SURVEY ARTICLE

A Review on Machine Learning Approaches in Identification of Pediatric Epilepsy

Mohammed Imran Basheer Ahmed[1] · Shamsah Alotaibi[2] · Atta-ur-Rahman[2] · Sujata Dash[3] · Majed Nabil[2] · Abdullah Omar AlTurki[2]

Abstract

Epilepsy is the second most common neurological disease after Alzheimer. It is a disorder of the brain which results in recurrent seizures. Though the epilepsy in general is considered as a serious disorder, its effects in children are rather dangerous. It is mainly because it reasons a slower rate of development and a failure to improve certain skills among such children. Seizures are the most common symptom of epilepsy. As a regular medical procedure, the specialists record brain activity using an electroencephalogram (EEG) to observe epileptic seizures. The detection of these seizures is performed by specialists, but the results might not be accurate and depend on the specialist's experience; therefore, automated detection of epileptic pediatric seizures might be an optimal solution. In this regard, several techniques have been investigated in the literature. This research aims to review the approaches to pediatric epilepsy seizures' identification especially those based on machine learning, in addition to the techniques applied on the CHB-MIT scalp EEG database of epileptic pediatric signals.

Keywords Machine learning · Disease identification · Pediatric epilepsy · Pediatric seizures

Introduction

A neurological disease is a disorder in nervous system in the human's body, and it is composed of three parts: brain, human body, and spinal cord. Causes of neurological problems are varied, but they could include a genetic disorder, congenital abnormalities, brain injury, or environmental health problems. The most significant symptom in neurological diseases is chronic pain, which affects 20–40% of patients and reaches 60% in the case of Parkinson's disease [1]. From a medical perspective, diagnosing neurological diseases is challenging, because the nervous system is complex. Many types of neurological diseases exist, and thankfully, all of them have treatments that can allow the patients to live a normal life and lead them to a comfortable level. The second most common disease after Alzheimer's is epilepsy. This is a type of neurological disorder characterized by a disorder of the brain causing recurrent seizures [2]. Epilepsy's history begins in the twentieth century when American Canadian neurologists noted a connection between brain and personality changes in the 1950s. The first antiepileptic, phenytoin, was introduced in that time, and it is still used alongside other drugs [3]. It is thus very important to find this correlation that could potentially result in an early identification and severity of the disease. In this regard, several studies have been conducted for epilepsy identification in the literature. Out of which machine learning (ML)-based approaches are getting popularity due to their promising nature and effectiveness. This study focuses on ML-based approaches in disease identification, especially for epilepsy and further dig down for applications of ML-based approaches in pediatric epilepsy identification. This area is important in many ways; for instance, an earlier identification of the disease in the minors could help in prevention and their better care after. Moreover, it is a long and sequential

✉ Sujata Dash
sujata238dash@gmail.com

1 Department of Computer Engineering, College of Computer Science and Information Technology (CCSIT), Imam Abdulrahman Bin Faisal University (IAU), P.O. Box 1982, Dammam 31441, Saudi Arabia

2 Department of Computer Science, College of Computer Science and Information Technology (CCSIT), Imam Abdulrahman Bin Faisal University (IAU), P.O. Box 1982, Dammam 31441, Saudi Arabia

3 Department of Computer Application, Maharaja Srirama Chandra Bhanj Deo University, Baripada, India

process involving several aspects that can affect the overall classification results. For instance, dataset, the way signals were obtained (health of BCI), the way signals were processed (a range of options available like DWT, IWT etc.), types of filters applied, the way missing details were added, investigating wide range of classifiers and their effectiveness, and finally applying various evaluation metrics such as accuracy, precision, recall, etc. to contrast the performance of various classifiers. Therefore, analyzing the best combination for a specific dataset is really a challenging task. This study aims to uncover such aspects in the pediatric epilepsy identification.

The organization of paper as follows: the background of pediatric epilepsy is given in the section "Background"; the section "Literature Review" contains the comprehensive literature review. Analysis by contrasting outcomes of various techniques is given in the section "Analysis of Feature Extraction Methods of EEG Signals", while the section "Conclusion" concludes the paper.

Background

Epilepsy in Adult and Children

According to the "International League against Epilepsy" (ILAE) in 2014, epilepsy definition is "a transient occurrence of signs and/or symptoms due to abnormal, excessive or synchronous neuronal activity in the brain" [4]. Epilepsy can be diagnosed by three main symptoms: at least two seizures having occurred over 24 h, reflex seizures having occurred twice or more over 10 years, or/and epilepsy syndrome [4]. The most common symptom is called "epileptic seizure" [5], which is a seizure caused by a disrupting episode with the brain's activities, and not all epileptic patients have seizures [6]. Approximately 50 million people over the world have epilepsy [7], or 1% of the population. According to the Saudi Epilepsy Society, 0.654% of people in Saudi Arabia suffer from this disease. Epilepsy affects all ages; the symptoms and signs differ by age group. For example, in newborns, the symptoms involve a lack of oxygen during delivery and/or abnormal brain development, in which the symptoms in infants are brain tumors and/or genetic disorders [2]. Therefore, an accurate diagnosis is important. As childhood is the stage of brain formation, epilepsy occurs within a more dynamic nervous system and might thus be interfering with brain development that could affect the individual in several ways, such as the failure to develop skills, a slower rate of development, and the possibility of losing previously developed skills [8]. Every 4.8/1000 children worldwide have epilepsy. In the US, 0.94/1000 children under 18 have epilepsy [9]. In recent years, epilepsy diagnosis in infants and children has improved, and new methods

to identify epilepsy have become hot topics for researchers. Electroencephalogram (EEG) signals can reflect the state of brain activities with time. It is a complex and non-linear interconnection between a billion neurons [10]. In addition, it is frequently the way to diagnose epilepsy by analyzing the EEG data; in this regard, several studies have focused on aiding epileptic patients to find suitable treatments.

Brain–Computer Interface

Brain–computer interface (BCI) is a fast-growing technology that helps in the medical field in terms of diagnosis. It contains hardware and software that can monitor the brain's activities to diagnose neurological diseases [11]. The BCI translate the brain signals into commands to send it into an output device to make decision. In addition, it can also store the output data [12]. The cycle of BCI can be summarized in five steps, as illustrated in Fig. 1, and details of each step provided afterward. After acquiring, the signal should be processed for further steps, such as feature extraction and classification, and it is eventually demonstrated on the application interface [13]. There are many applications of BCI (e.g., artificial hands using a brain-machine interface, systems to recognize driver drowsiness, brain fingerprint, and EEG) [13]. The challenges that BCI faces are many, especially when it is employed in the real world. The primary challenge is extracting the signals from the brain, which is not easy, because the strength of signals and the rate of data (bandwidth) are low [13], but in general, it is useful for researchers, because it does not require deep knowledge of programming languages. In a neurological disorder, BCI can connect the brain with the devices, and it allows the specialist to observe the brain activities and evaluate the disease.

Interfacing and Signal Acquisition

The first step in EEG signal processing is to remove artifacts. To do so, specific software must be used, such as MATLAB or Python. Then, the feature from the signals data
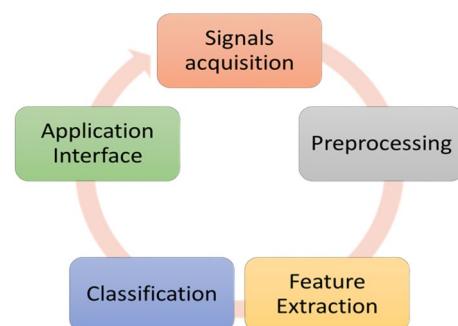

Fig. 1 Cycle of brain–computer interface

Table 1 Frequency bands of EEG signals

| EEG signals	Frequency bands (Hz)
"Delta"	"0.5–4"
"Theta"	"4–8"
"Alpha"	"8–13"
"Beta"	"13–30"
"Gamma"	"> 30"

must be extracted using feature extraction methods, such as wavelet transform. To remove artifacts from signals, FIR filter was applied. The FIR filter consisted of "high-pass" and "low-pass" filters. The "high-pass" filter was used to "allow frequencies higher than the border to pass through it while blocking low frequencies", and the "low-pass" filter was used to allow frequencies lower than the border to pass through it while blocking high frequencies". In this work, the high-pass filter border was 0.5 Hz, and the low-pass filter border was 40 Hz. These boundary frequencies were selected, because 0.5–40 Hz represents the range of the five frequency bands, from delta to gamma (Table 1). However, any frequencies lower than 0.5 Hz and higher than 40 Hz are regarded as noisy signals.

