Biomedical Signal Processing and Control 117 (2026) 109518

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Deep learning for epileptic seizure prediction from EEG signals: A review

Xindi Huang [a], Hongying Meng [a,*], Zhangyong Li [b]

[a] Department of Electronic and Electrical Engineering, Brunel University of London, London, UK
[b] School of Optoelectronic Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China

ARTICLE INFO

Keywords:
Epileptic seizure
Epileptic seizure prediction
Deep learning
Electroencephalogram

ABSTRACT

Epilepsy, a chronic noncommunicable brain disease affecting nearly 1% of the global population across all ages, manifests through seizures caused by abnormal electrical activity in the brain. Electroencephalogram (EEG) records the spontaneous electrical activity of the brain which is more suitable for analysing Epileptic Seizure (ES) than other modalities such as functional Near-Infrared Spectroscopy (fNIRS) and functional Magnetic Resonance Imaging (fMRI). ES prediction aims to provide advanced warning to patients, allowing timely intervention and preventing dangerous situations. Deep Learning (DL) has emerged as a promising approach for ES prediction due to its superior noise removal capabilities, nonlinear feature representation, and strong classification ability. This paper presents a comprehensive review of DL-based approaches for ES prediction in last 5 years, highlighting current research trends, identifying existing challenges, and suggesting potential future research directions.

Contents

* Corresponding author.
E-mail address: hongying.meng@brunel.ac.uk (H. Meng).

https://doi.org/10.1016/j.bspc.2026.109518
Received 11 March 2025; Received in revised form 29 November 2025; Accepted 8 January 2026
Available online 17 January 2026

1. Introduction

Epilepsy is a chronic neurological disorder affecting individuals of all ages globally, with an estimated 50 million people affected worldwide according to the most recent data from the World Health Organization and the World Health Assembly for 2021 [1]. Epileptic Seizure (ES) is the clinical manifestation of an abnormal, excessive, hypersynchronous discharge of a population of cortical neurons [2]. ES can affect a person in different ways, depending on which part of the brain is affected. Symptoms may occur include uncontrollable convulsions and tremors, loss of consciousness, blank stares, fainting and so on [3,4].

The major danger of epilepsy is that it is lethal and can cause death in humans or animals. The mortality rate for epilepsy in adults is as high as 3.6%. Although anti-epileptic drugs provide sufficient control of seizures in approximately 70% of patients with epilepsy, the remaining 30% are resistant to monotherapy [5]. Therefore, there are about 20 million drug-resistant patients globally, which means they are exposed to severe life-long adverse outcomes, such as accidental death from seizures, increased risk of injury, and learning and developmental disabilities at school age.

Electroencephalogram (EEG) records brain activity using electrophysiological indicators. It is closely associated with ES caused by abnormal brain discharges. In addition, as one of the brain wave activity measurements, EEG has attractive properties such as high temporal resolution, relatively low cost, high portability, and less risks to users [7,8] in comparison with other modalities. In clinical recordings, clinician analysis EEG signals by two stages (ictal and inter-ictal) as shown in Fig. 1. The ictal state in EEG signal is the situation during an ES. The inter-ictal state is defined as the situation between one seizure and the next [9].

ES prediction refers to the use of advanced technologies to anticipate the onset when an ES occurs. After the onset is predicted before the ictal stage, an warning alert can be issued. Through this alert, patients can get timely treatment to avoid life-threatening conditions.

For the development of ES prediction algorithms, EEG recordings are segmented firstly in two ways. One commonly used way is to divide the raw EEG signals into Seizure Occurrence Period (SOP) and Seizure Prediction Horizon (SPH) as shown in Fig. 1. SOP is defined as a time period during which the seizure is to be expected. In addition, for therapeutic intervention, there must be a minimum time window between when an alert is issued and when the SOP begins. This time window is referred to as the SPH. Considering the two time periods, SPH and SOP, the definition of a correct ES prediction is as follows: no seizure has occurred during the SPH after the alarm signal and a seizure during the SOP period [10]. Different ES prediction algorithms choose different lengths of time for SPH and SOP. For clinical use, the SPH must be long enough to allow for appropriate interventions or preventive measures. On the contrary, the duration of SOP should not be too long to minimise patient anxiety. Another way is to define the pre-ictal and ictal states as well as an intervention period. The pre-ictal state is only apparent for a period of time prior to the onset. The duration of the pre-ictal state is usually an hour [11] or 30 min [12]. Also, an intervention time needs to be defined between the ictal and the pre-ictal state, often defined as 1 min [12] or 5 min [11]. This enables the algorithm to predict seizures by analysing the different characteristics of the pre-ictal and inter-ictal states and to treat the patient within the intervention time. As illustrated in Fig. 2, we present an close-loop flowchart spanning from EEG acquisition and preprocessing to the integration of deep learning models with predictive outputs, serving as an overview of the paper's structure.

Table 1
List of abbreviations.

Abbreviation	Full form
AES	American Epilepsy Society
AUC	Area Under the Curve
Bi-LSTM	Bi-directional Long Short-Term Memory
CHB-MIT	Children's Hospital Boston–MIT Database
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DL	Deep Learning
DWT	Discrete Wavelet Transform
EEG	Electroencephalography
EDF	European Data Format
ES	Epileptic Seizure
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
FPR	False Prediction Rate
GAN	Generative Adversarial Network
GCN	Graph Convolutional Network
GNN	Graph Neural Network
iEEG	Intracranial Electroencephalography
LOPOCV	Leave-One-Patient-Out Cross-Validation
LOSOCV	Leave-One-Seizure-Out Cross-Validation
LSTM	Long Short-Term Memory
ResNet	Residual Neural Network
RNN	Recurrent Neural Network
Siena	Siena sEEG Database
sEEG	Scalp Electroencephalography
SNN	Spiking Neural Network
SOP	Seizure Occurrence Period
SPH	Seizure Prediction Horizon
STFT	Short-Time Fourier Transform
SWEC-ETHZ	SWEC-ETHZ iEEG Database
TUSZ	TUH EEG Seizure Corpus
ViT	Vision Transformer

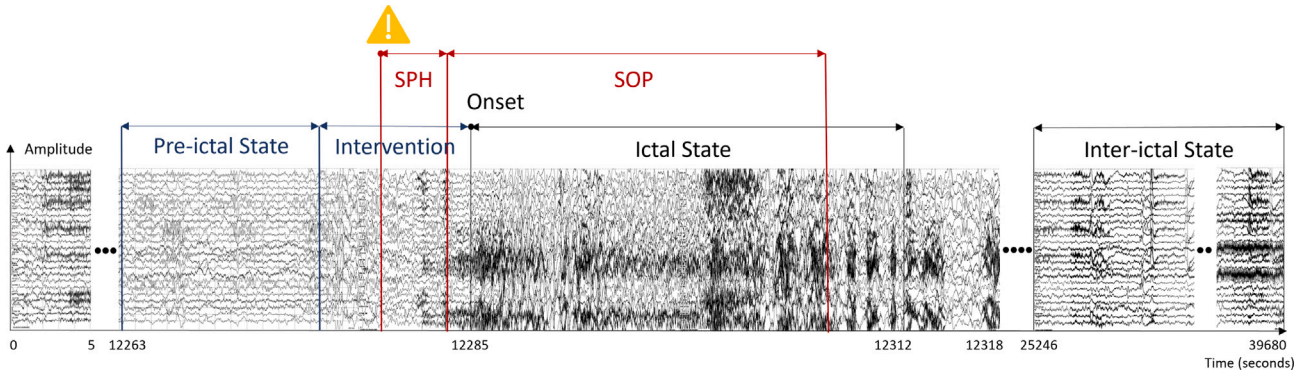

Fig. 1. An EEG recording example from CHB-MIT database [6] (e.g. the second seizure of patient 1). The ictal stage is labelled from 12285 to 12312 s, and the inter-ictal stage is chosen from 25246 to 39680 s. The correct prediction represents a ES that did not occur in the SPH time period after the warning was issued, but was found in the subsequent SOP time period.

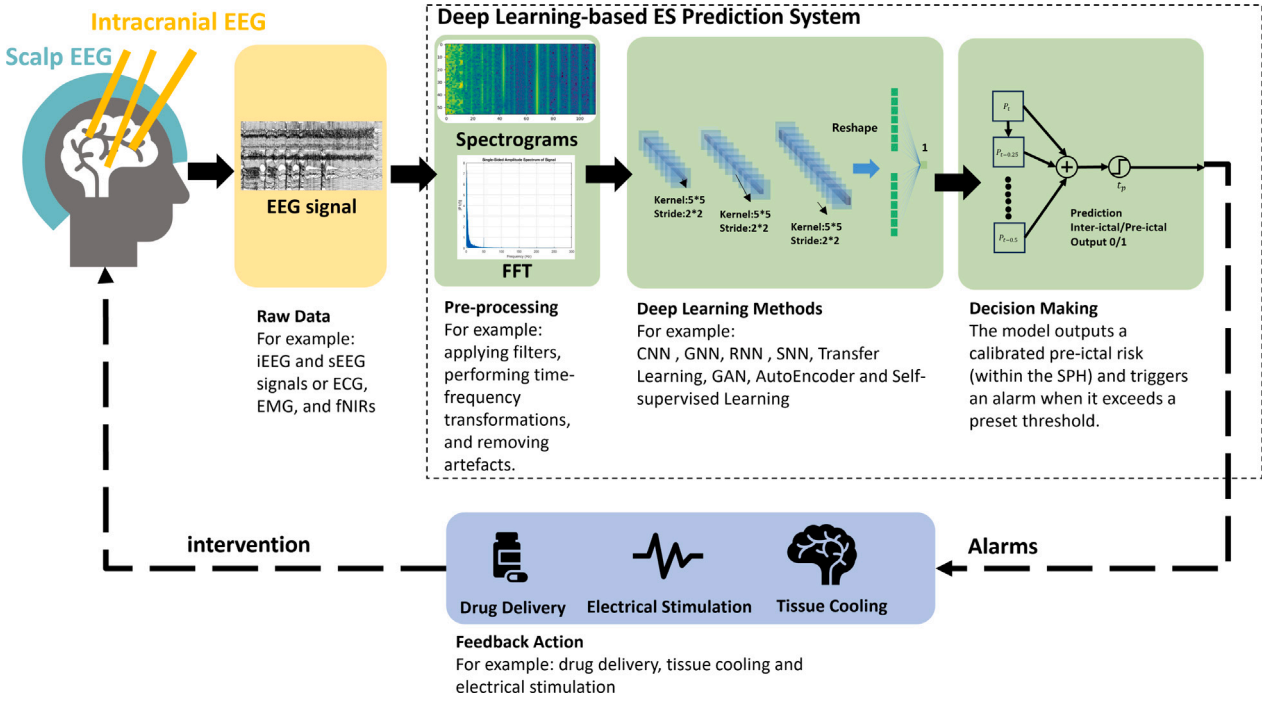

Fig. 2. Unified workflow diagram for ES prediction: From raw signal and preprocessing, the data inputted deep learning methods. This outputs risk within the specified SPH threshold, triggering a threshold-based alert. The controller then coordinates with external devices to record the event. Then, controller will interfere the brain back to the normal state.

Table 2
Comparison between this work and previous review articles on ES prediction across different years.

Paper	Year	Focused area	Database summary	DL	Future direction	Period
Mormann et al. [13]	2007	ES Prediction	✗	✗	✓	1975–2007
Kuhlmann et al. [14]	2018	ES Prediction	✗	✗	✓	2013–2018
Rasheed et al. [15]	2020	ES Prediction	✗	✓	✓	2018–2020
Baud et al. [16]	2022	ES Prediction	✗	✗	✓	2018–2022
Shoeibi et al. [17]	2022	ES Prediction and Detection	✓	✓	✓	2018–2021
This Work	2024	ES Prediction	✓	✓	✓	2020–2024

Structure: Section 2 provides an overview of EEG databases for ES prediction. Section 3 provides a comprehensive review of DL methods from supervised and unsupervised learning. Section 4 outlines prospects for future research in this area. Finally, Section 5 summarises the main contributions. Due to the frequent use of abbreviations in this work, a list of abbreviations is provided for clarity shown in Table 1.

Contribution: Previous reviews on ES prediction have focused on machine learning applications with a cut-off date of around 2020 on generic databases separately [15–17]. Rasheed et al. [15] provided an overview of machine learning algorithms applied to ES prediction up to 2020, while Baud et al. [16] pointed out significant advances in ES prediction over the past decade as well as persistent challenges. Shoeibi et al. [17] summarised the DL algorithms but they did not categorise the algorithms related to ES prediction separately. This paper aims to provide a complete overview of the current state on ES prediction using DL algorithms. Specifically, We summarise detailed DL

Table 3
Summary of publicly available iEEG and sEEG datasets used for ES prediction.

Type	Database	No. of Patients	No. of Channels	Sampling Frequency (Hz)	No. of Seizures	Duration Average (hr)	Electrode Location Info	Algorithms
iEEG	Freiburg	21	128	256	87	708	Yes	GNN, CNN, GAN
	AES Prediction Challenge	5 Dogs, 2 Patients	16	400, 5000	48	622	No	CNN, RNN, Self-supervised
	Epilepsy-ecosystem (Melbourne-Univ. AES-MathWorks)	3	16	400	1139	10 608	No	GAN
	Epilepsy-ecosystem (My Seizure Gauge)	10	–	–	–	–	No	–
	SWEC-ETHZ (Long-Term)	18	32–88	512/1024	116	–	No	CNN, RNN
	SWEC-ETHZ (Short-Term)	16	36–100	512	100	–	No	CNN, RNN
	Bonn Dataset (Epileptic Patients)	5	1	173.61	–	–	No	CNN
sEEG	CHB-MIT	22	23	256	182	844	Yes	CNN, RNN, GNN, GAN, AutoEncoder
	TUSZ	10 874	20–31	Minimum 250	–	–	Yes	Self-Supervised
	Siena	16	21/29	512	47	–	Yes	–
	Bonn Dataset (Healthy People)	5	1	173.61	–	–	No	CNN

algorithms for ES prediction after 2020, widely used databases in the field, and explores potential future research directions. Table 2 shows the comparative analysis of this work with existing review publications.

2. Database

For DL algorithm development on ES prediction, high-quality data is very important [18] and there are a few datasets have been produced. EEG signals can be divided into two categories: one is intracranial EEG (iEEG) signals that is invasive with higher Signal-to-Noise Ratio (SNR) and fewer artefacts. However, it requires implantation of EEG sensors into the brain through a craniotomy or drilling through the skull, which may cause long-term inflammation of the patient's brain [19]. Another one is a non-invasive method called scalp EEG (sEEG). Compared to iEEG signal, sEEG signal is collected through a wearable device mounted on the scalp with relatively simple and safe way for patients. Since sEEG capture leaky electricity from neurons and it is in a lower SNR and greater susceptibility to artefacts [20].

Starting with the first International Collaborative Workshop on ES Prediction in Bonn in 2002, which provided a set of five epilepsy-related EEG recordings over multiple consecutive days [13], more and more data have been collected for analysing the problem of ES prediction. Both the acquisition and labelling of EEG signals need to be handled by medical specialists, especially when acquiring iEEG signals. Also, most EEG signals are collected by hospitals. The reason is that the database related to ES requires the support of the patient. The existing databases such as Freiburg Hospital EEG Database (Freiburg) [21], CHB-MIT sEEG Database (CHB-MIT) [6], Epilepsy-ecosystem [22], SWEC-ETHZ iEEG Database (SWEC-ETHZ) [23], Bonn Dataset [24], Siena sEEG Database (Siena) [25], American Epilepsy Society (AES) Prediction Challenge [26] and TUH EEG Seizure Corpus (TUSZ) [27] were proposed (see Table 3).

