Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review

Khansa Rasheed, Adnan Qayyum, Junaid Qadir, Shobi Sivathamboo, Patrick Kwan, Levin Kuhlmann, Terence O'Brien, and Adeel Razi

(Clinical Application Review)

***Abstract*—With the advancement in artificial intelligence (AI) and machine learning (ML) techniques, researchers are striving towards employing these techniques for advancing clinical practice. One of the key objectives in healthcare is the early detection and prediction of disease to timely provide preventive interventions. This is especially the case for epilepsy, which is characterized by recurrent and unpredictable seizures. Patients can be relieved from the adverse consequences of epileptic seizures if it could somehow be predicted in advance. Despite decades of research, seizure prediction remains an unsolved problem. This is likely to remain at least partly because of the inadequate amount of data to resolve the problem. There have been exciting new developments in ML-based algorithms that have the potential to deliver a paradigm shift in the early and accurate prediction of epileptic seizures. Here we provide a comprehensive review of state-of-the-art ML techniques in early prediction of seizures using EEG signals. We will identify the gaps, challenges, and pitfalls in the current research and recommend future directions.**

***Index Terms*—Epileptic seizure, EEG, machine learning.**

Manuscript received February 17, 2020; revised May 31, 2020; accepted June 27, 2020. Date of publication July 13, 2020; date of current version January 22, 2021. *(Corresponding author: Khansa Rasheed.)*

Khansa Rasheed, Adnan Qayyum, and Junaid Qadir are with the Information Technology University (ITU)-Punjab, Lahore 54000, Pakistan (e-mail: msee18016@itu.edu.pk; adnan.qayyum@itu.edu.pk; junaid.qadir@itu.edu.pk).

Shobi Sivathamboo, Patrick Kwan, and Terence O'Brien are with the Department of Neuroscience, Central Clinical School, Monash University, Melbourne, VIC 3800, Australia, with the Department of Neurology, Alfred Health, Melbourne, VIC 3004, Australia, and also with the Departments of Neurology and Medicine, The Royal Melbourne Hospital, The University of Melbourne, Melbourne, VIC 3010, Australia (e-mail: shobi.sivathamboo@monash.edu; patrick.kwan@monash.edu; te.obrien@alfred.org.au).

Levin Kuhlmann is with the Faculty of Information Technology, Monash University, Clayton, VIC 3800, Australia, and also with the Department of Medicine, St. Vincent's Hospital, The University of Melbourne, Parkville, VIC 3010, Australia (e-mail: levin.kuhlmann@monash.edu).

Adeel Razi is with the Turner Institute for Brain and Mental Health, Monash University, Clayton, VIC 3800, Australia, with the Monash Biomedical Imaging, Monash University, Clayton, VIC 3800, Australia, with the Wellcome Centre for Human Neuroimaging, University College London, London WC1E 6BT, U.K., and also with the Department of Electronic Engineering, NED University of Engineering and Technology, Karachi 75270, Pakistan (e-mail: adeel.razi@monash.edu).

Digital Object Identifier 10.1109/RBME.2020.3008792

I. Introduction

EPILEPSY is a group of neurological disorders that are characterized by an enduring predisposition to generate recurrent seizures and can affect individuals of any age. Epilepsy arises from the gradual neurobiological process of '*epileptogenesis*' [1], which causes the normal brain network to fire neurons in a self-sustained hyper-synchronized manner in the cerebral cortex. According to the World Health Organization (WHO), 70 million people worldwide have epilepsy and epilepsy trails only migraine, stroke, and Alzheimer's disease in the list of the most widespread brain diseases [2]. The seizures caused by epilepsy are debilitating and disrupt the day-to-day activities of the patients, and are associated with an increased risk of premature mortality. The dearth of neurologists in many countries, particularly in developing countries, further complicates the management of epilepsy.

Even though epilepsy and seizures are sometimes referred to synonymously in some literature, it is worth noting that not all seizures are epileptic and convulsions and seizures may also occur due to acute neurological insults (such as stroke, brain trauma, metabolic disturbances, and drug toxicity) without necessarily reflecting a long term predisposition to recurrent unprovoked seizures (i.e. epilepsy).

An epileptic seizure (ES) is caused by a sudden abnormal, self-sustained electrical discharge that occurs in the cerebral networks and usually lasts for less than a few minutes. ES attacks are hard to predict, moreover, severity and duration of attack also cannot be anticipated. Therefore, injuries and safety issues from the events are a major concern for patients and their families. Hence, early prediction of epilepsy attacks is crucial to avoid and counter their adverse consequences. The brain activity of patients with epilepsy can be categorized as different states: *pre-ictal* (immediately preceding seizure), *ictal* (during a seizure), *post-ictal* (immediately following a seizure), and *interictal* (in-between seizures). Further details of these terms are provided in the section of the paper. ES prediction is a classification problem focused on differentiating between the pre-ictal and interictal states. Due to the recurrent nature of epilepsy, ES occurs in groups and patients afflicted from seizure clusters can acquire advantage through the forecasting of follow-on seizures.

Electroencephalography (EEG) is a particularly effective diagnostic tool to study the functional anatomy of the brain

TABLE I
COMPARISON OF THIS PAPER WITH EXISTING SURVEYS. LEGENDS: $\surd$= DISCUSSED, $\times$= NOT DISCUSSED, $\approx$ = PARTIALLY DISCUSSED

Reference	Year	Focused Area	EEG Analysis Techniques	Features	ML	DL	Pitfalls	Future Direction
Mormann et al. [3]	2006	Epilepsy prediction	$\times$	$\times$	$\surd$	$\times$	$\approx$	$\surd$
Subha et al. [4]	2008	EEG signals	$\surd$	$\surd$	$\times$	$\times$	$\approx$	$\times$
Yuedong et al. [5]	2011	EEG signals	$\surd$	$\approx$	$\times$	$\times$	$\times$	$\times$
Acharya et al. [6]	2013	EEG signals for epilepsy	$\surd$	$\surd$	$\surd$	$\times$	$\times$	$\surd$
Gadhoumi et al. [7]	2016	ES prediction	$\times$	$\surd$	$\surd$	$\times$	$\surd$	$\approx$
M.Iftikhar et al. [8]	2018	DL for EEG signals	$\surd$	$\approx$	$\times$	$\surd$	$\times$	$\times$
Acharya et al. [9]	2018	ES prediction	$\times$	$\surd$	$\surd$	$\approx$	$\times$	$\surd$
Kuhlmann et al. [10]	2018	ES prediction	$\times$	$\times$	$\surd$	$\approx$	$\times$	$\surd$
Roy et al. [11]	2019	DL for EEG	$\surd$	$\times$	$\times$	$\surd$	$\surd$	$\surd$
Li et al. [12]	2019	DL for EEG	$\times$	$\times$	$\approx$	$\surd$	$\surd$	$\surd$
This paper	2020	ES prediction	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$

during an ES attack. The prediction and medication of epilepsy have been broadly studied through EEG. EEG signals, which are non-Gaussian and non-stationary, measure the electrical activity in the brain which are in turn used to diagnose the type of the brain disorders. The analysis of EEG measurements helps segregate normal and abnormal function of the brain. For an accurate prediction of epilepsy, it is necessary to examine EEG recordings of longer duration. Expert neurologists examine epilepsy by studying continuous EEG signals recorded over several days, weeks, or even months, which requires a huge amount of human effort and time. Over the years, various studies have employed machine learning (ML)-based prediction methods to address this issue. Deep learning (DL) is an advanced ML technology that is capable of learning patterns more precisely from large collections of data by processing it through a multi-layer hierarchical architecture. The ability of DL to produce very accurate results has influenced the researchers to tackle numerous real-world applications by employing DL techniques with various researchers proposing DL-based approaches for the ES prediction in the last five to six years.

The objective of this paper is to review and elaborate upon the primary advances in the employment of ML methods for epilepsy prediction. We consider DL as a subset of the broader class of ML.[1] The advancement in the DL methods of ES prediction is also summarized in the paper. We will provide a brief introduction to neuroscience, various tools used for studying brain, and how they have been or could be used for the prediction of epilepsy.

Contributions of this paper: Although there exist several reviews that specifically cover epilepsy seizure prediction using EEG signals, to the best of our knowledge, there does not yet exist a review that covers in depth the application of ML methods for predicting epileptic seizures. For instance, Mormann *et al.* have provided an overview of the evolution of seizure predicting methods since the 1970 s till 2006 [3] and have covered the major issues related to methodology of ES prediction. Gadhoumi *et al.* have provided a brief overview of valid methods used for ES prediction and comprehensively described the statistical significance of the results of the prediction [7]. In the recently published review, Kuhlmann *et al.* have briefly described the advancement in the field of ES prediction and ES prediction competitions. They concluded that these advancements with standard statistical evaluations are opening ways for the development of ES prediction methodologies and they refined the existing guidelines to achieve this development [10]. This survey is unique because it provides comprehensive answers to questions like why there is a need for ML techniques for ES prediction, how the evolution of relatively newer techniques like DL is proving highly useful for ES prediction, and discusses directions for future research in this area. The comparison of this paper with existing surveys is presented in Table I.

Organization of this paper: The organization of this paper is depicted in Fig. 1. In Section II, a brief background of neuroscience, EEG, and epilepsy prediction is presented. Section III covers data-driven ML approaches for ES prediction. Section IV comprises of identifying several pitfalls in applying these methods. Future directions and open research problems are presented in Section V. Finally, the paper is concluded in Section VI. List of acronyms used in the paper is provided in Table II.

II. BACKGROUND ON EEG, NEUROSCIENCE, AND EPILEPSY PREDICTION

A. A Brief Introduction to Neuroscience and Neuroimaging

Neuroscience is the multidisciplinary study of the brain. It integrates multifarious disciplines including *neuroanatomy* (in which neuroanatomists engage with the structures of the human brain), *neurochemistry* (where chemists observe the chemical properties of intercommunication in the brain), *neurophysiology* (where the neurophysiologists investigate the electrical properties of the brain) and *neuropsychology* (where psychologists endeavor to interpret the cognitive domains and the structures that sustain those cognitive domains in neuroscience) [14]. Neuroscience also has further divisions—e.g., molecular neuroscience, cognitive neuroscience [15], clinical neuroscience, computational neuroscience [16], developmental neuroscience, and cultural neuroscience, to name just a few.

[1]This is a standard assumption made in mainstream works [13] interested readers are referred to https://bit.ly/3gg71Jc for more details. We make the distinction between DL and ML where necessary to separately talk about general ML approaches for ES prediction and to differentiate them from specific DL-based approaches.

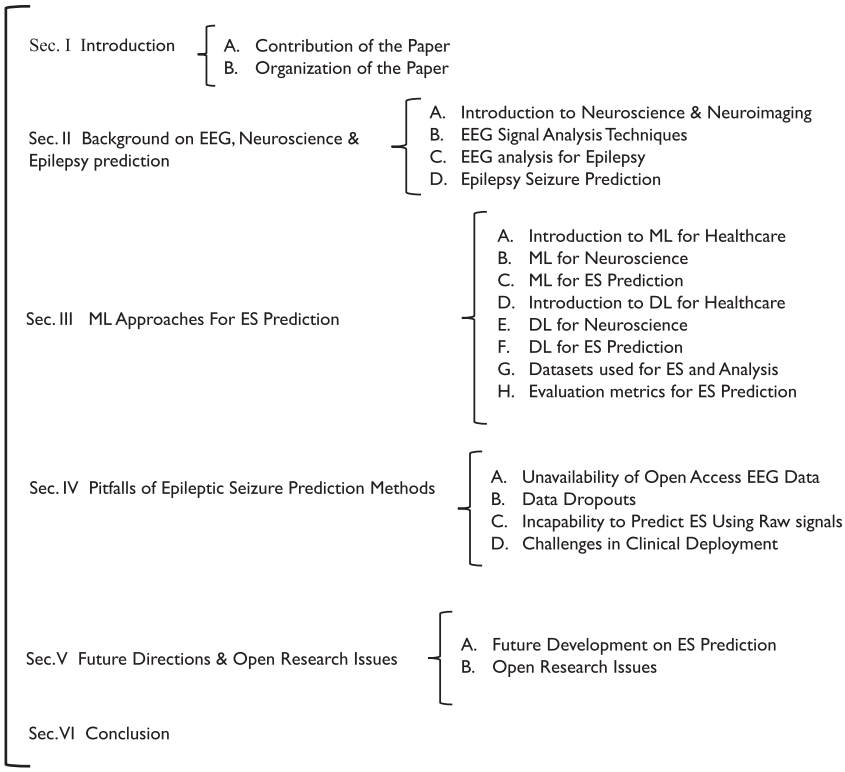

Fig. 1. Organization of the paper.

