This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

A Review on Machine Learning for EEG Signal Processing in Bioengineering

Mohammad-Parsa Hosseini *Senior Member, IEEE*, Amin Hosseini *Member, IEEE*, Kiarash Ahi *Member, IEEE*

***Abstract*—Electroencephalography (EEG) has been a staple method for identifying certain health conditions in patients since its discovery. Due to the many different types of classifiers available to use, the analysis methods are also equally numerous. In this review, we will be examining specifically machine learning methods that have been developed for EEG analysis with bioengineering applications. We reviewed literature from 1988 to 2018 to capture previous and current classification methods for EEG in multiple applications. From this information, we are able to determine the overall effectiveness of each machine learning method as well as the key characteristics. We have found that all the primary methods used in machine learning have been applied in some form in EEG classification. This ranges from Naive-Bayes to Decision Tree/Random Forest, to Support Vector Machine (SVM). Supervised learning methods are on average of higher accuracy than their unsupervised counterparts. This includes SVM and KNN. While each of the methods individually is limited in their accuracy in their respective applications, there is hope that the combination of methods when implemented properly has a higher overall classification accuracy. This paper provides a comprehensive overview of Machine Learning applications used in EEG analysis. It also gives an overview of each of the methods and general applications that each is best suited to.**

***Index Terms*—Machine Learning, EEG, Survey, Medical Applications, Signal Processing, Signal Analysis.**

I. Introduction

Electroencephalography (EEG) is a method of testing electrical signals in the brain. It is often applied as a technique for data analysis such as time and frequency series analysis. The brain's neurons contain ionic current, which creates voltage fluctuations that EEG can measure. This electrical activity is spontaneous and recorded over a period of time from many scalp electrodes to form an EEG signal. [24] Traditionally, EEG signals are taken on the surface of the scalp, but there also exists iEEG signals, which are taken inside the brain. In this paper, we will be focusing primarily on conventional scalp EEG signals.

Conventionally, EEG recordings may be obtained by connecting electrodes to the scalp with the use of a conductive gel. A differential amplifier is then used to amplify each active electrode compared to the reference before it is sent through an anti-aliasing filter. Finally, this filtered signal is converted with an analog-to-digital converter.

[1] M.P. Hosseini is with Bioengineering Department, Santa Clara University, CA, USA and AI Research, Silicon Valley, CA, USA mhosseini@scu.edu, parsa@cac.rutgers.edu

[3] A. Hosseini is with Electrical and Computer Engineering Department, Azad University, Central Tehran Branch, Tehran, Iran ami.hosseini.eng@iauctb.ac.ir

[2] K. Ahi is An alumnus of the University of Connecticut, CT, USA kiarash.ahi@uconn.edu

Clinically, EEG signals are used primarily to diagnose and treat various brain disorders such as epilepsy, tremor, concussions, strokes, and sleep disorders. More recent applications of EEG include using machine learning as a method of analysis. In particular, there is much research on epileptic seizure detection and sleep disorder research in combination with machine learning. Additionally, there is also a growing interest in studying EEG signals for gaming to control and manipulate objects using brainwaves due to EEG monitoring for brain activity during tasks [40].

EEG signals were first discovered in 1875 by Richard Caton, a physician who was studying electrical brain activity in rabbits and monkeys. Later on in the 1900s, the first human EEG recordings were made and studied with a focus on absent seizures. In the 1930s, epileptic spikes and seizure patterns became noticed, which sparked EEG as a new field of interest.

EEG waveforms vary based on the band, which denotes the frequency range. The delta band is the slowest wave with the highest amplitude, having a frequency range below 4Hz. For adults, it is located frontally, while for children it is located posteriorly. The theta band is between 4 to 7Hz and is most common in young children while signifying drowsiness or arousal in adults. This band tends to spike due to an active inhibition of a movement or response. The alpha band is between 8 to 14Hz, and it is correlated to eye muscle movements. It is located on both sides of the head's posterior regions. The beta band is above 14Hz and is correlated with general motor behavior. It is located on both sides of the head's frontal regions [48].