2.1. iEEG

2.1.1. Freiburg hospital EEG dataset

The Freiburg contains iEEG signals from 21 patients with medically refractory focal epilepsy and records the location of epileptic focal in different patients. The database was collected by the Neurofile NT digital video EEG system with 128 channels and 256 Hz sampling rate. The total number of seizures was 87, with an average of 708 h of EEG signals recorded. This database is provided in European Data Format (EDF). Preprocessing steps commonly include band-pass filtering (e.g., 0.5–150 Hz) and notch filtering at 50 Hz to remove power-line interference. Additionally, artefact rejection may be required to handle artefacts caused by motion, electrode drift, or clinical procedures. Since the records were collected in a preoperative setting, patients may have been in a specific clinical state that may not reflect their daily seizure activity, which may affect the performance of the algorithm.

2.1.2. AES prediction challenge dataset

The AES Prediction Challenge was posted to Kaggle in 2014. Compared to Freiburg Hospital, It recorded the iEEG signal from epileptic patients and dogs. This database used a dynamic monitoring system to record signals from dogs with epilepsy. Recorded iEEG signals were sampled at 400 Hz from 16 electrodes. The recorded voltages were referenced to the group average for dogs. The recordings were of long duration, ranging from a few months to a year, with some dogs recording up to hundreds of seizures. For iEEG signals from epileptic patients, electrodes with different patients were sampled at a frequency of 5000 Hz. The recorded voltages were referenced to electrodes outside the brain for patients. Although the dataset contains both dogs and human recordings, differences in acquisition settings (e.g., sampling frequency) raise concerns about the cross-species generalisability of models trained on mixed data.

2.1.3. Epilepsy-ecosystem dataset

The Epilepsy-ecosystem is a crowd-sourcing ecosystem for ES prediction. There are two subsets in Epilepsy-ecosystem. The first subset contains inter-ictal and pre-ictal signals from the 'Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge' that was hosted on Kaggle in 2016. The competition focuses on predicting seizures using long-term human EEG signals obtained from the world's first clinical trial of the implantable NeuroVista Epilepsy Advisory System [19]. The iEEG signals were sampled from 16 electrodes at 400 Hz. These recordings range in duration from months to year. The recorded voltages are referenced to the average of the electrode. It contains 1139 seizures from three patients. This is currently the only database of continuous iEEG signals collected from a clinical situation.

The second subset in Epilepsy-ecosystem is 'My Seizure Gauge' data which is a unique wearable device database that contains long-term records of epileptic patients [28]. The data was collected with funding from the Epilepsy Foundation of America to develop ES prediction algorithms for non-invasive wearable devices. The database provides continuous wearable device recordings from 10 patients, along with seizure duration and recording metadata (number of channels,

channel/biosignal type, sampling rate, total number of samples per recording, and timestamp). There are 3–5 days of data available for each patient. For the first time, this database collects biosignals other than EEG signals from patients with epilepsy.

2.1.4. SWEC-ETHZ dataset

In the SWEC-ETHZ, there are two subsets containing long-term and short-term iEEG signals. Both of them were recorded intracranially by strip, grid, and depth electrodes. All iEEG recordings were visually inspected by an experienced board-certified EEG epileptologist (K.S.) to determine seizure onset and end times and to exclude channels consistently interfered with by artefacts.

The long-term dataset included 116 seizures from 2656 h of anonymised continuous iEEG signal in patients with pharmacologically resistant epilepsy. The sampling frequency is 512 or 1024 Hz and the number of electrodes ranged from 32 to 88. Each file holds one hour of continuous iEEG signals. The detailed information for every subject is contained defining the sampling frequency, and beginning and end of the seizures. As a dataset containing long-term and continuous iEEG signals, it fits well with the requirements proposed in the review [14].

The short-term dataset consisted of 100 anonymised iEEG from 16 patients with drug-resistant epilepsy sampled at 512 Hz. Each recording consisted of a 3-minute pre-ictal segment, a ictal segment, and a 3-minute post-ictal segment. The number of electrodes ranged from 36 to 100.

The SWEC-ETHZ is a valuable resource for ES prediction research. However, it is important to be aware of its limitations. First, although the long-term subset contains 116 seizures over 2656 h, the relatively small number of patients limits the generalisability of models trained on this data. Second, electrode implantation varies widely across patients, and detailed electrode location metadata are not consistently standardised. This can hinder the reproducibility of research and complicate comparisons across patients. Finally, the short-term dataset provides only fixed pre-ictal, ictal, and post-ictal segments, which may not accurately reflect real-world continuous monitoring scenarios and could artificially simplify the prediction task.

In terms of preprocessing, SWEC-ETHZ recordings are distributed in EDF. Common steps reported in the literature include band-pass filtering (e.g., 0.5–150 Hz) and notch filtering at 50 or 60 Hz to remove power-line noise. For the long-term dataset, continuous iEEG is segmented into fixed-length epochs (typically 30 s or 60s), while the short-term dataset is already pre-segmented. After artefact removal, normalisation methods are applied to reduce inter-patient and inter-channel variability. This is particularly important because the number of channels varies significantly among patients in this dataset.

2.2. sEEG

2.2.1. CHB-MIT dataset

The CHB-MIT is one of the most widely used public databases for ES prediction research. It is collected by Boston Children's Hospital contained in the Massachusetts Institute of Technology. It provides recordings in EDF with a sampling rate of 256 Hz and 23 channels. The database consists of sEEG recordings from children with intractable seizures. A total of 23 recordings from 22 patients were included. Patients were monitored for up to several days after discontinuing anti-epileptic drugs to characterise seizures and assess their suitability for surgical treatment. A total of 182 seizure onset and end times were recorded. The International 10–20 system [29] of EEG electrode positions and nomenclature was used for these recordings. Pre-processing steps include segmenting continuous EEG into fixed-length windows (e.g., 30 s or 60s), applying band-pass filtering (typically 0.5–70 Hz) to remove low-frequency drift and high-frequency noise, and using a 60 Hz notch filter to eliminate power-line interference. Artefacts like eye blinks and muscle activity are usually removed either manually or through Independent Component Analysis (ICA). Additionally, normalisation techniques such as Z-score or min–max scaling are often applied to standardise the input data across patients. Although this dataset is widely used, several limitations should be noted. First, the dataset only includes paediatric patients, limiting its generalisability to adult and elderly populations. Second, the relatively small number of patients and seizures makes it difficult to train large-scale deep learning models, and for some patients, the number of seizure events is too limited to allow for reliable cross-patient evaluation. Third, the recording conditions are relatively homogeneous (23-channel EEG, sampling rate of 256 Hz), which may not reflect variability in other clinical settings, particularly those using iEEG or high-density sEEG. Additionally, although seizure onset and termination annotations are provided by clinical experts, inter-rater variability and the lack of a standardised definition of pre-ictal states may affect the consistency of model evaluations. Finally, while models trained on CHB-MIT often achieve high performance, their accuracy often decreases significantly when applied to other datasets, indicating limited external validity. These factors should be carefully considered when interpreting results from studies based on CHB-MIT.

2.2.2. TUSZ

The TUSZ consists of 16,986 records from 10,874 patients, each containing at least one EDF file. If the long-term monitoring record is divided into multiple files, more than one physician report is included. The EDF file usually contains EEG-specific channels and auxiliary channels. The number of specific channels ranges from 20 to 31. The auxiliary channels contain detected bursts, ECG, EMG and light stimuli. The EEG signals are mostly sampled at a frequency of 250 Hz. This database contains the largest number of patients and the largest number of records compared to other databases. It also includes recordings of biological signals other than EEG signals such as EMG, ECG.

2.2.3. Siena dataset

The Siena contains 14 patients with epilepsy. Subjects were monitored using video-electroencephalography with a sampling rate of 512 Hz, with electrode arrangements based on the International 10–20 System [29]. In total, the database contains 47 seizures on about 128 recording hours. In contrast to other databases, all recordings in Siena includes 1 or 2 EKG signals.

2.3. iEEG and sEEG

2.3.1. Bonn dataset

The Bonn dataset is composed of EEG data from five healthy individuals and five patients with epilepsy and contains a total of five data subsets which are A, B, C, D, and F. Single-channel EEG data were collected, where each subset dataset contained 100 data segments. Each data segment has a time length of 23.6 s, a sampling frequency of 173.61 Hz, a resolution of 12 bits, and 4096 data points.

Subsets A and B are sEEG signals collected from five healthy individuals, constituting the control group. The segment in A is the EEG when the subject's eyes are open, and the EEG when the subject's eyes are closed in B.

Subsets C, D, and E are iEEG taken from five patients with preoperative diagnosis. C and D were collected during the inter-ictal period, and E was collected during the ES period.

Unlike other datasets, the Bonn dataset includes sEEG recordings from healthy individuals. This enables more robust comparative analyses of differences between healthy and epileptic patients across EEG segments.

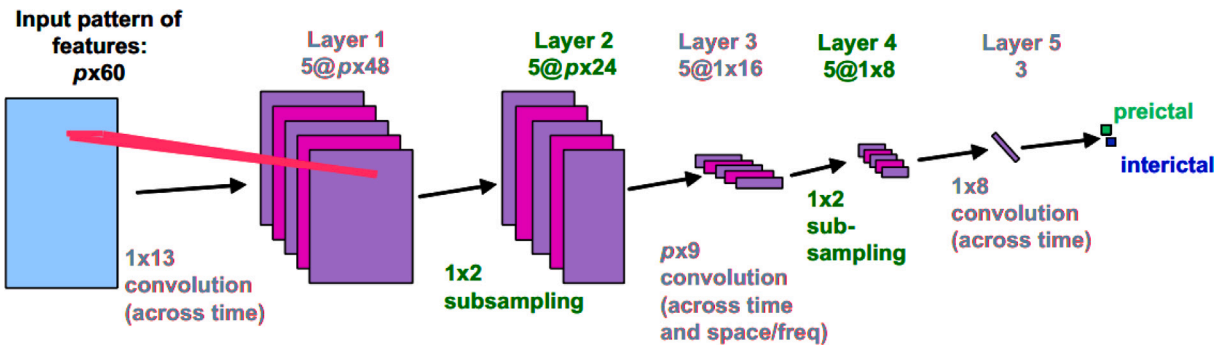

Fig. 3. The input to this convolutional network is in the form of an EEG of 60 consecutive frames (5 min) containing p features. The first and third layers of this network are convolutional layers with dimensions $1 * 13$ and $p * 9$, respectively. The second and fourth layers are sub-sampling layers that perform $1 * 2$ sub-sampling operations. The fifth layer is a fully connected layer. The output of this layer is the classification result of pre-ictal and inter-ictal periods [30].

3. Deep learning

Over several decades, Deep Learning (DL), which is a branch of Artificial intelligence (AI) and Machine Learning (ML) [31], has been applied to solve the ES prediction problem. DL algorithms can identify the current stage of the brain by analysing trends in EEG signals and the features of different stages. Thus, warning signals are sent before the ictal stage [14,32]. Also, previous research demonstrated that DL algorithms can significantly improve the accuracy and utility of biomedical signal analysis, especially in areas such as disease diagnosis and treatment [33]. Fig. 2 demonstrates the categorisation of DL algorithms for ES prediction. DL algorithms can be categorised into two types: supervised learning and unsupervised learning. Sections 3.1 and 3.2 respectively review the principal advances in supervised and unsupervised learning for ES prediction over the period 2020–2024.

3.1. Supervised learning

Supervised learning algorithms predict ES by distinguishing between pre-ictal and inter-ictal EEG signals. These algorithms are trained on labelled data where seizures are known to exist, allowing them to learn patterns that predict impending ES. From a comparative perspective, the supervised learning methods reviewed in Table 4 can be categorised into four major classes: (i) convolutional architectures, (ii) recurrent models, (iii) graph neural networks, and (iv) Transformers and attention-based approaches. When classified by dataset, the commonly used datasets primarily comprise three: CHB-MIT, AES, and Freiburg. We have summarised statistics according to these two classification approaches, consolidating the results in Tables 5 and 6. The following presents a statistical analysis of these tables.

The average sensitivity across all models was 93.75% (standard deviation = 6.19%, n = 43). However, a systematic comparison of deep learning models revealed significant trade-offs between different model families. Transformer and attention-based models exhibited the lowest average sensitivity (90.92%) and lowest average false positive rate (FPR = 0.033/h), indicating a more conservative alarm behaviour that suppresses false positives at the cost of missed pre-ictal events. The high standard deviation suggests poorer stability across different experimental settings. In contrast, the CNN architecture exhibits robust statistical stability with a lower standard deviation (SD = 4.41), yet a relatively high average false positive rate (FPR = 0.167/h), potentially highlighting a trade-off between robust feature extraction and potential sensitivity to non-epileptiform artefacts. Furthermore, the GNN approach maintained a favourable equilibrium, achieving high sensitivity (94.05%) and low false positive rates (0.075/h), with minimal variance in false positives (SD = 0.022/h), thereby underscoring its potential for effectively modelling stable epileptic network dynamics.

Furthermore, performance analysis by dataset revealed critical issues concerning generalisation capability. Although studies based on CHB-MIT reported high average sensitivities (94.16%), methods evaluated using more stringent, heterogeneous datasets (such as AES) reported significantly lower average FPR. This substantial disparity in FPR confirms that dataset homogeneity and cross-validation protocols critically influence the generalisability of reported results.

In summary, supervised deep learning approaches demonstrate considerable potential for ES prediction, yet their apparent high performance must be interpreted in light of the architecture's inductive bias, dataset homogeneity, and the critical balance between sensitivity, false positive rate, and clinical generalisability. The clinical viability of a model hinges not only on its high average sensitivity but also on the critical balance between stability (low standard deviation) and false positive rate (FPR < 0.05/h). Furthermore, regarding FPR robustness, architectures utilising GNNs or Transformers inherently exhibit superior FPR suppression (FPR < 0.08/h) and lower FPR variance, indicating their inductive biases are more adept at filtering clinical noise. However, due to the limited statistical data available, further work is required to validate this finding. Crucially, despite CNNs exhibiting higher FPR, they demonstrate the highest statistical robustness (lowest sensitivity standard deviation, approximately 4.41). This analysis thus provides a viable approach for future hybrid architecture deep learning models. Such models could combine CNNs' reliable local feature extraction capabilities with GNNs' or Transformers' global low-FPR contextual modelling abilities, thereby achieving high stability and clinical utility.

3.1.1. Convolutional Neural Networks (CNNs)

From a neurophysiological perspective, the inherent inductive biases of CNNs towards locality and translation invariance render them exceptionally well-suited to capturing the spatio-temporal patterns characteristic of events such as inter-ictal spikes and sharp waves, as well as pre-ictal rhythmic discharges. The core of this approach lies in treating convolutional filters as learnable feature detectors. As filters scan the raw EEG or its spectrograms along both the temporal and channel dimensions, they effectively identify and encode recurring waveform features in a manner analogous to visual inspection by clinicians. This efficient extraction of local, context-dependent features constitutes the key physiological basis enabling relatively shallow CNN models to achieve high performance across numerous epilepsy prediction benchmarks.

