19 Machine Intelligence-Based Epileptic Seizure Forecasting

Vasily Grigorovsky, Uilki Tufa, Daniel Jacobs, and Berj L. Bardakjian

Abstract

Epilepsy is one of the most common neurological disorders globally, and the decrease in quality of life associated with it includes – among other things – fear and uncertainty over when the next seizure would manifest itself. The most common way to treat epilepsy is by using antiepileptic drugs; however, around 30% of all patients develop refractory epilepsy, where medication fails to control seizures, and patients have to resort to surgical resection of epileptogenic zones. While manual techniques exist to detect epileptic seizures, and come up with the appropriate regiment of antiepileptic drugs, they are generally limited by the skill of the human operator and can be applied only to a particular application. Arguably, a better approach is to use machine intelligence to identify patterns in data unseen to the human eye and perform identification of seizure states, and medicine regiments in an automated objective manner. In this chapter, we will discuss such machine learning algorithms. We will explore the most widely used algorithms and their variations – both in the context of seizure prediction and detection (arguably the most widely used application of machine intelligence in epilepsy), as well as in other applications, such as antiepileptic drug efficacy. We will also talk about common techniques of feature extraction – particularly focusing on wavelet phase coherence and cross-frequency coupling. While much of work has been done to improve current machine learning algorithms in the context of epilepsy, challenges still remain to be solved, and potential future directions for machine intelligence applications in epilepsy are discussed at the end of the chapter.

Keywords

Epilepsy · Seizure prediction · Seizure detection · Machine learning · Cross-frequency coupling · Machine intelligence · State classification

19.1 Introduction

Epilepsy is a dynamical disease, and its effects are evident in up to 1% of the population, or over 60 million people worldwide. It ischaracterized

V. Grigorovsky · U. Tufa · D. Jacobs
Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

B. L. Bardakjian (✉)
Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada
e-mail: berj.bardakjian@utoronto.ca

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_19

by transient interruptions of brain function caused by abnormal temporal and spatial coherent firing of a neuronal population, often referred to as a seizure, paroxysmal discharge, or ictal event [1]. Beyond a number of comorbidities associated with epilepsy, patients with epilepsy are usually unable to predict when they will have a seizure and thus are often unable to drive, have difficulty engaging in the workforce, are at increased risk of head injury due to seizure-related fall, and typically carry a stigma associated with having epilepsy. All of these factors contribute to a reduced quality of life in patients with epilepsy and are largely attributed to the debilitating as well as the unpredictable nature of seizures. Furthermore, patients with refractory epilepsy are also at an elevated risk of sudden unexpected death in epilepsy (SUDEP), which might be preventable if one could anticipate a seizure occurrence [2]. Hence, there exists a need for monitoring systems that detect preclinical seizure states in the EEG to alert patients and caregivers to oncoming seizures.

The pathophysiology of seizures is an enhanced cortical excitability, leading to paroxysmal depolarization shifts, an enhanced probability of hypersynchronous activity of small neuronal networks, and an abnormal spreading of this pathological activity along cortico-cortical and cortico-subcortical neuronal connections [3, 4]. Thus, the common feature of antiepileptic therapies is the reduction of any pathological hyperactivity by either enhancing neuronal inhibition or reducing excitation. Current methods for seizure treatment include either the use of antiepileptic drugs (AEDs) or surgical removal of epileptic tissues. While usually the first treatment option to be used, AEDs require a regiment tailor-made for a given patient and have a wide range of side effects associated with them [5] – thus being able to predict whether a given regiment of AEDs will be successful will improve epilepsy therapy strategies.

In this chapter, we will describe EEG-based machine learning approaches for classification and detection of preclinical seizure states in epileptic patients, as well as look at some other applications of machine intelligence in context of epilepsy.

19.2 Feature Extraction

All machine learning techniques rely on input data to find underlying patterns and develop data-based models. This input data consists of measurable quantities designated as features, and choosing appropriate features is one of the main challenges in machine intelligence. Manual feature selection and tuning is a task that can be time-consuming and often requires expertise in the application. Feature engineering is the process of finding these features from our knowledge of the origin of scalp EEG recordings and deciphering the physiological and pathological basis of their oscillations.

19.2.1 Rhythms of the Brain

Scalp EEG is a noninvasive recording method that has been widely used by neurologists to identify epileptiform activity in patients. Human scalp EEG recordings are measures of electrical fields with contributions from all transmembrane currents in the brain. EEG reflects the summation and superposition of similarly oriented, synchronous neuronal and glial electrical activity favoring superficial sources rather than subcortical deeper structures [7]. The nature of volume integration in the brain leads to spatial averaging in EEG as compared to local field potential (LFP) recordings which can pick out local activity [7]. Nonetheless, EEG signals show brain rhythms relating to neuronal network effects and oscillations with high temporal resolution, and temporal and spectral analysis of these signals forms a large and important set of features for machine learning techniques.

As information in the brain is transmitted using neural coding, spectral information or rhythms at different frequency ranges recorded in EEG have been the target for analysis in perceptual binding and transient short- and long-range coordination. The rhythms of the brain were noticed by Penttonen and Buzsaki to show frequency ranges at an arithmetic progression on the natural logarithmic scale (Fig. 19.1) [6].

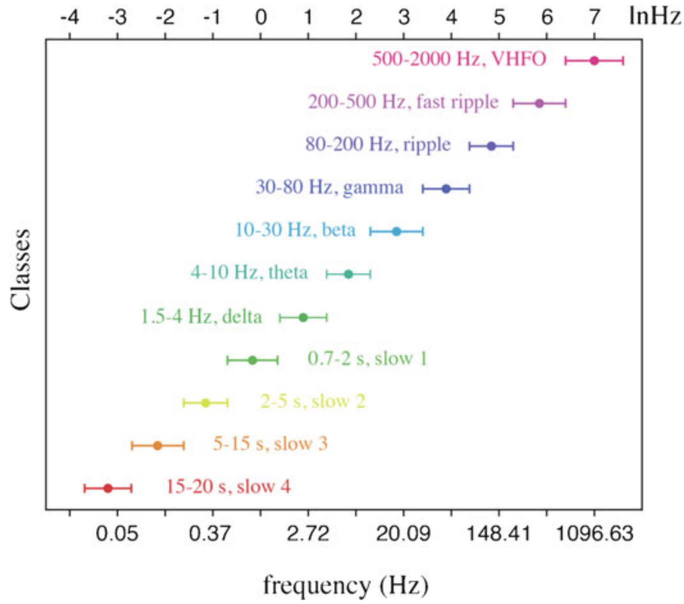

Fig. 19.1 Brain rhythm frequency range following a logarithmic scale. (Figure adapted from Penttonen and Buzsáki [6])

Lower-frequency oscillations allow for longer delays and communication between larger areas. Higher-frequency oscillations facilitate acute and spatially limited communication. These oscillations are concurrent with one another suggesting that the brain works at different time scales [8].

While low-frequency oscillations (LFOs) are important – e.g., the shape and synchronicity of beta (13–30 Hz) waveforms was shown to improve detection of Parkinson's disease pathophysiology in noninvasive recordings [9] – in the past decade, higher frequencies have gained prominence. High-frequency oscillations (HFOs) are defined as frequencies in EEG ranging from 100 to 500 Hz. More specifically, HFOs ranging from 100 to 150 Hz are described as *ripple* and 250 to 500 Hz as *fast ripple* [10]. HFOs have been identified occurring during interictal epileptiform discharges (IEDs) with fast ripples more restricted to seizure-onset zone. Jacobs et al. showed analysis of HFO rate independent of IEDs for identifying seizure-onset zone [10]. Fast ripples during IEDs and in absence showed higher sensitivity in finding the seizure-onset zone while keeping a specificity value of 95%.

The challenge of using HFOs in EEG for highlighting the seizure-onset zone (SOZ) is that often there is an overlap between physiological and pathological activity in the range of high-frequency oscillations. Brazdil et al. showed a higher specificity in locating the zone using frequency ranges from 600 Hz up to 2000 Hz (see Fig. 19.2) [11]. These very-high-frequency oscillations (VHFO) were shown to be present in patients with focal epilepsy [12]. Patients whose resected brain regions more closely corresponded to EEG channels containing VHFOs showed significantly better surgery outcomes indicating that this may be a superior biomarker.

To examine the power of different spectral bands, including the VHFOs, the Fourier transform has enabled us to transform EEG recorded signals from the time domain to the frequency domain. The Fourier transform is given as

$$\Im\{f(t)\} = \hat{f}(\xi) = \int_{-\infty}^{\infty} f(t)\, e^{-2\pi j t \xi} dt, \tag{19.1}$$

where ξ is the frequency in hertz. Applying it to discrete data and using a finite window, the short-term Fourier transform takes the form

$$F(\xi, k) = \sum_{n'=1}^{N} f_k w\left(n' - k\right) e^{-j\xi k}, \tag{19.2}$$

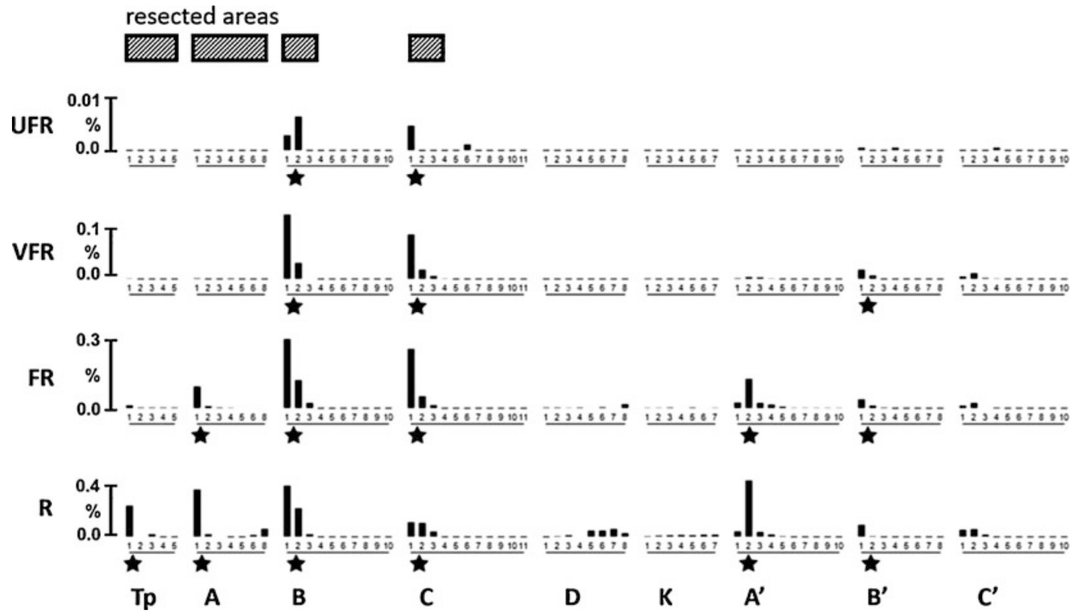

Fig. 19.2 Localization of epileptogenic zone using VHFO activity on iEEG data. Each vertical bar shows the length of duration exceeding threshold in different ranges of frequency bands (R ripple, FR fast ripple, VFR very fast ripple, UFR ultrafast ripple). Stars identify selected regions. (Figure taken from Bradzil et al. 2017)

where f_k is the value of the signal at $t_k = k\delta t$ and $w(n' - k)$ is a window function. The *wavelet transform* follows this transform using a wavelet basis instead of a sinusoidal basis function. Wavelets are a family of functions used as a basis for wavelet transforms which have the property of integrating to zero and are expressed as

$$W(s,n) = \sum_{n'=1}^{N} f_{n'} \psi^* \left[\frac{\left(n' - n\right) \delta t}{s} \right], \quad (19.3)$$

where $\psi(s,n)$ is the wavelet function used with scaling factors s and n. We can convert the scaling factor s into frequencies by scaling the central frequency of the mother wavelet by $1/s$. Continuous wavelet transform (CWT) is preferred over short-time Fourier transforms (STFT) for two distinct reasons. The chosen mother wavelet of the CWT can better extract the preferred frequencies of EEG signal which do not typically follow sinusoidal functions, and the CWT has better temporal resolution increasing with frequency. Complex wavelet transforms are a type of CWT which uses complex mother wavelets. The real and imaginary wavelet coefficients can be used to extract phase information of specific frequency bands in EEG signals.

19.2.2 Wavelet Phase Coherence

Wavelet phase coherence (WPC) is a measure of phase coherence that uses complex wavelet transform to extract the phase information of different frequency bands in EEG data. WPC describes how the phases of two EEG signals change with respect to one another within a time window. Unlike other coherence measurement, WPC is not related to the power of the frequency bands. The relative phase difference $\Delta\phi$ is extracted from wavelet coefficients of two signals $W_1(s,\tau)$ and $W_2(s,\tau)$, with s as the wavelet scaling coefficient and τ as the time shift, as follows:

$$\Delta\phi(s,\tau) = \arctan \left(\frac{W_1^*(s,\tau)\, W_2(s,\tau) - W_1(s,\tau)\, W_2^*(s,\tau)}{W_1(s,\tau)\, W_2(s,\tau) - W_1^*(s,\tau)\, W_2^*(s,\tau)} \right) \quad (19.4)$$

where W^* indicates the complex conjugate. The relative phase coherence is then measured as the

$$\rho(s,\tau) = \left|\left\langle e^{j\Delta\phi(s,\tau)}\right\rangle\right| \qquad (19.5)$$

and ranges from zero to one, with a value of one indicating complete coherence or a constant phase difference within a time window.

Wavelet phase coherence (WPC) of high-frequency oscillations was shown by Cotic et al. to be a useful feature in the localization of the epileptogenic zone [13]. Although the power of HFOs increased during seizures and could roughly locate the epileptogenic zone, WPC was better able to identify electrodes within this zone as confirmed using ROC curve analysis.

