the final predictions. The root mean square error (RMSE) obtained in the training and testing phases of the ~~final~~ validation model were 0.127 and 0.134, respectively. It was observed that the performance of the developed ensemble model was superior to those existing in literature~~(Raghukanth et al., 2017)~~. Further, the developed algorithm was employed for the seismic energy prediction in the active Western Himalayan region for a comprehensively compiled catalogue and the mean forecasted seismic energy for year 2024 is $7.21 \times 10^{14} J$. This work is a pilot project that aims to create a ~~forecast model for the release of seismic energy globally and further application at a regional level~~robust, scalable framework for forecasting seismic energy release globally and regionally. The findings of our investigation demonstrate the ~~possibility of the established method in the accurate seismological~~ promise of the ensemble approach in delivering reliable seismic energy forecast, which can help with appropriate hazard preparedness.

1 Introduction

Earthquakes are among the most disastrous natural calamities due to the release of accumulated strain energy from continuous tectonic movements. Like other natural disasters, ~~it~~ they can cause destruction both in financial terms and loss of life (Jain, 2016). The devastating potential of earthquakes is increased by their fundamentally unpredictable character due to both aleatory and epistemic uncertainties (Kramer, 1996; Baker et al., 2021). ~~Whereby, the inherent randomness in the process makes it a real challengeto accurately predict these events.~~ Because the problem at hand is unpredictable, creating an accurate forecasting model is a unique challenge. There are several attempts by seismologists to quantify the activity of the regions based on several seismicity indicators. Some of the studies for the Himalayan region are by performing paleo-seismic study (Lavé et al., 2005; Rajendran et al., 2013), statistical inferences (Bilham and Ambraseys, 2005), Global Positioning System (GPS) measurements (Banerjee and Bürgmann, 2002; Ader et al., 2012), numerical (Ismail-Zadeh et al., 2007; Jayalakshmi and Raghukanth, 2017), satellite ~~imagery based~~ imagery-based data (Bhattacharya et al., 2013; Misra et al., 2020), and ~~,~~ Global Navigation Satellite System (GNSS) studies (Sharma et al., 2023b; Kumar et al., 2023a). However, the inadequacy in precisely monitoring stress changes, pressure, material variability, and temperature variation deep beneath the ~~earth~~Earth's crust using scientific instruments leads to a lack of comprehensive data regarding accurate seismic characteristics. Subsequently, ~~these~~ this lack of information ~~had~~ has contributed to the uncertainty in earthquake occurrence, which ~~had resulted to major risk~~ has resulted in major risks to life and property. Hence, a robust quantification approach is essential considering the increasing vulnerability of the active regions due to developmental activities (Bilham, 2019). However, the variability in seismic ~~behavior~~behaviour, the worldwide occurrence of earthquakes, and the paucity of historical data all hamper predictive ~~modeling~~modelling. The ethical and practical consequences of delivering earthquake forecasts, the diversity in earthquake magnitudes, and the differences between human and geological timelines all add to the ~~tremendous problem of~~ challenges in reliable earthquake prediction (Mignan and Broccardo, 2020; Sun et al., 2022).

While progress is being made, the emphasis in earthquake research has turned toward establishing effective earthquake forecast models and early warning systems, understanding seismic risks, and improving preparation to lessen the effects of these deadly occurrences (Bose et al., 2008; Tiampo and Shcherbakov, 2012; Mousavi and Beroza, 2018; Mousavi et al., 2020; Tan et al., 2022). Nevertheless, with the advancements in field instrumentation, once an event occurs, we have attained the knowledge to estimate and record its information like magnitude, location, extent of ground shaking, etc., immediately (USGS (2023), IMD (2023)). The robustness of this data has also improved significantly over the years. An intriguing question here shall be, is it possible to predict and be better prepared for a forthcoming event using these information? This ~~work attempts to answer this question by compiling the available earthquake data and implementing it in advanced~~ study tackles this problem by compiling an extensive seismic dataset and building predictive models using state-of-the-art machine learning (ML) algorithms.

ML has evolved so much that its potential is widely explored to address numerous real-world problems (Schmidt et al., 2019; Kaushik et al., 2020; Sarker, 2021; Bertolini et al., 2021; Kumar et al., 2023b). Appropriate data processing using advanced ML algorithms led to successful prediction models. However, ML algorithms have only recently gained popularity in engineering seismology (Xie et al., 2020; Mousavi and Beroza, 2023). The most comprehensive application is in developing efficient Ground Motion Prediction Equations (GMPEs) (Alavi and Gandomi, 2011; Derras et al., 2014; Dhanya and Raghukanth, 2018; Gade et al., 2021; Seo et al., 2022; Sreenath et al., 2024). ~~In another direction, Paolucci et al. (2018) proposed a simple MLP model that should efficiently generate broadband ground motions. Sharma et al. (2023a) improved the model by incorporating source, path and site characteristics. Even though the results from machine learning approaches show promising applications in earthquake engineering, the full potential of ML is yet to be explored in earthquake forecasting. Moreover, the advancements in instrumentation that continuously capture seismic data alongside efficient learning algorithms are expected to reduce the prediction variability considerably. Concerning earthquake prediction, Adeli and Panakkat (2009) employed a probabilistic neural network (PNN) to predict maximum magnitude using eight predefined seismic indicators and found it efficient in predicting low-magnitude events. Further, Narayanakumar and Raja (2016) divided the events into 15 classes and attempted to predict them employing past 128-year data for the Himalayan Belt using a Back Propagation (BP) algorithm. However, the model seemed inefficient for large events. Furthermore, a few attempts were made with the multilayer perceptron (MLP) model to forecast yearly regional Kavitha and Raghukanth (2016) and global seismic energy releases(Raghukanth et al., 2017). Later, Asim et al. (2018) performed a more extensive study incorporating almost sixty seismic features suitably in the Support vector regression-Hybrid neural network (SVR-HNN) to develop a classification-based prediction model for the occurrence of an earthquake greater than 5. Furthermore, Yousefzadeh et al. (2021) attempted to perform spatiotemporal earthquake magnitude prediction (a classification problem) for Iran using a deep neural network and reported promising model performance. Their analysis was based on shallow neural network (SNN), support vector machine SVM), decision tree (DT), and deep neural network (DNN) models. Additionally, Salam et al. (2021) analyzed earthquake magnitudes for the southern California region using hybrid ANN models and found flower pollination algorithm-extreme learning machine (FPA-ELM) and FPA-SVM gave better predictions. Furthermore, Ridzwan and Yusoff (2023) has also discussed the relevance and evolution of machine learning applications in earthquake predictions.Zhang et al. (2023) presented EPT, a totally data-driven deep learning model. The model use gated feature extraction blocks (GFEB) to extract possible crustal motion and plate movement patterns from worldwide historical seismic data. It utilises them to help anticipate mainshocks in each local and provincial region. Also, Bhatia et al. (2023) proposed a cloud-based edge computing collaborative Internet of Things (IoT) monitoring and prediction system for earthquake prediction. Real-time sensor data was collected using Internet of Things technology and sent to the edge layer, where a unique Bayesian belief model approach was applied to feature categorization. Moreover, the cloud layer earthquake magnitude was predicted using the Adaptive Neuro-Fuzzy Inference System (ANFIS) mechanism. It is also observed that the seismic energy release pattern is one of the prominent indicators in the forecast or prediction models. Hence, considering the uncertainties in the seismic moment release, energy-based models are expected to provide reliable forecasts. Additionally, one can also note that most of the forecast models are based on a specific architecture or hybrid models combining optimization and ML formulations. However, there have been more advancements in the field of ML, and one of the promising addition in the forecast category are the ensemble models (Gastinger et al., 2021). Hence, this work attempts to develop robust forecast model for seismic energy release. The nonstationarity in the time series data is addressed by performing ensemble empirical model decomposition as suggested in Raghukanth et al. (2017). The proposed approach shall be first verified with a global database.~~

Upon suitable validation, the approach shall be extended to developing a forecast model for the active Western Himalayan province. The proposed work is the first of its kind attempt to predict earthquake occurrence in the Himalayan region. The study results are critical in identifying the potential seismic hazard and formulating swift policies for better preparedness for the impending earthquake. A detailed background and formulation of the proposed approach are discussed in subsequent sections.

2 Background

The expansion of the global seismograph network in recent decades has sparked a significant increase in the examination of seismic activity. Recognising that defining earthquake size in terms of seismic energy released has greater physical relevance than magnitude alone in understanding the cumulative seismic activity. Hence, the scientific community has worked diligently to predict seismic energy releases on a regular basis. Tsapanos and Liritzis (1992) investigated the relationship between seismic energy release for three seismic regions: Chile, Kamchatka, and Mexico. Tsapanos (1998) used released strain energy to assess seismic hazards in eleven regions around the world. Moreover, when seismic activities occur in the region where there is potential for other hazards, such as landslides and volcanic eruptions, it leads to multi-hazard scenarios. In this regard, Yokoyama (1988) used the seismic energy release from the precursory earthquake to predict the time of the eruption of dacitic or andesitic volcanoes. Nakamichi et al. (2019) studied the pattern of the rate of seismic energy release for Kelud volcano, Indonesia, before the 2007 effusive dome-forming and 2014 Plinian eruptions. Jaumé and Sykes (1999) have reviewed how seismic energy release accelerates prior to a great earthquake. They observed that a growing number of cases have been reported in which the occurrence of a large or great earthquake is preceded by an increase in the frequency of moderate-sized earthquakes in the surrounding region. A power-law time-to-failure relationship can be used to model the rate of moment and/or energy release in these sequences. Varga et al. (2012) examined a declustered catalogue of major earthquakes ($M >= 7.0$) that occurred between 1960 and 2011 and discovered that the latitude distribution of released seismic energy is bimodal with respect to the equator. Later, Kavitha and Raghukanth (2016) and Raghukanth et al. (2017) have attempted to forecast the seismic energy for the regional and globe scales, respectively, using the artificial neural network technique on mode decomposed energy series. Researchers have also exploredthe application of seismic energy to determine the probability of the occurrence of earthquake events. Whereby, Zarola and Sil (2018) have predicted the magnitude and time of occurrence of earthquakes for northeast India using four distribution models (Lognormal, Weibull, Gamma, and Log-logistic). Later, Asim et al. (2018) predicted the target earthquake by employing various seismic features in different machine-learning techniques like support vector regression and hybrid neural networks. Furthermore, Shimony et al. (2020) have studied seismic energy release along the Dead Sea transform (DST) from Intra-Basin Source and its influence on regional ground motions.Recently, Spassiani and Marzocchi (20 has also described that the energy-based approach is more appropriate in defining the magnitude frequency relation of seismic events due to its physical relevance embedded in rebound theories. In essence, seismic energy is one of the most important markers of earthquake occurrence, and it follows a pattern in which locations that have been seismically quiet for a long period are more likely to have significant earthquakes. So we can readily say that the previous release of seismic energy can be efficiently used to anticipate the future energy release. As a result, a reliable seismic energy forecasting model can help you prepare for an upcoming occurrence. Several researchers in the past have extensively explored and utilized various machine learning techniques in earthquake engineering and related fields. Xie et al. (2020) highlighted the adoption of Multilayer Perceptrons (MLPs) Moreover, several machine learning techniques have been explored, among those, Multilayer Perceptrons (MLPs) is the most widely used model in earthquake engineering . . Raghukanth et al. (2017) utilized applications (Xie et al., 2020) . Raghukanth et al. (2017) utilised a similar model for suitably combining stationary and non-stationary parts of energy series to forecast seismic energy. This MLP technique is also widely used in developing ground motion prediction equations, as evidenced by the works of Derras et al. (2014); Dhanya and Raghukanth (2018, 2020); Douglas (2021). In another direction, Paolucci et al. (2018) proposed a simple MLP model that should efficiently generate broadband ground motions. Sharma et al. (2023a) improved the model by incorporating source, path and site characteristics. Another architecture, such as Linear Regression (LR), has also been applied in various seismological studies due to its simplicity and efficiency. Pairojn and Wasinrat (2015) used LR for ground motion prediction in Thailand, while Cho et al. (2022) compared Artificial Neural Networks (ANN) and LR for predicting earthquake-induced slope displacement. The Random Forest (RF) technique has similarly motivated researchers across different fields, including seismology. Asim et al. (2017) used RF to predict earthquake magnitude in the Hindukhush region. More recently, Agarwal et al. (2023) developed a hybrid algorithm by combining ANN and RF to predict the earthquake magnitude. Pyakurel et al. (2023) utilized Apart from the plain linear regression, studies have also used ridge regression for earthquake forecast problems Ahmed et al. (2024). Pyakurel et al. (2023) utilised five supervised algorithms, including RF, to predict earthquake-induced landslides for the 2015 Gorkha earthquake. Additionally, (Li and Goda, 2023) extended the application of RF to tsunami early warning systems and loss forecasting. Furthermore, Support Vector Machines (SVM) with the optimized optimised version named as Sequential Minimal Optimization Optimisation for regression (SMOreg), as proposed by Shevade et al. (2000), are widely used for parameter learning. This approach has been applied to various natural hazard contexts, such as flood susceptibility mapping Saha et al. (2021)(Saha et al., 2021), ground motion prediction equations (Altay et al., 2023), and landslide monitoring (Kumar et al., 2023b). Similar to SMOreg, Instance based instance-based learning is also well-explored in earthquake prediction problems, as its reliability and accuracy owing from are owed to the algorithm's resistance to noise and outliers, as well as its versatility in the use of distance measures. Its applicability in seismic prediction is well demonstrated by Reyes et al. (2013), Ghaedi and Ibrahim (2017),Al Banna et al. (2020), and Ridzwan and Yusoff (2023).

Ensemble models have found versatile applications across numerous domains, including medicine, materials science, Apart from the individual machine learning techniques, ensemble learning is a mature and widely adopted methodology in the ML literature, which is renowned for averaging several models to enhance prediction accuracy and generality (Dietterich, 2000b; De Gooije . Ensemble models combine different base learners based on techniques such as bagging, boosting, and stacking, and so maximize the variance in data, reduce overfitting, and environmental science. These models , such as those employed by (Tan et al., 2022) for cancer classification using boosted and bagged decision trees, and by (Rezaei et al., 2022) for predicting gastric cancer using Gradient Boosting Decision Trees (GBDTs), Random Forest (RF), Linear Regression (LR), Elastic Net, and LASSO, often outperform individual models. In heart disease prediction Pouriyeh et al. (2017) (Pouriyeh et al., 2017) and COVID-19 case forecasting Maaliw III et al. (2021)(Maaliw III et al. , 2021), ensemble techniques have also shown superior performance. Their popularity extends to various fields, including earthquake prediction in seismology, as seen in (Shishegaran et al., 2019; Joshi et al., 2022).Despite their success , modeling the forecast of annual seismic energy release remains underexplored, presenting a potential research area. Ensemble models combine different ML approaches to capture more data variance, thereby outperforming individual models in reliability, precision, and comprehensibility (Dietterich, 2000a; De Gooijer . These models integrate several base learners and produce diverse ones through generative and non-generative methods(Re and Valentini, 20 . improve model reliability. They have been successful across a variety of applications including medical diagnosis and climate modelling to financial forecasting (Re and Valentini, 2012; Tan et al., 2022; Rezaei et al., 2022). Although extensively used, their use in seismic energy forecasting is still not well exploited, making the current research a timely and new addition in the geophysical hazard field.

In light of this, the present study attempts to forecast the annual seismic energy using the advanced ensemble model technique. In the proposed modelling approach, five non-parametric supervised machine-learning approaches, including Artificial neural network (ANN), Linear regression (LR), Random forest (RF), Sequential Minimal Optimization regression (SMOreg), and Instance-Bases learning with parameter k (IBk), are used to construct an individual model to forecast annual seismic energy separatelyThe research is based on the success of ensemble learning with the application of a stacked ensemble framework that is specific to the challenge of seismic energy forecasting. Even though ensemble models are extensively applied in other areas such as health, climate, and finance, their niche use in seismic energy prediction remains limited and largely untapped. The predictions of five individual machine learning models—MLP, RF, LR, SMOreg, and IBk—are combined and stacked in this research through a Random Forest meta-learner. In this study, we employ non-generative ensemble methods, combining predictions from multiple individual modelsfor enhanced accuracy. These foretasted values from each approach are stacked and fed into the ensembled random forest model to get a more robust and capable model. This method leverages the synergy of findings from many machine-learning algorithms, demonstrating proficiency in seismic energy predictions. Through the incorporation of pre-learned values as well as expert views from other approaches, the ensembled random forest model greatly improves its predicting performance. The validity/performance of the approach shall be tested against the global seismic energy forecast model developed by Raghukanth et al. (2017). Recognising the substantial interest and confidence in such forecasting models by local authorities, policymakers, and government agencies, the study expands its application to the active Western Himalayan area of the Indian subcontinent. This novel approach has the potential to improve our knowledge of and capacity to anticipate seismic energy releases, as well as to benefit stakeholders and support earthquake-prone areas preparedness initiatives.

setup, the final ensemble model uses the Random Forest as a meta-learner that synthesizes predictions from these individual models, each trained using domain-informed design choices and preprocessing. As such, the RF model mimics a consensus-based expert system, combining diverse perspectives across learning paradigms to enhance forecasting robustness.

The improved long-term seismic energy prediction capability—essential for forward-looking hazard mitigation—is the main contribution of this work. The proposed model showed versatility across diverse tectonic environments by working well for both global and western Himalayan data sets. The objective of this project is to apply stacked ensemble learning to develop a reliable model for annual seismic energy predictions. The Empirical Mode Decomposition (EMD) method is applied to decompose global seismic energy time series, considering stationary and non-stationary components as inputs. The model is compared to existing research, and its predictive ability is established via a case study for the Western Himalaya region.

2 Global Seismic Energy (GSE) time series

~~A comprehensive earthquake catalogis essential for making reliable earthquake predictions. For the present~~ Making accurate earthquake predictions requires a thorough earthquake catalog. We have used two global earthquake catalogs for this study. Raghukanth et al. (2017) used the ISC-GEM catalog (http://www.isc.ac.uk/) as the primary resource. In the current study, the ~~ISC-GEM catalogue () utilized by Raghukanth et al. (2017) is considered for model development, comparison~~model construction, comparison, and validation of the ~~proposed approach . We~~ suggested approach are done using this catalog. A more thorough and current worldwide earthquake catalog (up to 2023) was created using the USGS seismic database (https://earthquake.usgs.gov/earthquakes/search/) and the ISC-GEM catalog after the methodology was verified using this data. For our analysis, we used the same inputs as ~~described in Raghukanth et al. (2017)for our analysis. However, for better clarity on the data, we are explaining in brief the processing involved in arriving at~~ those mentioned in Raghukanth et al. (2017). We provide a brief explanation of the processing that goes into creating the final time series ~~used in modeling .The global earthquake catalogue contained data spanning from 1900-2015, having a total of 24375 events. The unification of the event magnitude was ensured by converting all~~ that is used for modeling in order to improve understanding of the data. The validation catalog, which is the worldwide earthquake catalog that was obtained from Raghukanth et al. (2017), includes data from 1900 to 2015, totaling 24,375 occurrences with a minimum magnitude of 4.98 Mw. The new global catalog created for this study, which will be referred to as the Global catalog, is compiled from both USGS and ISC-GEM sources and contains data from 1900 to 2023. Duplicates were thoroughly examined and eliminated because the data was obtained from two distinct sources. There were 988,812 distinct occurrences with a minimum magnitude of 1.09 after duplicate events were removed. All reported event magnitudes ~~to moment magnitudes~~ were converted to moment magnitude (M_w) using ~~suitable empirical relations. Further,~~ empirical relationships from Scordilis (2006) and Yenier et al. (2008) to guarantee magnitude uniformity across datasets. We next determined the magnitude of completeness (~~The~~ M_c) for both catalogs, which is the lowest magnitude above which ~~all earthquakes are reliably recorded) of the catalogue was estimated as~~ earthquakes are consistently documented. According to Raghukanth et al. (2017), using the maximum curvature approach (Wiemer and Wyss, 2000), M_c was determined to be 6.4 ~~(Figure 1~~ for the validation catalog (Figure S1(a) ~~) following the maximum curvature method proposed by Wiemer and Wyss (2000). The distribution of the final complete catalogconstituting 4619 events used in the analysis performed in this study is shown in Figures~~ in the supplemental document). Using the MATLAB-based program ZMAP version 7.1, M_c was calculated to be 4.9 M_w for the Global catalog, as seen in Figure 1(a).

