

## Research Paper

## Real-time earthquake magnitude prediction using designed machine learning ensemble trained on real and CTGAN generated synthetic data

Anushka Joshi <sup>a,\*</sup>, Balasubramanian Raman <sup>a</sup>, C. Krishna Mohan <sup>b</sup><sup>a</sup> Department of Computer Science and Engineering, Indian Institute of Technology, Roorkee, Uttarakhand, India<sup>b</sup> Department of Computer Science and Engineering, Indian Institute of Technology, Hyderabad, Telangana, India

## ARTICLE INFO

## Article history:

Received 2 July 2024

Received in revised form

1 October 2024

Accepted 16 October 2024

Available online 30 January 2025

## Keywords:

Magnitude  
Synthetic data  
Machine learning  
Earthquake  
CTGAN

## ABSTRACT

The earthquake early warning (EEW) system provides advance notice of potentially damaging ground shaking. In EEW, early estimation of magnitude is crucial for timely rescue operations. A set of thirty-four features is extracted using the primary wave earthquake precursor signal and site-specific information. In Japan's earthquake magnitude dataset, there is a chance of a high imbalance concerning the earthquakes above strong impact. This imbalance causes a high prediction error while training advanced machine learning or deep learning models. In this work, Conditional Tabular Generative Adversarial Networks (CTGAN), a deep machine learning tool, is utilized to learn the characteristics of the first arrival of earthquake P-waves and generate a synthetic dataset based on this information. The result obtained using actual and mixed (synthetic and actual) datasets will be used for training the stacked ensemble magnitude prediction model, MagPred, designed specifically for this study. There are 13295, 3989, and 1710 records designated for training, testing, and validation. The mean absolute error of the test dataset for single station magnitude detection using early three, four, and five seconds of P wave are 0.41, 0.40, and 0.38  $M_{JMA}$ . The study demonstrates that the Generative Adversarial Networks (GANs) can provide a good result for single-station magnitude prediction. The study can be effective where less seismic data is available. The study shows that the machine learning method yields better magnitude detection results compared with the several regression models. The multi-station magnitude prediction study has been conducted on prominent Osaka, Off Fukushima, and Kumamoto earthquakes. Furthermore, to validate the performance of the model, an inter-region study has been performed on the earthquakes of the India or Nepal region. The study demonstrates that GANs can discover effective magnitude estimation compared with non-GAN-based methods. This has a high potential for wide application in earthquake early warning systems.

© 2024 Editorial office of Geodesy and Geodynamics. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Earthquakes are the worst natural disasters due to their unpredictability and the widespread destruction caused by them. Earthquakes can additionally trigger other natural disasters such as tsunamis, fires, floods, and landslides that further exacerbate the impact of the earthquake. The deployment of earthquake early warning systems (EEWs) allows us to achieve early notification of approaching strong motion shaking in the vicinity of the epicenter

[1,2]. The EEWs system detects the magnitude of an earthquake based on the arrival of the first few seconds of the P wave at single or multiple stations [1,3–8]. The empirical relation between the magnitude of an earthquake and P wave onset parameters are used in the recent EEWs systems [9–13]. The application of machine learning and deep learning algorithms in EEWs and source parameter estimations for magnitude estimation has been found to be promising because of the recent studies conducted by Refs. [12–21]. These techniques require a good balanced dataset of higher magnitude events greater than 6.5  $M_{JMA}$ , which are very low in frequency. The Japan seismic events were measured using the  $M_{JMA}$  scale, the Moment Magnitude scale standardized by the Japan Meteorological Agency. The drawback of these techniques is that the data on higher-magnitude events is very limited.

\* Corresponding author.

E-mail address: [anushka\\_j@cs.iitr.ac.in](mailto:anushka_j@cs.iitr.ac.in) (A. Joshi).

Peer review under the responsibility of Institute of Seismology, China Earthquake Administration

Data augmentation is an approach in the field of Machine Learning (ML) and Deep Learning (DL) that is used to expand the original data [22]. The data augmentation method has resulted in outstanding achievements in the fields of environmental sound classification [23], text recognition, image recognition [24–26], biomedical image segmentation [27], etc. The data augmentation includes supervised and unsupervised data augmentation. The supervised data augmentation includes single-sample and multi-sample data augmentation [22]. The single sample includes the color transformation class and geometric operation class. The multi-sample supervised data augmentation includes the Synthetic Minority Oversampling Technique (SMOTE) [28], Pairing Samples [29], and mixup [30]. The unsupervised data augmentation method includes the Generative Adversarial Networks (GANs) [31] and AutoAugment [32]. GANs have several variants, such as InfoGAN [33], CycleGAN [34], and f-GAN [35]. GANs have been applied in several research areas, such as natural language processing [25], texture synthesis [36], image processing, and computer vision [37].

According to several research studies, GANs have proved to achieve success in many computer vision applications. The GANs can also be applied to tabular data generation. Several variants of GANs, such as Tabular GAN (TGAN) [38], CTGAN [24], and Medical Generative Adversarial Network (MedGAN) [39] can be used for tabular data synthesis. The conditional table GAN (CTAB-GAN) proposed by Zhao et al. [40] was used to model data types consisting of complex distribution. The MedGAN proposed by Choi et al. [39] has combined autoencoder and GANs. It was used to generate high-dimensional data. CTGAN, proposed by Chen et al. [24] is a conditional GAN for generating tabular data. Compared with other GAN variants for tabular datasets, CTGAN is used in the present work. The CTGAN utilizes a mode-specific normalization solution to represent information carried by continuous variables. In order to force the generator to produce samples from specific classes, conditional vectors are used in CTGAN [22]. Data augmentation is rarely used in magnitude detection problems of earthquake engineering. Therefore, CTGAN is used in this work to generate tabular synthetic data. This will provide leads for synthetic data generation for earthquake magnitude prediction.

Recently, Joshi et al. [14] has utilized a supervised ML model for EEWs. The model utilizes eight features extracted from the early P phase of the earthquake record that are applicable for predicting 3.0 to 7.4  $M_{JMA}$  earthquakes. In this study, the magnitude has been predicted by training a supervised ML model on actual and synthetic data generated by CTGANs. The ensemble of various machine learning models has been used for the study that includes Sanitized Grey Wolf Optimizer (SGWO) [41] as a meta-heuristic optimizer for hyperparameter tuning of XGBoost (XGB) [42]. The ensemble also includes standard implementations of Categorical Boosting (CatB) [43], Light Gradient Boosting Machine (LGBM) [44], and Random Forest (RF) [45]. The paper aims to investigate the learning capability of an ensemble model and the use of a synthetic dataset generated from CTGAN in regression analysis to estimate the magnitude of an earthquake.

The parameters extracted from the initial record after the onset of the P phase play an important role in predicting magnitude in EEWs. In this study, the three, four, and five seconds of the P wave window were selected for earthquake magnitude prediction. The duration of 3 s recorded after the arrival of the P phase for prediction of the magnitude of the earthquake has been used in several studies [1,4,8,46,47]. The numerical experiment for the moment-rate functions of Sato and Hirasawa [48] served as the basis for the choice of the minimum 3-s window of the P wave [4]. Additionally, Kanamori [46] demonstrated that  $\tau_c$  rises with a magnitude up to 7 (beyond which it saturates) if a length of 3 s is used [4]. This suggests that the features that correlate with the magnitude

before the completion of rupture can be found in the initial phases of the arrival of the P wave [49]. It has been found that the longer the duration, the more reliable the magnitude estimate is, but the warning time gets shorter [4].

In this study, thirty-four carefully chosen features were extracted from the first few seconds of the P-wave signal. There are 13295, 3989, and 1710 records designated for training, testing, and validation. The prediction concludes: 1) The data imbalance problem can be removed using a deep machine learning tool named CTGANs for synthetic data prediction in earthquake engineering. 2) The magnitude records above the strong level that are available in fewer numbers can be detected with better accuracy by training models using real and mixed datasets. 3) The magnitude predicted from the paper's pipeline gives better results than traditional magnitude prediction regression techniques. 4) Further, the comparison has been performed of the proposed model with the recent state-of-the-art method given by Zhu et al. [11], previously given regression relations for magnitude estimation, and established  $\tau_c$  and  $P_{GD}$  regression relation for this dataset. This comparison demonstrates the efficacy of the model. The source codes, along with the test dataset, are available for download at <https://github.com/anushka-joshi/MagPred>.

## 2. Method

The study utilizes the CTGAN for synthetic data generation and ensemble machine learning model for magnitude prediction. The entire framework of the proposed early warning system is shown in Fig. 1 from event triggering to single station magnitude prediction using early few seconds of P wave for warning dissemination. The current study predicts the magnitude in  $M_W$  units. Since, the dataset consists of Japan is in  $M_{JMA}$  unit, and converting it to  $M_W$  may introduce some error. Therefore,  $M_W$  has been used as the output metric of the model.

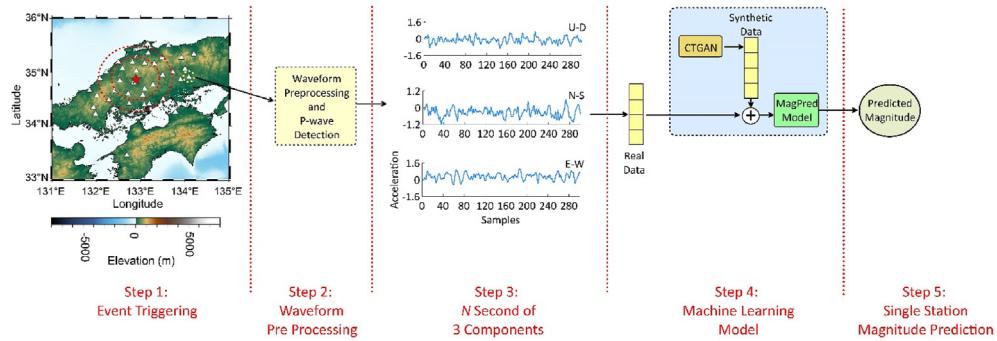
Obtaining the earthquake dataset is a very challenging task. The dataset of higher earthquakes is highly imbalanced, which means that the evaluation metrics will be biased toward the majority of the class. Therefore, adding the generated dataset using CTGAN in training set for higher events will reduce model error. In earthquake parameter prediction, data balancing is especially useful in predicting higher magnitudes, such as those above 6.5  $M_{JMA}$ , as it allows the model to be trained on a higher magnitude dataset that may contain fewer records. The magnitude prediction stage involves: one is generating a new synthetic dataset to add more synthetic records of input features for magnitude estimation. The other is training the model in that dataset.

### 2.1. Generative adversarial network (GANs)

GANs consist of a generator ( $G$ ) and a discriminator ( $D$ ). GANs tries to optimize the dual minimax game with the value function given by  $V(G, D)$  [31].  $G$  tries to match the real data distribution during fake data creation. The distinction between real and fake data is made using the discriminator  $D$ .  $D$  and  $G$  have adversarial goals in training. The real sample must be matched by  $D$  to maximize the probability (maximize  $\log D(x)$ ). The loss of  $D$  must be minimized by  $G$  as given by (minimum  $\log(1 - D(G(x)))$ ). The achievement of the Nash equilibrium [50] between  $G$  and  $D$  is the aim of the optimization process of GANs. The result of this process is to approximate the distribution of the real sample by  $G$ .

### 2.2. CTGAN

“Data imbalance” refers to a dataset scenario where the examples of different classes are not evenly distributed [28]. In this case



**Fig. 1.** Overall framework of the proposed work. The value of  $N$  is selected as three, four, and five seconds.

of earthquake magnitude prediction, this imbalance focus on the infrequency of higher magnitude records in training dataset. This imbalance implies that certain magnitude contain very high frequencies, potentially leading to bias in the model trained on this data. Specifically, the model may become skewed towards the majority magnitude values [51,52], neglecting the minority values, such as high earthquake magnitudes. Consequently, this imbalance can degrade the performance of ML models trained on such data.

CTGAN is employed as a solution to address this issue. It generates synthetic parameters representing earthquake records by analyzing the distribution patterns of parameters found in actual records. These are designed to mitigate the imbalance problem. Notably, the mixed dataset from CTGAN is balanced, containing an equal distribution of earthquake parameters of 500 records for each magnitude category. This balanced dataset helps alleviate the bias towards majority values, thus improving the performance of ML models trained on the data. This section elucidates the operational mechanics of the CTGAN model.

CTGAN follows mode-specific normalization to handle multi-modal distribution and non-Gaussian distribution. The columns of continuous values in CTGAN are displayed using mode-specific normalization. The mode-specific normalization takes the help of the variational Gaussian mixture model (VGM) [53] for each continuous column  $\mathbf{C}_i$ . The modes estimated for a column  $\mathbf{C}_i$  are represented by  $(\eta_1, \eta_2, \dots, \eta_k)$  where  $k$  is the number of modes given by  $m_i$ . The learned Gaussian mixture denoted by  $P_{\mathbf{C}_i}(c_{i,j})$  is given below:

$$P_{\mathbf{C}_i}(c_{i,j}) = \sum_{p=1}^{m_i} \mu_p \mathcal{N}(c_{i,j}; \eta_p; \theta_p) \quad (1)$$

In this equation,  $\mu_p$  and  $\theta_p$  are the weight and standard deviation of mode, respectively. The probability density denoted by  $\rho_p$  for each  $c_{i,j}$  is computed as follows:

$$\rho_p = \mu_p \mathcal{N}(c_{i,j}; \eta_p; \theta_p) \quad (2)$$

Then, the mode denoted by the integer  $\eta_s$  is sampled to normalize the value. There are two parameters obtained from the process denoted by  $\alpha_{i,j}$  and  $\beta_{i,j}$  as given in Eq. (3) and Eq. (4).

$$\alpha_{i,j} = \frac{c_{i,j} - \eta_s}{4\theta_s} \quad (3)$$

$$\beta_{i,j} = [0, 0, 0, \dots, 1] \quad (4)$$

In this equation,  $\beta$  is one hot encoded vector. Then, the learned row is represented by continuous and discrete columns, as shown below:

$$\mathbf{r}_j = \alpha_{1,j} \oplus \beta_{1,j} \oplus \dots \oplus \alpha_{N_c,j} \oplus \beta_{N_c,j} \oplus \mathbf{d}_{i,j} \oplus \dots \oplus \mathbf{d}_{N_d,j} \quad (5)$$

In this equation,  $d_{i,j}$  is the one hot representation of discrete values and  $\oplus$  means the concatenation operation. Further, the CTGAN deals with the problem of data imbalance in discrete columns by using generator loss, training by sampling, and conditional vector. The aim of a conditional generator is to reconstruct the original distribution as:

$$\mathbb{P}(\text{row}) = \sum_{k \in D_{i^*}} \mathbb{P}_G(\text{row}/\mathbf{D}_{i^*} = k^*) \mathbb{P}(\mathbf{D}_{i^*} = k) \quad (6)$$

In this equation,  $k^*$  is the value of the  $i^{\text{th}}$  discrete column as represented by  $\mathbf{D}_{i^*}$ . The vectors  $[\mathbf{d}_1, \dots, \mathbf{d}_{N_d}]$  is the one hot encoded representation of the discrete columns denoted by  $[\mathbf{D}_1, \dots, \mathbf{D}_{N_d}]$ , where  $N_d$  is the number of discrete columns. The  $\mathbf{d}_i = [\mathbf{d}_i^{(k)}]$  is the  $i^{\text{th}}$  one hot encoded vector with  $k = \{1, \dots, |D_i|\}$ . The  $\mathbf{m}_i = [\mathbf{m}_i^{(k)}]$ , for  $k = \{1, \dots, |D_i|\}$  is the mask associated with that vector denoted by:

$$\mathbf{m}_i^{(k)} = \begin{cases} 1 & \text{if } i = i^* \text{ and } k = k^* \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

The conditional vector is denoted by:

$$\mathbf{cond} = \mathbf{m}_1 \oplus \mathbf{m}_2 \oplus \dots \oplus \mathbf{m}_{N_d} \quad (8)$$

The conditional vector tries to penalize its loss by adding cross entropy between  $\mathbf{m}_{i^*}$  and  $\widehat{\mathbf{d}}_{i^*}$ , where  $\{\widehat{\mathbf{d}}_1, \dots, \widehat{\mathbf{d}}_{N_d}\}$  is the vector generated by the generator. The critic function finally evaluates the output. It tries to estimate the difference between the learned conditional distribution and conditional distribution on a real dataset.