Preprocessing

The second step of EEG data processing is to determine the channel location on the EEG scalp. Determining the location of the channels is significant to plot the EEG scalp map in 2D or 3D or to plot the data component in the brain area [97]. The location channels file is in the location format that should add to an EEG signal in case the file does not have any electrode locations. By knowing the locations of the channels, identifying the epileptic area becomes clear in addition to the types of seizures. To add the channel location file to the EEG data, the number of channels might differ from one patient to another depending about the EEG recording. In this study, the numbers of the channels in each patient could differ; for example, patient number 16 has 28 channels, while patient number 8 has 23 channels.

Fig. 2 EEG electrode measures the signal through the brain surface [15]
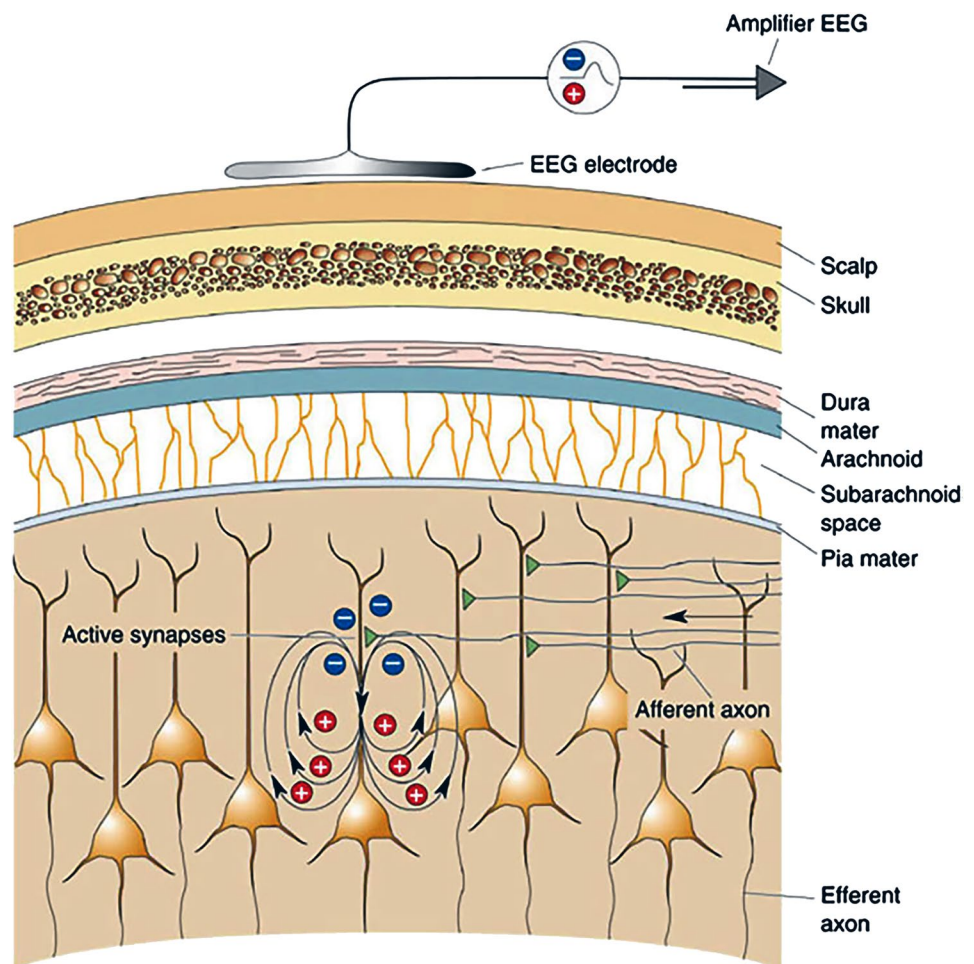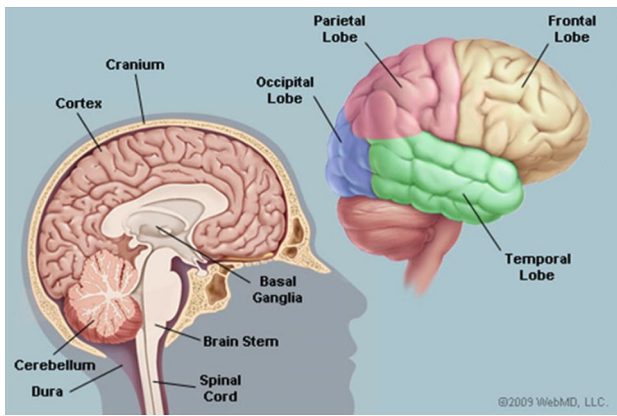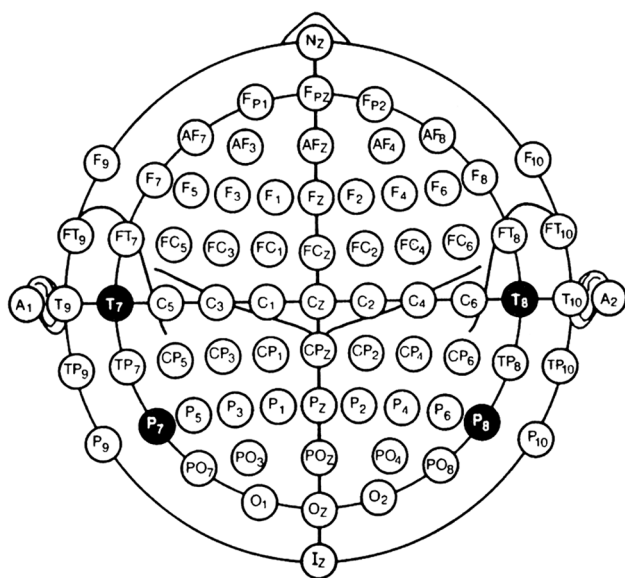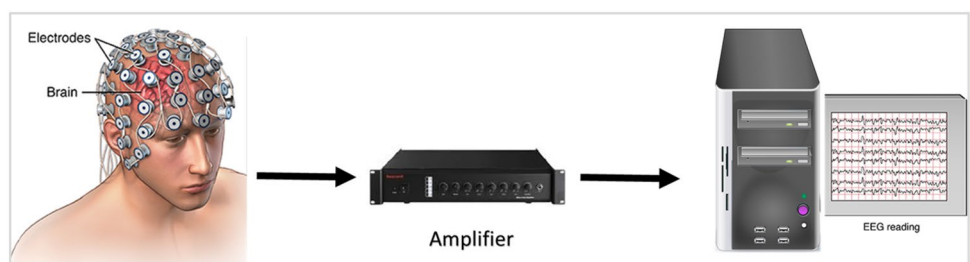
Fig. 3 Brain regions [15]

Fig. 4 Electrode location [16]

Feature Extraction

Since the EEG data are non-stationary signals, using DWT in the "wavelet transform" is a suitable method for an EEG signal, since it captures features in both the "time domain" and "frequency domain". To select a convenient level of decomposition in the Daubechies family, wavelet decomposition splits the original signal into a different band of frequencies called the A's and D's, which are Approximations and Details of coefficient information, respectively. In each stage, two types of coefficients exist: details as high-pass frequencies and approximations as low-pass frequencies, together with the number for the level (e.g., in level one there are D1 and A1). This procedure is repeated on the approximation side until it reaches the low-pass frequency. Statistics over the sets of coefficients are used to decrease the dimensionality of the extracted feature vectors; therefore, statistics features are used to represent the distribution of the time–frequency of the EEG signals. The statistics features selected in this study are "maximum", "minimum", "mean", "median", and "standard deviation" (STD) of the wavelet coefficients in each original single signal. Based on the feature extraction, 9-dimensional feature sets (from D1 to D9 and A9) in the five statistics over the sets are calculated. The average of the statistic features of the wavelet coefficients for each channel is calculated.

Classification

The classification technique is used when the data have an output, and the researcher is attempting to predict. The classification problem appears as when signals were classified as an epileptic pediatric seizure or not. This type of classification is called binary classification, in which there are two types of classes.

EEG Signals

The electroencephalogram (EEG) is a clinical method for recording the brain's activities signals from the brain surface. Based on the state of the brain, the characteristics of the EEG change [14], as presented in Fig. 2 [15]. The recording of high-quality EEG signals depends on the selection of positions of electrodes in the brain surface based on the brain regions. The brain can be divided into four regions based on functionality, as depicted in Fig. 3. These regions are the "occipital lobe", "temporal lobe", "parietal lobe", and "frontal lobe". The occipital region is the back region of the brain and is responsible for vision; the temporal region is on both side of the brain above the ears, and it is responsible

Fig. 5 EEG recording set-up [17]

for the brain skills such as hearing, memory, meaning, and language; the frontal region is up the "temporal lobes", and it is responsible for movement, speech, and emotion. The parietal regions are behind the "frontal lobes" in the top back of the brain and are responsible for the senses, such as pain, touch, and taste [15].