Furthermore, the Global catalog's year of completeness was established using the Stepp (1973) approach, showing that the catalog is complete starting in 1953 (Figure 1(b)~~and~~). Figures S1(b) and S1(c) display the distribution of events from the full validation catalog (4,619 events), while Figure 1(c) ~~. The catalogue had four great events having an $M_w \geq 9$ and the~~

largest event was the 1960 M_w 9.6 Valdiya earthquakeshows the full Global catalog with 217,751 events. Because partial representation increases statistical bias, events with magnitudes < M_c were not included in the analysis. Furthermore, to estimate the yearly occurrence time series of seismic events it becomes illogical to sum up the corresponding M_w. A suitable alternative in agreement with the physics of earthquake occurrence is to convert the given M_w to energy contribution is dominated by large occurrences since the connection between seismic energy and earthquake magnitude is logarithmic. According to the corresponding seismic energyand accumulate it to obtain a yearly manuscript's Equation 1, for instance, the seismic energy associated with a 3 M_w event is $5.0596 \times 10^8 J$, but the seismic energy associated with a 5 M_w event is $5.0596 \times 10^{11} J$. This indicates that it takes roughly 1,000 smaller 3 M_w events to equal the energy output of a single 5 M_w event. Therefore, from a hazard standpoint, smaller magnitude events are not of major interest and do not considerably contribute to the total seismic energy. The 1960 Mw 9.6 Valdivia event is the greatest of four major earthquakes ($M_w \geq 9$) listed in the database. Instead of using monthly or weekly aggregation, which can result in deceptive zeros because of few occurrences in smaller windows, annual accumulation was employed to generate the seismic energy time series. Hence, first the Following Hanks and Kanamori (1979), moment magnitudes ($M_w$was converted) were first transformed to seismic moment (M_0)following Hanks and Kanamori (1979), then the $M_\circ$). Choy and Boatwright (1995) then estimated seismic energy (SE) was calculated SE) using the relationproposed by Choy and Boatwright (1995); as expressed further :

$$SE = 1.6 \times 10^{-5} M_0 \quad \text{where, } M_0 = 10^{1.5 \times (M_w + 6)} \tag{1}$$

Using the corresponding expression, the yearly For both catalogs, annual seismic energy time series were estimated and the corresponding data is illustrated in Figure 2a. The great events like computed using this formula. Figure 2(a) for the Global catalog and Figure S2(a) for the validation catalog display the generated graphs. In the energy time history, significant occurrences like the earthquakes in Japan in 2011, Alaska in 1964, Sumatra in 2004, and the Valdivia in 1960 M_w9.6 Valdiya, 1964 M_w9.2 Alaska , 2004 M_w9.0 Sumatra , and 2004 M_w9.1 Japan earthquakes can be clearly identified from the distinct peaks in the time history. Furthermore, since it might be difficult to capture the energy appropriately wherever there is a sudden jump, the corresponding seismic energy had been converted to log scale as shown emerge as peaks. The seismic energy was also presented in logarithmic form, as seen in Figure 2b. Additionally, the distribution of the corresponding data is reported to be (b) for the Global catalog and Figure S2(b) for the validation catalog, because sudden shifts in the time series can skew scale interpretation. The resulting time series is non-Gaussian and non-stationary(Raghukanth et al., 2017). To perform the predictions better, the corresponding time history is split , according to Raghukanth et al. (2017). As shown in the sections that follow, the signal was broken down into stationary and non-stationary parts following an components using ensemble empirical mode decomposition , as discussed further. It should be noted that the seismic energies for each incident accumulated annually rather than on any other predetermined time frame like monthly or weekly because when attempted to collect seismic energy on a monthly basis, for $M \geq 6.4$, certain months had no events, resulting in zero seismic energy. Using a logarithmic representation of seismic energy time series may result in unrealistic results. So accumulation of seismic energy on annual basis is preferred. (a) Magnitude of completeness. (b) Magnitude distribution over years. (c) Global distribution of the events from the global

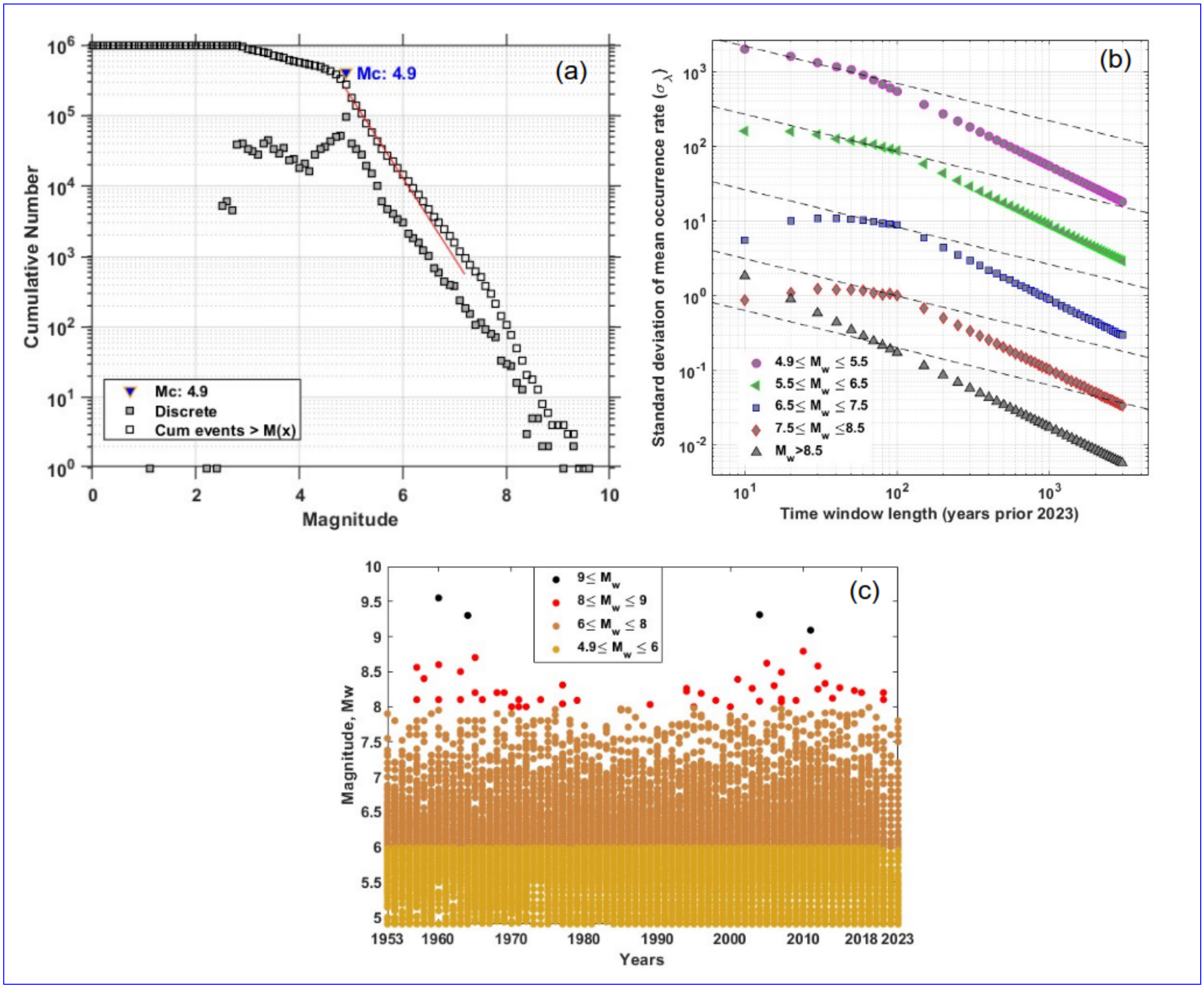

Figure 1. (a) Magnitude of completeness. (b) Year of completeness using Stepp (1973) approach. (c) Distribution of the events from the complete global earthquake catalog.

earthquake catalogue considered in Raghukanth et al. (2017) adopted for testing the performance of the proposed approach in this work in order to better capture trends and cycles.

2.1 Mode decomposition of GSE

The final seismic energy time series were split into orthogonal modes following the empirical mode decomposition (EMD) technique proposed by Huang et al. (1998). The basic functions termed as intrinsic modes were obtained following an iterative procedure on the data directly without any predefined functional form. Hence, the corresponding methodology is reported to be better adaptive of more adaptive to the features in the data. Furthermore, to avoid the issue of mode mixing in conventional

EMD Wu and Huang (2009) proposed ensemble EMD (EEMD) where a finite white noise is added to the data while performing decomposition. The basic steps in the mode extraction involve: (1) Adding finite white noise to the data, (2) Using cubic spline to construct lower and upper ~~envelops~~ envelopes connecting consecutive peaks as the respective sides, (3) At each time step estimating the average of positive and negative ~~envelops~~ envelopes and then subtract that from the data from step 1, (4) With the data from step 3 steps 2-3 is repeated till we obtain IMF (the time history having the number of extremas and zero-crossing differs by one and the mean is zero), (5) Once first IMF (IMF_1) is extracted then the corresponding value is subtracted from the time history in step 1, further follow steps 2-4 to extract next IMF (6) ~~Repeated~~ This is repeated until there are no zero-crossings left in the data. To perform ensemble empirical mode decomposition, steps 1-6 are repeated multiple times by adding different white noise~~to further~~, and the mean of IMFs at each level is identified as the final ~~modes~~mode. The observations that served as the foundation for the above strategy are (1) If we take an average of white noise in a time domain, it cancels out in the ensemble mean. Hence, in the final noise-added ensembled signal, when averaged, only the signal survives, not the noise. (2) To drive the ensemble to exhaust all viable solutions, finite, not infinitesimal, amplitude white noise is required. Finite magnitude noise causes the distinct scale signals to reside in the corresponding IMF, as mandated by the dyadic filter banks and therefore improves the meaning of the final ensemble mean. For the data of log seismic energy time series for validation and global catalog shown in Figure S2(b) and Figure 2b, we were able to extract six IMFs each following the described procedure. ~~It is noted that~~ These IMFs are mostly uncorrelated and orthogonal. Also, for the physical interpretation of IMFs, various methods exist in the literature to estimate the periodicity of time series, including the instantaneous frequency method and Fourier-based approaches. The resulting IMFs in the present study are comparable to sine/cosine waves. Hence, counting the number of extremes in an IMF allows for easy estimation of the time period. Table 1 lists the periods of all six IMFs in log-scaled seismic energy time series from both the catalogs. Table 1 also includes an estimate of the percentage variance for all IMFs, which is a statistical parameter, and calculated as the ratio of ~~IMF variance to data variance~~the variance of each IMFs to the variance of the data (Eqn. 2). The %(variance) denotes the contribution of each IMF to annual earthquake energy release. It can be noted that ~~the~~ IMF_1 ~~constitute~~ constitutes the maximum variance of the time series, and the IMF_6 represents the non-stationary trend in the data. ~~(a) Estimated Global seismic energy (J) time series from ISC gem catalogue used in developing the models (b) log scaled Global seismic energy time series (ln(GSE)) (c) First intrinsic mode estimated from ln(GSE) by performing ensemble empirical mode decomposition (EEMD) (d) Sum of second to last intrinsic modes estimated from ln(GSE) by performing EEMD.~~

$$P_{var_n} = \frac{\text{Var}(X_{IMF_n})}{\text{Var}(X_{data})} \times 100 \qquad (2)$$

where P_{var_n} is the % variance of nth IMF.

~~For the forecast of seismic energy , autoregressive modes can be adopted. However, as just using the actual data might cause greater complexity, the reasonably~~ Moreover, the IMFs are simple and well-behaved ~~IMF can also be included as variables in the prediction model~~when compared to original seismic energy data hence can capture the physics of occurrence of annual seismic energy when used as the input instead of complex seismic energy time series. Considering more about the physical

Table 1. Period observed and the variance captured by the IMFs obtained for log scaled seismic energy time series

~~IMFs~~**IMFs**	~~log(Global seismic energy time series)~~**Validation***		~~Log seismic energy time series for Western Himalaya~~**Global****		**Western H**
	Period (Years)	% (Variance)	Period (Years)	% (Variance)	Period (Years)
IMF_1	2.95	49.18	3.05	59.82	2.76
IMF_2	6.29	7.02	6	15.39	8.29
IMF_3	11.55	5.61	15.5	6.01	10.6
IMF_4	31.00	6.43	34	10.26	26
IMF_5	91	6.60	56	1.40	—
IMF_6	—	25.80	—	~~—~~7.69	

* Validation catalog is the catalog sourced from Raghukanth et al. (2017);

** Updated Global catalog prepared in the present study

interpretation of the IMFs study performed by Liritzis and Tsapanos (1993) have calculated the periodicity of global shallow seismic events from conventional approaches like Fourier method and they have got dominant period as 3(±0.5), 4.5, 6.5, 8-9, 14-20 and 31-34 years. The IMF_1 being the predominant period with contribution of around 50-60 % to annual seismic energy release has mean period of 3 years which is also reported by Liritzis and Tsapanos (1993) as one of the period. The IMF_2 having period in range of 6 to 6.29 is also conforming with the period 6.5 reported by Liritzis and Tsapanos (1993). The IMF_3 with period ranging from 11 to 15.5 is in the range of 11 year sunspot cycle as reported by Raghukanth et al. (2017) they have also found out that annual seismic energy release follow the sunspot period with 2 year delay and the standardized correlation coefficient between IMF_3 and sunspot cycle is 0.3024 which is significant. The IMF_4 and IMF_5 has 6-10 % and 1-6 % contribution to annual seismic energy respectively. Also, Wu and Huang (2004) have proposed a methodology to assess the importance of IMFs by comparing them with the Intrinsic mode functions of white noise. The suggested test we have performed on the IMFs obtained by the log seismic energy from updated global catalog and results are present in Figure 3. For pure noise, the energy and associated periods of IMFs will fluctuate linearly on the log-log plot, with all IMFs falling inside the confidence zone. It can be clearly inferred from Figure 3 that all the five IMFs (excluding IMF_6 which shows the trend) lie within the confidence interval conforming that IMFs are signal. Hence adopting the IMFs to forecast seismic energy instead of Complex seismic energy time series itself will better capture the underlying physics. For the forecast of seismic energy, autoregressive modes can be adopted. Thus, linear and nonlinear parts are separated and while modelling IMF_1 is taken ~~separate~~ separately and remaining IMFs separately as proposed by Iyengar and Raghukanth (2005) and Raghukanth et al. (2017). The corresponding IMF_1 ~~is~~ to IMF_6 are shown in Figure 2c ~~and the sum of IMF_2 to IMF_6 in Figure 2d~~for the global catalogue (Corresponding Figures for the validation catalogue are present in Figure S2(c) and S2(d) of supplementary material). The correlation of the IMFs obtained by utilising the log global time series of the global catalogue is shown in Figure 4 from which it is evident that all the IMFs are almost orthogonal. In the present study, this information was suitably incorporated into more advanced machine learning algorithms to make one step ahead of seismic energy forecast. A detailed description of the ML algorithms and the corresponding implementation is discussed further

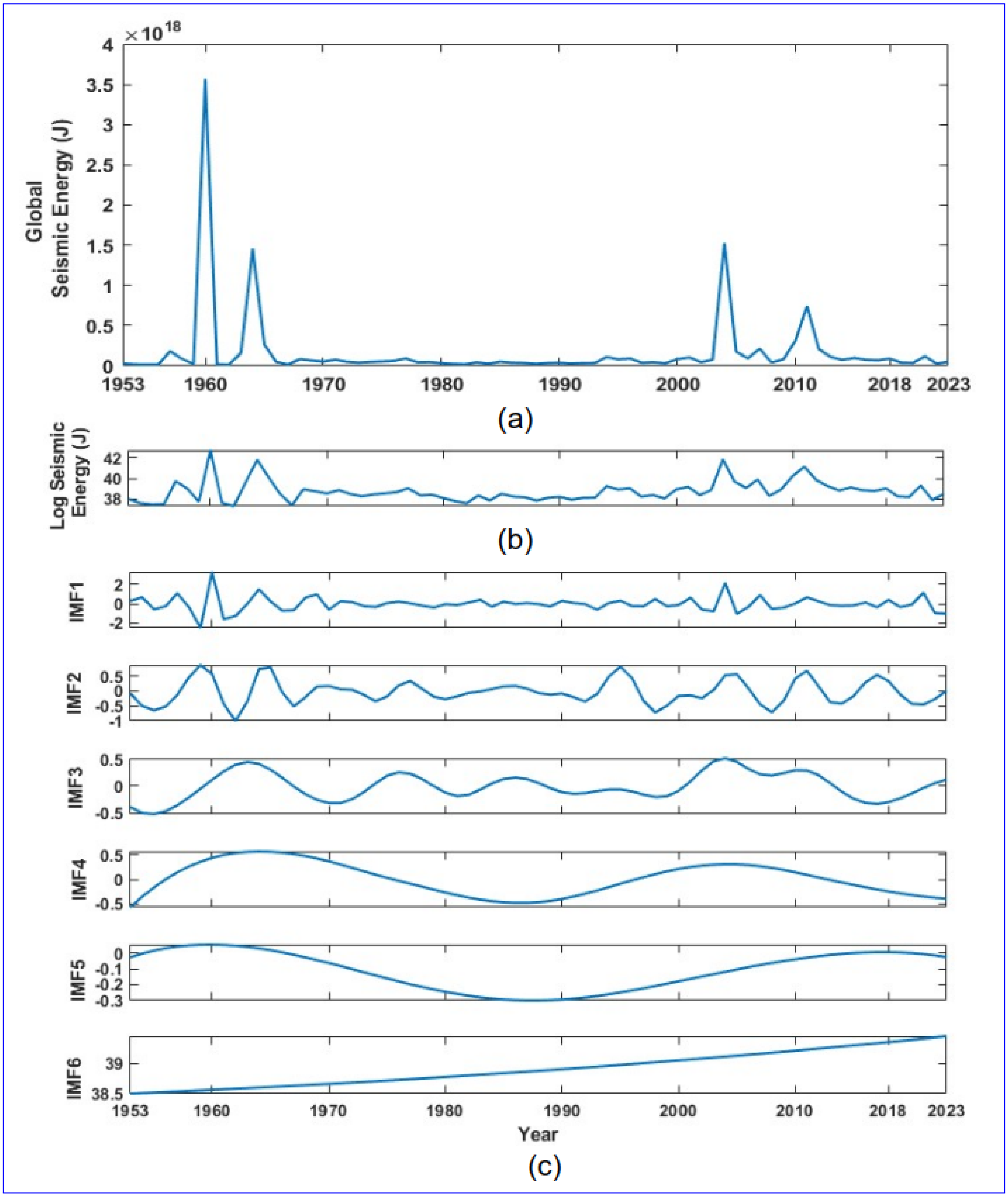

Figure 2. (a) Estimated Global seismic energy (J) time series from global catalog used in developing the models (b) log scaled Global seismic energy time series (ln(GSE)) (c) Intrinsic modes estimated from ln(GSE) by performing ensemble empirical mode decomposition (EEMD)

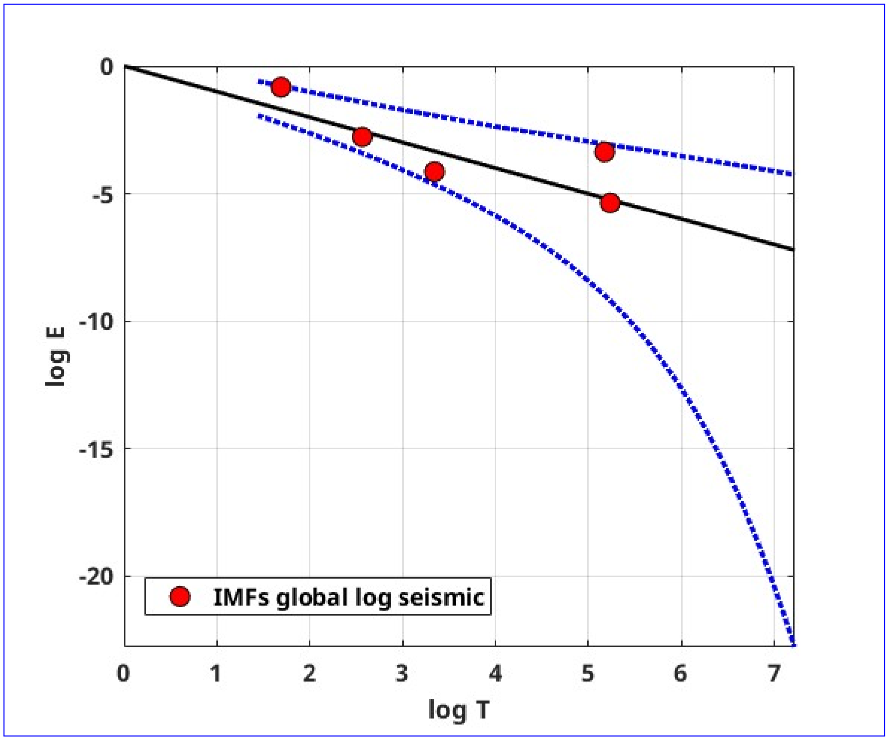

Figure 3. white noise test proposed by Wu and Huang (2004) for log seismic energy IMFs obtained from global catalog. The black line represents the expected line for white noise, and the dotted blue line shows 95 % confidence band.