19.2.3 Cross-Frequency Coupling

We have thus far introduced brain rhythms and how different regions can show phase coherence within specific frequency ranges. Cross-frequency coupling (CFC) pertains to the communication or brain code observed as a function of two or more interacting frequencies. Phase-amplitude CFC (PAC) has been observed in humans under a variety of conditions [14]. PAC refers to the relationship where the phase of a low-frequency oscillation modulates the amplitude of a high-frequency rhythm. The most popular example of PAC is the theta-gamma code and its role in spatial memory [15]. Distinct neural ensembles observed to fire in the gamma range were encoded within specific phases of theta cycles cued by positional information and long-term memory. One of the most common measures of PAC was developed by Tort et al. [16]. A variation of the algorithm uses complex wavelet transforms to extract phase and amplitude information in contrast to using band pass filtering with Hilbert transforms [17]. The amplitude of the high-frequency rhythm is computed using (Fig. 19.3)

$$A\left(\hat{t}, f_H\right) = \left|\mathrm{Re}\left\{W\left(\hat{t}, f_H\right)\right\} + j\,\mathrm{Im}\left\{W\left(\hat{t}, f_H\right)\right\}\right|. \qquad (19.6)$$

The phase of the low frequency can easily be computed from the analytic wavelet transform representation.

$$\phi\left(\hat{t}, f_L\right) = \arctan\frac{\mathrm{Im}\left\{W\left(\hat{t}, f_L\right)\right\}}{\mathrm{Re}\left\{W\left(\hat{t}, f_L\right)\right\}}. \qquad (19.7)$$

The mean amplitude is normalized in order to have an amplitude-independent measure of CFC

$$p_j\left(\hat{t}, f_H, f_L\right) = \frac{\left\langle A\left(\hat{t}, f_H\right)\right\rangle_j}{\sum_{k=1}^{N}\left\langle A\left(\hat{t}, f_H\right)\right\rangle_k}. \qquad (19.8)$$

The cross-frequency coupling index is then computed as a measure of entropy normalized to a uniform distribution.

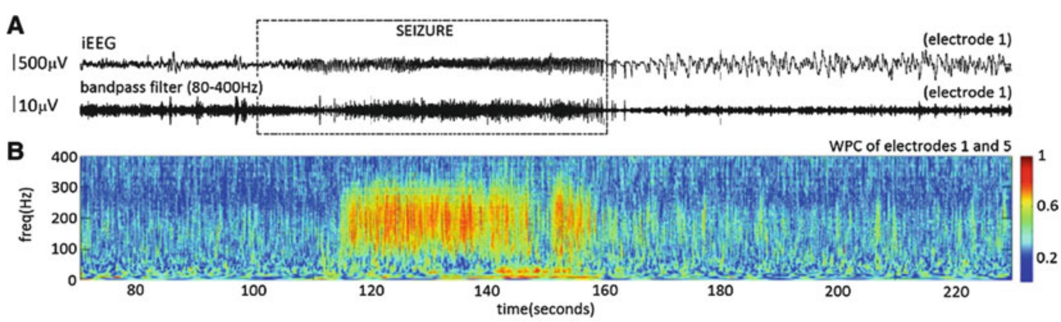

Fig. 19.3 Example wavelet phase coherence between electrodes 1 and 5 during a seizure. (Adapted from Cotic et al. [13])

$$H\left(\hat{t}, f_H, f_L\right) = -\sum_{j=1}^{N} p_j\left(\hat{t}, f_H, f_L\right) \log \left(p_j\left(\hat{t}, f_H, f_L\right)\right) \tag{19.9}$$

$$I_{CFC}\left(\hat{t}, f_H, f_L\right) = \frac{\log N - H\left(\hat{t}, f_H, f_L\right)}{\log N}. \tag{19.10}$$

PAC has been used as a biomarker of both physiological and pathological conditions. Guirgis et al. [18] showed PAC captured seizure dynamics and identified regions of interest for surgical resection in seven patients (Fig. 19.4). Modulation of high-frequency oscillations by delta activity showed higher specificity in selecting the seizure-onset zone (SOZ) as compared with regions determined by neurologists as well as considering the Engel class of the patient (i.e., how seizure-free is the patient after the surgery; EC I–IV denote a progressively worse surgical resection outcome). Conversely, in Amiri et al. [19], theta modulation of high-frequency oscillations was shown to best identify seizure-onset patterns.

19.2.4 Model Performance

Before a given machine learning algorithm can be trained on a set of features, those features need to be tested for reliability. Surrogate analysis is a common way to assess this reliability of nonlinear measures and how they differ from noise and inherent trends in the data. A common way to create surrogate data, described by Theiler et al. [20], is to shuffle phase while having an amplitude adjusted Fourier transform. This method preserves spectral information while removing the original temporal information. In the case of cross-frequency coupling, surrogate analysis consists of shuffling the phase information and recomputing the CFC index. Although we might expect a uniform distribution when binning the amplitude of high-frequency rhythms to phases of low frequencies, there may be an inherent CFC based on the noise of the data.

Table 19.1 Selection of algorithm performance metrics

Sensitivity	$\frac{TP}{TP+FN}$	**Accuracy**	$\frac{TP+TN}{TP+TN+FP+FN}$
Specificity	$\frac{TN}{TN+FP}$	**F1 score**	$\frac{2TP}{2TP+FP+FN}$
False-positive rate	$\frac{FP}{FP+TN}$	**Precision**	$\frac{TP}{TP+FP}$

Once the machine learning model is created, its performance needs to be evaluated. In case of a two-state classification (e.g., seizure vs. non-seizure), a number of metrics can be used; however, first we need to introduce the basic terminology:

True positive (TP) – The algorithm has classified and identified the state.
False positive (FP) – The algorithm has incorrectly identified the state (Type I error).
True negative (TN) – The algorithm has correctly rejected the state.
False negative (FN) – The algorithm has incorrectly rejected the state (Type II error).

From these definitions, several metrics can be established (see Table 19.1).

Sensitivity and specificity are commonly used in evaluating algorithms' performance in general [21]; however, other metrics – especially false-positive rate and accuracy – are also widely used. While it is not necessary to show all of these metrics (some of them can be derived from others), each measure offers different information on how the algorithm performs. For example, high sensitivity indicates a high probability of correctly identifying the diseased state, and high specificity indicates a high probability of correctly rejecting the diseased state. In classification tasks with more than two classes (e.g., interictal, preictal, ictal EEG state classification), these measures can be used for each individual class in a one-vs.-all approach – for example, interictal vs. non-interictal, preictal vs. non-preictal, and so forth.

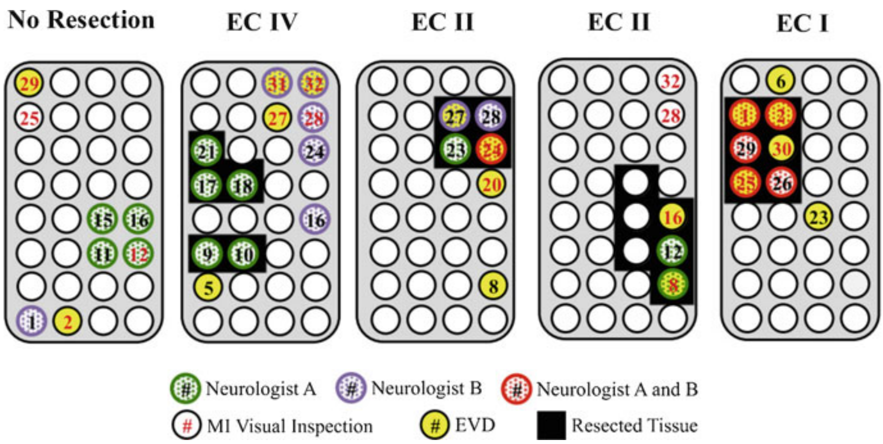

Fig. 19.4 Localization of epileptogenic zone using phase-amplitude cross-frequency coupling of iEEG data. Delta-HFO modulation index (MI) used along with eigenvalue decomposition (EVD) to localize epileptogenic zones in patients who underwent surgical resection with varying outcomes. (Figure adapted from Guirgis et al. [18])

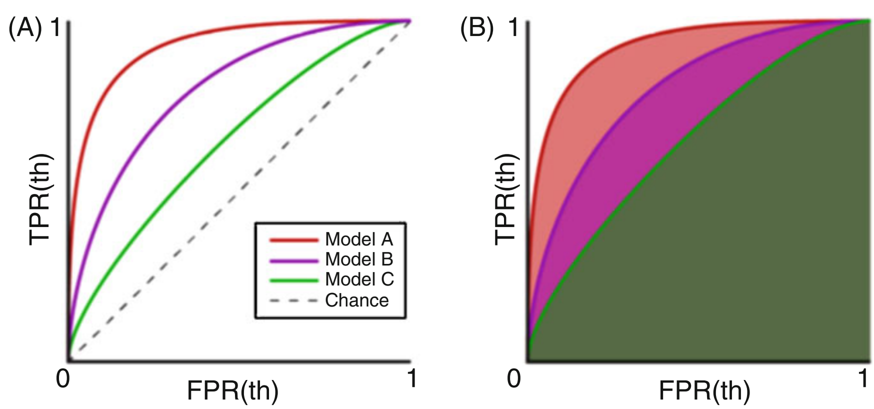

Fig. 19.5 (**a**) Example of receiver operating curve of three different models. (**b**) Area under the curve shows model A as the best performing classification model. (Figure used with author's permission)

These measures can be better visualized as a receiver operating characteristic (ROC), which is a plot of sensitivity against false-positive rate while ranging over values of a parameter of an algorithm with binary classification such as a threshold [22]. ROCs explore the trade-off between high sensitivity and high specificity. We can compare different classification models using the area under the curve (AUC) and find the best parameter to maximize sensitivity and specificity, giving bias to meet the requirements of classification problem (see Fig. 19.5).

19.3 Seizure Detection and Forecasting

The ability to reliably detect, classify, and forecast seizures in epileptic patients can have a profound impact on state-of-the-art therapies for epilepsy and patients' quality of life. Successful classification of EEG signals into a number of states – such as interictal, preictal, or potentially several seizure states – can identify different epilepsy etiologies, predict potential complications, and aid in classifying the severity of seizures. Being able to detect the seizure early (as opposed to after the fact) or even forecast the event before it happens can provide an alert or therapeutic intervention for epileptic patients. People with chronic epilepsy report decreased quality of life and common fear of future seizures due to uncertainty [23], which an early warning system could reduce or eliminate. However, the question whether reliable seizure detection and, forecasting are possible has long been left unanswered. While the difficulty varies significantly based on the task (detection, classification, or forecasting), the quality of data, and the overall goal, it was only in the last decade that computer algorithms became sophisticated enough to be able to forecast epileptic seizures with above chance accuracy (compare Mormann et al. [24] and Kuhlmann et al. [25]).

The algorithms that enabled this breakthrough belong to the area of machine intelligence, especially deep learning, that train on large amounts of data to extract underlying features and patterns which might not be noticeable to the human eye. Generally machine learning algorithms can be split into supervised and unsupervised learning; in this section we will mostly focus on the former category, while still presenting some examples of clustering algorithms used for EEG signal classification. In supervised learning, the algorithm is presented with a training set of inputs and corresponding outputs, based on which it attempts to infer an underlying input-output map – with its performance evaluated on the never-before-seen test set of data. Supervised learning could be further broken down into two areas – classification tasks with categorical outputs, such as seizure detection, and regression tasks with numerical answers, for example, predicting the duration of the seizure. The former dominates epilepsy research, as it is important to determine the current and the next state the patient is in; so in this section, we will exclusively focus on classification algorithms. Another way supervised learning can be divided is into linear models (e.g., logistic regression and support vector machines) and nonlinear models (e.g., decision trees and deep neural networks). We will first look at linear models and how they are used in epilepsy research and then at both tree-based methods and deep neural networks.

19.3.1 Linear Methods

The underlying feature of all linear methods is that, as the name implies, at the core they create a boundary to distinguish between two or more classes (in case of classification tasks) based on some linear combination of input features. For example, a *logistic regression* model applies an activation function to an otherwise linear summation of inputs:

$$z = b + w_1 x_1 + w_2 x_2 + \cdots = b + \boldsymbol{x}^T \boldsymbol{w} \tag{19.11}$$

$$y' = \frac{1}{1 + \exp(-z)}, \tag{19.12}$$

where $\boldsymbol{x}$ is a vector of inputs, $\boldsymbol{w}$ is a weight vector, and b is a bias term. In one case, logistic regression was used for seizure prediction with EEG data from 9 patients with an average of 320 days of recording and 116 seizures each [26]. The signal energy features from four frequency bands (8–16 Hz, 16–32 Hz, 32–64 Hz, and 64–128 Hz) were used, and the algorithm showed the average sensitivity for seizure prediction of 0.55 and an average AUC of 0.79. The authors have suggested to augment the logistic regression by integrating patient-specific circadian information, which increases average sensitivity to 0.61.

On its own, logistic regression is only suited for binary classification – e.g., whether an EEG signal is a seizure or not. However, it can be generalized to a multiclass classification using *softmax regression*, where a softmax function is used to calculate probability of every class occurring given the input (and the class with the largest probability is selected):

$$\sigma(z) = \frac{\exp(z_i)}{\sum_{j=1}^{K} \exp(z_j)}. \tag{19.13}$$

One group utilized softmax regression in a so-called mixture-of-experts model to classify EEG signals into normal or epileptic using the University of Bonn dataset (*Bonn dataset*), which consists of scalp EEG data obtained from five healthy volunteers and five individuals suffering from epilepsy [27]. A mixture-of-experts model consists of a population of simple linear classifiers (such as logistic regression) and a gating network (which contains a softmax function). The gating network mixes outputs from linear classifiers, and during training, it eventually learns to partition inputs such that each classifier is an "expert" in one subset of features. The model used features such as mean, standard deviation, and average power of wavelet coefficients from six distinct frequency bands covering the entire range up to 86.8 Hz and showed an improvement over a basic multilayer perceptron neural network (which we will cover in more detail in a later section) with an increased accuracy (94.5%), specificity (94%), and sensitivity (95%).

Support vector machines (SVMs) are another family of linear models, where the objective is to find the optimal hyperplane separating two classes by maximizing the space between the closest points (or *support vectors*) of these classes (see Fig. 19.6). A linear SVM is very similar to the logistic regression and can be adapted from Eqs. (19.11) and (19.12) to look like this:

$$y' = \sum_i^N w_i y_i k\left(\boldsymbol{x}_i, \boldsymbol{x}'\right) + b \qquad (19.14)$$

$$k\left(\boldsymbol{x}_i, \boldsymbol{x}'\right) = \boldsymbol{x}_i^T \boldsymbol{x}', \qquad (19.15)$$

where y' is the predicted class for the input x' and *k()* is the so-called kernel. Kernels are a transformation of the (potentially nonlinear) feature space associated with a classification problem. A linear SVM is very similar to the logistic regression, but has a few advantages over it, since SVM (a) ensures that the found solution is as fair as possible and (b) less sensitive to outliers compared to logistic regression. In one case, Bonn dataset was used to construct features such as dominant frequency, mean of power spectrum, and coefficient of variation [29]. These features were fed into a linear SVM to classify the given EEG signal as either normal or epileptic. The authors found that while each individual feature had about a 50% accuracy, combining the features led to a 98% accuracy.