### 2.3. Meta heuristic optimizer

The enhancement of model performance is achievable by adjusting hyperparameters [54] and fine-tuning these hyperparameters involves employing metaheuristic algorithms. Hyperparameters are distinct from internal weights or coefficients learned by the algorithm and play a crucial role in the model configuration. The XGB algorithm utilized in this work selects the optimized hyperparameter set using the meta heuristic optimization technique named Sanitized Grey Wolf Optimizer (SGWO). The SGWO algorithm has been utilized in this study. The SGWO is a variant of the Grey wolf optimizer (GWO) proposed by Mirjalili et al. [41]. The idea behind the GWO algorithm is taken from the social interaction and hierarchy of grey wolves. There are various members denoted by  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\omega$  in a pack of grey wolf. The domination power of the wolf in the pack decreases from  $\alpha$  to  $\omega$ .

Therefore, in this analysis, the first optimal hyperparameter set given by  $\alpha$  is selected.

The process of simulating the behavior of hunting in a computer is shown in Fig. 2 in the  $x$  and  $y$ -axis. The equation shows different positions that can be modelled from the location of the grey wolf and the location of the prey.

$$\mathbf{X}(t+1) = \mathbf{X}_{\text{prey}}(t) - \mathbf{A}\mathbf{P}_{\text{loc}} \quad (9)$$

In this equation,  $\mathbf{X}(t+1)$  and  $\mathbf{X}(t)$  are the next and current location of the wolf.  $\mathbf{X}_{\text{prey}}$ ,  $\mathbf{P}_{\text{loc}}$ , and  $\mathbf{A}$  are the location of the prey, vector depending on the prey and coefficient vector. The prey location vector  $\mathbf{P}_{\text{loc}}$  is calculated as follows:

$$\mathbf{P}_{\text{loc}} = |\mathbf{C}\mathbf{X}_{\text{prey}}(t) - \mathbf{X}(t)| \quad (10)$$

From Eq. (9) and Eq. (10), we get:

$$\mathbf{X}(t+1) = \mathbf{X}_{\text{prey}}(t) - \mathbf{A}|\mathbf{C}\mathbf{X}_{\text{prey}}(t) - \mathbf{X}(t)| \quad (11)$$

The random components of  $\mathbf{A}$  and  $\mathbf{C}$  are:

$$\mathbf{A} = 2\mathbf{a}\mathbf{r}_1 - \mathbf{a} \quad (12)$$

$$\mathbf{C} = 2\mathbf{r}_2 \quad (13)$$

In this equation,  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are random vectors in the interval  $[0, 1]$ . In the Eq. (12),  $\mathbf{a}$  is a vector that takes values from 2 to 0 and shows linear decrement as shown below:

$$\mathbf{a} = 2 - t \frac{2}{T} \quad (14)$$

In this equation,  $t$  and  $T$  are the current and maximum number of iterations. The position of each wolf is updated based on the defined hierarchy of wolves, the encircling equation and the position of the prey.

$$\mathbf{X}(t+1) = \frac{\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3}{3} \quad (15)$$

$$\mathbf{X}_1 = \mathbf{X}_\alpha(t) - \mathbf{A}_1 \mathbf{P}_\alpha \quad (16)$$

$$\mathbf{X}_2 = \mathbf{X}_\beta(t) - \mathbf{A}_2 \mathbf{P}_\beta \quad (17)$$

$$\mathbf{X}_3 = \mathbf{X}_\delta(t) - \mathbf{A}_3 \mathbf{P}_\delta \quad (18)$$

The equation  $\mathbf{P}_\alpha$ ,  $\mathbf{P}_\beta$ , and  $\mathbf{P}_\delta$  are as follows:

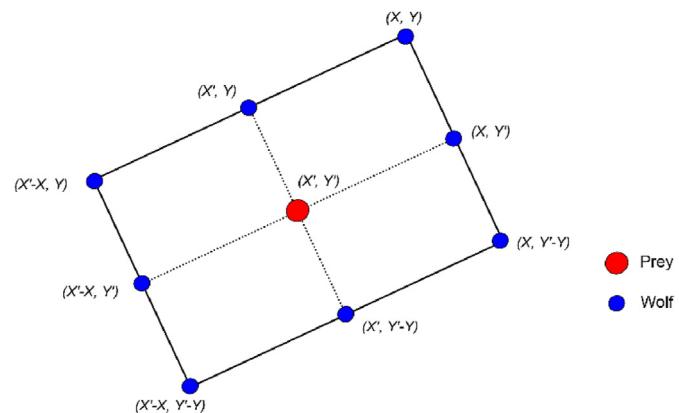


Fig. 2. The 2D mesh representing the position of prey and wolves.

$$\mathbf{P}_\alpha = |\mathbf{C}_1 \mathbf{X}_\alpha - \mathbf{X}| \quad (19)$$

$$\mathbf{P}_\beta = |\mathbf{C}_2 \mathbf{X}_\beta - \mathbf{X}| \quad (20)$$

$$\mathbf{P}_\delta = |\mathbf{C}_3 \mathbf{X}_\delta - \mathbf{X}| \quad (21)$$

The Sanitized Grey Wolf Optimizer (SGWO) algorithm is a modified version of the Grey Wolf Optimizer (GWO) algorithm based on the principles of chaos theory. The SGWO algorithm uses chaos theory to generate random values for the exploration and exploitation of the search space, which can improve the algorithm's convergence rate and global optimization capability. Additionally, the SGWO algorithm employs a sanitization process that ensures the feasibility and validity of the candidate solutions during the search process. In the present work, the SGWO algorithm is used to optimize the parameters of a model. The algorithm is applied to find the optimal values of the model parameters that minimize or maximize a given objective function. This optimization process can improve the performance and accuracy of the model, making it more suitable for practical applications.

#### 2.4. Boosting algorithms

Boosting algorithms, including gradient boosting, aim to minimize an objective function  $L$  that measures the difference between the true values ( $y$ ) and the predicted values ( $\hat{y}$ ). The objective function can be expressed as:

$$L(y, \hat{y}) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{i=1}^K \Omega(f_i) \quad (22)$$

In the above equation,  $n$  is the number of samples in the training data,  $l(y_i, \hat{y}_i)$  is the loss function that measures the difference between the true value  $y_i$  and predicted values  $\hat{y}_i$  for the  $i^{\text{th}}$  sample,  $K$  is the number of weak learners in the ensemble,  $\Omega(f_i)$  is a regularization term that penalizes complex models, and  $f_i$  is the weak learner at the  $i^{\text{th}}$  iteration. The boosting algorithm iteratively updates the weights of the training samples and the parameters of the weak learners to minimize the objective function. The results from the four models are averaged in the second layer, and several conditions are checked before the final result is predicted. The second layer also includes a bagging machine learning model, Random Forest (RF) [45], which is trained on both the actual and mixed datasets. In bagging, the objective function for each tree  $t$  in the ensemble is to minimize the loss function,  $L(y_i, \hat{y}_i)$ . The predicted value for a given sample is calculated as the average of the predicted values from all trees in the ensemble:

$$\hat{y}_i = \frac{1}{T} \sum_{t=1}^T f_t(x_i) \quad (23)$$

In this equation,  $f_t(x_i)$  is the predicted value for the  $i^{\text{th}}$  sample using the  $t^{\text{th}}$  tree in the ensemble.  $T$  is the number of trees in the ensemble. The objective function for bagging can be written as:

$$\min_{f_t} \frac{1}{N} \sum_{i=1}^N L(y_i, \hat{y}_i) \quad (24)$$

In this equation,  $N$  is the number of samples in the training data.

#### 2.5. Proposed framework

Fig. 3 describes the process of magnitude prediction using synthetic data generated by the CTGAN algorithm. The pipeline

consists of four distinct stages: (1) Preprocessing and feature extraction, which involves the preparation of the raw earthquake waveform data and extraction of relevant features; (2) Synthetic data generation, which uses the CTGAN algorithm to generate synthetic data resembling the underlying data distribution; (3) Magnitude prediction, which utilizes the machine learning models to predict the magnitude of the target variable using the synthetic data; and (4) Statistical evaluation, which involves the use of various statistical measures to evaluate the accuracy of the magnitude prediction. In this study, a model named MagPred is designed to predict the magnitude of an earthquake based on the first few seconds of P wave features.

**Problem Statement:** The input to the model is the feature set,  $\mathcal{X} \in \mathbb{R}^{N \times F}$  with  $N$  records. The feature set ( $F$ ) consists of thirty-four features given in the Data Study section extracted from the early few seconds (3, 4, or 5 s) of the P-wave signal. The output of the study is the earthquake magnitude ( $\hat{y}_i$ ) of  $i^{\text{th}}$  single station in  $M_{\text{JMA}}$  unit.

The first step in the MagPred model is the distribution of input training data  $\mathcal{X}$  into three components  $\mathcal{X}_{\beta_1}$ ,  $\mathcal{X}_{\beta_2}$ , and  $\mathcal{X}_{\beta_3}$ . These three components involved dividing  $\mathcal{X}$  dataset based on the range [5–6], [6–7], and [7–9]  $M_{\text{JMA}}$  and named as  $\mathcal{X}_{\beta_1}$ ,  $\mathcal{X}_{\beta_2}$ , and  $\mathcal{X}_{\beta_3}$ . Then, the synthetic dataset ( $\mathcal{X}_S$ ) is built using the concatenation of synthetic datasets generated by  $\text{CTGAN}_{\beta_1}$ ,  $\text{CTGAN}_{\beta_2}$ , and  $\text{CTGAN}_{\beta_3}$  as shown below:

$$\mathcal{X}_S = [\text{CTGAN}_{\beta_1}(\mathcal{X}_{\beta_1}), \text{CTGAN}_{\beta_2}(\mathcal{X}_{\beta_2}), \text{CTGAN}_{\beta_3}(\mathcal{X}_{\beta_3})] \in \mathbb{R}^{N_s \times (F \times 1)} \quad (25)$$

The  $N_s$  denote the synthetic records generated and  $(F \times 1)$  shows the number of input features plus one output magnitude. The mixed dataset ( $\mathcal{X}_M$ ) is passed to the models involving concatenation of  $\mathcal{X}_M = \{\mathcal{X}_S, \mathcal{X}\} \in \mathbb{R}^{(N_s+N) \times (F \times 1)}$ . The actual and mixed training datasets are passed to four boosting machine

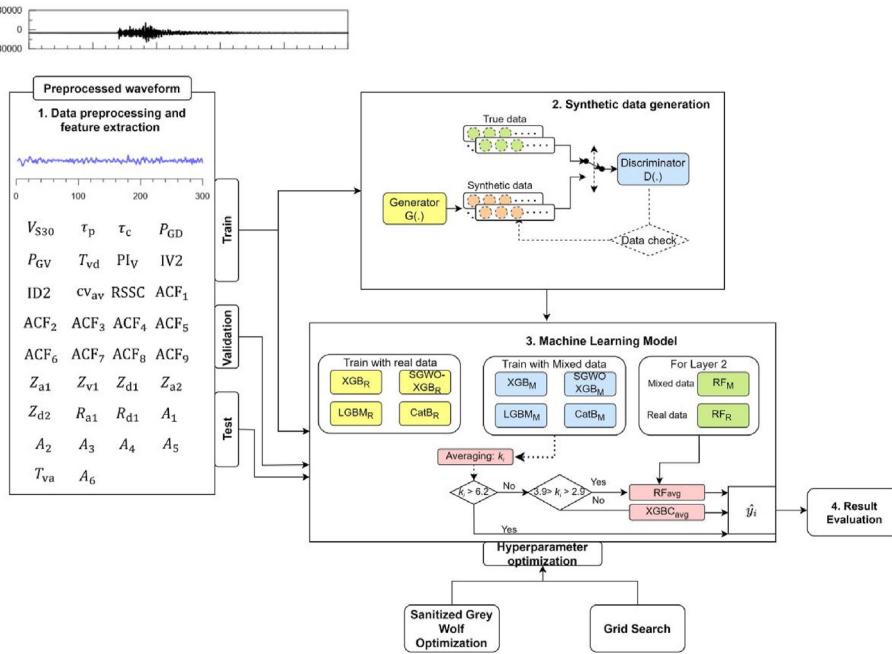
learning models in the first stage: XGB [42], SGWO-XGB, LGBM [44], and CatB [43]:

$$k_i = (\text{XGB}_R(\mathcal{X}) + \text{XGB}_M(\mathcal{X}_M) + \text{SGWO-XGB}_R(\mathcal{X}) + \text{SGWO-XGB}_M(\mathcal{X}_M) + \text{LGBM}_R(\mathcal{X}) + \text{LGBM}_M(\mathcal{X}_M) + \text{CatB}_R(\mathcal{X}) + \text{CatB}_M(\mathcal{X}_M))/8 \quad (26)$$

The models are subscripted with  $R$  and  $M$  based on the observation that they are trained with the real or mixed dataset. The model's first stage yields the average magnitude for the  $i^{\text{th}}$  index, denoted as  $k_i$ . The final magnitude selection is based on specific conditions that are given below:

$$\hat{y}_i = \begin{cases} k_i & k_i \geq 6 \\ \frac{(\text{RF}_M + \text{RF}_R)}{2} & 3.9 \geq k_i \geq 2.9 \\ \frac{(\text{XGB}_R + \text{CatB}_R + \text{SGWO-XGB}_M)}{3} & \text{otherwise} \end{cases} \quad (27)$$

In this condition, if  $k_i$  is greater than 6.2, then the value of  $k_i$  is directly assigned to the result variable denoted by  $\hat{y}_i$ . If  $k_i$  is not 6.2, the algorithm checks if  $3.9 \geq k_i \geq 2.9$ . In this case, the predictions from  $\text{RF}_M$  and  $\text{RF}_R$  are averaged denoted by  $\text{RF}_{\text{avg}}$ , and the result is assigned to  $\hat{y}_i$ . If  $k_i$  does not meet either condition, the average of the predictions ( $\text{XGB}_{\text{avg}}$ ) from  $\text{XGB}_R$ ,  $\text{CatB}_R$ , and  $\text{SGWO-XGB}_M$  is assigned to  $\hat{y}_i$ . Finally, statistical evaluation metrics are used to evaluate the results. CTGAN was trained for 50 epochs. The scikit-learn library is utilized to obtain machine learning models [55].



**Fig. 3.** The figure illustrates the pipeline utilized for magnitude prediction based on synthetic data generated by CTGAN. The pipeline consists of four main stages. The raw data is preprocessed in the first stage, and relevant features are extracted. The second stage involves the generation of synthetic data using CTGAN. The third stage comprises the utilization of machine learning models to perform the magnitude prediction. Finally, the fourth stage involves a statistical evaluation of the predicted results.