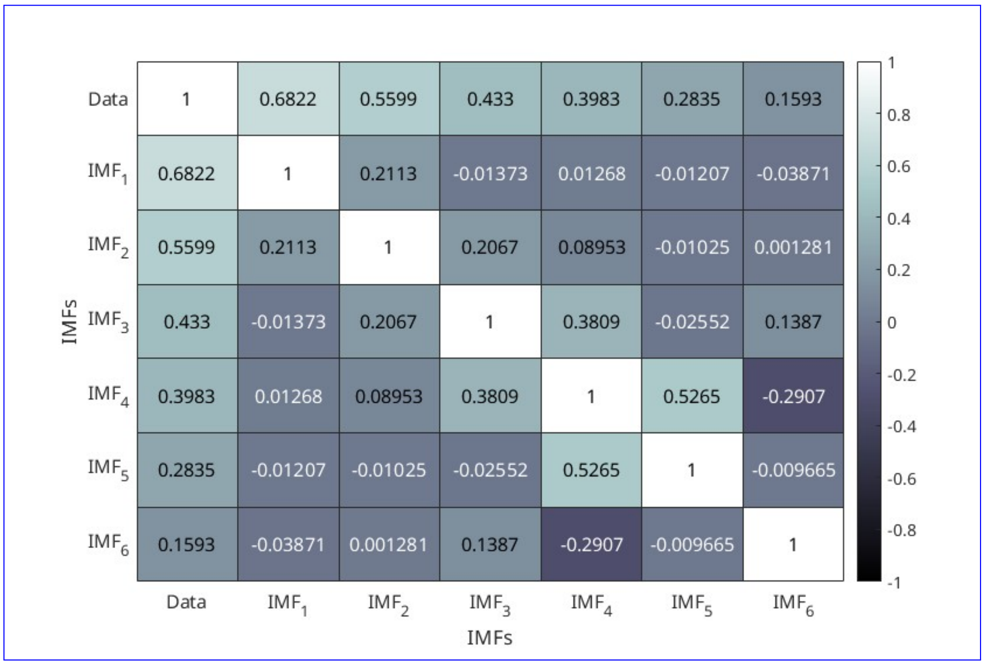

Figure 4. Correlation of the intrinsic mode functions obtained from global seismic energy time series of

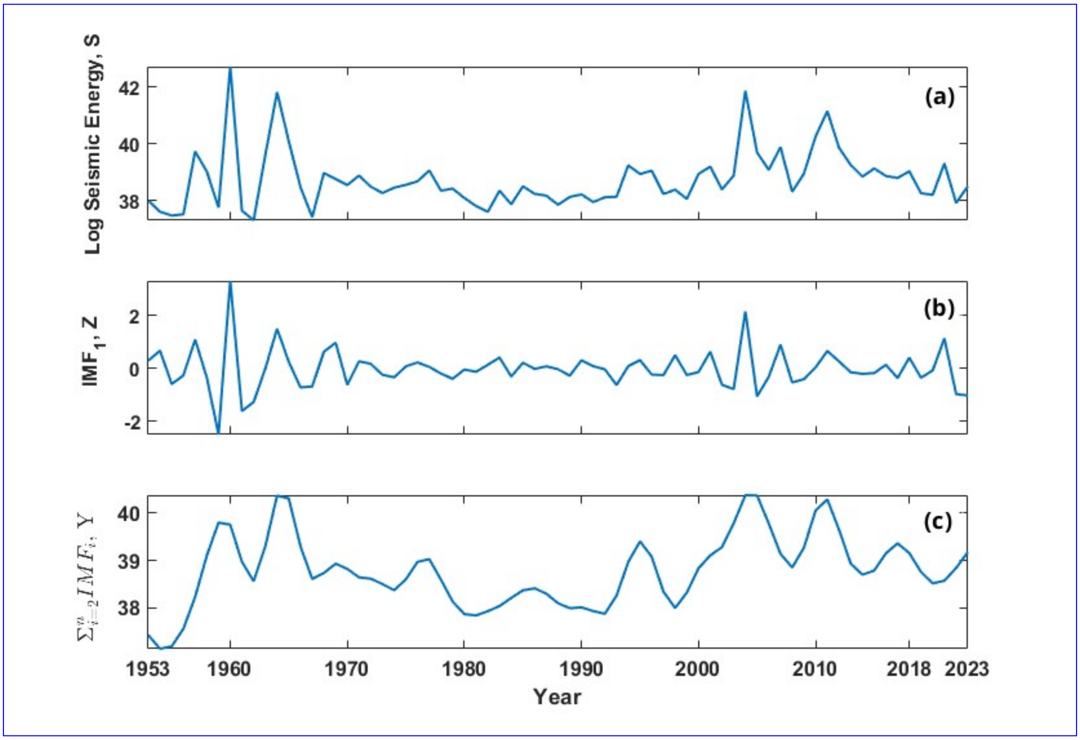

Figure 5. (a) Log-scaled global seismic energy time series (S) from global catalogue used as one of the inputs to the individual machine learning models. (b) First intrinsic mode function (Z) estimated from the log global seismic energy of the global catalogue and used as an input to individual models. (c) Summation of the second-to-last intrinsic mode functions obtained for the log global seismic energy of the global catalogue

3 Methodology

There are numerous advanced machine-learning techniques available in the literature. Some of the widely used variants include Artificial neural networks (ANN), Decision trees, Instance-based ~~learnings~~learning, classification and regression models ~~Bishop (2016)~~(Bishop, 2016). In this study, we attempted to include each of these flavours by including one representative algorithm for the analysis and further combining them using a suitable ensemble formulation. ~~The description of the models~~ Furthermore, for seismic energy forecasting, four input parameters are employed, which are log seismic energy i.e., the original time series data for log seismic energy denoted as "S", First Intrinsic mode function IMF1 denoted as "Z", Sum of remaining Intrinsic mode functions i.e, $\sum_{i=2}^{n} IMFi$ denoted as "Y", and the year of occurrence of seismic energy. Furthermore, the description of each model utilised in the study is provided further.

3.1 Multi-Layer Perceptron

Multi-Layer Perceptron (MLP) comes as part of ANN ~~Bishop (2016)~~(Bishop, 2016). A typical MLP architecture constitutes 3 layers, which are input, hidden, and output, respectively, mutually interconnected with weights. The typical functional form of an MLP from a single layer can be represented as

$$\widehat{y} = f(Wx + b) \tag{3}$$

where $\widehat{y}$ is the output from the layer, f the activation function, W the weights, x the vector of inputs corresponding to the values at the previous layer, and b the bias. The number of hidden layers, nodes, and activation functions (e.g., linear, logistic, tanh, ReLU) depend on the non-linearity between predicted and predictor variables (Kumar et al., 2023b). Once the architecture is finalized, parameters are estimated using back-propagation with mean squared error (MSE) or mean absolute error (MAE) as the cost function. Our MLP uses four input features: log seismic energy (S), IMF1 (Z), ~~$\sum_{i=1}^{n} IMFi$~~ $\sum_{i=2}^{n} IMFi$ (Y), and the year of occurrence of seismic energy. The hyperparameters for the model were optimized by varying the parameters as shown in Table 2. In time-series forecasting, the concept of lag values is fundamental. Lag values refer to the number of past observations used to predict future values in a time series (Surakhi et al., 2021). By incorporating information from previous time points, models can capture temporal dependencies and trends, leading to more accurate forecasts. The lag values were varied from 1 to 15, the number of neurons in hidden layers from 1 to 15, the learning rate (L) from 0.1 to 1.0, momentum (M) from 0.1 to 1.0, and the number of epochs from 100 to 2000. ~~The batch size was consistently set to 100.~~ A batch size of 100 was used for both the global and Himalayan models for consistency. For the Himalayan dataset (48 samples), this effectively resulted in full-batch training, which is appropriate given the small data size.

3.2 ~~Linear~~ Ridge Regression (RR)

~~LR~~

Ridge Regression (RR) is one of the widely used statistical machine learning ~~model Brownlee (2023)~~models (Hoerl and Kennard, 1970). It extends linear regression by introducing an L2 regularization term to penalize large coefficients, thereby improving generalization and reducing overfitting, especially in cases of multicollinearity or small datasets. The model establishes a linear ~~relation~~ relationship between the target variable and the input features~~. The~~ , and the general form can be expressed as:

$$\widehat{y} = \beta_{o0} + \beta_1 x_1 + ...\beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{4}$$

$$MSE = \sum_{i=1}^{p} (y - \widehat{y})^2$$

~~where y~~

where y is the target variable~~and~~ , $\widehat{y}$ is the predicted ~~variable by LR, $x_1, \ldots, x_n$~~ value, $x_1, \ldots, x_n$ are the input variables, β_0 is the ~~intersect, β_1 ,..., β_n are the slope or LR coefficient, n~~ intercept, $\beta_1, \ldots, \beta_n$ are the regression coefficients, n is the number of features, ~~p~~ p is the total number of ~~datapoints~~data points, and ϵ is the error term.

In Ridge Regression, the ~~error. Further, based on the number of input variables, there are two variants of LR, which are single-LR and multiple-LR. The unknowns β_0 ,...,β_n in the model are estimated by~~ coefficients β_i are estimated by minimizing a regularized loss function:

$$\min_{\beta} \left\{ \sum_{i=1}^{p} (y_i - \widehat{y}_i)^2 + \lambda \sum_{j=1}^{n} \beta_j^2 \right\} \tag{5}$$

Here, λ is the ridge regularization parameter that controls the strength of the penalty. The unknown parameters $\beta_0, \ldots, \beta_n$ are estimated using the gradient descent algorithm with ~~MSE~~ the Mean Squared Error (MSE) as the cost function.

$$\text{MSE} = \frac{1}{p} \sum_{i=1}^{p} (y_i - \widehat{y}_i)^2 \tag{6}$$

Based on the number of input variables, there are two common variants of linear models: single-variable linear regression and multiple-variable linear regression.

~~Linear Regression (LR)~~ RR architectures are employed for developing the forecast model. The inputs are the same as for the MLP. The lag values were varied from 1 to 15, which resulted in transformed input parameters from the higher order of the time variable and the product of time and different lagged variables. Attribute selection methods, such as the M5 method introduced by Quinlan et al. (1992) and the Greedy method can be used to reduce the number of attributes. In this work, the M5 method was used, retaining only the time steps from input variables that significantly affect the results for regression. The hyperparameter Ridge was varied from 1.0×10^{-6} to 1.0×10^{-9}, and the batch size was consistently set to 100, as shown in Table 2.

3.3 Random Forest

Random Forest (RF) is a supervised learning model used for classification and regression (Breiman, 2001a). It combines multiple decision trees to improve predictive accuracy (Cutler et al., 2012). A decision tree has decision nodes (with branches) and leaf nodes (with no branches). Trees start from a root node containing the entire dataset, splitting at each node based on attribute selection measures (ASM) like information gain or Gini index. Pruning removes unnecessary nodes to prevent overfitting.

RF addresses overfitting by building multiple decision trees on different data subsets and averaging their predictions. This ensemble of uncorrelated trees uses bootstrap sampling and feature randomness. RFs have lower computational costs, handle missing data, and can manage larger datasets efficiently. Randomness in tree generation is controlled by a fixed seed (set to 1). The number of trees (iterations) was varied from 30 to 120, and tree depth is unlimited (depth of 0). Features at each split are calculated by $(\log_2(\mathrm{N})+1)$, where N is the number of predictors (Breiman, 2001b). The lag values were varied from 1 to 15, and the batch size and bag size were consistently set to 100, as shown in Table 2.

3.4 Sequential Minimal Optimization Optimisation regression

Sequential Minimal Optimization Optimisation (SMO) is an iterative algorithm proposed by Platt (1998) for solving regression problems using support vector machines (SVM). SMO simplifies the optimization optimisation problem by breaking it down into smaller sub-problems that can be solved analytically, which makes it more efficient for training SVMs. Further improvements to SMO for regression were proposed by Shevade et al. (1999), who introduced modifications to enhance its efficiency. These improvements address the way SMO updates and maintains threshold values, resulting in two significantly more efficient versions for regression tasks. The corresponding algorithm effectively solves the quadratic optimization optimisation problem inherent in SVM training. SMOreg employs four input features: log seismic energy (S), IMF1 (Z), $\sum_{i=1}^{n} IMFi$ $\sum_{i=2}^{n} IMFi$ (Y), and the year of occurrence of seismic energy. The model is optimized optimised by fine-tuning hyperparameters such as the complexity number (C) and epsilon (ϵ). In this study, the complexity number was varied from 1 to 9 to balance minimizing minimising training error and problem complexity. Epsilon was varied from 1.0×10^{-9} to 1.0×10^{-15} to determine the allowable error within the epsilon tube. The kernel function, crucial for SVMs, impacts the ability to manage complex relationships in the data. Various kernel functions, including polynomial (Polykernel), puk, RBF, and string kernel, were considered. The lag values were varied from 1 to 15, and the batch size was consistently set to 100, as shown in Table 2.

3.5 Instance-Bases Instance-Based learning with parameter k

Instance-based learning (IBL), also called instance-based learning with parameter k (IBk), is a type of unsupervised supervised learning used for both classification and regression problems Aha et al. (1991); Jo and Jo (2021). It falls under lazy learning algorithms, which memorize memorise the training data and make predictions based on the similarity between new and training datasets. The parameter k represents the number of nearest neighbors neighbours considered for predictions. IBk searches for the k most similar instances from the training dataset based on the similarity measures using Manhattan distance, Euclidean distance or other distance matrices. More accurately, let the given instance x be described by the feature vector

$\langle a_1(x), a_2(x), ...a_n(x)\rangle$. where $a_r(x)$ denotes the value of rth attribute of instance x. The euclidean distance between x_i and x_j is given by

$$d(x_i, x_j) \equiv \sqrt{\sum_{r=1}^{n} ((a_r(x_i) - a_r(x_j))^2)} \tag{7}$$

Here in the K-Nearest Neighbour algorithm, the target function may be either real-valued or ~~discrete- valued~~ discrete-valued, defined by $\widehat{f}(x_q)$, which is just the most common value of f among k training example nearest to x_q.

$$\widehat{f} \leftarrow \frac{\sum_{i=1}^{k} f(x_i)}{k}) \tag{8}$$

In this study, the k value was varied between 1 and 12, using Euclidean and Manhattan distances. The lag values were varied from 1 to 15, resulting in transformed input features. The number of nearest neighbors for prediction was set using these variations, with a consistent batch size of 100, as shown in Table 2. The LinearNNSearch algorithm, suitable for small datasets, was adopted for nearest neighbor search, employing a linear search across all data points.

4 Ensemble Models

Hyndman and Athanasopoulos (2018) suggested that in time series forecasting approaches there is a need to include relevant characteristics to increase accuracy. Also, the wider notion of adding time-related characteristics is a well-known approach in machine learning and forecasting. Hence, the inclusion of year as one of the input features is decided in the present study. Now using only IMFs as the inputs for developing forecasting model might not be that effective because ~~Finding the IMF 1 at the conclusion of~~ finding IMF1 at the boundary of the data can be challenging ~~because to the lack of defined~~ due to undefined envelopes on both sides. To address this challenge, the previous value of the end point might be used as the next value. However, this strategy is not effective for anticipating issues. If n data is provided, IMFs can only be extracted for i = 2, 3, 4,..., n - 1. As the distance between extrema increases, extrapolation errors might permeate into the signal, misrepresenting greater IMFs at the end points. Hence~~The~~ , the inclusion of S in the input ensures that the deficiency of the empirical mode decomposition to estimate the end value does not affect the model predictions. Hence, ~~The~~ the time histories from S, Y, and Z (see Figure ~~2~~5 for global catalog and corresponding figure for validation catalog is present in Figure S2 of supplementary document) were provided simultaneously as inputs to the model along with the year to predict S as the output variable. The time series from up to 1995 is taken for the training phase and the remaining data from 1996-2015 is considered for the testing phase for validation catalog and for updated global catalog time series upto 2009 is taken for training and 2009-2023 for testing. Additionally, the combination of the inputs is determined such that it gives the best prediction performing beta coefficient analysis. Whereby the approach tries different indices and retains only those that are significant and ~~decreases the~~ minimizes the prediction error. To determine the optimal lag period, an analysis was performed by varying the lag from 1 to 15 time points. The optimal lag value was identified by evaluating the model's performance and selecting the lag period that minimizes the prediction error. The

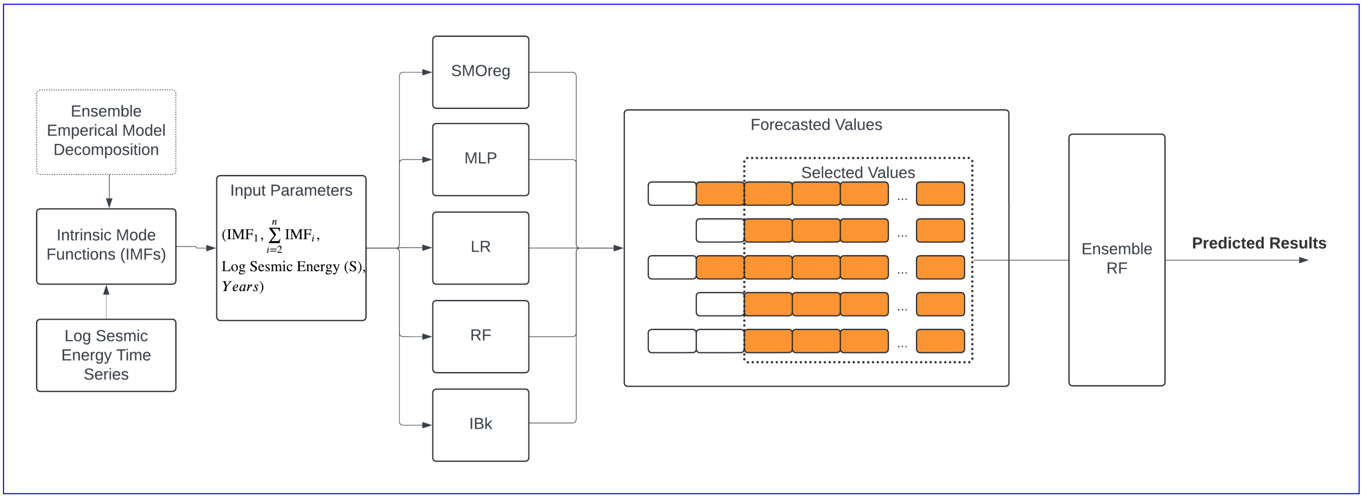

Figure 6. The flow chart representing the various steps and modelling approaches adopted for seismic energy forecast

range of lag values, along with the other hyperparameters varied to get the final model, are presented in Table 2. Also, different lag values for different models were obtained, and these values, along with the other hyperparameters used in the prediction models, are present in Table 3. The overall flow of the proposed approach is described through Figure ~~4~~6. Furthermore, a detailed description of the model architectures of ensemble model is furnished further under section 4.1.