The earliest CNN was proposed by Lecun et al. [34] in 1998. As a DL algorithm inspired by the visual working mechanism of biology, CNN is widely used in the field of computer vision [35]. As early as 2009, Mirowski et al. applied CNNs to the problem of ES prediction [30]. They compared the performance of CNNs, logistic regression and SVM in ES prediction from iEEG signals. The features used were manually designed inter-channel bi-variate features (cross-correlation, nonlinear interdependence, dynamical entrainment or wavelet synchrony) that encoded relationships between pairs of EEG channels. Then, the CNN

is applied to extract the features to classify the pre-ictal and inter-ictal periods. The detailed network architecture is illustrated in Fig. 3.

Then, Khan et al. [36] applied CNN into sEEG signals. They performed a continuous wavelet transform (CWT) on each channel of the EEG to obtain a tensor of wavelet coefficients with three modes: time, scale and channel. The CWT can use windows of different sizes. For example, a narrow window is used for high frequencies and a wide window is used for low frequencies. This approach allows for better analysis of time-frequency domain features. Then, the transformed signals are inputted into the CNN to extract the features. The CNN is trained using labelled data (i.e. inter-ictal, pre-ictal and ictal states). The output of the CNN is a probability distribution of these three categories, which is used to predict when a ES is likely to occur.

After that, Truong et al. [37] proposed firstly transformed the EEG signals to spectrograms by short-time Fourier transform (STFT). Then, they applied CNN in iEEG and sEEG databases. In 2020, Xu et al. [38] proposed not to process and transform EEG signals and used the original EEG signal as input. An end-to-end CNN was constructed to ES prediction. This method was applied in the AES Prediction Challenge and CHB-MIT as the datasets. This end-to-end CNN network is also applied in subsequent papers [12]. After 2020, researchers have attempted to solve the ES prediction problem using CNNs of different dimensions.

Using 2D-CNNs for processing biological signals offers advantages. Firstly, due to the time-frequency technology, the original EEG signals can be transformed into spectrograms which are 2D images. 2D-CNNs can efficiently handle these kinds of 2D images, allowing for the use of image-based DL models [39]. In order to better extract the features of the EEG signal after CWT, Hussein et al. [40] proposed an efficient data preprocessing method to convert time-series EEG signals into an image-like form (“scalogram”) using the CWT. Then, a convolution module called “semi-expansive convolution” is developed to better utilise the wavelet spectrogram and the geometric properties of non-square images. Some other studies [41,42] have similarly applied time-frequency transformation with 2D-CNN networks to improve the performance of ES prediction.

How to generate 2D images from 1D EEG signals and how to evaluate the quality of 2D images remain a key issue. In addition, brain rhythms suitable for ES analysis have not been fully explored. Shankar et al. [43] used 2D Recursive Picture (RP) images to generate EEG signals for specific brain rhythms and analysed them using 2D-CNN to address these issues. Ibrahim et al. [44] proposed the 2D-CNN with combining phase space reconstruction (PSR). The advantage of PSR is the direct projection from the time domain, thus preserving the main trends of the different signalling activities. Mu et al. [45] conducted by calculating the transfer entropy and phase transfer entropy of EEG signals and slicing them into features. These features are then classified pre-ictal or inter-ictal period using a deeply separable CNN with low parameters and computational.

Compared to higher dimensional CNNs, 1D-CNNs have fewer parameters, so the processing speed is faster in a single epoch [46]. They operate directly on multi-channel EEG time series [47,48] or time–frequency methods to extract the features such as STFT, DWT, SET/SVD [49–51] (see Fig. 4). These features distributed in different channels are used as inputs to the 1D CNN. Then, 1D convolutional kernel extracts features for multiple channels by stride. The features are classified after pooling and fully connected layers. Owing to their compactness, hardware-friendly 1D-CNN implementations with very low parameter counts have also been demonstrated [52].

Unlike 2D CNNs, 3D CNNs add correlations between images as features [53]. Qi et al. [54] presented a patient-specific approach based on EEG data to predict ES using spatial depth features of the 3D–2D Hybrid CNN (3D–2D HyCNN) model. This approach contributes to the acquisition of abundant and reliable depth features from multichannel EEG signals. The authors firstly reconstructed time-series EEG signals into 3D feature images. Then, correlation features between multichannel EEG signals are extracted using a 3D-2D HyCNN, and these features are automatically utilised by the network to improve ES prediction.

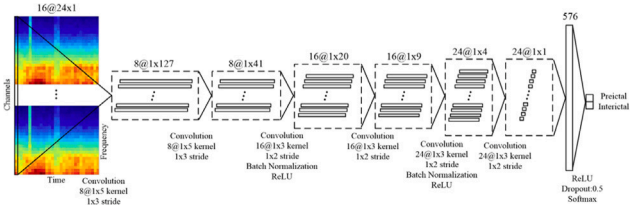

Fig. 4. The EEG signals were fed into a tiny CNN [49] after passing through the STFT for 20 s. The 1D-CNN consists of six convolutional layers, the first two of which are of size 1×5 and stride 1×3, and the last four are of size 1×3 and stride 12. The features are then passed through a fully connected layer and classified by the model into pre-ictal and inter-ictal periods.

3.1.2. Graph Neural Networks (GNNs)

Network-based methods for predicting ES have recently been recognised as a promising research direction [14]. These methods treat the brain as a dynamic network, in which ES activity is reflected in the constantly evolving connection patterns between different regions. Graph neural networks (GNNs) were firstly proposed in 2009 as a supervised learning framework [55] and are well-suited to this paradigm due to their ability to model dependencies between graph nodes through iterative information exchange. In the context of ES prediction, EEG channels can be viewed as nodes, while functional or structural connections between channels can be modelled as edges. This representation enables GNNs to capture spatial dependencies that traditional temporal models may overlook.

GNN-based approaches are motivated by the growing view of epilepsy as a network disorder rather than a purely focal phenomenon. iEEG and high-density sEEG studies consistently show that seizures emerge from and propagate through distributed cortico–subcortical networks, with abnormal patterns of functional connectivity long before clinical onset. By representing channels or cortical regions as nodes and their functional or structural connections as edges, GNNs encode this network structure directly into the model architecture. Message-passing operations in GNNs can be interpreted as modelling the spread of pathological activity along these connections, enabling the network to learn which subnetworks or hubs are most predictive of impending seizures. This inductive bias towards graph-structured interactions may partly explain why GNNs tend to be more robust than purely grid-based CNNs on heterogeneous, multi-centre datasets.

Building on the success of CNNs, graph convolutional networks (GCNs) further extend convolutional operations to irregular graph domains [56]. By aggregating node features and their neighbours’ features, GCNs can learn high-level representations of brain networks and have been applied to distinguish pre-ictal and inter-ictal states. However, several challenges remain. Firstly, the construction of graph structures (e.g., based on electrode placement, coherence, or correlation measurements) lacks standardisation, and different choices may lead to significantly different results. Secondly, most available datasets have limited spatial coverage and small patient sample sizes, which may affect the robustness and generalisation ability of GNN/GCN-based models. Third, patient-specific variability (e.g., electrode implantation location or signal quality differences) poses additional challenges for cross-patient evaluation.

Lian et al. [57] found that the effectiveness of GCNs is highly dependent on a priori maps describing the intrinsic relationships of EEG regions. However, due to the complex mechanisms of seizure evolution, the underlying relationships may vary from patient to patient before the seizure, making it nearly impossible to build appropriate a priori maps in general. To address this problem, they proposed the Joint Graph Structure and Representation Learning Network (JGRN) to automatically learn patient-specific graphs in a data-driven manner. The architecture of JGRN is shown in Fig. 5. Then, in order to explore the spatial and temporal dependencies in patient brains, Li et al. [58]

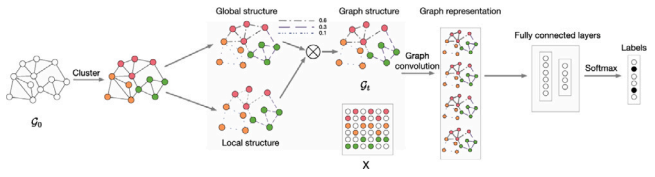

Fig. 5. The iteration process of JGRN Structure [57]. Firstly, the result after subgraph clustering is divided into two structures, global and local. After that, the obtained graph structure is subjected into graph convolution to get a graph representation. Finally a fully connected layer is added to classify the features.

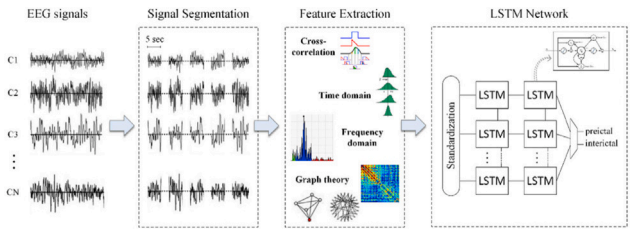

Fig. 6. The model first extracts four different types of features from the EEG signals of each segment of 5s: time domain, frequency domain, cross-correlation and graph theory. Later, after normalisation, it ends up in a two-layer LSTM network in which 128 memory units per layer are combined with a fully connected layer for classification [66].

proposed the spatio-temporal-spectral hierarchical GCN with a active pre-ictal interval learning scheme (STS-HGCN-AL). They first extracted the features of the different rhythms. After that, a residual GCN was used to capture the local and global dependencies within each rhythm for classification.

In 2022, Jia et al. [11] have attempted to apply a low-power implementable GCN on wearable devices. They reduced the computational effort of the model by extracting the necessary node and edge features. After that, Wang et al. [59] pointed out that most of the existing methods for ES prediction based on GCN focus only on the construction of static graphs. They first extracted the temporal, spatial and spectral features of the EEG signal. Then, they designed a point-wise Dynamic Multi-Graph Convolutional Network (dMGCN) to dynamically learn the graph structure to efficiently extract high-level features from multi-domain maps.

GCN is also widely used to solve the problems of data scarcity, diversity and privacy face by ES prediction systems. Dissanayake et al. [60] proposed a subject-independent seizure predictor using geometric deep learning (GDL) to address the problem of limited training data for the target patients. Saemaldahr et al. [61] proposed a Spiking Encoder (SE) integrated with GCNN (Spiking-GCNN). The network is able to utilise a large number of seizure patterns from a globally distributed patient population through federated learning (FL) while maintaining data privacy.

3.1.3. Recurrent neural networks (RNNs)

Recurrent Neural Networks (RNNs) [62] is a neural network suitable for processing sequence data, which can effectively capture the temporal information in the sequence. Therefore, as a dynamic time series signal, the feature of EEG signals can be analysed by RNN through its cyclically connected structure.

The clinical rationale for using recurrent and attention-based architectures lies in the inherently temporal nature of pre-ictal dynamics. Prior to an ES, gradual changes occur in synchronisation, spectral content, and network state, evolving over tens of seconds to several minutes. This enables RNNs to discern the precise moment of change by observing features of long-term information. When combined with temporal attention mechanisms, these models can further highlight specific time points within the pre-ictal window that are most informative for prediction. This approach facilitates the forecasting of ES.

Long Short-Term Memory (LSTM). One of the problem with original RNN is mainly short-term memory. LSTM can mitigate the gradient vanishing problem when dealing with sequential data through a gating mechanism [63]. Meanwhile, LSTM can also capture long-term dependencies through cell state and gating mechanism, which can retain the farther context information when processing sequential data [64]. LSTM has a sensitivity to time and can learn patterns and features in temporal data, which is well suited for temporal signal processing such as EEG. LSTM contains three different gates. The forget gate determines how much information stored in the previous cell state needs to be deleted. The input gate determines what new information is stored in the current cell state. The output gate determines what information is output from the cell state [65]. Therefore, the features of the EEG signal can be processed sequentially by these three gate structures according to different time steps.

In 2018, LSTM was first applied to address ES prediction using EEG signals. Tsiouris et al. [66] proposed an LSTM-based classifier to label sEEG segments as pre-ictal or inter-ictal. The model utilises a variety of features extracted prior to classification, including time and frequency domain features, cross-correlation between EEG channels, and graph theoretic features. The detailed structure is shown in Fig. 6

Ryu et al. [67] demonstrated a hybrid DL model with LSTM and Dense Convolutional Network (DenseNet). The model utilises the discrete wavelet transformed EEG signal as input and combines it with a DenseNet to extract features. LSTM applied to analysis the feature by every time step in order to classify the pre-ictal and inter-ictal states which detailed construction is shown in Fig. 8. Usman et al. also attempted to combine LSTM networks with other networks in an ensemble learning approach to predict ES [68]. The authors first generated pre-ictal segments using a GAN after pre-processing the signal through empirical pattern decomposition. A three-layer CNN was then used to automatically extract signal features and combine them with artificial features. Finally, model-independent meta-learning is utilised to combine the outputs of SVMs, CNNs and LSTMs. After that, Wu et al. [65] proposed a different idea by applying only the Gamma bands of the raw EEG signals as the input of an end-to-end LSTM model. The end-to-end LSTM network can directly utilise the input data without pre-processing. Such a network simplifies the process of designing and training DL models, while being able to improve the performance and generalisation of the models. Recently, Abdulwahhab et al. [69] proposed a parallel CNN–LSTM architecture. They fed the time–frequency map obtained from EEG via CWT/STFT into a CNN, while simultaneously processing the original time series using LSTM, and fused the output in a fully connected layer. This method achieved 99.75% accuracy on the Bonn dataset and approximately 97% on CHB-MIT.

Compared to LSTM, Bi-directional LSTM (Bi-LSTM) can compute both forward and backward at the same time and can better capture the information in the sequence. This also helps to extract better feature representations, which improves the robustness and generalisation of the model [70]. However, since Bi-LSTM requires simultaneous forward and backward computation, it requires a larger amount of computation and longer training time.

In 2018, Daoud et al. [71] used Bi-LSTM for processing the features through the pre-trained encoder. The detailed model structure is illustrated in Fig. 7. After that, Yan et al. [72] also applied Bi-LSTM to process the features from DenseNet based on the pre-processing method STFT. Then, Zhang et al. [73] presented that Bi-LSTM can learn multi-dimensional sample entropy (M-SampEn) features to achieve ES prediction. Also, some innovative LSTM networks have also been proposed such as multiplicative long short-term memory (MLSTM) [74].

Transformer. The attention mechanism was first proposed in Transformer [75]. As the core of Transformer, the logic of the attention

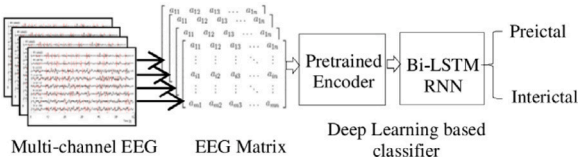

Fig. 7. The input to the model is a multichannel EEG signal. The EEG signal is formatted as a matrix where one dimension represents the EEG channel and the other dimension represents the time step. The extracted features are fed into the Bi-LSTM model for training by using a pre-trained encoder as a feature extractor. Finally the features are classified by the Bi-LSTM model [71].