Furthermore, SVMs can be used to extend linear modelling to a nonlinear domain, using kernels such as:

$$polynomial : k\left(\boldsymbol{x}_i, \boldsymbol{x}'\right) = \left(\boldsymbol{x}_i^T \boldsymbol{x}' + \mathbf{1}\right)^d \qquad (19.16)$$

$$radial\ basis\ function :$$

$$k\left(\boldsymbol{x}_i, \boldsymbol{x}'\right) = \exp\left(-\gamma \left\|\boldsymbol{x}_i - \boldsymbol{x}'\right\|^2\right). \qquad (19.17)$$

This allows the capture of some of the nonlinear dynamics of the brain. One group used a patient-specific radial basis function (RBF) SVM on intracranial EEG data of 19 out of 21 patients with epilepsy from Epilepsy Center of Univer-

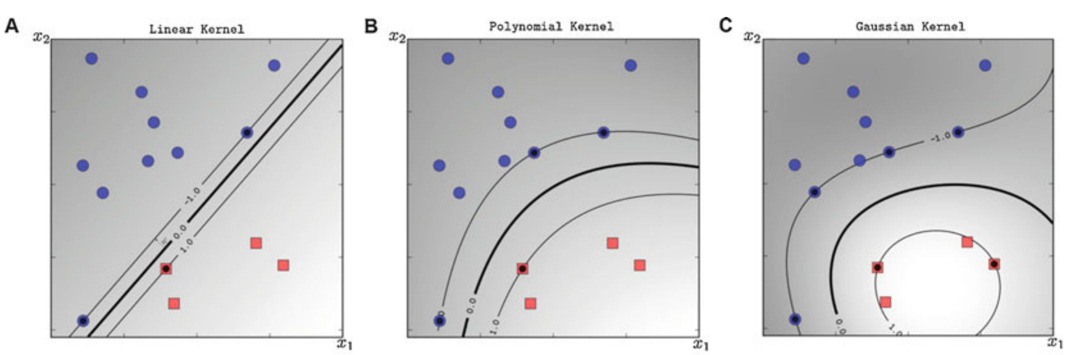

Fig. 19.6 Hyperplane and support vectors (−1 and 1) in a two-class SVM with linear, polynomial, and Gaussian (RBF) kernels. (Figure adapted from Ben-Hur et al. [28])

sity of Freiburg dataset (*Freiburg dataset*), using features based on correlation patterns and space/time delays to forecast seizures [30]. SVM outputs were also averaged over time to reduce noise, and the resultant algorithm, depending on alarm threshold values, had a sensitivity of 0.86–0.95 and false prediction rate (FPR) of 0.03/h to 0.07/h. Additionally, the algorithm spent between 3% and 9% of time in the seizure warning state. As feature selection is an important element of designing seizure detection and forecasting algorithms, another study used RBF SVM to test two ways of identifying the most important features for predicting seizures [31]. The authors used a combined dataset of scalp EEG (sEEG) of 16 patients and intracranial EEG (iEEG) of another 8 patients to extract absolute and relative spectral power from several frequency bands – delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–15 Hz), beta (15–30 Hz), and gamma (30–128 Hz). They compared a method of maximum difference of amplitude distribution histogram (MDAD) between preictal and non-preictal feature samples with minimum redundancy maximum relevance (mRMR) method and found that the former outperformed in seizure prediction with average sensitivity of 75.8% and FPR of 0.1/h, while mRMR showed sensitivity of 64.4% but marginally lower FPR.

Several studies have compared the performance of different commonly used kernels for SVM in the context of seizure detection and forecasting using EEG data. In a work by Zhang and Parhi [32], polynomial and RBF SVM classifiers were compared using iEEG from two patients and spectral power-based features calculated from 10 frequency bands covering the range from 3 Hz to 200 Hz. While RBF SVM classifier showed slightly better performance for predicting a seizure (AUC of 0.9985 compared to 0.9795 of the polynomial SVM), the second degree polynomial SVM classifier used fewer number of features, potentially increasing the computational efficiency of classification. Other studies have compared the performance of linear and nonlinear SVM classifiers. In Shiao et al. [33], the authors found that both linear and nonlinear SVMs can perform with similar sensitivity and FPR (attributing it to a carefully prepared training set), while another study showed that when using permutation entropy (a complexity measure based on neighboring values in the time series), whether nonlinear SVM outperformed linear one or vice versa depended on the state associated with the EEG [34].

While SVMs show adequate results for EEG classification, work is being done to further improve their performance. In the study by Park et al. [35], so-called cost-sensitive SVMs (CSVMs) are proposed, which penalize misclassification of preictal data higher than interictal data in an effort to address the imbalance of preictal and interictal samples in the training set. The authors used the algorithm on the Freiburg dataset and found that it achieved a sensitivity of 97.5% and a FPR of 0.27/h for seizure prediction. Another strategy to improve SVM performance was to utilize a group of different classifiers (an *ensemble*) each trained with a different set of weights – using Bonn dataset and extracted Teager energy among other features. The algorithm achieved an accuracy of 98.72% for seizure detection [36].

As with many machine learning algorithms, one concern with SVMs is that the algorithm will *overfit* the training set, meaning that it will model not just the underlying pattern of the data but also the noise specific to the training set – reducing its performance on the test set. In order to reduce the chance of overfitting, *regularization* is used where the weights or coefficients of the algorithm are kept small, which discourages learning a more complex model. While there are many ways to accomplish that, in one study Kalman filters were used to regularize SVM classifier on coefficients of autoregressive models (AR) of EEG signals to predict seizures, which achieved FPR of as low as 0.02/h [37].

19.3.2 Tree-Based Methods

Another family of machine learning algorithms are tree-based methods. A decision tree is a flowchart-like structure, consisting of branches and nodes, traversing which allows the algorithm to make a conclusion about the class of a new data point based on a recursive analysis of features

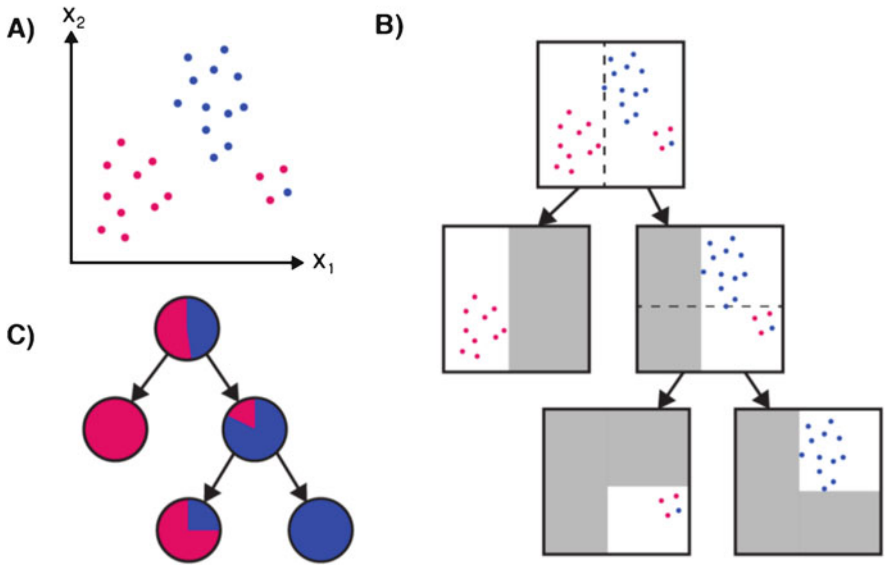

Fig. 19.7 Example of a decision tree for classification. (**a**) Sample training set with two classes. (**b**) Splitting of the original space into decision tree nodes. (**c**) Probability that a point belongs to either class in each node of the tree. (Figure used with author's permission)

associated with the data (see Fig. 19.7). In a decision tree, the nodes represent a feature, the branches connecting the nodes are a decision, and a terminal node (or a leaf) is the probable outcome. Thus, each path from the top of the tree to the leaf is a classification rule which the algorithm applies to the input vector. The trees are constructed by a recursive algorithm of binary splitting, which divides the training set data into two along a feature based on some cost function, with the goal of minimizing the cost. By splitting the data along each feature to come up with least-cost classification rules, decision tree-based algorithms are able to successfully capture nonlinear dynamics of the EEG signals and have been used for seizure detection and forecasting.

In one study, empirical mode decomposition (EMD) has been used to separate scalp EEG signals from the Bonn dataset into mono-rhythmic intrinsic mode functions (IMFs), and corresponding features such as spectral peaks, entropy, and energy of these IMFs were fed into a decision tree algorithm for seizure detection [39]. The algorithm was able to achieve the accuracy of 95.33%, sensitivity of 98%, and specificity of 97%. These results were confirmed by a long-term seizure advisory system, which was implanted into 15 patients with drug-resistant epilepsy for up to 24 months [40]. In that study, features from a range of frequencies from 8 to 128 Hz, such as average energy, Teager energy, and line length, were used in a combination of decision tree and *k-nearest neighbors* (an algorithm where a class of a given data point is determined by plurality vote of *k* of its neighbors) classifiers. The final algorithm showed a patient-specific sensitivity of 54–100% with time spent in "high" alert state of between 3% and 41%. While the algorithm showed a large variability in performance depending on the patient, it was one of the first results from a long-term real-life patient trial where the authors found little to no significant reduction in clinical effectiveness after 4 months of implantation. Several studies have also compared decision trees to other machine learning algorithms. In one, decision trees were compared with SVM classifiers with various kernels (linear, polynomial, RBF) and probabilistic neural networks (which will be briefly covered in the next section) for seizure detection task using features derived from intrinsic time-scale decomposition

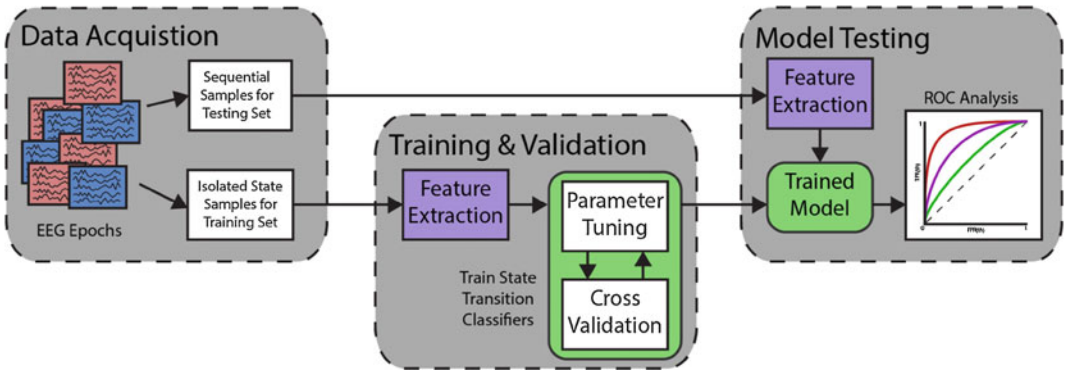

Fig. 19.8 Sample strategy for multistate classifier based on random forest. (Figure adapted from Jacobs et al. [38])

(ITD) – an adaptive data-driven method similar to EMD to decompose a complex signal [41]. The authors found that decision trees performed slightly better than the rest with accuracy of 96%, sensitivity of 99%, and specificity of 99.5%. This finding was confirmed by another study using different set of frequency-related features and an extended algorithm comparison, which found decision trees to have average sensitivity of 99% and specificity of 94% [42].

An iteration on the decision tree algorithm is a logistic model tree, where each of the leaves (terminal nodes) of the tree consists of a logistic regression. Logistic model trees have been reported to be accurate classifiers, combining high performance with ease of interpretability [43]. In one study, they have been used on the Bonn dataset for seizure detection and outperformed both logistic regression and SVM, with an overall AUC of 0.988 (compared to 0.932 and 0.52, respectively) [44].

In an effort to improve the performance of decision trees, an ensemble technique of *random forest* has been developed. As with real forests, a random forest algorithm consists of a number of trees (in this case, decision tree algorithms). In a random forest, each decision tree has access only to a random subset of features while making the decision to split the node and a random subset of training data points. The random forest, then, takes a majority vote (for the classification task) of all individual tree decisions as the final class. The large number of classifiers with, ideally, low correlation between any two trees results in the low error rate of the random forest. Random forest algorithms have been used extensively for seizure forecasting (a sample strategy for random forest use shown in Fig. 19.8). In the work by Tzimourta et al. [45], energy coefficients, entropy, and other frequency-based features were extracted from Bonn and Freiburg datasets and used with a random forest classifier achieving accuracy of 95% with FPR of 0.21/h. In Donos et al. [46], 11 time- and frequency-domain features have been extracted from intracranial EEG of 8 patients and fed into a random forest classifier, which showed 1.75 s median delays of seizure prediction and 0.07/h FPR. As the authors suggested, “For closed loop stimulation devices, an early detection is necessary if termination of epileptic activity prior to first ictal manifestations is aimed at,” which the short median delay of seizure-onset prediction enables. However, with correct input feature selection, the time of advance seizure forecasting can be extended. In Jacobs et al. [38], a global index of cross-frequency coupling computed from scalp EEG was used as an input to a multistage state classifier based on random forest, and the algorithm achieved a 45 ± 16 second advance alarm with AUC of 0.934. Robustness of a classifier to input features is also an important consideration, and in the same study, the authors found that the performance of random forest did not significantly change with reduced electrode ring configuration.

Random forest classifier has also been used for seizure detection. As an example, in Zhang et al. [47], a combination of variational mode decomposition (VMD; an extension of EMD technique, with an advantage of decomposing a multicomponent signal into a number of band-limited intrinsic mode functions non-recursively and synchronously) and AR was used on the Bonn dataset for feature extraction. These features were then fed into a three-state random forest classifier, which delivered an accuracy of 97.4%. In another study, a random forest classifier was compared with both SVM and an existing closed loop neuromodulation device for seizure detection and showed better performance compared to the other two strategies, while maintaining low detection delay and good energy efficiency [48].