### 3. Data Study

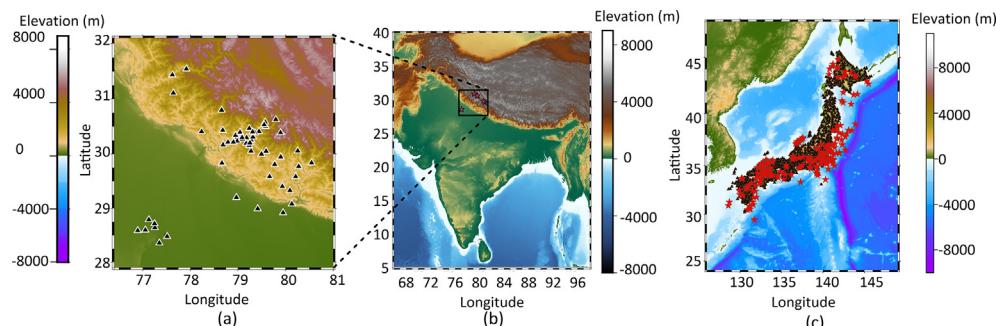
The strong motion data from Kyoshin Network (K-NET) seismic network archives that are operated by the National Research Institute for Earth Science and Disaster Resilience (NIED), Japan [56], since June 1996, has been collected for this work. The earthquake dataset is available at <http://www.kyoshin.bosai.go.jp/>. There were 18994 real event dataset records utilized in this work. This total dataset is divided into 70%:21%:9% of train, test, and validation sets consisting of 13295, 3989, and 1710 records. Three-component waveforms, east-west, north-south, and up-down, were first collected by the Japanese Meteorological Agency (JMA). The up-down waveform is utilized for this work since it consists of features corresponding to the magnitude. The waveforms are pre-processed using baseline correction, band pass filtering and scaling. This study uses the magnitude range of 3.0–9.0 to train the machine learning model. As seismic waves travel further, energy loss increases at a faster rate, causing a reduction in the quality of the waveform characteristics [57]. Therefore, the records with epicenter distance less than 200 km are selected for this study [58].

The Pearson correlation matrix introduced by Pearson [59] provides a quantitative evaluation of the linear association between two sets of continuous variables. Table 1 summarizes the datasets employed in this study for earthquake magnitude prediction. The SE3, SE4, and SE5 datasets correspond to training sets that utilize the initial 3, 4, and 5 s of the P-wave, respectively. For cross-regional evaluation, five earthquake event from India or Nepal were selected from the database of the Program for Excellence in Strong Motion Studies (PESMOS) provided by the Indian Institute of Technology, Roorkee, India (<http://www.pesmos.in>). This event is labelled as India-Nepal, with station and epicenter shown in Fig. 4(a) and Fig. 4(b). The location of the epicenter and stations of the selected earthquake from Japan region are shown in Fig. 4(c). The inter-region earthquake dataset is the testing dataset, evaluated against models trained with the SE3, SE4, and SE5 training sets. Consequently, the inter-region dataset is also partitioned into inter-region earthquake (IRE) datasets, named IRE3, IRE4, and IRE5, for 3, 4, and 5-s windows of P-wave data for testing.

**Table 1**

Parameter descriptions of SE3, SE4, and S5 training dataset. Cross-region testing of the model utilizes earthquake data from the India-Nepal region. Notably, no synthetic data is included for the India-Nepal region, as it serves exclusively as the testing dataset.

Dataset	Description	Magnitude	Ed (km)	Fd (km)	Real Data	Synthetic Data
SE3	Initial 3 s record	M <sub>JMA</sub> 3-9	0.02–200	0–256	18994	6612
SE4	Initial 4 s record	M <sub>JMA</sub> 3-9	0.02–200	0–256	18994	5211
SE5	Initial 5 s record	M <sub>JMA</sub> 3-9	0.02–200	0–256	18994	6621
India	India or Nepal region	M <sub>W</sub> 4.0–5.7	4.10–191.1	8–13	66	—



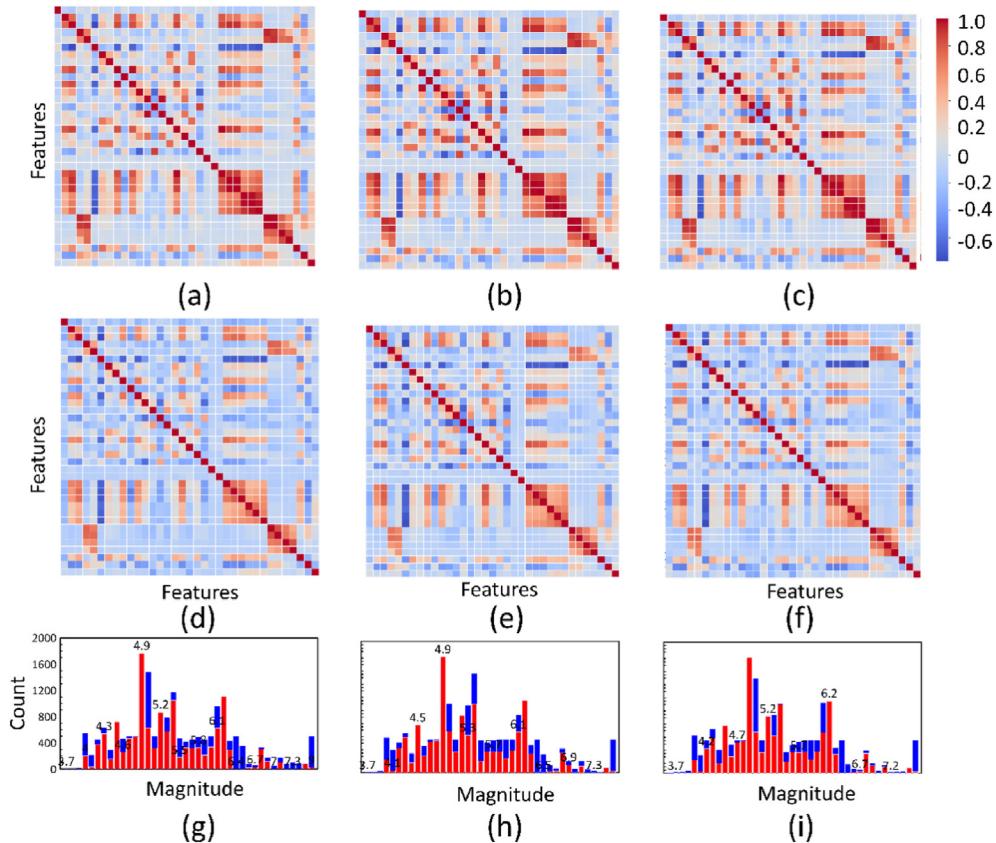
**Fig. 4.** Graphical illustration of the geospatial positioning of stations and events in (a) India and Nepal region and (c) Japan region. Seismic monitoring stations are represented by triangles and earthquake epicenters are represented by stars.

The Pearson correlation matrix consists of a square matrix with elements representing the Pearson correlation coefficients between all pairs of variables in a dataset. Pearson correlation plot can be useful in feature selection, identifying redundant features, and exploring the relationships between variables in the data set. The mathematical formulation for computing the Pearson correlation coefficients is presented as follows:

$$r_{XY} = \frac{\sum_{j=1}^N (X_j - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^N (X_j - \bar{X})^2} \sqrt{\sum_{j=1}^N (Y_j - \bar{Y})^2}} \quad (28)$$

In this equation,  $X_j$  and  $Y_j$  are the values for the  $j^{th}$  sample in the two variables being compared.  $\bar{X}$  and  $\bar{Y}$  are the means of the two variables.  $N$  is the number of samples.  $r_{XY}$  is the Pearson correlation coefficient between the two variables. The resulting value ranges from  $-1$  to  $1$ , indicating the strength and direction of the linear relationship. Fig. 5 (a) illustrates the Pearson correlation matrix for the actual training records. The Pearson correlation for mixed dataset (consisting of actual and synthetic training datasets generated via CTGAN) obtained from SE3, SE4, and SE5 is shown in Fig. 5 (b–d). The Pearson plot for the actual training dataset will remain the same for SE3, SE4, and SE5 training data. Therefore, it is represented by Fig. 5 (a). It can be visualized from the plots comparison that the actual and mixed data plots hold similar correlations. Thus, maintaining feature correlation even after generating synthetic records using CTGAN.

The distribution of magnitude and its corresponding count are represented for the actual training dataset in Fig. 5 (e). In Fig. 5 (f–h), the distribution count increases for larger earthquake events, particularly for SE3, SE4, and SE5. However, achieving a perfectly equal distribution across all events is not feasible, as it would disrupt the original correlation between features and magnitude. Therefore, the distribution was adjusted to balance feature correlation with magnitude while addressing data imbalance. The input dataset consists of synthetic and actual datasets. Therefore, the input matrix size passed to the model is  $34 \times N$ , where  $N$  is the total number of samples.



**Fig. 5.** The figure compares the Pearson correlation matrix for (a) real and (b,c,d) the mixed dataset of three, four, and five seconds of the training dataset. (e, f, g, h) The bar chart illustrates the frequency distribution of earthquake magnitudes in the training dataset's of actual and mixed records of SE3, SE4, and SE5. The sequence in which the features are represented in the Pearson plot is  $[V_{S30}, \tau_p, \tau_c, P_{CV}, P_{GD}, T_{vd}, T_{va}, ACF_1, ACF_2, ACF_3, ACF_4, ACF_5, ACF_6, ACF_7, ACF_8, ACF_9, Z_{a1}, Z_{v1}, Z_{d1}, Z_{a2}, Z_{d2}, R_{a1}, R_{d1}, A_1, A_2, A_3, A_4, A_5, A_6, cv_{av}, RSSC, IV2, ID2, PI_V, M_{JMA}]$ . Red and blue in bar plot represent the count of actual and mixed dataset records with respect to the corresponding earthquake.

### 3.1. Data and methodology

The primary waveform has been picked from the noise data using the short term to long-term average (STA-LTA) algorithm [2] on the UD component of earthquake waveform. The threshold of 2.0 has been selected for the STA-LTA algorithm with short time and long time as 10 and 70 samples, respectively. The acceleration of the record in UD components (*a*) has been converted to velocity (*v*) and displacement (*d*) waveform for parameter prediction. The four-order Butterworth filter [60] with a lower and upper-frequency range of 0.5 and 45 Hz is applied after the integration of the waveform. This has been applied to remove ‘low-frequency drift’ after the first integration. The higher order of four is selected since it leads to sharper attenuation [61].

The raw dataset from which the features have been extracted is typical time series data, which also consists of site-specific information. The extracted features used in this work mainly include site-specific parameters, time domain features, frequency domain features, and wavelet transform features. Most of the extracted features from the early P phase have been utilized by Joshi et al. [14] for the EEWs system. There are 13295, 3989, and 1710 records utilized for training, testing, and validating the MagPred model.

There were thirty-four parameters passed as input to the MagPred model. Since a single parameter may not provide magnitude information. To improve magnitude estimation, several parameters are required in the model that represents the amplitude, frequency, or energy information of the selected window of P waveform [11]. The more variable parameters the model provides, the better machine learning will perform. This will lead to accurate magnitude estimation for hazard mitigation. The grid search was performed on various combinations of parameters and found out that these thirty-four parameters provided better result.

$V_{S30}$ : The average shear wave velocity of the 30 m depth ( $V_{S30}$ ) is the site-specific feature that has been utilized in the present work as calculated below:

$$V_{S30} = \frac{\sum_{i=1}^N h_i}{\sum_{i=1}^N h_i / V_{si}} \quad (29)$$

In this equation,  $h_i$  and  $V_{si}$  denote the thickness in meter and shear velocity in  $\text{ms}^{-1}$  of  $i^{\text{th}}$  layer. The  $N$  denotes the top 30-m depth.

Tauc ( $\tau_c$ ): The  $\tau_c$  parameter was initially introduced by Kanamori [46] and is a modified version of the method originally given by Nakamura [62].  $\tau_c$  is calculated for EEWs from the vertical component of the waveform as shown below:

$$r = \frac{\int_0^T |v^2(t)| dt}{\int_0^T |d^2(t)| dt} \quad (30)$$

$$\tau_c = \frac{2\pi}{\sqrt{r}} \quad (31)$$

In this equation,  $v(t)$  and  $d(t)$  are the velocity and displacement calculated from the vertical component of the first few seconds of the P wave.

Taup ( $\tau_p$ ): The  $\tau_p$  parameter is used in earthquake magnitude detection as a measure of the duration of the seismic waves generated by an earthquake that was applied by Wu and Kanamori [8]. It is defined as:

$$\tau_p = \frac{\int_0^{T_p} t^{p-1} |h(t)|^2 dt}{\int_0^{T_p} |h(t)|^2 dt} \quad (32)$$

In the above equation,  $h(t)$  is the seismogram,  $T_p$  is the period of the P-wave, and  $p$  is a positive integer exponent that is typically set to 2 or 3.

$P_{GV}$  and  $P_{GD}$ : The  $P_{GV}$  and  $P_{GD}$  refers to the absolute peak velocity and displacement given by Wu and Kanamori [63]. These parameters are calculated from the preprocessed waveform from few seconds of the P wave.

$T_{vd}$  and  $T_{va}$ : The parameter  $T_{vd}$  is calculated using a similar relationship as given by Zhu et al. [13] as given below:

$$T_{vd} = \frac{P_{GV}}{P_{GD}} \quad (33)$$

The equation of parameter  $T_{va}$  is shown below:

$$T_{va} = 2\pi \frac{P_{GV}}{P_{GA}} \quad (34)$$

PI<sub>V</sub>, IV2, and ID2: The parameter PI<sub>V</sub> is the P wave index value that was initially given by Nakamura [64] as shown below:

$$PI_V = 0 \leq t \leq T_{max}(\log|a(t)v(t)|) \quad (35)$$

In this equation,  $a(t)$  and  $v(t)$  are the individual acceleration and velocity instances obtained from the early few seconds of the P wave. The parameters IV2 and ID2 are velocity and displacement squared integral given by Festa et al. [65]. These parameters are calculated as shown in Eq. (36) and Eq. (37).

$$IV2 = \int_0^T v^2(t) dt \quad (36)$$

$$ID2 = \int_0^T d^2(t) dt \quad (37)$$

cv<sub>av</sub> and RSSC: The parameter cv<sub>av</sub> is the cumulative vertical absolute velocity utilized by Zhu et al. [13] for magnitude estimation as shown below:

$$cv_{av} = \sum_{0 \leq t \leq T} |v(t)| \quad (38)$$

The parameter RSSC is given below:

$$RSSC = \sum \sqrt{v^2(t)} \quad (39)$$

ACF<sub>1</sub>, ACF<sub>2</sub>, ACF<sub>3</sub>, ACF<sub>4</sub>, ACF<sub>5</sub>, ACF<sub>6</sub>, ACF<sub>7</sub>, ACF<sub>8</sub>, and ACF<sub>9</sub>: The parameter ACF<sub>1</sub>, ACF<sub>2</sub>, and ACF<sub>3</sub> denote the time of the first negative peak in the autocorrelation function (ACF) of acceleration, velocity, and displacement record. The parameter ACF<sub>4</sub>, ACF<sub>5</sub>, and ACF<sub>6</sub> denote the time of the second positive peak in the ACF of acceleration, velocity, and displacement. ACF<sub>7</sub> and ACF<sub>8</sub> denote the time of the second negative peak in the ACF of velocity and displacement. The ACF<sub>9</sub> denotes the time of the first positive peak in ACF of the displacement record.