4.1 Ensembled model architecture

Ensembling can be performed by second-level trainable combiners through meta-learning techniques Duin and Tax (2000) In the present study, the stacking method was employed, wherein the output results of the base or weak learners were used as features in an intermediate space. These features were subsequently fed as input to a second-level meta-learner to perform a trained combination of weak learners. The base learners in this study included MLP, LR, RF, SMOReg, and IBk, which forecasted the values of seismic energy, as depicted in ~~the figure 5~~Figure 6. These base learners were optimized models with varying parameters, as detailed in Table 2. Each base learner produced predictions of different lengths due to the use of varying lag values in their optimization process. To address this discrepancy and ensure consistency across all learners, ~~we considered~~ the shortest prediction length among the ~~base learners~~models was used to align inputs. This approach was visually represented in Figure ~~4~~6, where the orange blocks indicated the forecasted values and the empty blocks denoted absent values. Here, forecasted values from all five techniques were then used as input features for the ensemble RF model, with the actual log seismic energy as the target for regression. This ensemble RF model ultimately predicted the log seismic energy, integrating the results from the base learners to ~~enhance~~ improve ensemble prediction accuracy.

To ensure optimal performance of the models, we employed grid search techniques for hyperparameter tuning. Grid search is an exhaustive search method that tests all possible combinations of specified hyperparameters to identify the best-performing configuration for each model. The process involves defining a parameter grid for each model, specifying a range of values for each hyperparameter. For example, for the Multi-Layer Perceptron (MLP), the grid includes different numbers of neurons in the hidden layers, learning rates, and momentum values. For the Random Forest (RF), the grid includes the number of trees and maximum depth. Each combination of hyperparameters is evaluated using an 80:20 train-test split method. The dataset is divided into 80 % training and 20 % testing sets, with the model trained on the training set and evaluated on the testing set. The performance of each hyperparameter combination is assessed using an appropriate evaluation metric, such as mean squared error (MSE) or root mean square error (RMSE). The combination of hyperparameters that results in the lowest error (or highest accuracy) on the training set is selected as the optimal configuration for the model. The final model, with ~~optimized~~ optimised hyperparameters, is then tested on the testing dataset to evaluate its ~~generalization~~ generalisation performance. This ensures that the selected model configuration not only performs well on the training data but also maintains its accuracy~~and robustness on unseen data~~, generalization and robustness.

Table 2. Combinations considered for the optimization of hyper-parameters in the model architecture

~~Model~~**Model**	~~Parameter~~**Parameter**	~~Range of parameter~~**Range of parameter**	
		~~Global~~**Global**	~~Western Himalaya~~**Western Himalaya**
MLP	Lag	1 to 15	1 to 12
	Hidden layers	1,2	1,2
	Neuron in hidden layers	1 to 15	1 to 15
	Learning Rate	0.1 to 1.0	0.1 to 1.0
	Momentum	0.1 to 1.0	0.1 to 1.0
	Batch size	100	100
	Epochs	100 to 2000	100 to 2000
~~Linear~~Ridge Regression	Lag	1 to 15	1 to 12
	Ridge	1.0E-6 to 1.0E-9	1.0E-6 to 1.0E-9
	Batch size	100	100
Random Forest	Lag	1 to 15	1 to 12
	Batch size	100	100
	Bag size	100	100
	Number of trees	30 to 120	30 to 120
SMOreg	Lag	1 to 15	1 to 12
	Kernel	Poly, puk, RBF, string	Poly, puk, RBF, string
	Epsilon (E)	1.0E-9 to 1.0E-15	1.0E-9 to 1.0E-15
	Complexity	1 to 9	1 to 9
	Batch size	100	100
Ibk	Lag	1 to 15	1 to 12
	K	1 to 12	1 to 12
	Distance Function	Euclidean and Manhattan	Euclidean and Manhattan
	Batch size	100	100

Table 3. Optimized hyper-parameter of the models for the data under consideration

Model	**Parameters**	**Validation**	**Global**	**Western Himalaya**
MLP	Lag	8	8	6
	Hidden layers	2	2	1
	Neuron in hidden layer	4,2	21,18	11
	Learning Rate	0.3	0.2	0.3
	Momentum	0.2	0.1	0.2
	Batch Size	100	100	100
	Epochs	1500	1000	500
~~Linear~~ Ridge Regression	Lag	6	9	6
	Ridge	1.0E-8	1.0E-12	1.0E-8
	Batch Size	100	100	100
Random Forest	Lag	7	5	7
	Batch Size	100	100	100
	Bag size	100	100	100
	Number of Trees	100	70	100
SMOreg	Lag	8	4	5
	Kernel	Poly	Poly	Poly
	Epsilon (E)	1.0E-12	1.0E-12	1.0E-12
	Complexity	1	1	1
	Batch Size	100	100	100
Ibk	Lag	8	8	8
	K	2	9	9
	Distance Function	Euclidean	Euclidean	Euclidean
	Batch Size	100	100	100

4.2 Input and output to different models

The proposed two-level ensemble forecasting framework operates on a structured input-output configuration where the input vector comprises four fundamental components: log seismic energy (S), IMF_1 (Z),$\sum_{i=2}^{n} IMF_i$ (Y), and temporal information i.e., the year of occurrence of seismic energy, systematically organised into sequential packets i.e., $(X_1, X_2, ..., X_m)$ for training, and $(X_{m+1}, ..., X_n)$ for testing dataset with suitable temporal dependency structure (i.e., lag value). All four features (S,Z,Y, and year) have lagged values in each input packet, which are structured so that past observations are utilized to forecast future S values. A generalised flow of input and output is presented in Fig. 7 by considering the lag value as 8. However, different base models have different lag values. These lag values are decided based on the optimised model on varying parameters as presented in Table 2. The final values of lag for different base learners are presented in Table 3. To be clear, the model is trained to predict the seismic energy at the subsequent time step (e.g., S9), and each input packet contains eight previous time steps of S,Y,Z and year as features for lag value of 8. This sliding-window method guarantees that the learning algorithm employs logical events sequences and catches temporal patterns in the data. The first level of the ensemble processes these input packets with different lag values through five different base models (Linear Regression, Multi-Layer Perceptron, SMOreg, IBK, and Random Forest) that make initial predictions. This changes the original four-dimensional input space into a five-dimensional prediction space, where each observation is represented by the outputs from all base models. After that, the second level uses a Random Forest meta-learner to process these stacked five-dimensional prediction vectors and make final consensus predictions (S_9, S_{10}, S_{11}, S_{12}, ..., S_n). In order to ensure temporal consistency and assess prediction performance, the same input structure is employed during the testing phase by shifting the lag window across unseen sequences. To prevent input mismatches, the output from only the overlapping prediction region (shortest sequence) is sent to the meta-learner because each base learner employs a distinct lag length. This is done by using the collective intelligence of the base models and learning the best combination weights and interaction patterns to make better predictions than individual algorithmic approaches. This is done by systematically combining different temporal pattern recognition capabilities across different seismic conditions.

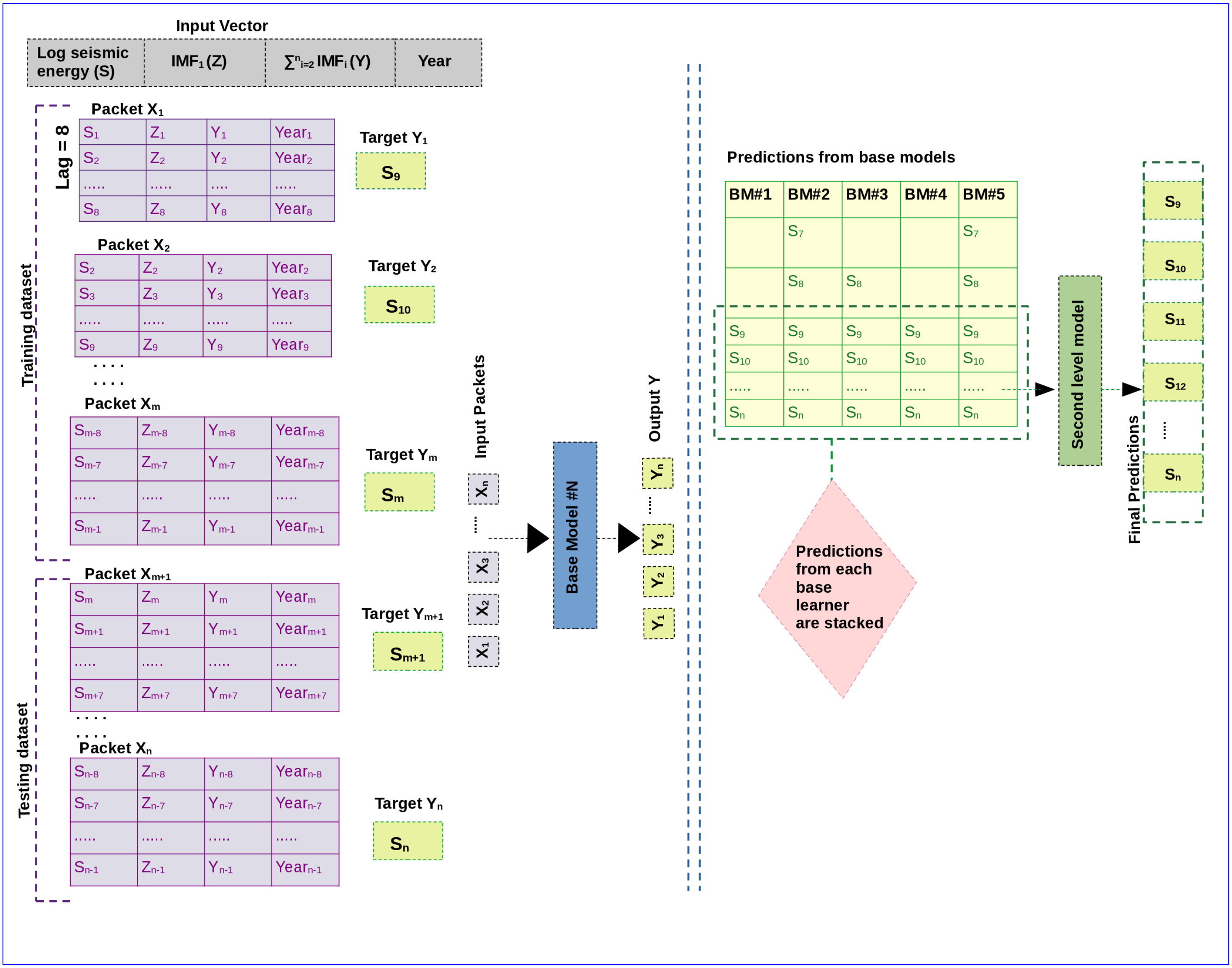

Figure 7. A general diagram showing the input and Output at different stages of the methodology with a representative value of lag=8 .

5 Validation

Based on the optimized hyper-parameters, the predictions from the models trained on the data derived from validation and global catalog in the training and testing phases are summarized in Figure ~~5~~ 8 and Figure 9 for the individual models. The performance is observed to vary between models both in training and testing parts. Hence, to have a quantitative evaluation of the model performances, the following indicators are estimated for both the training and testing phases:

1. Standard deviation of error ($\sigma(\epsilon)$)

$$\sigma(\epsilon) = \sqrt{\frac{\sum_{i=1}^{N}(S_i - \hat{S}_i) - \overline{(S_i - \hat{S}_i)})}{N-1}} \tag{9}$$

2. Pearson correlation coefficient (R)

$$R = \frac{\sum_{i=1}^{N}(S_i - \overline{S_i})(\hat{S}_i - \overline{\hat{S}_i})}{\sqrt{\sum_{i=1}^{N}(\hat{S}_i - \overline{\hat{S}_i})^2 \sum_{i=1}^{N}(S_i - \overline{S_i})^2}} \tag{10}$$

3. Performance Parameter (PP)

$$PP = 1 - \frac{\langle \parallel S - \hat{S} \parallel^2 \rangle}{\sigma_{\hat{S}}^2} \tag{11}$$

4. Root Mean Squared Error ($RMSE$)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(S_i - \hat{S}_i)^2}{N}} \tag{12}$$

To ensure reliable model selection while mitigating overfitting, hyperparameter tuning was performed using the training dataset. This approach avoids information leakage from the testing set and ensures unbiased performance estimates. Rather than reserving a third dataset exclusively for validation, this strategy allows efficient use of available data while maintaining robust generalization capability. Final model performance was then evaluated on the separate hold-out test set. The corresponding estimations for the weak-learners and ensemble models are summarised Table 4. ~~Out of the models ,~~ For models developed with validation catalog log seismic energy data the MLP model performs well in the training phase; however, at the testing phase, it is relatively weak. But the MLP model developed on global catalog data has performed well in both training and testing phase. The RF model also shows a similar trend to that of MLP. However, LR and SMOreg models are observed to perform consistently in both the training and testing phases. However, IBk architecture is the least-performing model for ~~the data~~ both the data of validation and global catalog under consideration. Nevertheless, according to a detailed literature review explained earlier, ensemble models are expected to improve the model's performance. Thus, a suitable ensemble model is developed as described in section 4. The corresponding model performance is summarized in Figure 6 and 10 for validation and Figure 11 for global catalog, and in Table 4.

It is interesting to note that the resultant Interestingly, across both datasets, the ensemble model outperformed the weak learnersany single base learner in terms of RMSE and correlation coefficients. Furthermore, the model performance is good and consistent in both the training and testing phases. Additionally, from a comparison of performance with the previous study (Raghukanth et al., 2017) (Table 4) on the same data i.e., validation catalog, we were able to conclude that the ensemble model performs significantly better, having lesser variability. Thus, the corresponding model can be suitably implemented in making reliable As a result, the relevant model may be a good fit for real-time seismic energy forecasting applications and can be appropriately used to produce accurate seismic energy predictionsand thus has significant application in earthquake forecasts.. With this motivation, we further attempted to explore the proposed approach for regional-level seismic energy forecast. The active Western Himalayan province is chosen for the evaluations. A detailed description of the corresponding data, processing and modelling is provided under section 6.

Table 4. The performance evaluation of individual models and the final ensemble random forest model at the training and testing phases of Global seismic energy

Models	Training				Tes	
	$\sigma(\epsilon)$	R	PP	$RMSE$	$\sigma(\epsilon)$	R
Raghukanth et al. (2017)	0.285	0.968	0.920	0.284	0.361	0.940
Validation						
MLP*	0.259	0.971	0.887	0.362	0.497	0.862
Linear Regression (LRRidge Regression (RR)	0.347	0.946	0.896	0.345	0.373	0.926
Random Forest (RF)	0.378	0.974	0.877	0.377	0.789	0.693
SMOreg**	0.309	0.958	0.919	0.307	0.400	0.916
IBk***	0.642	0.819	0.649	0.639	0.931	0.323
Ensemble RF	**0.127**	**0.994**	**0.986**	**0.127**	**0.136**	**0.992**
Global						
MLP*	0.109	0.993	0.985	0.108	0.482	0.861
Ridge Regression (RR)	0.361	0.915	0.840	0.358	0.579	0.776
Random Forest (RF)	0.359	0.964	0.884	0.356	0.577	0.825
SMOreg**	0.570	0.881	0.698	0.575	0.595	0.725
IBk***	0.733	0.617	0.324	0.738	0.783	0.648
Ensemble RF	**0.077**	**0.997**	**0.992**	**0.077**	**0.185**	**0.978**

* MLP- Multi-Layer Perceptron; **SMOreg - Sequential Minimal Optimization regression;

***IBk - Instance-Bases learning with parameter k

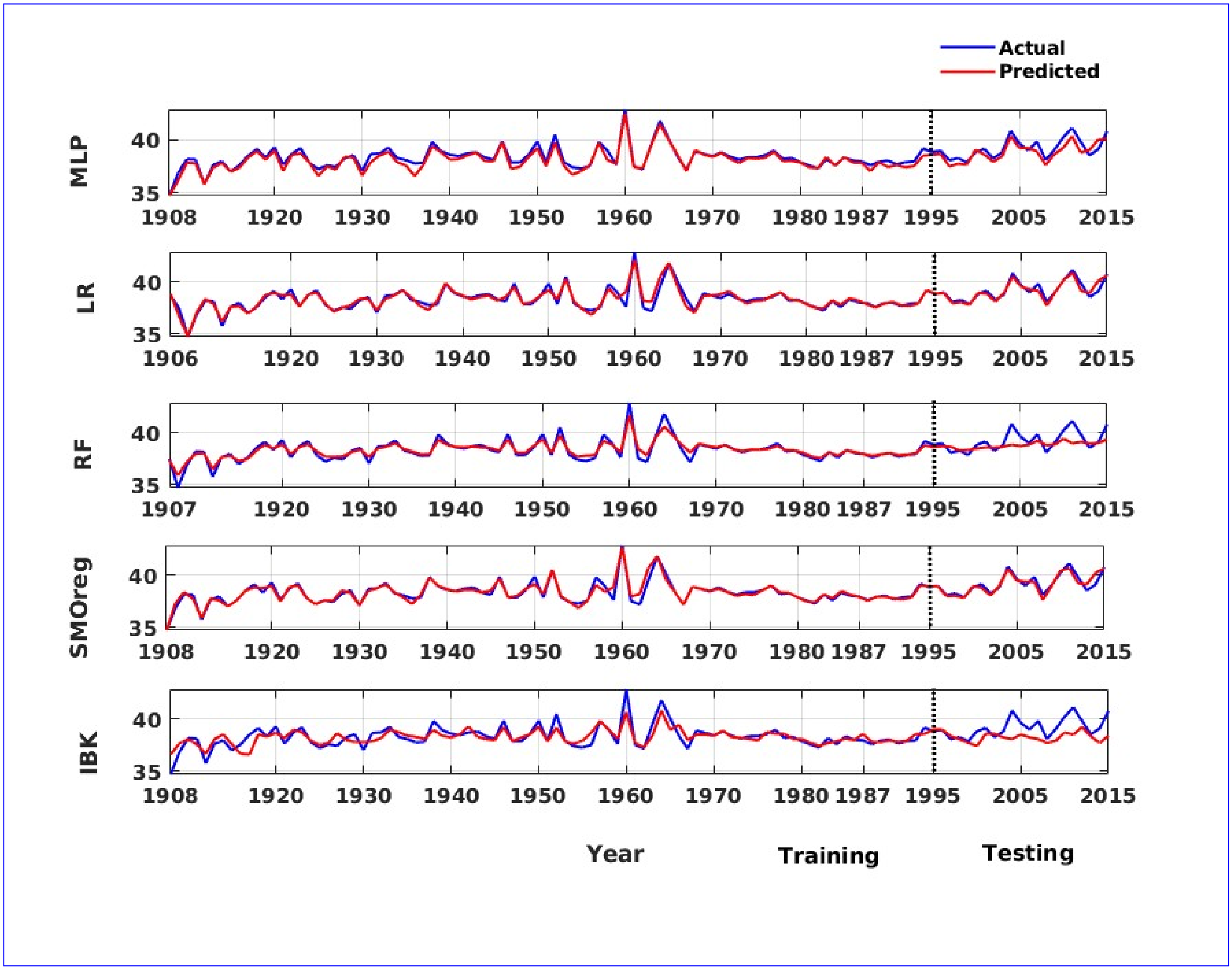

Figure 8. Actual vs Predicted values of global log seismic energy from individual machine learning technique adopted in this work for validation catalog data