TABLE II
LIST OF ACRONYMS

ANN	Artificial Neural Network
ApEn	Approximate Entropy
BLDA	Bayesian Linear Discriminant Analysis
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DNN	Deep Neural Network
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
ES	Epileptic Seizure
EEG	Electroencephalography
FD	Fractal Dimension
FPR	False Prediction Rate
FT	Fourier Transform
HE	Hurst Exponent
HHT	Hilbert-Huang Transform
HP	Hjorth Parameter
LLE	Largest Lyapunov Exponent
LSTM	Long-Short Term Memory
MLP	Multi-layer Perceptron
RNN	Recurrent Neural Network
SEF	Spectral Edge Frequency
SBP	Spectral Band Power
SM	Statistical Moment
SOM	Self Organizing Map
SpM	Spectral Moment
TPR	True Positive Rate
WFT	Wavelet Fourier Transform
WT	Wavelet Transform

The brain is anatomically segregated to communities which make up a functionally specialized brain network (functional segregation). These functionally segregated communities are functionally interconnected (functional integration) to perform very complex tasks like they implement cognition [17]. Neuroimaging uses various ways to directly or indirectly image the structure and the function of the central nervous system. Two broad categories are structural imaging that pertains to anatomy, pathology or injury and functional imaging that deals with metabolism, pharmacology or cognition. Some of the important and widely used neuroimaging techniques are namely: computed tomography (CT) that computes the absorbed amount of X-rays to provide a series of cross-sectional images of the brain; positron emission tomography (PET) that generates the image of active molecule binding; structural magnetic resonance imaging (MRI) that examines the anatomy and pathology of the brain; functional MRI that examines the brain activity; diffusion MRI that maps the diffusion of water molecules in the brain to reveal macroscopic details of brain tissues; and magnetic resonance spectroscopy (MRS) imaging is used to study the metabolic changes in brain tumors, stroke, and seizure, etc. The brain's electrical activity in different physiological situations can be measured using EEG, which falls under the category of functional imaging.

B. EEG Signal Analysis Techniques

1) Introduction to EEG: In 1923, Hans Berger contrived EEG, a non-invasive functional imaging methodology to study the brain. EEG records the electrical signals from the cerebral cortex by measuring the electrical activity of the group of neurons. Compared to the functional MRI, EEG provides a higher temporal insight into neural activity but has a lower

spatial resolution. Typically, five frequency bands are analysed for processing EEG signals, *Delta* (up to 4 Hz), *Theta* (4–8 Hz), *Alpha* (8–12 Hz), *Beta* (12–26 Hz), and *Gamma* (26–100 Hz). The amplitude of EEG range from 10 μV–100 μV while its frequency ranges from 1 Hz–100 Hz. To diagnose a disease, or to decode brain activity by using EEG data, one initially extracts features or uses spectral information of raw EEG data by applying Fourier transform (FT) or wavelet transform (WT). These extracted features or transformed raw data is then used to train an ML-based classifier while DL algorithms have been proved efficient for automatic extraction of feature for training.

There are two methods of EEG recording based on the position of the reference electrode.

- *Bipolar Montage:* In a bipolar montage, both electrodes are placed on an electrically active region of the scalp and the voltage difference between electrodes is measured.
- *Monopolar Montage/Unipolar Montage:* In a monopolar montage, one electrode is active and the other one (or two connected electrodes) serves as a reference electrode. The reference electrode should be as electrically neutral as possible in comparison with brain activity. The recorded signals are the difference between the active brain regions and reference electrodes. Regularly used reference sites are the ear lobe, the left or right mastoid, the tip of the nose, the chest, and the balanced non-cephalic sterno-vertebral lead [18].

The traditional method for the recording of EEG signals is to place the electrodes on the surface of the skull, which is known as scalp EEG. The main drawback of scalp EEG is that the recorded signals become distorted owing to a large distance between neurons inside the skull and the electrodes. For the quality of signals to be enhanced in terms of distortion and amplitude, intracranial electroencephalography (iEEG) signals are recorded by placing the electrodes on the exposed surface of the brain.

EEG possesses several characteristics which makes it quite preferable to use for ES prediction research. Along with its ability to track the various changes occurring in the brain during epilepsy, another main feature it provides is the relatively lower hardware cost which makes it able to be used for a large number of patients and to record for longer duration. Multiple other techniques such that fMRI or MEG require bulky and immobile equipment which piles up the cost to millions of dollars. Heart rate variability (HRV) analysis using electrocardiography (ECG) (recording of the electrical activity of the heart) provides a good means of predicting epilepsy [19]. To achieve the best results for the current practical approaches to predict epileptic seizures, *My Seizure Gauge* is the most example of a wearable device created to work as a personalized advisory device for seizure prediction [20]. This device can cover intracranial EEG recordings, scalp EEG, ECG, electromayography (EMG),[2] electrodermal activity (EDA), photoplethysmography (PPG), and respiration. Liu *et al.* [21] used a combination of EEG, ECG, and respiratory signals to predict ES. They mentioned a slight increment in the results by using the multi-bio signals as compared to uni-data (EEG signals only). Using a 1D convolutional neural network, they achieved a 0.5931 F score for EEG signals only and 0.6106 F score for combined data. Valderrama *et al.* [22] and Keider *et al.* [23] also used a fusion of EEG and ECG signals for ES prediction, but they did not compare their results with EEG signals only.

2) Analysis Techniques: EEG analysis methods can mainly be classified into the time domain methods, frequency domain methods, time-frequency domain methods, and linear or non-linear methods [4].

a) Time domain methods: EEG recordings are non-stationary and non-linear functions of time. *Linear prediction* is a time-domain method in which the output is calculated from the input and earlier outputs. *Principal component analysis (PCA)*, *linear discriminant analysis (LDA)* and *independent component analysis (ICA)* are widely used unsupervised time-domain methods to summarize EEG data. PCA is used to transform the high-dimensional data (in case of epilepsy high-dimensional feature vectors) to a low-dimensional data [24] while ICA decomposes high-dimensional data into linear statistically independent components [25]. In EEG data analysis, ICA is most commonly used to remove artifacts. Whereas, LDA is used to reduce dimensions of feature sets by finding linear combinations of feature vectors [26].

b) Frequency domain methods: During an epileptic seizure, there is a sudden change in the frequency of EEG signals, which is measurable by applying frequency-domain methods, e.g., using Fourier transform (FT). One can used either *parametric* or *non-parametric methods* to estimate the power spectrum using FT [27]. Welch (a non-parametric) method, a modified version of widely used *periodogram* method, is generally used for the estimation of PSD. But this has a disadvantage of spectral leakage and is overcome by employing parametric methods. Parametric methods provide better frequency resolution by assuming the EEG signal is a stationary random process. Moving average (MA), auto-regression (AR), and auto-regressive moving average (ARMA) are commonly applied parametric methods [28].

c) Time-frequency domain methods: Above mentioned time-domain and frequency-domain methods have limitations of providing exact frequencies involved at a particular time instant and the information of time moment respectively. To overcome these limitations, wavelet transform (WT), a time-frequency based analysis technique, is widely used to obtain multi-resolution decomposed sub-band signals by passing the EEG signal through filter banks [29].

d) Non-linear methods: Non-linear analysis methods are applied to detect the coupling among harmonics in signal's spectrum. *Higher order spectra (HOS)*, various measures of entropy—e.g., *approximate entropy (ApEn)*; *Kolmogorov entropy*; *sample entropy (SampEn)*—along with the *Hurst exponent (H)*, *largest Lyapunov exponent (LLE)* are widely used non-linear parameters for EEG analysis [30]. Entropy and LLE are commonly used as features for epilepsy classification. Entropy provides clues about information stored in the probability distribution of a signal and measures the uncertainty or randomness in the patterns of the data. A higher value of the entropy refers to

[2]EMG signals record the electrical activity produced by skeletal muscles.

highly random patterns of data. LLE provides the information of the dependence of the system on initial conditions. For a more detailed review of the analysis techniques, the interested readers are referred to [6] and [4].

C. EEG Signal Analysis for Epilepsy

Analysis of the EEG signals is the primary method to identify ES activities in the brain. EEG recordings are an important clinical tool for distinguishing ES from non-ES. EEG signals recorded, before and during a seizure, contains characteristics that can be used to identify the different stages of an epileptic seizure, and the pre- and post-seizure periods. During the past two decades, various studies have shown experimental confirmation that seizures are preceded by spatial and temporal changes in EEG (i.e., change in short-term maximum Lyapunov exponent (STLmax) of EEG studied by Iasemidis *et al.* [31], Le Van Quyen *et al.* [32] examined change in phase synchronization and showed the existence of different seizure states, Kalitzin *et al.*, [33] used relative phase clustering index as a measure to show the pre-seizure changes). Change in spike rate of EEG signals before epilepsy is also widely studied and it shows a logical existence of pre-ictal state [34]. However, there does not exist one or more definite features to characterize the EEG states. How to best define these states and to find the definite characteristic is still an open issue [10]. These stages are briefly described below.

a) *Pre-ictal State:* A pre-ictal state becomes apparent during a said time period before the occurrence of a seizure and does not occur at the rest of the times. It might not necessarily be visually apparent. However, it will reflect changes in the underlying signals and would be predictive of seizures within a specific range of values. For a pre-ictal state to be of use clinically in a warning system, it has to be detected early enough so that the time under false warning is minimized [10].
b) *Pro-Ictal State:* In this state seizures are more likely but not guaranteed to happen.
c) *Ictal and Interictal State:* The ictal state is a change in EEG signals during a seizure and interictal is the stage between two following seizure onsets. For the same person, the number of epileptogenic neurons, cortical region, and the span of seizure can be altered.
d) *Post-Ictal State:* This state is after the occurrence of a seizure.

The wave pattern may hold valuable information about brain activity. Experienced neurologists can detect disorders by visually observing the EEG signals. However, this procedure is time-consuming and is prone to faulty detection due to high temporal and spatial aspects of the dynamic non-linear EEG data. Therefore, computerized techniques, EEG signal parameters extraction, and analysis can be profoundly beneficial in the diagnostics.

D. Epilepsy Seizure Prediction

In the 1970 s, early research of ES prediction carried out using linear approaches of feature extraction [35]. While in 1980 s, the development of non-linear methods helped researchers to employ these techniques for feature extraction because of the non-linear nature of EEG signals [36] [46] . With the recognition of EEG patterns of epilepsy—i.e, pre-ictal, ictal, and interictal patterns—the use of the pre-ictal stage for ES detection was also applied in this decade. In 1998, early prediction of ES almost 6 sec before the seizure onset, was carried out by Salant *et al.* [49] which was further developed by Drogenlen *et al.* in 2003 [50]. They used Kolmogorov entropy as a feature to predict ES 2–40 min before onset. First international workshop on ES prediction was held in 2002 in which dataset of multi-day recordings of EEG provided by different epilepsy centers. Later, several studies were carried out on this dataset [51]. In 2003, Mormann *et al.* used the fact that the hyper-synchronous firing of neurons in the brain is a cause of ES and found that the phase synchronization of different EEG channels decrease before seizure onset [52]. In the first decade of the present century, studies based on extensive EEG data have raised doubts about the performance of measures calculated in the previous century. Researchers found that the results of earlier studies based on a selected and inadequate amount of data could not be reproduced on extensive and unseen data.