$Z_{a1}, Z_{v1}, Z_{d1}, Z_{a2}$ , and  $Z_{d2}$ : The parameter  $Z_{a1}, Z_{v1}$ , and  $Z_{d1}$  refers to the time of the first zero crossing in the ACF function of acceleration, velocity, and displacement. The parameter  $Z_{a2}$  and  $Z_{d2}$  refers to the second zero crossing in the ACF function of the acceleration and displacement. These parameters are utilized by Sinvhal and Khattri [66] to differentiate oil-bearing structures from seismic signals.

$R_{a1}$  and  $R_{d1}$ : The parameters  $R_{a1}$  and  $R_{d1}$  reflects the proportion of the area on both the positive and negative side of the abscissa in the ACF function, derived from the acceleration and displacement record as shown below:

$$R_{a1} = \frac{\sum_{i=0}^N a^+(i)}{\sum_{i=0}^N a^-(i)} \quad (40)$$

In the equation,  $a^+(i)$  represents the positive value of  $a$  at the  $i^{th}$  index and is set to 0 if the original value of  $a$  is negative. Conversely,  $a^-(i)$  represents the negative value of  $a$  at the  $i^{th}$  index and is set to 0 if the original value of  $a$  is positive. The same principle applies to the parameter  $R_{d1}$  and its corresponding positive and negative values.

$A_1, A_2, A_3, A_4, A_5$ , and  $A_6$ : The parameters  $A_1, A_2$ , and  $A_3$  correspond to the values of the ACF at lags one, two, and three, respectively. The parameter  $A_4, A_5$ , and  $A_6$  corresponds to the ratio of the second, third, and fourth index of ACF function with respect to the zero index value.

## 4. Results

### 4.1. Evaluation metrics

The evaluation of the predictions is based on two statistical parameters: mean absolute error (MAE) and root mean square error (RMSE). The mathematical expression for the MAE metric is depicted as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (41)$$

In this equation,  $N$  is the total number of samples.  $y_i$  is the true value for the  $i^{th}$  sample.  $\hat{y}_i$  is the predicted value for the  $i^{th}$  sample. The equation for RMSE is exemplified in the aforementioned mathematical expressions.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (42)$$

In this equation, variable  $N$  represents the number of samples present and two distinct variables associated with each sample, namely  $y_i$ , the true value, and  $\hat{y}_i$ , the predicted value. The formula for calculating individual error ( $\mu_i$ ) is shown below:

$$\mu_i = y_i - \hat{y}_i \quad (43)$$

$\text{ME}$  represents mean error obtained using  $\mu_i$ , and  $\sigma$  represents the standard deviation obtained from  $\mu_i$ .

#### 4.2. Model ablation study

The dataset utilized in this study comprises magnitudes recorded in the JMA scale. To reconcile this difference between inter-region, a conversion formula derived from Scordilis [67] has been employed to convert magnitude values in  $M_{\text{JMA}}$  scale ( $\text{mag}(M_{\text{JMA}})$ ) to the magnitude values in moment magnitude scale ( $\text{mag}(M_w)$ ). This conversion is necessary to ensure consistency and compatibility with the chosen scale for analysis and modelling.

$$\text{mag}(M_w) = 0.58\text{mag}(M_{\text{JMA}}) + 2.25, 3.0 \leq \text{mag}(M_{\text{JMA}}) \leq 5.5 \quad (44)$$

$$\text{mag}(M_w) = 0.97\text{mag}(M_{\text{JMA}}) + 0.04, 5.6 \leq \text{mag}(M_{\text{JMA}}) \leq 8.2 \quad (45)$$

**Table 2** represents the selected optimized hyperparameters corresponding to the machine learning models trained on the SE3, SE4, and SE5 datasets. The hyperparameters of all three datasets are the same. These hyperparameters are selected using grid search and a five-fold cross-validation set. The result obtained from the average of all the models predicted after the first layer is denoted by  $k_i$ . The final magnitude selection is based on specific selection conditions as shown in Eq. (27). The Eq. (27) shows that if  $k_i$  is between 2.9 and 3.9  $M_{\text{JMA}}$ , the result obtained from  $\text{RF}_{\text{avg}}$  is selected as the final magnitude of  $i^{\text{th}}$  record. **Table 3** represents the ablation study that shows the justification for assigning the  $\text{RF}_{\text{avg}}$  value to the predicted  $M_{\text{JMA}}$  value denoted by  $\hat{y}_i$ . The overall error of the test

and validation set of  $\text{RF}_{\text{avg}}$  is less than other model units. If the value of  $k_i$  in Eq. (27) is between 4.0 and 6.0  $M_{\text{JMA}}$ , then the  $\text{XGBC}_{\text{avg}}$  model is selected. The  $\text{XGBC}_{\text{avg}}$  consists of the average of  $\text{XGB}_R$ ,  $\text{CatB}_R$ , and  $\text{SGWO-XGB}_M$  models prediction selected as final magnitude value for ranges between 4 and 6 (with 4 and 6 inclusive)  $M_{\text{JMA}}$  as shown in **Table 4**. **Table 4** shows that  $\text{XGBC}_{\text{avg}}$  method outperforms other methods in this case. However, if  $k_i$  is larger than 6.0  $M_{\text{JMA}}$ , the average of all the models is selected. **Table 5** shows the justification for this selection criteria. The overall reduction in error of the test and validation set of Average is less than other model units. In **Tables 3–5**, the test datasets selected for comparison are those where the predictions  $k_i$  fall within the ranges specified by conditions II, III, and I of Eq. (27). The size of the test and validation dataset in **Tables 3–5** are utilized with a real dataset consisting of 3989 and 1710 records.

**Table 6** represents the model ablation study results, which compare the proposed method's performance with several other machine learning models, including SGWO-XGB, XGB, CatB, LGBM, tabular data learning architecture (Tab-Net) [68], Neural Network with 2 hidden layers (NN) [69], Support Vector Regressor model with radial basis kernel (SVR) [70] and RF. The study evaluates the effectiveness of each model by measuring its performance on a set of relevant metrics, such as RMSE and MAE. It can be visualized from **Table 6** that the NN, SVR and Tab-Net give more error compared with other models. Thus, they are not included in the MagPred pipeline. Another observation that can be noted from **Table 6** is that the models trained on mixed datasets give less error compared with the ones trained on only real datasets. This proves that when machine learning models are trained on mixed datasets, which consists of CTGAN generated and real input features for handling data imbalance, they give less error.

#### 4.3. Quantitative analysis

**Table 7** shows the quantitative result analysis on SE3, SE4, and SE5 datasets. The predictions obtained from the trained models were compared with several regression relations from 2006 to 2023 as shown in **Table 7**. The three regression models for  $P_{\text{GD}}$  proposed by Kumar et al. [10] and Wu et al. [9], for  $\tau_p^{\max}$  by Kumar et al. [10], and for CAA by Wu et al. [9] describe the relationships when these parameters are extracted from the initial 3, 4, and 5 s of

**Table 2**

The values of the optimized hyperparameters selected for the corresponding model, with the hyperparameter names consistent with those used in the Scikit-Learn library parameters.

Model	Hyperparameter Names	Value
SGWO – XGB <sub>M</sub>	max_depth, learning_rate, n_estimators, subsample, min_child_weight, scale_pos_weight, reg_alpha, reg_lambda	[4, 0.1, 500, 300, 0.7, 2, 1, 0, 1]
SGWO – XGB <sub>R</sub>	max_depth, learning_rate, n_estimators, subsample, min_child_weight, scale_pos_weight, reg_alpha, reg_lambda	[3, 0.1, 500, 300, 0.7, 2, 1, 0.005, 1]
XGB <sub>M</sub>	max_depth, learning_rate, n_estimators, subsample, min_child_weight, scale_pos_weight, reg_alpha, reg_lambda	[3, 0.1, 100, 3000, 0.86, 2, 1, 0.005, 1]
XGB <sub>R</sub>	max_depth, learning_rate, n_estimators, subsample, min_child_weight, scale_pos_weight, reg_alpha, reg_lambda	[3, 0.1, 100, 3000, 0.86, 2, 1, 0.005, 1]
CatB <sub>M</sub>	loss_function, max_depth, n_estimators	[RMSE, 4, 500]
CatB <sub>R</sub>	loss_function, max_depth, n_estimators	[RMSE, 5, 500]
LGBM <sub>M</sub>	Objective, max_depth, base_score, learning_rate, lambda_l1, num_boost_round	[Regression, 15, 500, 0.05, 0.01, 1500]
LGBM <sub>R</sub>	Objective, max_depth, base_score, learning_rate, lambda_l1, num_boost_round	[Regression, 15, 200, 0.05, 0.02, 300]
RF <sub>M</sub>	max_depth, n_estimators	[9, 500]
RF <sub>R</sub>	max_depth, n_estimators	[8, 500]

**Table 3**

The model ablation study to select the model for  $M_{JMA}$  when  $k_i$  is between 2.9 and 3.9  $M_{JMA}$  prediction. The results are reported in units of  $M_{JMA}$ . The predicted dataset within the range of 2.9 and 3.9  $M_{JMA}$  is selected for comparison. The models with subscript M consist of mixed datasets. The mixed datasets for SE3, SE4, and SE5 have 25606, 24205, and 25615 records, respectively.

Models	Test						Validation					
	SE3		SE4		SE5		SE3		SE4		SE5	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
SGWO	0.25	0.36	0.24	0.34	0.24	0.33	0.27	0.41	0.23	0.30	0.23	0.29
— XGB <sub>R</sub>	0.30	0.39	0.29	0.37	0.28	0.37	0.28	0.40	0.25	0.31	0.27	0.35
SGWO	0.28	0.40	0.27	0.38	0.27	0.37	0.29	0.42	0.25	0.30	0.22	0.27
— XGB <sub>M</sub>	0.26	0.38	0.25	0.37	0.24	0.36	0.27	0.39	0.26	0.33	0.26	0.32
XGB <sub>R</sub>	0.28	0.40	0.27	0.38	0.27	0.37	0.29	0.42	0.25	0.30	0.22	0.27
CatB <sub>R</sub>	0.31	0.41	0.29	0.40	0.28	0.39	0.27	0.37	0.26	0.32	0.22	0.30
CatB <sub>M</sub>	0.27	0.37	0.36	0.36	0.36	0.38	0.28	0.38	0.29	0.36	0.28	0.35
LGBM <sub>R</sub>	0.30	0.42	0.30	0.40	0.29	0.40	0.27	0.37	0.25	0.31	0.23	0.28
LGBM <sub>M</sub>	0.26	0.38	0.25	0.37	0.23	0.37	0.28	0.40	0.25	0.29	0.22	0.28
RF <sub>R</sub>	0.25	0.36	0.24	0.35	0.25	0.34	0.27	0.39	0.27	0.33	0.27	0.33
RF <sub>M</sub>	0.29	0.40	0.28	0.39	0.28	0.39	0.28	0.38	0.37	0.44	0.36	0.43
Average (All)	0.26	0.36	0.25	0.35	0.24	0.34	0.25	0.36	0.24	0.29	0.31	0.38
RF <sub>avg</sub>	0.16	0.24	0.15	0.23	0.14	0.21	0.13	0.23	0.12	0.21	0.11	0.20

**Table 4**

The model ablation study to select the model for  $M_{JMA}$  when  $k_i$  is between 4.0 and 5.9  $M_{JMA}$  prediction. The results are reported in units of  $M_{JMA}$ . The predicted dataset within the range of 4.0 and 6.0  $M_{JMA}$  is selected for comparison. The models with subscript M consist of mixed datasets. The mixed datasets for SE3, SE4, and SE5 have 25606, 24205, and 25615 records, respectively. The XGBC<sub>avg</sub> refers to the average of XGB<sub>R</sub>, XGB<sub>M</sub>, CatB<sub>R</sub> and CatB<sub>M</sub> models.

Models	Test						Validation					
	SE3		SE4		SE5		SE3		SE4		SE5	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
SGWO	0.53	0.71	0.51	0.70	0.51	0.69	0.51	0.68	0.50	0.66	0.49	0.65
— XGB <sub>R</sub>	0.51	0.68	0.48	0.66	0.46	0.65	0.49	0.67	0.48	0.65	0.47	0.63
SGWO	0.56	0.72	0.55	0.71	0.54	0.70	0.48	0.67	0.47	0.66	0.46	0.66
— XGB <sub>M</sub>	0.51	0.69	0.48	0.67	0.47	0.66	0.47	0.66	0.46	0.64	0.46	0.62
XGB <sub>R</sub>	0.54	0.73	0.53	0.72	0.51	0.70	0.52	0.73	0.51	0.72	0.50	0.71
XGB <sub>M</sub>	0.51	0.67	0.51	0.64	0.49	0.63	0.50	0.72	0.50	0.71	0.48	0.71
CatB <sub>R</sub>	0.50	0.66	0.49	0.65	0.49	0.64	0.51	0.74	0.50	0.74	0.50	0.73
CatB <sub>M</sub>	0.48	0.64	0.48	0.63	0.48	0.60	0.49	0.71	0.48	0.71	0.47	0.70
LGBM <sub>R</sub>	0.53	0.71	0.53	0.69	0.52	0.68	0.52	0.74	0.50	0.74	0.50	0.73
LGBM <sub>M</sub>	0.57	0.75	0.56	0.73	0.54	0.72	0.55	0.76	0.56	0.75	0.54	0.72
Average (All)	0.54	0.72	0.54	0.70	0.53	0.68	0.49	0.72	0.49	0.71	0.48	0.70
RF <sub>R</sub>	0.55	0.72	0.54	0.70	0.54	0.70	0.55	0.78	0.53	0.75	0.52	0.73
RF <sub>M</sub>	0.41	0.53	0.40	0.52	0.38	0.49	0.37	0.60	0.36	0.58	0.35	0.64

**Table 5**

The model ablation study to select the model when  $k_i$  is greater than 6.0  $M_{JMA}$  (excluding 6.0) prediction. The results are reported in units of  $M_{JMA}$ . The predicted dataset within the range greater than 6.0  $M_{JMA}$  is selected for comparison. The models with subscript M consist of mixed datasets. The mixed datasets for SE3, SE4, and SE5 have 25606, 24205, and 25615 records, respectively.

Models	Test						Validation					
	SE3		SE4		SE5		SE3		SE4		SE5	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
SGWO	0.41	0.60	0.40	0.58	0.40	0.56	0.40	0.60	0.40	0.59	0.39	0.58
— XGB <sub>R</sub>	0.39	0.57	0.39	0.56	0.38	0.53	0.38	0.58	0.37	0.57	0.37	0.56
SGWO	0.43	0.62	0.42	0.61	0.42	0.60	0.44	0.66	0.43	0.65	0.43	0.65
— XGB <sub>M</sub>	0.45	0.67	0.45	0.64	0.44	0.63	0.41	0.64	0.40	0.63	0.39	0.62
XGB <sub>R</sub>	0.46	0.64	0.46	0.62	0.45	0.60	0.39	0.61	0.38	0.60	0.36	0.59
XGB <sub>M</sub>	0.45	0.62	0.43	0.61	0.42	0.59	0.41	0.63	0.40	0.62	0.40	0.61
CatB <sub>R</sub>	0.43	0.61	0.42	0.60	0.42	0.60	0.48	0.69	0.48	0.68	0.46	0.67
CatB <sub>M</sub>	0.42	0.58	0.42	0.56	0.42	0.55	0.45	0.64	0.43	0.63	0.43	0.62
LGBM <sub>R</sub>	0.47	0.61	0.46	0.60	0.46	0.60	0.45	0.64	0.44	0.63	0.43	0.62
LGBM <sub>M</sub>	0.44	0.62	0.43	0.61	0.42	0.58	0.47	0.69	0.46	0.67	0.45	0.67
RF <sub>R</sub>	0.47	0.61	0.46	0.60	0.46	0.60	0.45	0.64	0.44	0.63	0.43	0.62
RF <sub>M</sub>	0.44	0.62	0.43	0.61	0.42	0.58	0.47	0.69	0.46	0.67	0.45	0.67
XGBC <sub>avg</sub>	0.41	0.56	0.38	0.55	0.38	0.54	0.40	0.65	0.40	0.64	0.39	0.64
RF <sub>avg</sub>	0.44	0.60	0.44	0.59	0.43	0.58	0.43	0.67	0.42	0.65	0.42	0.64
Average (All)	0.32	0.38	0.31	0.36	0.31	0.34	0.29	0.35	0.28	0.33	0.27	0.34

**Table 6**

The model ablation study based on RMSE and MAE for the test dataset. The comparison is presented for the best performing models. The results are reported in units of  $M_{JMA}$ .