Some of the advantages of using EEG compared to other techniques to study brain function are low costs, tolerance to motion from subjects, and no radiation exposure risks. Some of the disadvantages of using EEG include low spatial resolution and poor signal-to-noise ratio.

II. Machine Learning Methods for EEG

A. Overview

Machine learning is the use of a set of mathematical models and algorithms to gradually improve the performance of a singular task. It takes training data sets as input to use as a guide for making estimates without being specifically programmed to. The tasks vary widely in this space and can be categorized into two main groups: supervised and unsupervised learning. Unsupervised learning is the case when the algorithm builds a pattern of recognition from a data set containing only inputs with no set outputs. Supervised learning has a subsection being semi-supervised learning. They are identical in the sense that they both learn from data sets

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

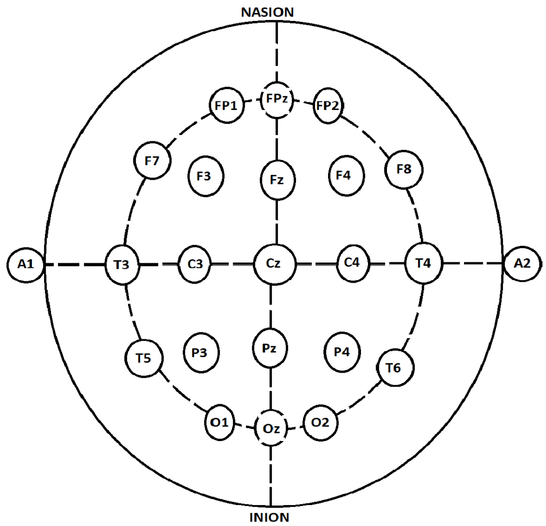

Fig. 1. Electrode placements in 10–20 system for EEG acquisition which describe the location of scalp electrodes in an EEG test.

with given inputs and known outputs with the exception that semi-supervised has parts of the data set missing. Supervised learning is primarily used in applications of classification and regression while unsupervised learning lends itself to feature learning and the inverse, dimensionality reduction. This paper will discuss some of the most popular machine learning methods and categorize them based on the type of learning with some practical applications in EEG.

EEG signals can be used as indicators of harder to detect medical conditions with the assistance of machine learning methods. In Fig. 2 the applications of machine learning on EEG signals are shown based on supervised and unsupervised learning. Supervised learning develops a predictive model using both input and desired output data is categorized to classification and regression which produce discrete and continuous accordingly. Unsupervised learning develops a predictive model using just input data is categorized to clustering and dimensionality reduction which produce discrete and continuous accordingly.

Fig. 3 describes the general flow of how machine learning is implemented to get the desired classification of the data sets. The first step is signal acquisition. This is essentially the raw data, unedited. Pre-processing involves the removal of noise and other outliers in the data set. Feature extraction determines the spectrum of the data point groupings and what features they correspond to. Feature selection is the isolation of the desired classifiers that the machine learning method will be testing for the following training. Machine learning training involves the use of training data sets, whether with or without known outputs to refine the classification method. Lastly, the testing phase is the processing of true test data sets and comparing the overall accuracy of the desired feature.

B. Regression

Regression modeling is a popular tool in statistics because it is a simple way to create a functional relationship between variables. Various types of regression include: univariate and multivariate for quantitative response variables; simple and multiple for predictor variables; linear for linearly transformable data; nonlinear for nonlinearly transformable data; analysis of variance for qualitative variable predictors; analysis of covariance for the combination of qualitative and quantitative variable predictors; and logistic for qualitative response variable [90].

Legendre and Gauss first applied regression using the Method of Least Squares. This method makes approximations by summing the squares of each equation residual to best fit the data, and it is applied in Linear Regression. as shown in the equation below.

$$_i = B_0 + B_1 x_i + e_{i,i=1,\ldots,n} \tag{1}$$

Linear Regression is one of the most common regression techniques. In this model, the parameters are specified in the form of a linear combination, while each independent variable is not necessarily linear. Multiple linear regression is similar, except that there are several independent variables rather than just one. When the parameters are not linear, nonlinear regression must be used. This also uses a sum of squares technique, though it uses an iterative procedure to minimize the function.

