Recent Advances in Early Earthquake Magnitude Estimation by Using Machine Learning Algorithms: A Systematic Review

Article *in* Applied Sciences · March 2025

DOI: 10.3390/app15073492

applied sciences

MDPI

Systematic Review

Recent Advances in Early Earthquake Magnitude Estimation by Using Machine Learning Algorithms: A Systematic Review

Andrés Navarro-Rodríguez [1,2], Oscar Alberto Castro-Artola [2], Enrique Efrén García-Guerrero [1], Oscar Adrian Aguirre-Castro [1], Ulises Jesús Tamayo-Pérez [1], César Alberto López-Mercado [1] and Everardo Inzunza-Gonzalez [1,*]

[1] Facultad de Ingeniería, Arquitectura y Diseño, Universidad Autónoma de Baja California, Carrt. Tijuana-Ensenada No. 3917, Ensenada 22860, Baja California, Mexico; navarro.andres@uabc.edu.mx (A.N.-R.); eegarcia@uabc.edu.mx (E.E.G.-G.); oscar.aguirre@uabc.edu.mx (O.A.A.-C.); ulises.tamayo@uabc.edu.mx (U.J.T.-P.); lopez.cesar7@uabc.edu.mx (C.A.L.-M.)

[2] Centro de Investigación Científica y de Educación Superior de Ensenada, Ensenada 22860, Baja California, Mexico; oscar.castro@cicese.mx

* Correspondence: einzunza@uabc.edu.mx

Abstract: Earthquakes are among the most destructive natural phenomena, leading to significant loss of human life and substantial economic damage that severely impacts affected communities. Rapid detection and characterization of seismic parameters, including location and magnitude, are crucial for real-time seismological applications, including Earthquake Early Warning (EEW) systems. Machine learning (ML) has emerged as a powerful tool to enhance the accuracy of these applications, enabling more efficient responses to seismic events of different magnitudes. This systematic review aims to provide researchers and professionals with a summary of the current state of ML applications in seismology, particularly on early earthquake magnitude estimations and related topics such as earthquake detection and seismic phase identification. A systematic search was conducted in Scopus, ScienceDirect, IEEE Xplore, and Web of Science databases, covering the period from early 2014 to 7 March 2025. The search terms included the following: (“earthquake magnitude” OR “earthquake early warning”) AND (prediction OR forecasting OR estimation OR forecast OR classification) AND (“machine learning” OR “deep learning” OR “artificial intelligence”). Out of the 472 articles initially identified, 28 were selected based on pre-defined inclusion criteria. The described methods and algorithms illustrate the strong performance of ML in earthquake magnitude estimation despite limited implementation in real-time systems. This highlights the need to develop standardized benchmark datasets to promote future progress in this field.

Keywords: earthquake magnitude estimation; early warning systems; real-time applications; artificial intelligence; machine learning; deep learning

check for updates

Academic Editors: Ke Jia, Wenhuan Kuang, Kai Deng and Shiyong Zhou

Received: 23 February 2025
Revised: 18 March 2025
Accepted: 19 March 2025
Published: 22 March 2025

Citation: Navarro-Rodríguez, A.; Castro-Artola, O.A.; García-Guerrero, E.E.; Aguirre-Castro, O.A.; Tamayo-Pérez, U.J.; López-Mercado, C.A.; Inzunza-Gonzalez, E. Recent Advances in Early Earthquake Magnitude Estimation by Using Machine Learning Algorithms: A Systematic Review. *Appl. Sci.* **2025**, *15*, 3492. https://doi.org/10.3390/app15073492

1. Introduction

Natural disasters result in significant casualties, damage, and injuries. Although humans cannot prevent them, timely prediction and appropriate safety measures can help save lives and protect valuable assets [1]. Earthquakes are among the most catastrophic natural events and many scientists have concluded that their prediction is currently unfeasible [2]. An EEW system can significantly reduce both the casualties and the economic losses caused by earthquakes [3]. Rapid identification and evaluation of seismic source parameters, such as magnitude, are essential for real-time seismological applications [4]. Magnitude provides the public with critical information about earthquakes and is also used for scientific investigations. Since Charles F. Richter proposed the magnitude scale, it has been known as the Richter magnitude or local magnitude (M_L) [5]. Today, there are various magnitude scales, such as body wave magnitude (M_b) and surface wave magnitude (M_S), which are empirical and saturate at different magnitude values [6]. Magnitude values measured on different scales can vary by more than one unit, especially for large earthquakes due to saturation effects. Even for the same type of magnitude, the values reported by different seismic agencies can vary by up to 0.5 units [7].

Currently, methods such as the predominant period (τ_p) [8], the effective average period (τ_c) [9], and the peak displacement (P_d) [10], which use the first 3 s of the P-wave, have demonstrated high reliability for magnitude estimation and are applied in EEW systems in various parts of the world [11].

On the other hand, artificial intelligence (AI), particularly ML and deep learning (DL), has experienced accelerated growth in recent years, specifically in data analysis and computing, facilitating the efficient operation of applications in various domains [12]. Today, ML is used in object recognition, speech-to-text conversion, and product recommendations, among others [13]. ML has a long history with seismology but has not yet been applied as widely as it is today. Early studies used supervised learning for seismic data processing, using labeled training datasets to establish relationships between input and output variables [14]. These studies were limited by several factors, such as the amount of available data, computational capacities, and algorithmic constraints, which prevented the training of high-performance models. These barriers have been overcome with the emergence of DL, which is based on essential nonlinear mathematical functions applied to raw data, making some models more efficient at performing classification or regression tasks.

In the field of seismology, the exponential growth in data availability and computational processing has enabled the application of ML and DL algorithms to various tasks, including event detection [15–20], phase detection [21–26], ground motion estimation [27–30], and magnitude estimation [3,4,7,31,32], among other applications.

To address the growing demands of seismic data analysis, various deep learning architectures have been employed to enhance the accuracy and efficiency of magnitude estimation. Among them, convolutional neural networks (CNNs) have been widely adopted due to their ability to extract spatial and frequency-based features from seismic waveforms [33]. These networks capture local waveform patterns by applying convolutional layers, making them particularly effective for event classification and magnitude estimation. Additionally, integrating CNNs with attention mechanisms (AMs) improves feature extraction for large-magnitude events, which are often underrepresented in training datasets.

While CNNs effectively capture spatial features, seismic signals also contain strong temporal dependencies that require specialized architectures. Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM) networks, have been employed to model these temporal relationships [4]. Unlike CNNs, which analyze spatial structures, RNNs process continuous time series data, making them well-suited for magnitude estimation and early earthquake classification. These architectures can capture long-term dependencies in P-wave signals, which are crucial for rapid magnitude prediction. Furthermore, multi-station LSTM networks have been developed to aggregate data from multiple seismic sensors, improving regional magnitude estimation and reducing prediction uncertainty.

More recently, Transformer-based models have emerged as alternatives to both CNNs and RNNs for seismic applications. Unlike previous architectures, Transformers utilize self-attention mechanisms to learn complex dependencies across entire seismic records [33].

This capability makes them highly effective for early magnitude estimation and event classification. For instance, the DFTQuake [34] model incorporates Fourier-based attention layers to improve the accuracy of magnitude estimation by capturing frequency-domain correlations in seismic signals. Additionally, techniques such as low-rank adaptation (LoRA) [33] have been employed to fine-tune Transformer architectures for regional seismic datasets, enhancing their adaptability while maintaining computational efficiency. These improvements offer a promising avenue for real-time seismic analysis, ensuring scalability and generalization across different tectonic environments.

Recent advancements in computational capabilities, particularly in GPU-accelerated methodologies, have enabled large-scale simulations for seismic analysis and structural response evaluation [35]. The integration of such high-performance computing strategies with ML-based magnitude estimation can improve both prediction accuracy and response time, facilitating the deployment of EEW systems in high-risk zones.

Beyond magnitude estimation, ML and DL models are also being explored to assess secondary earthquake-related hazards. For example, a spatiotemporal casualty assessment method that evaluates the impact of earthquake-caused debris from building clusters, considering human emergency behaviors, has been proposed [36]. These approaches integrate real-time seismic data with urban infrastructure models and human behavior simulations to enhance disaster risk assessments. Future research efforts may benefit from fusing ML-driven magnitude estimation with predictive models for secondary disaster impacts, ultimately contributing to more comprehensive earthquake response strategies.

The rapid evolution of ML and DL methodologies has demonstrated their potential for improving the accuracy and efficiency of earthquake magnitude estimation. However, despite these advancements, several challenges persist in fully integrating these techniques into operational real-time seismic monitoring systems. Issues such as data variability across different seismic networks, the need for extensive computational resources, and the generalization of models to diverse tectonic environments remain critical research areas.

This systematic review analyzes literature on the application of machine learning and deep learning techniques to early earthquake magnitude estimation. It is a valuable tool for researchers and professionals in this field. Additionally, it provides a comprehensive overview of recent advancements, compares relevant models, and outlines key challenges. Moreover, it highlights potential future research directions to enhance earthquake magnitude estimation and its integration into real-time seismic monitoring.

The remaining sections of the article are structured as follows: Section 2 describes the methods used in this work. Section 3 presents early earthquake magnitude estimation findings as a classification and regression task. Section 4 shows the trends and future work. Finally, in Section 5, the conclusions of this systematic review are presented.

2. Methodology

2.1. Search Strategy

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [37] to ensure a clear and transparent methodology. A bibliographic search was conducted covering the period from 2014 to 7 March 2025 in four major databases (Scopus, Elsevier, IEEE Xplore, and Web of Science) to identify relevant articles. The following keywords were used, combined with Boolean operators: ("earthquake magnitude" OR "earthquake early warning") AND (prediction OR forecasting OR estimation OR forecast OR classification) AND ("machine learning" OR "deep learning" OR "artificial intelligence").

In the Scopus database, the search fields used were as follows: Article Title, Abstract, and Keywords. Then, the following filters were applied: Document Type, Language, and Final Stage, specifically Article, English, and Final. In IEEE Xplore, the search field used was Journals & Magazines, followed by the Journals filter. In Web of Science, the search field used was Topic (Title, Abstract, Keyword Plus, and Author Keywords). Then, the following filters were applied: Document Type and Language, specifically Article and English. Finally, in Elsevier, the search fields used were Title, Abstract, or Author-Specified Keywords, and the Research Articles filter was applied.

It is important to note that Elsevier limits the use of Boolean operators to eight. Since nine operators were required in this study, the search was divided as follows:

- First, the following set of keywords was used: ("earthquake magnitude" OR "earthquake early warning") AND (prediction OR forecasting OR estimation OR forecast) AND ("machine learning" OR "deep learning" OR "artificial intelligence").
- Subsequently, an additional search was performed with the following keywords: ("earthquake magnitude" OR "earthquake early warning") AND (prediction OR forecasting OR estimation OR classification) AND ("machine learning" OR "deep learning" OR "artificial intelligence").
- Finally, duplicate articles from both searches were identified and removed, resulting in a unified set of relevant articles.

2.2. Screening and Eligibility Results

Initially, duplicate articles identified across the four databases were removed. Then, articles were excluded based on their title and abstract, and the following inclusion and exclusion criteria were applied:

Inclusion criteria:

- Articles focusing on machine learning methods for early earthquake magnitude estimation.
- Articles published in English.
- Articles published between 2014 and 7 March 2025.

Exclusion criteria:

- Conference papers.
- Systematic review articles.
- Articles in which machine learning was not applied to early earthquake magnitude estimation.

After applying these criteria, a complete analysis was performed on the selected articles. Data such as publication year, first author's name, primary research objective, database size, target variable, ML model, performance metrics, key findings, conclusions, and proposed future work were recorded. The PRISMA diagram used for this work is shown in Figure 1.

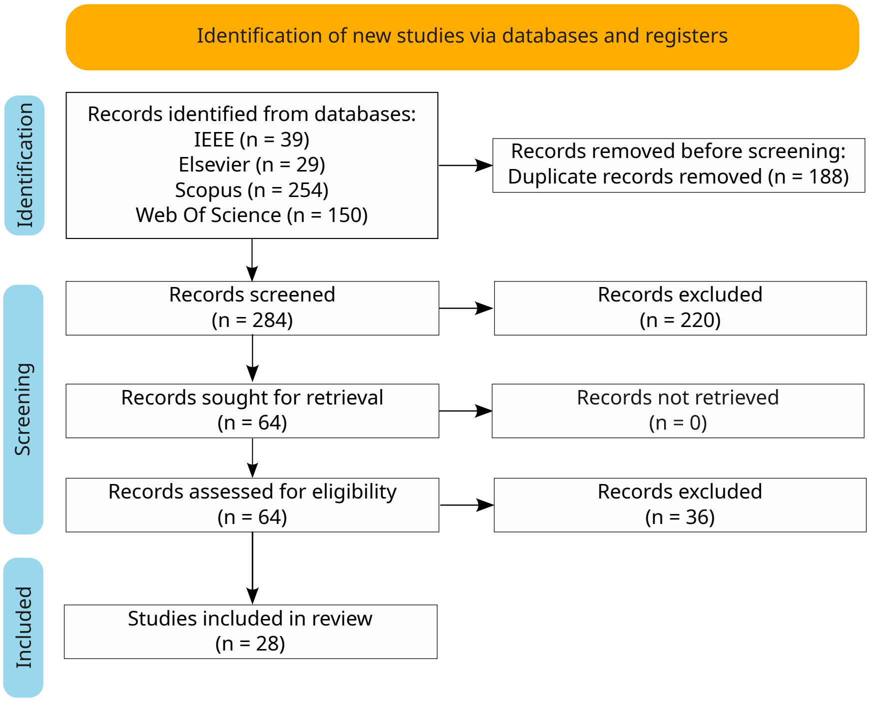

Figure 1. PRISMA diagram used in this systematic review [37].

3. Results

According to the bibliographic search conducted in the four databases, i.e., Scopus, Elsevier, IEEE Xplore, and Web of Science, 254, 29, 39, and 150 articles were found, respectively, totaling 472 works. Subsequently, 188 duplicates were identified and removed, leaving 284 articles for review.