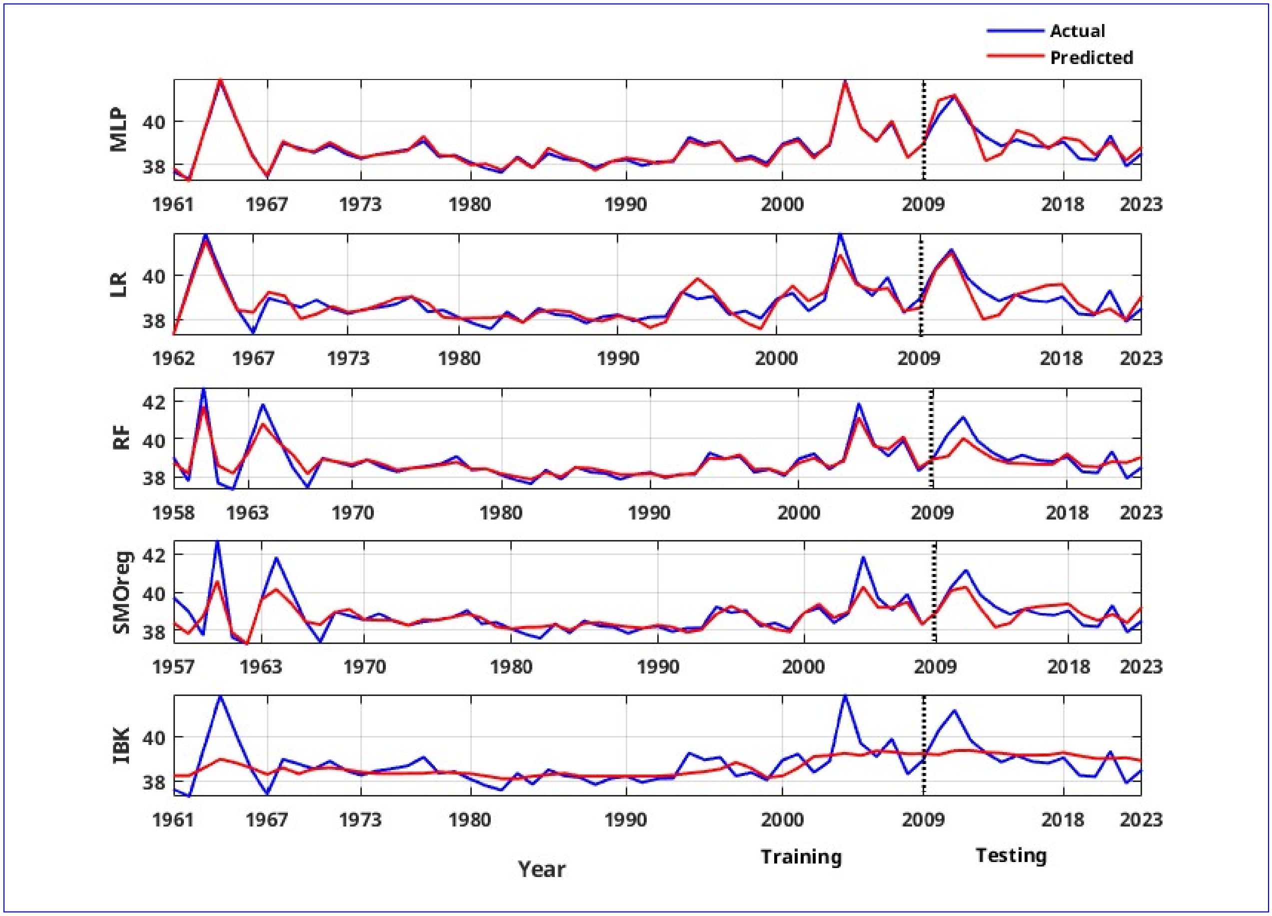

Figure 9. Actual vs Predicted values of individual machine learning technique for global log seismic energy calculated from ~~proposed ensembled random forest model~~global catalog

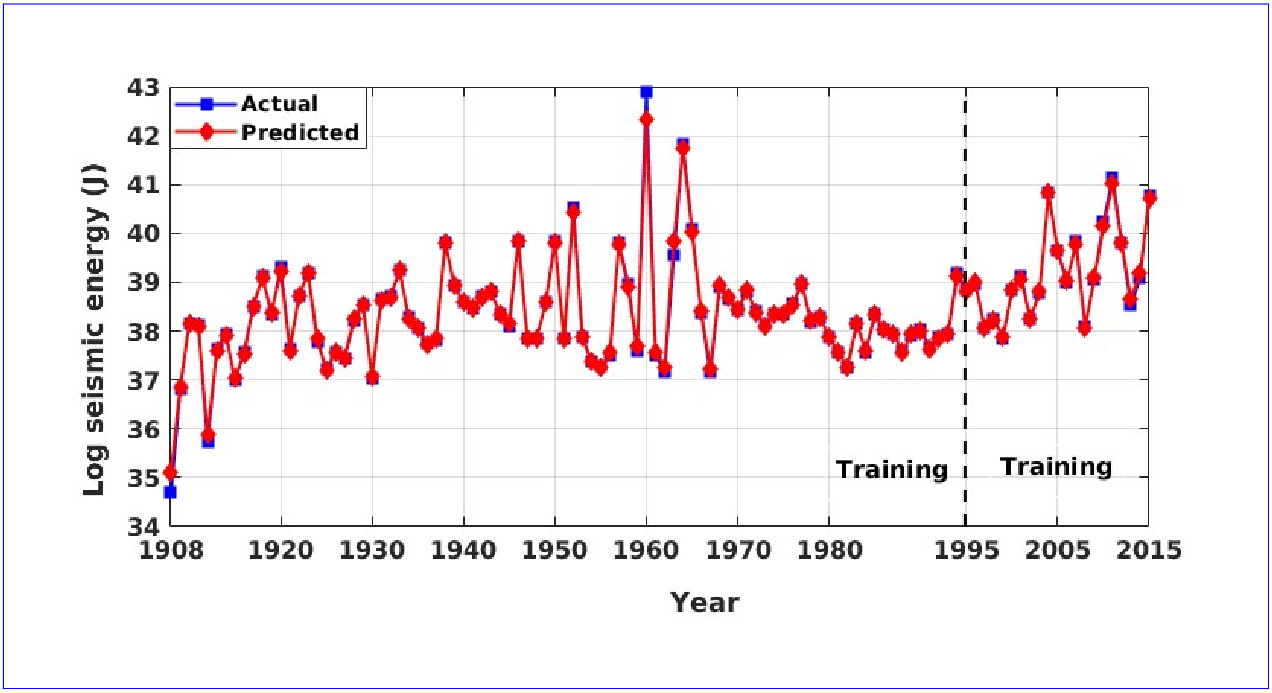

Figure 10. Actual vs Predicted values for global seismic energy from proposed ensembled random forest model for validation catalog

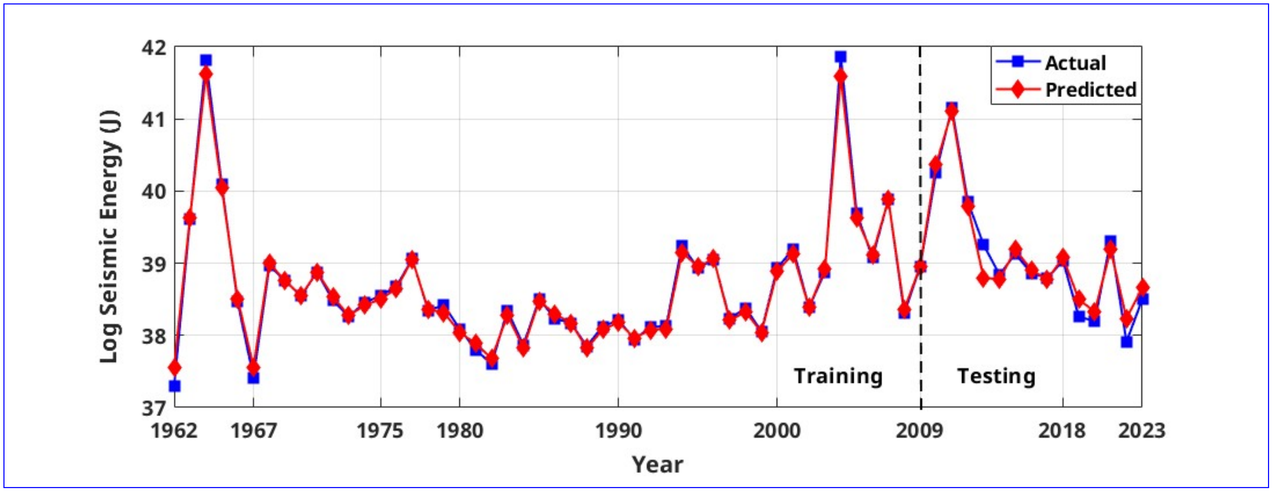

Figure 11. Actual vs Predicted values of proposed ensembled random forest model for global log seismic energy calculated from global catalog

6 Application to Western Himalayas

The vast Indian subcontinent region, which includes Bangladesh, India, Nepal, Bhutan, Pakistan, and Sri Lanka, is prone to frequent and severe earthquakes. This region is particularly vulnerable to seismic activity because of the tectonic ~~moments~~ movements and the proximity to the convergent margin of the Indian and Eurasian plates. Whereby the subsequent collision has resulted in a vast mountain belt known as the Great Himalayas, where frequent earthquakes are caused by ongoing tectonic activity. The uplift due to collision has caused linear zones of deformation, leading to crustal shortening along major boundary faults. These faults are Himalayan Frontal Thrust (HFT), Main Boundary Thrust (MBT) and Main Central Thrust (MCT), which resulted in some large paleo-earthquakes in the region (Geological Survey of India (GSI), 2000). India's northern and northeastern parts are more vulnerable; they are classified majorly as seismic zone IV and V in IS:1893-1 (2016), which indicates the highest degree of seismic hazard. As the Indian plate slowly sinks and subducts beneath Asia at a pace of around 47 mm/year, this collision tectonics makes the area very vulnerable to catastrophic earthquakes due to energy accumulation and subsequent release Bendick and Bilham (2001). Several large earthquakes ~~has~~ have been observed in this region in last two decades, such as the Uttarkashi earthquake in 1991 (m_b 6.6), the Chamoli earthquake in 1999 (m_b 6.3), the Kashmir earthquake in 2005 (M_w 7.8), the Sikkim earthquake in 2011 (M_w 6.9), Nepal earthquake in 2015 (M_w 7.8). Moreover, the region between Kangra earthquake in 1905 (M_w 7.8) and Bihar-Nepal earthquake in 1934 (M_w 8) is relatively silent and hence is identified as the central seismic gap (CSG) region having potential of generating $>$ 8 magnitude event (Bilham et al., 1997; Khattri, 1999; Bilham, 2019). It is, therefore, essential to have a reliable quantification of hazard in order to reduce the related seismic risks in the active western Himalayan area. The current approach, designed especially for this seismically active and dynamic area, is critical. Its alignment with the distinct geological features of the western Himalayas emphasises its significance and makes it an indispensable instrument for reducing the possible effect of seismic occurrences in this susceptible region. Furthermore, seismic hazard studies have also reported high values for design parameters for the region due to its active tectonics and recurrence rate (NDMA, 2010; Nath and Thingbaijam, 2012; Dhanya and Raghukanth, 2022; Sreejaya et al., 2022). Considering the tectonics and the risk due to exposure an efficient forecast model is critical for this region. However, ~~such a model is not attempted for the region~~a dedicated model for annual seismic energy forecasting is still lacking. Hence, the present work aims to develop a robust forecast model for annual seismic energy release in the Western Himalayas. The ensemble model algorithm validated in Section 5 shall be utilized for the work. A detailed description of the data compiled for the region and the resultant forecast is furnished further. This application demonstrates the extension of our ensemble methodology to a regional scale with real hazard implications.

6.1 Study region and data preparation

From the tectonics described earlier, the Hindukhush region and the adjoined region are seismically very active, as evident from Figure ~~8~~12. This observation is consistent with both the frequency of documented earthquakes and the geographical distribution of faults and lineaments (Figure ~~8~~12). This has motivated researchers to divide the whole geographical area of the country and the adjoining region into different seismic zones. For instance, Khattri et al. (1984), used seismotectonic and seismicity data to split the nation into 24 source zones. While Bhatia et al. (1999) found 85 source zones in India, 40 seismic zones were detected by another research Parvez et al. (2003). Considering all these past efforts The National Disaster Management Authority (NDMA) of India created a thorough study in 2010 that further divided the whole Indian subcontinent into 32 seismic zones, designated SZ-1 to SZ-32 NDMA (2010). This division of seismic zones was done by considering factors such as regional geodynamics, fault alignment, and recurrence parameters for the regions. Among these 32 zones, the present study focuses on earthquakes in SZ-1, SZ-2, SZ-6, and SZ-30, which lie in the Western part of the Himalayan belt. As this is preliminary work towards energy prediction for the regions, the comprehensive catalogue is combined together for all zones in the Western Himalayas. Here, the earthquake catalog for the region has been taken from Dhanya et al. (2022) and was updated until December 31, 2023, via the USGS seismic database (https://earthquake.usgs.gov/earthquakes/search/). There are 25,769 events (for the western Himalayan region considered for this work) in the final updated catalog spanning from 1250 BC to 2023 AD. Furthermore, the updated catalogue has been checked for both completeness for year and magnitude. For completeness of year, the method suggested by Stepp (1972) has been adopted in which the standard deviation of mean rate is plotted as a function of sample length and the period where this value deviated from tangent, i.e., $(1/\sqrt{(T)})$ is consider as completeness for considered magnitude. Furthermore, the magnitude of completeness was identified from the maximum curvature method proposed by Wiemer and Wyss (2000). Whereby the catalogue compiled for the region in the present work is observed to have a magnitude of completeness of M_w 4 and the corresponding year of completeness as 1964 (Figure ~~9~~ 13 (a) and ~~9~~ 13 (b). Hence, events spanning from 1964 and having a magnitude greater than 4 M_w have been considered for further input preparation. The distribution of the event in the final compiled catalogue can be identified from Figure ~~10~~14.

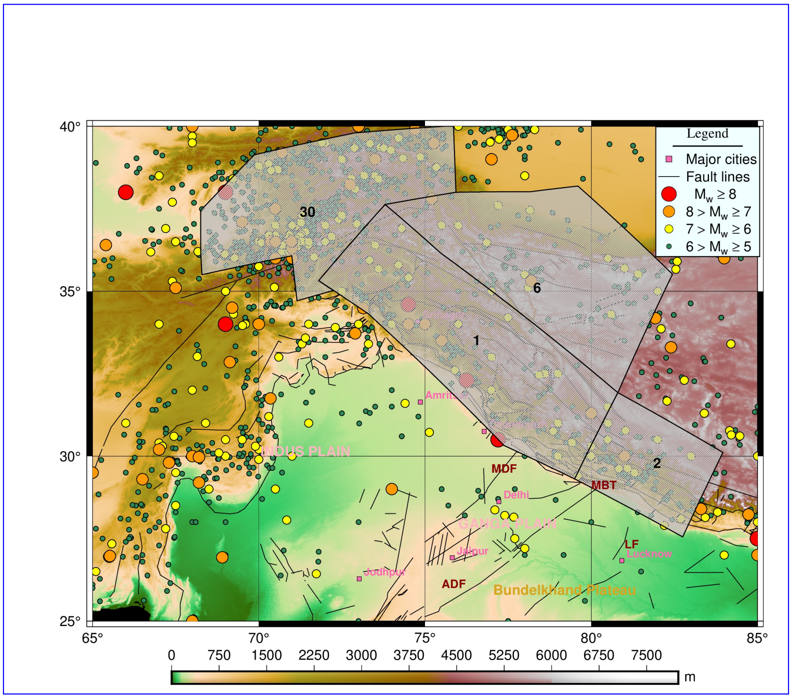

Figure 12. The regional level tectonics and the seismogenic zones as per NDMA (2010) in Western Himalayas

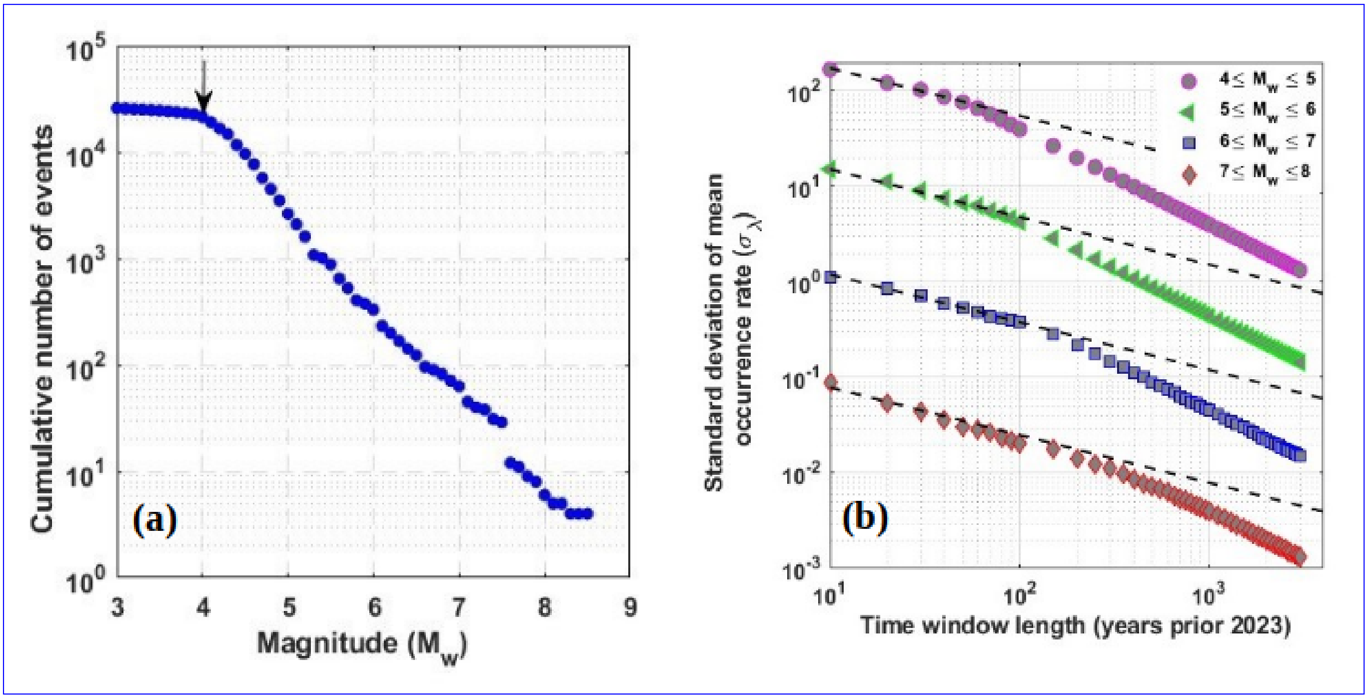

Figure 13. (a) The magnitude of completeness (Wiemer and Wyss, 2000) and (b) year of completeness (Stepp, 1972) obtained for the catalogue compiled for Western Himalaya region