19.3.3 Deep Neural Networks

Artificial neural networks (ANN) are a large family of machine learning algorithms inspired by biological neurons. The simplest ANN is system of multiple *perceptrons*, or individual artificial neurons which behave very similarly to the logistic regression described in an earlier section – the only difference being a slightly different activation function. In fact, unlike logistic regression, ANNs can use any activation function, and several have been commonly used. While the activation function cannot be linear (otherwise an ANN will collapse into a single perceptron), both sigmoid (with range 0 to 1) and hyperbolic tangent (with range -1 to 1) functions have been used. More recently, *rectified linear unit* (ReLU) and "leaky ReLU" functions have been designed to improve upon some of the issues with the sigmoid and tanh functions and are defined as

$$f(z) = \begin{cases} z, & z > 0 \\ \alpha z, & z \leq 0 \end{cases}, \tag{19.18}$$

where α is zero for ReLU and a small value (e.g., 0.01) for leaky ReLU. While these activation functions have enjoyed wide adoption as typical activations used in ANNs, a few other functions have been occasionally used, such as a radial basis function and a nonlinear cube function.

While on its own a perceptron is a linear classifier, a system with multiple perceptrons arranged in several layers becomes nonlinear. A typical multilayer perceptron network (MLP, also called a *feedforward network*) has at least three layers – an input layer of features, a hidden layer, and an output layer (a typical MLP is shown in Fig. 19.9). While the input layer has the same number of units as input features, and the number of units in the output layer is restricted by however many classes there are in the classification task, the number of hidden layers and units in each layer is dependent on algorithm design. Too few hidden layers/units lead to poor differentiation of complex patterns in the data, while too many units can lead to overfitting, and too many layers can make training time-consuming – so a careful consideration for MLP parameters is necessary. In the work by Sriraam et al. [50], a three-layer MLP with 10 hidden units was used with spectral power and energy features from scalp EEG of 20 patients for seizure detection and achieved a sensitivity of 97.1%, specificity of 97.8%, and FPR of 1/h. In Subasi and Erçelebi [51], a similar MLP with one hidden layer and 21 hidden units was compared with logistic regression using wavelet-extracted features from 500 scalp EEG segments for seizure classification, and it outperformed the latter algorithm with an accuracy of 92% and AUC of 0.889. MLPs with more than one hidden layer have also been used, for example, in the study by Abbasi and Esmaeilpour [52], where a neural network with two hidden layers (with 4 units in the first and 5 units in the second hidden layer) was used with wavelet-derived features from the Bonn dataset and achieved 98.3% accuracy in seizure detection.

While multilayer perceptron network is the most common among the simpler ANN designs, there are many iterations that attempt to improve the algorithm's performance. *Probabilistic neural network* (PNN) is a neural network with an exponential as an activation function which computes the distance from the test input to the training input vectors and produces a net output as a vector of probabilities. PNNs are characterized by fast training and have been compared with decision trees and SVM classifiers in Martis et

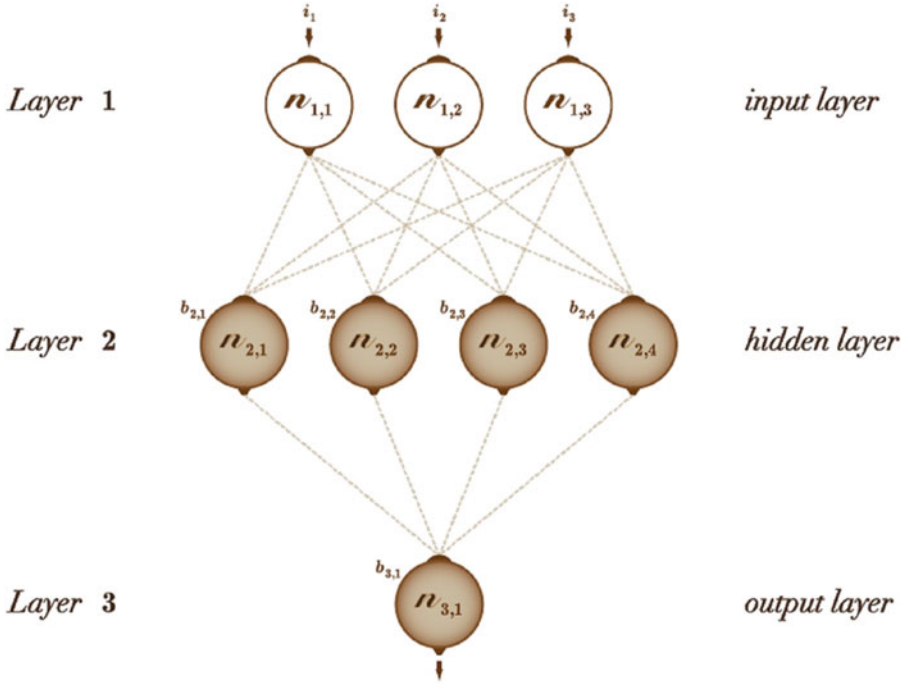

Fig. 19.9 Example of a multilayer perceptron with one hidden layer and four hidden units. (Figure adapted from Acharya et al. [49])

al. [41] and Acharya et al. [42] showing comparable accuracy, sensitivity, and specificity. *Continuous neural networks* are ANNs where each unit is described by ordinary differential equations (ODEs), and in one case, they were trained on Freiburg dataset as well as 90 scalp EEG trials, and the overall correct classification percentage was 97.2%, using features that, unlike most other noncontinuous classifiers, take into account the continuous nature of EEG signals [53]. *Extreme learning machines* (ELM) are a generalized single hidden layer MLP network where the parameters of hidden units (and not just the weights) are randomly generated. A sparse ELM has been shown to perform comparably to SVM classifiers and traditional ANN on a seizure detection task with accuracy of 98.4%, while requiring less storage space and training time [54].

An early comparison of several types of artificial neural networks for EEG state classification was shown in Costa et al. [55]. In the study, the authors investigated (1) a traditional feedforward network, (2) a radial basis function neural network, (3) a layer-recurrent network (with a feedback loop around each layer), and (4) a distributed time-delay network (where the output of a layer also depends on past outputs) using energy-based and complexity-based features extracted from the Freiburg dataset. The comparison showed that in a patient-specific task (i.e., both testing and training data came from the same patient), all of ANNs showed great performance with accuracy of close to 100% – with RBF network performing slightly worse than others. However, when the system was trained on one patient and tested on another, the performance of ANNs dropped significantly.

Perhaps the two patients used for comparison had two drastically different epilepsy etiologies (as the authors suggested), or there was not enough data to properly tune ANNs to successfully classify EEG signals across different patients. However, it is equally likely that ANNs used were unable to capture the full complexity of the provided EEG signals. Deep learning is a subfield of machine learning which is rapidly gaining prominence due to the ability of deep neural networks to better capture complexity associated with real-life data without the necessary fine-tuned feature selection. While the precise definition of what makes a neural network deep is elusive, the generally accepted criterion is having at least three hidden layers. An example of deep neural network is the multilayer perceptron with three hidden layers used with bispectral entropy features for seizure prediction using intracranial EEG data, where it achieved a test accuracy of 78.11% [56].

One specific class of deep neural network is a *convolutional neural network* (CNN or ConvNet, example shown in Fig. 19.10) which was inspired by and is highly correlated with the organization of the visual cortex [58] and has been extensively used on image classification tasks. In general, ConvNets consist of a feature learning stage and a classification stage. The feature learning stage is comprised of convolutional and pooling layers. The former consists of filters or *kernels*, matrices that convolve with the image (a spectrogram, a matrix of wavelet coefficients, or a compilation of EEG signals for seizure detection and forecasting tasks) to extract spatial features and create a feature map. The latter, pooling layer, down-samples the input data and reduces its dimensions, decreasing the necessary computational power as well as extracting dominant features. Two types of pooling layers exist – a *max pooling* returns the maximum value from the subregion of the data, while the *average pooling* returns the average of all values from the subregion. As max pooling can also act as a de-noising filter, it is the preferred choice when designing the CNN. Due to existence of the feature learning stage, CNNs require little preprocessing or manual feature selection, unlike other machine learning algorithms. Features extracted from the input data are then fed into the classification stage, which is typically a multilayer perceptron trained for a classification task.

Recently, convolutional neural networks have been used for seizure prediction and EEG state classification. In a work by Khan et al. [59], a CNN with six convolutional layers (with max pooling) and two dense (or MLP) layers was used with wavelet-transformed scalp EEG signals for seizure prediction and performed with sensitivity of 87.8% and FPR of 0.142/h. In another study, a sequence of short-time Fourier transforms was used with a CNN with three convolutional layers with max pooling and two MLP layers for seizure prediction with FPR of 0.06/h and sensitivity of 81.4% [57]. In both examples, a two-dimensional convolutional neural network was used on, a spectrogram image; however, this need not be the case. In Acharya et al. [49], for example, a one-dimensional CNN with five convolutional layers, five max-pooling layers, and three MLP layers was used on a normalized EEG trace. The algorithm was used to classify scalp EEG into normal, preictal, and seizure states and achieved an accuracy of 88.7%, sensitivity of 95%, and specificity of 90%. On the other hand, in the study by Wei et al. [60], a multichannel scalp EEG data was fed into a three-dimensional CNN with nine total layers, and the seizure detection performance was compared with a two-dimensional CNN and a SVM-based classifier. With average accuracy of 92.4%, the 3-D CNN outperformed the other two classifiers. Further comparison of CNNs to other classifiers also showed that CNNs outperformed SVM and logistic regression classifiers for seizure prediction [61] and achieved zero-false-alarm seizure prediction in 20 out of 21 patients of the Freiburg dataset, while SVMs only had 11 such predictions [62].

Another commonly used class of a deep neural network is a recurrent neural network (RNN), designed specifically for sequential data. RNNs take as an input not only the current training/testing example but also previous information they have encountered – so they are said to have *memory*. Adding this memory can be advantageous since there is information in the sequence itself (e.g., the sequence of interictal $\rightarrow$ preictal $\rightarrow$ ictal EEG states) that other ANNs cannot capture. Recurrent neural networks for seizure prediction were first used in 2000, when an RNN with one hidden layer of between 10 and 15 units was used with intracranial and scalp EEG of two patients for seizure prediction [63]. Both EEG time-series data and wavelet-decomposed spectral bands were fed into the RNN which resulted in up to 15 second early warning of seizure onset.

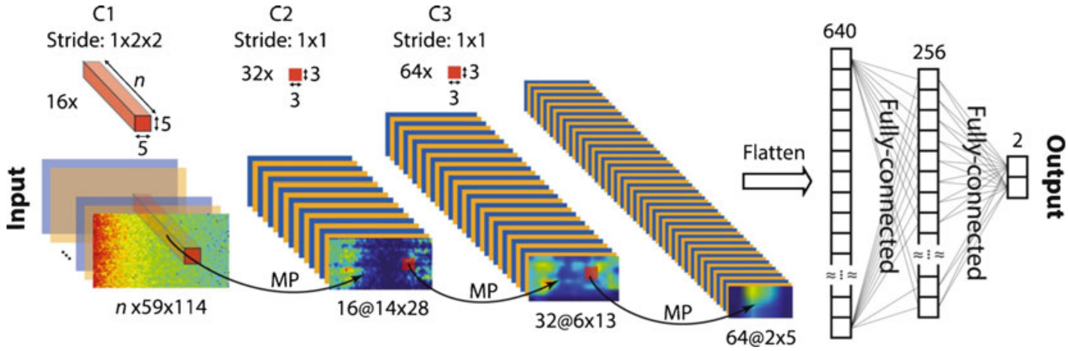

Fig. 19.10 Example of a convolutional neural network, with three convolutional kernels, three max-pooling layers, and two fully connected MLP layers. (Figure from Truong et al. [57])

More work has been done since with RNNs, including classification of epileptic seizures using wavelet energy and norm entropy as features, resulting in average accuracy of 99.8% [64], and using a recurrent *cellular neural network* (an ANN with geometric arrangement of units with the restriction that the communication is only allowed between neighboring units) on EEG time-series data to successfully detect 100% of seizures with an average detection delay of 7.0 seconds [65].

Regular recurrent neural networks have some limitations on their memory, and improved RNNs have been developed – namely, *gated recurrent unit* (GRU) and *long short-term memory* (LSTM, schematic shown in Fig. 19.11) networks. Both networks have units which contain so-called gates, mechanisms regulating the flow of information and allowing the unit to learn which data in the sequence is important to keep. These gates improve the performance, for example, when an LSTM network was used on frequency-domain, time-domain, and cross-correlation features extracted from scalp EEG for seizure prediction [67]. The algorithm achieved average sensitivity of 100% and a false prediction rate of 0.11/h – the authors also noted that increasing the window of preictal data available to LSTM reduced the FPR to as low as 0.03/h.

There has been some effort put into combining recurrent neural networks and ConvNets to take advantage of both automated feature learning and sequential memory in one algorithm. In one study, a CNN-LSTM hybrid algorithm was used on scalp EEG of 23 patients, with three frequency bands covering 0–49 Hz and 2-D projection of electrode placements as features [68]. The proposed hybrid algorithm achieved sensitivity of 95–100%, FPR of 0.1/h for the same patient, and 0.8/h for cross-patient trials. Furthermore, it proved to be more robust to missing electrodes than previous algorithms.