An EEG is recorded by placing electrodes in different positions on the brain surface, and each electrode has a name based on its place on the brain surface, as illustrated in Fig. 4 [16]. This arrangement is called the 10–20 system; it is a relationship between the locations of each electrode measuring the total distances between points on the head [10]. For example, from the bridge of the nose to the lowest point on the back of the head [12], each electrode has a number and letter to identify its location on the head. Here, "F" is frontal region, "T" is temporal region, "P" is parietal region, and "O" is occipital region, while even numbers mean the right half of the head, and odd numbers mean the left half of the head [16].

The equipments needed to record EEG signals are electrodes, amplifiers, and filters, an analog-to-digital converter, and recording hardware and software. Figure 5 illustrates the EEG recording set-up [17].

An EEG signal can be divided into five frequency bands based on rhythm [18]: "delta", "theta", "alpha", "beta", and "gamma". The frequency bands range of "delta" between 0.5 and 4 Hz, "theta" between 4 and 8 Hz, "alpha" between 8 and 13 Hz, "beta" between 13 and 30 Hz, and "gamma" more than 30 Hz (see Table 1). The amplitudes of alpha waves are low; these waves are recorded primarily from the occipital and parietal regions. For beta waves, the amplitudes are low and are typically recorded in the temporal and frontal regions. By contrast, delta waves are high where theta waves are slow–medium, and gamma waves are the fastest frequency bands and could reach up to 100 Hz [14].
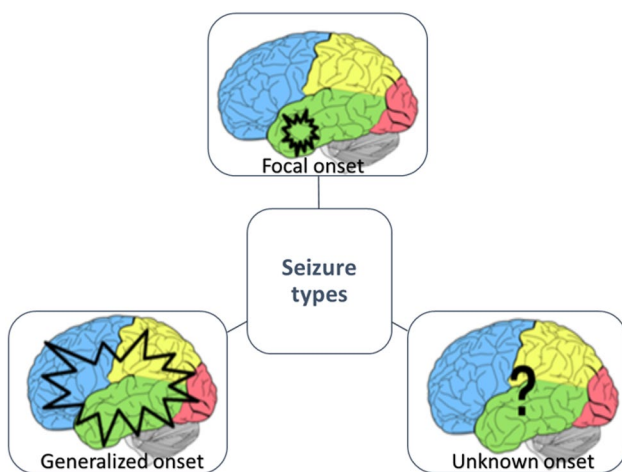
In an actual EEG signal, there is a mixture of these frequencies, and it is a complex signal. Many techniques exist for extracting features from signals; the two most known are the Fourier transform and the wavelet transform.

Classification of Epileptic Seizures

Classifying epileptic seizures is important for researchers to conduct investigation, for clinicians to enhance the treatment and drugs, and for the patient [4]. According to the classification by the ILAE performed in 2017, seizure types are divided into three sections, illustrated in Fig. 6. Based on the area of onset, seizures are classified into three major sections: "focal", "generalized", and "unknown onset". Focal onset is a new term for "partial," and can occur in any region of the brain [13]. Generalized onset occurs when seizures are activated by patient behavior and EEG, and unknown onset refers to an "unknown" onset, where other symptoms are "known," meaning that seizures cannot be diagnosed as having focal or generalization onset [4, 13]. Epilepsy is classified into types as well as a seizure types, creating a fourth type that is a combination of focal and generalized onset [19].

Many techniques have been used to classify epileptic seizures, most of which are reviewed in the subsequent sections. The goal of this study is to review the application of machine learning techniques in epileptic seizures identification. The paper reviews the epilepsy approaches systematically. From general epilepsy identification in adults using AI and ML to specifically children. To highlight the research gaps and shortcoming in terms of dataset availability, signal preprocessing models, and selection of appropriate model.

Literature Review

In the past decade, many studies have published on EEG signal classification for detecting epilepsy and other neurological diseases. In this section, some of the previous studies are discussed, and it is divided into several sections based on area of specialization.

Machine Learning Techniques in Epilepsy

Several studies have applied data mining techniques to detect epilepsy in humans and animals, and they achieved high accuracy. In Jaiswal et al. [20], two techniques are proposed: "sub-pattern" and "cross-sub-pattern correlation-based principal component analysis" ("SpPCA" and "SubXPCA") with the support vector machine algorithm. The authors compared their result with other studies that used different

Fig. 6 Classification of seizure types based on ILAE [13]

algorithms. The results of the proposed techniques achieved the highest accuracy, 100%, and this average was better than that of other techniques, such as "k-nearest neighbor", "support vector machine" (SVM), "decision tree", and "naïve Bayes". The authors in [20] proposed method for detect seizures of epilepsy, the type of the problem is a binary classification, which the EEG signals are classified into two classes: seizure and non-seizure signals. The EEG signals can be classified as seizure and non-seizure using a key point computation-based "local binary pattern" (LBP) [21]. Processing of the EEG signals is divided into three phases, the first phase called "key point localization", the second called "key point-based local binary pattern computation", and the third called histogram feature. Afterward, the "support vector machine" classifier categorizes the signals into (seizure/non-seizure). These methodologies were easy to use and simple to implement; it was achieved good performance compared with existing methods in other studies.

In Tharayil et al., [22], the authors developed a method to predict epileptic seizures in adult and child patients together. This method called linear mixed model; they applied these techniques to more than 1.2 million seizures recorded. The primary discovery was that all developed models achieved higher accuracy in adults than in children. The authors stated the reasons for this, such as that seizure patterns of children and adults are different, or that undercounted seizures in children not being available. The early discover of epileptic seizures can help the patient to avoid any side effects in the brain. In Usman et al. [23], data preprocessing converted 23 EEG signals channels into a single signal channel to improve the "signal-to-noise ratio" (SNR) and then applied "empirical mode decomposition" (EMD) to increase the SNR. The authors selected SVM as a classifier, and the result revealed that the model predicted the seizure prior to its occurrence by 23.6 min, up to a maximum prediction time of 33 min. According to Kabir et al. [24], the proposed approach is performed by dividing the EEG signals into groups based on time period and then drawing a sample from each group of the class via optimum-allocated techniques (OTA) before combining all the samples; afterward, the features are extracted from the OTA set.

The authors in [25] used three techniques in learning: "logistic model trees" (LMTs), "multinomial logistic regression" (MLR), and SVM. The results reveal that the LMT classifier achieved higher accuracy than the others, and that the OAT can be used as an ideal scheme for extracting the features. A few other algorithms were applied in the epilepsy data, such as SVM and neural network, relevance vector machine (RVM), and fractal dimension [26]. In addition, there are random forest, Gaussian mixture models, k-means, and other methods [27]. The SVM algorithm is efficient in the binary classification problem. In research by Sohaib [28], the authors evaluated different machine learning techniques to classify EEG signals. The study extracted four features: "minimum", "maximum", "mean", and "standard deviation" (STD). The result reveals that the SVM classifier was the best classifier for EEG signals, with 85.81%, while the regression tree was second best, with 83.50%. In Patrick et al., [29], applied machine learning techniques were used to predict and detect an epileptic seizure by applying three techniques. The first was the phase-space adjacency spectrum, which used the graph of phase-space adjacency as a biomarker of seizure prediction; this technique achieved 97% accuracy. The second technique was the phase-space Laplacian spectrum, which used the metrics of the phase-space Laplacian spectrum as a biomarker of seizure prediction; this achieved 93% accuracy. The last technique was the hypergraph analysis of phase-space graph, which analyzed a subset of edges of phase graph as a hyperedge in hypergraph and used it as a biomarker for seizure prediction and training accuracy; this technique achieved 93% and 80% for testing accuracy. For seizure detection, the author combined the phase-space analysis method with deep learning using the CNN algorithm to detect the onset of seizures, and this achieved 100% accuracy. An ensemble classifier was used in many studies to classify epileptic seizures.