It was decided to conduct competitions on seizure prediction in international workshops conducted on the said topic. The purpose of these competitions was to standardize the comparison of the performance of algorithms trained on a common dataset. The first seizure prediction competition was held in collaboration with International Workshop on Seizure Prediction 3 (IWSP3) in 2007 while the second competition conducted in 2009 was in collaboration with IWSP4. In both the competitions, the contestants were provided with the continuous iEEG recording from three epileptic patients. However, the performance results of the algorithms were not satisfactory. American Epilepsy Society Seizure Prediction Challenge which was held in 2014, involved short-term human iEEG containing 942 seizures recorded over more than 500 days and long-term iEEG recordings of dogs with epilepsy. All contestants were provided with the same 10 min long training and testing data. The Area Under the Curve (AUC) was used as a performance evaluation metric. With the same structure, another contest held by Melbourne University which involved long-term iEEG recording with 1139 seizures. For more details of the contest see [53]. The contests were open to any algorithm computing basic features of EEG signals for ES prediction or ML methods trained on these basic features. In any case we still do not really know what features or algorithms are best. In the contests, people submitted algorithms, that were too complicated. So it is difficult to say which feature or ML algorithm was best. The organizers of the contests are working towards dissecting it now with Epilepsyecosystem.org. Recent work of Matias Maturana *et al.* [54] presents a solution that might work well across patients. They identified the critical slowing of brain signals as an indicator for ES prediction. A timeline for the development of the EEG data measures is depicted in Fig. 2, the interested readers can refer to [55] for more detailed information about the history of these developments.

e) Types of ES: In 1981 International League Against Epilepsy (ILAE) made classes of epilepsy to facilitate clinicians diagnose and treat patients with epilepsy. They revised this classification in 2017. According to the new classification scheme

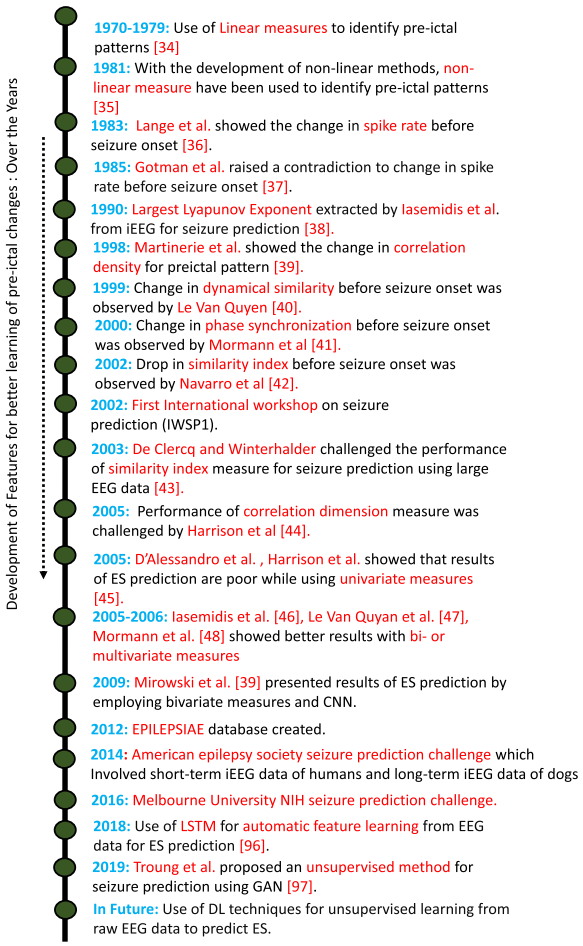

Fig. 2. The timeline for the development of EEG measures used for ES prediction.

of ILAE, epilepsy is categorized as focal epilepsy, generalised epilepsy, and unknown epilepsy. These are further classified in motor onsets and non-motor onsets. Interested readers may consult [56] for further reading. To classify [57], detect [58], and predict [59] these ES types, researchers are using ML methods. Researchers are also working on predicting the surgical outcomes for these types of epilepsy [60]. To predict the drug-resistant epilepsy cook *et al.* held a study by implanting advisory devices in patients for almost one year [59]. Sungtae *et al.* evaluated different ML algorithms to predict drug-resistant ES [61].

III. ML Approaches for ES Prediction

In this section, we provide a comprehensive review of the literature using ML-based methods for ES prediction and we start this section by first highlighting the potential of using ML techniques for healthcare and neuroscience applications.

A. Introduction to ML for Healthcare

ML is proliferating across research areas over the past few decades by using statistical methods to recognize patterns in large collections of data. The availability of large-scale biomedical data is turning over a new leaf for healthcare researchers. Development of effective medical tools relies on data analysis approaches and the advancements of ML techniques. Because the manual detection of representations is not possible due to the complex structure of medical data and that is why ML is extensively used in healthcare for the diagnosis of diseases, e.g., detection of breast cancer [62], classification of skin cancer [63], diagnosis of Alzheimer disease [64], prediction of epilepsy [65], and diagnosis of diabetic retinopathy in retinal images [66]. Electronic health records (EHR)-based ML algorithms have proved beneficial for prediction of future diseases and are capable of automatically diagnosing patients given their clinical status [67] although still much work is needed. Biomedical fields with large image datasets—such as radiology [68], cardiology [69], pathology [70], and genomics [71]—are using various ML methods for automatic diagnosis, classification, and prediction of various disease.

B. ML for Neuroscience

Learning about the structure and functional anatomy of the human brain has been the foremost focus of neuroscientists in recent years. The advancements in technology have enabled the neuroscientists to acquire, process, and analyse the neuroimaging data at unprecedented detail, while ML and DL are the paramount examples for such enabling technologies that can be used as a potential exploratory source for building theories about brain functioning for neuroscientists [72]. In this section, we provide a general introduction to various ML techniques (e.g., *supervised learning*, *unsupervised learning*, and *reinforcement learning*) that have been used in the field of neuroscience.

1) Supervised Learning: In supervised learning, training data accompanied by labels assigned by human experts is fed to the learning algorithm for extracting the relation between data and labels so that the system can classify the unseen data accurately to their respective categories. For instance, a training data consists of images with labels of house, a dog, a cat and we want an algorithm that can predict the label of an image previously unknown to the system. These algorithms have wide applications in the field of computational and theoretical neuroscience—an example technique is support vector machine (SVM), a supervised learning algorithm generally used for prediction of ES (described in a subsection III-C). Analysis of neural mechanisms under stress is carried out using a supervised ML approach [73].

2) Unsupervised Learning: Our brain receives most of the information in a day without any guidance. The brain develops a working model from the repetition of information and uses this model to make a perception. This perception is then used for detecting the patterns in new information. Unsupervised learning algorithms are motivated by how the brain studies new things through perceptions. Unsupervised learning applies unclassified or unlabeled data for training of the algorithms. These algorithms are extensively used in the identification and classification of diseases from neurophysiological data. As a representative example, we refer to the work of Drysdale *et al.*

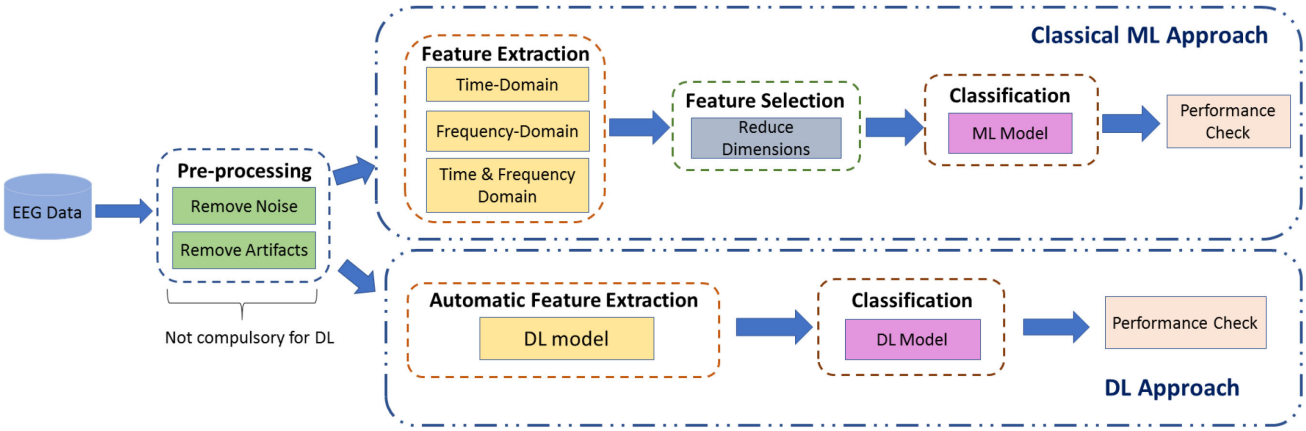

Fig. 3. Process of epilepsy prediction using EEG data and classification algorithm.

who classified depression types using fMRI [74] and the work of O'Donnell *et al.* who used a clustering algorithm for the identification of white matter tracts from diffusion MRI [75].

3) Reinforcement Learning: Animal psychology, how animals communicate with each other and with the environment, helped to develop reinforcement learning (RL) [76]. RL is a significant illustration of the advancement of technology due to the collaboration of neuroscience and AI. Reinforcement Learning is the process of developing a policy to maximize the rewards of interaction between an agent and its environment. Central factors of a reinforcement learning system are a policy, reward signal, value function, and model of the environment.

C. ML for Epileptic Seizure Prediction

Since the last century, researchers are working to overcome the hurdles related to the detection and prediction of epilepsy. As EEG signals are a key source for monitoring brain activity before, during, and after ES so the first focus of ES prediction research was on the analysis of EEG recordings. EEG signals are vitiated by eye-movements, blinks, cardiac signals, and muscle noise. Several filtering and noise reduction methods are used to decrease the effect of these various sources of noise and artifacts [77]. After the removal of artifacts, significant features are needed for building ML methods for the identification and classification of pre-ictal and interictal stages. Fig. 3 shows the classical ML methodology for the epilepsy prediction and also highlight the major difference between the use of ML and DL technique. Basically one can give the raw data or minimally processed data (i.e., without extraction of features from the raw data) to a DL model for pattern learning.

1) Signal Processing: Noise and artifact identification is a crucial procedure in raw biomedical signals. To reduce the influence of these artifacts in feature extraction, filtering of these artifacts is needed. Multiple techniques have been employed for filtering, e.g., band-pass filter, wavelet filter, finite impulse response filter, and adaptive filter. This processing is also performed to normalize the data to make it comparable with the recording of other patients. There are also many data dropouts or corrupted data in the EEG recording due to limitations of implanted electrodes which lead to the insignificant performance of algorithms. Due to muscle artifacts and environmental noise, there also exist some outliers in data. The presence of these outliers badly influences the extracted features.

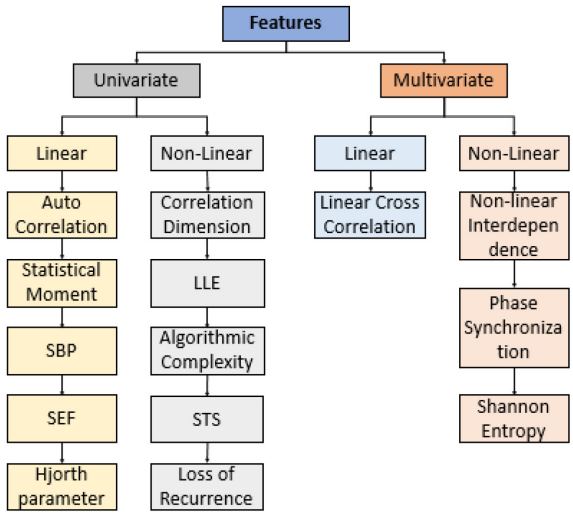

Fig. 4. Feature classification based on number of channels of EEG data.

2) Feature Extraction and Selection: All prediction methods need reliable features, well correlated with pre-ictal and interictal stages. One can categorize these features based on the number of EEG channels as univariate (measures taken on each EEG channel separately) and multivariate (measures taken on two or more EEG channels) features. Further categorization of each of these is as linear or nonlinear features. Florian *et al.* compared the performance of univariate and bivariate measures containing both linear and non-linear strategies for ES prediction [86]. They noted that while using univariate measures, pre-ictal variations transpired 5–30 min before ES onset. While bivariate measures performed better by capturing pre-ictal changes at least 240 min before an ES onset. Fig. 4 shows some of the linear and nonlinear measures used in the literature for ES prediction. Linear measures performed better or some times similar to nonlinear measures.