19.3.4 Improving Model Performance

In the previous subsections, we have outlined the main classes of machine learning algorithms used for seizure detection and forecasting from intracranial and scalp EEG signals. However, across all types of algorithms, some strategies exist to further improve classification performance. One of the ways to improve algorithm performance is through using ensemble techniques, where a combination of weak learners is used to create an overall strong learner with better performance. We have briefly mentioned examples of ensemble techniques before, such as random forests, or ensemble of SVM classifiers in the work by Tang and Durand [36]. Ensemble learning can also be extended to ANNs and deep learning, such as using three groups of five neural networks each for three-way EEG signal classification, which improved the performance by 10% compared to

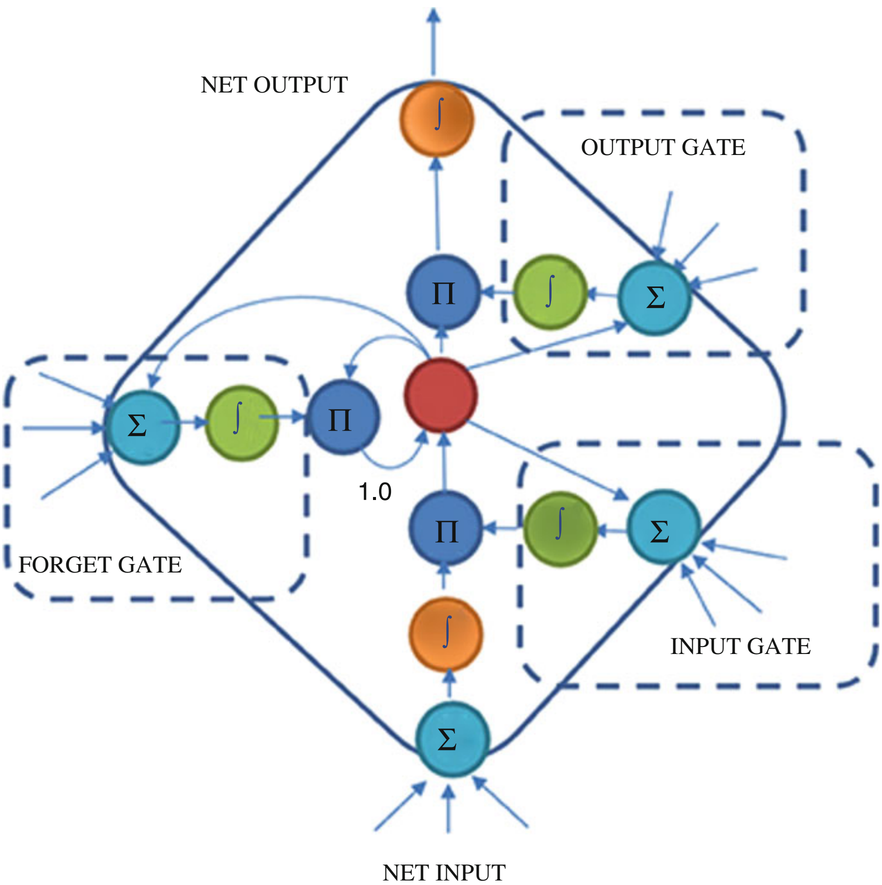

Fig. 19.11 Schematic of a long short-term memory unit for RNN with internal gates for memory management. (Figure from Yu et al. [66])

an individual ANN (98.78% vs. 88%) [69]. In another study, an ensemble of so-called pyramidal convolutional networks (CNNs with smaller kernel size at each layer) was used with raw EEG signals and achieved an accuracy of 99.1% for epilepsy detection task [70].

Ensemble learning is not limited to using multiple copies of the same algorithm. In a work by Abdulhay et al. [71], *k*-nearest neighbor, RBF-SVM classifier, and naïve Bayes (a conditional probability supervised learning method based on Bayes' theorem) classifiers were combined into an ensemble model, and the performance for each base classifier increased by around 3% for EEG state classification. A large study of different ensemble models for seizure forecasting in human and canine epilepsy in an online competition was done by Brinkmann et al. [72], where several of the top 10 algorithms utilized ensemble learning (see Table 19.2) and showed higher AUC than, for example, a ConvNet; moreover, the first algorithm improved its performance AUC by up to 10% compared to its base classifiers. In any ensemble model, the final decision has to be reached from the combination of individual classifiers – one of the most widely used ways of determining the final decision in a classification is a majority vote. However, other ensemble methods exist, such as weighted average, Platt scaling (combining all of the outputs into a probability distribution over all classes), or Bayesian combination of classifiers.

Table 19.2 Details of seizure forecasting classifiers

Selected features used	Machine learning algorithm	Ensemble method	AUC
Spectral power, correlation, distribution statistics, signal variance	Generalized linear model; SVM classifier; random forest	Weighted average	0.82
Log spectral power, covariance	SVM	Platt scaling	0.8
Spectral power, correlation, signal derivative	Neural network; *k*-nearest neighbor	Bayesian combination	0.79
Spectral power, statistical measures, covariance matrices	SVM; generalized linear model	Weighted average of rank scores	0.79
Spectral power, signal standard deviation	Convolutional neural network	N/A	0.78

Adapted from Brinkmann et al. [72]

Ensemble models are not the only strategy for improving classification results – correctly selecting features to feed into a machine learning algorithm is equally important. One way to reduce the algorithm's reliance on correctly selected features is to utilize a CNN with its feature learning stage, which was covered earlier. Another is to rely on unsupervised learning algorithms to automatically identify useful features. Instead of building an input-output map from a training set, unsupervised algorithms find patterns in the data without being provided the "correct" answers. In one study, *k-means* clustering algorithm was used for feature extraction from scalp EEG (Bonn dataset) together with an MLP model to achieve an overall accuracy of 98.3%, about 5–8% increase compared to MLP used with manual features [73]. K-means algorithm finds *k* number of clusters, or collections of data points aggregated together based on some similarity, by reducing the in-cluster distance between every data point and the center of the cluster. Another unsupervised learning technique for feature extraction is *bag-of-words*, originally developed for natural language processing, where each feature vector (a so-called bag) is described by the distribution of unique features ("words") – or how many times each feature has appeared in the input. In the study by Martinez-del-Rincon et al. [74], bag-of-words technique was used with an SVM classifier for seizure detection and showed an overall 10% improvement in the F1 score over the second-best ranked method, likely due to more linear and discriminative feature space.

Careful consideration for the type of machine learning algorithm and the feature selection is necessary for good classification performance. In Fig. 19.12, ROC curves show that even for the same algorithm, using different features can lead to vastly different AUC – in the example, a random forest algorithm using time-based and co-modulogram features led to an increase of 0.226 in AUC compared to power-based features [17]. Deep neural network-based unsupervised learning algorithms also have been used for feature extraction. *Autoencoders* are a type of unsupervised neural networks with two stages – an encoder and a decoder – which attempt to learn an identity function by adjusting hidden layer(s) such that the input and the output are as close to each other as possible – in essence create a reduced representation of the data which can be used as features. The underlying type of neural network used for an autoencoder can vary, for example, in one study, a CNN-based autoencoder feature learning was used with various classifiers (SVM, decision tree, random forest, MLP) for EEG state classification and showed more than 10% improvement in average accuracy compared to other, non-machine-learning, feature extraction techniques [75]. Another study used a stack of two autoencoders to extract the features to the extent that only a supervised learning softmax function was needed, and it achieved accuracy of

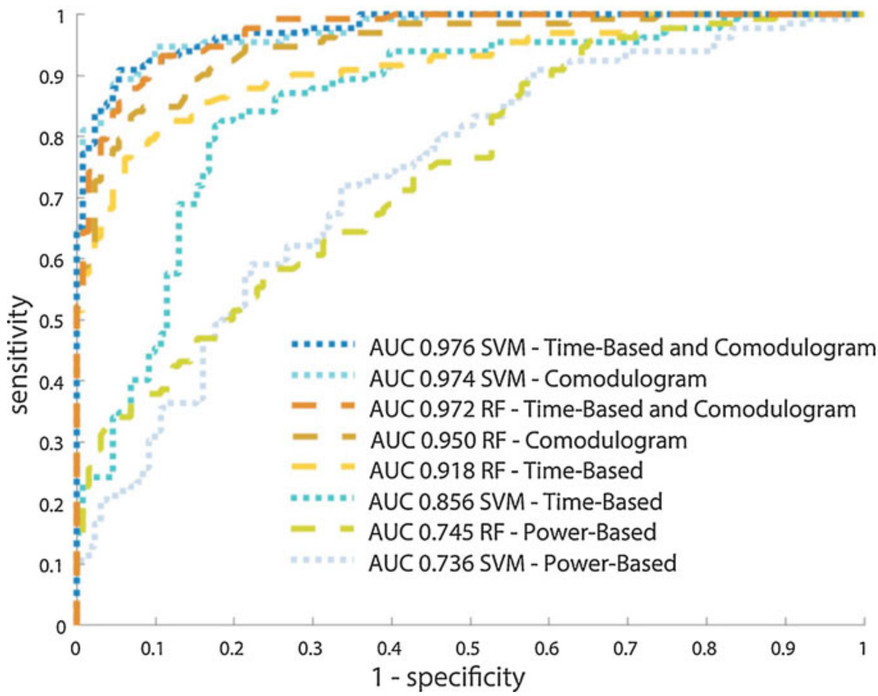

Fig. 19.12 ROC curves of different machine learning algorithms (SVM and RF) using varying sets of features, used to predict AED treatment efficacy. (Figure from Colic et al. [17])

94% (around 15% points higher than the next best method) and FPR of 0.05/h for seizure forecasting [76]. A recent paper improved on that approach by using a deep convolutional autoencoder coupled with bidirectional LSTM and showed an increased per-patient prediction accuracy of 99.6% with false alarm rate of 0.004/h and prediction time of 1 h prior the seizure onset [77].

Occasionally, in addition to feature selection, unsupervised learning algorithms can also be used for seizure prediction and forecasting in their own right. K-means algorithm has been with entropy-based features extracted from the Bonn dataset for seizure detection and showed a 6% higher accuracy with 97% less execution time compared to the SVM classifier [78]. Another type of unsupervised learning used for seizure detection and forecasting is a *hidden Markov model* (HMM) – a probabilistic algorithm used to model a sequence of underlying hidden states based on observable variables. A very common example of an HMM is predicting the weather state (rain, cloudy, sunny) based on the type of clothes people wear without being able to look outside. In context of EEG analysis, HMMs can identify the underlying EEG state based on some observable feature set. In a work by Baldassano et al. [79], an autoregressive hidden Markov model was used with intracranial EEG recordings from six dogs with naturally occurring epilepsy, and the method showed a reduced false-positive rate compared to a previously used random forest classifier with manually selected features (0.0012/h vs. 0.058/h FPR) with an average 12.1 second advance seizure detection. In another study, an HMM with observable states that were assumed to be a combination of Gaussian distributions (a *Gaussian mixture model*) was used with pediatric scalp EEG data to predict seizures with sensitivity of 0.95 and specificity of 0.86 [80].

It is evident from this section that a large variety of machine intelligence algorithms have

Table 19.3 Summary of strengths and weaknesses of a number of common machine learning algorithms

Algorithm	Strengths	Weaknesses
Logistic regression	Output can be interpreted as probability Easy to train	As a linear model, cannot handle nonlinear relationships in the data
SVM	Works well for nonlinear classification Deals well with outliers	Hard to pick the right kernel Memory intensive Poor performance on noisy data
Decision tree	Easy to understand and visualize Require less data preprocessing	Can create complex trees that do not generalize well
Random forest	Improves the performance of decision trees Works well in high-dimensional feature spaces	Output can be hard to interpret Predictions are slow to create Does not work well with sparse datasets
Deep neural networks	Can learn complex input-output mapping of the data Can perform feature extraction On large datasets generally outperform most other algorithms	Require a lot of data Computationally expensive Very hard to interpret the resultant classifier itself and the internal workings of the algorithm

been used for seizure detection, classification, and forecasting. While some studies and strategies discussed above have compared their performance to other classifiers, an astute reader can notice that no one particular method has been identified as the "gold standard" to be used for EEG signal classification. In part, it is due to the fact that EEG signals are inherently complex due to their nonlinear, dynamic, and non-Gaussian nature, making classification difficult. Another reason is the so-called *no free lunch* theorem which states that there is no one machine learning model that works best for every problem due to underlying assumptions one has to make during algorithm design. Deep convolutional neural networks, for example, can perform better than some other classifiers due to fewer number of parameters and CNN's property of rotational and positional invariance; however, that same invariance can prove detrimental when the position or rotation of a feature is important. Furthermore, deep learning models in general are not very good at handling imbalanced data, a situation frequently encountered in EEG signal classification. With that in mind, some of the strengths and weaknesses of machine learning algorithms discussed in this chapter are presented in Table 19.3.

19.4 Other Applications of Machine Intelligence with EEG

In the previous section, we have discussed at length the application of several types of machine learning algorithms to seizure prediction task. However, while these algorithms are effective, they are not the only approach – in one case, effective connectivity of brain networks was used for seizure prediction, achieving sensitivity of 80% and FPR of 0.33/h [81]. Another area where machine intelligence performance is steadily improving is seizure localization. In one study, using intracranial EEG signals, an SVM classifier was trained and tested on patients with Engel class I to class IV outcomes, demonstrating superior performance in the class I patients in Fig. 19.13 [82]. The classification using features based upon both high-frequency and low-frequency oscillations was best able to identify channels suited for resection. This study demonstrates a novel approach to region of interest identification and provides a path for developing tools to improve outcomes in epilepsy surgery [17]. Another SVM classifier was used in identifying SOZ based on phase locking value (PLV) [83]. The study showed that more than 96% of electrodes identified as the SOZ were within the

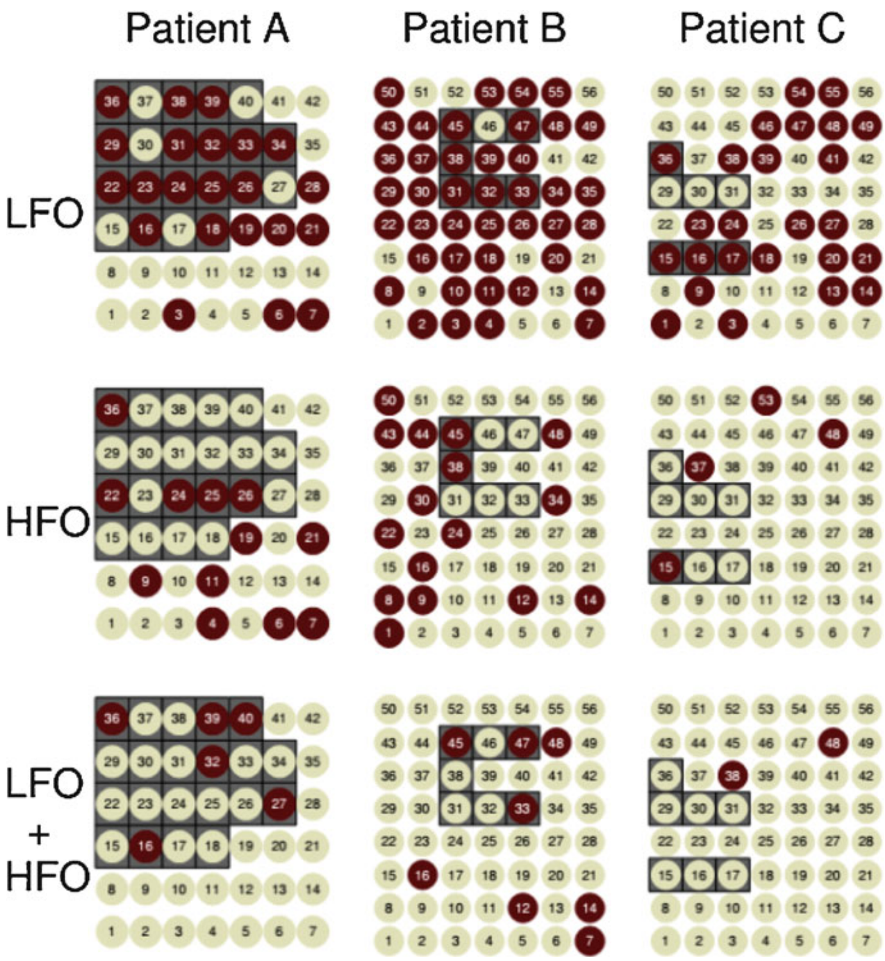

Fig. 19.13 SVM-classified region of interest channels (shown in brown) coincide with resected area (shown in gray) in a seizure-free patient (Patient A, EC I), while in patients where SVM-identified channels are outside of resected area, surgical resection resulted in poor control of seizures (Patients B and C, EC III and IV, respectively). (Figure from Dian et al. [82])

resected area in six seizure-free patients. In four non-seizure-free patients, more than 31% of the identified SOZ electrodes were outside the resected area. Furthermore, in the same study the outcome in non-seizure-free patients correlated with the number of non-resected SOZ electrodes identified. In the study by Tomlinson et al. [84], an SVM classifier was used on iEEG data from 17 pediatric patients, and it was able to predict surgical outcome using global synchrony and local heterogeneity features with 94.1% accuracy.