In Abualsaud et al. [30], the ensemble classifier was applied in incomplete EEG data to examine the efficiency of the classifier. In a comparison with other experiments, the ensemble classifier achieved 90%, while the other experiments achieved 85%, 85.9%, and 89.5%, respectively. In Raghu et al. [31], a new tool called "Computerized Automated Detection of Focal Epileptic Seizure" (CADFES) was introduced. Its primary function was to preprocess EEG data to detect focal and non-focal epileptic seizures. After extracting 28 features from the data set, the authors optimized the number of features using "Neighborhood Component Analysis" (NCA), and then, Four classifiers used to evaluate algorithm performance were SVM, KNN, random forest, and the AdaBoost classifier. The result demonstrated that SVM achieved the best accuracy, with 95.9%. In [32], the authors established a new method to detect an epileptic seizure. They used hybrid SVM by combining a "genetic algorithm" (GA) with "particle swarm optimization" (PSO) to determine the SVM parameters; this model achieved 99.38% accuracy. Lekshmy et al. [84] provided a comparative analysis of the ML methods in epileptic seizure prediction along with their effectiveness. It was concluded that the Random Forest (RF) and long short-term memory (LSTM) algorithms achieved the highest accuracies as 97% and 98%, respectively. Similarly, Nair et al. [85] concluded that the AI-based approaches have been tremendously contributed to epilepsy detection, prediction, and management for an improved healthcare society. Likewise, Natu et al. [86] presented a wholistic view of AI and ML applications in epilepsy detection. It encompasses, data preprocessing approaches, channel selection

Table 2 Summary of machine learning in epilepsy

Publication	Type of data	Algorithms	Accuracy in percentage
Sohaib et al., 2012 [28]	EEG record	"K-Nearest Neighbor" (KNN), "Regression Tree" (RT), "Bayesian Network" (BNT), "Support Vector Machine" (SVM), "Artificial Neural Networks" (ANN)	SVM with 85.81%
Tiwari et al., 2016 [21]	EEG record	"Local Binary Pattern" (LBP) and SVM	98.80%
Kabir and Siuly, 2016 [24]	EEG record	"Logistic model trees" (LMT) "Multinomial Logistic Regression" (MLR), "Support Vector Machine" (SVM)	LMT with 95.33%
Lima et al. 2016 [26]	EEG record	SVM and Relevance Vector Machine (RVM)	SVM with 97%
Tharayil et al., 2017 [22]	EEG record—adults and children	Linear Mixed Model	82% for adults 76% for children
Usman and Usman, 2017 [23]	EEG record	SVM	92.23%
Kumar, 2017 [25]	EEG record	SVM Multilayer Preceptor Neural Network Recurrent of Neural Network Radial Basis Function Neural Network Ensemble of Machine Learning	MPNN and EL with 100%
Jaiswal and Banka, 2018 [20]	EEG record	"Sub-pattern and cross-sub-pattern correlation-based PCA" (SpPCA and SubXPCA) with "support vector machine algorithm"	100%
Patrick and Luckett 2018 [29]	EEG record	For seizure prediction: phase-space adjacency spectrum, phase-space Laplacian spectrum, and hypergraph analysis of phase-space graph For seizure detection: phase-space analysis method with deep learning by using the CNN algorithm	For seizure prediction: phase-space adjacency spectrum with 97% Seizure detection method with 100%
Abualsaud and Mahmuddin, 2015 [30]	EEG record	Ensemble classifier	90%
Raghu and Sriraam, 2018 [31]	EEG record	Computerized Automated Detection of Focal Epileptic Seizure (CADFES)	95.9%
Subasi, 2019 [32]	EEG record	GA and PSO with SVM	99.38%
Lekshmy et al., 2021 [84]	EEG record	ML, DL, RF, LSTM	RF: 97% LSTM: 98%
Nair et al., 2021 [85]	EEG record	kNN and other AI, ML algorithms	–
Natu et al., 2022 [86]	EEG record	AI, ML, and DL algorithms	–
Ilakiyaselvan et al., 2020 [87]	RPS of EEG record	DL	98.5% (binary) 95% (tertiary)
Brari and Belghith, 2021 [88]	EEG record	ML with correlation dimension (CD)	100%

methods followed by the classifier or the prediction model. They further provided the research gap and the limitations in this regard. For instance, the dataset related issues and shortcomings as well as the incurred computation cost of the prediction model. As a remedy, they suggested dataset labeling, deep learning algorithms investigation alongside the appropriate feature selection. In [87], authors proposed a deep learning (DL) approach to seizer detection by investigating the reconstructed phase space (RPS) instead of direct EEG signals that exhibits a chaotic and non-linear behavior make them inadequate for analysis. The approach exhibits accuracies of 98.5% and 95% for binary and tertiary classification, respectively. In [88] authors proposed a novel ML approach to epilepsy prediction using correlation dimension (CD) and achieved 100% accuracy. The proposed model exhibits

a much faster convergence in contrast to similar approach in the literature over the same dataset due to a small set of features and specific combination of subsets. Table 2 presents a summary of the studies investigating AI and ML in epilepsy. It includes, type of dataset used, algorithm investigated and the obtained metrics like accuracy, precision, and recall. It shows that ML remain quite promising in terms of accuracy and other metrics for epilepsy identification.

Machine Learning Techniques in Pediatric Epilepsy

In Shoeb [33] and [34], the researchers applied the SVM algorithm to classify seizure onset in pediatric epilepsy patients. The data set from the Boston children's Hospital, it contains 23 patient's files of EEG recording, and all

patients under 18 years. It consisted of 18 channels that contained 163 seizures and was separated into records of one hour each. The records that did not contain a seizure were called "non-seizure" records, and those that had a seizure were called "seizure" records. For 163 seizures [33], the algorithm achieved 96% [33] and 173 seizures as well [34] with a 3-s delay. In [35], the authors designed an "Advance Seizure Prediction via Preictal Relabeling" (ASPPR) to predict epileptic pediatric seizures; the model achieved 96.30% for predicting seizures between 1 and 6 min, 96.13% for 8 and 13 min, 94% for 14 and 19 min, and 94.2% for 20 and 25 min.

In [36], the authors proposed a technique to extract features from the epileptic pediatric data set from the CHB-MIT EEG database. In their proposed method, EEG signals are mapped into 2D space that leads to texture image and the gray-level image domain. Furthermore, they compared their method with other methods and achieved 97.74% using the SVM classifier with linear kernels. In addition, a study was conducted using the CHB-MIT EEG database [37]. The data mining of EEG signals was analyzed using a time-series approach, which calculated peak points (the lowest point in a part of the signals) and valley points (the highest point in a part of the signals) and then calculated the distances between them. The study consisted of three experiments; the first experiment was conducted of the basis of an optimized setting of one patient file, while the second experiment was performed by identifying common sub-settings to predict the onset of the seizure for one patient seizure file. The third experiment was applied to the setting of the first experiment on three other patients. All the experiments predicted onset seizures early with minimum latency. Identifying seizures in children is different than in adults, because the seizures in children have lack characteristics of EEG features. The study in [38] proposed two linear methods to detect seizures in newborns (39–42 weeks) namely correlation-based feature methods and a relief F method. The feature selector operated in several steps: select subset feature, relevance and redundancy evaluator, and optimized feature vector prior to classification. The average detection achieved 93%. Some studies that focused on newborn data proposed alternative methods for detecting seizures ([39–43], and [44]) and achieved a high-level accuracy compared with previous studies. The number of electrodes may affect pediatric seizure detection. The authors in [42] attempted to decrease the electrode's number in the terms to achieve a high-quality result of detecting seizures; the experiment revealed that 19 electrodes had higher agreement than less than 19. When

Table 3 Summary of machine learning in pediatric epilepsy

| Publication	Type of data	Algorithms	Accuracy in percentage
Aarabi et al. 2006 [38]	Newborn data	"Linear correlation-based feature selection "methods and the relief method	93%
Deburchgraeve et al., 2008 [39]	Newborn data	"Neonatal seizure detection mimicking a human observer reading EEG"	96.3%
Shoeb, 2009 [33], 2010 [34]	Children	SVM	96%
Temko et al. 2011 [40]	Newborn data	"Multi-channel patient-independent neonatal seizure detection system" based on the "Support Vector Machine" (SVM) classifier	96.3%
Temko et al. 2011 [44]	Newborn data	SVM	89%
Cherian et al., 2011 [41]	Newborn data	patients classified into two groups: "mild to moderate" (grades 1–5) and "severe" (grades 6–8) EEG abnormalities	Group 1: 73.7% Group 2: 89.5%
Moghim et al., 2014 [35]	Children	ASPPR (Advance Seizure Prediction via Preictal Relabeling)	95%
Kiranyaz and Ince, 2014 [45]	Children	A "collective network of binary classifiers" (CNBC) using "multi-dimensional particle swarm optimization" (MD PSO)	93%
Samiee and Kiranyaz, 2015 [36]	Children	SVM with linear kernel	97.74%
Mathieson et al., 2016 [43]	Newborn	Seizure detection algorithm (SDA)	70%
Suhani et al. 2018 [37]	Children	Time-series approach	–
Kinney-lang et al., 2019 [46]	Children	Network analysis	85%
Alotaibi et al. 2021 [74]	Children	Ensemble learning approach	100%
Abdelhameed and Bayoumi [89]	Children	DL, 2D deep convolution autoencoder (2D-DCAE) linked to a neural network-based classifier to form a unified system that is trained in a supervised deep convolutional autoencoder (SDCAE) using LSTM	$98.79 \pm 0.53\%$ accuracy, $98.72 \pm 0.77\%$ sensitivity, $98.86 \pm 0.53\%$ specificity, $98.86 \pm 0.53\%$ precision, and F1-score of $98.79 \pm 0.53\%$
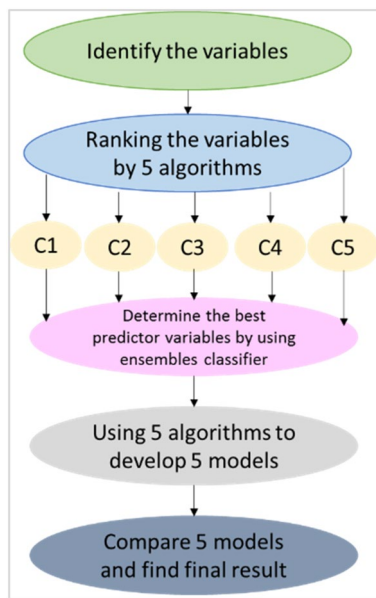
Fig. 7 Flowchart of the proposed techniques in [49]