Next, articles were selected based on their titles and abstracts, eliminating 220, resulting in 64 works for a full review. Of these, 28 articles met the inclusion criteria, with the most relevant being those that applied a machine learning method for early earthquake magnitude estimation.

Given the specific focus of this review, all the works were grouped into two categories—the first addresses the estimation of magnitude as a classification task, and the second addresses it as a regression task. Only five articles fall into the classification group, while the remaining 23 correspond to regression. Table 1 presents the relevant aspects of each work included in this review.

Figure 2 illustrates the number of articles published per year. It was observed that 2022 and early 2024 saw the highest publication rates, while 2020 and 2021 registered fewer publications. Two articles were published in 2025. While these data are included in the figure, it is important to note that the comparison may not be entirely fair, as the search was conducted only until March of this year.

Figure 3 shows the top eight countries that have contributed to publications on this topic. China leads the list, followed by India and Germany. Countries such as the United States, South Korea, Romania, Chile, and Japan have also made significant contributions to research on using machine learning in early earthquake magnitude estimation.

Figure 4 presents the journals in which the selected articles were published for this review. Examples include *IEEE Geoscience and Remote Sensing Letters* with five articles and *Geophysical Journal International* with three. Journals such as *Artificial Intelligence in Geosciences*, *Seismological Research Letters*, and *Bulletin of the Seismological Society of America* each contributed two articles. Table A1 provides a list of journals with one article in this review.

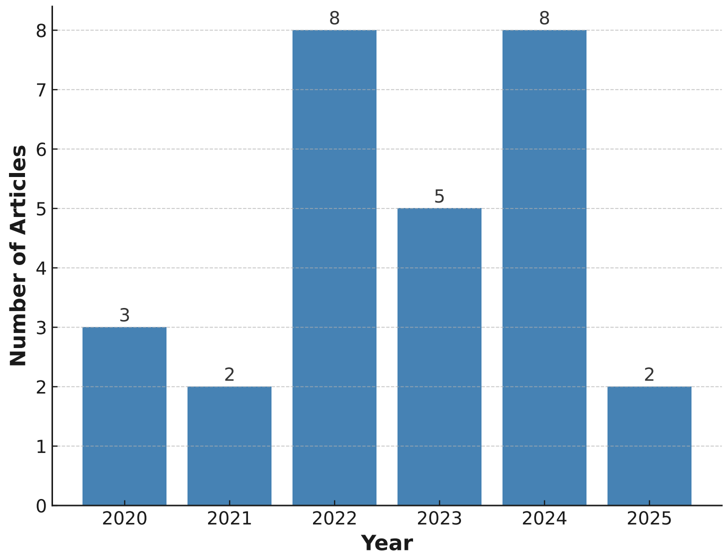

Figure 2. Articles published per year.

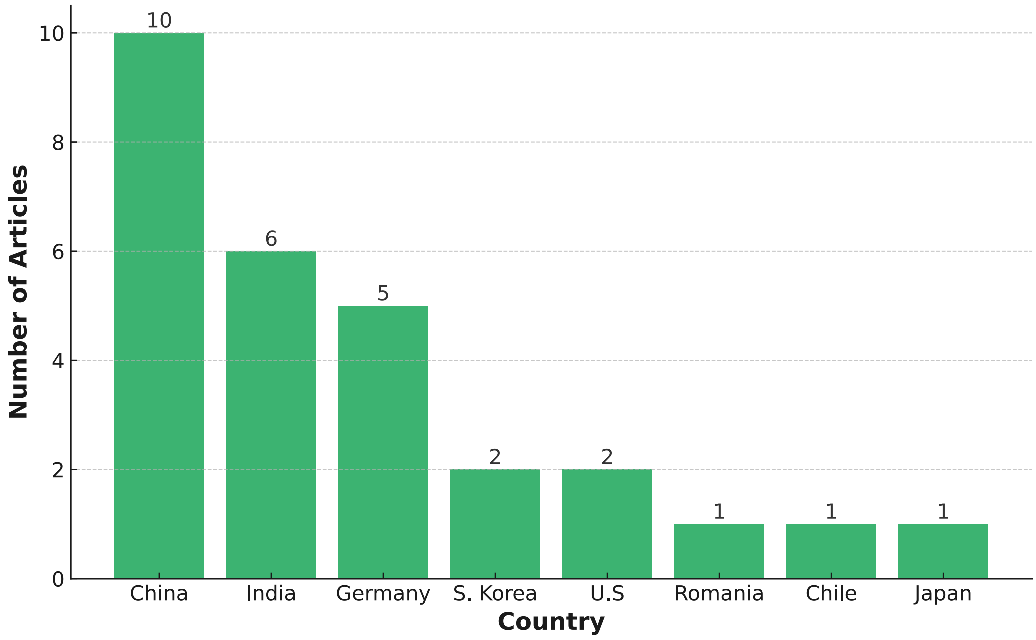

Figure 3. Articles published by country.

Figure 5 highlights the authors with the most publications. Joshi has five articles, followed by Zhu with three. Chakraborty and Münchmeyer each have two. Table A2 lists the authors with one article in this review.

Figure 6 shows the countries of origin for the datasets used in the studies. Japan provides the most widely used datasets, including KiK-net and K-NET, followed by the United States with STEAD [38]. In addition, various studies have used data from China, Italy (INSTANCE) [39], and Chile. Table A3 details cases where multiple datasets or synthetic data, not tied to a specific region, were used.

Figure 7 illustrates the most commonly used algorithms for early magnitude estimation. CNNs are the most frequently employed, often in combination with RNNs and Transformers. Support vector machines (SVMs) have also been applied to this task, along with ensembles of ML models and hybrid approaches that combine ML with DL, such as

XGBoost, LightGBM, CatBoost, random forest (RF), deep neural network (DNN), and AM, which have been explored to improve performance.

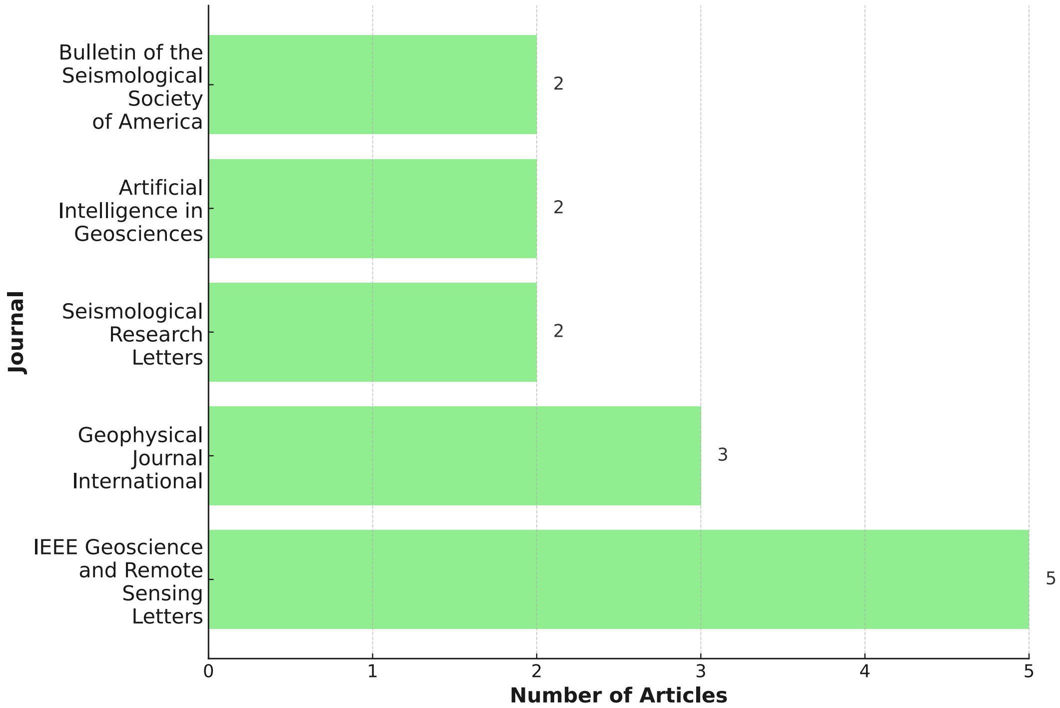

Figure 4. Articles published by journal.

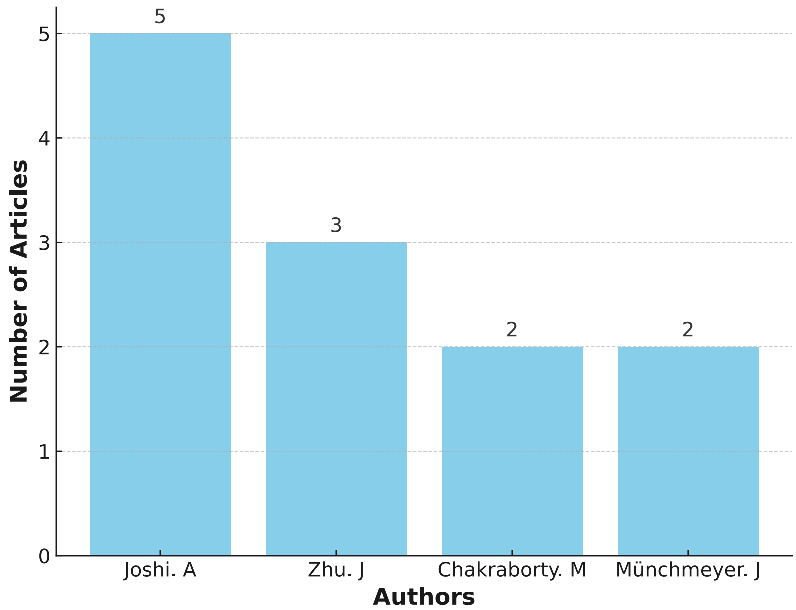

Figure 5. Articles published by author.

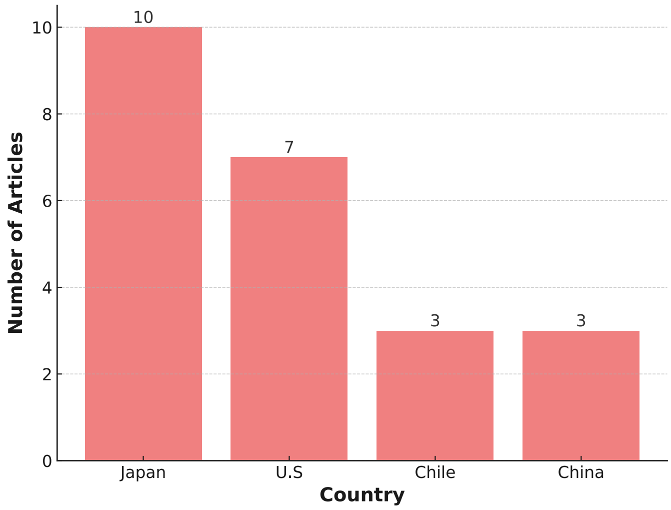

Figure 6. Country of dataset origin.

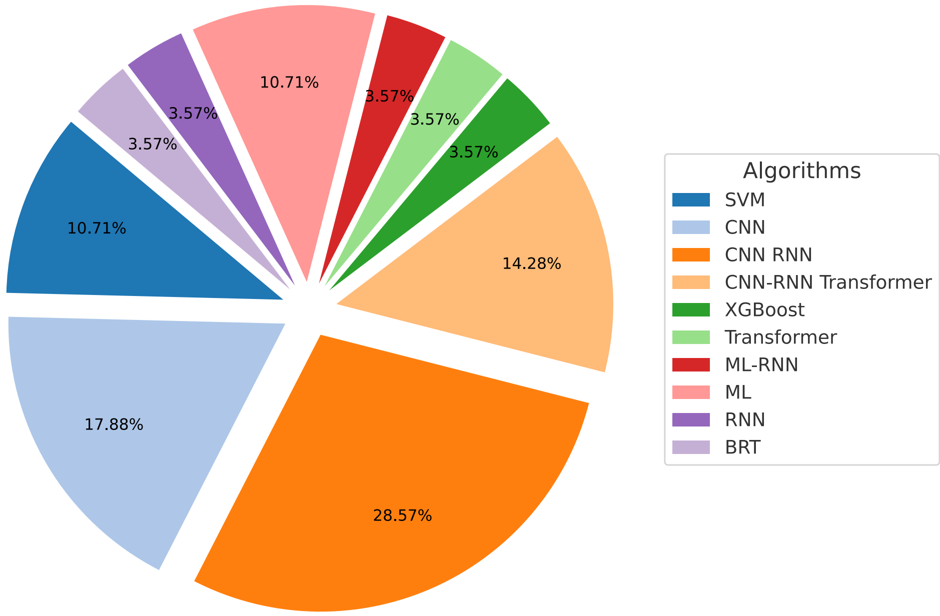

Figure 7. Percentage use of SVM, CNN, RNN, and other DL and ML techniques.

Table 1. Characteristics of the included studies.