Both random forest and SVM classifiers were used to distinguish between resection and non-resection areas of 94 patients, using interictal magnetoencephalogram (MEG) recordings. MEG is a technique very similar to scalp EEG, though better suited to source localization, and with features such as delta frequency power, power ratio, and phase lag index extracted from MEG data, both classifiers distinguished the resection areas from non-resection areas with 59.94% accuracy for SVM and 60.34% for random forest (however, the above method was not able to differentiate seizure-free from not seizure-free patients) [85]. Overall, as with seizure prediction, the accuracy of epileptogenic source localization techniques varies based on data modality and features selected. Although machine learning methods showed improvement over manual SOZ identification, they are still facing challenges to properly identify epileptogenic sources especially in noninvasive recordings due to low signal-to-noise ratio (SNR).

19.4.1 Prediction of Antiepileptic Drug Treatment Outcomes

Frequently, antiepileptic drug (AED) treatments produce inconsistent outcomes, so patients may need to go through several drug trials until a successful treatment can be found. There are dozens of commonly used AEDs, and many more experimental drugs, available to treat the disorder. Determining the efficacy of one drug for a specific patient often involves a trial-and-error procedure. There are 20–40% of epileptic patients with drug-resistant epilepsy [86], though they only become aware of this after having already participated in numerous AED trials. Antiepileptic drugs can also make the seizures worse and more frequent, which are associated with numerous side effects that can affect patients' cognition and functioning [87]. Unsuccessful drug trials and delayed treatments highly impact patients' quality of life and are expensive for both patients and the health-care system. Determining a priori the most effective treatment using machine learning methods would go a long way in improving the lives of patients and reducing the financial burden.

While using patients' scalp or intracranial EEG is the gold standard for epilepsy research, sometimes it is easier to do preliminary assessment on computer or animal models before transitioning the methodology to humans. One example of this is the use of rodent models of Rett syndrome – a neurological disorder characterized in part by neural network hyperexcitability and spontaneous epileptiform-like discharges, similar to epilepsy [88]. In Rett syndrome model, an examination of different feature sets showed that, like other classification tasks, the selection of features is vital in achieving class separation and thus has a profound effect on determining treatment outcome [89]. In the study by Colic et al. [17], the normalized power feature projections did not show any clustering by individual animal subjects and were the least useful features in terms of separating responders and non-responders, while *ensemble empirical mode decomposition* (EEMD) time-based and comodulogram features achieved the best separability with distinct clusters for each of the animal subjects. These features were then used with both SVM and random forest classifiers to predict treatment efficacy of an antiepileptic drug, and the results showed that comodulogram features (AUC 0.974) outperformed those of EEMD time-based (AUC 0.918) and normalized power (AUC 0.745) – see Fig. 19.12.

When the two machine learning methods were evaluated to predict the treatment outcome of four different AEDs, SVM was found to predict the treatment outcome of outliers found in random forest predictions (see Fig. 19.14). In the same study by Colic et al. [17], random forest prediction of treatment outcome for ganaxolone applied on mouse 2 was close to 100%, when it should have been closer to 0%, whereas SVMs predicted 44%. Similarly, for phenytoin, the prediction for mouse 1 was 84% when it should have been closer to 0%, whereas SVMs predicted 59%. Generally, SVMs estimated 90% or greater likelihood scores only for successful treatments.

Patient variability is a serious challenge to selecting treatments for epilepsy. Often antiepileptic drug treatments are cycled through until an effective treatment can be found, and with over two dozen commonly prescribed AEDs available, it can be a cumbersome process. There are certain AEDs that have been found to be statistically more likely to lead to a successful treatment outcome, and it is those AEDs that typically are tried first. However, the likelihood of a successful treatment reduces with each round of AED application [90], possibly due to patient desensitization to AEDs which happens over time. By indicating which patients would be unresponsive to certain AEDs, and what AEDs are most likely to be successful – using machine intelligence – epileptologists could choose the most appropriate therapy for the patient without unnecessary testing of AEDs, and the treatment is more likely

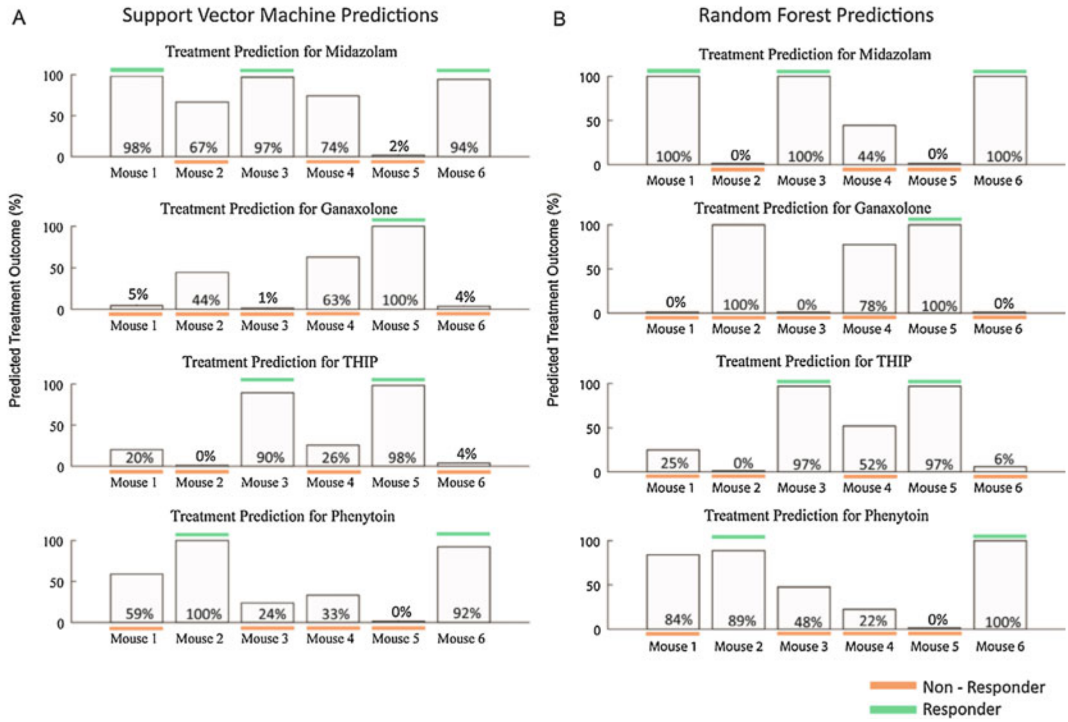

Fig. 19.14 Predicted likelihood of favorable treatment outcome across four commonly used AEDs using SVM and RF machine learning algorithms. Green bars indicate the patient was successfully treated by specific AED, and brown bars indicate unsuccessful treatment. (**a**) SVM predictions accurately predicted treatment outcome for all AEDs. (**b**) RFs had comparable prediction results, with misclassifications for ganaxolone treatment for mice 2 and 4 and phenytoin for mouse 1. (Figure from Colic et al. [17])

to show a positive improvement in a patient's quality of life.

19.5 Current Challenges and Future Directions

In this chapter, we have focused on the use of machine intelligence for seizure detection and forecasting, and prediction of antiepileptic drug treatment outcomes, as well as feature extraction and selection to be used for machine learning algorithms – including wavelet phase coherence and cross-frequency coupling. While a lot of progress has been made in the past several years to improve EEG-based techniques with cutting-edge algorithms, several important challenges still remain. Frequently, EEG data (especially obtained from scalp) is imbalanced, favoring one class (e.g. interictal EEG state) over others, and characterized by relatively low signal-to-noise ratio, which can significantly impair a given classifier's performance – so any classifier should be designed to be robust to high noise and class balance issues. In situations where it is important to understand how the classifier reached its decision, low interpretability of machine learning algorithms (especially deep neural networks) might prove that it is difficult to get the necessary insight.

Other major challenges of EEG-based machine learning algorithms include issues concerning EEG data, namely amount of data, source of data, accurate data labels, and artifacts. Due to the constraints of human EEG acquisition, there is typically a relatively small amount of heterogeneous data available for a particular task – usually on the order of a couple of dozen to a couple of hundred EEG segments from 5 to 20 patients. While it might seem like a lot of data for manual analysis, this amount of data could make it difficult for machine learning algorithms (especially deep networks) to achieve reliable and highly accurate classification. One way to circumvent this issue is to use a technique called transfer learning, where a model trained in the domain with a lot of general data (e.g., all of scalp EEG available) is repurposed to a more specific task (e.g., antiepileptic drug efficacy). This improved the classifier's performance, since the algorithm can learn more basic features on a larger dataset. One such strategy was used in Liang et al. [91] where six available EEG datasets not related to seizure prediction were used as auxiliary information to a one dataset for seizure prediction and found that the prediction performance improved. A related issue is the source of EEG data – it is easier to get the data for analysis from animal models; however, one must be careful to ensure that features or classes they identified are transferable to humans.

Sometimes, parts of the data are unlabelled, or there is some uncertainty about how reliable labels are. This poses an issue for the classifier, since it is given incorrect or missing training data. In this situation, so-called *semi-supervised learning* techniques can be used, such as semi-supervised version of extreme learning machines (ELM) which, despite having unlabelled data, outperformed a fully supervised ELM model [92]. The most common issue with real-world EEG signals is the presence of artifacts. Artifacts in EEG can be very diverse, from not relevant physiological signals (e.g., EMG, ECG) to cable and electrode movement, environmental interference, and recording equipment; they can be present in multiple electrodes or only in one and can be periodic or irregular. Most of publicly available EEG datasets manually remove artifacts ahead of time, which means that algorithms trained on them will not perform as well on the non-processed data. Islam et al. [93] presented a thorough review of methods for artifact detection and removal, but, in short, artificial neural networks, SVM classifiers, and *k*-means clustering can be used to detect unwanted signals, while other techniques, such as independent component analysis, EMD, wavelet transform, and neural network-based algorithms, have been used for artifact correction.

Specific uses of machine intelligence can also have their unique challenges. As an example, for seizure prediction, it can be complicated to compare patient-specific algorithms that are in the 95–100% sensitivity range. For one, patient-specific algorithms require new training for every new patient, so optimally cross-patient algorithms should be prioritized. Another issue is the potential discrepancy between reported benchmarks and real-life expectations. For example, the best seizure prediction algorithms report around 0.05/h false-positive rate, which appears low especially compared to previous methods. However, that translates to roughly one false alarm every day. For some uses, such as warning the person about the upcoming seizure, this might not be an acceptable rate; for others, such as neurostimulation system, it might be within tolerance – though long-term effects of routine daily neurostimulation should probably be investigated. Sometimes, parameters that normally are not a main priority (such as the latency of the algorithm or its energy efficiency) become crucially important, as they are in mobile seizure prediction systems. All these challenges – both general and specific – are the reasons why a recent seizure prediction system designed for a wearable device achieved mean sensitivity of only 69% [94].

At their current stage, machine learning algorithms can be used to augment existing techniques, such as providing an opinion on the potential location of the epileptogenic zone, or identifying seizures for further processing. While it is not yet clear whether machine intelligence will completely eliminate the need for manual intervention, some future directions of EEG-based algorithms can be suggested. One likely potential development is integration of more probabilistic modelling into machine learning algorithms. Estimating seizure probability as a way to detect seizures has already been investigated by Kuhlmann et al. [95], a circadian probability as subclassifier was used in Karoly et al. [26] for seizure forecasting, and we have briefly described a probabilistic neural network in Sect. 19.3. A natural extension to all those is a *Bayesian neural network* (BNN) or Bayesian deep learning, which, for example, was used recently with scalp EEG for mental fatigue detection [96]. BNN is a neural network that uses a prior probability distribution on its weights in order to incorporate uncertainty about the prediction. This gives an advantage of BNN to work better on smaller datasets, prevent overfitting, and give an overall insight over how reliable the given prediction is. Another way to incorporate probability into machine intelligence is to use *restricted Boltzmann machines* (RBM). A Boltzmann machine (BM) is a type of unsupervised fully connected stochastic recurrent neural network with a visible input layer and at least one hidden layer, while an RBM has a *restriction* that connections can exist only between layers. In context of EEG signals, one interpretation is that the units in the visible layer represent observable attributes, while the hidden layer units act as nonlinear feature detectors, and recently, an RBM-based technique has been successfully evaluated for detection of epileptogenic lesions [97].

Another potential development is the integration of *genetic algorithms* with machine learning techniques to improve feature or hyperparameter (a parameter with a value set before the learning process) selection. Genetic algorithm belongs to a family of evolutionary computation algorithms inspired by biological evolution – mirroring the biological inspiration between various types of artificial neural networks. In short, the genetic algorithm generates multiple candidate solutions with various parameters and after some training assesses their "fitness." Each new generation of algorithms is produced by removing less fit solutions and introducing small random changes (mimicking biological concepts of mutation and crossover) – this eventually creates a subset of high-quality optimized solutions to a given problem. A recent review thoroughly examined a number of evolutionary computation algorithms for EEG feature selection, including the genetic algorithm [98], while another study found that using genetic algorithm with an MLP for a major depressive disorder classification task increased accuracy and AUC by 10% [99]. In a work by Mesejo et al. [100], an evolutionary computation algorithm was combined with an *artificial neuron-glia network* (ANGN) – an extension of a regular ANN to include longer-term dependencies for weight adjustments which mirror effects of astrocytes (dominant glial cells in the brain) in biological neural networks. Astrocytes have been shown to be involved in neuronal firing [101], particularly that their activity has an effect on neuronal codes similar to those seen in the human brain [102]. These findings make astrocytes an attractive target for more biologically inspired machine learning algorithms. While in the study by Mesejo et al. [100] the resultant algorithm performed comparably to existing ANNs, introducing more biomimetic algorithms for machine intelligence tasks could result in better performance in complex problems.