3) Classification: Identification of pre-ictal and interictal patterns from EEG data is carried out using ML algorithms, e.g, artificial neural network (ANN), k-means clustering, decision trees, SVM, and fuzzy logic. Mostly threshold-based on features values are utilized to make conclusions. However, ML-based studies broadly focused on the extraction of optimized features for prediction.

a) Use of bispectral features to predict seizure: Higher-order spectrum (HOS) features of iEEG recordings used to detect the seizure in earlier studies [87]. However, Assi *et al.* [85] used the HOS features to present that the bispectrum analysis of EEG provides significant phase information. They showed that the normalized bispectral entropy and the normalized squared bispectral entropy decreased during the pre-ictal state of seizure. They extracted these features from the 30 sec non-overlapping windows of iEEG recordings of epileptic dogs. They trained a 5-layer multilayer perceptron (MLP) for the classification of pre-ictal and interictal classes. The input layer of MLP consisted of 16 nodes as there are 16 channels of iEEG signals. They added 3 hidden layers of 30, 60 and 30 nodes of ReLu activation function. They computed the F1 score and p-value corresponding to pre-ictal and interictal distribution using each feature. However, researchers prefer to analyse the performance of the algorithm in terms of sensitivity and specificity for defined seizure prediction horizon and seizure occurrence period. This aspect is missing in this study.

Permutation entropy (PE) has been used in various early studies to characterize the EEG states of epilepsy [88], [89]. In 2007, Li *et al.* [90] used PE to distinguish pre-ictal states in rats. Recently, Yang *et al.* [65] used PE as a feature extracted from the iEEG data of Freiburg hospital data. They analysed 83 seizures from 19 patients. They trained an SVM classifier with RBF kernel using 5 sec segments of features as input. Sensitivity and false prediction rate (FPR) used as performance analysis measures. They achieved 94% sensitivity and 0.11 FPR on average with a mean SPH of 61 min.

b) Use of non-negative matrix factorization to predict seizure: To make a computationally efficient patient-specific ES predicting algorithm Stojanovic *et al.* [82] proposed the use for non-negative matrix factorization (NMF) for extracting features from the power spectra of interictal and pre-ictal classes of iEEG data. They decompose the power spectra of the Freiburg and Epilepsyecosystem dataset into time and frequency components. These time and frequency components then used as features (input) to SVM for classification. To get better results, they used synthetic minority over-sampling technique (SMOTE) to overcome the issue of imbalanced data. They achieved 97.42% accuracy, 95.2% sensitivity, and 99.4% specificity on average for the Freiburg dataset. They achieved 75.2% accuracy, 69% sensitivity, 78.6% specificity on average for Epilepsyecosystem dataset.

c) Use of selected amount of data to predict seizure: To reduce the dimensions of EEG is one of the foremost concerns of researchers for the processing of data and predicting seizure using this data. Various dimension reduction techniques have been proposed with some pros and cons. In the resent work, Kitano *et al.* [80] proposed the use of a small amount of data to predict seizure. They used only 20 min data of 9 patients out of the hours-long recording of 24 patients of CHBMIT database. 20 min data consisted of the 10 min pre-ictal and 10 min interictal data. They applied DWT on 4 sec non-overlapping windows of this 20 min data and extracted zero-crossings of level 1 detailed coefficients of DWT. They used a self-organizing map (SOM), formerly introduced by Teuvo Kohonen in 1982 [91], for mapping the input data in clusters of pre-ictal and interictal states. They achieved 98% sensitivity using the selected amount of data.

Although Kitano *et al.* achieved highly significant results on the selected amount of data, there are various flaws with this selection. Pre-ictal and interictal patterns have temporal variations across patients and with inpatient. To randomly select 10-min pre-ictal data is not a significant approach. Training of neural networks on a small amount of data leads to the over-fitting of results. With the training on the randomly selected small amount of data, the model might not be able to show significant performance in real-time scenarios. Summary of recent work on ES prediction using ML techniques is in Table III. Comparison of sensitivity and prediction time for ES prediction using ML methods is shown in Fig. 5.

D. Introduction to DL For Healthcare

DL methods are the result of advancements in ML research that provide an ability to process raw data. DL methods comprise of multiple layers of computational (non-linear) modules that work mutually to process data and produce an ultimate result. These multiple layers help in extraction of appropriate features and their examination or analysis for the output result. For example, in the classification task, higher layers of representation amplify features of the input that are significant for discrimination and subdue unnecessary variations. The core of DL methods is that they contain modular layers that are designed to learned data using general-purpose algorithms [100]. These layers are building blocks of deep neural networks (DNN). Commonly used neural networks are *convolutional neural network* (CNN) and *recurrent neural network* (RNN) [100]. The structure of CNN is similar to that of the connectivity pattern of neurons in the brain. Convolution operation of a CNN is just like a filter with weights for extracting the features from multi-dimensional input data. While the RNNs are used to find logical sequences in input data. The output of each hidden layer passed to the next layer and also fed back to itself. Simply, the current output is a combined experience of the present moment and history. The key difference between CNN and RNN architecture is that CNNs only consider the current input while RNN considers current input and as well as the previous input, i.e., it contains memory logic. RNN performs significantly better on time series data while CNN is good for tasks like image classification.

DL architectures have been used in many medical domains, e.g., in clinical imaging [101], genomics, and proteomics [102], computational biology [103], and disease prediction [92]. DL algorithms are turned out to be adequate in detecting intricate patterns in high-dimensional data for classification, especially in EEG data. CNN is a widely used neural network for the

TABLE III
SUMMARY OF ML METHODS USED FOR ES PREDICTION. N/M INDICATES THE NOT MENTIONED ENTRIES IN THE TABLE

Year	Ref	Predictive characteristics	Model	EEG Type	No. of Patients	No. of Seizure per Patient	Prediction Time	Sensitivity	False Positive per hr
MIT Database									
2017	Usman et al. [78]	Entropy, ApEn, HP, SpM, SM	SVM	Scalp	24	3.5	23.48 min	92.23%	N/M
2018	Usman et al. [79]	Variance, Skewness, SD, HP, Entropy, Kurtosis	KNN Naive Bayes SVM	Scalp	24	3.5	34 min	97.44% 90.66% 97.07%	N/M
2018	Kitano et al. [80]	Zero-crossing of DWT coefficients	SOM	Scalp	9	>4	N/M	98%	N/M
Freiburg Database									
2017	Sharif et al. [81]	Distribution of 6 fuzzy rules	SVM	iEEG	19	4.4	42 min	96.6%	0.05-0.08
2018	Yang et al. [65]	Permutation Entropy	SVM	iEEG	21	>2	61.93 min	94%	0.111
2020	Stojanovic et al. [82]	NMF	SVM	iEEG	5	>2	N/M	95.2%	N/M
EPILEPSIAE Database									
2015	Bandarabadi et al. [83]	Amplitude distribution histogram & Spectral power	N/M	iEEG/EEG	24	3.6	8 sec	73.98%	0.06
2017	Direito et al. [84]	22 univariate features	SVM	iEEG/EEG	216	5.6	N/M	38.5	0.2
IEEG.org Database									
2018	Assi et al. [85]	Bi-spectral Entropy Bi-spectral Squared Entropy Mean magnitude of bispectrum	MLP	iEEG	3 Dogs	N/M	N/M	N/M	N/M
Epilepsyecosystem Database									
2020	Stojanovic et al. [82]	NMF	SVM	iEEG	3	>200	N/M	69%	N/M

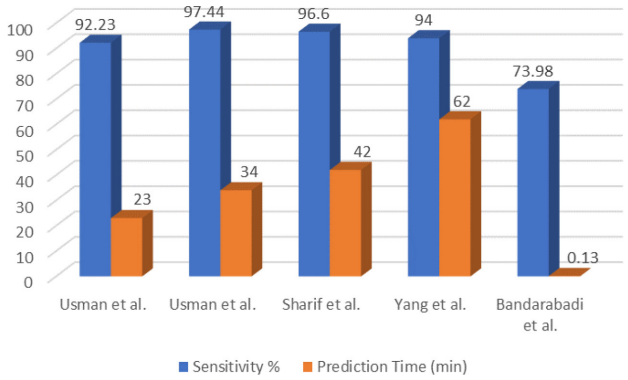

Fig. 5. Comparison of sensitivity and prediction time for ES prediction using ML methods.

training using EEG data because it can be very effective to reduce noise [104].

E. DL for Neuroscience

DL is solving problems in many fields, however, a potent relation exists between DNN and the study of the nervous system. ANNs were considered as a model for brain activity computations [105], while CNNs are used for visual information processing and the activations of hidden layers of CNN are considered as the activity of neurons in connected brain regions associated with the processing of visual sensory motors. Deep networks are a valuable mean of computation in neuroscience as these are statistical time-series models of neural activity in the brain, e.g, the CNN can act as an encoding model of computational neuroscience. In connectomics, to understand the mapping of the connectivity of neural networks in the brain, deep networks are used to understand the connectivity of neural units from 3D electron microscopic images [106]. The existing era of advancement is accelerating the research of neuroscience-inspired ML tools [105].

F. DL for ES Prediction

ML classification algorithms use feature vectors, derived from traditional signal processing methods for training and provide good accuracy but a generalised model can not be anticipated from these techniques. For seizure prediction through an ML approach, script writing requires feature extraction stage that takes a lot of time. The presence of noise and artifacts in data makes feature extraction very complex to handle. Hence it is a challenging problem to produce a generalised automatic system with loyal performance especially even when limited training samples are available. On the other hand, the ability of DL algorithms to automatic feature learning is opening new ways of research in ES prediction [107]. Features learned through DL methods are more distinguishing and robust than hand-crafted features [108].

1) Use of CNN for ES Prediction: To introduce a method that can be applied for all patients with minimum pre-processing of EEG data Troung *et al.* [93] proposed a CNN based prediction method. They used the Freiburg hospital iEEG database and CHBMIT scalp EEG database for training and testing of the CNN model. Short-term Fourier transform (STFT) used to transform the raw EEG data into a two-dimensional matrix. This image is then fed to the CNN for feature learning and classification of pre-ictal and interictal states. For evaluation of

the performance of the algorithm, they set the seizure prediction horizon (SPH) to 5 min and seizure occurrence period (SOP) to 30 min and used sensitivity and false prediction rate as evaluation metrics. While following a leave-one-out cross-validation, they reached 79.7% sensitivity with 0.24 FPR on raw EEG and 89.8% sensitivity with 0.17 FPR on standardized data.

For real-time clinical use of ES predictor, SPH must be long enough to allow the patient to come out of a dangerous situation and take precautionary measures and SOP should not be too long. The work of Haider *et al.* [92] performed better than previous work by giving 87.8% sensitivity and 0.142 FPR with 10 min SPH. They used CHBMIT and MSSM databases for the training and the testing of the model. Raw EEG converted into wavelet tensors and CNN used to extract features from transformed data for classification of pre-ictal and interictal data.

After the establishment of the feasibility of ES prediction in a clinical setting by demonstrating the success of implantable recording system by Cook *et al.* [59], new avenues of further research have been opened. To take the work of Cook *et al.* forward, Isabell *et al.* [107] presented a portable seizure prediction system with tunable parameters according to the patient's need. They transformed iEEG data into spectrograms and used frequency transformed data as an input to the deep learning model for automatic pre-ictal feature learning. These tunable parameters are the sensitivity of the system, duration, and the number of alarms. For the tuning of these parameters, the authors added a processing layer in the model. They deployed their prediction algorithm on a low-power TrueNorth chip to introduce a wearable device. Their prediction system performed an average sensitivity of 69% and average time in warning of 27%, significantly exceeding a comparable random predictor for all patients by 42%.

Motivated by the work of Cook *et al.* [59] and Karoly *et al.* [109], demonstrating that the seizure prediction algorithm could not produce satisfactory prediction sensitivity for some patients, Ramy Hussain *et al.* [94] worked on some part of the data of these patients. They applied a technique of downsampling to reduce the dimension of data by a factor of 4. They explained that handcrafted features are not suitable for authentic ES prediction because the EEG data not only varies between patients but also varies for the same patient over time. They transformed the EEG data by applying STFT and fed this transformed data to CNN. To learn local features they used 1×1 convolutional layers and for abstract feature learning, they used larger convolutions. They obtained 87.85% sensitivity and 0.84 AUC on average as a performance measure of the prediction algorithm. They also explained that reasons for the limited performance of ES prediction algorithms are data drop-outs, data mismatch, imbalance distribution and outliers in data.