Author, Year, Country, Ref.	Dataset Size Country	Target Features	AI Model	Performance Index Name	Performance Value (Unit)
Zhu et al., 2022, China, [40]	K-NET 1837 earthquakes 57,789 waveforms Japan	Magnitude	SVM-M	σ	0.297
Song et al., 2022, China, [32]	K-NET 1205 earthquakes 69,033 waveforms Japan	Magnitude	MEANet: CNN-RNN-AM	0 < error < 0.5 0.5 < error < 1.0 1.0 < error	0.9422 0.0525 0.42
Wang et al., 2023, China, [41]	KiK-net 30,756 waveforms Japan	Magnitude	CNN	σ PME Mean	0.40 0.8421 −0.06

Table 1. *Cont.*

Author, Year, Country, Ref.	Dataset Size Country	Target Features	AI Model	Performance Index Name	Performance Value (Unit)
Quinteros-Cartaya et al., 2024, Germany, [42]	Synthetic 36,800 waveforms Chile	Magnitude	CNN	Synthetic data: RMS Real data: RMS	Synthetic data: 0.06 Real data: 0.09
Münchmeyer et al., 2020, Germany, [43]	IPOC 101,601 earthquakes Chile	Magnitude	Boosting Regression Tree	RMSE	0.117
Kuang et al., 2021, China, [44]	CSES 21,700 synthetic earthquakes China	Magnitude	Fully Convolutional Network (FCN)	$3 < M < 5.9$: Mean, σ $2.3 < M < 3.5$: Mean, σ	$3 < M < 5.9$: -0.017, 0.21 $2.3 < M < 3.5$: -0.011, 0.14
Mousavi et al., 2020, USA, [7]	STEAD 300 k waveforms Global	Local magnitude Duration magnitude	CNN-LSTM	Mean σ	-0.1 0.24
Joshi et al., 2024, India, [45]	K-NET 20 k waveforms Japan	Magnitude	LSTM-Bi-LSTM-XGBoost RF-LightGBM-SVR	MAE RMSE σ	0.24 0.29 0.17
Joshi et al., 2024, India, [31]	K-NET 2960 waveforms Japan	Magnitude	XGBoost	APE	0.004 ± 0.57
Jin et al., 2024, South Korea, [33]	STEAD 1.2M waveforms KPED 335k waveforms Global	Earthquake/Noise Magnitude	Conformer: Convolutional-augmented Transformer	STEAD Dataset: Classification: Accuracy Precision Recall F1 Score Magnitude: MAE KPED Dataset: Classification: Accuracy Precision Recall F1 Score Magnitude: MAE	STEAD Dataset: Classification: 0.9999 0.9999 0.9999 0.9999 Magnitude: 0.1278 KPED Dataset: Classification: 1 1 1 1 Magnitude: 0.1925
Münchmeyer et al., 2021, Germany, [46]	Chile: 1.6 M waveforms Italy: 494,183 waveforms Japan: 372,661 waveforms	Location Magnitude	TEAM-LM (Transformer)	Chile: RMSE (magnitude), Mean error (location) Italy: RMSE (magnitude), Mean error (location) Japan: RMSE (magnitude), Mean error (location)	Chile: 0.08 19 km a 0.5 s 2 km a 25 s Italy: 0.20 2 km a 7 km Japan: 0.22 14 km a 22 km
Chakraborty et al., 2022 , Germany, [47]	STEAD 1.2 M waveforms Global	Earthquake/Noise Magnitude	CNN-Bi-LSTM	Accuracy	0.9386
Ristea et al., 2022, Romania, [48]	STEAD 1.2 M waveforms Global	Epicentral distance Depth Magnitude	Complex CNN	Epicentral distance: MAE Depth: MAE Magnitude: MAE	Epicentral distance: 4.51 km Depth: 6.15 km Magnitude: 0.26
Joshi et al., 2022, India, [49]	K-NET 2951 waveforms Japan	Magnitude	EEWPEnsembleStack: AdaBoost-XGBoost LightGBM-DT Lasso regression	MAE R^2	0.419 0.63
Cofre et al., 2022, Chile, [50]	CSN 7580 earthquakes Chile	Magnitude	LSTM	$M > 4$: MAPE $M < 4$: MAPE	$M > 4$: 0.401 $M < 4$: 0.804

Table 1. *Cont.*

Author, Year, Country, Ref.	Dataset Size Country	Target Features	AI Model	Performance Index Name	Performance Value (Unit)
Wang et al., 2024, China, [51]	STEAD 200k waveforms Global	Local magnitude Duration magnitude	Graph Neural Network CNN-RCGL	M_L: R^2 RMSE Mean σ M_D: R^2 RMSE Mean σ M_L (RMSE): 0–1, 1–2, 2–3, 3–4, $\geq$4 M_D (RMSE): 0–1, 1–2, 2–3, $\geq$3 M_L (R^2 add-SNR dB): -2, -1, 0, 1, 2, 3, 5, 15 M_D (R^2 add-SNR dB): -2, -1, 0, 1 , 2, 3, 5, 15	M_L: 0.9303 0.1844 0.0054 0.1843 M_D: 0.8621 0.2575 0.0422 0.2540 M_L: 0.1512, 0.1788, 0.2619, 0.3640, 0.6324 M_D: 0.3020, 0.2289, 0.2586, 0.3083 M_L: 0.872, 0.872, 0.873, 0.875, 0.876, 0.880, 0.885, 0.892 M_D: 0.792, 0.796, 0.797, 0.802, 0.808, 0.815, 0.837, 0.850
Zhu et al., 2024, China, [52]	K-NET 129,513 waveforms 2794 earthquakes Japan	Magnitude	MCFrame: (CNN-RNN-AM) SVM, RF, DNN	$M < 5.5$ a 3 s: Accuracy $M \geq 5.5$ a 3 s: Accuracy $M < 5.5$ a 1 s: Accuracy $M \geq 6$ a 1 s: Accuracy $M \geq 6$ a 3 s: Accuracy $M \geq 6$ a 5 s: Accuracy $5 < M < 6$ a 5 s: Accuracy	$M < 5.5$ a 3 s 0.98 $M \geq 5.5$ a 3 s 0.89 $M < 5.5$ a 1 s 0.99 $M \geq 6$ a 1 s 0.90 $M \geq 6$ a 3 s 0.95 $M \geq 6$ a 5 s 0.97 $5 < M < 6$ a 5 s 0.78
Yoon et al., 2023, South Korea, [53]	STEAD 260k waveforms KiK-net 130 k waveforms Global/Japan	Magnitude Epicentral distance	CRNN	STEAD Dataset (epicentral distance, magnitude): MAE KiK-net (epicentral distance, magnitude): MAE Inference time: (GPU, CPU)	STEAD Dataset (epicentral distance, magnitude): 2.2736, 0.1337 KiK-net (epicentral distance, magnitude): 5.0040, 0.1448 Inference time (ms): 796.4, 5.68
Shakeel et al., 2022, Japan, [54]	STEAD 93,144 waveforms Global	Magnitude	3D Convolutional Recurrent Neural Network (3D-CNN-RNN)	$0 < M < 1$ Precision Recall F1 Score $1 < M < 2$ Precision Recall F1 Score $2 < M < 3$ Precision Recall F1 Score $3 < M < 4$ Precision Recall F1 Score $4 < M < 8$ Precision Recall F1 Score Earthquake/Noise: Precision Recall F1 Score	$0 < M < 1$ 0.97 0.50 0.66 $1 < M < 2$ 0.98 0.69 0.81 $2 < M < 3$ 0.83 0.51 0.63 $3 < M < 4$ 0.93 0.90 0.91 $4 < M < 8$ 0.84 0.81 0.82 Earthquake/Noise: 0.99 0.87 0.92

Table 1. *Cont.*

Author, Year, Country, Ref.	Dataset Size Country	Target Features	AI Model	Performance Index Name	Performance Value (Unit)
Ren et al., 2023, China, [55]	STEAD/CENC 1097 earthquakes 4166 waveforms China/Global/Italy	Magnitude	CNN	General accuracy Medium and large earthquake: Precision Recall F1 Score Small earthquake: Precision Recall F1 Score	0.9765 Medium and large earthquake: 0.9827 0.9693 0.9769 Small earthquake: 0.9796 0.9834 0.9770
Dybing et al., 2024, USA, [56]	USGS (MLAAPDE) 2.4 M waveforms Global	Magnitude	AIMag: CNN-RNN	Mean Precision (M2.3 to M7.6)	±0.5
Meng et al., 2023, China, [57]	CENC 324,266 waveforms China	Magnitude	EEWMagNet: CNN	Accuracy Precision Recall F1 Score	0.9023 0.8935 0.9108 0.9021
Hou et al., 2024, China, [58]	8144 earthquakes Japan	Magnitude	Transformer LSTM-CNN	First 3 s: RMSE MAE σ First 14 s: RMSE MAE σ	First 3 s: 0.38 0.29 0.38 First 14 s: 0.20 0.15 0.20
Chanda et al., 2021, India, [59]	SPECFEM3D 400 earthquakes Synthetic data	Magnitude Location	SVM	Magnitude: RMSE R^2 MSE MAE Hypocentral dist: RMSE R^2 MSE MAE Azimuth: RMSE R^2 MSE MAE Elevation: RMSE R^2 MSE MAE	Magnitude: 0.0412 1.0 0.00169 0.009419 Hypocentral dist: 485.53 1.0 235,700 268.64 Azimuth: 68.85 0.58 4741.5 58.94 Elevation: 0.0056422 1.0 0.00003 0.0015
Chakraborty et al., 2022, Germany, [4]	STEAD 32,356 waveforms INSTANCE 135,347 waveforms Global/Italy	P-wave arrival time Earthquake/Noise Magnitude	CNN Bi-LSTM	STEAD Dataset Classification: Accuracy Precision Recall F1 Score Magnitude: Mean error σ RMSE MAE P-wave arrival time: Mean error σ RMSE MAE INSTANCE Classification: Accuracy Precision Recall F1 Score Magnitude: Mean error σ RMSE MAE P-wave arrival time: Mean error σ RMSE MAE	STEAD Dataset Classification: 0.9858 0.9964 0.9831 0.9897 Magnitude: −0.06 0.60 0.61 0.46 P-wave arrival time: −0.05 0.10 0.12 0.05 INSTANCE Classification: 0.9759 0.9866 0.9753 0.9810 Magnitude: −0.02 0.69 0.69 0.54 P-wave arrival time: 0.01 0.52 0.52 0.29

Table 1. *Cont.*

Author, Year, Country, Ref.	Dataset Size Country	Target Features	AI Model	Performance Index Name	Performance Value (Unit)
Zhu et al., 2022, China, [60]	CSMNC 7236 waveforms 461 earthquakes China	Magnitude	SVM	Estimation error: $M4.2$, $M5.3$ and $M6.3$ Estimation error: $M6.6$ and $M7$ Average error σ	Estimation error: $M4.2$, $M5.3$ and $M6.3$ $\pm$0.3 units 1 s Estimation error: $M6.6$ and $M7$ $\pm$0.3 units 13 s 0.31 0.41
Joshi et al., 2025, India, [61]	18,994 waveforms Japan	Magnitude	MagPred XGBoost-LightGBM CatBoost-RF	First 3 s: MAE RMSE First 4 s: MAE RMSE First 5 s: MAE RMSE	First 3 s: 0.42 0.56 First 4 s: 0.40 0.54 First 5 s: 0.39 0.53
Joshi et al., 2025, India, [34]	26,279 waveforms Japan	Magnitude PGA	DFTQuake AM-NN-LightGBM XGBoost-RF	Magnitude: MAE RMSE $\mathcal{M}_U$ $\mathcal{M}_L$ Training time (s) Parameters PGA: MAE RMSE $\mathcal{M}_U$ $\mathcal{M}_L$ Training time (s) Parameters	Magnitude: 0.66 0.85 3.81 0.12 62.22 ~2.4 M PGA: 0.25 0.32 95.19 99.78 62.22 ~2.4 M

3.1. Early Earthquake Magnitude Estimation as a Classification Task

Ref. [47] explored the impact of input data length on a deep-learning-based magnitude classifier, evaluating durations from 1 to 30s of P-wave data. The model employed a combination of CNNs and Bi-LSTMs for feature extraction and classification, distinguishing between noise, low-magnitude, and high-magnitude events ($M \geq 5.0$). Data augmentation techniques, including polarity inversion and sliding windows, were applied specifically to high-magnitude events to address class imbalance, improving classification performance. Despite testing various data lengths, the results indicate that extending the input window beyond 3 s does not significantly enhance classification accuracy, suggesting that early waveform information carries sufficient discriminative features for magnitude classification. The model consistently achieved an accuracy range of 90.04% to 93.86%, demonstrating robustness across different hypocentral distances and signal-to-noise ratio (SNR) conditions. However, classification errors were concentrated around the $M5.0$ decision threshold, with 65% of events between $M4.5$ and $M5.0$ misclassified as high-magnitude. This trend suggests a tendency toward conservative magnitude estimation, prioritizing the reduction of false negatives over false positives. The study highlights the potential for integrating such classifiers into EEW systems, as rapid magnitude classification using minimal waveform data is crucial for timely alerts. However, the misclassification near the decision boundary raises questions about whether a probabilistic approach or a regression-based framework could improve classification confidence and mitigate the overestimation of events near $M5.0$. Future work could explore hybrid approaches combining classification and regression methods to enhance both speed and reliability in real-time seismic applications.

Ref. [55] proposed a CNN model to estimate earthquake severity. Instead of predicting the exact magnitude, the model classifies earthquakes as medium-to-large ($M_L \geq 5$) or small ($M_L < 5$), using only the first 4 s of the P-wave. The CNN architecture consists of blocks with convolutional layers, Max-pooling, and 1 × 1 convolutional layers to enhance feature abstraction. The training uses binary cross-entropy loss and the Adam optimizer with a 10^{-4} learning rate. The model achieved an accuracy of 97.90%, outperforming the traditional method based on P_d amplitude (94.54%). Tests were conducted with real earthquakes, such as the Changning (M6.0) and Tangshan (M5.1) swarms, with an accuracy of 98.63% in identifying small aftershocks. The model demonstrated strong generalization when tested with the INSTANCE dataset (accuracy of 95.52%) and in real-time at the China Earthquake Networks Center (CENC), monitoring 16 earthquakes. One key advantage of this approach is its ability to operate with minimal waveform information while maintaining high accuracy, making it suitable for early warning applications in regions with sparse seismic networks. However, the binary classification scheme may limit its applicability for detailed magnitude estimation, particularly for borderline cases around the M5.0 threshold.