Models	SE3		SE4		SE5	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
SGWO – XGB <sub>M</sub>	0.55	0.66	0.53	0.65	0.53	0.62
SGWO – XGB <sub>R</sub>	0.54	0.69	0.53	0.68	0.53	0.66
XGB <sub>M</sub>	0.52	0.67	0.52	0.65	0.51	0.63
XGB <sub>R</sub>	0.58	0.69	0.55	0.66	0.55	0.64
CatB <sub>M</sub>	0.55	0.68	0.54	0.67	0.54	0.66
CatB <sub>R</sub>	0.52	0.67	0.51	0.64	0.52	0.64
LGBM <sub>M</sub>	0.53	0.66	0.51	0.65	0.51	0.64
LGBM <sub>R</sub>	0.63	0.80	0.62	0.79	0.61	0.79
RF <sub>M</sub>	0.52	0.62	0.51	0.60	0.52	0.60
RF <sub>R</sub>	0.55	0.58	0.55	0.57	0.53	0.55
SVR <sub>M</sub>	0.63	0.83	0.60	0.82	0.58	0.80
SVR <sub>R</sub>	0.61	0.82	0.60	0.81	0.60	0.80
Tab – Net <sub>M</sub>	0.62	0.84	0.63	0.82	0.61	0.80
Tab – Net <sub>R</sub>	0.61	0.85	0.61	0.82	0.59	0.80
NN <sub>M</sub>	0.79	1.01	0.77	0.98	0.79	1.00
NN <sub>R</sub>	0.81	1.01	0.80	1.00	0.78	0.99
MagPred	0.42	0.56	0.40	0.54	0.39	0.53

the P-wave. Initially the  $P_{GD}$  relation for magnitude estimation was given by Wu and Zhao [63] and  $\tau_c$  relation given by Kanamori [46], for Japan dataset. Comparison is also carried out on ML based method named SVM-M by Zhu et al. [11]. The model proposed by Zhu et al. [11] utilizes a support vector machine (SVM-M) model for magnitude estimation. Twelve parameters are passed as input to the model. The SVM-M model has been trained and tested on the same dataset used in this study. This has been done to better compare the proposed model with the model given by Zhu et al. [11] on the same dataset.

In Table 7,  $P_{GD}$  is the peak ground displacement that was mentioned in Section 2.5.1, and  $R$  is the hypocenter distance in km. The  $\tau_c$  is the same as the feature discussed in Section 2.5.1. It can be observed from Table 7 that the proposed MagPred model outperforms other compared ML models and regression relations. The other second and third best-performing models on this dataset are  $P_{GD}$  relation by Kumar et al. [10] and Bhardwaj et al. [71].

It can be seen that in Table 7, the MAE are 0.43, 0.41, and 0.39 for SE3, SE4, and SE5, respectively. This indicates that the error decreases as the selected time window for the P wave in magnitude prediction increases. However, the reduction in error from SE3 to SE5 is minimal, suggesting that selecting an early 3-s window can still provide accurate magnitude predictions while offering more lead time for earthquake early warning systems.

The data demonstrate that the MagPred model outperforms traditional regression methods. Traditional methods typically rely on linear relationships with a limited set of parameters, whereas machine learning models, like MagPred, can uncover complex patterns within the dataset through optimized weight training. Additionally, this work includes the derivation of linear relationships involving  $\tau_c$  and  $P_{GD}$  specific to this dataset, complementing the traditional regression approaches. These new relations which fit the dataset utilized in this study are denoted by  $\tau_c^\dagger$  and  $P_{GD}^\dagger$ , respectively, as shown in Table 7. The results from Table 7 also indicates that the MagPred model outperforms the  $\tau_c^\dagger$  and  $P_{GD}^\dagger$  regression relations for magnitude estimation.

Table 8 presents the prediction errors for earthquakes in India and Nepal, derived from the PESMOS datasets. The results shown in Table 8 were obtained using the MagPred model, which was trained on earthquake data from Japan. Notably, the model demonstrates better predictive accuracy for events with magnitudes above  $M_W$  5.0. This trend is evident from the higher errors observed for earthquakes with magnitudes below  $M_W$  5.0. These findings

highlight the limitations of the MagPred model when applied to inter-regional studies, especially for smaller magnitude earthquakes.

The scope of the study has broadened to explore the prediction of earthquake magnitude based on earthquake events. Table 9 represents the weighted and unweighted magnitude obtained for five selected earthquakes. These earthquakes are the Kumamoto (Foreshock and Main), Osaka, Off-Shore, and Off Fukushima Prefecture earthquakes. These impactful earthquakes are added to enhance the clarity regarding the inter-region, multi-station magnitude prediction of the MagPred model. There are two columns named ‘Unweighted’ and ‘Weighted’ magnitude in Table 9. The ‘Unweighted’ magnitude represents the mean of all the magnitude obtained from all the stations of the selected earthquake. While ‘Weighted’ magnitude depicts the weighted mean of magnitudes, where weights are assigned based on the arrival time of the waveform in the station. The weights are assigned based on time intervals of 3 s and are increased by 0.01 for every element within each interval. The weight is calculated using the following equation:

$$w = 1.00 + 0.01i \quad (46)$$

Here,  $i$  represents the  $i^{\text{th}}$  element within a 3 s interval. The results presented in Table 9 indicate that using weights improves the average magnitude estimation compared with the scenario without assigning weights. Therefore, the final weight equation as given by Joshi et al. [14] is calculated as follows:

$$M_j = \frac{\sum_{i=1}^n w_i \sum_{k=1}^{k_i} \frac{M_i}{k_i}}{N} \quad (47)$$

In this equation,  $M_j$  is the average weighted magnitude of the  $j^{\text{th}}$  earthquake. Parameter  $k_i$  is the number of stations in the  $i^{\text{th}}$  time interval, and  $w_i$  is the weight of the corresponding interval to which the station belongs.  $N$  is the total number of stations. Table 9 shows that the error is less in inland earthquake cases for the weighted magnitude prediction than for the unweighted magnitude prediction.

#### 4.4. Qualitative analysis

Fig. 6 shows the comparison between actual and predicted magnitudes for the test dataset across three different datasets: (a) SE3, (b) SE4, and (c) SE5. The plot illustrates that when the P-wave window size is set to 5 s, the predicted magnitudes exhibit reduced scatter compared with those obtained using a 3-s window. However, this improvement is only noticeable in a subset of cases. Consequently, the results indicate that for magnitude prediction using machine learning models, a 3-s P-wave window yields performance comparable to that of a 5-s window in most instances. This validates the use of a 3-s window as a viable choice for magnitude prediction. Another observation from Fig. 6 (a–c) is that error is increased as the magnitude reaches above 6.0  $M_{JMA}$ . This can be observed from the study by Noda and Ellsworth [49] that when the departure time is 0.38 s, 1.05 s, and 2.87 s, the arrival peak amplitude occurs at 1.5 s, 3–4 s, and 10 s, respectively, for 5  $M_W$ , 6  $M_W$ , and 7  $M_W$ , respectively [77]. This explains the high error in magnitude prediction above 6  $M_{JMA}$ . The error obtained with respect to the epicenter distance for (d) SE3, (e) SE4, and (f) SE5 test set is shown in Fig. 6. The error increases as epicenter distance increases. As seismic waves travel greater distances, energy loss accelerates, leading to a significant reduction in the clarity of waveform characteristics with earthquake source parameters [57,58].

**Table 7** Comparison of various methods for magnitude estimation. The  $M_W$  unit have been utilized to compare various models.

Method	Year	Relation		SE3	SE4	SE5	
				MAE	RMSE	MAE	RMSE
$\tau_c$ by Wu et al. [72]	2006	$\text{mag}(M_W) = 7 \log_{10}(\tau_c) + 5.9$		1.17	1.31	1.53	1.31
$\tau_c$ by Wurman et al. [73]	2007	$\text{mag}(M_W) = 5.22 \log_{10}(\tau_c) + 6.66$		2.8	3.14	3.49	3.12
$\tau_c$ by Wu and Kanamori [74]	2008	$\text{mag}(M_W) = 3.37 \log(\tau_c) + 5.78$		1.54	1.78	1.98	1.72
$\tau_c$ by Zollo et al. [75]	2010	$\log(\tau_c) = 0.21(\pm 0.01)\text{mag}(M_W) - 1.2$		1.94	2.22	2.52	2.2
$\tau_c$ by Bhardwaj et al. [71]	2012	$\text{mag}(M_W) = 2.465 \log(\tau_c) + 5.299$		1.34	1.55	1.47	1.46
$P_{CD}$ by Bhardwaj et al. [71]	2012	$\text{mag}(M_W) = 2.465 \log(\tau_c) + 6.4284$		0.59	0.78	0.57	0.75
$\tau_c$ by Jin et al. [76]	2013	$\text{mag}(M_W) = 2.94 \log_{10}(\tau_c) + 5.3$		1.88	2.13	2.08	2.36
$P_{CD}$ by Jin et al. [76]	2013	$\text{mag}(M_W) = 0.91 \log_{10}(P_{CD}) + 0.48 \log_{10}(R) + 5.65$		0.85	1.07	0.8	1.02
$\tau_c$ by Kumar et al. [10]	2020	$\log_{10}(\tau_c) = 0.161 \text{mag}(M_W) - 1.424$		2.05	2.2	6.01	6.36
$\log_{10}(\tau_c) = 0.152 \text{mag}(M_W) - 1.394 \log_{10}(\tau_c) = 0.151 \text{mag}(M_W) - 1.385$				0.65	0.84	0.61	0.81
$\log_{10}(P_{CD}) = 1.027 \text{mag}(M_W) - 7.624$				0.84	1.02	0.82	0.95
$\log_{10}(P_{CD}) = 1.075 \text{mag}(M_W) - 7.707 \log_{10}(P_{CD}) = 1.101 \text{mag}(M_W) - 7.703$				3.94	5.06	3.31	4.67
$SVM-M$ by Zhu et al. [11]	2022	Machine Learning		1.89	2.14	2.17	2.4
$\tau_c$ by Zhu et al. [11]	2022	$\tau_c = 0.10 \text{mag}(M_W) + 0.16$		0.84	1.02	0.82	0.95
CAA by Wu et al. [9]	2023	$\log(CAA) = -2.132 \log(R) + 0.773 \text{mag}(M_W) - 1.658$		3.3	4.68	3.3	4.68
$P_{CD}$ by Wu et al. [9]	2023	$\log(CAA) = -2.181 \log(R) + 0.764 \text{mag}(M_W) - 1.326 \log(CAA) = -2.227 \log(R) + 0.756 \text{mag}(M_W) - 1.049$		2.27	2.5	2.27	2.5
$P_{CD}$	2024	$\log(P_{CD}) = -1.749 \log(R) + 0.503 \text{mag}(M_W) - 1.780$		1.06	1.34	1.02	1.29
Proposed $\tau_{c\ddagger}$	2024	$\log(P_{CD}) = -1.807 \log(R) + 0.635 \text{mag}(M_W) - 1.771 \log(P_{CD}) = -1.901 \log(R) + 0.682 \text{mag}(M_W) - 1.820$		0.89	1.07	0.89	1.06
Proposed $P_{GD\ddagger}$	2024	$\text{mag}(M_W) = 32.01 \log_{10}(P_{CD}) + 0.01 \log_{10}(R) + 4.07$		0.86	1.04	0.86	1.03
MagPred	2024	Machine Learning		0.42	0.56	0.40	0.54

**Fig. 7** presents a scatter plot comparing the actual and predicted magnitudes of the SE3 test set using the  $P_{GD}$  method by (a) Kumar et al. [10], (b) Bhardwaj et al. [71] and (c) MagPred method. The regression relations using  $P_{GD}$  method by Bhardwaj et al. [71] and Kumar et al. [10] have been selected for further comparison since these methods get second and third best MAE, as shown in **Table 7**. The primary objective of the EEWs is to accurately determine earthquake magnitude using data from single recording stations located in proximity to the epicenter of the earthquake. Similarly, **Fig. 7** represents a comparison plot on SE4 test set using the  $P_{GD}$  method by (d) Kumar et al. [10], (e) Bhardwaj et al. [71], and (f) MagPred method. **Fig. 7** represents a comparison plot on SE5 test set using the  $P_{GD}$  method by (g) Kumar et al. [10], (h) Bhardwaj et al. [71], and (i) MagPred method. It can be observed that **Fig. 7(c,f,i)** performs better than the magnitude predicted using  $P_{GD}$  method by Kumar et al. [10] and Bhardwaj et al. [71].

The statistical parameters  $\mu$  and  $\sigma$  denote the average error and the standard deviation of the absolute errors. **Fig. 8** represents the error obtained from the magnitude predicted from  $P_{GD}$  method by (a) Kumar et al. [10], (b) Bhardwaj et al. [71], and (c) MagPred method with respect to the epicenter distance in km for the SE3 test dataset. The  $\mu$  and  $\sigma$  of  $0.26 \pm 0.79$ ,  $0.17 \pm 0.78$  and  $0.03 \pm 0.56 M_W$  are obtained from **Fig. 8(a–c)**, respectively. **Fig. 8** represents the error obtained from  $P_{GD}$  method by (d) Kumar et al. [10], (e) Bhardwaj et al. [71], and (f) MagPred method with respect to the epicenter distance in km for the SE4 test dataset. The  $\mu$  and  $\sigma$  of  $0.19 \pm 0.77$ ,  $0.11 \pm 0.74$  and  $0.01 \pm 0.54 M_W$  are obtained from **Fig. 8(d–f)**, respectively. **Fig. 8** represents the error obtained from the magnitude predicted from  $P_{GD}$  method by (g) Kumar et al. [10], (h) Bhardwaj et al. [71], and (i) MagPred method with respect to the epicenter distance in km for the SE5 test dataset. The  $\mu$  and  $\sigma$  of  $0.17 \pm 0.75$ ,  $0.09 \pm 0.72$ , and  $0.01 \pm 0.53 M_W$  have been predicted from **Fig. 8(g,h,i)**, respectively. It can be observed from  $\mu \pm \sigma$  that the error is reduced as the input window of the P wave increases for magnitude prediction in all three comparison models.

Similarly, **Fig. 9** illustrates the error distribution for the test set of the SE3 dataset, comparing the actual magnitude with the results from (a) Kumar et al. [10], (b) Bhardwaj et al. [71], and (c) the MagPred method. **Fig. 9** shows the error obtained for the SE4 test dataset, comparing the actual magnitude with the results from (d) Kumar et al. [10], (e) Bhardwaj et al. [71], and (f) the MagPred method. Similarly, **Fig. 9** depicts the error obtained for the SE5 test dataset, comparing the actual magnitude with the results from (g) Kumar et al. [10], (h) Bhardwaj et al. [71], and (i) the MagPred method.