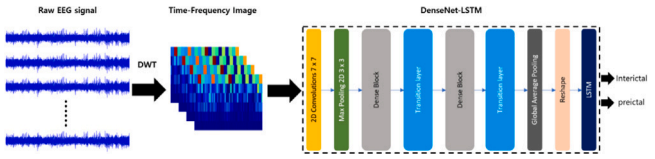

Fig. 8. The raw EEG signals are transformed to time–frequency images by DWT every 10 s. The multichannel time–frequency images are fed into DenseNet-LSTM [67]. The model has 2D convolutional layer of size is 7 × 7 with a stride size of 2. The model then processes the feature map using two dense blocks containing 6 sub-layers and 12 sub-layers (each sub-layer consists of a convolution of size 1 × 1 and a convolution of size 3 × 3). An average pooling layer containing convolution of size 1 × 1 and 2 × 2 is added after each dense layer with a stride size of 2. After that, the feature maps were transformed into 1D vectors after passing through the global average pooling layer and fed into the LSTM.

mechanism is from focusing on the whole sequence to focusing on the key points. Therefore, compared to other RNNs, Transformer saves more computational resources and quickly obtains the most effective information [76].

Some researchers combined the attention mechanism with the ResNet model. These algorithms first converted the raw EEG signal into a spectrogram by STFT by every channel. Combining the spectrogram of multiple channels results in an input format with three dimensions: frequency, time, and channel. Yang et al. [77] developed a generalised method for predicting ES in specific patients. They proposed a dual self-attention residual network (RDANet) that combines a spectral attention module with global features and a channel attention module that mines interdependence between channel mappings. After that, A model called TASM ResNet was proposed [78] based on a Temporal Attention Simulation Module and a pre-trained ResNet with iEEG signals which is shown in Fig. 9. In addition, a DL framework called Channel Attention Dual-Input CNN (CADCNN) was proposed by Sun et al. [79].

The Attention mechanism is also used in the gated recurrent unit (GRU) network. GRU has one less gate inside GRU compared to LSTM. So GRU has fewer parameters than LSTM, but it can also realise the function of LSTM [80]. To address the effective properties of EEG that may not be adequately evaluated, Wang et al. [81] proposed a synchronisation-based spatiotemporal graphical attention network (STGAT). The spatial and functional connectivity information between EEG channels is first extracted using phase-locked values (PLVs), thus modelling multichannel EEG signals as graph signals. The STGAT model then utilises GAT and GRU to dynamically learn the temporal correlation of EEG sequences and explore the spatial topological information of multiple channels. Ji et al. [82] proposed GAMRNN which combines a two-layer GRU model with a convolutional attention module. GAMRNN aims to capture intricate spatiotemporal features by highlighting informative feature channels and spatial pattern dynamics. They employed Lion optimisation algorithms to enhance the model's generalisation capability and prediction accuracy. The GAMRNN can achieve ES prediction within a lead time of 5 to 35 min.

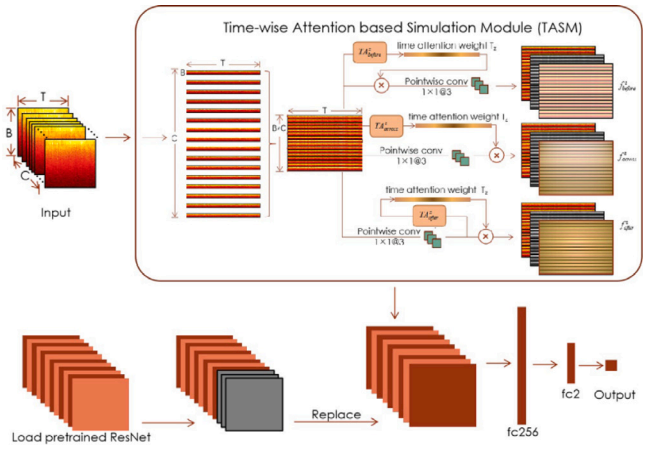

Fig. 9. Architecture of the TASM ResNet model. The input data shape of this model is (C,T,B) where C is the number of channels, T is the number of time samples and B is the number of bands. TASM is used to convert the input EEG data into image like data and extract the temporal features [78].

Not only can the Attention mechanism be applied to other networks, but the Transformer can be used to extract features from EEG signals. Transformer's outstanding ability to capture long-term dependencies and interactions is particularly attractive for time series modelling [83]. Yan et al. [84] proposed a transformer model applied in ES prediction. By transforming the EEG signals in the CHB-MIT Scalp EEG Database to spectrogram using the STFT, the features of the spectrogram are fused and classified using a three-transformer tower model. Then, the model combined the Transformer with graph attention network (GAT) called Gatformer was proposed by Wang et al. [85]. Temporal and spatial attention are combined to extract EEG information from the perspective of spatio-temporal interactions. Cseker et al. compared Transformers and CNNs for classifying mild cognitive impairment on resting-state EEG spectrograms, reporting significantly superior performance to CNNs [86]. Although not an early-stage scenario, the results suggest that Transformers based on spectrograms can better capture complex spatiotemporal dynamics, thus holding potential value for future ES prediction.

Vision Transformer (ViT) is a model proposed by the Google team in 2020 to apply Transformer to image classification. ViT divides the image into a series of patches and converts each patch into a vector representations as an input sequence. These vectors are then processed through multiple layers of Transformer encoders. The model can capture the contextual dependencies at different locations in the image [87]. Since EEG signals can be converted into the spectrograms by time-frequency conversion techniques, ViT has also been attempted to be applied in ES prediction tasks.

Zhang et al. first applied ViT in ES prediction [88]. In this paper, the raw EEG signals of each patient in CHB-MIT were filtered, and the pre-ictal and inter-ictal periods were extracted and labelled which were converted into 2D spectrograms by STFT. Then, the spectrogram was input into the ViT model to complete the feature extraction and ES prediction. Subsequently, Hussein et al. [89] proposed the Multi-channel Vision Transformer (MViT). They enhanced this approach by automatically learning the spatio-temporal and spectral features of multi-channel EEG signals. The methodology involves first converting the EEG into a scalogram using the CWT, then segmenting it into fixed-size, non-overlapping patches. These patches are subsequently concatenated across channels before being fed into the MViT for classification (see Fig. 10).

3.1.4. Spike neural networks (SNNs)

SNNs use models that fit biological neuronal mechanisms to perform computation. By constructing impulse transfer information between different neurons, SNNs are able to learn features similarly to Multilayer

Table 4
Summary of ES prediction performance using supervised learning methods. Accuracy represents the proportion of seizures correctly classified as either predicted or non-predicted. Sensitivity corresponds to the true positive rate, i.e., the proportion of seizures correctly predicted before their onset. Specificity denotes the true negative rate, i.e., the proportion of inter-ictal periods correctly identified as non-seizure. The false prediction rate (FPR) refers to the average number of false alarms per hour. The area under the ROC curve (AUC) quantifies the overall discriminative ability of the model. The seizure prediction horizon (SPH) is the time window between the prediction alarm and the actual seizure onset.

Ref	Year	Database	Method	Evaluation	Cases	Accuracy	Sensitivity	Specificity	FPR	AUC	SPH
[90]	2020	CHB-MIT	2D-CNN	LOSOCV	23	90.00%	92.00%	–	0.120/h	0.9200	30 min
[91]	2020	CHB-MIT	Transfer learning	–	11	92.60%	92.30%	97.00%	–	–	10 min
[38]	2020	CHB-MIT	2D-CNN	–	7	–	98.80%	–	0.074/h	0.9880	5 min
		AES			5		93.50%		0.063/h	0.9810	
[57]	2020	Freiburg	GNN	5-F CV	9	95.67%	83.68%	–	–	–	10 min
[60]	2021	CHB-MIT	GNN	10-F CV	6	95.38%	94.47%	94.16%	–	0.9878	60 min
[73]	2021	CHB-MIT	LSTM	–	–	80.09%	86.67%	–	0.260/h	–	5 min
[79]	2021	CHB-MIT	Attention	–	17	97.10%	95.60%	–	0.029/h	0.9170	3 min
[67]	2021	CHB-MIT	LSTM	K-F CV	24	93.28%	92.92%	93.65%	0.063/h	–	–
[77]	2021	CHB-MIT	Attention	5-F CV	13	92.07%	89.33%	93.02%	–	0.9126	–
[68]	2021	CHB-MIT	LSTM	K-F CV	24	–	96.28%	95.65%	–	–	33 min
		AES			5D, 2P		94.20%	95.80%			
[40]	2021	CHB-MIT	2D-CNN	LOPOCV	–	98.82%	98.90%	98.75%	0.060/h	–	–
		AES					88.45%			0.9280	
[92]	2021	CHB-MIT	SNN	–	7	95.10%	–	–	0.080/h	0.9140	–
[42]	2021	CHB-MIT	2D-CNN	LOSOCV	20	–	85.00%	–	0.14/h	–	30 min
		Freiburg			16		91.00%		0.06/h		
[47]	2022	Freiburg	1D-CNN	K-F CV	16	95.13%	98.65%	–	0.080/h	–	5 min
[88]	2022	AES	1D-CNN	5-F CV	5D, 2P	94.44%	–	–	0.011/h	0.9790	–
[44]	2022	CHB-MIT	2D-CNN	–	–	99.89%	99.89%	99.95%	–	–	–
[12]	2022	CHB-MIT	end-to-end CNN	LOSOCV 5-F CV	16	93.30%	–	–	0.007/h	–	30 min
[74]	2022	CHB-MIT	LSTM	–	–	–	89.47%	–	0.340/h		
		SWEC-ETHZ					95.56%		0.270/h		
[88]	2022	CHB-MIT	Transformer	–	14	81.20%	75.59%	81.78%	–	0.8570	–
[47]	2022	CHB-MIT	Attention	–	24	98.74%	98.87%	99.21%	–	–	–
[89]	2022	CHB-MIT	Transformer	–	–	99.80%	99.80%	99.70%	0.004/h	–	–
[78]	2022	AES	Attention	5-F CV	7	80.50%	76.10%	81.00%	–	0.8980	–
[84]	2022	CHB-MIT	Transformer	–	23	–	96.01%	96.23%	0.047/h	–	3 min
[45]	2022	CHB-MIT	2D-CNN	–	–	98.87%	98.45%	99.26%	–	–	15 min
[52]	2022	CHB-MIT	1D-CNN	5F-CV	5	99.01%	99.24%	98.68%	0.470/h	–	5 min
		SWEC-ETHZ			5	97.54%	98.22%	97.02%	0.990/h		5 min
[11]	2022	CHB-MIT	GNN	LOSOCV	18	–	96.51%	–	–	0.9200	60 min
[58]	2022	CHB-MIT	GNN	LOSOCV	19	–	95.50%	–	–	0.9380	15–90 min
[93]	2022	CHB-MIT	Transfer learning	–	20	93.80%	91.20%	93.80%	–	–	5 min
[41]	2022	CHB-MIT	2D-CNN	5-F CV	23	96.99%	96.48%	97.46%	–	–	–
[48]	2023	CHB-MIT	1D-CNN	–	23	96.55%	96.47%	–	–	–	–
[51]	2023	CHB-MIT	1D-CNN	–	23	99.71%	99.75%	99.56%	–	–	–
		Bonn			8	99.97%	100.00%	99.95%			
[50]	2023	CHB-MIT	1D-CNN	–	–	98.09%	99.04%	–	–	–	–
[65]	2023	CHB-MIT	LSTM	10-F CV	13	–	91.76%	–	0.290/h	–	60 min
[85]	2023	CHB-MIT	Transformer	5-F CV	24	98.74%	98.87%	99.21%	–	–	–
[82]	2023	CHB-MIT	Attention	LOSOCV	–	91.73%	88.09%	92.09%	0.053/h	–	5–35 min
[72]	2023	CHB-MIT	LSTM	LOSOCV	24	92.45%	92.66%	–	0.066/h	0.9360	–
[94]	2023	CHB-MIT	LSTM	LOSOCV	13	–	82.84%	85.97%	–	0.9080	–
[59]	2023	CHB-MIT	GNN	LOSOCV	23	–	97.81%	–	0.059/h	–	–
[61]	2023	CHB-MIT	GNN	LOSOCV k-F CV	24	96.28%	96.33%	96.14%	0.090/h	0.8960	–
[95]	2023	EPILEPSIA	Transfer learning	–	10	98.47%	–	–	0.031/h	–	40 min
[54]	2023	CHB-MIT	3D-CNN	LOSOCV	13	98.43%	98.58%	96.86%	–	–	–

Table 5
Quantitative synthesis of DL architectures for ES prediction. For each architecture, we report the unweighted mean ± SD of Sensitivity, FPR (/h), and AUC across the reviewed studies. "Best paper (by AUC)" lists the study with the highest reported AUC within that family. Parentheses indicate the number of studies contributing to a given metric.

DL architecture	Papers (N)	Avg. Sensitivity (%) (Mean ± SD)	Avg. FPR (/h) (Mean ± SD)	Avg. AUC (Mean ± SD)	Best paper
CNNs (1D/2D/3D)	19	96.25 ± 4.41	0.167 ± 0.274	0.9670 ± 0.0316	Xu et al. [96]
RNN/LSTMs	9	91.37 ± 4.36	0.215 ± 0.120	0.9220 ± 0.0198 ($N = 2$)	Yan et al. [72]
GNNs	6	94.05 ± 5.20	0.075 ± 0.022 ($N = 2$)	0.9355 ± 0.0389	Dissanayake et al. [60]
Transformer/Attention	9	90.92 ± 9.48	0.033 ± 0.022 ($N = 4$)	0.8962 ± 0.0273 ($N = 4$)	Sun et al. [79]
Overall Aggregate	43	93.75 ± 6.19	0.149 ± 0.211	0.9313 ± 0.0397	Xu et al. [96]

Table 6
Quantitative synthesis by database for ES prediction. Values are unweighted mean ± SD across included studies. "Best model (by AUC)" denotes the highest reported AUC on that dataset. If the AUC is not reported, then the highest sensitivity should be used as the evaluation metric. "N/A" = not reported.

Database	Studies (N)	Avg. Sensitivity (%) (Mean ± SD)	Avg. FPR (/h) (Mean ± SD)	Avg. AUC (Mean ± SD)	Best model
CHB-MIT	36	94.16 ± 5.61	0.125 ± 0.129	0.9255 ± 0.0377	Xu et al. [96]
AES	5	88.06 ± 8.38	0.047 ± 0.029	0.9465 ± 0.0406	Xu et al. [96]
Freiburg	3	91.11 ± 7.49	0.070 ± 0.014	N/A	Wang et al. [47]

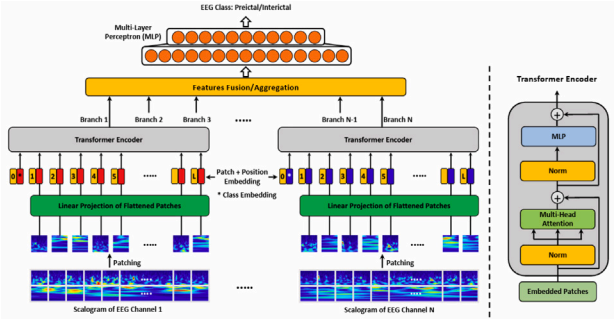

Fig. 10. Framework of MViT for multi-channel EEG feature learning [89] The EEG signal was first transformed to Scalograms by CWT. Then, the different patches in the scalogram are combined with the corresponding position codes and fed into the encoder of the transformer. Different channels are labelled as different branches. After passing through the features aggregation layer, these features are classified by Multi-Layer Perception.

Perceptron (MLP) [97]. Therefore, SNN is called the third-generation neural network. Since the concept of SNN is closely related to neurons in the brain, it is applied to the ES prediction problem. The EEG signal is first passed through a pulse encoder that converts the continuous EEG signal into a time-dependent pulse sequence. SNNs make ES prediction by learning different [92]. The computational effort is greatly reduced due to the encoding of raw signals into pulse features. So SNN is an energy efficiency and hardware friendly network.