In [57], the EEWMagNet model was introduced for rapid seismic magnitude classification in EEW systems. Its goal is to identify whether an earthquake's magnitude is above or below 4, which is critical for issuing warnings without calculating the exact magnitude. EEWMagNet employs CNN layers, a Dense Block with a bottleneck for feature extraction, and multi-head attention for temporal feature correlation. The epicentral distance is also incorporated to improve the accuracy of the magnitude classification. Models were trained with different P-wave window lengths from CENC data, and results showed that a 7 s window provided the best performance, even compared to longer ones. The dataset includes recordings from 1104 stations, with $M \geq 4$ events augmented using sliding windows and noise overlay techniques to address the imbalance. Experiments demonstrated that EEWMagNet achieves an accuracy of 90.23%, with a precision of 89.35%, recall of 91.08%, and an F1 score of 90.21%. Incorporating epicentral distance improved classification by 18%, and avoiding normalization improved amplitude retention, increasing accuracy by 8%. The model demonstrates the potential for deployment in early warning systems, offering rapid and accurate magnitude classification with minimal data length requirements.

Ref. [52] proposed a machine learning magnitude classification framework (MCFrame), a rapid earthquake magnitude classification method based on ML. Using single-station data, it classifies seismic events as high magnitude ($M \geq 5.5$) or low magnitude ($M < 5.5$). The model combines CNN layers, RNNs, and an AM to extract features from P-wave signals during the first seconds after an earthquake. The model was trained with different P-wave time windows, and the results showed that a 5 s window provided the best performance. Different classifiers were analyzed, and a probability threshold of 0.5 was established to determine whether an event is $M \geq 5.5$ or lower. To eliminate the effect of epicentral distance, it was fixed at 10 km during training. It was trained and tested using 129,513 Japanese K-NET seismic network records, corresponding to 2794 seismic events with magnitudes ranging from M3 to M8. Results showed an accuracy of over 99% for low-magnitude events and more than 90% for high-magnitude events. MCFrame was validated with 496 independent seismic events, achieving 87% accuracy for high-magnitude events. MCFrame outperforms traditional baseline models such as standalone CNN and LSTM networks. These conventional methods focus on regression-based magnitude estimation but suffer from limitations in accuracy and generalization, particularly near classification boundaries (e.g., M5.5).

In [54], a 3D-CNN-RNN model was proposed to classify earthquake magnitudes as a multi-label problem, allowing a seismic event to belong to multiple categories simultaneously. The model extracts temporal and spatial features from seismic signals using log-mel spectrograms and combining 3D convolutional layers with gated recurrent units (GRUs). Unlike traditional multi-class approaches, this method better captures the continuous nature of earthquake magnitudes, reducing misclassification at class boundaries. Earthquake magnitudes were grouped into five classes (0 to 8) and an additional category for non-earthquake events. Key results showed significant improvements in multi-label classification compared to multi-class approaches, achieving an accuracy of up to 97% for small earthquakes (0–1) and 99% for non-earthquake events. However, performance declined for lower magnitudes due to their weaker signal characteristics and spectral overlap with noise, highlighting a challenge in detecting small events. Regarding performance, the model achieved an F1 score of 0.91 for earthquakes of magnitude 3–4, although performance was lower for smaller magnitudes. Tests also demonstrated improved classification of earthquakes across multiple simultaneous magnitudes, which is crucial for precise and early disaster response. This suggests that integrating additional data preprocessing techniques or hybrid architectures could further enhance classification reliability, particularly for low-magnitude events.

3.2. Early Earthquake Magnitude Estimation as a Regression Task

Of the reviewed studies, 23 of 28 approached earthquake magnitude estimation as a regression task. The work presented in [7] estimates M_L and duration magnitude (M_D) using a single seismic station without requiring input data normalization. The network's input consists of three-component seismograms of 30 s duration (3000 samples), which includes both P and S waves, allowing the model to capture a more complete representation of seismic signals for improved magnitude estimation. The architecture consists of two convolutional layers for feature extraction, each followed by dropout and max-pooling layers for dimensionality reduction and regularization. The LSTM layer captures temporal dependencies in the waveforms, and the final fully connected layer predicts the earthquake magnitude. The model was trained using the Adam optimizer and minimizes mean squared error (MSE) as the loss function. The study demonstrates that neural networks can learn attenuation relationships and site effects necessary for magnitude estimation, achieving high prediction accuracy with a mean error close to zero and a standard deviation of $\sim$0.2. When using only a P-wave window without the S-wave, the model struggled to make accurate predictions, limiting its effectiveness for EEW applications. Similarly, in [45], magnitude estimation was performed, but using a single seismic station with only 3 s of P-wave data. The AMagDN model combines autocorrelation attention, LSTM, Bi-LSTM, and ML blocks to produce precise magnitude estimations. The ML block employs models like XGBoost, RF, and support vector regression (SVR) for final magnitude estimation. Unlike in [7], robust scaling was applied to normalize data features. The model performance indicates overestimation for low magnitudes (5.0–6.5 M_{JMA}) and underestimation for high magnitudes (6.5–8.0 M_{JMA}). However, the prediction's standard deviation (σ) remains within $\pm$0.55, indicating relatively low error. Moreover, the error decreases with greater epicentral distances, while inaccuracies occur for closer distances due to the rapid arrival of the P-wave.

In [31], the first 3 s of the P-wave was also used with acceleration records. A machine learning-based approach using the XGBoost model was employed and compared to well-known methods such as τ_c and P_d. The model's improved accuracy, achieved through 29 extracted features, suggests that early P-wave characteristics contain enough information for rapid and reliable magnitude estimation. Similar to [45], tests were conducted on two earthquakes in Japan with magnitudes 4.5 and 6.1 M_{JMA}. Based on the P-wave, the model predicted magnitudes of 4.58 $\pm$ 0.33 and 6.32 $\pm$ 0.29, respectively. This supports the feasibility of using short-window P-wave data for early warning applications. These results demonstrate that early ground motion (3 s of P-wave data) during an earthquake can effectively predict magnitude with reasonable accuracy. However, the model's reliance on a single seismic station and limited waveform duration could lead to underestimation or overestimation in complex ruptures or varying epicentral distances.

Ref. [33] introduced a conformer-based architecture enhanced with LoRA for seismic event classification and magnitude estimation. Instead of using raw waveforms, the model employed spectrograms derived from Z-component seismograms via short-time Fourier transform (STFT), which may introduce a dependence on preprocessing techniques. The input is processed through convolutional subsampling, linear layers, and dropout. The conformer encoder is simplified to six layers with reduced dimensionality, drastically lowering parameters from 1.07 M to 12 K while preserving performance. The decoder differs for each task, for example, fully connected layers for classification and a Bi-LSTM for magnitude estimation. LoRA is employed during fine-tuning, freezing pre-trained weights, and introducing low-rank decomposition matrices, which minimizes trainable parameters while adapting the model to regional datasets. The training utilizes SpecAugment for data augmentation, the Adam optimizer, and mean squared error or cross-entropy loss, depending on the task. The study used the STEAD and KPED datasets for classification and magnitude estimation. The resulting model outperformed state-of-the-art models such as EQTransformer [62], CRED [17], DetNet [63], Yews [64], and STA/LTA [65] in classification, as well as complex CNN [48], MagNet [44], and a deeper convolutional recurrent neural network (CRNN) [53] in magnitude estimation. Similarly, for magnitude estimation, it surpasses previous models, achieving a 7% reduction in mean absolute error (MAE) on the STEAD dataset and a 48% decrease on the KPED dataset. These improvements suggest that the conformer model, combined with LoRA fine-tuning, is well-suited for both global and regional seismicity studies. However, while LoRA reduces trainable parameters, its effectiveness relies on pre-trained weights from global datasets, which may not always capture local seismic characteristics. Future work could explore hybrid approaches integrating raw waveform processing alongside spectrogram-based feature extraction, as well as multi-station input to further enhance model robustness.

Ref. [48] introduced a complex CNN-based model to estimate epicentral distance, depth, and magnitude from raw three-component seismograms without requiring P- and S-wave arrivals. By leveraging STFT representations, the model processes both magnitude and phase information using complex-valued convolutional layers (CReLUs), improving accuracy over real-valued CNNs. Unlike traditional methods that depend on explicit arrival-time picking, this model directly learns seismic attributes from waveform data, enhancing robustness against picking errors. Results show that complex CNN surpasses traditional models, achieving an MAE of 4.51 km for distance, 6.15 km for depth, and 0.26 for magnitude. The use of skip connections and residual complex layers enhances feature extraction and convergence, demonstrating the potential of complex-valued deep learning for single-station seismic analysis. Additionally, the model was tested on multiple datasets, confirming its adaptability to different seismic networks. However, eliminating P- and S-wave arrival constraints may introduce uncertainties in complex waveforms, and the higher computational cost could limit real-time applications. Future work should explore lightweight implementations of complex CNNs to reduce inference time while maintaining accuracy.

Ref. [41] introduced a new approach for seismic magnitude estimation using a CNN called EEWNet. The novelty of this method lies in directly processing vertical accelerograms without preprocessing, adapting its architecture based on the P-wave length received as input. EEWNet's architecture includes sequential convolution and pooling layers, with ReLU activations, and a final fully connected layer. Unlike traditional approaches that rely on predefined P-wave parameters, EEWNet takes raw waveform data as input and adapts its architecture accordingly, eliminating the need for hyperparameter tuning. The model predicts the logarithm of the maximum displacement in horizontal components, which is used to calculate magnitude via established relationships. Training employed the Adam optimizer with a learning rate of 0.0001, batch size of 200, and 400 epochs. The dataset consists of 30,756 accelerograms recorded by 688 KiK-net borehole sensors, covering earthquakes with magnitudes from 4 to 9 and hypocentral distances between 25 km and 200 km. EEWNet's performance is compared to classic magnitude estimation approaches, such as $\tau_{\log}$, τ_c, and P_d, demonstrating a lower standard deviation and a higher percentage of predictions with magnitude errors between -0.5 and 0.5, indicating significantly improved accuracy. Additionally, experiments show that increasing the P-wave duration used as input enhances prediction accuracy, reinforcing the importance of capturing more waveform information.

In [49], the EEWPEnsembleStack model was proposed to estimate magnitude using the first 3 s of the P-wave, similar to previous models that also rely on this initial time window for rapid magnitude estimation. This ensemble model combines ML techniques such as AdaBoost, XGBoost, LightGBM, decision tree (DT), and Lasso regression. Predictions from earlier layers serve as inputs to subsequent layers, culminating in the final magnitude estimates. The model was trained and validated using a 10-fold cross-validation approach to prevent overfitting, and hyperparameters are optimized via grid search. In specific tests, the model correctly predicted a magnitude of 6.3 as 6.0 ± 0.5, outperforming traditional methods such as τ_c and P_d, which underestimated magnitude by approximately two units. Additionally, the study highlights the impact of feature ablation analysis, demonstrating that the combination of all extracted features enhances model performance. Notably, P_d, $ID2$, and $IV2$ were identified as the most significant features for magnitude estimation. Furthermore, the model was tested in various hypocentral scenarios, showing that its accuracy was unaffected by epicentral distance. However, while the ensemble model achieves lower prediction errors compared to conventional methods, its computational complexity and inference speed could be potential challenges for real-time applications in EEW systems. Future work could explore reducing the number of base models while maintaining accuracy.

Ref. [46] presented TEAM-LM, a Transformer-based model designed for real-time magnitude and hypocentral location estimation. Unlike previous deep learning models that relied on fixed station sets or time windows, TEAM-LM processes seismic data from a dynamically varying set of stations, leveraging self-attention mechanisms to capture inter-station dependencies. The model integrates a mixture density network to provide probabilistic estimates of magnitude and location. Experiments were conducted with three diverse datasets from Chile, Italy, and Japan, showcasing the model's robustness across different tectonic environments. TEAM-LM demonstrated superior performance compared to classical magnitude estimation approaches and previous deep learning models, achieving root mean square error (RMSE) values of 0.08 in Chile, 0.20 in Italy, and 0.22 in Japan for magnitude estimation. For epicentral location, the model improved over time, reaching an error of 2 km in Chile and Italy after 25 s. However, challenges remain, particularly in the estimation of large magnitudes, where limited training data on major seismic events can lead to underestimation. To address this, transfer learning techniques were explored, incorporating events from different regions into the training process, which improved large-magnitude predictions. Additionally, while location estimates are precise in well-sampled regions, errors increase significantly outside the training distribution, suggesting potential limitations in generalization. Future research could focus on refining the model's ability to generalize beyond the training distribution and incorporating physics-based constraints to enhance reliability in real-world applications.

In [51], the EQGraphNet model was presented for magnitude estimation using a single station. The architecture combines deep CNN layers with a graph-based approach, incorporating 11 CNN layers and 10 residual connection graph layers (RCGLs), designed to mitigate over-smoothing issues typical in deep graph neural network (GNN) models and improve temporal feature extraction. EQGraphNet processes seismic signals as temporal graphs, where each time step is represented as a node connected to adjacent time steps through a symmetric graph structure. This enables efficient modeling of temporal dependencies while preserving local amplitude variations. The RCGL integrates a GNN and residual connection to enhance signal denoising and retain essential amplitude and coda duration features, reducing noise interference. The performance of EQGraphNet is compared to other models such as MagNet [44], CREIME [4], and CNQI [66]; EQGraphNet demonstrates higher accuracy in estimating M_L and M_D, with determination coefficients (R^2) of 0.9303 for M_L and 0.8621 for M_D, as well as lower estimation errors. Unlike previous approaches that rely solely on convolutional or recurrent architectures, EQGraphNet integrates graph structures to leverage spatial-temporal relationships in the data. EQGraphNet is robust against natural noise, performing well even in signals with low SNR, making it ideal for adverse conditions. The model was also tested in regions such as Puerto Rico, Alaska, the Salish Sea, and the Hawaiian Islands, where EQGraphNet demonstrated its ability to generalize across different geological environments. However, the study highlights that performance deteriorates for very large magnitudes, suggesting that additional training data or transfer learning techniques could further enhance generalization. The proposed model is a promising tool for real-time magnitude estimation and early warning applications, especially in areas with limited seismic networks or noisy signals.