The case study has been performed for various earthquakes that occurred in Japan and five in India or Nepal. On April 14, 2016 at 9:26 p.m., an  $M_W 6.3$  earthquake hit the Japanese prefecture of Kumamoto and neighboring regions with JMA seismic intensity category of 7 in Mashiki and 6 in Kumamoto, Tamana, Uki, and Nishihara locations [78]. **Fig. 10 (a)** represents the location of the epicenter and 48 stations near the  $M_W 6.0$  earthquake. **Fig. 10 (a)** also represents the overlapped contour plot obtained with single station magnitude prediction using the MagPred model. It is evident from **Fig. 10 (a)** that in most cases, the predicted magnitude is above  $M_W 6.0$ . **Fig. 10 (b), 10(c), and 10(d)** show the error (actual-predicted magnitude) in actual magnitude with respect to the predicted magnitude using MagPred for the SE3, SE4, and SE5 datasets.

**Fig. 11 (a)** shows the epicenter and station location of the mainshock of the  $M_W 5.6$  Osaka earthquake that occurred on June 18, 2018 (07:58 of Japan Standard Time) in Japan. The event caused the instrumental intensity in the epicentral area to reach a level of 6-lower on the 7-degree JMA scale [79]. **Fig. 11 (a)** also displays the

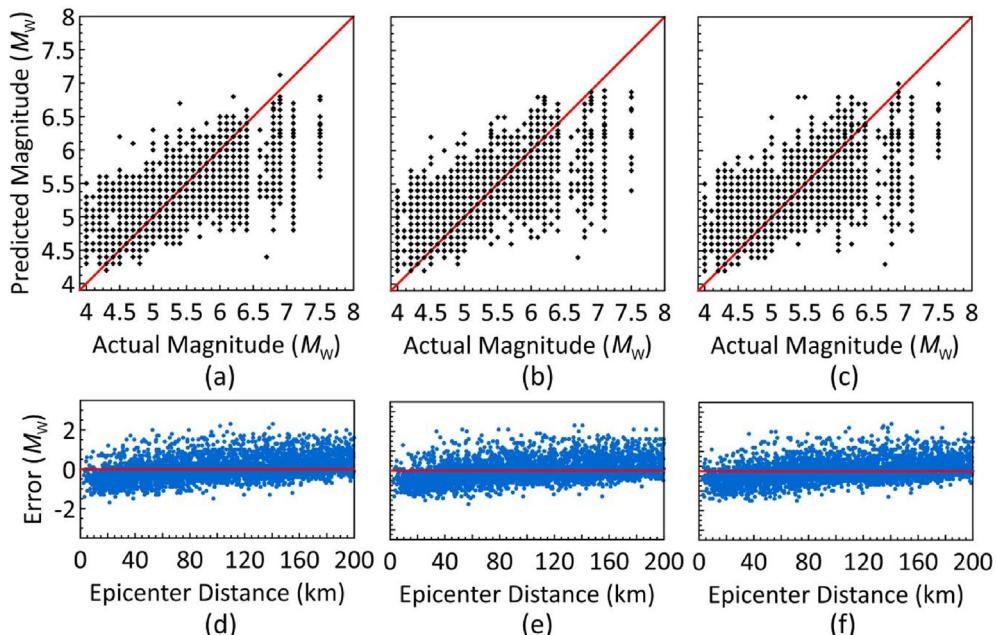
**Table 8**

Result of five earthquakes that occurred in India and Nepal region.

Region	$M_w$	Date	Depth (km)	Stations	IRE3			IRE4			IRE5		
					MAE	RMSE	Mean	MAE	RMSE	Mean	MAE	RMSE	Mean
Uttarkashi	4.7	21-09-2009	13	12	0.77	0.64	5.4	0.76	0.87	5.5	0.91	0.94	5.6
Sonipat	4.2	7-9-2011	8	7	1.28	1.29	5.4	1.29	1.31	5.5	1.37	1.4	5.6
Nepal	5.2	01-12-2006	10	17	0.32	0.38	5.6	0.31	0.37	5.6	0.31	0.39	5.6
India-Nepal	5.7	04-04-2011	10	16	0.28	0.34	5.7	0.3	0.35	5.7	0.28	0.36	5.9
Chamoli	4	29-11-2015	10	14	1.54	1.56	5.5	1.61	1.63	5.6	1.68	1.69	5.68

**Table 9**Result comparisons for multi-station weighted and unweighted magnitude prediction for specific earthquakes. The weighted and unweighted magnitude results are in  $M_{JMA}$  and the focal depth is in km.

Earthquake Name	$M_{JMA}$	Date	Focal Depth	Stations	Unweighted Magnitude ( $M_{JMA}$ )			Weighted Magnitude ( $M_{JMA}$ )		
					SE3	SE4	SE5	SE3	SE4	SE5
Kumamoto (Foreshock)	6.5	April 14, 2016	11	48	5.7	5.8	5.9	6.4	6.5	6.5
Kumamoto (Main)	7.3	April 16, 2016	12	45	6.3	6.4	6.6	6.8	7	7
Osaka Earthquake	6.1	June 18, 2018	13	58	5.6	5.6	5.6	6	6.4	6.6
Off-Shore	7.1	November 29, 2004	48	13	6.3	6.4	6.4	7.3	7	7
Off Fukushima	7.3	February 13, 2021	55	18	6.5	6.6	6.7	6.9	7	7

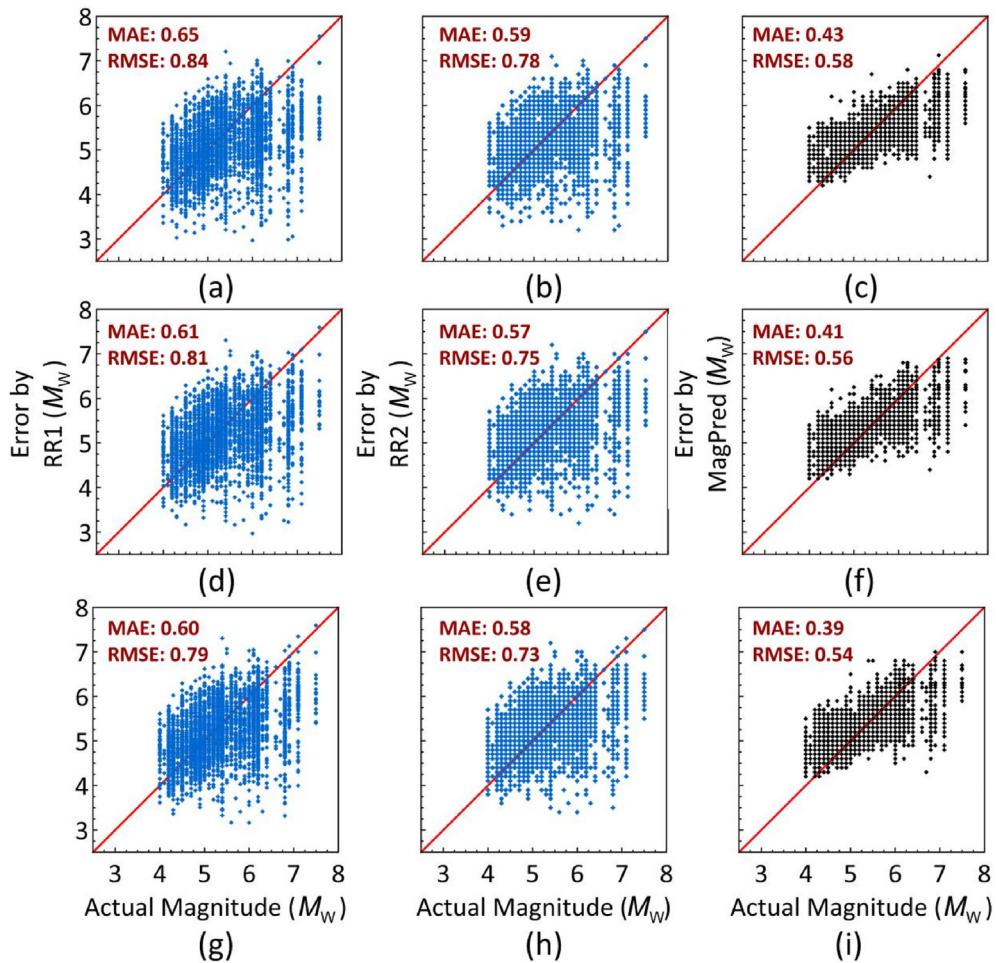
**Fig. 6.** The comparison of actual and predicted magnitude for (a) SE3, (b) SE4, and (c) SE5 test sets represented by black dots. The error (actual-predicted) with respect to the epicenter distance of (d) SE3, (e) SE4, and (f) SE5 test sets is shown by blue dots.

overlapped contour plot of predicted magnitude using the MagPred model. The error obtained from the magnitude prediction using MagPred concerning hypocenter distance has been shown in Fig. 11 (b-d) for SE3, SE4, and SE5 dataset. The  $\mu \pm \sigma$  of  $0.02 \pm 0.35$ ,  $-0.01 \pm 0.36$ , and  $-0.01 \pm 0.44 M_w$  has been obtained from all the stations for SE3, SE4, and SE5.

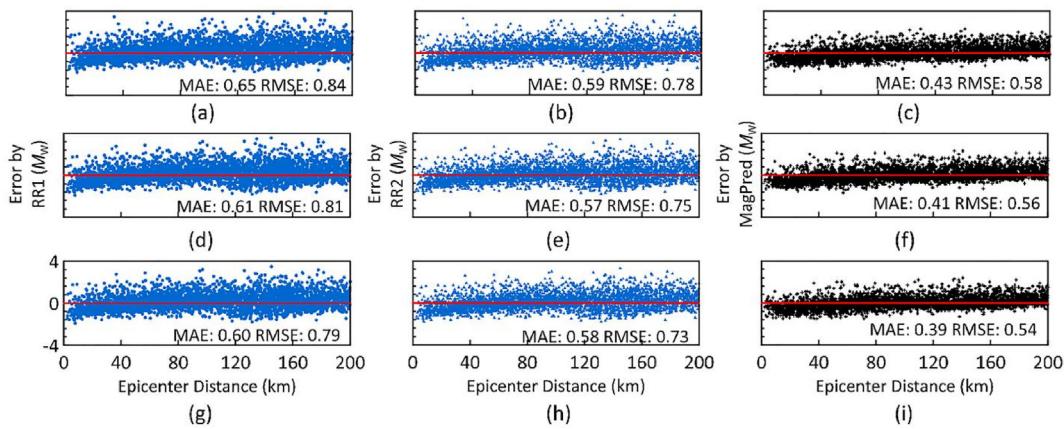
In this case study, an offshore earthquake with less impact is selected. Fig. 12 (a) shows the epicenter location and stations near the  $M_w 6.9$  earthquake. Fig. 12 (a) also displays the contour plot of predicted magnitude using the MagPred model based on single station responses. The error (predicted-actual magnitude) obtained from the magnitude prediction using MagPred concerning hypocenter distance has been shown in Fig. 12 (b-d) for SE3, SE4, and

SE5 datasets. The  $\mu \pm \sigma$  of  $0.58 \pm 0.58$ ,  $0.54 \pm 0.58$ , and  $0.61 \pm 0.54 M_w$  has been obtained from all the stations for SE3, SE4, and SE5.

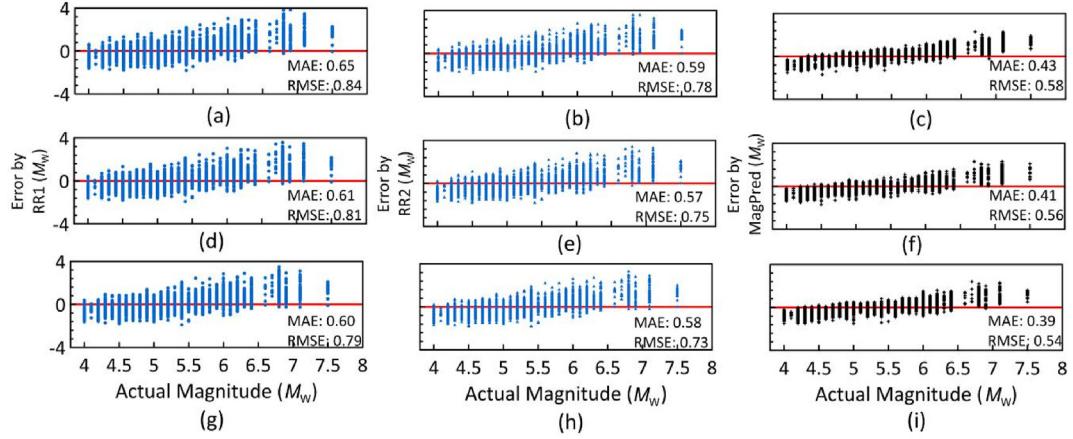
The station location and epicenter of the  $M_w 5.2$  India-Nepal earthquake that occurred on December 01, 2016 at 16:52 UTC in Nepal-India border is represented in Fig. 13 (a). It is a moderate inland earthquake that occurred in a completely different tectonic region. Fig. 13 (a) also displays the contour plot of predicted magnitude using the MagPred model based on single station responses. The error obtained from the magnitude prediction using MagPred with respect to hypocenter distance has been shown in Fig. 13 (b-d) for IRE3, IRE4, and IRE5 datasets. The mean error of  $-0.32 \pm 0.19$ ,  $-0.31 \pm 0.20$ , and  $-0.30 \pm 0.24 M_w$  has been obtained from all the stations. The analysis demonstrates the efficacy of the



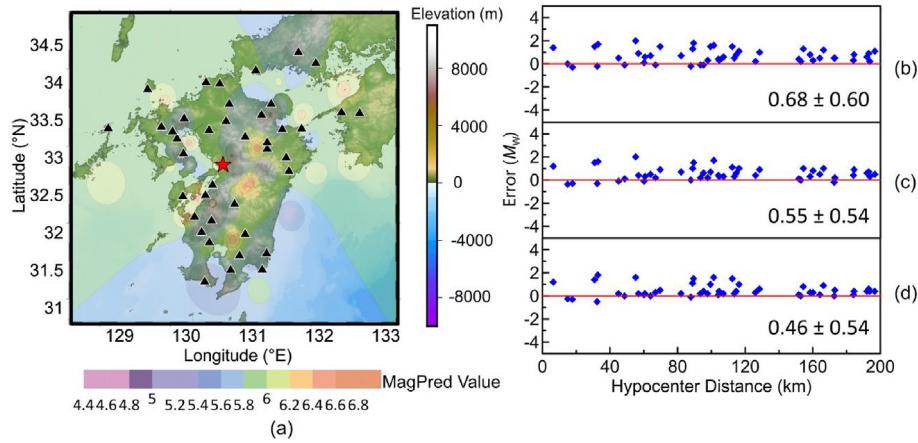
**Fig. 7.** The actual and predicted magnitude plot for the SE3 test set using magnitude predicted from  $P_{GD}$  regression relation given by (a) Kumar et al. [10], (b) Bhardwaj et al. [71], and (c) MagPred model. The actual and predicted magnitude plot for the SE4 test set using magnitude predicted from  $P_{GD}$  regression relation given by (d) Kumar et al. [10], (e) Bhardwaj et al. [71], and (f) MagPred model. The actual and predicted magnitude plot for the SE5 test set using magnitude predicted from  $P_{GD}$  regression relation given by (g) Kumar et al. [10], (h) Bhardwaj et al. [71], and (i) MagPred model. The magnitude is represented in  $M_w$  scale. RR1 and RR2 represent  $P_{GD}$  regression relation by Kumar et al. [10] and Bhardwaj et al. [71].