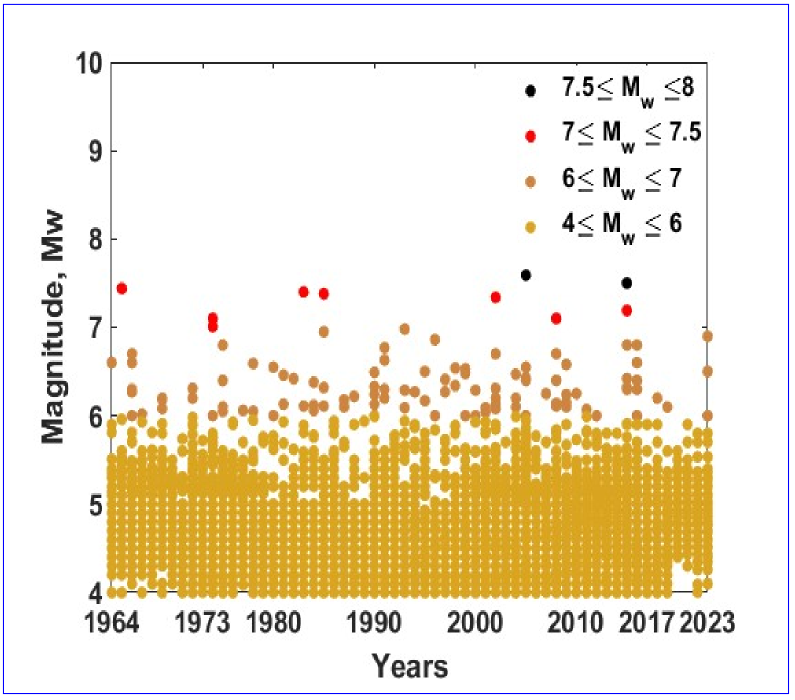

Figure 14. Distribution of events for the complete catalog for the western Himalayan region. Events with magnitude equal to or greater than 7.5 M_w are shown with black legends.

6.2 Seismic time series and mode decomposition

The same approach discussed under section 2 has been adopted for Western Himalaya, where the complete catalogue spanning from 1964-2023 with 20774 events having a magnitude in different scales is unified by converting all the earthquake magnitude into moment magnitude M_w. The catalogue has two major earthquakes having a magnitude $\geq$ 7.5 M_w, the 2005 Kashmir earthquake (7.6 M_w) and the 2015 Afghanistan earthquake (7.5 M_w) (Figure ~~10~~14). After unification, magnitudes are converted to seismic energy using Eqn. 1 discussed under section 2 considering the physical significance of the parameter in earthquake occurrence. Furthermore, energies are added annually to get the seismic energy time series. From Figure ~~11~~ 15 (a), one can note two distinct peaks at 2005 and 2015 indicator of the major earthquakes explained earlier. Furthermore, to enhance the predictability of the time series, these values are converted to log scale to remove sudden jumps similar to that described in Section 2. After obtaining the seismic energy time series, it further decomposes into intrinsic mode functions using the EEMD technique. The corresponding division is expected to account for the linear and non-linear components of time histories appropriately. The decomposition allows the model to separately learn dominant cycles (IMF1) and low-frequency, non-linear patterns (IMF2–IMF5). Thus, by applying the EEMD technique as described in section 2.1 on the log-scaled Western Himalaya seismic energy time series (Figure ~~11~~ 15 (b), first subplot), we were able to obtain five intrinsic mode functions (Figure ~~11~~ 15 (b)). Furthermore, the correlation coefficients between the models are presented in Figure ~~12~~16. It is noted that IMFs are mostly uncorrelated and orthogonal. Table 1 lists the periods and variance of all five IMFs in log-scaled seismic energy time series. Similar to the global seismic energy modes, for the regional model also, the period of the IMFs seems to increase. Furthermore, the IMF_1 is observed to capture maximum variance in the data. To incorporate these IMFs as input for the machine learning techniques, IMF_1 and the sum of IMF_2 to IMF_5 ($\sum_{i=2}^{5}$ IMF$_i$) have been taken separately as suggested in Raghukanth et al. (2017). Furthermore, considering the limitation of ~~the EEMD approach in predicting the last value appropriately as similar to that discussed in section 5~~EEMD at time series boundary, the log seismic energy and year ~~of occurrence is also considered as inputs into the machine learning models. The corresponding information is illustrated in Figure13.~~ are also used as model inputs (see Figure17). A detailed description of the model architecture that was found optimal for the regional database is discussed further

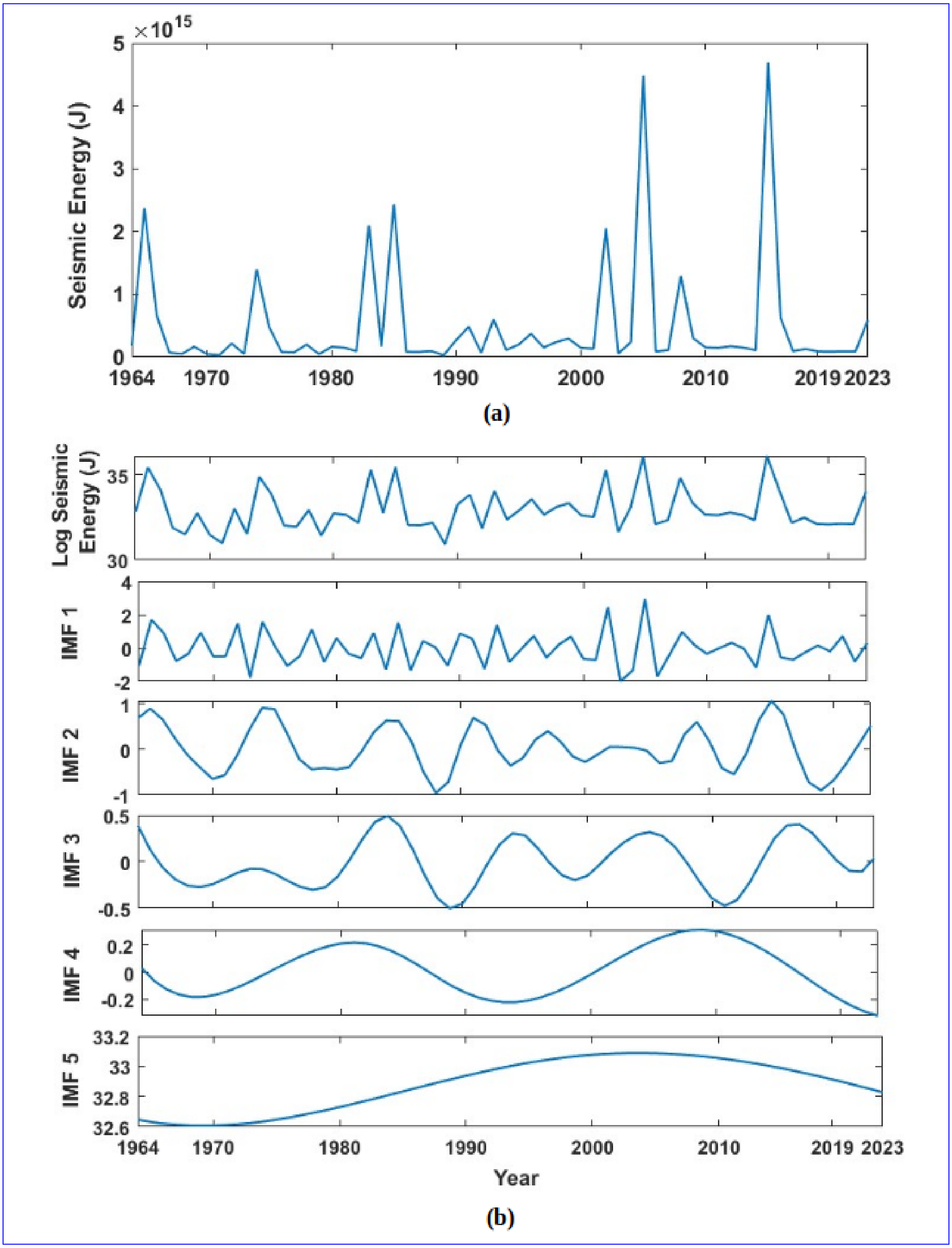

Figure 15. (a) Annual Seismic energy estimated for the catalogue compiled for western Himalayas (b) Log scaled seismic energy for western Himalayas and the corresponding intrinsic mode function obtained by employing ensemble empirical mode decomposition (EEMD)

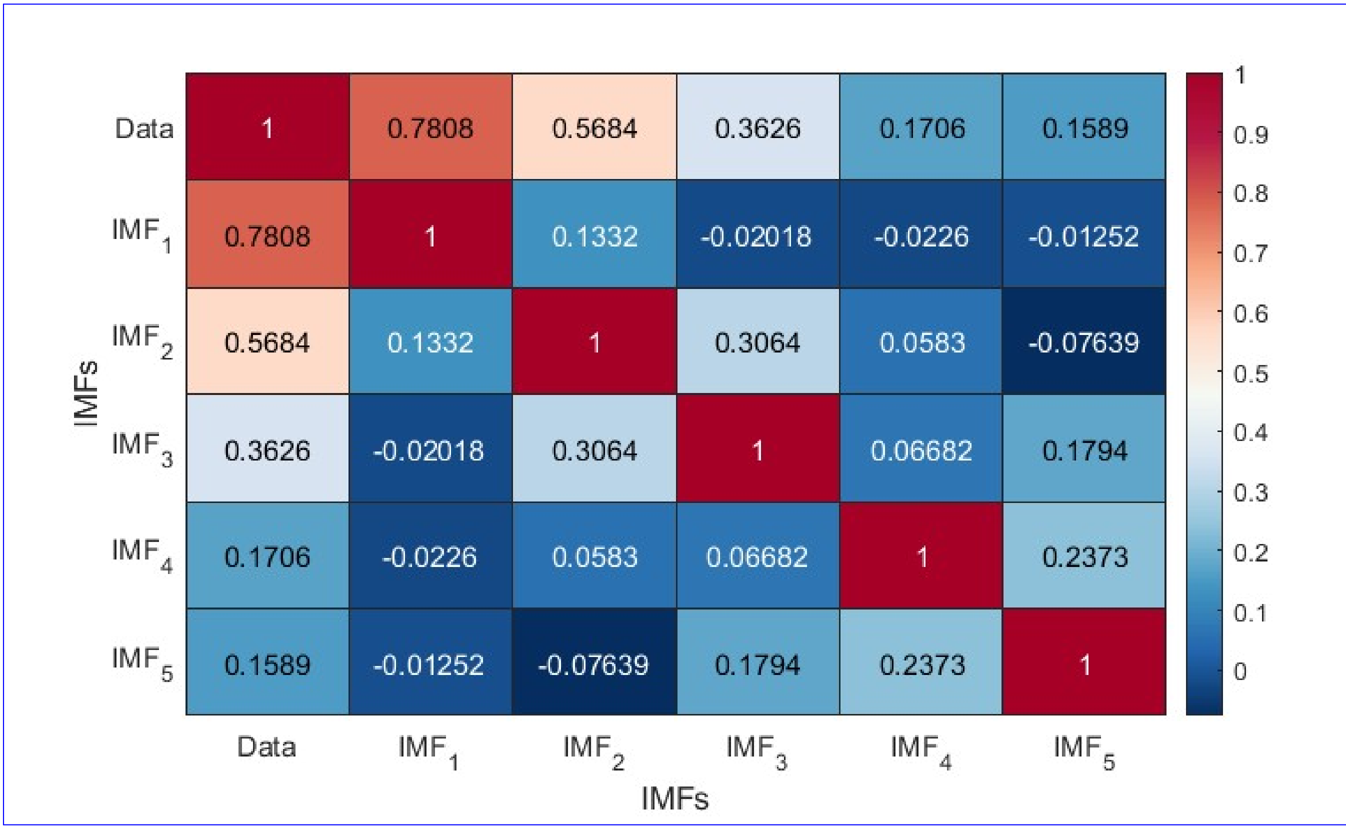

Figure 16. The correlation coefficient estimated for the intrinsic mode functions extracted from log scaled seismic energy time series of Western Himalayas

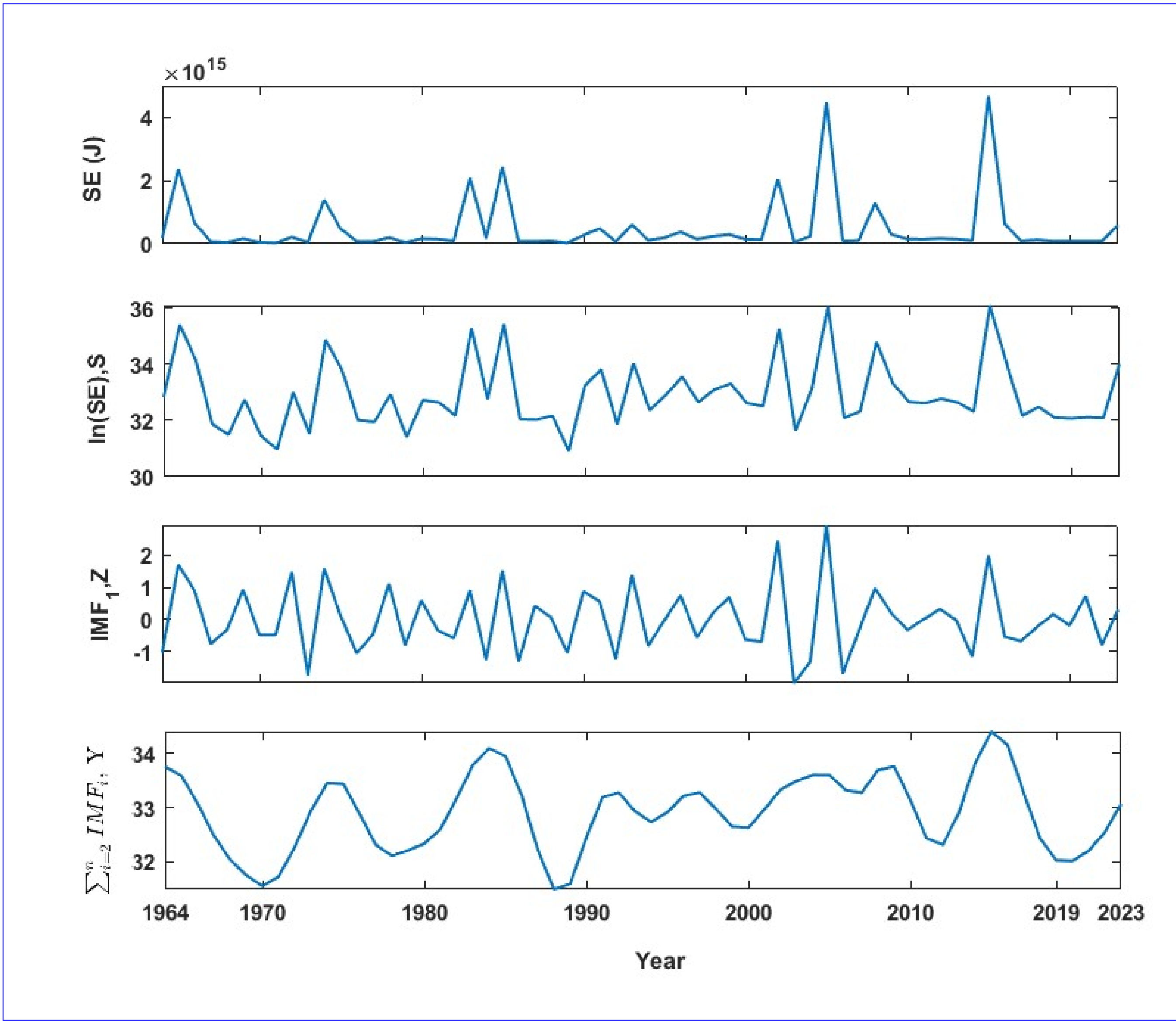

Figure 17. Estimated seismic energy series from updated catalogue for Western Himalayan region used in developing the models (a) Global seismic energy (GSE) time series having units as joules (b) log scaled GSE (c) First intrinsic mode estimated from ln(GSE) by performing ensemble empirical mode decomposition (EEMD) (d) Sum of second to last intrinsic modes estimated from ln(GSE) by performing EEMD

6.3 Model architecture for Western Himalayas

After input preparation, the approaches discussed under Section 3 are also tested for the active Western Himalayan region, i.e., obtained IMFs along with the log seismic energy and year of occurrence are taken as input for the first-level individual machine learning techniques (MLP, LR, RF, SMOreg, and IBk) and the forecasted results of these techniques as an input to the final ensemble random forest technique (Figure ~~4~~6). For lag consideration look back period is varied from 1-15 for the individual models and the value for which results are optimum is selected. Other ~~hyper parameteres~~ hyperparameters were also suitably iterated to find the best model in each individual architecture for the data under consideration. The lag value for individual models, along with the other hyper-parameters used to optimise the model predictions for the regional dataset corresponding to various techniques, are present in Table 3. For the final ensemble RF model 100 trees were considered along with the depth of the tree as 0 and both bag and batch size also as 100. Furthermore, for the ensemble model, the overall lag value of 8 is adopted. For base learners, the time series data is divided into 80 to 20 % for training and testing, respectively, i.e., time series up to 2011 is used to train the model, and from 2012 to 2023 is used to test the model. The division into training (up to 2011) and testing (2012–2023) was performed to preserve the time-dependence of the sequence and to evaluate model generalization.