Ref. [53] proposed a deep CRNN model with a multitask learning (MTL) approach for estimating epicentral distance and seismic magnitude. It processes three-component seismograms, combining deep convolutional blocks and Bi-LSTMs to extract static and temporal features. The feature extractor applies spectral normalization, leaky ReLU activations, and dropout to improve stability and prevent overfitting, while Bi-LSTMs model temporal dependencies, ensuring sensitivity to waveform variations over time. The model incorporates an embedding network based on a multilayer perceptron (MLP) to process the maximum amplitude of the waveform, transforming it into feature vectors that enhance magnitude estimation and epicentral distance predictions. Unlike previous models, this approach explicitly integrates amplitude features into the feature representation, reducing bias in magnitude estimation. The estimator employs a shared MTL structure with task-specific output layers for magnitude and distance, optimizing a joint loss function that combines L1 loss for accuracy and weight decay for regularization. The model was implemented using PyTorch (version 2.0.0) and compared to the following three existing models: Mousavi and Beroza [67], complex CNN [48], and MagNet [44]. Experimental results show that explicitly modeling amplitude enhances distance estimation, particularly for low-amplitude events, while Bi-LSTMs improve temporal feature extraction. Compared to previous models, the proposed approach achieves lower MAE for both magnitude and distance, with higher R^2 values, indicating strong prediction accuracy. The study also suggests that incorporating additional waveform attributes, such as envelope functions or coda duration, could further improve generalization and robustness to regional variability. Ref. [42] presented a CNN-based deep learning model for earthquake magnitude estimation using high-frequency GNSS (HR-GNSS) displacement data. The proposed architecture consists of six 2D convolutional layers, three max-pooling layers for dimensionality reduction, and three fully connected layers for final magnitude estimation. The model processes three-component displacement data, utilizing kernel sizes of (1,3) and ReLU activation functions. Unlike traditional approaches that rely on empirical scaling relationships, this model directly learns patterns from raw HR-GNSS signals. The model was primarily trained on a synthetic dataset consisting of 36,800 rupture scenarios generated for the Chilean subduction zone. These synthetic signals provided a controlled environment to analyze performance without noise interference. However, applying the model to real earthquake data revealed that the lack of noise in training could affect generalization. Tests with synthetic data yielded a root mean square (RMS) error between 0.06 and 0.11, while real data tests, including events from regions such as Cascadia and Chile, showed RMS values ranging up to 0.09 when station configurations were consistent with the training data. One of the key findings is the impact of temporal signal window length on magnitude estimation accuracy. Shorter time windows sometimes led to the underestimation of large events ($M_w > 9.5$), while smaller earthquakes ($M_w < 6.9$) tended to be overestimated. Additionally, factors such as SNR and maximum signal amplitude significantly influenced performance when working with real data. The study highlights the potential of deep learning in HR-GNSS-based earthquake magnitude estimation but also emphasizes the challenges of adapting models trained on synthetic data to real-world applications. Future work could focus on incorporating realistic noise during training and refining the model's ability to generalize across different seismogenic regions.

Ref. [43] introduced a method for estimating seismic magnitudes with low uncertainty, even when data from only a few stations is available. It combines three-dimensional (3D) physical corrections with boosted tree regression to reduce variability in single-station magnitude estimates. The approach was tested on the IPOC seismic catalog from northern Chile, which included over 100,000 events recorded between 2007 and 2014. Unlike traditional methods, this approach extracts 110 waveform features, such as displacement, velocity, acceleration, and cumulative energy. These features are corrected for source, station, and path effects using a non-parametric function that depends on epicentral distance and depth, while also incorporating station-specific adaptive 3D corrections. The boosting tree regression further enhances prediction stability by integrating multiple features, reducing standard deviation by 57%, with 40% of this improvement attributed to the regression model itself. Compared to standard correction models, this approach achieves up to a 47% reduction in RMSE for horizontal displacement and 39% for vertical acceleration. P-wave-derived features and envelope values prove crucial in reducing uncertainties. However, despite these improvements, the method's reliance on a large dataset for training and the computational cost of 3D corrections may limit its application in real-time scenarios. Future studies could explore hybrid models combining boosting trees with neural networks to further enhance efficiency and adaptability.

Ref. [40] proposed an SVM-based model (SVM-M) for magnitude estimation within the first three seconds of the P-wave arrival, similar to other models that utilize this early time window for rapid magnitude estimation. The model operates without constraints on SNR or epicentral distance, enhancing its flexibility across different seismic conditions. It utilizes a Gaussian radial basis function kernel to optimize a regression function mapping 12 input parameters to magnitude estimates. Compared to traditional methods like τ_c and P_d, SVM-M significantly reduces prediction error, achieving a standard deviation of 0.297, in contrast to 1.637 for τ_c and 0.425 for P_d. Notably, while increasing the time window from 1 s to 10 s improves performance, particularly for large earthquakes, using only 3 s of P-wave data still provides reliable magnitude estimates, suggesting that essential information is present even in the early portion of the waveform. Additionally, the model demonstrates strong generalization capabilities, reinforcing its potential for earthquake early warning applications.

Ref. [32] presented MEANet, a model for magnitude estimation that integrates physics-based features with deep learning. The input consists of a combination of acceleration, velocity, and displacement waveforms with a length of 3 s after the P-wave arrival. The architecture includes an input block, a convolutional-based feature extractor, a bidirectional gated recurrent unit (Bi-GRU) with an AM for sequential learning, and an output layer for final magnitude estimation. Compared to traditional τ_c and P_d methods, MEANet demonstrates superior accuracy, achieving an error of ± 0.25 magnitude units at a single station. Tests indicate that extending the time window from 3 to 10 s improves magnitude prediction, as additional waveform data enhances event characterization, particularly for larger earthquakes. The study also highlights the importance of certain physical parameters, such as P_d, P_v, and P_a, in improving robustness. By leveraging a physics-guided approach combined with deep learning, MEANet reduces the underestimation of large magnitudes, a common issue in early warning models. This hybrid strategy suggests potential applications in early warning systems, particularly in regions where rapid and precise magnitude estimation is critical, highlighting its ability to generalize across different seismic scenarios.

Ref. [44] introduced MagNet, a fully convolutional network (FCN) designed for magnitude estimation. The architecture consists of two main components, namely, a compression module for feature extraction and an expansion module for magnitude prediction. The compression module applies sequential convolutional and pooling layers to extract key features from 2D waveform data, reducing dimensionality while increasing feature depth. The expansion module uses upsampling and convolutional layers to generate a Gaussian probability distribution representing the magnitude. Unlike conventional CNNs, MagNet employs a probability distribution approach rather than a single value estimation, enhancing robustness against noise and outliers. The model was trained on 700 real seismic events from the China Seismic Experimental Site (CSES) and augmented to 21,700 datasets using synthetic data generation techniques. It processes waveform records from multiple seismic stations to generate probabilistic magnitude estimates. MagNet performs well for moderate earthquakes (magnitudes 3 to 5.9), achieving an average error of -0.017 and a standard deviation of 0.21, with a 99.8% recovery rate. For small earthquakes (magnitudes 2.3 to 3.5), the model maintains high accuracy with an error of -0.011 and a standard deviation of 0.14, demonstrating its capability for lower-magnitude events. The model also shows resilience against anomalous noise, outperforming traditional magnitude estimation methods. However, the model struggles with large earthquakes ($M > 6.5$), as the dataset lacks sufficient high-magnitude training samples. This limitation suggests the need for future improvements, such as transfer learning or hybrid models incorporating physical constraints.

Ref. [56] proposed AIMag, a deep learning model for single-station magnitude estimation based on the MagNet architecture. AIMag employs a CRNN structure that integrates convolutional layers for feature extraction and recurrent layers for temporal modeling. Unlike traditional approaches, AIMag estimates magnitudes using the first P-waves without requiring location information or instrument response corrections. It was tested across various time window lengths ranging from 1 to 114 s, revealing that performance improves as the window length increases. At its optimal configuration (114 s), AIMag achieves a median estimation error within ± 0.5 magnitude units for events ranging from $M2.3$ to $M7.6$. However, for magnitudes above $M7$, the model tends to underestimate, likely due to P-wave saturation and the scarcity of large-magnitude events in the training set. To enhance generalization, data augmentation techniques were applied, including P-phase shifting, noise injection, and random channel removal. AIMag was tested on historical earthquakes, such as the 2004 Sumatra ($M9.1$) and 2011 Tōhoku ($M9.1$) events, correctly identifying them as large but underestimating their exact magnitudes. Despite this limitation, AIMag provides rapid magnitude estimates that could improve early warning systems, particularly in tsunami-prone regions.

Ref. [58] introduced a DL model for real-time magnitude estimation. The model integrates three heterogeneous modalities, namely, three-component acceleration seismograms, differential P-wave arrivals, and differential sensor locations, using a combination of CNN-LSTM and Transformer architectures to process these data. Waveform inputs are processed using CNN layers to downsample dimensions and LSTM layers to capture temporal dependencies. The Transformer encoder extracts features from differential P-wave arrivals, representing travel time relationships, while the Transformer decoder combines these with differential seismometer locations to infer implicit distance information. A global average pooling (GAP) layer fuses features from multiple seismometers for final magnitude estimation through a fully connected layer. By incorporating location information, the model aims to enhance magnitude estimation accuracy, particularly in early warning scenarios where limited data are available. Data augmentation strategies were implemented to enhance model robustness. Key evaluation metrics included RMSE, MAE, and standard deviation, with values of 0.38, 0.29, and 0.38 at 3 s and 0.20, 0.15, and 0.20 after 14 s of the P-wave arrival. The model demonstrated greater accuracy for earthquakes with magnitudes between 3.0 and 5.0. However, performance deteriorated for larger magnitudes ($\geq$7.0) due to data limitations and saturation effects. The model also proved robust to errors of up to 0.2 s in P-wave detection.

Ref. [59] proposed an SVM-based method to estimate earthquake magnitude and location using only the vertical component of a single seismic station. The SVM model employs a Gaussian kernel for regression, transforming the input features into a higher-dimensional space to separate non-linear relationships. Bayesian optimization was used to determine optimal hyperparameters, and tenfold cross-validation was applied to avoid overfitting. The model was trained on 400 synthetic seismograms generated with SPECFEM3D, simulating earthquakes with magnitudes between 4 and 7 at a fixed depth of 40 km. The study highlights the advantage of using single-component data, making it applicable in regions with sparse seismic networks. Six features were extracted from the seismic signal: mean amplitude, maximum amplitude, RMS, P- and S-wave arrival times, and the P-S interval. The model's performance was evaluated using various metrics. For magnitude estimation, the RMSE was 0.0412, with a perfect R^2 of 1.0, indicating exact predictions. Despite these strong results, the model's reliance on synthetic data raises concerns about its generalizability to real seismic events, especially in high-noise environments. For hypocentral distance, the model achieved an RMSE of 485.53 m and an R^2 of 1.0, with high accuracy. The elevation angle estimation was equally accurate, with an RMSE of 0.0056 radians and R^2 of 1.0. However, azimuth angle estimation was less accurate, with an RMSE of 68.85 and R^2 of 0.58. The model was tested under noise conditions, adding 1%, 3%, and 5% noise levels to synthetic seismograms. While the model maintained robustness for $M > 5$ events, its accuracy significantly declined for smaller magnitudes (M_w 4), suggesting a strong dependency on SNR.

Ref. [60] developed a model for rapid seismic magnitude estimation in the Sichuan-Yunnan region, China, using SVM and transfer learning. The TLSVM-M model adjusts a pre-trained model to improve accuracy in magnitude estimation within the first 3 s of the P-wave. By leveraging transfer learning, the model benefits from prior knowledge, reducing the need for extensive labeled training data and accelerating convergence during training. Unlike conventional SVM approaches, TLSVM-M incorporates regional data adaptation to enhance generalization. The model exhibited an average error of $\pm$0.3 units for minor to moderate earthquakes (M_s 4.2, 5.2, 6.3) within the first second. For larger magnitudes (M_s 6.6, 7.0), an initial underestimation improved within the first 13 s, achieving an error of $\pm$0.3 units. Compared to other methods, such as SVM without transfer learning (error of 0.44), τ_c (1.35), and P_d (0.44), the model achieved an average error of 0.31 and a standard deviation of 0.41, making it the most precise. Despite these improvements, the reliance on single-station data might introduce uncertainties in complex seismic scenarios, suggesting the potential benefit of multi-station integration.

In [4], CREIME, a multitask model for seismic event identification and magnitude estimation based on P-waves, was presented. The architecture integrates three convolutional layers, max-pooling for feature extraction and dimensionality reduction, and two Bi-LSTM layers to capture temporal dependencies in seismic signals. The output layer generates 512 values corresponding to the input samples. The model was trained with a custom loss function combining MSE, MAE, and a magnitude-specific penalty term to address the underestimation of high magnitudes. The primary objective of CREIME is to detect seismic signals, identify P-wave arrival, and estimate earthquake magnitude using only the first 2 s of P-wave data. Compared to conventional models, CREIME avoids explicit feature extraction by leveraging an end-to-end deep learning approach, reducing the need for prior knowledge about waveform characteristics. The model performed well for low and medium magnitudes but underestimated events with magnitudes above 5.5. To address this issue, oversampling and undersampling techniques were applied to balance the magnitude distribution. Another key advantage of CREIME is its robustness against seismic noise, achieving an event detection accuracy of 98% even in low-SNR environments. CREIME outperforms traditional methods like STA/LTA and can be implemented in early warning systems.

In [50], an LSTM-based method was proposed to estimate magnitude using only data from a single seismic station, similar to previous approaches that have explored single-station magnitude estimation. The main objective is to enhance EEW systems. This model focuses on earthquakes with magnitudes above $M6$ due to their more significant destructive potential and follows an end-to-end approach. Designed features are used instead of raw signals to overcome the limitation of large-event data. The dataset includes events with magnitudes ranging from $M1.8$ to $M8.3$, with an average distance of 196 km between the epicenter and the nearest station. The model showed a relative error of 4.01% for earthquakes above $M4$ and 8.04% for smaller events. Results significantly improved when using nearby stations, with a mean absolute percentage error (MAPE) of 6.77% compared to 14.34% for distant stations. This suggests that distance plays a critical role in model accuracy, reinforcing the importance of single-station proximity in early warning systems. Tests included earthquakes greater and smaller than $M4$ at varying station distances. The model achieved an overall MAE of 0.23, excelling in large events with an MAE of 0.21. The results are competitive with other methods in the literature. However, performance may degrade in regions with sparse seismic networks, where station distance increases uncertainty in magnitude estimation.