CNN methods are not only used for classification but also widely used for feature extraction. These features then used as input to a simple classifier. This aspect of the CNN is used by Usman *et al.* [95] for ES prediction. They proposed a three-layered CNN architecture to extract feasible features from pre-ictal and interictal classes of the CHB-MIT dataset. They used SVM as a classifier for the classification of pre-ictal and interictal classes based on features extracted from CNN. They anticipated ES 21 min before onset with 92.7% sensitivity and 90.8% specificity.

2) Unsupervised DL Method for ES Prediction: One problem in the ES prediction is the availability of labeled data. To overcome this problem the first step taken by Troung *et al.* [98] is the use of a generative adversarial network (GAN) to do unsupervised training. They fed the spectrogram of STFT of EEG to GAN and used trained discriminator as a feature to predict seizures. This unsupervised training is significant because it not only provides real-time prediction using EEG recording also does not require manual effort for feature extraction. They used AUC as a performance measure with 5 min SPH and SOP of 30 min. They compared their results with supervised methods of model training, and their approach performed well with 77.68% AUC for CHBMIT scalp EEG data, 75.47% AUC on Freiburg hospital data and 65.05% AUC with EPILEPSIAE database. A summary of these works is presented in Table IV.

Another work suggested by Ahmed *et al.* [110] to predict ES with unlabeled data is the use of a 2D deep convolutional autoencoder for learning the appropriate features from the unlabeled raw EEG signals of CHB-MIT database. The proposed architecture of the encoder contains four 2D convolutional layers and three max-pooling layers and the decoder consists of four 2D convolutional layers and three upsampling layers. They trained the autoencoder for 100 epochs with the selected amount of data and saved the weights of the encoder. These weights are later used to initialize weights while training other patients' networks. This transfer learning technique is used to facilitate fast learning. The features learned from the trained encoder are used as an input to a Bidirectional LSTM for classification of pre-ictal and interictal EEG intervals. For one hour of SPH, they scored an average sensitivity of 94.6% and an average low false prediction alarm rate of 0.04FP/h.

3) Use of RNN for ES Prediction: Tsiouris *et al.* [97] used long short-term memory (LSTM) for the first time for the prediction of an epileptic seizure. They compared the performance of different architectures of LSTM for randomly selected input segment size of 5–50 sec. They compared the performance of three architectures of LSTM using feature vectors of EEG segments as input to LSTM, where the feature vector consists of various features from the time domain, frequency domain, and local and global measures from graph theory. LSTM-1 architecture consisted of a single layer with 32 memory units, while the number of memory units increased to 128 in LSTM-2 architecture. The number of memory units preserved at 128 but an extra layer of equal dimension added to LSTM-3. The performance of LSTM-3 was the best among the three considered architectures. Using the pre-ictal window of 15 min, they also evaluated the performance of LSTM-3 for raw EEG as input as compared to the performance of the feature vector. They showed that deep architecture with raw EEG input and satisfactory performance is still an open issue in the said field. On average, LSTM-3 performed better with 99.28% sensitivity for 15 min pre-ictal period, 99.35% sensitivity for 30 min pre-ictal period, 99.63% sensitivity with 60 min pre-ictal period, and 99.84% sensitivity with 120 min pre-ictal period. However, too much feature engineering needed for these results. Comparison

TABLE IV
SUMMARY OF DL METHODS USED FOR ES PREDICTION

Year	Ref	Predictive characteristics	Database	EEG Type	No. of Patients	No. of Seizure per Patient	Prediction Time	Sensitivity	False Positive/hr
CNN									
2017	Haider et al. [92]	Wavelet Transform	MSSM CHB-MIT	Scalp	47	2.78	8 min 6 min	87.8%	0.142
2018	Truong et al. [93]	STFT	Freiburg CHB-MIT American Epilepsy Society	iEEG/ Scalp	28 humans 5 canines	N/M	5 min	81.4% 81.2% 82%	0.06 0.16 0.22
2019	Ramy Hussain et al. [94]	STFT	Melbourne seizure prediction competition dataset	iEEG	3	380	5 min	87.8%	N/M
2020	Usman et al. [95]	Feature extracted from CNN	CHB-MIT	Scalp	24	3.7	21 min	92.7%	N/M
2020	Ranjan et al. [96]	Feature extracted from CNN	CHB-MIT	Scalp	24	3.7	20 min	68%	0.05
LSTM									
2018	Tsiouris et al. [97]	Various time and frequency features	CHB-MIT	Scalp	24	7.7	15-120 min	99.28%	0.11-0.02
GAN									
2019	Troung et al. [98]	STFT	Freiburg CHB-MIT EPILEPSIAE	iEEG/ Scalp	56	6.8	5 min	N/M	N/M
DCAE + Bi-LSTM									
2019	Hisham et al. [99]	Raw data	CHB-MIT	Scalp	8	5.37	1 hr	99.72%	0.004

TABLE V
OVERVIEW OF EEG DATABASES

Database	No. of Subjects	No. of Channels	Recording Type	No. of ES	Duration of Each Recording (Hour)	Sampling Frequency (Hz)
CHB-MIT	24	23	Scalp EEG	198	1 (some cases have 2-4 hours of recording)	256
MSSM	28	22	Scalp	61	48-192	256
Freiburg	21	128	iEEG	88	At least 24	256
Bonn	25 (5 sets each consists of recording of 5 subjects)	1 (100 flies of single channel data in each set)	Scalp/iEEG	Dataset E is the recording of ictal stage	23.6	173
EPILEPSIAE	30	122	Scalp/iEEG	1800+	96	250-2500
TUH	10874	24-36	iEEG	≈ 14777	-	250

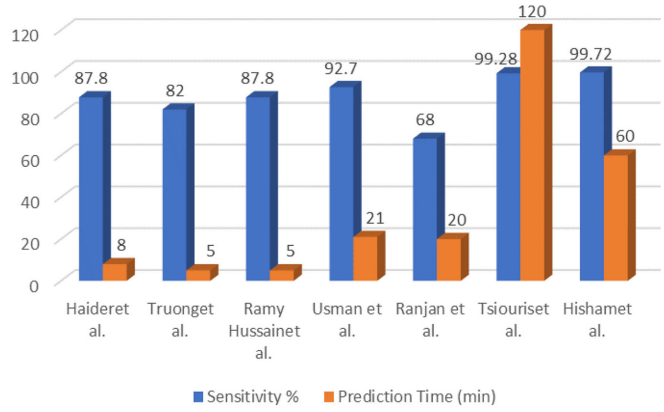

Fig. 6. Comparison of sensitivity and prediction time for ES prediction using DL methods.

of sensitivity and prediction time for ES prediction using DL methods is shown in Fig. 6.

G. Datasets Used for ES Prediction and Analysis

EEG is becoming a prevailing mean of acquiring brain signals to detect and predict ES. To this end, various open-access databases have been published by various hospitals and research centers. For instance, the Center of Epilepsy at Children's Hospital, Boston and Temple University Hospital have made their EEG databases publicly available to the researchers who aim to develop ML/DL methods and other statistical analysis based methods at physionet.org. In Table V, a summary of such widely used and publicly available databases is provided. Although, the database of Bonn University is not large enough but is extensively used for the detection of ES in the literature. It consists of 5 datasets A, B, C, D and E. CHB-MIT database has data of 22 patients with 9–24 recordings of each patient and every recording is 1 hour long with some discontinuities due to hardware limitation (some cases have 2–4 hours long recordings). Freiburg Hospital's database was one of the considerable databases which contained iEEG data of 21 subjects with around 88 seizures but recently it has been merged into EPILEPSIAE database to provide more larger datasets due to which this database is not open-source now.

H. Evaluation Metrics for ES Prediction

The clinical employment of ES prediction methods requires a sufficient performance and quality check and different evaluation metrics have been proposed in the ES prediction literature.

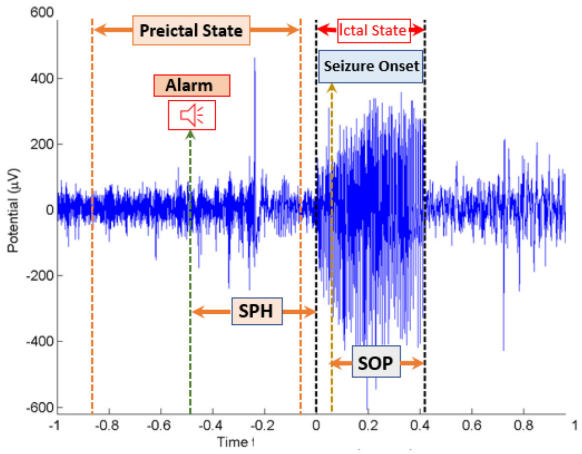

Fig. 7. Concept of seizure occurrence period (SOP) and Seizure prediction horizon (SPH). With a precise prediction, a seizure must occur after SPH and within the SOP.

For instance, Osorio *et al.* proposed sensitivity and false prediction rate as performance parameters of ES predictors [111]. Sensitivity is measured as the ratio of correctly predicted seizures to all seizures. Moreover, contrary to the ideal situation, one can not prevent false prediction and with the increase in sensitivity, the false prediction rate also increases. The widely used evaluation metrics are described below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

Where

- True positive (TP) is the number of correctly predicted ES.
- False Negative (FN) is the number of ES that are incorrectly predicted as not seizures.
- True negative (TN) is the number of correctly predicted no-seizure.
- False Negative (FP) is the number of non-ES that are incorrectly predicted as seizures.

An ES predictor generates an alarm before seizure onset and according to ideal situation the predictors must anticipate the exact time of onset. In practical applications, a predictor anticipates a duration of the high probability of occurrence of seizure. So, another performance check metric is a *seizure occurrence period* (SOP), the time duration in which there is a possibility of seizure. Another metric is the *seizure prediction horizon* (SPH), the duration of time between the alarm and the inception of SOP [112], [113]. In Fig. 7 the concept of SPH and SOP is illustrated. Another widely used evaluation metric for the detection and prediction of ES is the Receiver Operator Characteristic (ROC) curve. ROC is a plot used to exhibit the diagnostic efficacy of classifiers. ROC evaluates the true positive rate against false positive rate during the interictal and pre-ictal states. The use of ROC curves as evaluation matric is evident from [95], [114] Area under the ROC curve (AUC) is also used for performance evaluation of ES predicting algorithms [53].

IV. Pitfalls of Epileptic Seizure Prediction Methods

ML has solved several challenges for ES prediction that include manual, tedious, and time-consuming analysis methods. Model interpretation is crucial and pattern identification in data is as significant as data fitting. A fundamental difficulty in bio-medicine is the correct classification of ailment and its sub-types. Enormous available biomedical data can lead to the identification of more comprehensive sub-types. One can easily find various ways in which ML, specifically DL has improved the EEG analysis. The hierarchical nature of neural networks has significantly developed the potential of learning features from raw data or minimally processed data. Automatically learned features through DL methods are more powerful and effective than those extracted by analytical tools. This shows that DL has the potential to give high performance on analysis tasks. Research on epilepsy prediction has been going for many years and much progress has been made using different approaches, but there also exist many problems. Some potential pitfalls related to ES prediction using ML methods are discussed next.

A. Unavailability of Open Access EEG Data

A core problem in the ES prediction and analysis research is the unavailability of long-term EEG data. In 2005 Iasemidis *et al.* performed the prediction alarm almost 91 min before the ES onset on private EEG data [47]. However, no one has been able to reproduce these results since then on any publicly open EEG data. so, there is an urgent need for open access sharing of EEG databases with long-term recordings and also code sharing (using Github or similar repositories) for reproducibility of findings.