3.1.5. Transfer learning

Due to the low data amount of publicly available EEG databases for ES prediction, the researchers applied deep transfer learning algorithms to solve the issue. The advantage of deep transfer learning algorithms is that only a small amount of data is fed into the DL network, and the input data is classified using the network parameters that have already been learned from the large database [98].

Residual Neural Network (ResNet). ResNet was proposed by He et al. [99] won the champion of ImageNet Large Scale Visual Recognition Challenge in 2015. The main contribution of ResNet is the discovery of "Degradation" and the invention of "Shortcut connection" to address the degradation phenomenon, which greatly eliminates the problem of training neural networks with too much depth [100]. For the first time, the depth of neural networks exceeded 100 layers, and the largest neural networks even exceeded 1000 layers.

Mohammad et al. [93] applied ResNet and transfer learning (SPERTL) for a patient-specific ES prediction model using EEG data. The model used is trained on a dataset of >2.5 million patients [101]. The EEG signals are fed into the pre-trained ResNet. The convolutional, residual blocks and fully connected layers in the model are fine-tuned through training. On the CHB-MIT Scalp EEG Database, 20 patients with a SPH of 5 min were trained in this model.

Inception. The Inception network is an milestone in the history of CNN classifiers. Before Inception, most popular CNNs simply stacked more and more convolutional layers to make them deeper and deeper in the hope of getting better performance. The feature of GoogLeNet [102] is the use of the Inception module, which aims at designing a network with an excellent local topology, i.e., performing multiple convolution operations or pooling operations on the input image in parallel and stitching all the outputs into a very deep feature map [103].

Gao et al. [91] proposed method of Epileptic EEG Signal Classification. This method firstly converts epileptic EEG signals into power spectral density energy maps (PSDED). Then, the PSDED are fed into the pre-trained Inception-v3 [100], ResNet152 and Inception-ResNet-v2 [99]. EESC learns PSDED features by fine-tuning the entire model. Sarvi Zargar et al. [95] applied three different models for transfer learning, namely Xception [104], EfficientNet-B0 [105] and MobileNet-V2 [106]. These models learn from features by freezing the convolutional layer weights and adding new classification layers.

Yang et al. [107] demonstrates a new application of Inception-V3 in ES prediction, using a transformed network-based data augmentation approach together with multi-weighted fuzzy-granular recurrence graphs. This method generates synthetic pre-ictal signals via random walks and converts EEG data into non-linear recurrent images, enabling robust modelling of noise. On the CHB-MIT and AES datasets, the method achieved average sensitivity, specificity, and accuracy exceeding 96%. These results indicate that Inception-V3 is not only effective in transfer learning tasks but can also be combined with novel feature representations to explore more complex ES prediction scenarios. However, the method relies on synthetic data and complex preprocessing, and its clinical feasibility requires further validation.

3.2. Unsupervised learning

Unsupervised learning algorithms have different learning methods in ES prediction. Some unsupervised learning algorithms augment the original dataset by generating signals similar to the original dataset by studying the distribution of the data, thus improving the prediction performance of the dataset. The other unsupervised learning algorithms learn the original features of the data before classifying the signals. They do not require labels when learning the features of the data.

A statistical overview of the unsupervised learning studies (Table 7) reveals that this research direction is still in its early stages but demonstrates significant potential. These unsupervised learning approaches have been primarily applied to the CHB-MIT and AES datasets. Reported sensitivities are often above 90%, and in some cases approach 97%, indicating strong predictive ability.

However, accuracies are less consistently reported, and AUC values range widely from 0.83 to 0.98, reflecting less stable performance compared with supervised approaches. FPR also show large variation, with some methods achieving very low rates (0.009/h) while others remain substantially higher (0.7/h). Furthermore, the limited number of studies and datasets employed makes it difficult to draw firm statistical conclusions.

These results suggest that unsupervised learning holds significant promise, particularly in reducing dependence on labelled data. Nevertheless, further systematic evaluations on larger and more diverse datasets are required to validate its clinical potential.

3.2.1. Generative Adversarial Networks (GANs)

GAN is one of the unsupervised DL algorithms proposed by Goodfellow et al. [108] in 2014. It consists of two parts: generator and discriminator. The distribution of real datasets can be captured by the unsupervised learning procedure of the generator. Then, the generator can forge the data according to the learned distribution by inputting

the random noise. The discriminator corrects the generated data based on the difference between the generated data and the real data.

Since 2018, scholars have consecutively proposed the use of GANs in EEG signals. Hartmann et al. [109] described a framework called EEG-GAN to generate the EEG signals. The construction of the proposed GAN model used the DCGAN [110]. They compared the performance with several upsampling methods within the CNN-based generator in different metrics such as Inception Score (IS), Frechet Inception Distance (FID) and sliced Wasserstein Distance (WD). These evaluation indicators demonstrates were used as an assessment of the differences between the generated model and the real data distribution.

In 2019, Truong et al. [111] first applied GAN to predict ES from EEG signals. They proposed a method which was utilised by three public EEG datasets including CHB-MIT, Freiburg, and EPILEPSIAE respectively. The data was divided into two stages, inter-ictal and pre-ictal. Then, they defined SOP as 30 min and SPH as 5 min. To analyse the data, Spectrograms were generated using the STFT on the original datasets. The GAN was trained through unsupervised learning on EEG signal. The trained discriminator of the GAN is then used as a feature extractor. The features generated by the feature extractor are classified by two fully connected layers on the labelled EEG signals. The results of the studies have shown that the GAN-based approach outperforms previous CNN [112] in predicting ES. The detailed structure of GAN is shown in Fig. 11.

After that, Pascual et al. [113] proposed EpilepsyGAN as a synthetic brain activity generator for epileptic patients to preserve privacy. This is the first time that GAN has been used as a synthetic EEG signal for epileptic patients. Then, Rasheed et al. [114] used the CHB-MIT and Epilepsy-ecosystem to synthesise the EEG signals for ES prediction. They used STFT to process EEG signals in 30-second segments. These spectrograms were fed into the DCGAN for learning. The trained generator produced spectrograms that were used for data augmentation. They used a CNN as a classifier for ES prediction.

Xu et al. [96] proposed a pre-ictal signal synthesis algorithm based on GANs to generate muti-channel EEG pre-ictal samples. They fed the temporal signals into four GANs with different structures. The performance of the GAN was evaluated by comparing the difference between the samples generated by the trained generator and the orginal samples using FDRMSE (Frequency domain root mean square error), FID and WD. They found that Deep Convolutional Wasserstein Generative Adversarial Network (DCWGAN) outperformed the rest of the models. Subsequently, they expanded the DCWGAN-generated samples to the original dataset. Compared to the original dataset, the expanded dataset improved both the accuracy of ES prediction and the area under the receiver operating characteristic curve (AUC-ROC). Yu et al. [115] synthesised EEG signals by Conditional GAN (CGAN) for data augmentation. Compared to other GAN models, they added refiner to CGAN to solve the problem of data scarcity and imbalance. The refiner further reduced the difference between the generated data and the real data by learning the data generated by CGAN. After that, Yang et al. [94] first utilised unsupervised learning to train the discriminators in the Wasserstein GAN as feature extractor. Then, they applied Bi-LSTM to predict the ES by incorporating the features of EEG signals.

3.2.2. AutoEncoder

AutoEncoder as an unsupervised learning model is mainly trained based on a back propagation algorithm with optimisation methods [121]. The model can be divided into two main parts: an encoder and a decoder. The role of the encoder is to encode the high-dimensional input vector into a low-dimensional hidden variable thus forcing the neural network to learn the most informative features. The role of the decoder is to reduce the hidden variables of the hidden layer to their initial dimensions, ideally in a state where the output of the decoder perfectly or approximately recovers the original input. The advantage of the AutoEncoder is that it is highly generalisable and does not require data annotation. Abdelhameed et al. [122] proposed an

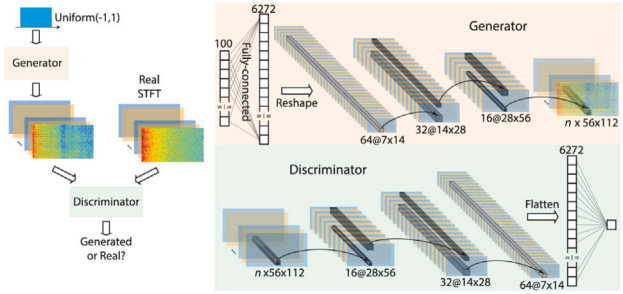

Fig. 11. The GAN structure [111]. The generator takes 100 data points randomly drawn from a uniform distribution from −1 to 1 as input. The input is fully connected to a hidden layer with an output size of 6272, which is then reshaped to 64 × 7 × 14. The number of filters in the three deconvolution layers is 32, 16, and n, respectively. The discriminator consists of three convolutional layers with a filter size of 5 × 5 and a stride of 2 × 2. The number of filters in the three convolutional layers is 16, 32, and 64, respectively.

ES prediction system. The system combines a 2D deep convolutional AutoEncoder and Bi-LSTM. First, they used the AutoEncoder as a feature extractor. Then the extracted features are fed into Bi-LSTM. Bi-LSTM performs ES prediction based on temporal information. After that, a network was proposed by Gözütok et al. [116]. The authors used convolutional AutoEncoder for feature extraction. LSTM was then used for ES prediction.

As one of the AutoEncoder, Variational AutoEncoder (VAE) is also a generative network structure based on Variational Bayes (VB) inference proposed by Kingma et al. [123]. It introduces random variables and probabilistic models on top of the AutoEncoder, enabling the model to learn the probability distribution of the data. Therefore, the model has great application in data generation. He et al. [119] proposed a data augmentation method based on a stochastic transformation strategy to solve the problem of insufficient datasets in EEG signals without adding additional noise. They proposed an improved unsupervised feature learning method, Residual Convolutional VAE with Randomised Translation Strategy (RTS-RCVAE). Residual learning is embedded into the VAE model, which improves the model's convergence ability in the unsupervised learning phase and reduces the loss of useful information. The proposed model is validated by training and simulation using the CHB-MIT.

3.2.3. Self-supervised learning

The supervised information of self-supervised learning is not manually labelled, but automatically constructed by the algorithm in large-scale unsupervised data. Therefore, self-supervised learning is also a kind of unsupervised learning. Since Self-supervised Learning uses unlabelled datasets in the pre-training phase, this greatly reduces the time doctors have to spend labelling EEG data [124]. This has led to the application of self-supervised learning in the field of ES prediction. In 2022, Yang et al. [118] proposed a system for continuously improving ES prediction through self-supervised learning using auto-correlation of time-series EEG signals. The system reduces the burden of manual labelling by generating weak labels and training them as targets. The method allows the development of personalised prediction models while eliminating the need to label long sequences of physiological signals.

Contrastive learning (CL) is a part of self-supervised learning. It makes similar samples close together in space and then makes different samples as far away from each other as possible [125]. Zhao et al. [117] proposed a method called adder net work and supervised CL (AddNet-SCL) that reduces computational cost by using addition instead of multiplication in the network (see Fig. 12). In addition, the method uses CL to efficiently utilise label information to cluster data points of the

Table 7
Summary of ES prediction performance using unsupervised learning methods. Accuracy represents the proportion of seizures correctly classified as either predicted or non-predicted. Sensitivity corresponds to the true positive rate, i.e., the proportion of seizures correctly predicted before their onset. Specificity denotes the true negative rate, i.e., the proportion of inter-ictal periods correctly identified as non-seizure. The false prediction rate (FPR) refers to the average number of false alarms per hour. The area under the ROC curve (AUC) quantifies the overall discriminative ability of the model. The seizure prediction horizon (SPH) is the time window between the prediction alarm and the actual seizure onset.

Paper	Year	Database	Method	Evaluation	Cases	Accuracy	Sensitivity	FPR	AUC	SPH
	2021	CHB-MIT	GAN	10-F CV	13	–	96.00%	0.050/h	–	5 min
[114]		Epilepsy-ecosystem			3		92.87%	0.150/h		
[116]	2021	CHB-MIT	AutoEncoder	LOSOCV	5	92.20%	–	–	–	–
[117]	2022	CHB-MIT	Self-supervised	LOSOCV	19	–	93.00%	0.094/h	0.929	1 min
		AES			5D		89.10%	0.120/h	0.831	5 min
[118]	2022	TUSZ	Self-supervised	–	-	–	63.16%	6.6/24h	–	–
		EPILEPSIAE			30					
[96]	2022	CHB-MIT	GAN	LOSOCV	7	78.00%	–	0.704/h	–	5 min
[119]	2023	CHB-MIT	VAE	–	12	98.43%	97.32%	0.009/h	0.984	–
[120]	2023	CHB-MIT	Self-supervised	LOSOCV	19	–	96.70%	0.072/h	0.918	–
[115]	2023	CHB-MIT	GAN	–	22	–	–	–	0.610	5 min

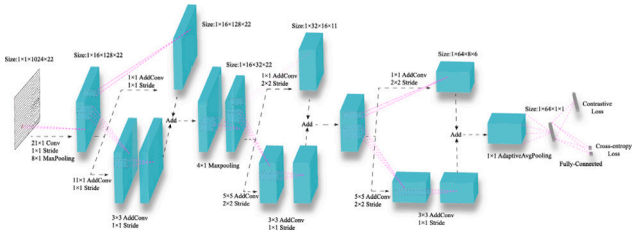

Fig. 12. The Self-supervised Learning Structure [117]. Firstly, the model extracts features in the temporal dimension of the data by one-dimensional convolution. Then, the input features are learnt using additive convolution and combined with residual concatenation. A 64-dimensional features vector is obtained after processing by CNN. Finally, the obtained vector is updated and classified by contrast loss and cross entropy loss.

same category together and separate data points of different categories. After that, Guo et al. [120] presented a CL method for ES prediction using Spatio-Temporal-Spectral Network (STS-Net).

The method understands the inherent epileptic EEG patterns of different patients through comparative learning and extracts multi-scale temporal and spectral representations of different rhythms from the raw EEG signals. Then, these features are fed into the triple attention layer (TAL) for processing. Finally, ES prediction was performed by spatial dynamic Graph Convolutional Network (sdGCM)

It is worth noting that self-supervised learning has also achieved significant success in other areas of medical image analysis. Addulaa et al. [126] employed a hierarchical self-learning framework combined with transfer learning on a convolutional neural network-based model on the classification of breast cancer histopathology images, achieving an accuracy rate of 99.1% after iterative label correction. This evidence suggests that self-supervised techniques cannot only reduce the burden of manual labelling but also significantly improve the robustness and generalisation capabilities of models. Therefore, applying such methods to ES prediction may be a promising future direction.

3.3. Comparison with traditional machine learning approaches

Compared to deep learning, traditional machine learning methods were applied earlier in epileptic seizure prediction systems. These primarily employed a workflow combining handcrafted features with shallow classifiers. This included extracting temporal features (line length, variance), frequency-band power and ratios, time-frequency domain representations, and non-linear dynamical metrics or connectivity features such as Lyapunov exponents. Subsequently, seizure detection was performed by training support vector machines (SVM), random forests (RF), k-nearest neighbours, logistic regression, or shallow multilayer perceptrons [127–129]. Thereafter, machine learning approaches advanced further in feature engineering and optimisation. For instance, Wang et al. [130] combined multiple time-frequency domain transformations with principal component analysis in an RF classifier for seizure detection. Truong et al. [131] proposed integrating automatic channel selection with RF classification for efficient iEEG detection. Concurrently, Burrello et al. introduced hyperdimensional computation of Local Binary Patterns (LBP) and one-shot learning [132,133], the energy-efficient Laelaps algorithm for long-term iEEG with near 'zero false alarms' [23], and related hyperdimensional classifier ensemble methods [134]. These algorithms offer sound approaches for hardware integration. It is worth emphasising that the pioneering work by Mirowski [30] and Shoeb [6], which distinguished pre-ictal from inter-ictal states through engineered EEG features and supervised learning, established methodological precedents for subsequent deep learning workflows.