In [61], MagPred, an ML ensemble model, was introduced for real-time earthquake magnitude estimation. The model addresses data imbalance by incorporating synthetic data generated with the conditional tabular generative adversarial network (CTGAN). It predicts magnitudes using the first 3, 4, and 5 s of P-wave signals, extracting 34 features related to waveform characteristics and site-specific parameters. The ensemble integrates XGBoost, random forest, LightGBM, and CatBoost, optimized with the sanitized gray wolf optimizer (SGWO). Unlike conventional ML models, MagPred leverages stacked ensemble learning, combining multiple algorithms to enhance prediction accuracy and robustness. The model outperforms traditional approaches like τ_c and P_d, achieving an MAE of 0.41, 0.40, and 0.38 for 3, 4, and 5 s of P-wave data, respectively. Additionally, it demonstrates generalization capabilities by performing well on earthquakes from different regions, such as India and Nepal, suggesting its applicability beyond the original training dataset. While effective for moderate earthquakes (M_s 3.0–5.0), it underestimates larger events, likely due to the limited presence of high-magnitude records. This limitation highlights the need for improved high-magnitude data representation, possibly through more advanced augmentation techniques or hybrid training strategies combining real and synthetic data. A key contribution is the use of GAN-based synthetic data, which improves magnitude estimation for early warning applications. However, the reliance on synthetic data may introduce biases, particularly if real earthquake distributions differ significantly from generated samples, potentially affecting performance in extreme cases. Future work could integrate real-time adaptive learning techniques to enhance robustness.

Finally, Ref. [34] introduced DFTQuake, a deep learning model designed for real-time early magnitude and peak ground acceleration (PGA) prediction, integrating tripartite Fourier transform attention and dendritic neural networks. The architecture incorporates a Fourier transform attention mechanism that captures seismic signal features in the frequency domain, enhancing interpretability and improving magnitude estimation accuracy. Unlike previous approaches, DFTQuake replaces traditional fully connected layers with a dendritic neural network, which introduces localized non-linearity, improving feature representation and reducing computational overhead. The model was trained using strong motion datasets from Japan, India, and other regions, demonstrating its ability to generalize across seismic environments. It processes the first three seconds of P-wave data and generates predictions with an MAE of 0.66 M_{JMA} for magnitude and 0.25 gal for PGA. Comparisons against existing models, including Vision Transformers (ViT) and LSTM-based networks, show that DFTQuake achieves a 13.16% improvement in magnitude estimation and a 34.21% improvement in PGA prediction. One key advantage of DFTQuake is its robustness to variations in epicentral distance, as it does not impose upper limits on distance ranges, unlike other deep learning models. However, despite its advantages, the model exhibits higher uncertainty for large-magnitude events due to the limited number of high-magnitude samples in training data. Future enhancements could involve integrating additional regional datasets and refining the dendritic neural network structure to further improve real-time prediction accuracy.

3.3. Key Insights into Machine Learning and Deep Learning Models for Magnitude Estimation

The analysis of the 28 selected articles has identified diverse approaches for earthquake magnitude estimation using ML and DL techniques. Most studies utilize deep neural networks, supervised learning models, and hybrid architectures that combine different techniques to improve the accuracy and robustness of predictions.

Table 2 presents the advantages, disadvantages, and applications of some representative models used in earthquake magnitude estimation. These models were selected based on their relevance in the literature, covering both traditional ML techniques and advanced DL architectures. The analysis of the models reveals several key trends in current research:

- Dominance of DL models: Architectures such as Bi-LSTM [7], Transformers [46], CNN-based networks [32], and hybrid approaches like MEANet [32] and MCFrame [52] have demonstrated high capability in capturing temporal and spatial features of seismic signals. These models improve early magnitude prediction by leveraging complex feature extraction mechanisms.
- Optimization strategies: Some studies employ ensemble learning models, such as EEWPEnsembleStack [49], which combine multiple predictors to reduce variance and enhance generalization across diverse seismic datasets.
- Computational considerations: While DL models offer superior accuracy, their computational demands can pose challenges for real-time applications, particularly in resource-constrained environments. In contrast, simpler models like SVM-M [40] perform well on small datasets but may struggle to maintain accuracy and scalability for complex, large-scale earthquake events.
- Balance between speed and accuracy: In early warning systems, rapid and accurate predictions are crucial. Models like MagNet [44] and real-time CNN-based approaches have been designed to operate efficiently, making them suitable for emergency response scenarios.

- Challenges and future directions: Despite advances in magnitude estimation, accurate prediction of large earthquakes ($M \geq 6$) remains challenging due to limited training data in this range. Additionally, generalizing models to different geological regions is still an open issue that requires further investigation.

Table 2. Comparison of machine learning models for magnitude estimation.

ML/DL Model	Advantages	Disadvantages	Application in Magnitude Estimation	Real-Time Feasibility
Bi-LSTM [7]	Captures long-term dependencies in seismic signals, robust to temporal distortions, effective for time-series prediction.	Higher computational cost than simple RNNs, requires large training datasets, prone to vanishing gradients in long sequences.	Used for direct magnitude regression based on waveform sequences, enhances feature extraction in hybrid architectures.	Efficient if optimized with pruning/quantization.
TEAM-LM [46]	Captures global dependencies, highly parallelizable, effective in handling large datasets.	High memory consumption, complex architecture, requires extensive pre-training.	Optimized for fast magnitude estimation, potential in regional/global earthquake monitoring.	Requires substantial GPU resources. LoRA fine-tuning improves feasibility.
MEANet [32]	Fast feature extraction, robust to noise, minimal manual preprocessing.	Limited to spatial features, lacks the ability to capture temporal dependencies.	Designed for rapid magnitude estimation from the initial P-wave signals, with potential applications in EEW.	Demonstrates high-speed processing, but has potential for EEW applications.
SVM-M [40]	Works well with small datasets, interpretable, effective for binary magnitude classification.	Scalability issues with large datasets, less effective in highly non-linear relationships.	Applied in magnitude classification rather than direct regression.	Provides a rapid initial estimate and can guide early response decisions.
MagNet [44]	Reduces noise through deep convolutions, and generates probabilistic magnitude estimates.	Requires large labeled datasets, sensitive to data distribution changes.	Used in probabilistic magnitude estimation and uncertainty quantification.	Deployable in real-time but requires fine-tuning for regional variations.
MCFrame [52]	Combines CNN feature extraction with RNN for temporal dependencies.	High computational cost, difficult to fine-tune hyperparameters.	Used in event-based magnitude classification and real-time seismic monitoring.	Requires optimizations for real-time use, but has potential for EEW applications.
EEWPEnsembleStack [49]	Combines multiple models to reduce bias and variance; improves accuracy.	Computationally expensive; requires careful feature engineering/hyperparameter tuning.	Reduces uncertainty in magnitude estimation, used in multi-source seismic data fusion.	Requires substantial processing power, limiting real-time deployment.

3.4. Comparison Between Machine Learning Techniques and Traditional Methods

The estimation of earthquake magnitude has traditionally relied on empirical methods such as τ_c [9] and P_d [10], which analyze the first few seconds of the P-wave to infer the event's magnitude. These approaches are widely used in operational EEW systems due to their simplicity, real-time applicability, and minimal data requirements for calibration, making them practical for deployment in different tectonic settings [68]. However, they also have limitations, including limited effectiveness for large-magnitude events and sensitivity to noise, which can affect their reliability in complex seismic environments.

On the other hand, recent advances in ML and DL have significantly improved the accuracy and reliability of magnitude estimation [4,7,33,61]. ML-based models, including XGBoost, RF, and SVMs, leverage extracted seismic features to enhance predictive performance, while DL architectures such as CNNs, LSTMs, and Transformers process raw waveform data without requiring feature engineering. These models have demonstrated enhanced generalization capabilities and robustness against noisy signals, making them highly promising for EEW applications.

3.4.1. Quantitative Comparison

Several studies have benchmarked ML and DL approaches against traditional empirical methods. Table 3 presents a comparative analysis of different magnitude estimation methods based on their MAE and standard deviation, highlighting the enhanced accuracy of ML and DL techniques compared to traditional approaches.

Table 3. Performance comparison between traditional and ML-/DL-based magnitude estimation methods.

Method	MAE (Magnitude Units)	Standard Deviation (σ)
τ_c [9]	1.637	1.45
P_d [10]	0.425	0.41
SVM-M [40]	0.297	0.31
CNN-LSTM [32]	0.25	0.25
Transformer [33]	0.15	0.20

As shown in Table 3, ML and DL models have demonstrated lower MAE and standard deviation compared to traditional approaches. The τ_c method presents the highest error among the methods considered, while modern Transformer-based architectures tend to exhibit lower uncertainty, suggesting that attention mechanisms may contribute to improving magnitude estimation.

3.4.2. Considerations

While ML and DL models offer advantages compared to traditional methods, their widespread adoption in EEW systems requires further validation across diverse seismic networks. Real-time deployment introduces additional constraints, including the need for low-latency inference and robust performance under varying seismic conditions. Challenges such as data imbalance, limited availability of large-magnitude event records, and computational demands for real-time deployment must be addressed to ensure reliable performance in operational settings [34,56].

Beyond computational limitations, several key factors explain the continued preference for traditional methods in operational seismic warning systems:

- Interpretability and trustworthiness: Traditional methods like τ_c and P_d are based on well-established physical principles, making them transparent and easy to validate. In contrast, ML/DL models function as "black boxes," where the decision-making process is not easily interpretable. However, recent advancements in explainable AI (XAI) [69] have introduced techniques to enhance interpretability, such as Grad-CAM [70], SHapley Additive Explanations (SHAP) [71], and local interpretable model-agnostic explanations (LIME) [72]. Although these techniques have been applied in various fields, such as healthcare [73], finance [74], and image recognition [75], their potential in seismology remains largely unexplored. Further research is needed to assess their effectiveness in improving the transparency of ML/DL models for seismic applications, ensuring that these methods can be reliably integrated into real-time earthquake early warning systems.
- Generalization and adaptability across different seismic regions: Traditional empirical models have well-defined calibration parameters that can be adapted with minimal regional data, whereas ML/DL models require large datasets to prevent overfitting. The variability in tectonic conditions across different regions can significantly impact the

performance of ML/DL approaches [4,68]. To improve model adaptability, researchers have explored techniques such as fine-tuning [76] and domain adaptation [77].

- Robustness in data-limited or noisy conditions: In regions with low-density seismic networks, traditional methods can still function with minimal data, whereas ML/DL models require large-scale datasets for training and may fail when encountering unseen or noisy data. To address these limitations, researchers have explored techniques such as data augmentation [78] and self-supervised learning [79].
- Regulatory and approval constraints in EEW systems:The deployment of new ML/DL models in operational seismic warning systems requires extensive validation, certification, and regulatory approval. Traditional methods have been used and optimized for decades, whereas ML-based approaches must undergo a rigorous evaluation before being integrated into national or regional networks [34,56].
- Practical benefits and real-world impact: Although ML-based models have demonstrated superior accuracy in magnitude estimation, their actual contribution to improving EEW performance and reducing earthquake-related losses remains an open research topic. Some studies suggest that ML models can reduce false alarms and missed detections, thereby improving public trust in EEW systems [4,7]. However, large-scale evaluations and real-world implementation studies are still needed to quantify their effectiveness in reducing casualties and economic damage. Future research should focus on systematic field tests that assess the real-time performance of ML-driven EEW systems in operational environments [61].

In summary, the integration of ML/DL models into EEW systems presents significant opportunities, but overcoming computational, data-related, and interpretability challenges will be essential for their widespread adoption. Additionally, further research is needed to quantify the practical benefits of ML-based approaches in reducing earthquake-induced losses, ensuring that these technologies not only improve accuracy but also contribute meaningfully to disaster risk mitigation.

4. Trends and Future Work

The use of advanced preprocessing techniques, such as noise reduction and time-frequency domain transformations [80], enables more robust analysis, especially for large-magnitude events that are underrepresented in databases due to their low-frequency content. Methods like the S-transform and generalized S-transform [81]have been used to enhance seismic data representation.

To address data scarcity, GAN-based data augmentation has been proposed [61]. However, further research is required to ensure that synthetic waveforms accurately capture real seismic patterns and do not introduce biases into ML models [34,56]. Hybrid approaches combining real and synthetic seismic data could further improve dataset balance, reducing bias toward specific magnitude ranges.

One of the main challenges in magnitude estimation is ensuring model applicability to different geological regions. While traditional models can be calibrated with minimal local data, ML/DL models require large datasets to prevent overfitting. Transfer learning allows pre-trained models on large seismic databases to be adapted to regions with limited data. For example, CNNs trained on global seismic records have shown significant improvements in magnitude estimation when fine-tuned for specific seismic events. Similarly, Transformer-based and recurrent models have leveraged transfer learning to refine their performance in predicting large magnitudes, where training data are scarce.

One promising strategy involves combining DL models with physics-based constraints. Incorporating wave propagation equations and geological characteristics into neural network architectures enhances estimation accuracy, particularly in under-instrumented regions. Recent studies have shown that integrating physical constraints into ML models reduces dependency on large labeled datasets and improves prediction stability, mitigating errors associated with training data biases [4]. Some ML/DL models exhibit overconfidence in their predictions, which can be problematic for early warning systems. To improve probabilistic prediction calibration, uncertainty quantification techniques, such as Bayesian deep learning, have been proposed [82]. These methods provide confidence intervals for magnitude estimates, increasing their reliability for EEW applications. ML models are highly sensitive to noisy or distorted data, which can impact magnitude estimation accuracy. To address this, strategies such as including noise in training data and using adaptive neural network architectures have been explored. These architectures dynamically adjust the time window size and the number of seismic stations used for prediction.