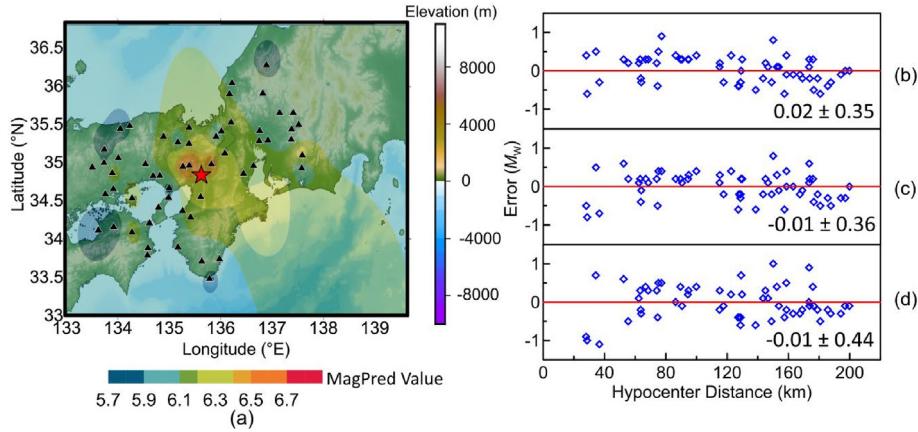
**Fig. 8.** The error with respect to epicenter distance using  $P_{GD}$  method by (a) Kumar et al. [10], (b) Bhardwaj et al. [71], and (c) MagPred method on SE3 dataset. The error with respect to epicenter distance using  $P_{GD}$  method by (d) Kumar et al. [10], (e) Bhardwaj et al. [71], and (f) MagPred method on SE4 dataset. The error with respect to epicenter distance using  $P_{GD}$  method by (g) Kumar et al. [10], (h) Bhardwaj et al. [71], and (i) MagPred method on SE5 dataset. RR1 and RR2 represent  $P_{GD}$  regression relation by Kumar et al. [10] and Bhardwaj et al. [71].



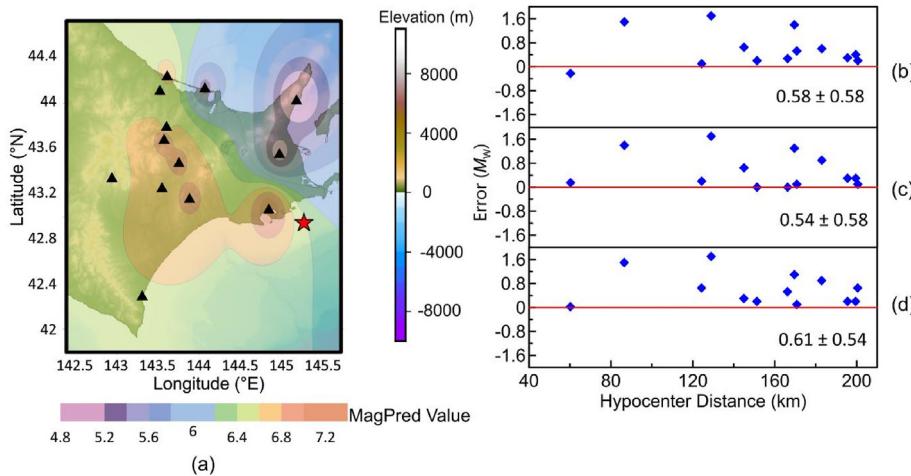
**Fig. 9.** The error with respect to magnitude using (a) Kumar et al. [10], (b) Bhardwaj et al. [71], and (c) MagPred method on SE3 dataset. The error with respect to magnitude using (d) Kumar et al. [10], (e) Bhardwaj et al. [71], and (f) MagPred method on SE4 dataset. The error with respect to magnitude using (g) Kumar et al. [10], (h) Bhardwaj et al. [71], and (i) MagPred method on SE5 dataset. RR1 and RR2 represent  $P_{GD}$  regression relation by Kumar et al. [10] and Bhardwaj et al. [71].



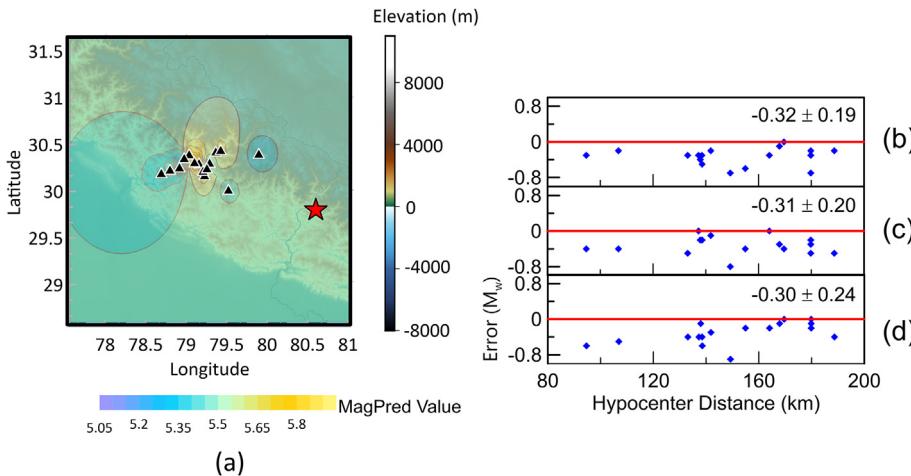
**Fig. 10.** (a) A geological map displaying the epicenter location (red star) and 48 stations (triangle) near Kumamoto  $M_w$  6.3 earthquake occurred on April 14, 2016 at 09:26 p.m. It is overlapped by the contour plot of MagPred predicted magnitude. The error plots illustrate the deviation between predicted and actual magnitudes concerning hypocenter distance on (b) SE3, (c) SE4, and (d) SE5 datasets using the MagPred method.



**Fig. 11.** (a) shows a geological map with the epicenter (red star) and 58 stations distribution (triangles) for the  $M_w$  5.6 Osaka earthquake on June 18, 2018. (b), (c), and (d) present error plots showing the deviation between predicted and actual magnitudes against hypocenter distance for the SE3, SE4, and SE5 datasets, respectively, using the MagPred method.



**Fig. 12.** (a) presents a geological map of the epicenter location and 13 stations near the  $M_w$  6.9 earthquake that occurred on November 29, 2004 at 03:32. It is overlapped by the predicted magnitude contour plot. The error plots of predicted magnitude with respect to hypocenter distance are shown using the (b) SE3, (c) SE4, and (d) SE5 datasets.



**Fig. 13.** The plot (a) presents a geological map of the epicenter location and 17 stations near the  $M_w$  5.2 earthquake that occurred on December 01, 2016 at 16:52 UTC. The error plots of predicted magnitude using the (b) IRE3, (c) IRE4, and (d) IRE5 dataset with respect to hypocenter distance are shown.

proposed method on single as well as multi-station earthquake events.

## 5. Discussion

Balancing the tradeoff between filling the data imbalance gap and maintaining the statistical properties of the original data is very important. The results from MagPred confirm that incorporating GANs to address data imbalance in earthquake datasets, along with machine learning for magnitude prediction, yields lower error rates. This approach outperforms models that do not account for data imbalance in earthquake datasets.

The primary purpose of the model is to predict the magnitude from the initial few seconds of the P wave of a single station. Various machine learning models are used in the study to avoid model overfitting. Tables 3–5 show the quantitative analysis for selecting the model for the final predicted magnitude. The model ablation study with respect to other machine learning models apart from the selected ones is shown in Table 6, which justifies the selection of MagPred meta-models.

The mean error of  $0.03 \pm 0.56$ ,  $0.01 \pm 0.54$ , and  $0.01 \pm 0.53$   $M_{JMA}$  has been obtained for the test set of SE3, SE4, and SE5 datasets.

Fig. 7(a), 7(b), and 7(c) shows the ability of machine learning to predict magnitude with better results than the second and third best-performing methods. These methods are  $P_{GD}$  regression relation given by Kumar et al. [10] and Bhardwaj et al. [71]. Table 7 proves that the machine learning model outperforms the given traditionally used methods.

The testing of the dataset has been performed based on single and multi-station recordings. The study has been extended to assess the model's capability for multi-station magnitude prediction, focusing on several prominent historical earthquakes. These events are detailed in Table 9. The challenge that persists in generating synthetic datasets using GAN is the substantial computational resources. Another challenge involves that the generated dataset should follow the statistical properties of the original dataset [80]. The future direction of this study is the introduction of more sophisticated methodologies for covering the data imbalance problem.

## 6. Conclusion

A machine learning model named MagPred has been developed and applied to predict the magnitude of single-station earthquakes

in Japan, one of the most seismically active regions in the world. The data imbalance problem has been overcome using synthetic data generated from CTGAN. The study uses an ensemble of machine learning models, including SGWO-XGB, XGB, LGBM, RF, and CatB that are trained on real and augmented datasets. The meta-heuristic optimizer, such as Sanitized Grey Wolf Optimizer, has been used for hyperparameter optimization of the XGB algorithm. The input to the model is the carefully curated thirty-four parameters.

This study shows that addressing synthetic data problems for earthquake magnitude prediction can provide better results. The MAE of 0.42, 0.40, 0.38  $M_W$  have been obtained on testing data that considers only the initial three, four, and five seconds of the P waveform. It is evident from the literature survey that increasing the time periods of the P wave for calculating magnitude will decrease the error, as can be seen from the proposed method results analysis. However, this increase is not very high. Thus, machine-learning algorithms can provide approximately the same accuracy with 3 s compared with 5 s of P wave to provide more lead-time.

The error comparison was conducted for single station-based magnitude prediction using the MagPred model, and regression relations were obtained for this study. The comparison shows that artificial intelligence techniques can replace traditional regression relations that are dependent on a few parameters. The multi-station earthquake study was performed using the weighted and unweighted magnitude. The weighted result gives better performance. The proposed model outperforms the selected comparison models. Furthermore, the study has been conducted on this dataset's recent state-of-the-art model given by Zhu et al. [11] and proposed regression relation of  $\tau_c$ , and  $P_{GD}$ . This observation shows that the GANs and machine learning can be applied for magnitude prediction and solving data imbalance problems in earthquake magnitude prediction.

#### CRediT authorship contribution statement

**Anushka Joshi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Balasubramanian Raman:** Writing – review & editing, Visualization, Supervision. **C. Krishna Mohan:** Visualization, Supervision.

#### Code availability

##### Name of the code: MagPred.

**Developers:** Anushka Joshi, Balasubramanian Raman and C Krishna Mohan; contact address: Machine Vision and Intelligence Labs, IIT Roorkee, Department of Computer Science and Engineering, IIT Roorkee, Uttarakhand, India, 247,667.

**First available year:** 2024.

**Contact address:** Machine Vision & Intelligence Labs, Indian Institute of Technology, Roorkee, Uttarakhand, India. E-mail: [anushka\\_j@cs.iitr.ac.in](mailto:anushka_j@cs.iitr.ac.in).

**Program language:** Python 3.0.

**Program size:** 24 kb.

The source codes along with the test dataset are available for download at the link: <https://github.com/anushka-joshi/MagPred>.

#### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Anushka Joshi reports that financial support was provided by Indian Institute of Technology Roorkee. Code availability (software application or custom code): Code used in this paper has

been developed in Python and the code github repository link is attached to the manuscript. Ethics Approvals: The work presented in this paper does not involve research using humans and/or animals. Availability of data and material: The data used in this research work is taken from National Research Institute for Earth Science and Disaster Resilience (NIED).

#### Acknowledgments

This work is related to grant PM-31-22-626-414 from the Prime Minister's Research Fellows (PMRF) of the Indian Institute of Technology Roorkee.

#### References

- [1] Y. Wu, L. Zhao, Magnitude estimation using the first three seconds p-wave amplitude in earthquake early warning, *Geophys. Res. Lett.* 33 (2006), <https://doi.org/10.1029/2006GL026871>.
- [2] R.V. Allen, Automatic earthquake recognition and timing from single traces, *Bull Seismol Soc Am* 68 (1978) 1521–1532, <https://doi.org/10.1785/BSSA0680051521>.
- [3] H. Mittal, Y. Wu, M.L. Sharma, B.M. Yang, S.C.D. Gupta, Testing the performance of earthquake early warning system in northern India, *Acta Geophys.* 67 (2018) 59–75, <https://doi.org/10.1007/s11600-018-0210-6>.
- [4] Y. Wu, H. Kanamori, R.M. Allen, E. Hauksson, Determination of earthquake early warning parameters,  $\tau_c$  and  $P_d$ , for southern California, *Geophys J Int* 170 (2007) 711–717, <https://doi.org/10.1111/j.1365-246X.2007.03430.x>.
- [5] H.M. Brown, R.M. Allen, V.F. Grasso, Testing elarms in Japan, *Seismol Res. Lett.* 80 (2009) 727–739, <https://doi.org/10.1785/gssrl.80.5.727>.
- [6] M. Olivieri, R.M. Allen, G. Wurman, The potential for earthquake early warning in Italy using elarms, *Bull Seismol Soc Am* 98 (2008) 495–503, <https://doi.org/10.1785/0120070054>.
- [7] J.-T. Shieh, Y. Wu, R.M. Allen, A comparison of  $\tau_c$  and  $\tau_{Pmax}$  for magnitude estimation in earthquake early warning, *Geophys. Res. Lett.* 35 (2008), <https://doi.org/10.1029/2008GL035611>.
- [8] Y. Wu, H. Kanamori, Experiment on an onsite early warning method for the taiwan early warning system, *Bull Seismol Soc Am* 95 (2005) 347–353, <https://doi.org/10.1785/0120040097>.
- [9] Y.M. Wu, H. Mittal, Y.H. Lin, Y.H. Chang, Magnitude determination using cumulative absolute absement for earthquake early warning, *Geosci. Letture* 10 (2023) 1–7, <https://doi.org/10.1186/s40562-023-00314-6>.
- [10] S. Kumar, H. Mittal, K.S. Roy, Y. Wu, R. Chaubey, A.P. Singh, Development of earthquake early warning system for kachchh, Gujarat, in India using  $\tau_c$  and  $P_d$ , *Arabian J. Geosci.* 13 (2020), <https://doi.org/10.1007/s12517-020-05353-3>.
- [11] J. Zhu, S. Li, J. Song, Magnitude estimation for earthquake early warning with multiple parameter inputs and a support vector machine, *Seismol Res. Lett.* 93 (2022) 126–136, <https://doi.org/10.1785/0220210144>.
- [12] J. Song, J. Zhu, S. Li, Meanet: magnitude estimation via physics817 based features time series, an attention mechanism, and neural networks, *Geophys* 88 (2023) 33–43, <https://doi.org/10.1190/geo2022-0196.1>.
- [13] J. Zhu, S. Li, J. Song, Y. Wang, Magnitude estimation for earthquake early warning using a deep convolutional neural network, *Front Earth Sci* 9 (2021), <https://doi.org/10.3389/feart.2021.653226>.
- [14] A. Joshi, V. Chalavadi, K. Mohan, Early detection of earthquake magnitude based on stacked ensemble model, *J. Asian Earth Sci.* X 8 (2022) 100122, <https://doi.org/10.1016/j.jaesx.2022.100122>.
- [15] A. Joshi, B. Raman, C.K. Mohan, An integrated approach for prediction of magnitude using deep learning techniques, *Neural Comput Appl* 36 (2024) 16991–17006, <https://doi.org/10.1007/s00521-024-09891-9>.
- [16] A. Joshi, V. Chalavadi, C.K. Mohan, B. Raman, Application of xgboost model for early prediction of earthquake magnitude from waveform data, *J. Earth Syst Sci* 133 (2023) 1–18, <https://doi.org/10.1007/s12040-023-02210-1>.
- [17] A. Joshi, P. Singh, B. Raman, A deep attention model for onsite estimation of earthquake epicenter distance and magnitude, *IEEE Trans. Geosci. Remote Sens.* 62 (2024) 1–11, <https://doi.org/10.1109/TGRS.2024.3459425>.
- [18] S.M. Mousavi, G.C. Beroza, A machine-learning approach for earth841 quake magnitude estimation, *Geophys. Res. Lett.* 47 (2019), <https://doi.org/10.1029/2019GL085976>.
- [19] T. Perol, M. Gharbi, M.A. Denolle, Convolutional neural network for earthquake detection and location, *Sci. Adv.* 4 (2018), <https://doi.org/10.1126/sciadv.1700578>.
- [20] A. Joshi, B. Raman, C.K. Mohan, L.R. Cenkeramaddi, Application of a new machine learning model to improve earthquake ground motion predictions, *Nat Hazards* 170 (2023) 729–753, <https://doi.org/10.1007/s11069-023-06230-4>.
- [21] A. Joshi, S.K. Peddoju, M. Pandey, Simulated annealing algorithm for subsurface shear wave velocity investigation using ground vibration data, in: *Local site effects and ground failures*, Springer Singapore, Singapore, 2021, pp. 23–38, [https://doi.org/10.1007/978-981-15-9984-2\\_3](https://doi.org/10.1007/978-981-15-9984-2_3).
- [22] G. He, Y. Zhao, C. Yan, Application of tabular data synthesis using generative adversarial networks on machine learning-based multiaxial fatigue life