Whilst deep learning models often achieve higher metrics when compared against machine learning on identical datasets, these advantages are modest—typically yielding accuracy or sensitivity gains of approximately 5–10 percentage points. This is particularly evident on large, homogeneous datasets such as the CHB-MIT corpus. Conversely, on smaller datasets, machine learning remains competitive when combined with carefully designed feature extraction approaches. For instance, within the SWEC-ETHZ iEEG short-term dataset, hyperdimensional algorithms demonstrate comparatively favourable performance in interpreting iEEG signals [23,133]. These observations suggest that deep learning models generally exhibit superior advantages and performance when sufficient data is available. Nevertheless, traditional machine learning models may prove more suitable under conditions of sparse data and resource constraints.

Concurrently, considering clinical requirements, we wish to discuss the computational burden differences between these two approaches. Deep learning networks, particularly 2D/3D CNNs or Transformer-based architectures, frequently require millions of parameters. In contrast, traditional classifiers operating on low-dimensional feature vectors can be implemented on low-power microcontrollers or custom neuromorphic hardware, rendering them more suitable for fully implantable or long-term wearable devices. Within closed-loop ES prediction systems, latency and battery life are critical considerations.

Differences also exist in interpretability and validation approaches. Traditional workflows centre on feature constructs with explicit neurophysiological significance, while simple classifiers like linear SVMs or logistic regression provide directly inspectable weights. This facilitates clinical interpretation, hypothesis generation, and regulatory scrutiny. In contrast, deep models learn more distributional representations that often resist direct mapping to established biomarkers, though post-hoc explanation techniques partially bridge this gap. Overall, both deep learning and traditional machine learning possess distinct advantages for seizure prediction tasks. Future work may explore architectures combining deep feature extractors with lightweight classifiers, as discussed in Sections 4.2.3 and 4.2.5 of this paper.

4. Current challenges and future directions

This section mainly focus on the current challenges and future directions in ES prediction which are divided into two parts. One is the clinical medicine and the other is the DL algorithms, corresponding to what clinicians and researchers of DL algorithms can contribute to the future development of ES prediction.

4.1. Clinical medicine

4.1.1. Next clinical trial

Since the first ES prediction clinical trial was initiated in 2013, a decade has passed. Although the trial was terminated prematurely due to technical and economic constraints, the data collected remains invaluable to the scientific community, as it includes long-term continuous iEEG recordings. Over the past decade, numerous high-performance deep learning algorithms have been developed using publicly available datasets, making it both feasible and timely to initiate new clinical trials. However, several challenges must be addressed before the next trial can be successful. These challenges include the lack of large-scale, demographically diverse data required to ensure model generalisation, the need for strict pseudo-prospective evaluation protocols, and the necessity of integrating multimodal signals into clinically deployable systems.

As suggested by previous research [14], developing a feasible closed-loop ES prediction system represents a critical next step. Such systems require close collaboration between artificial intelligence researchers, clinicians, and biomedical engineers to ensure technical robustness and clinical relevance. Conducting new clinical trials under these principles will mark the next milestone application in the field of ES prediction.

4.1.2. Ethical considerations

In future clinical trial designs, ethical considerations must be a core component. This includes ensuring patient privacy and data security, obtaining informed consent from participants, and carefully managing the clinical risks posed by both false positives (unnecessary interventions) and false negatives (missed seizures).

From an ethical and patient-centred perspective, ES prediction systems must not only achieve high accuracy but also behave consistently across different patient groups and provide well-calibrated risk estimates. Overconfident false alarms or systematically poorer performance in under-represented populations can undermine trust and exacerbate existing inequities in clinical care. These issues of algorithmic bias, predictive uncertainty and model calibration are examined in more detail in Section 4.2.2.

Another critical aspect is transparency and interpretability. Clinicians and patients should be able to at least partially understand how prediction outcomes are generated, how uncertainty is quantified, and what the main limitations of the system are, thereby building realistic expectations and trust. Furthermore, sharing clinical EEG data across institutions requires a balance between open scientific collaboration and strict patient privacy protection, necessitating robust data governance, de-identification and anonymisation protocols.

Finally, widespread clinical adoption of ES prediction systems will likely require regulatory approval. Therefore, future trials must not only demonstrate technical efficacy but also provide strong evidence of clinical safety, ethical compliance, fairness across patient subgroups, and patient-centred value. Integrating these ethical safeguards will be crucial for the safe, equitable and responsible application of ES prediction systems in clinical practice.

4.1.3. Challenges on data scarcity and patient heterogeneity

One of the biggest challenges facing ES prediction is the lack of large-scale, diverse datasets. Limited data availability hinders the generalisation ability of DL models, often leading to over-fitting in small patient populations and reduced robustness when applied to unseen data. Additionally, the high variability among patients—including differences in epilepsy type, electrode placement, brain anatomy, medication, and comorbidities further complicates model development. Models trained on small, homogeneous patient populations may perform well within that population but may fail to maintain accuracy when tested on data from different patients or clinical settings. Such variability undermines the effectiveness and reliability of predictive models, highlighting the urgent need for larger, more representative datasets with diverse demographic characteristics.

Although researchers have emphasised the importance of prospective or pseudo-prospective evaluation of algorithms [14], so far the only large-scale dataset that allows for such evaluation is the one collected in the first clinical trial in 2013 [135]. The partially available version of this dataset (Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge) provides continuous iEEG recordings, but the relatively small number of patients makes it difficult to reliably assess cross-patient generalisation. As a result, this dataset has not been widely adopted for the evaluation of ES algorithms. To address this limitation, future research should prioritise the collection of larger, more representative datasets. Several directions can be highlighted:

- **Long-term clinical monitoring data:** Continuous recordings from patients in real-world clinical environments are essential for capturing the variability of seizure dynamics over time.
- **Cross-species biological data:** Data from animal models of epilepsy can provide valuable insights into similarities and differences with human seizure mechanisms, thereby improving translational validity.
- **Multimodal datasets:** Combining EEG with other physiological signals such as ECG, EMG, and NIRS can enable more accurate and reliable prediction through complementary information sources.
- **Multi-patient demographic diversity:** Datasets should include comprehensive demographic and clinical metadata (e.g., age, gender, seizure type, disease duration, comorbidities) to improve model generalisability, fairness, and personalisation.
- **More detailed recorded information:** In addition to pre-ictal, ictal, and inter-ictal periods, future datasets should record information such as electrode channel locations in intracranial EEG. Such details are crucial for analysing the spatiotemporal nature of seizures and can inspire the design of new network architectures.

Expanding data collection along these lines will not only improve the robustness of DL-based ES prediction models but also enhance their clinical applicability and ethical soundness.

4.2. DL algorithm

4.2.1. Reproducibility and data leakage risks

Although numerous ES prediction algorithms report highly promising performance, our review indicates that the field remains subject to significant data leakage risks. This severely compromises the feasibility and reproducibility of research findings. Specifically, when pre-ictal and inter-ictal segments from the same patient appear concurrently in both training and testing sets, patient independence is compromised. This may cause the model to memorise patient-specific patterns rather than genuinely learning seizure-related features.

Analysis of previous literature in our review reveals only a few studies explicitly implementing patient-independent segmentation. Many rely on conventional k-fold cross-validation without clarifying whether patient identifiers were used for separation. Some studies entirely ignore segmentation strategy details. For ES prediction datasets, multiple

usable segments are typically available per patient. Consequently, k-fold cross-validation based on EEG signals almost inevitably mixes data from the same patient across training and testing sets. While this approach often yields high accuracy, it fails to reflect the clinically relevant scenario of predicting ES in previously unseen patients. Indeed, studies evaluating models under strictly patient-based protocols frequently observe diminished performance. This suggests that some reported improvements may stem from data leakage artefacts rather than genuine gains in generalisation capability.

To obtain more clinically meaningful estimates of generalisation performance, models should be evaluated under patient-independent validation schemes. Two approaches are particularly suitable: the first is Leave-One-Patient-Out Cross-Validation (LOPOCV), wherein all data from a single patient are reserved as the test set in each iteration, while data from all other patients are used for training and validation. This strategy ensures patient independence and is well-suited for datasets with a small number of patients and lengthy records per patient. The second approach is Leave-One-Seizure-Out Cross-Validation (LOSOCV). This evaluates each patient individually, using one ES for testing while employing the remaining ES for training and validation. This method estimates model performance for individual patients but suffers from inadequate estimation of the model's generalisation capability across all patients.

Beyond the selection of evaluation schemes, we contend that ES prediction research would benefit from a more standardised benchmarking framework. Current studies have already endeavoured to construct comprehensive benchmarks for ES detection [136]. For predictive algorithms, we recommend considering the following points. Firstly, temporal independence: for any given patient, distinct EEG signal recordings within the training and test sets should not overlap. Moreover, future data must not be employed to predict past events. Secondly, concerning preprocessing, all normalisation parameters and feature selection steps should be fitted solely on the training set, then directly applied to validation and test set data. For small datasets, LOPOCV or LOSOCV is recommended, reporting the mean performance and standard deviation across each fold. For larger patient datasets, patient partitioning is advised, with repeated experiments conducted across multiple random divisions. This evaluation methodology not only enables other researchers to replicate results under explicitly defined splits and preprocessing schemes, thereby enhancing reproducibility. Simultaneously, as models are assessed under comparable and leak-free conditions rather than specialised or opaque data divisions, it improves the fairness of method comparisons. Considering clinical applications, models evaluated through this approach also closely resemble real-world deployment scenarios.

4.2.2. Algorithmic bias, uncertainty and model calibration

Beyond issues with evaluation methods, deep learning-based ES prediction systems also raise concerns regarding algorithmic bias, uncertainty, and model calibration. Most studies reviewed in this paper primarily report threshold-related metrics such as sensitivity, specificity, and accuracy. However, scant attention is paid to whether predicted probabilities are well-calibrated or whether performance remains consistent across different patient cohorts and recording locations. In ES prediction, excessively high false positives may induce unnecessary anxiety, inappropriate behavioural restrictions, and alert fatigue. Conversely, excessively low false negatives could foster a false sense of security and delay appropriate interventions. These risks are particularly pronounced for wearable devices and closed-loop systems providing frequent real-time predictions.

Furthermore, distinctions in model testing across datasets must be considered. Most ES prediction studies suffer from training and evaluation datasets exhibiting demographic imbalances, hardware heterogeneity, and centre-specific acquisition protocols, posing significant risks of algorithmic bias. A specific example is the widely used CHB-MIT dataset, which predominantly contains paediatric data. Many approaches achieve excellent overall performance on the CHB-MIT dataset. However, these results do not guarantee good predictive performance in adult populations or patients with different electrode configurations. Also, certain algorithms have demonstrated favourable performance on datasets such as the SWEC-ETHZ iEEG and AES. However, due to differences between iEEG and sEEG in signal amplitude, acquisition methods, and recording duration, their efficacy cannot be guaranteed across other datasets. To address this, we recommend that ES prediction algorithms should be tested across multiple datasets where feasible. Furthermore, performance reporting should account for variations across patient cohorts, electrode placements, and epileptogenic regions.

Beyond potential issues during algorithm training, the datasets themselves present ethical and fairness concerns. Due to considerations of data privacy and de-identification policies by patients and hospitals during data collection, public datasets tend to be biased towards specific demographic characteristics and EEG recording configurations. Therefore, we recommend that public datasets provide more detailed information. This may include demographic details (age distribution, gender, seizure types), acquisition specifics (electrode configurations, sampling rates, hardware), and inclusion/exclusion criteria. This ensures that trained models are more equitable.

4.2.3. Lightweight model

In recent years, EEG collection systems have evolved rapidly in terms of removable devices, especially for scalp EEG signals. Commercially available mobile devices for EEG collection already exist [16]. So how to integrate AI algorithms on these removable or wearable devices is the current problem. For cheaper chips, today's models tend to be more computationally intensive and algorithmically complex. It becomes especially important to develop a lightweight network making it possible for current chips to quickly make predictions about seizures.

One of the future research direction is SNNs. Since the principle of SNN is by converting the input signal into pulses [92,137], such an approach can greatly reduce the computation of the network. Meanwhile, in terms of hardware implementation, SNN only requires the use of adders to build the hardware. These advantages make SNN be used frequently in the future in daily life. However, again due to the presence of only impulse computation among the network, SNN is not able to back propagate the gradient like traditional neural networks. This also requires more thinking from the researchers to solve this problem.

Another idea is to lighten the model by using transfer learning. One of the major advantages of transfer learning is the ability to reduce the amount of computation in the model by applying one-shot or few-shot learning. However, most DL algorithms currently achieve the effect of prediction ES by transferring large models trained on non-EEG datasets [91,95]. Current large models are often based on natural language or images as training data. These types of models have prior knowledge that is different from the signal as a feature. The performance when dealing with EEG data problem is often sub-optimal. Integrating existing EEG databases and training a large network for transfer learning may be one of the directions for future research.

4.2.4. Graph-based model

Future research on ES prediction based on GNN/GCN should go beyond general improvements and focus on customised methods tailored to network modelling. First, brain graph construction requires standardisation: current studies rely on metrics such as electrode distance or functional connectivity, while future research could explore learnable graph structures where edge weights are optimised in tandem with model training. Second, the dynamic evolution of ES necessitates the use of temporal graph models to capture connection patterns that change over time. Third, hierarchical GNNs can be employed to model both local electrode-level interactions and higher-level brain region modules within a unified framework. Fourth, integrating heterogeneous

graphs from EEG and auxiliary signals (e.g., ECG or EMG) enables multimodal expansion. Finally, incorporating interpretability methods (e.g., graph attention networks, GNNExplainer) is crucial for identifying key nodes and connections, thereby linking algorithmic predictions to clinically interpretable biomarkers.

4.2.5. Generalised prediction model

Building models for general or specific patients has been an unresolved disagreement in the field of ES prediction [16]. Due to the specificity of EEG across patients, many current seizure algorithms are still stuck on training algorithm weights for each patient. Some methods are now able to convert patient-specific models into models that work for different patients [138].

In the future, it is more productive to train a generalised prediction model. When ES prediction algorithms are used in the clinic and a new patient arises, the generalised model tends to predict better with a small amount of data. Patient-specific algorithms often struggle to achieve good results using a small training set. A novel research direction involves applying multimodal approaches to develop generalised seizure prediction algorithms. Multimodal fusion provides complementary physiological information sources that reduce false positives and enhance generalisation. Recent works such as DistilCLIP-EEG [139] and EFRM [140] provide concrete methodologies for applying multimodality to ES prediction. Building upon these approaches, future research may focus on cross-modal pre-training and domain-robust calibration across large-scale heterogeneous datasets, followed by patient-independent fine-tuning and evaluation, thereby establishing a generalised model workflow for clinical deployment.