Recent studies have demonstrated that multi-station integration significantly improves magnitude estimation, suggesting that future models should prioritize hybrid approaches that leverage both single- and multi-station inputs dynamically [43,58]. Additionally, extending these models to practical applications, such as urban personal seismometers, could expand their usability in seismic monitoring [83].

The next crucial step in ML/DL model evolution involves validating them in operational EEW systems. To facilitate widespread adoption, these techniques must be tested in diverse regions and over extended operational periods. Real-time optimization and adaptation to different sensor configurations and data types are critical for achieving a precise and robust seismic monitoring system.

Furthermore, implementing these technologies in mobile devices for real-time earthquake alerts presents a significant opportunity to enhance EEW accessibility [84]. Extending real-time EEW applications to large-scale environments, including mobile sensor networks, could benefit both high-risk urban areas and remote locations with limited seismic infrastructure. The integration of ML/DL models into EEW systems presents significant opportunities, but addressing computational, data-related, and interpretability challenges remains crucial for widespread adoption. Figure 8 illustrates the key advancements in ML and DL for earthquake magnitude estimation. It highlights the transition from traditional empirical methods to feature-based models, the adoption of DL architectures such as CNNs and Transformers, and emerging hybrid approaches that integrate physics-based constraints and probabilistic models. These trends outline the direction of future research in optimizing real-time EEW applications.

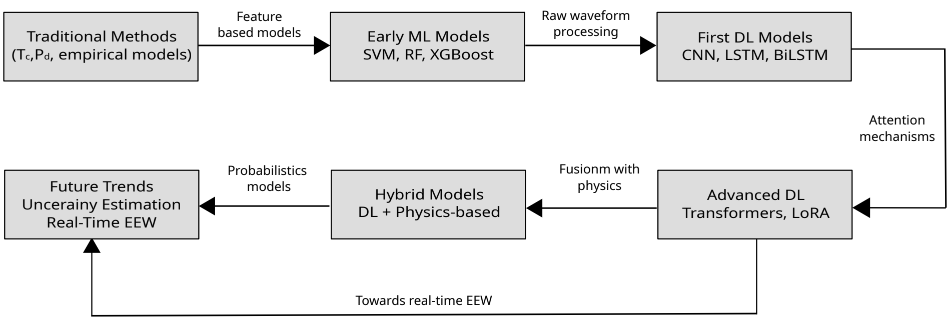

Figure 8. Evolution of ML and DL approaches for earthquake magnitude estimation, highlighting major advancements and future research directions.

Real-World Usage Scenario

A real-world usage scenario for the real-time estimation of earthquake magnitude using machine learning models is presented below. The process follows multiple stages, from data acquisition to result visualization. This scenario describes the workflow of an artificial intelligence-based system for early magnitude estimation.

The process begins with seismic stations continuously monitoring ground motion using broadband seismometers and accelerometers, along with digital seismic digitizers. These devices record waveforms in three components (Z, N, E) and transmit the data in real-time to a centralized processing system. There, the first 2 to 3 s of the P-wave are automatically extracted for analysis. To enhance signal quality before processing, filtering, and noise removal techniques are applied.

Once the preprocessed waveform data reaches the processing system, an AI-based model is applied for magnitude estimation. Deep learning architectures such as CNNs, Bi-LSTMs, and Transformers analyze waveform features to predict earthquake magnitude. The system is optimized for low-latency predictions, ensuring rapid magnitude estimation within milliseconds. Additionally, models are trained on regional seismic datasets to improve accuracy and adaptability to different tectonic settings. After magnitude estimation, the results are displayed and processed for further analysis. A real-time monitoring dashboard provides a software interface where raw and processed waveforms are visualized alongside estimated magnitude values and confidence levels. This scheme is graphically represented in Figure 9, illustrating the key stages from seismic data acquisition to magnitude estimation and visualization.

Figure 9. Graphical representation of a real-world usage scenario as a part of an EEW system.

5. Conclusions

Applying ML to early earthquake magnitude estimation is emerging as a transformative tool in modern seismology. In recent years, this field has experienced significant advancements, driven by the increasing availability of high-quality seismic datasets and enhanced computational capabilities. The findings of this review confirm that DL models, particularly CNN, RNN, and Transformer architectures, have demonstrated high accuracy in magnitude estimation using the first seconds of P-wave data. However, several critical limitations remain. One of the most pressing challenges is data scarcity for large-magnitude earthquakes, which not only increases uncertainty but also contributes to systematic underestimation in extreme events due to saturation effects in waveform features. Moreover, geological variability and seismic network heterogeneity introduce substantial generalization barriers, as models trained in specific regions often struggle to adapt effectively to areas with different tectonic characteristics. Addressing these issues requires advanced domain adaptation strategies, including transfer learning, fine-tuning techniques, and physics-informed corrections, to improve model robustness across diverse seismic environments.

Another key aspect influencing model performance is the quality of input data, including SNR, precise signal alignment, and waveform preprocessing methods. Incorporating noise augmentation and advanced preprocessing techniques, such as wavelet denoising and time-frequency domain transformations, has proven effective in enhancing model resilience against distorted signals and improving generalization capabilities. Additionally, data augmentation techniques have been essential in addressing dataset imbalances, particularly for small-magnitude events, which are often overrepresented in training data. One of the most promising applications of these technologies lies in EEW systems. Successfully integrating machine learning models into operational seismic networks has the potential to enhance disaster response strategies, reducing earthquake impacts on vulnerable communities. However, real-world validation remains imperative to ensure these models meet the reliability and precision standards required for critical applications. Interdisciplinary collaboration among scientific seismologists, engineers, and data scientists is essential to address computational limitations, scalability concerns, model interpretability, and future implementations in real-time.

This systematic review comprehensively analyzes machine learning-based approaches for earthquake magnitude estimation. Various architectures have been compared, evaluating their advantages and limitations concerning data availability, computational efficiency, and generalization capabilities. Additionally, key challenges and strategies for improving model robustness in real-time applications have been identified. These findings are crucial for future research and implementing ML-based models in seismic monitoring networks.

Author Contributions: Conceptualization, E.I.-G. and O.A.C.-A.; data curation, O.A.A.-C. and C.A.L.-M.; formal analysis, E.E.G.-G. and U.J.T.-P.; funding acquisition, E.I.-G.; investigation, A.N.-R. and O.A.C.-A.; methodology, A.N.-R., O.A.C.-A., and E.I.-G.; project administration, E.I.-G.; resources, E.E.G.-G. and U.J.T.-P.; software, A.N.-R. and O.A.A.-C.; supervision, E.I.-G. and O.A.C.-A.; validation, C.A.L.-M., E.E.G.-G., and U.J.T.-P.; visualization, A.N.-R. and O.A.C.-A.; writing—original draft, A.N.-R. and O.A.C.-A.; writing—review and editing, O.A.C.-A. and E.I.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Universidad Autónoma de Baja California (UABC) through the 25th internal call for research projects with grant number 402/6/C/53/25, and Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE). The authors also thank SECIHTI (Secretaría de Ciencia, Humanidades, Tecnología e Innovación) for the scholarship awarded to A. Navarro-Rodríguez.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Acknowledgments: The authors would like to express their sincere gratitude to Boris Rösler for his invaluable support in reviewing the technical information presented in this manuscript. His expertise, particularly in moment tensor analysis, provided crucial insights that contributed to the accuracy and robustness of this work. His constructive feedback and thoughtful suggestions were very helpful in refining the technical aspects of this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	artificial intelligence
AM	attention mechanism
APE	average prediction error
Bi-GRU	bidirectional gated recurrent unit
Bi-LSTM	bidirectional long short-term memory
BRT	boosted regression tree
CENC	China Earthquake Networks Center
CNN	convolutional neural network
CReLU	complex-valued convolutional layer
CRNN	convolutional recurrent neural network
CSES	China Seismic Experimental Site

CSN	community seismic network
CTGAN	conditional tabular generative adversarial network
DL	deep learning
DNN	deep neural network
DT	decision tree
EEW	earthquake early warning
FCN	fully convolutional network
GAP	global average pooling
GNN	graph neural network
GRU	gated recurrent unit
HR-GNSS	high-frequency global navigation satellite system
INSTANCE	The Italian Seismic Dataset for Machine Learning
IPOC	Integrated Plate Boundary Observatory Chile
K-NET	Kyoshin Network
KiK-net	Kiban Kyoshin Network
LIME	local interpretable model-agnostic explanations
LoRA	low-rank adaptation
LSTM	long short-term memory
MAE	mean absolute error
MAPE	mean absolute percentage error
MCFrame	machine learning magnitude classification framework
ML	machine learning
MLP	multilayer perceptron
MSE	mean squared error
MTL	multitask learning
M_w	moment magnitude
PGA	peak ground acceleration
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCGLs	residual connection graph layers
RF	random forest
RMS	root mean square
RMSE	root mean square error
RNN	recurrent neural network
SNR	signal-to-noise ratio
SGWO	sanitized gray wolf optimizer
SHAP	SHapley Additive Explanations
STA/LTA	short-time-average/long-time-average
STEAD	Stanford Earthquake Dataset
STFT	short-time Fourier transform
SVM	support vector machine
SVR	support vector regression
USGS	United States Geological Survey
ViT	Vision Transformer
XGBoost	extreme gradient boosting
M_b	body-wave magnitude
M_D	duration magnitude
M_L	local magnitude
M_s	surface-wave magnitude
M_{JMA}	Japan Meteorological Agency Magnitude
P_d	peak displacement
R^2	determination coefficient
σ	standard deviation
τ_c	effective average period
τ_p	predominant period

Appendix A

Table A1. Journals with one article.

Journal	Year
Geodesy and Geodynamics	2025
Engineering Applications of Artificial Intelligence	2025
Neural Computing and Applications	2024
Journal of Earth System Science	2024
Journal of South American	2024
Earth, Planets and Space	2024
Gondwana Research	2023
Solid Earth	2023
Journal of Asian Earth Sciences: X	2022
Geophysics	2022
Applied Sciences (Switzerland)	2022
Journal of Geophysical Research: Solid Earth	2022
Geophysical Research Letters	2020
Pure and Applied Geophysics	2020

Appendix B

Table A2. Authors with one article.

Author	Year
Jin Y	2024
Wang Z	2024
Quinteros-Cartaya C	2024
Hou B	2024
Dybing S N	2024
Wang Y	2023
Yoon D	2023
Ren T	2023
Meng F	2023
Ristea N	2022
Cofre A	2022
Song J	2022
Shakeel M	2022
Kuang W	2021
Mousavi S	2020
Chanda S	2020

Appendix C

Table A3. Multiple datasets in one article.

Country	Year
U.S.–Japan	2023
China–U.S.–Italy	2023
Chile–Italy–Japan	2021
Synthetics	2020
U.S.–Italy	2020