6.4 Results for Western Himalaya

The representation of the model results and the comparison with the data is illustrated in Figure ~~14~~ 18 for the individual model and Figure ~~15~~ 19 for the ensemble model. Similar to that observed for global data, we observed varied performance in the training and testing phases. Furthermore, the qualitative performance seemed to improve by adopting the ensemble model. ~~Furthermore, to qualitatively evaluate the model performances, the statistical parameters like standard deviation of error ($\sigma(\epsilon)$),~~ Standard statistical metrics such as the Pearson correlation coefficient (~~R~~R), Performance Parameter (~~PP), and~~ PP), Root Mean Squared Error (~~$RMSE$)values are obtained for~~ RMSE), and standard deviation of error ($\sigma(\epsilon)$) were employed to objectively assess model performance. Table 5 displays these values for the Western Himalayan model~~are presented in Table 5.~~ . In an ideal case, the value of (R) and (PP) should be unity, and the value of ($\sigma(\epsilon)$) should be zero. ~~From Table 5, it is quite clear that Multilayer Perceptron (MLP) has performed best among the individual models or weak learners with Performance parameters (PP)~~ With an R-value of 0.989 and 0.848, respectively, and a PP value of 0.968 in ~~the~~ training and 0.685 in ~~the testingphase.~~ testing, Table 5 clearly shows that the Multilayer Perceptron (MLP) outperformed the other models. Also, an R-value of 0.989 in training and 0.848 in testing. The same was evident from Figure 14, where for both the training and testing phases, the MLP model is observed to capture the data variations better than other architectures. Furthermore, when the prediction made from the individual techniques is employed as input for the ensembled random forest model, its performance increases significantly with RMSE values of 0.117 and 0.236 in the training and testing phases, respectively. It shows the satisfactory performance of the model, ensuring a reliable prediction. This improvement is also evident from the ensemble RF model presented in Figure 15, where the model is able to capture the peaks and troughs efficiently.

~~Furthermore, we attempted to forecast the expected annual seismic~~ Additionally, we made an effort to predict the anticipated yearly seismic energy release for ~~the upcoming year (i.e. 2024). Based on the model developed from this work, we can expect a total annual seismic energy in the range $9.11 \times 10^{14} J$ to~~ 2024. According to the ensemble model that was built, the estimated seismic energy falls between $5.69 \times 10^{14} J$ ~~Whereby, we can expect a maximum magnitude of~~ and $9.11 \times 10^{14} J$. This range highlights the region's potential for a destructive event and corresponds to a maximum predicted magnitude of roughly 7.17 M_w~~for the Western Himalayan region~~. Even with these encouraging outcomes, the performance of the existing model might be enhanced by adding more geophysical variables, increasing the catalog inputs geographic resolution, and adding real-time updates for operational forecasts. These are potential directions for future growth.

Table 5. The performance evaluation of individual models and the final ensemble random forest model at the training and testing phases for Western Himalayan region

Models	Training				Testing			
	$\sigma(\epsilon)$	R	PP	$RMSE$	$\sigma(\epsilon)$	R	PP	$RMSE$
MLP*	0.180	0.989	0.968	0.221	0.659	0.848	0.685	0.687
Linear Regression (LRRidge Regression (RR)	0.356	0.957	0.918	0.352	0.677	0.833	0.544	0.826
Random Forest (RF)	0.450	0.978	0.866	0.445	1.023	0.673	0.348	0.988
SMOreg**	0.475	0.924	0.835	0.492	0.574	0.903	0.688	0.684
IBk***	1.088	0.447	0.164	1.090	1.133	0.425	0.180	1.108
Ensemble RF	**0.114**	**0.996**	**0.990**	**0.117**	**0.235**	**0.991**	**0.963**	**0.236**

* MLP- Multi-Layer Perceptron; **SMOreg - Sequential Minimal Optimization regression;

***IBk - Instance-Bases learning with parameter k

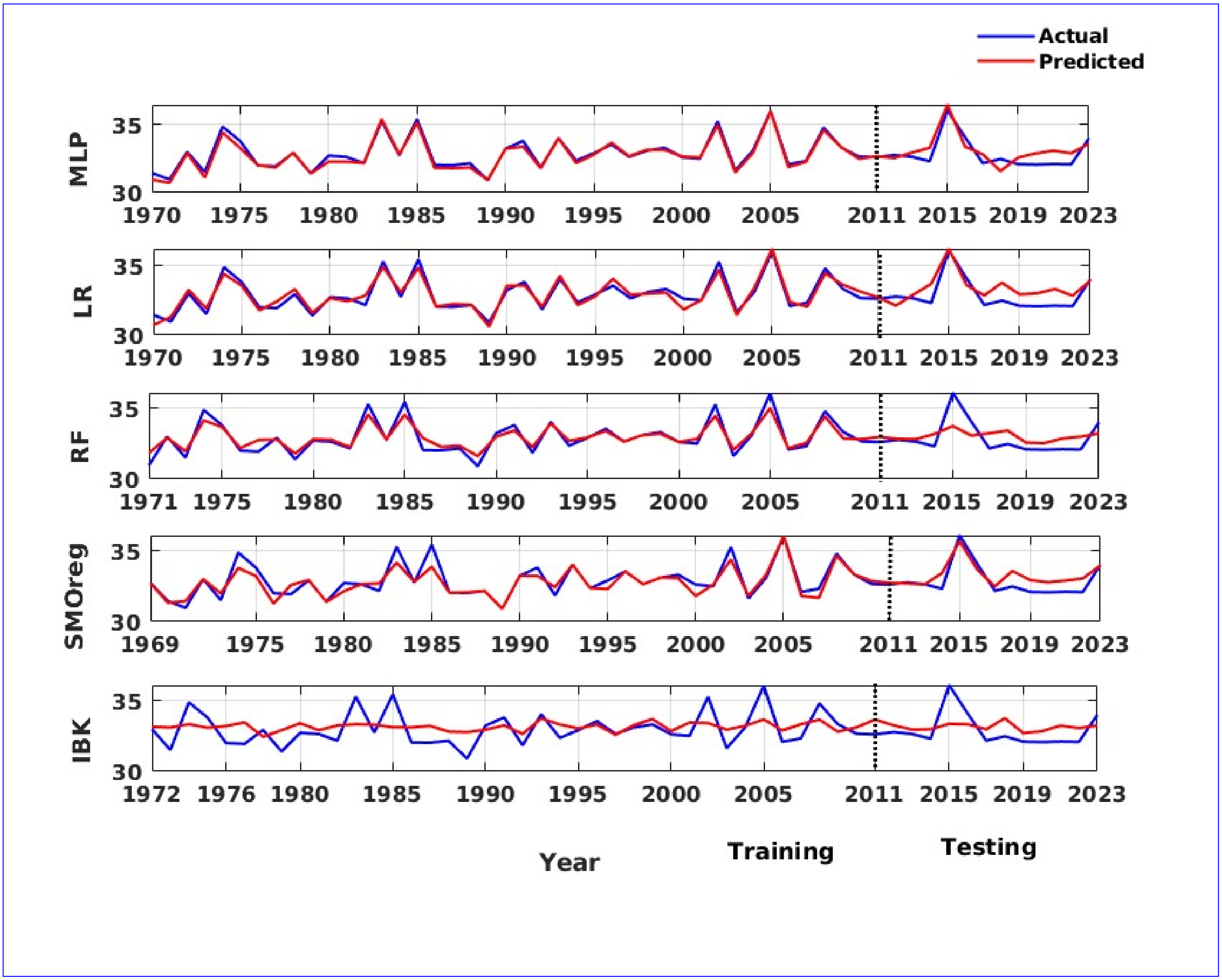

Figure 18. Actual vs Predicted values of log seismic energy from various individual techniques adopted for Western Himalayan region

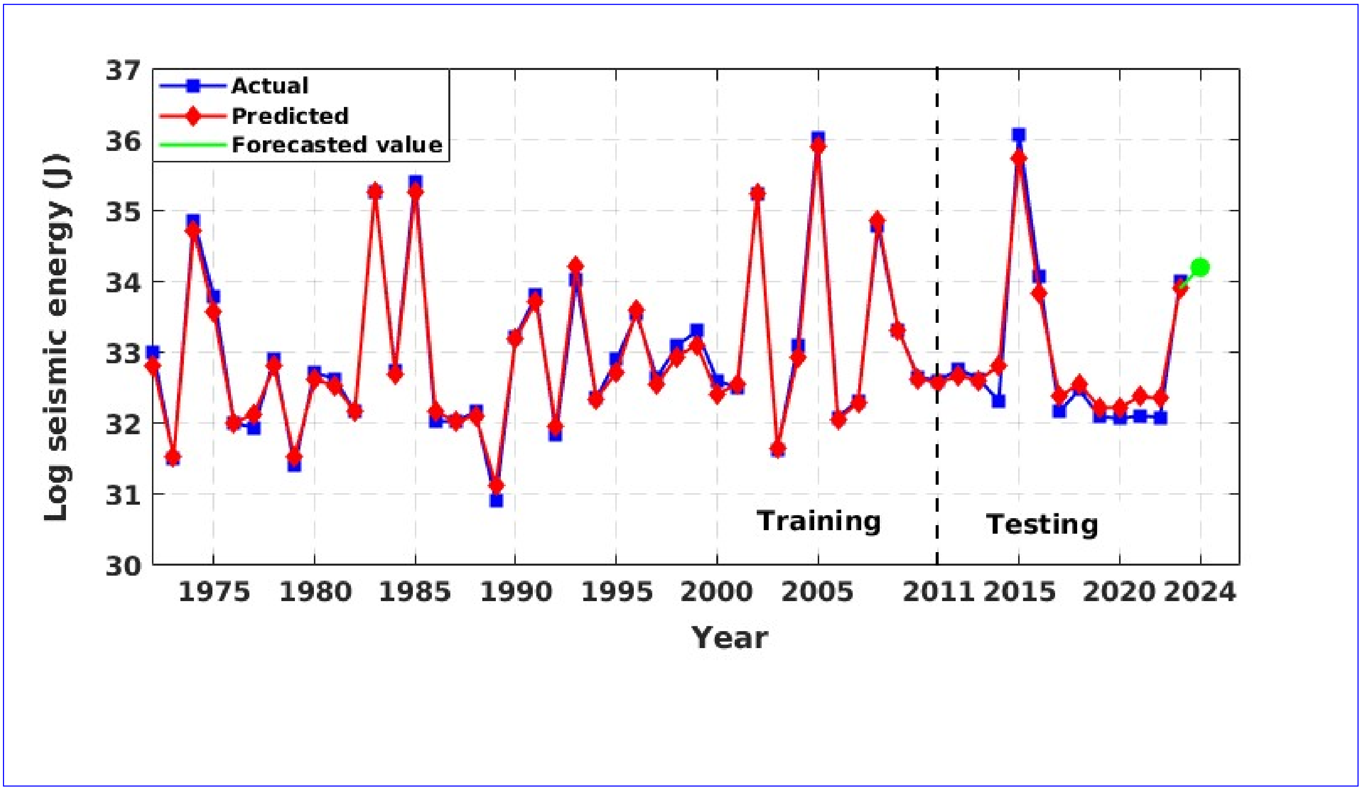

Figure 19. Actual vs Predicted values of seismic energy for Western Himalayan region from developed ensembled random forest model. Also the green marker shows the forecasted value for year 2024.

7 Discussion

This work investigated the application of sophisticated machine learning (ML) algorithms for seismic energy predictions. The proposed work is significant in quantifying the immediate hazard for the region. It is well known that seismic energy is a potential indicator of seismic activity in the region. Thus, a reliable forecast through robust algorithms shall aid in the enhancement of hazard preparedness. Thus, a thorough modelling using five separate architectures representative of different structures The research systematically investigates this hypothesis by creating and comparing various machine learning models to identify their potential in seismic prediction. Therefore, an exhaustive modelling with five different architectures reflective of various machineries in machine learning is attempted first . first attempted. The model approaches considered are Multilayer Perceptron (MLP), Linear Regression (LRRidge Regression (RR), Random Forest (RF), Sequential Minimal Optimization for Regression (SMOreg), and k-nearest Neighbors (IBk). Examining the separate models, it was found that MLP consistently performed better than the others during the training and testing stages. The strong performance of MLP indicates that it might be a good fit for encapsulating intricate connections in seismic data. The model predictions from different architectures were observed to vary at the training and testing phases. Thus, the different ways in which these models function highlight how crucial it is to choose an algorithm that is suitable for the features of the seismic data. To improve the robustness of the prediction, the separate models (weak models) were combined to create an ensemble RF model. The outcomes showed that the performance of the ensemble model outperformed that of any single model, highlighting the possible advantages of mixing several modelling techniques. Furthermore, when comparing the ensemble model to Raghukanth et al. (2017) model, the variance is lower, which suggests that the seismic energy forecasts are more stable and reliable. This research used train-test validation to ensure impartiality in model choice and prevent overfitting through hyperparameter optimization using the training data. The method delivers an unbiased measure of generalization performance since it protects against any test exposure during model tuning. Our ensemble method is more consistent and has less error compared to the baseline model of Raghukanth et al. (2017) with the utilization of the same validation catalog. This contributes to the evidence for the model's robust generalization and excellent performance even in the absence of a third separate independent validation set. To further enhance resilience and evaluate learning frameworks under robust conditions, a separate validation dataset can be included in future work. Even though the study used a worldwide time series that covered the years 1900 to 2015, it is important to recognize any potential limitations related to this temporal scope. It's possible that patterns of seismic activity change and that some recent occurrences go unrecorded. In order to overcome this constraint and maintain the predictive accuracy and relevance of results, future research should train models on an updated catalog. Updated research (e.g., Sharma et al. (2023a); Kumar et al. (2023b)) has established the advantage of using recent datasets in enhancing model generability. The study's encouraging findings provide opportunities for more investigation. Thus, as a sample study, the regional-level forecast model is developed for the Western Himalayan region. As similar to the global model, the regional data performance in forecast improved while adopting an ensemble architecture. Even though the results are promising the analysis is done on a larger cluster combining 4 seismogenic zones in the region. Such pooling, although streamlining model construction, can veil localised spatiotemporal features of prime importance to accurate hazard estimation. A more detailed physics-based clustering and further application to forecast modelling is expected to provide more insights into the spatiotemporal patterns of seismic activity. A more sophisticated knowledge of seismic energy trends may be obtained by extending the research to a more recent catalog and carrying out extensive regional-level investigations on regular basis. Furthermore, the present study acknowledges that Ensemble Empirical Mode Decomposition (EEMD) is adept at addressing non-stationarity in seismic energy time series; however, it also presents challenges, including edge effects, residual noise due to incomplete ensemble averaging, and constraints in accurately representing end-point behaviour. These limitations can affect the signal accuracy and can ultimately affect the performance of the seismic energy forecasts. Hence, in order to improve the forecasting capabilities and to address present limitations the future studies may explore more advanced time series decomposition techniques like wavelet-based denoising, Complete Ensemble EMD with Adaptive Noise (CEEMDAN), or hybrid filtering methods that enhance signal structure preservation and minimise noise interference. A further direction of potential is uncertainty-conscious modelling, where the confidence interval of predictions is explicitly defined to enable risk-based decision-making. Furthermore, combining various seismogenic zones into larger clusters, although beneficial for this initial analysis, could potentially mask specific localised spatiotemporal seismic patterns. Therefore, upcoming models ought to integrate physics-informed regional clustering to improve spatial resolution. From a machine learning perspective, the ensemble random forest model showed enhanced performance compared to individual models. Furthermore, the potential to improve the accuracy of the forecast model through a rigorous feature selection approach can also be adopted. These shall be taken as the future scope of this work. Additionally, there are intriguing prospects to improve forecast accuracy further and capture complex patterns in seismic data by exploring more sophisticated and hybrid machine learning techniques like Deep Learning, Extreme Learning Machine (ELM), and Generative Adversarial Networks (GANs). The ~~benefits of using sophisticated machine learning techniques highlight the potential real-world uses for seismic energy prediction. The dependability of~~ use of explainable ML approaches (e.g., SHAP or LIME) will further improve model interpretability—a crucial step towards establishing trust in model outputs for stakeholders concerned with policy and hazard response. The advancements in time series preprocessing, spatial modelling, and predictive algorithms are anticipated to significantly improve the accuracy, robustness, and practical application of seismic energy forecasting models. This, in turn, will facilitate more effective early warning systems ~~is increased by increased forecast accuracy and decreased variability, which helps improve preparedness and mitigation tactics in seismically active areas~~and strategies for hazard mitigation.

8 Conclusions

~~The adoption of a suitable ensemble model is observed to significantly improve~~ It has been found that using an appropriate ensemble model greatly increases the model's ~~accuracy towards forecasting seismic energy. Based on the confidence gained from testing with reported data and approach, the proposed approach was further employed to forecast~~ predicting accuracy. seismic energy, especially when handling the non-linearity and inherent complexity of geophysical data. The suggested method was then used to predict seismic energy for the western Himalayas ~~. Based on the developed model, we can expect a~~ based on the assurance provided by testing with published data and methodology. The outcomes demonstrate the model's applicability in situations involving regional hazards. According to the proposed model, the total annual seismic energy in ~~the range~~

~~$9.11 \times 10^{14} J$ to~~ 2024 is expected to be between $9.11 \times 10^{14} J$ and $5.69 \times 10^{14} J$, ~~which is equivalent to~~ or a magnitude range of 7.03~~-7.17 in 2024. Thus we can expect~~ –7.17. Therefore, we may anticipate that the Western Himalayan region would see a maximum magnitude of 7.17 M_w~~for the Western Himalayan region~~. This study ~~is a pilot study in the direction of~~ serves as a prototype project aimed at forecasting seismic energy release ~~and forecast for~~ in the Himalayan region. ~~A more detailed spatio-temporal investigation of the seismic energy patterns shall be taken up as the future scope~~ With the appropriate adjustments, the framework developed here can be expanded or modified for use in other tectonically active areas. The future focus of this work ~~. Nevertheless, the study results and formulations are critical in~~ will be on seismic energy patterns. However, the findings and recommendations of the study are essential for developing appropriate policy formulations, hazard preparedness, ~~immediate risk assessmentand suitable policy formulations.~~ urgent risk assessment, and seismic resilience specific to a given location

Data availability. All the data is with corresponding authors and will be available on reasonable request.

Author contributions. **Sukh Sagar Shukla:** Conceptualization, Data curation, Methodology, Validation, Visualization, Writing - original draft. **J Dhanya:** Funding acquisition, Conceptualization, Writing - original draft, Project administration, Supervision. **Priyanka:** Data curation, Formal analysis, Software. **Praveen kumar:** Data curation, Formal analysis, Software. **Varun Dutt:** Conceptualization, Resources, Software, Supervision.

Competing interests. 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.

Funding

The authors would like to acknowledge seed grand at IIT Mandi to support this research under the project titled “Earthquake forecast and prediction model for Himalayas using machine learning approaches” with project number: IITM/SG/DJ/98.

References

Adeli, H. and Panakkat, A. (2009). A probabilistic neural network for earthquake magnitude prediction. *Neural networks*, 22(7):1018–1024.

Ader, T., Avouac, J.-P., Liu-Zeng, J., Lyon-Caen, H., Bollinger, L., Galetzka, J., Genrich, J., Thomas, M., Chanard, K., Sapkota, S. N., et al. (2012). Convergence rate across the nepal himalaya and interseismic coupling on the main himalayan thrust: Implications for seismic hazard. *Journal of Geophysical Research: Solid Earth*, 117(B4).

Agarwal, N., Arora, I., Saini, H., and Sharma, U. (2023). A novel approach for earthquake prediction using random forest and neural networks. *EAI Endorsed Transactions on Energy Web*, 10.

Aha, D. W., Kibler, D., and Albert, M. K. (1991). Instance-based learning algorithms. *Machine learning*, 6:37–66.

Ahmed, F., Akter, S., Rahman, S. M., Harez, J. B., Mubasira, A., and Khan, R. (2024). Earthquake magnitude prediction using machine learning techniques. In *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, volume 2, pages 1–5. IEEE.

Al Banna, M. H., Taher, K. A., Kaiser, M. S., Mahmud, M., Rahman, M. S., Hosen, A. S., and Cho, G. H. (2020). Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges. *IEEE Access*, 8:192880–192923.