One final direction of future development is adapting alternative sequential models for EEG analysis. Since EEG data is sequential in nature, machine learning algorithms would benefit from having memory to be able to capture existing temporal dependencies within it. We have already described several variants of recurrent neural networks – deep neural networks adapted for sequential data – and their use in seizure prediction studies. One disadvantage of RNNs, however, is that they require a lot of resources (time and computational power) to train properly. *Autoregressive feedforward models*, such as a WaveNet [103] or gated convolutional networks [104], are being developed as an alternative to RNNs. In autoregressive neural networks, instead of relying on most of the history of the sequence for making predictions, the model only uses the finite number n of most recent inputs. While theoretically RNNs should be more flexible, in practice, Bai et al. [105] showed that autoregressive neural networks outperform comparable RNNs in a wide variety of tasks such as audio synthesis and machine translation while also benefitting from significantly easier and faster model training and prediction.

Acknowledgments The authors would like to acknowledge funding from Canadian Institutes of Health Research and National Sciences and Engineering Research Council.

Homework

Conceptual Questions

1. Given the discussion of feature engineering of both scalp and intracranial EEG data in this chapter, describe some useful features for seizure detection and prediction.
2. Given a relatively small dataset of 10 patients with a selection of interictal, preictal, and ictal recordings, a) suggest an approach to divide the dataset into training and test sets, and b) provide benefits and drawbacks of leaving one or more patients entirely for the test set.
3. For the same dataset as in previous question, suggest what machine learning algorithm you would use and why. Would your answer change if a) it was only two classes; b) the dataset contained 1000 patients; c) the algorithm needs to perform EEG state classification in real time.
4. What are common noise sources and artifacts in EEG recordings? Suggest a few ways to improve signal quality and eliminate these artifacts.
5. Frequently, EEG data is imbalanced, favoring one class over others. How does that impact classification performance? How would you overcome this issue?
6. In this chapter, we have briefly covered several network architectures where the targets are the same as their inputs. Name two and explain when you would likely use them.

Practical Analysis Questions

These questions are intended as introductory guides to your own practical implementation of the techniques outlined in this chapter.

For questions 7 and 8, use data from UPenn and Mayo Clinic's Seizure Detection Challenge (available at https://www.kaggle.com/c/seizure-detection).

7. Physicians and researchers working in epilepsy often review large quantities of EEG data to identify seizures, which in some patients may be quite subtle and hard to detect. Automated algorithms to detect seizures in large EEG datasets with low false-positive rates (FPR) and false-negative rates (FNR) would greatly assist both clinical care and preclinical research. Using a multilayer perceptron, classify windows of human EEG data as *seizure* or *non-seizure*. Use spectral power features computed from 1 second windows as inputs to the MLP (see figure below). Divide the data into a training set and a testing set using a ratio of 80% to 20%, respectively. Use the training set to train the MLP and the testing set to find the FPR and FNR. Compute an ROC curve and the area under the curve to compare network performance.
   (a) Using a MLP with one hidden layer, and gradient descent method with step size of 0.5, alter the number of units in the hidden layer (5, 10, 40) and explore whether increased number of hidden units will have a positive effect on the network performance. What are the pros and cons of having more hidden units?
   (b) Alter the number of hidden layer (no hidden layers, 1 hidden layer, or 2 hidden layers) in the feedforward neural network, using 10 units per hidden layer and gradient descent method with step size of 0.5. Determine whether increased number of hidden layers will have a positive effect on the network performance. What are the pros and cons of having more hidden layers?
   (c) Would you say that using a convolutional neural network is preferable over using a multilayer perceptron and why?
8. Using the same approach as in question 6, explore the effect of training parameters.
   (a) *Learning Rate* – Try different step sizes or learning rates ($lr = 0.1, 0.5, 1$) using gra-

dient descent training function on a neural network with one hidden layer network with 40 hidden units. Determine whether large step size will always expedite learning.

(b) *Momentum* – Investigate the effect of momentum using a network with 1 hidden layer (10 units) and gradient descent with momentum ($mc = 0.1, 0.5, 0.9$). Determine whether a strong momentum term will always expedite learning.

For questions 9 and 10, use data from American Epilepsy Society Seizure Prediction Challenge (available at https://www.kaggle.com/c/seizure-prediction).

9. Responsive neurostimulation (RNS) presents a possible therapy for abolishing seizures in epileptic patients that are drug-resistant and ineligible for surgery. Seizures that build and generalize beyond the area of origin are very difficult to abort; thus electrical stimulation must be applied as early as possible. Using the same algorithmic approach as in question 6, train your system to *predict* epileptic seizures in human patients. How does your performance (in terms of the AUC metric) compare to the seizure detection task as well as results shown in Table 19.2? Explain your results.
10. Suggest improvements to your seizure prediction algorithm. Select a few improvements, and implement them to see how much AUC is increased compared to results in question 8. If you know that sequential state changes are characteristic of seizure episodes, how does that change your suggested improvements?

References

1. D.M. Durand, M. Bikson, Suppression and control of epileptiform activity by electrical stimulation: A review. Proc. IEEE **89**, 1065–1082 (2001). https://doi.org/10.1109/5.939821
2. R. Surges, R.D. Thijs, H.L. Tan, J.W. Sander, Sudden unexpected death in epilepsy: Risk factors and potential pathomechanisms. Nat. Rev. Neurol. **5**, 492–504 (2009). https://doi.org/10.1038/nrneurol.2009.118
3. F.E. Dudek, T.P. Sutula. Epileptogenesis in the dentate gyrus: A critical perspective. *Progress Brain Res*. **153**, 755–773 (2007)
4. M. Steriade, Corticothalamic networks, oscillations, and plasticity. Adv. Neurol. **77**, 105–134 (1998)
5. E. St. Louis, Minimizing AED adverse effects: Improving quality of life in the interictal state in epilepsy care. Curr. Neuropharmacol. **7**, 106–114 (2009). https://doi.org/10.2174/157015909788848857
6. M. Penttonen, G. Buzsáki, Natural logarithmic relationship between brain oscillators. Thalamus Relat. Syst. **2**, 145–152 (2003). https://doi.org/10.1016/S1472-9288(03)00007-4
7. G. Buzsáki, C.A. Anastassiou, C. Koch, The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. **13**, 407–420 (2012). https://doi.org/10.1038/nrn3241
8. G. Buzsáki,*Rhythms of the Brain* (Oxford University Press, Oxford/New York, 2006)
9. N. Jackson, S.R. Cole, B. Voytek, N.C. Swann, Characteristics of waveform shape in Parkinson's disease detected with scalp electroencephalography. eNeuro (2019). https://doi.org/10.1523/ENEURO.0151-19.2019
10. J. Jacobs, P. LeVan, R. Chander, et al., Interictal high-frequency oscillations (80-500 Hz) are an indicator of seizure onset areas independent of spikes in the human epileptic brain. Epilepsia **49**, 1893–1907 (2008). https://doi.org/10.1111/j.1528-1167.2008.01656.x
11. M. Brázdil, M. Pail, J. Halámek, et al., Very high-frequency oscillations: Novel biomarkers of the epileptogenic zone: VHF oscillations in epilepsy. Ann. Neurol. **82**, 299–310 (2017). https://doi.org/10.1002/ana.25006
12. J. Jacobs, R. Staba, E. Asano, et al., High-frequency oscillations (HFOs) in clinical epilepsy. Prog. Neurobiol. **98**, 302–315 (2012). https://doi.org/10.1016/j.pneurobio.2012.03.001
13. M. Cotic, O.C. Zalay, Y. Chinvarun, et al., Mapping the coherence of ictal high frequency oscillations in human extratemporal lobe epilepsy. Epilepsia **56**, 393–402 (2015). https://doi.org/10.1111/epi.12918
14. R.T. Canolty, R.T. Knight, The functional role of cross-frequency coupling. Trends Cogn. Sci. **14**, 506–515 (2010). https://doi.org/10.1016/j.tics.2010.09.001
15. J. Lisman, The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme. Hippocampus **15**, 913–922 (2005). https://doi.org/10.1002/hipo.20121
16. A.B.L. Tort, R. Komorowski, H. Eichenbaum, N. Kopell, Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. J. Neurophysiol. **104**, 1195–1210 (2010). https://doi.org/10.1152/jn.00106.2010

17. S. Colic, R.G. Wither, M. Lang, et al., Prediction of antiepileptic drug treatment outcomes using machine learning. J. Neural Eng. **14**, 016002 (2017). https://doi.org/10.1088/1741-2560/14/1/016002
18. M. Guirgis, Y. Chinvarun, M. del Campo, et al., Defining regions of interest using cross-frequency coupling in extratemporal lobe epilepsy patients. J. Neural Eng. **12**, 026011 (2015). https://doi.org/10.1088/1741-2560/12/2/026011
19. M. Amiri, B. Frauscher, J. Gotman, Interictal coupling of HFOs and slow oscillations predicts the seizure-onset pattern in mesiotemporal lobe epilepsy. Epilepsia **60**, 1160–1170 (2019). https://doi.org/10.1111/epi.15541
20. J. Theiler, S. Eubank, A. Longtin, et al., Testing for nonlinearity in time series: The method of surrogate data. Physica D: Nonlinear Phenomena **58**, 77–94 (1992). https://doi.org/10.1016/0167-2789(92)90102-S
21. A.G. Lalkhen, A. McCluskey, Clinical tests: Sensitivity and specificity. Contin. Educ. Anaesth. Crit. Care Pain **8**, 221–223 (2008). https://doi.org/10.1093/bjaceaccp/mkn041
22. C.D. Brown, H.T. Davis, Receiver operating characteristics curves and related decision measures: A tutorial. Chemom. Intell. Lab. Syst. **80**, 24–38 (2006). https://doi.org/10.1016/j.chemolab.2005.05.004
23. S. Beyenburg, A.J. Mitchell, D. Schmidt, et al., Anxiety in patients with epilepsy: Systematic review and suggestions for clinical management. Epilepsy Behav. **7**, 161–171 (2005). https://doi.org/10.1016/j.yebeh.2005.05.014
24. F. Mormann, R.G. Andrzejak, C.E. Elger, K. Lehnertz, Seizure prediction: The long and winding road. Brain **130**, 314–333 (2007). https://doi.org/10.1093/brain/awl241
25. L. Kuhlmann, K. Lehnertz, M.P. Richardson, et al., Seizure prediction—Ready for a new era. Nat. Rev. Neurol. **14**, 618–630 (2018). https://doi.org/10.1038/s41582-018-0055-2
26. P.J. Karoly, H. Ung, D.B. Grayden, et al., The circadian profile of epilepsy improves seizure forecasting. Brain **140**, 2169–2182 (2017). https://doi.org/10.1093/brain/awx173
27. A. Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst. Appl. **32**, 1084–1093 (2007). https://doi.org/10.1016/j.eswa.2006.02.005
28. A. Ben-Hur, C.S. Ong, S. Sonnenburg, et al., Support vector machines and kernels for computational biology. PLoS Comput. Biol. **4**, e1000173 (2008). https://doi.org/10.1371/journal.pcbi.1000173
29. C.H. Seng, R. Demirli, L. Khuon, D. Bolger, Seizure detection in EEG signals using support vector machines, in *2012 38th Annual Northeast Bioengineering Conference (NEBEC)*, (IEEE, Philadelphia, 2012), pp. 231–232
30. J.R. Williamson, D.W. Bliss, D.W. Browne, J.T. Narayanan, Seizure prediction using EEG spatiotemporal correlation structure. Epilepsy Behav. **25**, 230–238 (2012). https://doi.org/10.1016/j.yebeh.2012.07.007
31. M. Bandarabadi, C.A. Teixeira, J. Rasekhi, A. Dourado, Epileptic seizure prediction using relative spectral power features. Clin. Neurophysiol. **126**, 237–248 (2015). https://doi.org/10.1016/j.clinph.2014.05.022
32. Z. Zhang, K.K. Parhi, Seizure prediction using polynomial SVM classification, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, Milan, 2015), pp. 5748–5751
33. H.-T. Shiao, V. Cherkassky, J. Lee, et al., SVM-based system for prediction of epileptic seizures from iEEG signal. IEEE Trans. Biomed. Eng. **64**, 1011–1022 (2017). https://doi.org/10.1109/TBME.2016.2586475
34. N. Nicolaou, J. Georgiou, Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. Expert Syst. Appl. **39**, 202–209 (2012). https://doi.org/10.1016/j.eswa.2011.07.008
35. Y. Park, L. Luo, K.K. Parhi, T. Netoff, Seizure prediction with spectral power of EEG using cost-sensitive support vector machines: Seizure prediction with spectral power of EEG. Epilepsia **52**, 1761–1770 (2011). https://doi.org/10.1111/j.1528-1167.2011.03138.x
36. Y. Tang, D.M. Durand, A tunable support vector machine assembly classifier for epileptic seizure detection. Expert Syst. Appl. **39**, 3925–3938 (2012). https://doi.org/10.1016/j.eswa.2011.08.088
37. L. Chisci, A. Mavino, G. Perferi, et al., Real-time epileptic seizure prediction using AR models and support vector machines. IEEE Trans. Biomed. Eng. **57**, 1124–1132 (2010). https://doi.org/10.1109/TBME.2009.2038990
38. D. Jacobs, T. Hilton, M. del Campo, et al., Classification of pre-clinical seizure states using scalp EEG cross-frequency coupling features. IEEE Trans. Biomed. Eng. **65**, 2440–2449 (2018). https://doi.org/10.1109/TBME.2018.2797919
39. R.J. Martis, U.R. Acharya, J.H. Tan, et al., Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals. Int. J. Neural Syst. **22**, 1250027 (2012). https://doi.org/10.1142/S012906571250027X
40. M.J. Cook, T.J. O'Brien, S.F. Berkovic, et al., Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study. Lancet Neurol. **12**, 563–571 (2013). https://doi.org/10.1016/S1474-4422(13)70075-9
41. R.J. Martis, U.R. Acharya, J.H. Tan, et al., Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction. Int. J. Neural Syst. **23**, 1350023 (2013). https://doi.org/10.1142/S0129065713500238