B. Data Dropouts

One of the main reasons for the low performance of prediction algorithms is the missing observations. There are many zero or nearly equal to zero values in the observed data because of the failure of communication between the wearable devices or implanted devices with limited storage capacity and storage device for several possible reasons. Learning from corrupt, or missing, data has not attained much attention to the machine learning community. However, there is a need to add missingness indicators in algorithms that can provide significant pieces of information for making predictions.

C. Incapability to Predict ES Using Raw Signals

Time consumption and computational cost will increase due to excessive feature extraction. We require a quick prediction with comparatively low-power hardware and cheaper computational cost so that the real-time system for ES prediction becomes feasible. Unfortunately, researchers are not able to build a model for learning feasible features from raw signals yet. Although DL has greatly rectified the problem of feature extraction by automatically extracting features from pre-processed data, the

limitation is that these methods require an abundant amount of data for effective prediction [115].

D. Challenges in Clinical Deployment

Despite the impressive performance of ML/DL techniques in ES prediction, these techniques are not yet translated for clinical deployment. Below we provide a non-exhaustive list of the roadblocks in taking ES prediction to clinic.

1) Reliability and Reproducibility of ES Predictions: While the ES prediction works well in a prototype, it suffers from the lack of reliability and reproducibility in clinical setting. One of main the reasons for the failure of the reproducibility of predictive results in clinical practice is due to the flaw in setting up the seizure prediction as a binary classification problem [116]. Epilepsy is a multi-scale problem [117] and also strongly depends on the circadian profile of the patient (wake-sleep routine of a patient, environment, time of the day, etc.) [118]. That is why stating the problem as a simple binary classification of seizure segments is not enough to design efficient predictive ML/DL methods. This challenge can be met by developing more complete understanding of the mechanisms of how epilepsy is caused [10].

2) Generalisability of ES Predictions: One of the main concerns with clinical implementation is the lack of generalisablity of ML-based seizure prediction algorithms across types of seizures and patients. ES-related biomarkers are patients specific and the findings may not generalise well due to the variability of seizure characteristics across patients (sometimes even the same individual patient can show different characteristics at different times). This variability and, as stated above, reflects the general lack of mechanistic understanding of epilepsy leading to poor predictive performance in clinical settings [119].

3) Heterogeneity of Seizures Types: The major reasons behind the failure in the use of the ES prediction algorithms in clinical practice are the variance in the causes of ES occurrence, difficulty in locating the area of occurrence [120], and a lack of understanding of how the seizures spread [121]. To develop a robust solution, researchers need to understand the different causes, and the ensuing consequences of epileptic seizure.

4) Development of Multimodal Framework: The development of multimodal framework is critical for clinical settings since reliance on a single modality may well not be enough for clinical translation. Studies have revealed several non-neurological changes during the pre-ictal state. Heart rate variability [19], change in local cerebral blood flow (CBR) [122], variation in inhibitory neurotransmitter (GABA) (experiment on mice neural network) [123], and an increase in blood-oxygen-level-dependent activities [124] are evidence for the need of a multimodal framework for ES prediction. However, these variations are also patient dependent, and also leads to generalization problems as mentioned above.

5) Effective Hardware Implementations: For the translation of the ES prediction algorithms into clinical practice, there is a dire need to develop cost-effective and power-efficient hardware implementation. In particular, computational complexity should be low so they can be deployed in real time for implantable or wearable devices. Although some initial work on hardware implementation has been performed [125], more work is required on manufacturing cost-effective devices at scale that produce optimized results.

6) Development of Automated and User Friendly Interfaces: The patients and clinicians are increasingly more concerned about the lack of customized user interface which is intuitive and easy to operate and maintain [126]. Current development in this area is still far from where the interaction between patients, researchers, and clinicians can happen in a meaningful way to develop an optimized user friendly interface. To develop a user-friendly predictive system, there is a need to get specifications from patients whether, for example, if they require invasive or non-invasive solution, how much accuracy and sensitivity will be sufficient for a given severity of disease and patient history, and whether patient would be better served with only a predictive signal or a closed-loop therapy based system would be more efficient [127], [128].

V. Future Directions and Open Research Issues

In this section, we present directions for future work in ES prediction and various open research issues that require further investigation.

A. Future Development on ES Prediction

The following are the promising possible future developments on ES prediction.

1) Curse of Data Dimensionality: EEG signals are recorded using multiple electrodes due to which the dimensions of the recorded signals increases and analyzing multi-channel EEG signal become difficult. An ideal approach is to convert multi-channel EEG data into a single channel by applying appropriate signal analysis techniques (e.g., converting them to spectrograms) or to make use of the single-channel EEG signals from the collection of brain signals. It has been observed that interesting seizure-related brain activities were weak in a few signals of multi-channel EEG [129]. Therefore, the selection of a good quality signal for effective prediction of seizure is crucial. In a recent study, signal quality index (SQI) based adaptive algorithm is presented for best channel selection in multi-channel EEG of nonconvulsive seizure patient [129] and substantial efforts have been made on developing adaptive algorithms for EEG channel selection using different dimensionality reduction techniques such as principal component analysis (PCA) [130], independent component analysis (ICA) [131], and discernibility matrix-based dimensionality reduction [132], etc. However, the development of an optimal technique for dimensionality reduction of EEG signals or an optimal way for the selection of most significant channel in multi-channel EEG recordings can be interesting future work in ES prediction. This will decrease time involved in feature extraction across different channels and will eventually enable timely prediction of ES onset.

2) Handling Data Annotation Problem: In the literature, the problem of ES prediction and detection is mostly formulated as a supervised learning task that requires labeled data. The EEG recordings are manually annotated by expert neurologists and

physicians which is a costly, time-consuming, and tidy task. The performance of the ML techniques significantly depends upon the quality of annotations and to increase the efficacy of ML techniques, and in particular DL the natural approach is to use more training data. The development of a true validation set for assessing the performance of the trained model is also important. However, the annotation of large-scale collections of EEG recordings into respective categories is practically not feasible, it hinders the applicability of ML/DL techniques. This necessitates the development of automated ways for data labeling such as active learning, data labeling using generative models. For instance, ES prediction using GAN is presented in [98]) that uses unsupervised training for ES prediction. However, such methods need the availability of enough variables to learn classification. Therefore, future work on ES prediction using ML/DL techniques should incorporate annotation efficiency in the model development.

3) Real-Time Monitoring of ES Patients: Real-time or near real-time monitoring of ES patients could be key for enabling timely interventions before the onset of ES, e.g., emergency services in case of severe onset. In this regard, continuous monitoring of ES patients using wearable EEG devices interconnected with smartphones, and other IoT devices have great potential to augment the capacity of ML/DL for ES prediction or forecasting. For example, EEG recordings collected using wearable sensors can be transmitted to fog/cloud servers for the analysis and ML/DL method(s) deployed on the fog/cloud server can transmit back its outcomes, e.g., prediction of ES. Similar system architecture has been proposed to diagnose the abnormal ECG using deep learning in [133], the idea of Wearable IoT-cloud-based health monitoring system (WISE) is also proposed by Wan *et al.* [134].

B. Open Research Issues

In this section, we briefly describe general open research issues related to ML/DL and its deployment in healthcare applications.

1) Distribution Data Sharing and Management: In clinical settings, patients' data is produced across different facilities and to develop efficient ML/DL techniques, sharing of distributed data across different departments and as well across different hospitals is required. Moreover, data from different domains can be integrated to extract knowledge required for different tasks, e.g., the annotation. Recurrent neural networks and different natural language processing (NLP) techniques can be used to extract rich knowledge from raw clinical notes and electronic health records (EHR) that can enhance the capability of data annotators. In addition, ML/DL methods can be developed that are capable of learning from heterogeneous sources and distributed data. However, the cost of data sharing and management can be huge and also this will lead to new challenges of data integrity, availability, and privacy. This direction of work requires innovative ways to encourage data pooling and sharing.

2) Interpretable ML: Despite the state of the art performance of DL techniques, these methods are black-box techniques and lack underlying theory about their learning behavior and thought process. Therefore, their decisions are not interpretable due to which uncertainty quantification of predictions becomes extremely difficult. In addition, the life-critical nature of healthcare applications demands that DL methods' decision should be explainable and interpretable at the same time. It has been argued that interpretable methods enable the extraction of most relevant and important features for the specific tasks for which they are developed [135]. In a recent study, a visualization framework named Deep-Tune is presented that enables neuroscientists to identify patterns that activate a certain neuron in a CNN model that was trained for the task of neural spike rate prediction [136]. The work on developing interpretable ML methods is catching up and is still an open research problem.

3) Secure and Private ML: ML methods in general and DL, in particular, suffers from various security and privacy issues such as adversarial attacks on DL based medical systems [137], poisoning attack [138], and privacy breaches. Such vulnerabilities hinder the smooth, robust, and efficient application of ML and DL systems in actual clinical settings. Privacy is a major challenge in predictive healthcare which is associated with the utilization of patient's data by ML/DL techniques that often contain personal attributes. Techniques used should not reveal any information about the identity of subject patients either at the training or inference phase. To ensure privacy-preserving ML, three types of methods have been used in literature, i.e., cryptographic approaches, differential privacy, and federated learning. On the other hand, various methods to make DL robust against different security attacks (such as adversarial attacks, model stealing, and poisoning attacks, etc.) have been proposed in the literature [139]. However, these methods only work in particular settings (i.e., for a specific type of attack) for which they were developed and fail to defend in different settings (i.e., defense does not generalise to other type of attacks). Therefore, the development of secure and privacy-preserving ML/DL models is still an open research problem.

VI. Conclusion

In this paper, we comprehensively reviewed the available literature and highlighted why early prediction of ES is required, how ML and DL techniques are used for ES prediction. In the context of EEG analysis, feature selection, ES detection, and prediction, and the evaluation of prediction or detection algorithms, ES prediction is a capacious topic. Contrary to the findings of this paper, most of the previous survey papers focused only on EEG analysis and a few of them covered the developments of prediction techniques; while we tried to provide insights by considering the aspects of the feature selection, prediction techniques, and evaluation methodologies, etc. In addition, we have also highlighted future work directions and open research problems that require further investigation.

References

[1] S. N. Rakhade and F. E. Jensen, "Epileptogenesis in the immature brain: Emerging mechanisms," *Nature Rev. Neurol.*, vol. 5, no. 7, pp. 380–391, 2009.

[2] M. J. England, C. T. Liverman, A. M. Schultz, and L. M. Strawbridge, "Epilepsy across the spectrum: Promoting health and understanding.: A summary of the institute of medicine report," *Epilepsy Behav.*, vol. 25, no. 2, pp. 266–276, 2012.

[3] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2006.

[4] D. P. Subha, P. K. Joseph, R. Acharya, and C. M. Lim, "EEG signal analysis: A survey," *J. Med. Syst.*, vol. 34, no. 2, pp. 195–212, 2010.

[5] Y. Song, "A review of developments of EEG-based automatic medical support systems for epilepsy diagnosis and seizure detection," *J. Biomed. Sci. Eng.*, vol. 4, no. 12, pp. 788–796, 2011.

[6] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowl. Syst.*, vol. 45, pp. 147–165, 2013.

[7] K. Gadhoumi, J.-M. Lina, F. Mormann, and J. Gotman, "Seizure prediction for therapeutic devices: A review," *J. Neurosci. Methods*, vol. 260, pp. 270–282, 2016.

[8] M. Iftikhar, S. A. Khan, and A. Hassan, "A survey of deep learning and traditional approaches for EEG signal processing and classification," in *Proc. IEEE 9th Annu. Inf. Technol., Electron. Mobile Commun. Conf.*, 2018, pp. 395–400.

[9] U. R. Acharya, Y. Hagiwara, and H. Adeli, "Automated seizure prediction," *Epilepsy Behav.*, vol. 88, pp. 251–261, 2018.

[10] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction—Ready for a new era," *Nature Rev. Neurol.*, vol. 14, pp. 618–630, 2018.

[11] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, 2019.

[12] G. Li, C. H. Lee, J. J. Jung, Y. C. Youn, and D. Camacho, "Deep learning for EEG data analytics: A survey," *Concurrency Comput.: Pract. Exp.*, 2019, Art. no. e5199.

[13] M. Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*. London, U.K.: Penguin, 2019.

[14] J. Weyhenmeyer and E. A. Gallman," *Rapid Review Neuroscience E-Book*. New York, NY, USA: Elsevier, 2006.