4.2.6. Unsupervised learning

The above literature review on DL algorithms shows that the number of unsupervised learning algorithms is relatively small. However, this does not mean that the future of unsupervised learning is slim. Data labelling is often a tricky issue for ES prediction algorithms [118]. There are many epileptic patients whose EEG signals are currently recorded, but manually labelling these EEG signals is a time-consuming and labour-intensive task if done through a doctor [124]. Unsupervised learning has the unique advantage of requiring little or no labelling of data to construct the model. Thus, building a ES prediction algorithm based on unsupervised learning can greatly save doctors' labour, enabling them to do other work.

Also, generative networks like GAN can address the problem of a small amount of data for ES prediction. With the development of generative models, the generated data is getting closer and closer to the real data. This can dramatically reduce the time it takes for a DL model to collect data on a different patient when applied to that patient. For example, a generative algorithm is utilised to first generate the collected data for a patient, and then other algorithms are utilised to predict it [111].

4.2.7. Interpretability

As a biological signal, EEG often needs to consider the interpretability of the algorithm when constructing the algorithm, especially the biological aspect. Meanwhile, as DL algorithms in the medical field, which is closely related to patients' health, higher interpretability can better quantify the uncertainty in predicting ES. In the past few years, some researchers have tried to improve the interpretability of DL through EEG rhythms [16,141,142]. With more and more of these types of analyses being incorporated into DL algorithms, the interpretability of DL could be improved in the future. However, current review [143] indicate that existing explainability approaches lack assessments of the trade-offs between interpretability and performance. Examples include: elucidating the most interpretable AI applications, describing the most useful waveforms learned within XAI models, documenting domains of interest, and identifying correlations between frequency bands and epilepsy.

4.2.8. Assessment of indicators

Currently, there is no complete system for evaluating algorithms for ES prediction. The evaluation metric should based on the algorithms currently available, or researchers should write down the specific settings in each method, as in Table 4 to compare their work with others. Some specific evaluation metrics can also be introduced for different networks. For example, WD, FID and MSE are metrics that can compare the performance of different GANs [96]. So these metrics can also be added when evaluating the performance of different GANs for the ES prediction task. At the moment, e.g. ecosystem.org [22] has started to build a complete evaluation system to compare algorithms using this database.

5. Conclusion

In conclusion, we comprehensively summarise articles on the application of DL algorithms to the ES prediction problem, highlighting current research trends, pointing out existing challenges, and suggesting potential future research directions. Also we review data that are widely used in ES prediction.

CRediT authorship contribution statement

Xindi Huang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Hongying Meng:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Zhangyong Li:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

Acknowledgements

The authors would like to thank Brunel University London for support this project. This work was funded by Royal Society, United Kingdom (IEC\NSFC\223285) and National Natural Science Foundation of China (General Program) No. 62171073.

Data availability

No data was used for the research described in the article.

References

[1] World Health Organization, et al., Intersectoral Global Action Plan on Epilepsy and Other Neurological Disorders 2022–2031, World Health Organization, 2023.
[2] E.B. Bromfield, J.E. Cavazos, J.I. Sirven, Basic mechanisms underlying seizures and epilepsy, in: An Introduction To Epilepsy [Internet], American Epilepsy Society, 2006.
[3] A. Kumar, K. Maini, K. Arya, S. Sharma, Simple partial seizure, StatPearls [Internet] (2022) https://www.ncbi.nlm.nih.gov/books/NBK500005/.
[4] J.R. Gavvala, S.U. Schuele, JAMA patient page: Epilepsy, JAMA 316 (2016) 2686.
[5] S.J. Czuczwar (Ed.), Epilepsy, Exon Publications, Australia, 2022, https://www.ncbi.nlm.nih.gov/books/NBK580617/.
[6] A.H. Shoeb, J.V. Guttag, Application of machine learning to epileptic seizure detection, in: Proceedings of the 27th International Conference on Machine Learning, ICML-10, 2010, pp. 975–982.
[7] P. Tangkraingkij, Significant frequency range of brain wave signals for authentication, in: Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 2015, Springer, 2016, pp. 103–113.
[8] L.F. Nicolas-Alonso, J. Gomez-Gil, Brain computer interfaces, a review, Sensors 12 (2) (2012) 1211–1279.

[9] P.J. Karoly, D.R. Freestone, R. Boston, D.B. Grayden, D. Himes, K. Leyde, U. Seneviratne, S. Berkovic, T. O'Brien, M.J. Cook, Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity, Brain 139 (4) (2016) 1066–1078.

[10] T. Maiwald, M. Winterhalder, R. Aschenbrenner-Scheibe, H.U. Voss, A. Schulze-Bonhage, J. Timmer, Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic, Phys. D: Nonlinear Phenom. 194 (3–4) (2004) 357–368.

[11] M. Jia, W. Liu, J. Duan, L. Chen, C. Chen, Q. Wang, Z. Zhou, Efficient graph convolutional networks for seizure prediction using scalp EEG, Front. Neurosci. 16 (2022) 967116.

[12] Y. Gao, X. Chen, A. Liu, D. Liang, L. Wu, R. Qian, H. Xie, Y. Zhang, Pediatric seizure prediction in scalp EEG using a multi-scale neural network with dilated convolutions, IEEE J. Transl. Eng. Health Med. 10 (2022) 1–9.

[13] F. Mormann, R.G. Andrzejak, C.E. Elger, K. Lehnertz, Seizure prediction: the long and winding road, Brain 130 (2) (2007) 314–333.

[14] L. Kuhlmann, K. Lehnertz, M.P. Richardson, B. Schelter, H.P. Zaveri, Seizure prediction—ready for a new era, Nat. Rev. Neurol. 14 (10) (2018) 618–630.

[15] K. Rasheed, A. Qayyum, J. Qadir, S. Sivathamboo, P. Kwan, L. Kuhlmann, T. O'Brien, A. Razi, Machine learning for predicting epileptic seizures using EEG signals: A review, IEEE Rev. Biomed. Eng. 14 (2020) 139–155.

[16] M.O. Baud, T. Proix, N.M. Gregg, B.H. Brinkmann, E.S. Nurse, M.J. Cook, P.J. Karoly, Seizure forecasting: bifurcations in the long and winding road, Epilepsia 64 (2023) S78–S98.

[17] A. Shoeibi, P. Moridian, M. Khodatars, N. Ghassemi, M. Jafari, R. Alizadehsani, Y. Kong, J.M. Gorriz, J. Ramírez, A. Khosravi, et al., An overview of deep learning techniques for epileptic seizures detection and prediction based on neuroimaging modalities: Methods, challenges, and future works, Comput. Biol. Med. (2022) 106053.

[18] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, D.P. Subha, Automated EEG-based screening of depression using deep convolutional neural network, Comput. Methods Programs Biomed. 161 (2018) 103–113.

[19] M.J. Cook, T.J. O'Brien, S.F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, L. Litewka, et al., Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study, Lancet Neurol. 12 (6) (2013) 563–571.

[20] M.W. Bekele, D.E. Abera, M.T. Hailemichael, K.Y. Dechasa, M.N. Fanos, M.Y. Getachew, Electroencephalography data denoising with deep neural networks, in: 2023 IEEE International Conference on Control, Electronics and Computer Technology, ICCECT, IEEE, 2023, pp. 1074–1078.

[21] T. Haddad, N. Ben-Hamida, L. Talbi, A. Lakhssassi, S. Aouini, Temporal epilepsy seizures monitoring and prediction using cross-correlation and chaos theory, Heal. Technol. Lett. 1 (1) (2014) 45–50.

[22] L. Kuhlmann, P. Karoly, D.R. Freestone, B.H. Brinkmann, A. Temko, A. Barachant, F. Li, G. Titericz Jr., B.W. Lang, D. Lavery, et al., Epilepsyecosystem. org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG, Brain 141 (9) (2018) 2619–2630.

[23] A. Burrello, L. Cavigelli, K. Schindler, L. Benini, A. Rahimi, Laelaps: An energy-efficient seizure detection algorithm from long-term human iEEG recordings without false alarms, in: 2019 Design, Automation & Test in Europe Conference & Exhibition, DATE, IEEE, 2019, pp. 752–757.

[24] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, Phys. Rev. E 64 (6) (2001) 061907.

[25] P. Detti, G. Vatti, G. Zabalo Manrique de Lara, EEG synchronization analysis for seizure prediction: A study on data of noninvasive recordings, Processes 8 (7) (2020) 846.

[26] bbrinkm, W. Cukierski, 2014, URL: https://www.kaggle.com/c/seizure-prediction,

[27] V. Shah, E. Von Weltin, S. Lopez, J.R. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, J. Picone, The temple university hospital seizure detection corpus, Front. Neuroinformatics 12 (2018) 83.

[28] S.B. Dumanis, J.A. French, C. Bernard, G.A. Worrell, B.E. Fureman, Seizure forecasting from idea to reality. Outcomes of the my seizure gauge epilepsy innovation institute workshop, Eneuro 4 (6) (2017).

[29] R.W. Homan, J. Herman, P. Purdy, Cerebral location of international 10–20 system electrode placement, Electroencephalogr. Clin. Neurophysiol. 66 (4) (1987) 376–382.

[30] P. Mirowski, D. Madhavan, Y. LeCun, R. Kuzniecky, Classification of patterns of EEG synchronization for seizure prediction, Clin. Neurophysiol. 120 (11) (2009) 1927–1940.

[31] I.H. Sarker, Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions, SN Comput. Sci. 2 (6) (2021) 420.

[32] N. Mathur, A. Gupta, S. Jaswal, R. Verma, Deep learning helps EEG signals predict different stages of visual processing in the human brain, Biomed. Signal Process. Control. 70 (2021) 102996.

[33] J. Wang, S. Zhao, C. Fang, J. Yang, M. Sawan, Live demonstration: Real-time analyses of biosignals based on a dedicated CMOS configurable deep learning engine, in: 2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems, AICAS, IEEE, 2023, pp. 1–2.

[34] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[35] M. Krichen, Convolutional neural networks: A survey, Computers 12 (8) (2023) 151.

[36] H. Khan, L. Marcuse, M. Fields, K. Swann, B. Yener, Focal onset seizure prediction using convolutional networks, IEEE Trans. Biomed. Eng. 65 (9) (2017) 2109–2118.

[37] N.D. Truong, A.D. Nguyen, L. Kuhlmann, M.R. Bonyadi, J. Yang, O. Kavehei, A generalised seizure prediction with convolutional neural networks for intracranial and scalp electroencephalogram data analysis, 2017, arXiv preprint arXiv:1707.01976.

[38] Y. Xu, J. Yang, S. Zhao, H. Wu, M. Sawan, An end-to-end deep learning approach for epileptic seizure prediction, in: 2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems, AICAS, IEEE, 2020, pp. 266–270.

[39] D. Gramlich, P. Pauli, C.W. Scherer, F. Allgöwer, C. Ebenbauer, Convolutional neural networks as 2-D systems, 2023, arXiv preprint arXiv:2303.03042.

[40] R. Hussein, S. Lee, R. Ward, M.J. McKeown, Semi-dilated convolutional neural networks for epileptic seizure prediction, Neural Netw. 139 (2021) 212–222.

[41] K. Song, J. Fang, L. Zhang, F. Chen, J. Wan, N. Xiong, An intelligent epileptic prediction system based on synchrosqueezed wavelet transform and multi-level feature CNN for smart healthcare IoT, Sensors 22 (17) (2022) 6458.

[42] P. Peng, L. Xie, H. Wei, A deep fourier neural network for seizure prediction using convolutional neural network and ratios of spectral power, Int. J. Neural Syst. 31 (08) (2021) 2150022.

[43] A. Shankar, H.K. Khaing, S. Dandapat, S. Barma, Analysis of epileptic seizures based on EEG using recurrence plot images and deep learning, Biomed. Signal Process. Control. 69 (2021) 102854.

[44] F.E. Ibrahim, H.M. Emara, W. El-Shafai, M. Elwekeil, M. Rihan, I.M. Eldokany, T.E. Taha, A.S. El-Fishawy, E.S.M. El-Rabaie, E. Abdellatef, et al., Deep-learning-based seizure detection and prediction from electroencephalography signals, Int. J. Numer. Methods Biomed. Eng. 38 (6) (2022) e3573.

[45] Y. Mu, X. Zhang, M. Zhang, H. Wang, Epilepsy prediction based on PTE and TE of EEG signals using DSC-cnn, in: 2022 12th International Conference on Information Science and Technology, ICIST, IEEE, 2022, pp. 193–200.

[46] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J. Inman, 1D convolutional neural networks and applications: A survey, Mech. Syst. Signal Process. 151 (2021) 107398.

[47] X. Wang, C. Zhang, T. Kärkkäinen, Z. Chang, F. Cong, Channel increment strategy-based 1D convolutional neural networks for seizure prediction using intracranial EEG, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2022) 316–325.

[48] R. Jana, I. Mukherjee, Efficient seizure prediction and EEG channel selection based on multi-objective optimization, IEEE Access (2023).

[49] Y. Zhang, Y. Savaria, S. Zhao, G. Mordido, M. Sawan, F. Leduc-Primeau, Tiny CNN for seizure prediction in wearable biomedical devices, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2022, pp. 1306–1309.

[50] D. Sunaryono, J. Siswantoro, R. Sarno, R.I. Susilo, S.I. Sabilla, Epilepsy detection using combination DWT and convolutional neural networks based on electroencephalogram, in: 2023 International Seminar on Intelligent Technology and Its Applications, ISITIA, IEEE, 2023, pp. 411–416.

[51] J.S. Ra, T. Li, et al., A novel epileptic seizure prediction method based on synchroextracting transform and 1-dimensional convolutional neural network, Comput. Methods Programs Biomed. 240 (2023) 107678.

[52] C. Li, C. Lammie, X. Dong, A. Amirsoleimani, M.R. Azghadi, R. Genov, Seizure detection and prediction by parallel memristive convolutional neural networks, IEEE Trans. Biomed. Circuits Syst. 16 (4) (2022) 609–625.

[53] S. Ji, W. Xu, M. Yang, K. Yu, 3D convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2012) 221–231.

[54] N. Qi, Y. Piao, P. Yu, B. Tan, Predicting epileptic seizures based on EEG signals using spatial depth features of a 3D-2D hybrid CNN, Med. Biol. Eng. Comput. (2023) 1–12.

[55] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Trans. Neural Netw. 20 (1) (2008) 61–80.

[56] J. Bruna, W. Zaremba, A. Szlam, Y. LeCun, Spectral networks and locally connected networks on graphs, 2013, arXiv preprint arXiv:1312.6203.

[57] Q. Lian, Y. Qi, G. Pan, Y. Wang, Learning graph in graph convolutional neural networks for robust seizure prediction, J. Neural Eng. 17 (3) (2020) 035004.

[58] Y. Li, Y. Liu, Y.Z. Guo, X.-F. Liao, B. Hu, T. Yu, Spatio-temporal-spectral hierarchical graph convolutional network with semisupervised active learning for patient-specific seizure prediction, IEEE Trans. Cybern. 52 (11) (2021) 12189–12204.

[59] Y. Wang, W. Cui, T. Yu, X. Li, X. Liao, Y. Li, Dynamic multi-graph convolution based channel-weighted transformer feature fusion network for epileptic seizure prediction, IEEE Trans. Neural Syst. Rehabil. Eng. (2023).

[60] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, C. Fookes, Geometric deep learning for subject independent epileptic seizure prediction using scalp eeg signals, IEEE J. Biomed. Health Inform. 26 (2) (2021) 527–538.