using a "collective network of binary classifiers" (CNBC) using MD PSO, the proposed method achieved 93% accuracy [45], this method was applied in long-term EEG data and extracted the seizures from the signals data. Table 3 shows a summary of studies that reviewed in this section are presented. The authors in [46] detected cognitive impairment in children with epilepsy using network analysis, and their proposed method achieved 85% accuracy. Authors in [74] investigated ensemble learning approaches to identify epilepsy in the children and achieve 100% accuracy over CHB-MIT children epilepsy dataset. Similarly, authors in [89] proposed a novel deep learning approach to detect epilepsy in children on CHB-MIT dataset. They proposed a 2D deep convolution autoencoder (2D-DCAE) linked to a neural network-based classifier to form a unified system that is trained in a supervised deep convolutional autoencoder (SDCAE) using LSTM. The scheme exhibits $98.79 \pm 0.53\%$ accuracy, $98.72 \pm 0.77\%$ sensitivity, $98.86 \pm 0.53\%$ specificity, $98.86 \pm 0.53\%$ precision, and F1-score of $98.79 \pm 0.53\%$. It can be seen that ML-based techniques are promising in the pediatric epilepsy identification as of Table 3.

Table 4 Summary of machine learning in other diseases

Publication	Type of data	Algorithms	Best accuracy
Perveen and Shahbaz, 2016 [48]	Diabetes data set	J48 tree algorithm using AdaBoost and bagging techniques	AdaBoost with 89%
Tseng et al., 2017 [49]	Ovarian cancer data set	"Ensemble Learning" "Support Vector Machine" (SVM), "C5.0" "Extreme Learning Machine" (ELM) "Multivariate Adaptive Regression Splines" (MARS), "Random Forest" (RF)	C5.0 with 90%
Mohamed and Waguih, 2018 [47]	Giza Egypt institute data set	J48 tree algorithm reducer errors pruning (REP) tree	J48 with 87.64%
Abdar et al.,2018 [50]	Breast cancer	"SV-BayesNet-3-MetaClassifier" and "SV-Naïve Bayes-3-MetaClassifier"	98.07%
Hart et al., 2019[51]	Forest data set	Artificial neural network and random forest	RF:>75%
Dash et al., 2018 [73]	UCI repository	Hybrid chaotic firefly algorithm with kernel based Naïve Bayes (CFA-KNB)	CFA-KNB with KNN~90%
Rahman et.al. [75]	Govt. Hospital data	Supervised Machine Learning	98.4%
Zagrouba et al. [76]	Hospital data	Supervised Machine Learning	96.79%
Ahmed et al. [77], 77	Hospital data	Fuzzy Rule Based System	88.78%
Sujata et al.2019 [79]	PD data from UCI repository	Kernel based chaotic Firefly model	90%
Dash et al., 2017 [80]	PD data from UCI repository	Enhanced chaos-based Firefly model	97.20%
Khan et al., 2020 [92]	Hospital data	SVM for heart disease prediction	93.33%
Rehman et al., 2020 [93]	Hospital data	Deep extreme learning machine for Diabetes Type II	92.8%
Ghazal et al., 2022 [95]	1920 images	Transfer learning	87.1%
Alqudaihi et al., 2021 [96]	Voice data	COVID-19 detection by Cough sound using ML	–
Alsunaidi et al., 2021 [97]	Big data	Big data analytics for COVID-19 detection	–
Alhaidari et al., 2021 [98]	E-triage data	E-triage of COVID-19 patients using e-triage	–

Machine Learning Techniques in Other Fields

In Mohamed et al. [47], the authors applied data mining techniques on data from the Egypt institute in Giza to build a model that helps students to select the track based on their features. Three types of decision tree algorithms were applied: J48, reducer error pruning (REP) tree, and random tree. The results revealed that the J48 tree achieved higher accuracy, with 87.64%. The authors in [48] applied the J48 tree algorithm using AdaBoost and bagging techniques in the Diabetes data set. The results demonstrated that Ada-Boost is superior to bagging. The authors in [49] applied five classifiers (after hit and trial) to data sets and took the average. They used an ensemble classifier to increase the accuracy of detecting the disease, develop five models, and compared the models' performance. These algorithms included support vector machine (SVM), C5.0, extreme learning machine (ELM), multivariate adaptive regression splines (MARS), and random forest (RF). Figure 7 presents various steps of the proposed techniques.

Using more than one classifier in the ensemble technique, the study in [50] achieved 98.07% accuracy for diagnosing breast cancer. The authors proposed two layers of the ensemble model by applying three Meta classifiers rather than single classifiers such as Bayes-Net and Naïve Bayes

to improve the accuracy of detecting breast cancer. Table 4 shows a summary of studies that reviewed in this section are presented. In [51], the authors tested the ability of machine learning techniques to predict which tree in a forest might be damaged during a storm. An "artificial neural network" and "random forest" were used as models; the random forest algorithm achieved better accuracy than NN, with more than 75%. In [75][75], authors proposed supervised machine learning methods for COVID-19 and achieved very high accuracy. Similarly, ML-based approaches have been investigated for classification in other areas [90][90]. Approaches in [92] and [93] investigate ML for heart disease prediction and diabetes type II, respectively.

Survey Studies in Machine Learning Techniques

The authors in [52] divided data mining techniques into two main categories: the predictive and the descriptive [52]. Each category has its own techniques. The predictive category has classification, predication, regression, time-series, and fractal coding. The descriptive category has association rules, summarization, clustering, and sequence discovery. According to the IEEE conference on data mining techniques in December 2006 [53], the top 10 algorithms tested were "AdaBoost", "Apriori", "Bagging", "C4.5",

Table 5 Summary of survey studies in machine learning techniques

Publication	Approach	Conclusion
Orosco and Laciar, 2013 [7]	A "survey of performance and techniques for automatic epilepsy"	The most popular techniques are time analysis and the wavelet technique, and it can be combined
Al-fahoum and Al-fraihat, 2014 [55]	Comparing between feature extractions methods in EEG signals	Each method has positive and negative sides for specific signals
Siuly and Zhang, 2016 [54]	Computer-aided diagnosis (CAD) approach in neurological diseases	Computer-aided diagnosis can help to diagnose all these types of diseases
Settouti and Bechar, 2016 [53]	Top 10 algorithms of data mining techniques	The best result of experimental models determined no one was better than another
Lashari et al., 2018 [52]	Investigates the existing practices of medical data classification based on data mining techniques	Medical data mining contributes to business intelligence, which is useful for diagnosing diseases
Kadhim, 2019 [56]	ML in Text Mining	An effective method for classifying a text is to combine related information into the classification process
Dash et al., 2019 [81]	A Modified Firefly-based Meta-Search Algorithm for feature selection: A Predictive Model for Medical Data	An efficient nature inspired hybrid model for selecting predictive features from medical database
Dash et al., 2021 [82]	Intelligent Computing on Time-Series Data Analysis and Prediction of Covid-19 Pandemics	Prophet Model provides understanding of the number of people affected daily by this disease and has achieved around 85% MAPE for all the six countries and the six states of India
Dash et al., 2021 [83]	BIFM: Big-Data Driven Intelligent Forecasting Model for COVID-19	The best ARIMA models are used for predicting the daily-confirmed cases for 90 days future values of six worst-hit countries of the world and six high incidence states of India. The goodness-of-fit measures for the model achieved 85% MAPE for all the countries and all states of India

"CART", "EM"," k-means", "KNN", "NB", and "SVM". No one algorithm was better than another; the best result of the experimental models determined which achieved the best accuracy. In Siuly et al. [54], the authors presented the survey in neurological diseases diagnosis, in which they compared between the most important computer-aided diagnosis (CAD) approaches to neurological diseases. They reviewed many articles on epilepsy and applied the classification methods in EEG signal classification, such as RBFNN, "support vector machine", "k-nearest neighbor", "artificial neural network", and "multiclass least square support vector machine" (MLS-SVM), among other methods. All these studies proved that computer-aided diagnosis can help to diagnose all these types of diseases. Many processing techniques exist for processing EEG signals before applying machine learning techniques. The most popular technique is time analysis, and it consists of two sections: "time domain" and "frequency domain". The wavelet technique is an analysis of signals into sub-signals that contain activity at different times [7], and it can be a combination of two or more techniques. The primary challenge in signal processing is selecting analysis methods; there are many, and each has pros and cons [55]. The most two methods are the Fourier transform, the wavelet transform, and the eigenvector have characteristics that help researchers to select a suitable method for their research. Table 5 presents a summary of the studies that reviewed in this section. The paper in [56] surveyed various supervised machine learning techniques of text classification. The results revealed that the most effective method for classifying a text is to combine related information into the classification process, because it enhances the result quality of the classification.