References

1. Asim, K.M.; Martínez-Álvarez, F.; Basit, A.; Iqbal, T. Earthquake magnitude prediction in Hindukush region using machine learning techniques. *Nat. Hazards* **2017**, *85*, 471–486.
2. Geller, R.J.; Jackson, D.D.; Kagan, Y.Y.; Mulargia, F. Earthquakes Cannot Be Predicted. *Science* **1997**, *275*, 1616.
3. Zhang, D.; Fu, J.; Li, Z.; Wang, L.; Li, J.; Wang, J. A Synchronous Magnitude Estimation with P-Wave Phases' Detection Used in Earthquake Early Warning System. *Sensors* **2022**, *22*, 4534.
4. Chakraborty, M.; Fenner, D.; Li, W.; Faber, J.; Zhou, K.; Rümpker, G.; Stoecker, H.; Srivastava, N. CREIME—A Convolutional Recurrent Model for Earthquake Identification and Magnitude Estimation. *J. Geophys. Res. Solid Earth* **2022**, *127*, e2022JB024595.
5. Richter, C.F. An instrumental earthquake magnitude scale. *Bull. Seismol. Soc. Am.* **1935**, *25*, 1–32.
6. Chung, D.H.; Bernreuter, D.L. Regional relationships among earthquake magnitude scales. *Rev. Geophys.* **1981**, *19*, 649–663.
7. Mousavi, S.M.; Beroza, G.C. A Machine-Learning Approach for Earthquake Magnitude Estimation. *Geophys. Res. Lett.* **2020**, *47*, e2019GL085976.
8. Nakamura, Y. On the Urgent Earthquake Detection and Alarm System (UrEDAS). In Proceedings of the Ninth World Conference on Earthquake Engineering, Tokyo/Kyoto, Japan, 2–9 August 1988; Volume VII, pp. 673–678.
9. Kanamori, H. Real-time seismology and earthquake damage mitigation. *Annu. Rev. Earth Planet. Sci.* **2005**, *33*, 195–214.
10. Kanamori, H.; Allen, R.M. Earthquake Early Warning Systems. *Science* **2003**, *300*, 786–789.
11. Böse, M.; Hauksson, E.; Solanki, K.; Kanamori, H.; Heaton, T.H. Real-time testing of the on-site warning algorithm in southern California and its performance during the July 29 2008 Mw5.4 Chino Hills earthquake. *Geophys. Res. Lett.* **2009**, *36*. https://doi.org/10.1029/2008GL036366.
12. Sarker, I.H. A machine learning based robust prediction model for real-life mobile phone data. *Internet Things* **2019**, *5*, 180–193.
13. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
14. Mousavi, S.M.; Beroza, G.C. Machine Learning in Earthquake Seismology. *Annu. Rev. Earth Planet.* **2023**, *51*, 105–129.
15. Wilkins, A.H.; Strange, A.; Duan, Y.; Luo, X. Identifying microseismic events in a mining scenario using a convolutional neural network. *Comput. Geosci.* **2020**, *137*, 104418.
16. Kuyuk, H.S.; Susumu, O. Real-Time Classification of Earthquake using Deep Learning. *Procedia Comput. Sci.* **2018**, *140*, 298–305.
17. Mousavi, S.M.; Zhu, W.; Sheng, Y.; Beroza, G.C. CRED: A Deep Residual Network of Convolutional and Recurrent Units for Earthquake Signal Detection. *Sci. Rep.* **2019**, *9*, 10267.
18. Wang, T.; Bian, Y.; Zhang, Y.; Hou, X. Classification of earthquakes, explosions and mining-induced earthquakes based on XGBoost algorithm. *Comput. Geosci.* **2023**, *170*, 105242.
19. Zainab, T.; Karstens, J.; Landsiedel, O. LightEQ: On-Device Earthquake Detection with Embedded Machine Learning. In Proceedings of the ACM International Conference Proceeding Series, San Antonio, TX, USA, 9–12 May 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 130–143.
20. Agathos, L.; Avgoustis, A.; Avgoustis, N.; Vlachos, I.; Karydis, I.; Avlonitis, M. Identifying Earthquakes in Low-Cost Sensor Signals Contaminated with Vehicular Noise. *Appl. Sci.* **2023**, *13*, 10884.
21. Zhao, Y.; Deng, P.; Liu, J.; Wang, M.; Wan, J. LCANet: Lightweight Context-Aware Attention Networks for Earthquake Detection and Phase-Picking on IoT Edge Devices. *IEEE Syst. J.* **2022**, *16*, 4024–4035.
22. Lim, J.; Jung, S.; JeGal, C.; Jung, G.; Yoo, J.H.; Gahm, J.K.; Song, G. LEQNet: Light Earthquake Deep Neural Network for Earthquake Detection and Phase Picking.*Front. Earth Sci.* **2022**, *10*, 848237.
23. Zhu, W.; Beroza, G.C. PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophys. J. Int.* **2019**, *216*, 261–273.
24. Choi, S.; Lee, B.; Kim, J.; Jung, H. Deep-Learning-Based Seismic-Signal P-Wave First-Arrival Picking Detection Using Spectrogram Images. *Electronics* **2024**, *13*, 229.
25. Sugondo, R.A.; Machbub, C. P-Wave detection using deep learning in time and frequency domain for imbalanced dataset. *Heliyon* **2021**, *7*, e08605.
26. Li, J.; Hao, M.; Cui, Z. A High-Resolution Aftershock Catalog for the 2014 Ms 6.5 Ludian (China) Earthquake Using Deep Learning Methods. *Appl. Sci.* **2024**, *14*, 1997.
27. Hsu, T.Y.; Wu, R.T.; Liang, C.W.; Kuo, C.H.; Lin, C.M. Peak ground acceleration estimation using P-wave parameters and horizontal-to-vertical spectral ratios. *Terr. Atmos. Ocean. Sci.* **2020**, *31*, 1–8.
28. Mandal, P.; Mandal, P. Peak ground acceleration prediction using supervised machine learning algorithm for the seismically hazardous Kachchh rift zone, Gujarat, India. *Nat. Hazards* **2024**, *120*, 1821–1840.
29. Somala, S.N.; Chanda, S.; Alhamaydeh, M.; Mangalathu, S. Explainable XGBoost-SHAP Machine-Learning Model for Prediction of Ground Motion Duration in New Zealand. *Nat. Hazards Rev.* **2024**, *25*. https://doi.org/10.1061/NHREFO.NHENG-1837.
30. Khosravikia, F.; Clayton, P. Machine learning in ground motion prediction. *Comput. Geosci.* **2021**, *148*, 104700.
31. Joshi, A.; Vishnu, C.; Mohan, C.K.; Raman, B. Application of XGBoost model for early prediction of earthquake magnitude from waveform data. *J. Earth Syst. Sci.* **2024**, *133*, 5.

32. Song, J.; Zhu, J.; Li, S. MEANet: Magnitude Estimation Via Physics-based Features Time Series, an Attention Mechanism, and Neural Networks. *Geophysics* **2022**, *88*, V33–V43.
33. Jin, Y.; Kim, G.; Ko, H. Classification and Magnitude Estimation of Global and Local Seismic Events Using Conformer and Low-Rank Adaptation Fine-Tuning. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5.
34. Joshi, A.; Vedium, N.R.; Raman, B. DFTQuake: Tripartite Fourier attention and dendrite network for real-time early prediction of earthquake magnitude and peak ground acceleration. *Eng. Appl. Artif. Intell.* **2025**, *144*, 110077.
35. Li, H.; Liu, C.; Wang, P.; Xiong, W.; Xu, X. Efficient GPU-accelerated seismic wave propagation simulations for nuclear structural safety assessment. *Comput. Methods Appl. Mech. Eng.* **2024**, *415*, 116305.
36. Xu, J.; Wang, Y.; Zhang, T.; Liu, R. A spatio-temporal model for real-time casualty estimation in earthquake-affected urban areas. *Nat. Hazards* **2025**, *110*, 203–221.
37. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int. J. Surg.* **2021**, *88*, 105906.
38. Mousavi, S.M.; Sheng, Y.; Zhu, W.; Beroza, G.C. Stanford Earthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access* **2019**, *7*, 179464–179476.
39. Michelini, A.; Cianetti, S.; Gaviano, S.; Giunchi, C.; Jozinović, D.; Lauciani, V. INSTANCE-the Italian seismic dataset for machine learning. *Earth Syst. Sci. Data* **2021**, *13*, 5509–5544.
40. Zhu, J.; Li, S.; Song, J. Magnitude Estimation for Earthquake Early Warning with Multiple Parameter Inputs and a Support Vector Machine. *Seismol. Res. Lett.* **2022**, *93*, 126–136.
41. Wang, Y.; Li, X.; Wang, Z.; Liu, J. Deep learning for magnitude prediction in earthquake early warning. *Data Driven Model.* **2023**, *123*, 164–173.
42. Quinteros-Cartaya, C.; Köhler, J.; Li, W.; Faber, J.; Srivastava, N. Exploring a CNN model for earthquake magnitude estimation using HR-GNSS data. *J. South Am. Earth Sci.* **2024**, *136*, 104815.
43. Münchmeyer, J.; Bindi, D.; Sippl, C.; Leser, U.; Tilmann, F. Low uncertainty multifeature magnitude estimation with 3-D corrections and boosting regression tree: Application to North Chile. *Geophys. J. Int.* **2020**, *220*, 142–159.
44. Kuang, W.; Yuan, C.; Zhang, J. Network-based earthquake magnitude determination via deep learning. *Seismol. Res. Lett.* **2021**, *92*, 2245–2254.
45. Joshi, A.; Raman, B.; Mohan, C.K. An integrated approach for prediction of magnitude using deep learning techniques. *Neural Comput. Appl.* **2024**, *36*, 16991–17006.
46. Münchmeyer, J.; Bindi, D.; Leser, U.; Tilmann, F. Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network. *Geophys. J. Int.* **2021**, *226*, 1086–1104.
47. Chakraborty, M.; Li, W.; Faber, J.; Rümpker, G.; Stoecker, H.; Srivastava, N. A study on the effect of input data length on a deep-learning-based magnitude classifier. *Solid Earth* **2022**, *13*, 1721–1729.
48. Ristea, N.C.; Radoi, A. Complex Neural Networks for Estimating Epicentral Distance, Depth, and Magnitude of Seismic Waves. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5.
49. Joshi, A.; Vishnu, C.; Mohan, C.K. Early detection of earthquake magnitude based on stacked ensemble model. *J. Asian Earth Sci. X* **2022**, *8*, 100122.
50. Cofre, A.; Marin, M.; Pino, O.V.; Galleguillos, N.; Riquelme, S.; Barrientos, S.; Yoma, N.B. End-to-End LSTM-Based Earthquake Magnitude Estimation with a Single Station. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5.
51. Wang, Z.; Chen, Z.; Zhang, H. EQGraphNet: Advancing single-station earthquake magnitude estimation via deep graph networks with residual connections. *Artif. Intell. Geosci.* **2024**, *5*, 100089.
52. Zhu, J.; Zhou, Y.; Liu, H.; Jiao, C.; Li, S.; Fan, T.; Wei, Y.; Song, J. Rapid Earthquake Magnitude Classification Using Single Station Data Based on the Machine Learning. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5.
53. Yoon, D.; Li, Y.; Ku, B.; Ko, H. Estimation of Magnitude and Epicentral Distance From Seismic Waves Using Deeper CRNN. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 3000305.
54. Shakeel, M.; Nishida, K.; Itoyama, K.; Nakadai, K. 3D Convolution Recurrent Neural Networks for Multi-Label Earthquake Magnitude Classification. *Appl. Sci.* **2022**, *12*, 2195.
55. Ren, T.; Liu, X.; Chen, H.; Dimirovski, G.M.; Meng, F.; Wang, P.; Zhong, Z.; Ma, Y. Seismic severity estimation using convolutional neural network for earthquake early warning. *Geophys. J. Int.* **2023**, *234*, 1355–1362.
56. Dybing, S.N.; Yeck, W.L.; Cole, H.M.; Melgar, D. Rapid Estimation of Single-Station Earthquake Magnitudes with Machine Learning on a Global Scale. *Bull. Seismol. Soc. Am.* **2024**, *114*, 1523–1538.
57. Meng, F.; Ren, T.; Liu, Z.; Zhong, Z. Toward earthquake early warning: A convolutional neural network for rapid earthquake magnitude estimation. *Artif. Intell. Geosci.* **2023**, *4*, 39–46.
58. Hou, B.; Zhou, Y.; Li, S.; Wei, Y.; Song, J. Real-time earthquake magnitude estimation via a deep learning network based on waveform and text mixed modal. *Earth Planets Space* **2024**, *76*, 1.

59. Chanda, S.; Somala, S.N. Single-Component/Single-Station–Based Machine Learning for Estimating Magnitude and Location of an Earthquake: A Support Vector Machine Approach. *Pure Appl. Geophys.* **2021**, *178*, 1959–1976.
60. Zhu, J.; Li, S.; Ma, Q.; He, B.; Song, J. Support Vector Machine-Based Rapid Magnitude Estimation Using Transfer Learning for the Sichuan-Yunnan Region, China. *Bull. Seismol. Soc. Am.* **2022**, *112*, 894–904.
61. Joshi, A.; Raman, B.; Mohan, C.K. Real-time earthquake magnitude prediction using designed machine learning ensemble trained on real and CTGAN generated synthetic data. *Geod. Geodyn.* **2025**, *17*, 18–29. https://doi.org/10.1016/j.geog.2024.10.001.
62. Mousavi, S.; Ellsworth, W.; Zhu, W.; Chuang, L.; Beroza, G. Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* **2020**, *11*, 3952.
63. Zhou, Y.; Yue, H.; Zhou, S.; Kong, Q. Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks. *Seismol. Res. Lett.* **2019**, *90*, 1079–1087.
64. Zhu, L.; Peng, Z.; McClellan, J.; Li, C.; Yao, D.; Li, Z.; Fang, L. Deep learning for seismic phase detection and picking in the aftershock zone of 2008 Mw7.9 Wenchuan Earthquake. *Phys. Earth Planet. Inter.* **2019**, *293*, 106261.
65. Allen, R.V. Automatic earthquake recognition and timing from single traces. *Bull. Seismol. Soc. Am.* **1978**, *68*, 1521–1532.
66. Lomax, A.; Michelini, A.; Jozinović, D. An Investigation of Rapid Earthquake Characterization Using Single-Station Waveforms and a Convolutional Neural Network. *Seismol. Res. Lett.* **2019**, *90*, 517–529.
67. Mousavi, S.M.; Beroza, G.C. Bayesian-Deep-Learning Estimation of Earthquake Location from Single-Station Observations. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8211–8224.
68. Zollo, A.; Amoroso, O.; Lancieri, M.; Wu, Y.-M.; Kanamori, H. A threshold-based earthquake early warning using dense accelerometer networks. *Geophys. J. Int.* **2010**, *183*, 963–974.
69. Doshi-Velez, F.; Kim, B. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv* **2017**, arXiv:1702.08608.
70. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.
71. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4766–4775.
72. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv* **2016**, arXiv:1602.04938.
73. Holzinger, A.; Langs, G.; Denk, H.; Zatloukal, K.; Müller, H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1312.
74. Samek, W.; Montavon, G.; Vedaldi, A.; Hansen, L.K.; Müller, K.R. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. *Lect. Notes Comput. Sci.* **2019**, *11700*, 1–8.
75. Zhang, Q.; Zhu, S.C. Visual Interpretability for Deep Learning: A Survey. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 27–39.
76. Howard, J.; Ruder, S. Universal language model fine-tuning for text classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, 15–20 July 2018; Volume 1, pp. 328–339.
77. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; March, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* **2016**, *17*, 2096–2030.
78. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60.
79. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning (ICML), Online, 13—18 July 2020; pp. 1575–1585.
80. Zhu, W.; Mousavi, S.M.; Beroza, G.C. Seismic Signal Denoising and Decomposition Using Deep Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9476–9488. https://doi.org/10.1109/TGRS.2019.2926772.
81. Cui, X.Z.; Hong, H.P. Use of discrete orthonormal s-transform to simulate earthquake ground motions. *Bull. Seismol. Soc. Am.* **2020**, *110*, 565–575.
82. Zhao, Q.; Rong, M.; Wang, J.; Li, X. An end-to-end multi-task network for early prediction of the instrumental intensity and magnitude in the north–south seismic belt of China. *J. Asian Earth Sci.* **2024**, *276*, 106369.
83. Löberich, E.; Long, M.D. Follow the Trace: Becoming a Seismo-Detective with a Campus-Based Raspberry Shake Seismometer. *Seismol. Res. Lett.* **2024**, *95*, 2538–2553.
84. Zheng, Z.; Wang, J.; Shi, L.; Zhao, S.; Hou, J.; Sun, L.; Dong, L. Generating phone-quality records to train machine learning models for smartphone-based earthquake early warning. *J. Seismol.* **2022**, *26*, 439–454.