[15] M. D. Lieberman, "Intuition: A social cognitive neuroscience approach," *Psychol. Bull.*, vol. 126, no. 1, pp. 109–137, 2000.

[16] T. J. Sejnowski, C. Koch, and P. S. Churchland, "Computational neuroscience," *Science*, vol. 241, no. 4871, pp. 1299–1306, 1988.

[17] A. Razi and K. J. Friston, "The connected brain: Causality, models, and intrinsic dynamics," *IEEE Signal Process. Mag.*, vol. 33, no. 3, pp. 14–35, May 2016.

[18] F. S. Tyner and J. R. Knott, *Fundamentals of EEG Technology: Basic Concepts and Methods*, vol. 1. Baltimore, MD, USA: Williams & Wilkins, 1983.

[19] K. Fujiwara *et al.*, "Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 6, pp. 1321–1332, Jun. 2016.

[20] S. B. Dumanis, J. A. French, C. Bernard, G. A. Worrell, and B. E. Fureman, "Seizure forecasting from idea to reality. Outcomes of the my seizure gauge epilepsy innovation institute workshop," *eNeuro*, vol. 4, no. 6, 2017.

[21] Y. Liu *et al.*, "Epileptic seizure detection using convolutional neural network: A multi-biosignal study," in *Proc. Australas. Comput. Sci. Week Multiconference*, 2020, pp. 1–8.

[22] M. Valderrama, S. Nikolopoulos, C. Adam, V. Navarro, and M. L. Van Quyen, "Patient-specific seizure prediction using a multi-feature and multi-modal EEG-ECG classification," in *Proc. XII Mediterranean Conf. Med. Biol. Eng. Comput.*, 2010, pp. 77–80.

[23] I. Mporas, V. Tsirka, E. I. Zacharaki, M. Koutroumanidis, M. Richardson, and V. Megalooikonomou, "Seizure detection using EEG and ECG signals for computer-based monitoring, analysis, and management of epileptic patients," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3227–3233, 2015.

[24] U. R. Acharya, S. V. Sree, A. P. C. Alvin, and J. S. Suri, "Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9072–9078, 2012.

[25] L. Sun, Y. Liu, and P. J. Beadle, "Independent component analysis of EEG signals," in *Proc. IEEE Int. Workshop VLSI Des. Video Technol.*, 2005, pp. 219–222.

[26] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 141–144, 2003.

[27] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Appl. Math. Comput.*, vol. 187, no. 2, pp. 1017–1026, 2007.

[28] W. Gersch and J. Yonemoto, "Parametric time series models for multivariate EEG analysis," *Comput. Biomed. Res.*, vol. 10, no. 2, pp. 113–125, 1977.

[29] M. Akin, "Comparison of wavelet transform and FFT methods in the analysis of EEG signals," *J. Med. Syst.*, vol. 26, no. 3, pp. 241–247, 2002.

[30] K. Natarajan, R. Acharya, F. Alias, T. Tiboleng, and S. K. Puthusserypady, "Nonlinear analysis of EEG signals at different mental states," *Biomed. Eng. Online*, vol. 3, no. 1, pp. 1–7, 2004.

[31] W. Chaovalitwongse, L. D. Iasemidis, P. M. Pardalos, P. R. Carney, D.-S. Shiau, and J. C. Sackellares, "Performance of a seizure warning algorithm based on the dynamics of intracranial EEG," *Epilepsy Res.*, vol. 64, no. 3, pp. 93–113, 2005.

[32] M. L. Van Quyen *et al.*, "Preictal state identification by synchronization changes in long-term intracranial EEG recordings," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 559–568, 2005.

[33] S. Kalitzin, D. Velis, P. Suffczynski, J. Parra, and F. L. Da Silva, "Electrical brain-stimulation paradigm for estimating the seizure onset site and the time to ictal transition in temporal lobe epilepsy," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 718–728, 2005.

[34] S. Li, W. Zhou, Q. Yuan, and Y. Liu, "Seizure prediction using spike rate of intracranial EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 6, pp. 880–886, Nov. 2013.

[35] S. S. Viglione and G. O. Walsh, "Proceedings: Epileptic seizure prediction," *Electroencephalography Clin. Neurophysiol.*, vol. 39, no. 4, pp. 435–436, 1975.

[36] Z. Rogowski, I. Gath, and E. Bental, "On the prediction of epileptic seizures," *Biol. Cybern.*, vol. 42, no. 1, pp. 9–15, 1981.

[37] H. H. Lange, J. P. Lieb, J. Engel Jr., and P. H. Crandall, "Temporo-spatial patterns of pre-ictal spike activity in human temporal lobe epilepsy," *Electroencephalography Clin. Neurophysiol.*, vol. 56, no. 6, pp. 543–555, 1983.

[38] J. Gotman and M. G. Marciani, "Electroencephalographic spiking activity, drug levels, and seizure occurence in epileptic patients," *Ann. Neurol.*, vol. 17, no. 6, pp. 597–603, 1985.

[39] L. D. Iasemidis, J. C. Sackellares, H. P. Zaveri, and W. J. Williams, "Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures," *Brain Topography*, vol. 2, no. 3, pp. 187–201, 1990.

[40] J. Martinerie *et al.*, "Epileptic seizures can be anticipated by non-linear analysis," *Nature Medicine*, vol. 4, no. 10, pp. 1173–1176, 1998.

[41] M. L. Van Quyen, J. Martinerie, M. Baulac, and F. Varela, "Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings," *Neuroreport*, vol. 10, no. 10, pp. 2149–2155, 1999.

[42] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, "Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients," *Physica D: Nonlinear Phenom.*, vol. 144, no. 3–4, pp. 358–369, 2000.

[43] V. Navarro *et al.*, "Seizure anticipation in human neocortical partial epilepsy," *Brain*, vol. 125, no. 3, pp. 640–655, 2002.

[44] W. De Clercq, P. Lemmerling, S. Van Huffel, and W. Van Paesschen, "Anticipation of epileptic seizures from standard EEG recordings," *The Lancet*, vol. 361, no. 9361, pp. 970–971, 2003.

[45] M. A. F. Harrison, M. G. Frei, and I. Osorio, "Accumulated energy revisited," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 527–531, 2005.

[46] M. A. F. Harrison, I. Osorio, M. G. Frei, S. Asuri, and Y.-C. Lai, "Correlation dimension and integral do not predict epileptic seizures," *Chaos: Interdisciplinary J. Nonlinear Sci.*, vol. 15, no. 3, 2005, Art. no. 033106.

[47] L. D. Iasemidis *et al.*, "Long-term prospective on-line real-time seizure prediction," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 532–544, 2005.

[48] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky, "Classification of patterns of EEG synchronization for seizure prediction," *Clin. Neurophysiol*, vol. 120, no. 11, pp. 1927–1940, 2009.

[49] Y. Salant, I. Gath, and O. Henriksen, "Prediction of epileptic seizures from two-channel EEG," *Med. Biol. Eng. Comput.*, vol. 36, no. 5, pp. 549–556, 1998.

[50] W. van Drongelen *et al.*, "Seizure anticipation in pediatric epilepsy: Use of Kolmogorov entropy," *Pediatric Neurol.*, vol. 29, no. 3, pp. 207–213, 2003.

[51] K. Lehnertz and B. Litt, "The first international collaborative workshop on seizure prediction: Summary and data description," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 493–505, 2005.

[52] F. Mormann, T. Kreuz, R. G. Andrzejak, P. David, K. Lehnertz, and C. E. Elger, "Epileptic seizures are preceded by a decrease in synchronization," *Epilepsy Res.*, vol. 53, no. 3, pp. 173–185, 2003.

[53] L. Kuhlmann *et al.*, "Epilepsyecosystem.org: Crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG," *Brain*, vol. 141, no. 9, pp. 2619–2630, 2018.

[54] M. I. Maturana *et al.*, "Critical slowing as a biomarker for seizure susceptibility," *Nature Commun.*, vol. 11, 2020, Art. no. 2172.

[55] B. Schelter, J. Timmer, and A. Schulze-Bonhag, Eds., *Seizure Prediction Epilepsy: From Basic Mechanisms to Clinical Application*. Hoboken, NJ, USA: Wiley, 2008.

[56] R. S. Fisher, "The new classification of seizures by the international league against epilepsy 2017," *Curr. Neurol. Neurosci. Rep.*, vol. 17, no. 6, p. 48, 2017.

[57] M. Berendt and L. Gram, "Epilepsy and seizure classification in 63 dogs: a reappraisal of veterinary epilepsy terminology," *J. Veterinary Internal Med.*, vol. 13, no. 1, pp. 14–20, 1999.

[58] W. J. C. Van Elmpt, T. M. E. Nijsen, P. A. M. Griep, and J. B. A. M. Arends, "A model of heart rate changes to detect seizures in severe epilepsy," *Seizure*, vol. 15, no. 6, pp. 366–375, 2006.

[59] M. J. Cook *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.*, vol. 12, no. 6, pp. 563–571, 2013.

[60] D. An, F. Fahoum, J. Hall, A. Olivier, J. Gotman, and F. Dubeau, "Electroencephalography/functional magnetic resonance imaging responses help predict surgical outcome in focal epilepsy," *Epilepsia*, vol. 54, no. 12, pp. 2184–2194, 2013.

[61] S. An *et al.*, "Predicting drug-resistant epilepsy—A machine learning approach based on administrative claims data," *Epilepsy Behav.*, vol. 89, pp. 118–125, 2018.

[62] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Structural Biotechnol. J.*, vol. 13, pp. 8–17, 2015.

[63] T. M. Jørgensen, A. Tycho, M. Mogensen, P. Bjerring, and G. B. E. Jemec, "Machine-learning classification of non-melanoma skin cancers from image features obtained by optical coherence tomography," *Skin Res. Technol.*, vol. 14, no. 3, pp. 364–369, 2008.

[64] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, Alzheimer's Disease Neuroimaging Initiative *et al.*, "Machine learning framework for early mri-based alzheimer's conversion prediction in MCI subjects," *Neuroimage*, vol. 104, pp. 398–412, 2015.

[65] Y. Yang *et al.*, "Epileptic seizure prediction based on permutation entropy," *Frontiers Comput. Neurosci.*, vol. 12, pp. 55, 2018.

[66] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[67] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Netw.*, vol. 21, no. 2–3, pp. 427–436, 2008.

[68] G. Zaharchuk, E. Gong, M. Wintermark, D. Rubin, and C. Langlotz, "Deep learning in neuroradiology," *Amer. J. Neuroradiol.*, vol. 39, no. 10, pp. 1776–1784, 2018.

[69] K. W. Johnson *et al.*, "Artificial intelligence in cardiology," *J. Amer. College Cardiol.*, vol. 71, no. 23, pp. 2668–2679, 2018.

[70] P. Kukharchik, D. Martynov, I. Kheidorov, and O. Kotov, "Vocal fold pathology detection using modified wavelet-like features and support vector machines," in *Proc. IEEE 15th Eur. Signal Process. Conf.*, 2007, pp. 2214–2218.

[71] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," *Nature Reviews Genetics*, vol. 16, no. 6, pp. 321–332, 2015.

[72] G. E. Hinton, "Machine learning for neuroscience," *Neural Syst. Circuits*, vol. 1, no. 1, p. 12, 2011.

[73] M.-A. T. Vu *et al.*, "A shared vision for machine learning in neuroscience," *J. Neurosci.*, vol. 38, no. 7, pp. 1601–1607, 2018.

[74] A. T Drysdale *et al.*, "Resting-state connectivity biomarkers define neurophysiological subtypes of depression," *Nature Medicine*, vol. 23, no. 1, pp. 28–38, 2017.

[75] L. J. O'Donnell, M. Kubicki, M. E. Shenton, M. H. Dreusicke, W. E. L. Grimson, and C.-F. Westin, "A method for clustering white matter fiber tracts," *Amer. J. Neuroradiol.*, vol. 27, no. 5, pp. 1032–1036, 2006.

[76] R. S. Sutton *et al. Introduction to Reinforcement Learning*, vol. 2. Cambridge, MA, USA: MIT Press, 1998.