- prediction, Int. J. Pres. Ves. Pip. 199 (2022) 104779, <https://doi.org/10.1016/j.ijipvp.2022.104779>.
- [23] Z. Mushtaq, S. Su, Environmental sound classification using a regular861 ized deep convolutional neural network with data augmentation, Appl Acoust 167 (2020) 107389, <https://doi.org/10.1016/j.apacoust.2020.107389>.
- [24] F. Chen, N. Wang, J. Tang, D. Liang, H. Feng, Self-supervised data augmentation for person re-identification, Neurocomputing 415 (2020) 48–59, <https://doi.org/10.1016/j.neucom.2020.07.087>.
- [25] M. Frid-Adar, E. Klang, M.M. Amitai, J. Goldberger, H. Greenspan, Synthetic data augmentation using gan for improved liver lesion classification, 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (2018) 289–293.
- [26] X. Peng, Z. Tang, F. Yang, R.S. Feris, D.N. Metaxas, Jointly optimize data augmentation and network training: adversarial data augmentation in human pose estimation, in: 2018 IEEE/CVF conference on computer vision and pattern recognition, 2018, pp. 2226–2234, <https://doi.org/10.1109/CVPR.2018.00237>.
- [27] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, 2015, [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28). ArXiv abs/1505.04597.
- [28] N. Chawla, K. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority over-sampling technique, 2002. ArXiv abs/1106.1813.
- [29] H. Inoue, Data augmentation by pairing samples for images classification, 2018, <https://doi.org/10.3390/jimaging9020046>. ArXiv 9.
- [30] H. Zhang, M. Cisse, Y. Dauphin, D. Lopez-Paz, mixup: beyond empirical risk minimization, 2017. ArXiv abs/1710.09412.
- [31] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.C. Courville, Y. Bengio, Generative adversarial nets, NIPS, 2014, pp. 2672–2680.
- [32] E.D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q.V. Le, Autoaugment: learning augmentation strategies from data, in: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2019, pp. 113–123doi, <https://doi.org/10.1109/CVPR2019.00020>.
- [33] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, Infogan: interpretable representation learning by information maximiz895 ing generative adversarial nets, in: NIPS, 2016, pp. 2180–2188, doi:89610.5555/3157096.3157340.
- [34] J.Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: 2017 IEEE international conference on computer vision (ICCV), 2017, pp. 2242–2251, <https://doi.org/10.1109/ICCV.2017.244>.
- [35] S. Nowozin, B. Cseke, R. Tomioka, f-gan: training generative neural samplers using variational divergence minimization, in: NIPS, Curran Associates Inc., Red Hook, NY, USA, 2016, pp. 271–279, <https://doi.org/10.5555/3157096.3157127>.
- [36] C. Li, M. Wand, Precomputed real-time texture synthesis with marko906 via generative adversarial networks, in: Computer vision - eccv 2016, Springer International Publishing, Cham, 2016, pp. 702–716.
- [37] L. Tran, X. Yin, X. Liu, Representation learning by rotating your faces, IEEE Trans Pattern Anal Mach Intell 41 (2017) 3007–3021.
- [38] L. Xu, K. Veeramachaneni, Synthesizing tabular data using generative adversarial networks, 2018. ArXiv abs/1811.11264.
- [39] E. Choi, S. Biswal, B.A. Malin, J.D. Duke, W.F. Stewart, J. Sun, Generating multi-label discrete patient records using generative adversarial networks, in: Machine learning in health care, 2017, pp. 286–305.
- [40] Z. Zhao, A. Kunar, H. vanderScheer, R. Birke, LY. Chen, Ctabgan: effective table data synthesizing, in: Proceedings of machine learning research -acml 2021, 2021.
- [41] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Software 69 (2014) 46–61, <https://doi.org/10.1016/j.advengsoft.2013.12.007>.
- [42] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, <https://doi.org/10.1145/2939672.2939785>.
- [43] L. Ostroumova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, Cat927 boost: unbiased boosting with categorical features, in: Neural information processing systems, 2017, pp. 6639–6649, <https://doi.org/10.5555/3327757.3327770>.
- [44] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.Y. Liu, Lightgbm: a highly efficient gradient boosting decision tree, in: NIPS, 2017, pp. 3149–3157, <https://doi.org/10.5555/3294996.3295074>.
- [45] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
- [46] H. Kanamori, Real-time seismology and earthquake damage mitigation, Annu. Rev. Earth Planet Sci. 33 (2005) 195–214, <https://doi.org/10.1146/annurev.earth.33.092203.122626>.
- [47] Y. Wu, H. Kanamori, Rapid assessment of damage potential of earthquakes in taiwan from the beginning of p waves, Bull Seismol Soc Am 95 (2005) 1181–1185, <https://doi.org/10.1785/0120040193>.
- [48] T. Sato, T. Hirasawa, Body wave spectra from propagating shear cracks, J. Phys. Earth 21 (1973) 415–431, <https://doi.org/10.4294/jpe1952.21.415>.
- [49] S. Noda, W.L. Ellsworth, Scaling relation between earthquake magnitude and the departure time from p wave similar growth, Geophys Res Lett 43 (2016) 9053–9060, <https://doi.org/10.1002/2016GL070069>.
- [50] L.J. Ratliff, S.A. Burden, S.S. Sastry, Characterization and computation of local nash equilibria in continuous games, in: 2013 51st annual allerton conference on communication, control, and computing (allerton), 2013, pp. 917–924, <https://doi.org/10.1109/Allerton.2013.6736623>.
- [51] O. Habibi, M. Chemmakha, M. Lazaar, Imbalanced tabular data modelization using ctgan and machine learning to improve iot botnet attacks detection, Eng Appl Artif Intell 118 (2023) 105669, <https://doi.org/10.1016/j.engappai.2022.105669>.
- [52] M. Mujahid, E. Kina, F. Rustam, M.G. Villar, E.S. Alvarado, I. de la Torre Diez, I. Ashraf, Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering, J. Big Data 11 (2024) 1–32, <https://doi.org/10.1186/s40537-024-00943-4>.
- [53] C.M. Bishop, Information science and statistics, Springer, New York, NY, 2006.
- [54] S. Nematzadeh, F. Kiani, M. Torkamanian-Afshar, N. Aydin, Tuning hyper-parameters of machine learning algorithms and deep neural networks using metaheuristics: a bioinformatics study on biomedical and biological cases, Comput Biol Chem 97 (2022) 107619, <https://doi.org/10.1016/j.combiolchem.2021.107619>.
- [55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, J Mach Learn Res 12 (2011) 2825–2830.
- [56] NIED, Nied k-net, kik-net, national research institute for earth science and disaster resilience, National Research Institute for Earth Science and Disaster Resilience, 2019, <https://doi.org/10.17598/NIED.0004>.
- [57] L. Jing, X. Zhuo, X. Wang, Effect of complex site on seismic wave propagation, Earthq. Eng. Struct. Dynam. 25 (2005) 16–23.
- [58] F. Meng, T. Ren, Z. Liu, Z. Zhong, Toward earthquake early warning: a convolutional neural network for rapid earthquake magnitude estimation, Artificial Intelligence in Geosciences 4 (2023) 39–46, <https://doi.org/10.1016/j.jaig.2023.03.001>.
- [59] K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, London, Edinburgh Dublin Phil. Mag. J. Sci. 50 (1900) 157–175, <https://doi.org/10.1080/1478640009463897>.
- [60] M. Zare, P.Y. Bard, Strong motion dataset of Turkey: data process989 ing and site classification, Soil Dynam. Earthq. Eng. 22 (2002) 703–718, [https://doi.org/10.1016/S0267-261\(02\)000283](https://doi.org/10.1016/S0267-261(02)000283).
- [61] M.T. Thompson, Chapter 14 – analog low-pass filters, in: Intuitive analog circuit design, 2006, pp. 531–583.
- [62] Y. Nakamura, On the urgent earthquake detection and alarm system (uredas), Proc Ninth World Conf Earthq Eng 7 (1988) 673–678, [https://doi.org/10.1007/978-3-540-72241-0\\_13](https://doi.org/10.1007/978-3-540-72241-0_13).
- [63] Y. Wu, L. Zhao, Magnitude estimation using the first three seconds p wave amplitude in earthquake early warning, Geophys. Res. Lett. 33 (2006) L16312, <https://doi.org/10.1029/2006GL026871>.
- [64] Y. Nakamura, A new concept for the earthquake vulnerability estimation and its application to the early warning system, in: early Warning Syst, Natural Disaster Reduct (2003) 693–699, [https://doi.org/10.1007/978-3-642-55903-7\\_92](https://doi.org/10.1007/978-3-642-55903-7_92).
- [65] G. Festa, A. Zollo, M. Lancieri, Earthquake magnitude estimation from early radiated energy, Geophys. Res. Lett. 35 (2008), <https://doi.org/10.1029/2008GL035576>.
- [66] A. Sinvhal, K.N. Khattri, Application of seismic reflection data to discriminate subsurface lithostratigraphy, Geophysics 48 (1983) 1498–1513, <https://doi.org/10.1119/1.1441433>.
- [67] E. Scordilis, Globally valid relations converting ms, mb and mjma to mw, in: earthquake monitoring and seismic hazard mitigation in balkan countries, Nato Advanced Research Workshop on “Earthquake Monitoring and Seismic Hazard Mitigation in Balkan Countries” (2005) 158–161.
- [68] S.O. Arik, T. Pfister, Tabnet: attentive interpretable tabular learning, in: Proceedings of the AAAI conference on artificial intelligence 35, 2021, pp. 6679–6687, <https://doi.org/10.1609/aaai.v35i8.16826>.
- [69] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull Math Biol 52 (1990) 99–115, <https://doi.org/10.1007/BF02478259>.
- [70] N. Cristianini, E. Ricci, Encyclopedia of algorithms, Springer US, Boston, MA, ch, Support Vector Machines (2008) 928–932, [https://doi.org/10.1007/978-0-387-30162-4\\_415](https://doi.org/10.1007/978-0-387-30162-4_415).
- [71] R. Bhardwaj, A. Kumar, M.L. Sharma, Analysis of tau(c) and pd attributes for earthquake early warning in India, in: 15th World conference on earthquake engineering, 2012, pp. 1–8.
- [72] Y. Wu, H. Yen, L. Zhao, B. Huang, W. Liang, Magnitude determination using initial p waves: a single-station approach, Geophys Res Lett 33 (2006), <https://doi.org/10.1029/2005GL025395>.
- [73] G. Wurman, R.M. Allen, P.N. Lombard, Toward earthquake early warning in northern California, J Geophys Res 112 (2007), <https://doi.org/10.1029/2006JB004830>.
- [74] Y.M. Wu, H. Kanamori, Development of an earthquake early warning system using real-time strong motion signals, Sens. 8 (2008) 1–9, <https://doi.org/10.3390/s801001>.
- [75] A. Zollo, O. Amoroso, M. Lancieri, Y. Wu, H. Kanamori, A threshold based earthquake early warning using dense accelerometer networks, Geophys J Int 183 (2010) 963–974, <https://doi.org/10.1111/j.1365-246X.2010.04765.x>.
- [76] X. Jin, H. Zhang, J. Li, Y. Wei, Q. Ma, Earthquake magnitude estimation using the tau c and pd method for earthquake early warning systems, Earthq Sci 26 (2013) 23–31, <https://doi.org/10.1007/s11589-013-0005-4>.
- [77] H. Kanamori, E. Brodsky, The physics of earthquakes, Rep Prog Phys 67 (2004) 1429–1496, <https://doi.org/10.1088/0034-4885/67/8/R03>.

- [78] N. Achour, M. Miyajima, Post-earthquake hospital functionality evaluation: the case of kumamoto earthquake 2016, *Earthq Spectra* 36 (2020) 1670–1694. <https://doi.org/10.1177/8755293020926180>.
- [79] M. Hallo, I. Oprsal, K. Asano, F. Gallovic, Seismotectonics of the 2018 northern osaka m6.1 earthquake and its aftershocks: joint movements on strike-slip and reverse faults in inland Japan, *Earth Planets Space* 71 (2019) 1–21. <https://doi.org/10.1186/s40623-019-1016-8>.
- [80] M. Miletic, M. Sariyar, Challenges of using synthetic data generation methods for tabular microdata, *Appl Sci* 14 (2024) 5975–5988, <https://doi.org/10.3390/app14145975>.



**Anushka Joshi** received her M.Tech. in Computer Science and Engineering from the Indian Institute of Technology (IIT) Hyderabad in 2022. She is currently pursuing a Ph.D. degree in Computer Science and Engineering at the Indian Institute of Technology Roorkee. Her research interests include deep learning, machine learning, and earthquake engineering. Ms. Joshi is a recipient of the prestigious Prime Minister's Research Fellowship (PMRF).



**Balasubramanian Raman** (Senior Member, IEEE) received his Ph.D. from IIT Madras. He is a Chair Professor in the Department of Computer Science and Engineering and a Joint Faculty in the Mehta Family School of Data Science

and Artificial Intelligence at IIT Roorkee. With over 200 research papers published in reputed journals and conferences, his research interests span Machine Learning, Image and Video Processing, Computer Vision, and Pattern Recognition. He has served as a Guest Professor and Visiting Researcher at prestigious institutions such as Osaka Metropolitan University, Curtin University, and the University of Cyberjaya. He has held postdoctoral positions at Rutgers University and the University of Missouri-Columbia. Under his coaching, teams have achieved notable rankings in the ACM International Collegiate Programming Contest (ICPC) World Finals. He has been recognized with several awards, including the BOYSCAST Fellowship and the Ramkumar Prize for Outstanding Teaching and Research.



**C Krishna Mohan** (Senior Member, IEEE) received the B.Sc.Ed. Degree from the Regional Institute of Education, India, in 1988, the M.C.A. degree from the S. J. College of Engineering, India, in 1991, the M.Tech. degree in system analysis and computer applications from the National Institute of Technology Surathkal, India, in 2000, and a Ph.D. in computer science and engineering from the Indian Institute of Technology Madras, India, in 2007. He is a Professor at the Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad, India. His research interests include video content analysis, pattern recognition, and neural networks.