Analysis of Feature Extraction Methods of EEG Signals

An EEG signals has many features; therefore, many studies have proposed techniques to extract this feature in terms to classify them. The authors in [57] used various methods to extract features: "time analysis", "frequency analysis", "time–frequency analysis", and "time–frequency space analysis". The features were classified using an artificial neuronal network algorithm, and 99% accuracy was achieved between two tasks and 96% between three tasks. Another study described how to extract features from EEG signals using the fast Fourier transform filter [58], a short-time Fourier transform [59].

The authors in [60] used two methods for extract features from EEG signals. The first method called "local neighbor descriptive pattern" (LNDP), and the second was "one-dimensional local gradient pattern" (1D-LGP), which achieved an accuracy of 99.82% and 99.80%, respectively.

The authors in [61] described the wavelet transform filter to extract features, which achieved 98%. The first step in terms of classifying EEG signals is reprocessing. Many approaches are available to do so, each of which has different results. In [62], the authors proposed an interesting approach for removing artifact signals called "fully automated statistical thresholding for EEG artifact rejection" (FASTER). This approach divides EEG signals into five aspects: channels, single-channel, epochs, single epoch, ICs, and aggregated data, each of which were calculated data of statistical parameters and defined the metrics. This method achieved more than 90% accuracy. Furthermore, the authors in [63] proposed a successful method for removing artifacts from signals. The method is known as ADJUST; it is automatic algorithm which identifies artifact components by combined temporal features with stereotype artifact-specific spatial and achieved 95.2% accuracy.

Electroencephalogram signals have noise signals such as eye movements called electrooculography (EOG), and to process EEG signals and achieve quality results in classification, a noise should be removed from signals. The authors in [64] proposed automated methods to removed artifacts from EEG signals by implementing an open-source software library called BioSig for processing a signal, which reduces artifact by 80%. The contribution of the authors in [65] was to achieve clean EEG signal data from the artifact, and these data were made available for the public to allow the researchers to manually add an artifact for their purposes. A method called "independent component analysis" (ICA) combined with dipole was applied in [66]. The advantages of ICA in terms of removed artifacts are that it is efficient; the technique can separate EEG signals and artifacts without needing references. According to a review the authors presented in [18], no perfect solution exists for cleaning EEG signals from the artifacts, and some of the methods have an application, while others do not, according to the comparison. Many studies have proposed methods to reduce the noisy signal from EEG signal, such as combining the wavelet transform with an adaptive thresholding mechanism [67] or independent component analysis [68]. In terms of using ICA to remove artifact signals from EEG, the study in [69] reveals that a good selection of variables in ICA results in enhanced EEG signals and reduces artifacts as accurately as possible for these variables. The advantage of using the ICA is the ability to examine information directly in the data [70]. Although it is an efficient method for rejecting artifacts, it has disadvantages; one of which is that ICA is a complex algorithm that requires suitable data to address [71]. The authors in [72] proposed new methods by combining two techniques, namely the "information in frequency domain" technique and the "information gain" technique. After extracting the features, SVM is used to classify the EEG. These proposed methods achieved 95.62%

Table 6 Summary of feature extraction methods of EEG signals

Publication	Analysis methods	Features extraction methods	Results
Schlo, 2007 [64]	BioSig—open-source library	–	80%
Suleiman and Toka, 2011 [57]	–	"Time analysis", "frequency analysis", "time–frequency analysis", and "time–frequency space analysis"	99%—two tasks 96%—three tasks
Mognon et al., 2011[63]	ADJUST method	–	95.2%
Yong and Fatourechi, 2012 [67]	Wavelet transform with adaptive thresholding mechanism	–	73.1%
Manajemen et al., 2013 [62]	FASTER: "fully automated statistical thresholding for EEG artifact rejection"	–	90%
Ullah et al., 2015 [61]	–	Wavelet transform	98%
Jaiswal and Banka, 2017 [60]	–	LNDP	99.82%
	–	1D-LGP	99.80%
Ratham et al., 2019 [72]	–	frequency domain and information gain techniques	95.62%
Geng et al., 2022 [94]	Independent component analysis (ICA), wavelet transform (WT), common spatial pattern (CSP)	Bayesian Linear Discriminative Analysis (BLDA), SVM, LDA, Bagging Tree (BT)	BLDA: 87.42% SVM: 81.75% BT: 79.42 LDA: 78.41%

accuracy. A technique in [94] proposed an improved feature extraction algorithm for EEG signals obtained from motor imagery BCI. They investigated ICA, WT, and common spatial pattern (CSP) jointly. For analysis purposes, they investigated Bayesian Linear Discriminative Analysis (BLDA), SVM, LDA, and Bagging Tree (BT). The accuracy obtained as BLDA: 87.42%, SVM: 81.75%, BT: 79.42, and LDA: 78.41%, respectively. Table 6 presents a summary of the studies reviewed in this section.

Research Findings

The undergoing study reviews the epilepsy detection, prediction, and identification approaches using AI, ML, and DL, systematically. From general epilepsy identification in adults using AI and ML to specifically children. To highlight the research shortcoming in terms of dataset availability, signal preprocessing models, and selection of appropriate model. Following research gap has been identified with suggestions to overcome in the future studies.

1. *Dataset* the dataset availability is big challenge in epilepsy in general and for children in specific. This has many aspects: first, the time period and frequency of the seizer occurrence and its proper recording. This plays a critical role in disease prediction and varies for patient to patient and the way data were recorded. Second, the equipment used can also be another issue. Third, there is a need to combine, normalize, and process multiple practices datasets around the globe to have a holistic, comprehensive, and uniform dataset.

2. *Dataset preprocessing* the preprocessing methods play a crucial role in the accuracy of prediction and success of the model. There are various methods with their own pros and cons. There is a need to find out the best preprocessing approach for the unified and normalized dataset concept state in point 1.

3. *Feature selection method* there are several feature selection methods from EEG signals in the literature with their own strengths and weaknesses. An investigation of such methods for the holistic dataset (as state in points 1 and 2) is much needed for the better understanding the nature of the disease in a particular age group. It is worth mentioning that there is still much attention needed to improve the accuracy of feature selection methods. Moreover, this selection plays and crucial role in seizer prediction.

4. *Selection of the model* instead of having a single model or method, an ensemble of ML and DL method should be investigated in contrast to the learning methods and aligned to the points 1–3. It is better to come up with a framework in this regard, which should comprehend all the aspects from dataset selection to the evaluation.

Conclusion

This paper presents a comprehensive review on machine learning (ML) techniques for identification of pediatric epilepsy, comprising of last decade (2009–2022) especially in children. The techniques are further categorized in terms of approaches for general disease identification, epilepsy identification in general and finally pediatric