Alavi, A. H. and Gandomi, A. H. (2011). Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. *Computers & Structures*, 89(23-24):2176–2194.

Alpaydin, E. (2007). Combining pattern classifiers: Methods and algorithms (kuncheva, l.i.; 2004) [book review]. *IEEE Transactions on Neural Networks*, 18(3):964–964.

Altay, G., Kayadelen, C., and Kara, M. (2023). Model selection for prediction of strong ground motion peaks in türkiye. *Natural Hazards*, pages 1–19.

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

Asim, K. M., Idris, A., Iqbal, T., and Martínez-Álvarez, F. (2018). Earthquake prediction model using support vector regressor and hybrid neural networks. *PloS one*, 13(7):e0199004.

Baker, J., Bradley, B., and Stafford, P. (2021). *Seismic hazard and risk analysis*. Cambridge University Press.

Banerjee, P. and Bürgmann, R. (2002). Convergence across the northwest himalaya from gps measurements. *Geophysical Research Letters*, 29(13):30–1.

Bendick, R. and Bilham, R. (2001). How perfect is the Himalayan arc? *Geology*, 29(9):791–794.

Bertolini, M., Mezzogori, D., Neroni, M., and Zammori, F. (2021). Machine learning for industrial applications: A comprehensive literature review. *Expert Systems with Applications*, 175:114820.

Bhatia, M., Ahanger, T. A., and Manocha, A. (2023). Artificial intelligence based real-time earthquake prediction. *Engineering Applications of Artificial Intelligence*, 120:105856.

Bhatia, S., Kumar, M., and Gupta, H. (1999). A probabilistic seismic hazard map of india and adjoining regions. *Annals of Geophysics*, 42.

Bhattacharya, A., Vöge, M., Arora, M. K., Sharma, M. L., and Bhasin, R. K. (2013). Surface displacement estimation using multi-temporal sar interferometry in a seismically active region of the himalaya. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 7(3):184–197.

Bilham, R. (2019). Himalayan earthquakes: a review of historical seismicity and early 21st century slip potential. *Geological Society, London, Special Publications*, 483(1):423–482.

Bilham, R. and Ambraseys, N. (2005). Apparent himalayan slip deficit from the summation of seismic moments for himalayan earthquakes, 1500–2000. *Current science*, pages 1658–1663.

Bilham, R., Larson, K., and Freymueller, J. (1997). Gps measurements of present-day convergence across the nepal himalaya. *Nature*, 386(6620):61–64.

Bishop, C. M. (2016). *Pattern Recognition and Machine Learning*. Springer New York.

Bose, S., Das, K., and Arima, M. (2008). Multiple stages of melting and melt-solid interaction in the lower crust: new evidence from uht granulites of eastern ghats belt, india. *Journal of Mineralogical and Petrological Sciences*, 103(4):266–272.

Breiman, L. (2001a). Random forest, vol. 45. *Mach Learn*, 1.

Breiman, L. (2001b). Random forests. *Machine Learning*, 45:5–32.

Brownlee, J. (2023). Linear regression for machine learning. https://machinelearningmastery.com/linear-regression-for-machine-learning/. Online; accessed 20-February-2024.

Cho, Y., Khosravikia, F., and Rathje, E. (2022). A comparison of artificial neural network and classical regression models for earthquake-induced slope displacements. *Soil Dynamics and Earthquake Engineering*, 152:107024.

Choy, G. L. and Boatwright, J. L. (1995). Global patterns of radiated seismic energy and apparent stress. *Journal of geophysical research: Solid earth*, 100(B9):18205–18228.

Cutler, A., Cutler, D. R., and Stevens, J. R. (2012). Random forests. *Ensemble machine learning: Methods and applications*, pages 157–175.

De Gooijer, J. G. and Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, 22(3):443–473.

Derras, B., Bard, P. Y., and Cotton, F. (2014). Towards fully data driven ground-motion prediction models for europe. *Bulletin of Earthquake Engineering*, 12(1):495–516.

Dhanya, J. and Raghukanth, S. T. G. (2018). Ground motion prediction model using artificial neural network. *Pure and Applied Geophysics*, 175:1035–1064.

Dhanya, J. and Raghukanth, S. T. G. (2020). Neural network-based hybrid ground motion prediction equations for western himalayas and north-eastern india. *Acta Geophysica*, 68:303–324.

Dhanya, J. and Raghukanth, S. T. G. (2022). Probabilistic fling hazard map of india and adjoined regions. *Journal of Earthquake Engineering*, 26(9):4712–4736.

Dhanya, J., Sreejaya, K. P., and Raghukanth, S. T. G. (2022). Seismic recurrence parameters for india and adjoined regions. *Journal of Seismology*, 26:1–25.

Dietterich, T. (2000a). Ensemble methods in machine learning. pages 1–15.

Dietterich, T. (2000b). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.*, 40.

Douglas, J. (2021). Ground motion prediction equations 1964–2021. *Department of Civil and Environmental Engineering University of Strathclyde, Glasgow, United Kingdom.*

Duin, R. P. W. and Tax, D. M. J. (2000). Experiments with classifier combining rules. In *Multiple Classifier Systems*, pages 16–29, Berlin, Heidelberg. Springer Berlin Heidelberg.

Gade, M., Nayek, P. S., and Dhanya, J. (2021). A new neural network–based prediction model for newmark's sliding displacements. *Bulletin of Engineering Geology and the Environment*, 80:385–397.

Gastinger, J., Nicolas, S., Stepić, D., Schmidt, M., and Schülke, A. (2021). A study on ensemble learning for time series forecasting and the need for meta-learning. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Geological Survey of India (GSI) (2000). Seismotectonic Atlas of India and its Environs.

Ghaedi, K. and Ibrahim, Z. (2017). Earthquake prediction. *Earthquakes-Tectonics, Hazard and Risk Mitigation*, 66:205–227.

Hanks, T. C. and Kanamori, H. (1979). A moment magnitude scale. *Journal of Geophysical Research: Solid Earth*, 84(B5):2348–2350.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971):903–995.

Hyndman, R. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts, Australia, 2nd edition.

IMD (2023). Indian meteorological department. https://riseq.seismo.gov.in/riseq/earthquake. Online; accessed 01-January-2024.

IS:1893-1 (2016). Criteria for Earthquake Resistant Design of Structures, Part 1: General Provisions and Buildings.

Ismail-Zadeh, A., Le Mouël, J.-L., Soloviev, A., Tapponnier, P., and Vorovieva, I. (2007). Numerical modeling of crustal block-and-fault dynamics, earthquakes and slip rates in the tibet-himalayan region. *Earth and Planetary Science Letters*, 258(3-4):465–485.

Iyengar, R. N. and Raghukanth, S. T. G. (2005). Intrinsic mode functions and a strategy for forecasting indian monsoon rainfall. *Meteorology and Atmospheric Physics*, 90(1):17–36.

Jain, S. K. (2016). Earthquake safety in india: achievements, challenges and opportunities. *Bulletin of Earthquake Engineering*, 14:1337–1436.

Jaumé, S. C. and Sykes, L. R. (1999). Evolving Towards a Critical Point: A Review of Accelerating Seismic Moment/Energy Release Prior to Large and Great Earthquakes. *Pure and Applied Geophysics*, 155(2-4):279–305.

Jayalakshmi, S. and Raghukanth, S. T. G. (2017). Finite element models to represent seismic activity of the indian plate. *Geoscience Frontiers*, 8(1):81–91.

Jo, T. and Jo, T. (2021). Instance based learning. *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*, pages 93–115.

Joshi, A., Vishnu, C., and Mohan, C. K. (2022). Early detection of earthquake magnitude based on stacked ensemble model. *Journal of Asian Earth Sciences: X*, 8:100122.

Kaushik, S., Choudhury, A., Sheron, P. K., Dasgupta, N., Natarajan, S., Pickett, L. A., and Dutt, V. (2020). Ai in healthcare: time-series forecasting using statistical, neural, and ensemble architectures. *Frontiers in big data*, 3:4.

Kavitha, B. and Raghukanth, S. T. G. (2016). Regional level forecasting of seismic energy release. *Acta Geodaetica et Geophysica*, 51:359–391.

Khattri, K. (1999). Probabilities of occurrence of great earthquakes in the himalaya. *Proceedings of the Indian Academy of Sciences-Earth and Planetary Sciences*, 108:87–92.

Khattri, K., Rogers, A., Perkins, D., and Algermissen, S. (1984). A seismic hazard map of india and adjacent areas. *Tectonophysics*, 108(1):93–134.

Kramer, S. L. (1996). *Geotechnical earthquake engineering*. Pearson Education India.

Kumar, P., Malik, J. N., Gahalaut, V. K., Yadav, R. K., and Singh, G. (2023a). Evidence of strain accumulation and coupling variation in the himachal region of nw himalaya from short term geodetic measurements. *Tectonics*, 42(8):e2022TC007690.

Kumar, P., Priyanka, P., Dhanya, J., Uday, K. V., and Dutt, V. (2023b). Analyzing the performance of univariate and multivariate machine learning models in soil movement prediction: A comparative study. *IEEE Access*.

Lavé, J., Yule, D., Sapkota, S., Basant, K., Madden, C., Attal, M., and Pandey, R. (2005). Evidence for a great medieval earthquake (˜ 1100 ad) in the central himalayas, nepal. *Science*, 307(5713):1302–1305.

Li, Y. and Goda, K. (2023). Risk-based tsunami early warning using random forest. *Computers & Geosciences*, 179:105423.

Liritzis, I. and Tsapanos, T. M. (1993). Probable evidence for periodicities in global seismic energy release. *Earth, Moon, and Planets*, 60:93–108.

Maaliw III, R., Ballera, M., Mabunga, Z., Mahusay, A., Dejelo, D., and Seño, M. (2021). An ensemble machine learning approach for time series forecasting of covid-19 cases. pages 0633–0640.

Mignan, A. and Broccardo, M. (2020). Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations. *Seismological Research Letters*, 91(4):2330–2342.

Misra, A., Agarwal, K., Kothyari, G. C., Talukdar, R., and Joshi, G. (2020). Quantitative geomorphic approach for identifying active deformation in the foreland region of central indo-nepal himalaya. *Geotectonics*, 54:543–562.

Mousavi, S. M. and Beroza, G. C. (2018). Evaluating the 2016 one-year seismic hazard model for the central and eastern united states using instrumental ground-motion data. *Seismological Research Letters*, 89(3):1185–1196.

Mousavi, S. M. and Beroza, G. C. (2023). Machine learning in earthquake seismology. *Annual Review of Earth and Planetary Sciences*, 51:105–129.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., and Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, 11(1):3952.

Nakamichi, H., Iguchi, M., Triastuty, H., Hendrasto, M., and Mulyana, I. (2019). Differences of precursory seismic energy release for the 2007 effusive dome-forming and 2014 plinian eruptions at kelud volcano, indonesia. *Journal of Volcanology and Geothermal Research*, 382:68–80. Lessons learned from the recent eruptions of Sinabung and Kelud Volcanoes, Indonesia.

Narayanakumar, S. and Raja, K. (2016). A bp artificial neural network model for earthquake magnitude prediction in himalayas, india. *Circuits and Systems*, 7(11):3456–3468.

Nath, S. and Thingbaijam, K. (2012). Probabilistic seismic hazard assessment of india. *Seismological Research Letters*, 83(1):135–149.

NDMA (2010). Development of probabilistic seismic hazard map of india. *The National Disater Management Authority*, page 86.

Pairojn, P. and Wasinrat, S. (2015). Earthquake ground motions prediction in thailand by multiple linear regression model. *Electronic Journal of Geotechnical Engineering*, 20.25:12124.

Paolucci, R., Gatti, F., Infantino, M., Smerzini, C., Özcebe, A. G., and Stupazzini, M. (2018). Broadband ground motions from 3d physics-based numerical simulations using artificial neural networks. *Bulletin of the Seismological Society of America*, 108(3A):1272–1286.

Parvez, I. A., Vaccari, F., and Panza, G. F. (2003). A deterministic seismic hazard map of india and adjacent areas. *Geophysical Journal International*, 155(2):489–508.

Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines.

Pouriyeh, S., vahid, s., Sannino, G., De Pietro, G., Arabnia, H., and Gutiérrez, J. (2017). A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease.

Pyakurel, A., Dahal, B. K., and Gautam, D. (2023). Does machine learning adequately predict earthquake induced landslides? *Soil Dynamics and Earthquake Engineering*, 171:107994.

Quinlan, J. R. et al. (1992). Learning with continuous classes. In *5th Australian joint conference on artificial intelligence*, volume 92, pages 343–348. World Scientific.

Raghukanth, S. T. G., Kavitha, B., and Dhanya, J. (2017). Forecasting of global earthquake energy time series. *Advances in Data Science and Adaptive Analysis*, 9(04):1750008.

Rajendran, C., Rajendran, K., Sanwal, J., and Sandiford, M. (2013). Archeological and historical database on the medieval earthquakes of the central himalaya: Ambiguities and inferences. *Seismological Research Letters*, 84(6):1098–1108.

Re, M. and Valentini, G. (2012). Ensemble methods. *Advances in Machine Learning and Data Mining for Astronomy*, pages 563–593.

Reyes, J., Morales-Esteban, A., and Martínez-Álvarez, F. (2013). Neural networks to predict earthquakes in chile. *Applied Soft Computing*, 13(2):1314–1328.

Rezaei, H., Amjadian, A., Sebt, M., Askari, R., and Gharaei, A. (2022). An ensemble method of the machine learning to prognosticate the gastric cancer. *Annals of Operations Research*, 328.

Ridzwan, N. S. M. and Yusoff, S. H. M. (2023). Machine learning for earthquake prediction: a review (2017–2021). *Earth Science Informatics*, 16(2):1133–1149.

Saha, T. K., Pal, S., Talukdar, S., Debanshi, S., Khatun, R., Singha, P., and Mandal, I. (2021). How far spatial resolution affects the ensemble machine learning based flood susceptibility prediction in data sparse region. *Journal of Environmental Management*, 297:113344.

Salam, M. A., Ibrahim, L., and Abdelminaam, D. S. (2021). Earthquake prediction using hybrid machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 12(5):654–6652021.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(3):160.

Schmidt, J., Marques, M. R., Botti, S., and Marques, M. A. (2019). Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials*, 5(1):83.

Scordilis, E. (2006). Empirical global relations converting ms and mb to moment magnitude. *Journal of seismology*, 10:225–236.

Seo, H., Kim, J., and Kim, B. (2022). Machine-learning-based surface ground-motion prediction models for south korea with low-to-moderate seismicity. *Bulletin of the Seismological Society of America*, 112(3):1549–1564.

Sharma, V., Dhanya, J., Gade, M., and Sivasubramonian, J. (2023a). New generalized ann-based hybrid broadband response spectra generator using physics-based simulations. *Natural Hazards*, 116(2):1879–1901.

Sharma, Y., Pasari, S., Ching, K.-E., Verma, H., Kato, T., and Dikshit, O. (2023b). Interseismic slip rate and fault geometry along the northwest himalaya. *Geophysical Journal International*, 235(3):2694–2706.

Shevade, S., Keerthi, S., Bhattacharyya, C., and Murthy, K. (1999). Improvements to the smo algorithm for svm regression. In *IEEE Transactions on Neural Networks*.

Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., and Murthy, K. R. K. (2000). Improvements to the smo algorithm for svm regression. *IEEE transactions on neural networks*, 11(5):1188–1193.

Shimony, R., Gvirtzman, Z., and Tsesarsky, M. (2020). Seismic energy release from intra-basin sources along the dead sea transform and its influence on regional ground motions. *Bulletin of the Seismological Society of America*, 111.

Shishegaran, A., Taghavizade, H., Bigdeli, A., and Shishegaran, A. (2019). Predicting the earthquake magnitude along zagros fault using time series and ensemble model. *Journal of Soft Computing in Civil Engineering*, 3(4):67–77.

Spassiani, I. and Marzocchi, W. (2021). An energy-dependent earthquake moment–frequency distribution. *Bulletin of the Seismological Society of America*, 111(2):762–774.

Sreejaya, K. P., Raghukanth, S. T. G., Gupta, I. D., Murty, C. V. R., and Srinagesh, D. (2022). Seismic hazard map of india and neighbouring regions. *Soil Dynamics and Earthquake Engineering*, 163:107505.

Sreenath, V., Basu, J., and Raghukanth, S. T. G. (2024). Ground motion models for regions with limited data: Data-driven approach. *Earthquake Engineering & Structural Dynamics*.

Stepp, J. (1972). Analysis of completeness of the earthquake sample in the puget sound area and its effect on statistical estimates of earthquake hazard. In *Proc. of the 1st Int. Conf. on Microzonazion, Seattle*, volume 2, pages 897–910.

Stepp, J. (1973). Analysis of completeness of the earthquake sample in the puget sound area. *Contributions to Seismic Zoning: US National Oceanic and Atmospheric Administration Technical Report ERL*, pages 16–28.

Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S. M., Wang, J., Lin, C., Cristea, N., Tong, D., Carande, W. H., Ma, X., et al. (2022). A review of earth artificial intelligence. *Computers & Geosciences*, 159:105034.

Tan, M. L., Becker, J. S., Stock, K., Prasanna, R., Brown, A., Kenney, C., Cui, A., and Lambie, E. (2022). Understanding the social aspects of earthquake early warning: A literature review. *Frontiers in Communication*, 7:939242.

Tiampo, K. F. and Shcherbakov, R. (2012). Seismicity-based earthquake forecasting techniques: Ten years of progress. *Tectonophysics*, 522:89–121.

Tsapanos, T. (1998). Seismic hazard for some regions of the world examined using strain energy release.

Tsapanos, T. and Liritzis, I. (1992). Time-lag of the earthquake energy release between three seismic regions. *Pure and Applied Geophysics PAGEOPH*, 139:293–308.

USGS (2023). The united states geological survey. https://earthquake.usgs.gov/earthquakes/search/. Online; accessed 01-January-2024.

Varga, P., Krumm, F., Riguzzi, F., Doglioni, C., Sule, B., Wang, K., and Panza, G. F. (2012). Global pattern of earthquakes and seismic energy distributions: Insights for the mechanisms of plate tectonics. *Tectonophysics*, 530:80–86.

Wiemer, S. and Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from alaska, the western united states, and japan. *Bulletin of the Seismological Society of America*, 90(4):859–869.

Wu, Z. and Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460(2046):1597–1611.

Wu, Z. and Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1(01):1–41.

Xie, Y., Ebad Sichani, M., Padgett, J. E., and DesRoches, R. (2020). The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. *Earthquake Spectra*, 36(4):1769–1801.

Yenier, E., Erdoğan, O., and Akkar, S. (2008). Empirical relationships for magnitude and source-to-site distance conversions using recently compiled turkish strong-ground motion database. In *The 14th world conference on earthquake engineering*, pages 1–8.

Yokoyama, I. (1988). Seismic energy releases from volcanoes. *Bulletin of Volcanology*, 50:1–13.

Yousefzadeh, M., Hosseini, S. A., and Farnaghi, M. (2021). Spatiotemporally explicit earthquake prediction using deep neural network. *Soil Dynamics and Earthquake Engineering*, 144:106663.

Zarola, A. and Sil, A. (2018). Forecasting of future earthquakes in the northeast region of india considering energy released concept. *Computers & Geosciences*, 113:1–13.

Zhang, B., Hu, Z., Wu, P., Huang, H., and Xiang, J. (2023). Ept: A data-driven transformer model for earthquake prediction. *Engineering Applications of Artificial Intelligence*, 123:106176.