[77] M. M. N. Mannan, M. A. Kamran, and M. Y. Jeong, "Identification and removal of physiological artifacts from electroencephalogram signals: A review," *IEEE Access*, vol. 6, pp. 30630–30652, 2018.

[78] S. M. Usman, M. Usman, and S. Fong, "Epileptic seizures prediction using machine learning methods," *Comput. Math. Methods Medicine*, vol. 2017, 2017.

[79] S. M. Usman and A. Hassan, "Efficient prediction and classification of epileptic seizures using EEG data based on univariate linear features," *JCP*, vol. 13, no. 6, pp. 616–621, 2018.

[80] L. A. S. Kitano *et al.*, "Epileptic seizure prediction from EEG signals using unsupervised learning and a polling-based decision process," in *Proc. Int. Conf. Artif. Neural Netw.*, 2018, pp. 117–126.

[81] B. Sharif and A. H. Jafari, "Prediction of epileptic seizures from EEG using analysis of ictal rules on poincaré plane," *Comput. Methods Programs Biomedicine*, vol. 145, pp. 11–22, 2017.

[82] O. Stojanović, L. Kuhlmann, and G. Pipa, "Predicting epileptic seizures using nonnegative matrix factorization," *PloS One*, vol. 15, no. 2, 2020, Art. no. e0228025.

[83] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clin. Neurophysiol.*, vol. 126, no. 2, pp. 237–248, 2015.

[84] B. Direito, C. A. Teixeira, F. Sales, M. Castelo-Branco, and A. Dourado, "A realistic seizure prediction study based on multiclass SVM," *Int. J. Neural Syst.*, vol. 27, no. 3, 2017, Art. no. 1750006.

[85] E. B. Assi, L. Gagliano, S. Rihana, D. K. Nguyen, and M. Sawan, "Bispectrum features and multilayer perceptron classifier to enhance seizure prediction," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 15491.

[86] F. Mormann *et al.*, "On the predictability of epileptic seizures," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 569–587, 2005.

[87] K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim, "Analysis of epileptic EEG signals using higher order spectra," *J. Med. Eng. Technol.*, vol. 33, no. 1, pp. 42–50, 2009.

[88] F. Mormann *et al.*, "Automated detection of a preseizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients," *Phys. Rev. E*, vol. 67, no. 2, 2003, Art. no. 021912.

[89] A. A. Bruzzo, B. Gesierich, M. Santi, C. A. Tassinari, N. Birbaumer, and G. Rubboli, "Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients. A preliminary study," *Neurol. Sci.*, vol. 29, no. 1, pp. 3–9, 2008.

[90] X. Li, G. Ouyang, and D. A. Richards, "Predictability analysis of absence seizures with permutation entropy," *Epilepsy Res.*, vol. 77, no. 1, pp. 70–74, 2007.

[91] T. Heskes, "Energy functions for self-organizing maps," in *Kohonen Maps*. New York, NY, USA: Elsevier, 1999, pp 303–315.

[92] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2109–2118, 2017.

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

[94] R. Hussein, M. O. Ahmed, R. Ward, Z. J. Wang, L. Kuhlmann, and Y. Guo, "Human intracranial EEG quantitative analysis and automatic feature learning for epileptic seizure prediction," 2019, *arXiv:1904.03603*.

[95] S. M. Usman, S. Khalid, and M. H. Aslam, "Epileptic seizures prediction using deep learning techniques," *IEEE Access*, vol. 8, pp. 39998–40007, 2020.

[96] R. Jana, S. Bhattacharyya, and S. Das, "Patient-specific seizure prediction using the convolutional neural networks," in *Intell. Enabled Res.* Berlin, Germany: Springer, 2020, pp. 51–60.

[97] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Medicine*, vol. 99, pp. 24–37, 2018.

[98] N. D. Truong, L. Kuhlmann, M. R. Bonyadi, D. Querlioz, L. Zhou, and O. Kavehei, "Epileptic seizure forecasting with generative adversarial networks," *IEEE Access*, vol. 7, pp. 143999–144009, 2019.

[99] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 5, pp. 804–813, 2019.

[100] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[101] J.-G. Lee *et al.*, "Deep learning in medical imaging: General overview," *Korean J. Radiol.*, vol. 18, no. 4, pp. 570–584, 2017.

[102] E. Asgari and M. R. K. Mofrad, "Continuous distributed representation of biological sequences for deep proteomics and genomics," *PloS One*, vol. 10, no. 11, Art. no. e0141287, 2015.
[103] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Mol. Syst. Biol.*, vol. 12, no. 7, 2016.
[104] G. Li, C. Lee, J. Jung, Y. C. Youn, and D. Camacho, "Deep learning for EEG data analytics: A survey," *Concurrency Comput.: Pract. Experience*, vol. 2, 2019.
[105] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.
[106] V. Jain, H. S. Seung, and S. C. Turaga, "Machines that learn to segment images: A crucial technology for connectomics," *Current Opinion Neurobiol.*, vol. 20, no. 5, pp. 653–666, 2010.
[107] I. Kiral-Kornek *et al.*, "Epileptic seizure prediction using big data and deep learning: Toward a mobile system," *EBioMedicine*, vol. 27, pp. 103–111, 2018.
[108] A. Antoniades, L. Spyrou, C. C. Took, and S. Sanei, "Deep learning for epileptic intracranial EEG data," in *Proc. IEEE 26th Int. Workshop Mach. Learn. Signal Process.*, 2016, pp. 1–6.
[109] P. J. Karoly *et al.*, "Interictal spikes and epileptic seizures: Their relationship and underlying rhythmicity," *Brain*, vol. 139, no. 4, pp. 1066–1078, 2016.
[110] A. Abdelhameed and M. Bayoumi, "Semi-supervised deep learning system for epileptic seizures onset prediction," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl.*, 2018, pp. 1186–1191.
[111] I. Osorio, M. G. Frei, and S. B. Wilkinson, "Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset," *Epilepsia*, vol. 39, no. 6, pp. 615–627, 1998.
[112] T. Maiwald, M. Winterhalder, R. Aschenbrenner-Scheibe, H. U. Voss, A. Schulze-Bonhage, and J. Timmer, "Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic," *Physica D: Nonlinear Phenomena*, vol. 194, no. 3-4, pp. 357–368, 2004.
[113] D. E. Snyder, J. Echauz, D. B. Grimes, and B. Litt, "The statistics of a practical seizure warning system," *J. Neural Eng.*, vol. 5, no. 4, pp. 392–401, 2008.
[114] A. S. Zandi, R. Tafreshi, M. Javidan, and G. A. Dumont, "Predicting epileptic seizures in scalp EEG based on a variational Bayesian Gaussian mixture model of zero-crossing intervals," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1401–1413, May 2013.
[115] M. Ghassemi, T. Naumann, P. Schulam, A. L. Beam, and R. Ranganath, "Opportunities in machine learning for healthcare," 2018, *arXiv:1806.00388*.
[116] D. R Freestone, P. J. Karoly, and M. J. Cook, "A forward-looking review of seizure prediction," *Current Opinion Neurol.*, vol. 30, no. 2, pp. 167–173, 2017.
[117] L. Kuhlmann, D. B. Grayden, F. Wendling, and S. J. Schiff, "The role of multiple-scale modelling of epilepsy in seizure forecasting," *J. Clin. Neurophysiol.*, vol. 32, no. 3, pp. 220–226, 2015.
[118] P. J. Karoly *et al.*, "The circadian profile of epilepsy improves seizure forecasting," *Brain*, vol. 140, no. 8, pp. 2169–2182, 2017.
[119] W. C Stacey, "Seizure prediction is possible—Now let's make it practical," *EBioMedicine*, vol. 27, pp. 3–4, 2018.
[120] S. S. Ho, S. F. Berkovic, M. R. Newton, M. C. Austin, W. J. McKay, and P. F. Bladin, "Parietal lobe epilepsy: Clinical features and seizure localization by ictal spect," *Neurology*, vol. 44, no. 12, pp. 2277–2277, 1994.
[121] W. W. Lytton, "Computer modelling of epilepsy," *Nature Rev. Neurosci.*, vol. 9, no. 8, pp. 626–637, 2008.
[122] S. Tewolde, K. Oommen, D. Y. C. Lie, Y. Zhang, and M.-C. Chyu, "Epileptic seizure detection and prediction based on continuous cerebral blood flow monitoring–A review," *J. Healthcare Eng.*, vol. 6, no. 2, pp. 159–178, 2015.
[123] K. P. Lillis, M. A. Kramer, J. Mertz, K. J. Staley, and J. A. White, "Pyramidal cells accumulate chloride at seizure onset," *Neurobiol. Disease*, vol. 47, no. 3, pp. 358–366, 2012.
[124] R. A. J. Masterton, P. W. Carney, D. F. Abbott, and G. D. Jackson, "Absence epilepsy subnetworks revealed by event-related independent components analysis of functional magnetic resonance imaging," *Epilepsia*, vol. 54, no. 5, pp. 801–808, 2013.
[125] D. J. DiLorenzo, K. W. Leyde, and D. Kaplan, "Neural state monitoring in the treatment of epilepsy: Seizure prediction—Conceptualization to first-in-man study," *Brain Sci.*, vol. 9, no. 7, pp. 156–174, 2019.
[126] V. Nagaraj *et al.*, "The future of seizure prediction and intervention: Closing the loop," *J. Clin. Neurophysiol.*, vol. 32, no. 3, pp. 194–206, 2015.
[127] A. S.-Bonhage *et al.*, "Views of patients with epilepsy on seizure prediction devices," *Epilepsy Behav.*, vol. 18, no. 4, pp. 388–396, 2010.
[128] C. Hoppe, M. Feldmann, B. Blachut, R. Surges, C. E. Elger, and C. Helmstaedter, "Novel techniques for automated seizure registration: patients' wants and needs," *Epilepsy & Behavior*, vol. 52, pp. 1–7, 2015.
[129] Y. Wang, X. Long, H. V. Dijk, R. Aarts, and J. Arends, "Adaptive EEG channel selection for nonconvulsive seizure analysis," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process.*, 2018, pp. 1–5.
[130] J. Birjandtalab, M. B. Pouyan, D. Cogan, M. Nourani, and J. Harvey, "Automated seizure detection using limited-channel EEG and nonlinear dimension reduction," *Comput. Biol. Medicine*, vol. 82, pp. 49–58, 2017.
[131] F. C. Viola, S. Debener, J. Thorne, and T. R. Schneider, "Using ICA for the analysis of multi-channel EEG data," *Simultaneous EEG fMRI: Recording, Anal., Appl.: Recording, Anal., Appl.*, pp. 121–133, 2010.
[132] R. Chatterjee, D. Guha, D. K. Sanyal, and S. N. Mohanty, "Discernibility matrix based dimensionality reduction for EEG signal," in *Proc. IEEE Region 10 Conf.*, 2016, pp. 2703–2706.
[133] X. Gao, "Diagnosing abnormal electrocardiogram (ECG) via deep learning," IntechOpen, 2019.
[134] J. Wan *et al.*, "Wearable iot enabled real-time health monitoring system," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 298, 2018.
[135] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Interpretable machine learning: Definitions, methods, and applications," 2019, *arXiv:1901.04592*.
[136] R. Abbasi-Asl *et al.*, "The deeptune framework for modeling and characterizing neurons in visual cortex area v4," *bioRxiv*, 2018, Art. no. 465534.
[137] S. G. Finlayson, J. D. Bowers, J. Ito, J. L. Zittrain, A. L. Beam, and I. S. Kohane, "Adversarial attacks on medical machine learning," *Science*, vol. 363, no. 6433, pp. 1287–1289, 2019.
[138] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. Nita-Rotaru, and B. Li, "Manipulating machine learning: Poisoning attacks and countermeasures for regression learning," in *Proc. IEEE Symp. Secur. Privacy*, 2018, pp. 19–35.
[139] A. Qayyum, M. Usama, J. Qadir, and A. Al-Fuqaha, "Securing connected & autonomous vehicles: Challenges posed by adversarial machine learning and the way forward," *IEEE Commun. Surv. Tut.*, vol. 22, no. 2, pp. 998–1026, 2020.