epilepsy identification. It further tabulates and classifies the results chronologically, based on the schemes/approaches/algorithms investigated, datasets, and highest benchmark achieved. That is in terms of accuracy mainly and/or precision, F-score, and recall rate. Moreover, this study segregates the analysis and feature extraction approaches investigated on electroencephalogram (EEG) signals, specifically. It further elaborates that one scheme may be better in way, while another be in another way. By summarizing this systematic literature review, the researcher, scholars, and academician in the field of applied machine learning for disease identification can have a clear overview about the total work done in the field of epilepsy identification in general and pediatric epilepsy identification in specific. Furthermore, research gaps have been identified and potential solution is suggested for the researchers as their future work.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Borsook D. Neurological diseases and pain. Brain. 2012;135(2011):320–44.
2. Gavvala JR, Schuele SU. Epilepsy. JAMA. 2016;316(24):2686. https://doi.org/10.1001/jama.2016.18623.
3. Jackson JH. Reproduced with permission of copyright owner. Further reproduction prohibited without permission. 2015.
4. Falco-walter JJ, Sche IE, Fisher RS. The new definition and classification of seizures and epilepsy. Epilep Res. 2018;139(17):73–9.
5. World Health Organization, NEUROLOGICAL DISORDERS public health challenges, 2006th ed. WHO Library Cataloguing-in-Publication Data.
6. Fox R, Heartshorne R, Kobylecki C, Murphy C. An unusual cause of seizures. 2015. pp. 145–147.
7. Orosco L, Laciar E. Review: a survey of performance and techniques for automatic epilepsy detection, no. September 2016, 2013.
8. Smith ML. Neuropsychology in epilepsy: children are not small adults. Epilepsia. 2010;51(SUPPL. 1):68–9.
9. Sharma P, Hussain A, Greenwood R. Precision in pediatric epilepsy [version 1; referees: 2 approved]. Refer Status. 2019;8:1–15.
10. Wen T, Zhang Z. Effective and extensible feature extraction method using genetic algorithm-based frequency-domain feature search for epileptic EEG multi-classification. Medicine. 2017;96:1–17.
11. Ramadan RA, Vasilakos AV. Brain computer interface control signals review. Neurocomputing. 2017;223(Ocotober 2016):26–44.
12. Jerry J, Dean J, Jonathan R. Brain-computer interfaces in medicine. Oxford: Elsevier; 2012.
13. Baliyan V, Das CJ, Sharma R, Gupta AK. Diffusion weighted imaging: Technique and applications. World J Radiol. 2016;8(9):785–98. https://doi.org/10.4329/wjr.v8.i9.785 PMID: 27721941; PMCID: PMC5039674.
14. Sarma P, Tripathi P, Sarma MP, Sarma KK. Pre-processing and feature extraction techniques for EEG-BCI applications—a review of recent research. ADBU J Eng Technol. 2016;5(0051604):1–8.
15. Yanchun Zhang A. EEG signal analysis and classification techniques and applications. Berlin: Springer; 2016.
16. Zhou J. EEG data analysis, feature extraction and classifiers. All Theses. 1075. 2011. https://tigerprints.clemson.edu/all_theses/1075.
17. Suhani Shrivastava BE. Detecting the onset of an epileptic seizure using a novel time-series approach. 2018.
18. Islam K, Rastegarnia A, Yang Z. Methods for artifact detection and removal from scalp EEG: a review Les méthodes de détection et de rejet d ' artefact de l ' EEG de. Neurophysiol Clin/Clin Neurophysiol. 2016;46(4–5):287–305.
19. Hospital LN, Delhi N. ILAE classification of seizures and epilepsies: an update for the pediatrician. Indian Pediatr. 2019;56(1):60–2.
20. Jaiswal AK, Banka H. Epileptic seizure detection in EEG signal using machine learning techniques. Australas Phys Eng Sci Med. 2018;41(1):81–94.
21. Tiwari A, Pachori RB, Kanhangad V, Panigrahi B. Automated diagnosis of epilepsy using key-point based local binary pattern of EEG signals. IEEE J Biomed Heal Informat. 2016;21(99):1.
22. Tharayil JJ, Chiang S, Moss R, Stern JM, Theodore WH, Goldenholz DM. A big data approach to the development of mixed-effects models for seizure count data. Epilepsia. 2017;58(5):835–44.
23. Usman SM, Usman M, Fong S. Epileptic seizures prediction using machine learning methods. Comput Math Methods Med. 2017;2017:9074759. https://doi.org/10.1155/2017/9074759. Epub 2017 Dec 19. PMID: 29410700; PMCID: PMC5749318.
24. Kabir E, Zhang Y. Epileptic seizure detection from EEG signals using logistic model trees. Brain Inform. 2016;3(2):93–100.
25. Kumar SS. Weighted majority voting-based ensemble of classifiers using different machine learning techniques for classification of EEG signal to detect epileptic seizure. Informatica. 2017;41:99.
26. Lima CAM, Coelho ALV, Madeo RCB, Peres SM. Classification of electromyography signals using relevance vector machines and fractal dimension. Neural Comput Appl. 2016;27(3):791–804.
27. Senders JT, et al. An introduction and overview of machine learning in neurosurgical care. Acta Neurochir (Wien). 2018;160(1):29–38.
28. Sohaib AT, Qureshi S. An empirical study of machine learning techniques for classifying emotional states from EEG data. Master's Thesis Computer Science, School of Computing Blekinge Institute of Technology, Sweden; 2012.
29. Patrick H, Luckett BS. Nonlinear methods for detection and prediction of epileptic seizures. A Dissertation submitted in University of South Alabama, July, 2018.
30. Abualsaud K, Mahmuddin M, Saleh M, Mohamed A. Ensemble classifier for epileptic seizure detection for imperfect EEG data. Sci World J. 2015;2015: https://doi.org/10.1155/2015/945689.
31. Raghu S, Sriraam N. Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms. Expert Syst Appl. 2018;113:18–32.
32. Subasi A. Epileptic seizure detection using hybrid machine learning methods. Neural Comput Appl. 2019;31:317–25.
33. SHOEB AH. Application of machine learning to epileptic seizure onset detection and treatment. Ph.D. Thesis, Harvard University; 2009. p. 1–162.

34. SHOEB A, GUTTAG J. Application of machine learning to epileptic seizure detection. Proceedings of the 27th international conference on machine learning (ICML-10). 2010. p. 975–982.

35. Moghim N, Corne DW. Predicting epileptic seizures in advance. PLoS ONE. 2014;9(6):e99334.

36. Samiee K, Kiranyaz S, Gabbouj M, Saramäki T. Expert Systems with Applications Long-term epileptic EEG classification via 2D mapping and textural features. Expert Syst Appl. 2015;42(20):7175–85.

37. "Chb-mit." [Online]. http://www.physionet.org/physiobank/database/chbmit/. Accessed 12 Jun 2019.

38. Aarabi A, Wallois F, Grebe R. Automated neonatal seizure detection: a multistage classification system through feature selection based on relevance and redundancy analysis. Clin Neurophysiol. 2006;117(2):328–40.

39. Deburchgraeve W, et al. Automated neonatal seizure detection mimicking a human observer reading EEG. Clin Neurophysiol. 2008;119(11):2447–54.

40. Temko A, Thomas E, Marnane W, Lightbody G, Boylan GB. Performance assessment for EEG-based neonatal seizure detectors. Clin Neurophysiol. 2011;122(3):474–82.

41. Cherian PJ, et al. Validation of a new automated neonatal seizure detection system: a clinician' s perspective. Clin Neurophysiol. 2011;122(8):1490–9.

42. Stevenson NJ, Lauronen L, Vanhatalo S. The effect of reducing EEG electrode number on the visual interpretation of the human expert for neonatal seizure detection. Clin Neurophysiol. 2018;129(1):265–70.

43. Mathieson SR, et al. Validation of an automated seizure detection algorithm for term neonates. Clin Neurophysiol. 2016;127(1):156–68.

44. Temko A, Thomas E, Marnane W, Lightbody G, Boylan G. EEG-based neonatal seizure detection with support vector machines. Clin Neurophysiol. 2011;122(3):464–73.

45. Kiranyaz S, Ince T, Zabihi M, Ince D. Automated patient-specific classification of long-term electroencephalography. J Biomed Inform. 2014;49:16–31.

46. Kinney-lang E, et al. Analysis of EEG networks and their correlation with cognitive impairment in preschool children with epilepsy. Epilepsy Behav. 2019;90:45–56.

47. Mohamed MH, Waguih HM. A proposed academic advisor model based on data mining classification techniques. Int J Adv Comput Res. 2018;8(36):129–36.

48. Perveen S, Shahbaz M, Guergachi A, Keshavjee K. Performance analysis of data mining classification techniques to predict diabetes. Procedia Comput Sci. 2016;82(March):115–21.

49. Tseng CJ, Lu CJ, Chang CC, Den Chen G, Cheewakriangkrai C. Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence. Artif Intell Med. 2017;78:47–54.

50. Abdar M, et al. A new nested ensemble technique for automated diagnosis of breast cancer. Pattern Recogn Lett. 2018;132:123–31.

51. Hart E, Sim K, Kamimura K, Meredieu C, Guyon D. Use of machine learning techniques to model wind damage to forests. Agric For Meteorol. 2019;265(October 2018):16–29.

52. Lashari SA, Ibrahim R, Senan N, Taujuddin NSAM. Application of data mining techniques for medical data classification: a review. MATEC Web Conf. 2018;06003:1–6.

53. Settouti N, Bechar MEA, Chikh MA. Statistical comparisons of the top 10 algorithms in data mining for classification task. Int J Interact Multimed Artif Intell. 2016;4(1):46.

54. Siuly S, Zhang Y. Medical big data: neurological diseases diagnosis through medical data analysis. Data Sci Eng. 2016;1(2):54–64.

55. Al-fahoum AS, Al-fraihat AA. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. ISRN Neurosci. 2014;2014:1–7.

56. Zaman G, Mahdin H, Hussain K, Rahman A, Abawajy J, Mostafa SA. An ontological framework for information extraction from diverse scientific sources. IEEE Access. 2021;9:42111–24. https://doi.org/10.1109/ACCESS.2021.3063181.