[61] R. Saemaldahr, M. Ilyas, Patient-specific preictal pattern-aware epileptic seizure prediction with federated learning, Sensors 23 (14) (2023) 6578.

[62] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536.

[63] P. Malhotra, L. Vig, G. Shroff, P. Agarwal, et al., Long short term memory networks for anomaly detection in time series., in: Esann, vol. 2015, 2015, p. 89.

[64] R.G. Hefron, B.J. Borghetti, J.C. Christensen, C.M.S. Kabban, Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation, Pattern Recognit. Lett. 94 (2017) 96–104.

[65] X. Wu, Z. Yang, T. Zhang, L. Zhang, L. Qiao, An end-to-end seizure prediction approach using long short-term memory network, Front. Hum. Neurosci. 17 (2023) 1187794.

[66] K.M. Tsiouris, V.C. Pezoulas, M. Zervakis, S. Konitsiotis, D.D. Koutsouris, D.I. Fotiadis, A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals, Comput. Biol. Med. 99 (2018) 24–37.

[67] S. Ryu, I. Joe, A hybrid DenseNet-LSTM model for epileptic seizure prediction, Appl. Sci. 11 (16) (2021) 7661.

[68] S.M. Usman, S. Khalid, S. Bashir, A deep learning based ensemble learning method for epileptic seizure prediction, Comput. Biol. Med. 136 (2021) 104710.

[69] A.H. Abdulwahhab, A.H. Abdulaal, A.H.T. Al-Ghrairi, A.A. Mohammed, M. Valizadeh, Detection of epileptic seizure using EEG signals analysis based on deep learning techniques, Chaos Solitons Fractals 181 (2024) 114700.

[70] Z. Huang, Bidirectional LSTM-CRF models for sequence tagging, 2015, arXiv preprint arXiv:1508.01991.

[71] H. Daoud, M.A. Bayoumi, Efficient epileptic seizure prediction based on deep learning, IEEE Trans. Biomed. Circuits Syst. 13 (5) (2019) 804–813.

[72] K. Yan, J. Shang, J. Wang, J. Xu, S. Yuan, Seizure prediction based on hybrid deep learning model using scalp electroencephalogram, in: International Conference on Intelligent Computing, Springer, 2023, pp. 272–282.

[73] Q. Zhang, J. Ding, W. Kong, Y. Liu, Q. Wang, T. Jiang, Epilepsy prediction through optimized multidimensional sample entropy and Bi-LSTM, Biomed. Signal Process. Control. 64 (2021) 102293.

[74] Z. Yu, L. Albera, R. Le Bouquin Jeannes, A. Kachenoura, A. Karfoul, C. Yang, H. Shu, Epileptic seizure prediction using deep neural networks via transfer learning and multi-feature fusion, Int. J. Neural Syst. 32 (07) (2022) 2250032.

[75] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[76] D. Bahdanau, Neural machine translation by jointly learning to align and translate, 2014, arXiv preprint arXiv:1409.0473.

[77] X. Yang, J. Zhao, Q. Sun, J. Lu, X. Ma, An effective dual self-attention residual network for seizure prediction, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 1604–1613.

[78] Y. Jiang, Y. Lu, L. Yang, An epileptic seizure prediction model based on a time-wise attention simulation module and a pretrained ResNet, Methods 202 (2022) 117–126.

[79] B. Sun, J.J. Lv, L.G. Rui, Y.X. Yang, Y.G. Chen, C. Ma, Z.K. Gao, Seizure prediction in scalp EEG based channel attention dual-input convolutional neural network, Phys. A 584 (2021) 126376.

[80] J. Chung, Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014, arXiv preprint arXiv:1412.3555.

[81] Y. Wang, Y. Shi, Y. Cheng, Z. He, X. Wei, Z. Chen, Y. Zhou, A spatiotemporal graph attention network based on synchronization for epileptic seizure prediction, IEEE J. Biomed. Health Inform. 27 (2) (2022) 900–911.

[82] H. Ji, T. Xu, T. Xue, T. Xu, Z. Yan, Y. Liu, B. Chen, W. Jiang, An effective fusion model for seizure prediction: GAMRNN, Front. Neurosci. 17 (2023).

[83] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, L. Sun, Transformers in time series: A survey, 2022, arXiv preprint arXiv:2202.07125.

[84] J. Yan, J. Li, H. Xu, Y. Yu, T. Xu, Seizure prediction based on transformer using scalp electroencephalogram, Appl. Sci. 12 (9) (2022) 4158.

[85] Y. Wang, Y. Shi, Z. He, Z. Chen, Y. Zhou, Combining temporal and spatial attention for seizure prediction, Health Inf. Sci. Syst. 11 (1) (2023) 38.

[86] M. Şeker, M.S. Özerdem, Investigating convolutional and transformer-based models for classifying mild cognitive impairment using 2D spectral images of resting-state EEG, Biomed. Signal Process. Control. 105 (2025) 107667.

[87] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.

[88] X. Zhang, H. Li, Patient-specific seizure prediction from scalp EEG using vision transformer, ITOEC, in: 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference, vol. 6, IEEE, 2022, pp. 1663–1667.

[89] R. Hussein, S. Lee, R. Ward, Multi-channel vision transformer for epileptic seizure prediction, Biomedicines 10 (7) (2022) 1551.

[90] Y. Zhang, Y. Guo, P. Yang, W. Chen, B. Lo, Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network, IEEE J. Biomed. Health Inform. 24 (2) (2019) 465–474.

[91] Y. Gao, B. Gao, Q. Chen, J. Liu, Y. Zhang, Deep convolutional neural network-based epileptic electroencephalogram (EEG) signal classification, Front. Neurol. 11 (2020) 375.

[92] F. Tian, J. Yang, S. Zhao, M. Sawan, A new neuromorphic computing approach for epileptic seizure prediction, in: 2021 IEEE International Symposium on Circuits and Systems, ISCAS, IEEE, 2021, pp. 1–5.

[93] U. Mohammad, F. Saeed, SPERTL: Epileptic seizure prediction using EEG with ResNets and transfer learning, in: 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics, BHI, IEEE, 2022, pp. 1–5.

[94] X. Yang, L. Liu, Z. Li, Y. Xia, Z. Fan, J. Zhou, Semi-supervised seizure prediction model combining generative adversarial networks and long short-term memory networks, Appl. Sci. 13 (21) (2023) 11631.

[95] B. Sarvi Zargar, M.R. Karami Mollaei, F. Ebrahimi, J. Rasekhi, Generalizable epileptic seizures prediction based on deep transfer learning, Cogn. Neurodynamics 17 (1) (2023) 119–131.

[96] Y. Xu, J. Yang, M. Sawan, Multichannel synthetic preictal EEG signals to enhance the prediction of epileptic seizures, IEEE Trans. Biomed. Eng. 69 (11) (2022) 3516–3525.

[97] A. Tavanaei, M. Ghodrati, S.R. Kheradpisheh, T. Masquelier, A. Maida, Deep learning in spiking neural networks, Neural Netw. 111 (2019) 47–63.

[98] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2009) 1345–1359.

[99] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[100] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, 2017.

[101] A.H. Ribeiro, M.H. Ribeiro, G.M. Paixão, D.M. Oliveira, P.R. Gomes, J.A. Canazart, M.P. Ferreira, C.R. Andersson, P.W. Macfarlane, W. Meira Jr., et al., Automatic diagnosis of the 12-lead ECG using a deep neural network, Nat. Commun. 11 (1) (2020) 1760.

[102] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[103] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[104] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.

[105] M. Tan, Efficientnet: Rethinking model scaling for convolutional neural networks, 2019, arXiv preprint arXiv:1905.11946.

[106] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[107] G. Yang, D. Long, K. Wang, S. Xia, J. Zou, Epileptic seizure prediction method based on transition network data augmentation and fuzzy granular recurrence plot, Biomed. Signal Process. Control. 107 (2025) 107837.

[108] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).

[109] K.G. Hartmann, R.T. Schirrmeister, T. Ball, EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals, 2018, arXiv preprint arXiv:1806.01875.

[110] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 2015, arXiv preprint arXiv:1511.06434.

[111] N.D. Truong, L. Kuhlmann, M.R. Bonyadi, D. Querlioz, L. Zhou, O. Kavehei, Epileptic seizure forecasting with generative adversarial networks, IEEE Access 7 (2019) 143999–144009.

[112] N.D. Truong, A.D. Nguyen, L. Kuhlmann, M.R. Bonyadi, J. Yang, S. Ippolito, O. Kavehei, Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram, Neural Netw. 105 (2018) 104–111.

[113] D. Pascual, A. Amirshahi, A. Aminifar, D. Atienza, P. Ryvlin, R. Wattenhofer, Epilepsygan: Synthetic epileptic brain activities with privacy preservation, IEEE Trans. Biomed. Eng. 68 (8) (2020) 2435–2446.

[114] K. Rasheed, J. Qadir, T.J. O'Brien, L. Kuhlmann, A. Razi, A generative model to synthesize EEG data for epileptic seizure prediction, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 2322–2332.

[115] T. Yu, B. Cui, Y. Xu, X. Liu, Refine EEG spectrogram synthesized by generative adversarial network for improving the prediction of epileptic seizures, in: 2023 11th International IEEE/EMBS Conference on Neural Engineering, NER, IEEE, 2023, pp. 1–4.

[116] B. Gözütok, A. Ademoülu, Epileptic seizure prediction using convolutional autoencoder based deep learning, in: 2021 29th Signal Processing and Communications Applications Conference, SIU, IEEE, 2021, pp. 1–4.

[117] Y. Zhao, C. Li, X. Liu, R. Qian, R. Song, X. Chen, Patient-specific seizure prediction via adder network and supervised contrastive learning, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 1536–1547.

[118] Y. Yang, N.D. Truong, J.K. Eshraghian, A. Nikpour, O. Kavehei, Weak self-supervised learning for seizure forecasting: a feasibility study, R. Soc. Open Sci. 9 (8) (2022) 220374.
[119] P. He, L. Wang, Y. Cui, R. Wang, D. Wu, Unsupervised feature learning based on autoencoder for epileptic seizures prediction, Appl. Intell. (2023) 1–19.
[120] L. Guo, T. Yu, S. Zhao, X. Li, X. Liao, Y. Li, CLEP: Contrastive learning for epileptic seizure prediction using a spatio-temporal-spectral network, IEEE Trans. Neural Syst. Rehabil. Eng. (2023).
[121] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[122] A.M. Abdelhameed, M. Bayoumi, Semi-supervised deep learning system for epileptic seizures onset prediction, in: 2018 17th IEEE International Conference on Machine Learning and Applications, ICMLA, IEEE, 2018, pp. 1186–1191.
[123] D.P. Kingma, M. Welling, Auto-encoding variational bayes, 2013, arXiv preprint arXiv:1312.6114.
[124] M.H. Rafiei, L.V. Gauthier, H. Adeli, D. Takabi, Self-supervised learning for electroencephalography, IEEE Trans. Neural Networks Learn. Syst. (2022).
[125] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607.
[126] A.H. Abdulaal, M. Valizadeh, M.C. Amirani, A.S. Shah, A self-learning deep neural network for classification of breast histopathological images, Biomed. Signal Process. Control. 87 (2024) 105418.
[127] B. Świderski, S. Osowski, A. Cichocki, A. Rysz, Epileptic seizure prediction using Lyapunov exponents and support vector machine, in: International Conference on Adaptive and Natural Computing Algorithms, Springer, 2007, pp. 373–381.
[128] J.R. Williamson, D.W. Bliss, D.W. Browne, J.T. Narayanan, Seizure prediction using EEG spatiotemporal correlation structure, Epilepsy Behav. 25 (2) (2012) 230–238.
[129] H.T. Shiao, V. Cherkassky, J. Lee, B. Veber, E.E. Patterson, B.H. Brinkmann, G.A. Worrell, SVM-based system for prediction of epileptic seizures from iEEG signal, IEEE Trans. Biomed. Eng. 64 (5) (2016) 1011–1022.
[130] X. Wang, G. Gong, N. Li, S. Qiu, Detection analysis of epileptic EEG using a novel random forest model combined with grid search optimization, Front. Hum. Neurosci. 13 (2019) 52.
[131] N.D. Truong, L. Kuhlmann, M.R. Bonyadi, J. Yang, A. Faulks, O. Kavehei, Supervised learning in automatic channel selection for epileptic seizure detection, Expert Syst. Appl. 86 (2017) 199–207.
[132] A. Burrello, K. Schindler, L. Benini, A. Rahimi, One-shot learning for iEEG seizure detection using end-to-end binary operations: Local binary patterns with hyperdimensional computing, in: 2018 IEEE Biomedical Circuits and Systems Conference, BioCAS, IEEE, 2018, pp. 1–4.
[133] A. Burrello, K. Schindler, L. Benini, A. Rahimi, Hyperdimensional computing with local binary patterns: One-shot learning of seizure onset and identification of ictogenic brain regions using short-time iEEG recordings, IEEE Trans. Biomed. Eng. 67 (2) (2019) 601–613.
[134] A. Burrello, S. Benatti, K. Schindler, L. Benini, A. Rahimi, An ensemble of hyperdimensional classifiers: Hardware-friendly short-latency seizure detection with automatic iEEG electrode selection, IEEE J. Biomed. Health Inform. 25 (4) (2020) 935–946.
[135] I. Kiral-Kornek, S. Roy, E. Nurse, B. Mashford, P. Karoly, T. Carroll, D. Payne, S. Saha, S. Baldassano, T. O'Brien, et al., Epileptic seizure prediction using big data and deep learning: toward a mobile system, EBioMedicine 27 (2018) 103–111.
[136] J. Dan, U. Pale, A. Amirshahi, W. Cappelletti, T.M. Ingolfsson, X. Wang, A. Cossettini, A. Bernini, L. Benini, S. Beniczky, et al., SzCORE: seizure community open-source research evaluation framework for the validation of electroencephalography-based automated seizure detection algorithms, Epilepsia 66 (2025) 14–24.
[137] K. Burelo, G. Ramantani, G. Indiveri, J. Sarnthein, A neuromorphic spiking neural network detects epileptic high frequency oscillations in the scalp EEG, Sci. Rep. 12 (1) (2022) 1798.
[138] Z. Deng, C. Li, R. Song, X. Liu, R. Qian, X. Chen, Centroid-guided domain incremental learning for EEG-based seizure prediction, IEEE Trans. Instrum. Meas. (2023).
[139] Z. Wang, L. Shi, H. Wu, J. Luo, X. Kong, J. Qi, DistilCLIP-EEG: Enhancing epileptic seizure detection through multi-modal learning and knowledge distillation, IEEE J. Biomed. Health Inform. (2025).
[140] E. Jung, J. An, EFRM: A multimodal EEG–fNIRS representation-learning model for few-shot brain-signal classification, Comput. Biol. Med. 199 (2025) 111292.
[141] M. Liu, J. Ding, X. Wang, The interaction between circadian rhythm and epilepsy, Acta Epileptol. 4 (1) (2022) 28.
[142] P.J. Karoly, H. Ung, D.B. Grayden, L. Kuhlmann, K. Leyde, M.J. Cook, D.R. Freestone, The circadian profile of epilepsy improves seizure forecasting, Brain 140 (8) (2017) 2169–2182.
[143] D.K. Gurmessa, W. Jimma, A comprehensive evaluation of interpretable artificial intelligence for epileptic seizure diagnosis using an electroencephalogram: A systematic review, Digit. Health 11 (2025) 20552076251325411.