42. U.R. Acharya, S.V. Sree, P.C.A. Ang, et al., Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals. Int. J. Neural Syst. **22**, 1250002 (2012). https://doi.org/10.1142/S0129065712500025
43. N. Landwehr, M. Hall, E. Frank, Logistic model trees. Mach. Learn. **59**, 161–205 (2005). https://doi.org/10.1007/s10994-005-0466-3
44. E. Kabir, Siuly, Y. Zhang, Epileptic seizure detection from EEG signals using logistic model trees. Brain Informatics **3**, 93–100 (2016). https://doi.org/10.1007/s40708-015-0030-2
45. K.D. Tzimourta, A.T. Tzallas, N. Giannakeas, et al., A robust methodology for classification of epileptic seizures in EEG signals. Heal. Technol. **9**, 135–142 (2019). https://doi.org/10.1007/s12553-018-0265-z
46. C. Donos, M. Dümpelmann, A. Schulze-Bonhage, Early seizure detection algorithm based on intracranial EEG and random Forest classification. Int. J. Neural Syst. **25**, 1550023 (2015). https://doi.org/10.1142/S0129065715500239
47. T. Zhang, W. Chen, M. Li, AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier. Biomed. Signal Process. Control **31**, 550–559 (2017). https://doi.org/10.1016/j.bspc.2016.10.001
48. F. Manzouri, S. Heller, M. Dümpelmann, et al., A comparison of machine learning classifiers for energy-efficient implementation of seizure detection. Front. Syst. Neurosci. **12** (2018). https://doi.org/10.3389/fnsys.2018.00043
49. U.R. Acharya, S.L. Oh, Y. Hagiwara, et al., Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput. Biol. Med. **100**, 270–278 (2018). https://doi.org/10.1016/j.compbiomed.2017.09.017
50. N. Sriraam, S. Raghu, K. Tamanna, et al., Automated epileptic seizures detection using multi-features and multilayer perceptron neural network. Brain Informatics **5** (2018). https://doi.org/10.1186/s40708-018-0088-8
51. A. Subasi, E. Erçelebi, Classification of EEG signals using neural network and logistic regression. Comput. Methods Prog. Biomed. **78**, 87–99 (2005). https://doi.org/10.1016/j.cmpb.2004.10.009
52. R. Abbasi, M. Esmaeilpour, Selecting statistical characteristics of brain signals to detect epileptic seizures using discrete wavelet transform and perceptron neural network. International Journal of Interactive Multimedia and Artificial Intelligence **4**, 33 (2017). https://doi.org/10.9781/ijimai.2017.456
53. M. Alfaro-Ponce, A. Argüelles, I. Chairez, Pattern recognition for electroencephalographic signals based on continuous neural networks. Neural Netw. **79**, 88–96 (2016). https://doi.org/10.1016/j.neunet.2016.03.004
54. Y. Wang, Z. Li, L. Feng, et al., Automatic detection of epilepsy and seizure using multiclass sparse extreme learning machine classification. Comput. Math. Methods Med. **2017**, 1–10 (2017). https://doi.org/10.1155/2017/6849360
55. R.P. Costa, P. Oliveira, G. Rodrigues, et al., Epileptic seizure classification using neural networks with 14 features, in *Knowledge-Based Intelligent Information and Engineering Systems*, ed. by I. Lovrek, R. J. Howlett, L. C. Jain, (Springer, Berlin/Heidelberg, 2008), pp. 281–288
56. E. Bou Assi, L. Gagliano, S. Rihana, et al., Bispectrum features and multilayer perceptron classifier to enhance seizure prediction. Sci. Rep. **8** (2018). https://doi.org/10.1038/s41598-018-33969-9
57. N.D. Truong, A.D. Nguyen, L. Kuhlmann, et al., Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. Neural Netw. **105**, 104–111 (2018). https://doi.org/10.1016/j.neunet.2018.04.018
58. I. Kuzovkin, R. Vicente, M. Petton, et al., Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. Commun. Biol. **1** (2018). https://doi.org/10.1038/s42003-018-0110-y
59. H. Khan, L. Marcuse, M. Fields, et al., Focal onset seizure prediction using convolutional networks. IEEE Trans. Biomed. Eng. **65**, 2109–2118 (2018). https://doi.org/10.1109/TBME.2017.2785401
60. X. Wei, L. Zhou, Z. Chen, et al., Automatic seizure detection using three-dimensional CNN based on multi-channel EEG. BMC Med. Inform. Decis. Mak. **18** (2018). https://doi.org/10.1186/s12911-018-0693-8
61. P. Mirowski, D. Madhavan, Y. LeCun, R. Kuzniecky, Classification of patterns of EEG synchronization for seizure prediction. Clin. Neurophysiol. **120**, 1927–1940 (2009). https://doi.org/10.1016/j.clinph.2009.09.002
62. P.W. Mirowski, Y. LeCun, D. Madhavan, R. Kuzniecky, Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG, in *2008 IEEE Workshop on Machine Learning for Signal Processing*, (IEEE, Cancun, 2008), pp. 244–249
63. A. Petrosian, D. Prokhorov, R. Homan, et al., Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG. Neurocomputing **30**, 201–218 (2000). https://doi.org/10.1016/S0925-2312(99)00126-5
64. S. Raghu, N. Sriraam, G.P. Kumar, Classification of epileptic seizures using wavelet packet log energy and norm entropies with recurrent Elman neural network classifier. Cogn. Neurodyn. **11**, 51–66 (2017). https://doi.org/10.1007/s11571-016-9408-y
65. L. Vidyaratne, A. Glandon, M. Alam, K.M. Iftekharuddin, Deep recurrent neural network for seizure detection, in *2016 International Joint Conference on Neural Networks (IJCNN)*, (IEEE, Vancouver, 2016), pp. 1202–1207
66. Z. Yu, D.S. Moirangthem, M. Lee, Continuous timescale long-short term memory neural network for human intent understanding. Front.

Neurorobot. **11** (2017). https://doi.org/10.3389/fnbot.2017.00042
67. K.M. Tsiouris, V.C. Pezoulas, M. Zervakis, et al., A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. Comput. Biol. Med. **99**, 24–37 (2018). https://doi.org/10.1016/j.compbiomed.2018.05.019
68. P. Thodoroff, J. Pineau, A. Lim, Learning robust features using deep learning for automatic seizure detection (2016). arXiv:160800220 [cs]
69. M. Li, W. Chen, T. Zhang, Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble. Biomed. Signal Process. Control **31**, 357–365 (2017). https://doi.org/10.1016/j.bspc.2016.09.008
70. I. Ullah, M. Hussain, E.-H. Qazi, H. Aboalsamh, An automated system for epilepsy detection using EEG brain signals based on deep learning approach. Expert Syst. Appl. **107**, 61–71 (2018). https://doi.org/10.1016/j.eswa.2018.04.021
71. E. Abdulhay, V. Elamaran, M. Chandrasekar, et al., Automated diagnosis of epilepsy from EEG signals using ensemble learning approach. Pattern Recogn. Lett. (2017). https://doi.org/10.1016/j.patrec.2017.05.021
72. B.H. Brinkmann, J. Wagenaar, D. Abbot, et al., Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. Brain **139**, 1713–1722 (2016). https://doi.org/10.1093/brain/aww045
73. U. Orhan, M. Hekim, M. Ozer, EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. Expert Syst. Appl. **38**, 13475–13481 (2011). https://doi.org/10.1016/j.eswa.2011.04.149
74. J. Martinez-del-Rincon, M.J. Santofimia, X. del Toro, et al., Non-linear classifiers applied to EEG analysis for epilepsy seizure detection. Expert Syst. Appl. **86**, 99–112 (2017). https://doi.org/10.1016/j.eswa.2017.05.052
75. T. Wen, Z. Zhang, Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. IEEE Access **6**, 25399–25410 (2018). https://doi.org/10.1109/ACCESS.2018.2833746
76. M.-P. Hosseini, H. Soltanian-Zadeh, K. Elisevich, D. Pompili, Cloud-based deep learning of big EEG data for epileptic seizure prediction, in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, (IEEE, Washington, DC, 2016), pp. 1151–1155
77. H. Daoud, M.A. Bayoumi, Efficient epileptic seizure prediction based on deep learning. IEEE Trans. Biomed. Circuits Syst. **13**, 804–813 (2019). https://doi.org/10.1109/TBCAS.2019.2929053
78. G. Zhu, Y. Li, P. Wen, et al., Unsupervised classification of epileptic EEG signals with multi scale K-means algorithm, in *Brain and Health Informatics*, ed. by K. Imamura, S. Usui, T. Shirao, et al., (Springer, Cham, 2013), pp. 158–167
79. S. Baldassano, D. Wulsin, H. Ung, et al., A novel seizure detection algorithm informed by hidden Markov model event states. J. Neural Eng. **13**, 036011 (2016). https://doi.org/10.1088/1741-2560/13/3/036011
80. C.E. Solorzano-Espindola, B. Tovar-Corona, A. Anzueto-Rios, Pediatric seizure forecasting using nonlinear features and Gaussian mixture hidden Markov models on scalp EEG signals, in *2018 15th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, (IEEE, Mexico City, 2018), pp. 1–6
81. M. Hejazi, A. Motie Nasrabadi, Prediction of epilepsy seizure from multi-channel electroencephalogram by effective connectivity analysis using granger causality and directed transfer function methods. Cogn. Neurodyn. (2019). https://doi.org/10.1007/s11571-019-09534-z
82. J.A. Dian, S. Colic, Y. Chinvarun, et al., Identification of brain regions of interest for epilepsy surgery planning using support vector machines, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, Milan, 2015), pp. 6590–6593
83. B. Elahian, M. Yeasin, B. Mudigoudar, et al., Identifying seizure onset zone from electrocorticographic recordings: A machine learning approach based on phase locking value. Seizure **51**, 35–42 (2017). https://doi.org/10.1016/j.seizure.2017.07.010
84. S.B. Tomlinson, B.E. Porter, E.D. Marsh, Interictal network synchrony and local heterogeneity predict epilepsy surgery outcome among pediatric patients. Epilepsia **58**, 402–411 (2017). https://doi.org/10.1111/epi.13657
85. I.A. Nissen, C.J. Stam, E.C.W. van Straaten, et al., Localization of the epileptogenic zone using interictal MEG and machine learning in a large cohort of drug-resistant epilepsy patients. Front. Neurol. **9** (2018). https://doi.org/10.3389/fneur.2018.00647
86. J. Jacobs, M. Zijlmans, R. Zelmann, et al., Value of electrical stimulation and high frequency oscillations (80-500 Hz) in identifying epileptogenic areas during intracranial EEG recordings. Epilepsia **51**, 573–582 (2010). https://doi.org/10.1111/j.1528-1167.2009.02389.x
87. E. Ben-Menachem, J.W. Sander, M. Privitera, F. Gilliam, Measuring outcomes of treatment with antiepileptic drugs in clinical trials. Epilepsy Behav. **18**, 24–30 (2010). https://doi.org/10.1016/j.yebeh.2010.04.001
88. R.G. Wither, S. Colic, C. Wu, et al., Daily rhythmic behaviors and thermoregulatory patterns are disrupted in adult female MeCP2-deficient mice. PLoS One **7**, e35396 (2012). https://doi.org/10.1371/journal.pone.0035396
89. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)

90. M. Brodie, S. Barry, G. Bamagous, J. Norrie, P. Kwan, Patterns of treatment response in newly diagnosed epilepsy. Neurology **78**(20), 1548–1554 (2012)
91. J. Liang, R. Lu, C. Zhang, F. Wang, Predicting seizures from electroencephalography recordings: A knowledge transfer strategy, in *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, (IEEE, Chicago, 2016), pp. 184–191
92. Q. She, B. Hu, H. Gan, et al., Safe semi-supervised extreme learning machine for EEG signal classification. IEEE Access **6**, 49399–49407 (2018). https://doi.org/10.1109/ACCESS.2018.2868713
93. M.K. Islam, A. Rastegarnia, Z. Yang, Methods for artifact detection and removal from scalp EEG: A review. Neurophysiol. Clin./Clin. Neurophysiol. **46**, 287–305 (2016). https://doi.org/10.1016/j.neucli.2016.07.002
94. I. Kiral-Kornek, S. Roy, E. Nurse, et al., Epileptic seizure prediction using big data and deep learning: Toward a Mobile system. EBioMedicine **27**, 103–111 (2018). https://doi.org/10.1016/j.ebiom.2017.11.032
95. L. Kuhlmann, A.N. Burkitt, M.J. Cook, et al., Seizure detection using seizure probability estimation: Comparison of features used to detect seizures. Ann. Biomed. Eng. **37**, 2129–2145 (2009). https://doi.org/10.1007/s10439-009-9755-5
96. R. Chai, Y. Tran, G.R. Naik, et al., Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network, in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, Orlando, 2016), pp. 4654–4657
97. Y. Zhao, B. Ahmed, T. Thesen, et al., A nonparametric approach to detect epileptogenic lesions using restricted Boltzmann machines, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD'16*, (ACM Press, San Francisco, 2016), pp. 373–382
98. B. Nakisa, M.N. Rastgoo, D. Tjondronegoro, V. Chandran, Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors. Expert Syst. Appl. **93**, 143–155 (2018). https://doi.org/10.1016/j.eswa.2017.09.062
99. T.T. Erguzel, S. Ozekes, O. Tan, S. Gultekin, Feature selection and classification of electroencephalographic signals: An artificial neural network and genetic algorithm based approach. Clin. EEG Neurosci. **46**, 321–326 (2015). https://doi.org/10.1177/1550059414523764
100. P. Mesejo, O. Ibáñez, E. Fernández-Blanco, et al., Artificial neuron–glia networks learning approach based on cooperative coevolution. Int. J. Neural Syst. **25**, 1550012 (2015). https://doi.org/10.1142/S0129065715500124
101. G.G. Somjen, H. Kager, W.J. Wadman, Computer simulations of neuron-glia interactions mediated by ion flux. J. Comput. Neurosci. **25**, 349–365 (2008). https://doi.org/10.1007/s10827-008-0083-9
102. V. Grigorovsky, B.L. Bardakjian, Low-to-high cross-frequency coupling in the electrical rhythms as biomarker for Hyperexcitable neuroglial networks of the brain. IEEE Trans. Biomed. Eng. **65**, 1504–1515 (2018). https://doi.org/10.1109/TBME.2017.2757878
103. A. van den Oord, S. Dieleman, H. Zen, et al., WaveNet: A generative model for raw audio (2016). arXiv:160903499 [cs]
104. Y.N. Dauphin, A. Fan, M. Auli, D. Grangier, Language modeling with gated convolutional networks (2016). arXiv:161208083 [cs]
105. S. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling (2018). arXiv:180301271 [cs]