57. Suleiman AB, Fatehi TA. Features extraction techniques of EEG signal for BCI applications. Iraq: Faculty of Computer and Information Engineering Department College of Electronics Engineering, University of Mosul; 2007.

58. Shakshi RJ, Jaswal R. Brain wave classification and feature extraction of EEG signal by using FFT on lab view. Int Res J Eng Technol. 2016;3(7):1208–12.

59. Zabidi A, Mansor W, Lee YK, Che Wan Fadzal CWNF. Short-time Fourier Transform analysis of EEG signal generated during imagined writing. In: Proc. 2012 Int. Conf. Syst. Eng. Technol. ICSET 2012, no. 2. 2012. pp. 1–4.

60. Jaiswal AK, Banka H. Local pattern transformation-based feature extraction techniques for classification of epileptic EEG signals. Biomed Signal Process Control. 2017;34:81–92.

61. Ullah H, Aamir A, Malik S, Fayyaz R. Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. Australas Phys Eng Sci Med. 2015;38(1):139–49.

62. Manajemen F, Puskesmas DI, Kabupaten L. FASTER: fully automated statistical thresholding for EEG artifact rejection. J Neurosci Methods. 2013;192(1):1–10.

63. Mognon A, Jovicich J, Bruzzone L, Buiatti M. ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology. 2011;48(2):229–40.

64. Schlo A. A fully automated correction method of EOG artifacts in EEG recordings. Clin Neurophysiol. 2007;118:98–104.

65. Klados MA, Bamidis PD. A semi-simulated EEG/EOG dataset for the comparison of EOG artifact rejection techniques. Data Br. 2016;8:1004–6.

66. Zhou W, Gotman J. Automatic removal of eye movement artifacts from the EEG using ICA and the dipole model. Prog Nat Sci. 2009;19(9):1165–70.

67. Yong X, Fatourechi M, Ward RK, Birch GE. Automatic artefact removal in a self-paced hybrid brain- computer interface system. J Neuroeng Rehab. 2012;9:1–21.

68. Burger C, Van Den Heever DJ. Removal of EOG artefacts by combining wavelet neural network and independent component analysis. Biomed Signal Process Control. 2015;15:67–79.

69. Pontifex MB, Gwizdala KL, Parks AC, Billinger M. Variability of ICA decomposition may impact EEG signals when used to remove eyeblink artifacts. Psychophysiology. 2017;54:386–98.

70. Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage. 2007;34:1443–9.

71. Hoffmann S, Falkenstein M. The correction of eye blink artefacts in the EEG: a comparison of two prominent methods. PLoS ONE. 2008;3(8):e3004.

72. Ratham H, Ghayab A, Li Y, Siuly S, Abdulla S. Epileptic seizures detection in EEGs blending frequency domain with information gain technique. Soft Comput. 2019;23(1):227–39.

73. Dash S, Abraham A, Rahman A. Kernel based chaotic firefly algorithm for diagnosing Parkinson's disease. In: 18th International conference on hybrid intelligent systems (HIS), advances in intelligent systems and computing (AISC), vol. 923. Springer, Cham. 2019. pp. 176–188.

74. Alotaibi SM, Rahman A, Basheer MI, Khan MA. Ensemble machine learning based identification of pediatric epilepsy. Comput Mater Continua. 2021;68(1):149–65.

75. Rahman A, Sultan K, Naseer I, Majeed R, Musleh D, et al. Supervised machine learning-based prediction of COVID-19. Comput Mater Continua. 2021;69(1):21–34. https://doi.org/10.32604/cmc.2021.013453.

76. Zagrouba R, Khan MA, Rahman A, Saleem MA, Mushtaq MF, et al. Modelling and simulation of covid-19 outbreak prediction using supervised machine learning. Comput Mater Continua. 2021;66(3):2397–407.

77. Ahmed MIB, Rahman A, Farooqui M, Alamoudi F, Baageel R, Alqarni A. Early identification of COVID-19 using dynamic fuzzy rule based system. Math Model Eng Probl. 2021;8(5):805–12.

78. Naqvi RA, Mushtaq MF, Mian NA, Khan MA, Rahman A, et al. Coronavirus: a "mild" virus turned deadly infection. Comput Mater Continua. 2021;67(2):2631–46.

79. Dash S, Abraham A, Luhach A, Mizera-Pietraszko J, Rodrigues JJPC. Hybrid chaotic firefly decision-making model for Parkinson's disease diagnosis. Int J Distrib Sens Netw. 2019;16(1):1–18.

80. Dash S, Thulasiram R, Thulasiram P. An enhanced chaos-based firefly model for Parkinson's disease diagnosis and classification. In: IEEE ICIT conference. 2017, IEEE Xplore. pp. 159–164. https://doi.org/10.1109/ICIT.2017.43.

81. Dash S, Thulasiram R, Thulasiram P. A modified firefly based meta-search algorithm for feature selection: a predictive model for medical data. IJSIR. 2019;10(2):2.

82. Dash S, Dash CS, Chakraborty C, Giri SK, Pani SK. Intelligent computing on time-series data analysis and prediction of Covid-19 pandemics. Pattern Recogn Lett. 2021;151:69–75. https://doi.org/10.1016/j.patrec.2021.07.027.

83. Dash S, Chakraborty C, Giri SK, Pani SK, Frnda J. BIFM: big-data driven intelligent forecasting model for COVID-19. IEEE Access. 2021;9:97505–17. https://doi.org/10.1109/ACCESS.2021.3094658.

84. Lekshmy HO, Panickar D, Harikumar S. Comparative analysis of multiple machine learning algorithms for epileptic seizure prediction. J Phys Conf Ser. 2022. https://doi.org/10.1088/1742-6596/2161/1/012055.

85. Nair PP, Aghoram R, Khilari ML. Applications of artificial intelligence in epilepsy. Int J Adv Med Health Res. 2021;8(2):41–8.

86. Natu M, Bachute M, Gite S, Kotecha K, Vidyarthi A. Review on epileptic seizure prediction: machine learning and deep learning approaches. Comput Math Methods Med. 2022;2022:17. https://doi.org/10.1155/2022/7751263.

87. Ilakiyaselvan N, Khan AN, Shahina A. Deep learning approach to detect seizure using reconstructed phase space images. J Biomed Res. 2020;34(3):240–50. https://doi.org/10.7555/JBR.34.20190043.

88. Brari Z, Belghith S. A novel Machine Learning approach for epilepsy diagnosis using EEG signals based on Correlation Dimension. IFAC-PapersOnLine. 2021;54(17):7–11.

89. Abdelhameed A, Bayoumi M. A deep learning approach for automatic seizure detection in children with epilepsy. Front Comput Neurosci. 2021;15(1):1–15. https://doi.org/10.3389/fncom.2021.650050.

90. Ibrahim NM, Gabr DGI, Rahman A, Dash S, Nayyar A. A deep learning approach to intelligent fruit identification and family classification. Multimedia Tools Appl. 2022. https://doi.org/10.1007/s11042-022-12942-9.

91. Gollapalli M, Rahman A, Musleh D, Ibrahim N, et al. A neuro-fuzzy approach to road traffic congestion prediction. Comput Mater Continua. 2022;72(3):295–310.

92. Khan MA, Abbas S, Atta A, Ditta A, Alquhayz H, Khan MF, Rahman A, Naqvi RA. Intelligent cloud based heart disease prediction system empowered with supervised machine learning. Comput Mater Continua. 2020;65(1):139–51.

93. Rehman A, Athar A, Khan MA, Abbas S, Rahman A, Saeed A. Modelling, simulation, and optimization of diabetes type II prediction using deep extreme learning machine. J Ambient Intell Smart Environ. 2020;12(2):125–38.

94. Geng X, Li D, Chen H, Yu P, Yan H, Yue M. An improved feature extraction algorithms of EEG signals based on motor imagery brain-computer interface. Alex Eng J. 2022;61(6):4807–20.

95. Ghazal TM, AlHamadi H, Nasir MU, Rahman A, Gollapalli M, Zubair M, Khan MA, Yeun CY. Supervised machine learning empowered multifactorial genetic inheritance disorder prediction. Comput Intell Neurosci. 2022;2022:10. https://doi.org/10.1155/2022/1051388.

96. Alqudaihi KS, et al. Cough sound detection and diagnosis using artificial intelligence techniques: challenges and opportunities. IEEE Access. 2021;9:102327–44. https://doi.org/10.1109/ACCESS.2021.3097559.

97. Rahman A, Alqahtani A, Aldhafferi N, Nasir MU, Khan MF, Khan MA, Mosavi A. Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning. Sensors. 2022;22(10):3833. https://doi.org/10.3390/s22103833.

98. Rahman A, Abbas S, Gollapalli M, Ahmed R, Aftab S, et al. Rainfall prediction system using machine learning fusion for smart cities. Sensors. 2022;22(9):1–15. https://doi.org/10.3390/s22093504.