C. SVM

SVM is a subcategory of supervised learning used for analyzing data for classification and regression analysis. The purpose is to map points in space such that the examples of the target categories are divided by the largest possible margin. This allows for SVM to have a general lower generalization error as a classifier [43]. The objective is to find a hyperplane or set of hyperplanes in an N-dimensional space. Support vectors are data points that are closer to a given hyperplane. They maximize the margin of the classifier by changing the position and orientation of the hyperplane. Additionally, within this space, it is also possible that the points are not separable linearly due to the position of the data. SVM is capable of utilizing generated kernel functions or more commonly known as "kernel trick" to the data set to remedy this issue. This trick involves the transformation of the existing algorithm from a lower-dimensional data set to a higher one. The amount of information remains the same, but in this higher dimensional space, it is possible to create a linear classifier. Several K kernels are assigned to each point which then help determine the best fit hyperplane for the newly transformed feature space. With enough K functions, it is possible to get precise separation. The only major concern is overfitting. [120]. Fig. 4 depicts a sample of data separation in both 2D and 3D.

$$\overrightarrow{w} \cdot \overrightarrow{x} - b = 1, -1 \tag{2}$$

Linear SVM classifier with hard margin

$$W(\alpha) = -\sum_{i=1}^{l} \alpha_i + \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_i \mathbf{x}_i \mathbf{x}_j \tag{3}$$

Kernel trick equation minimizing W subject to:

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

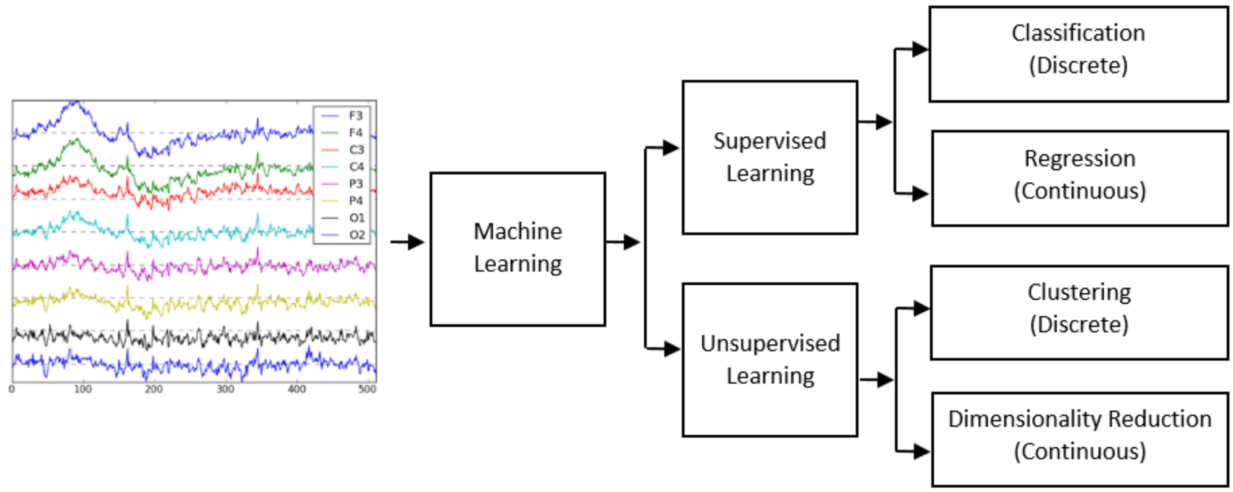

Fig. 2. Machine learning applications on EEG have been developed based on supervised and unsupervised learning in the literature. Supervised learning is categorized to classification and regression which produce discrete and continuous accordingly. Unsupervised learning is categorized to clustering and dimensionality reduction which produce discrete and continuous accordingly.

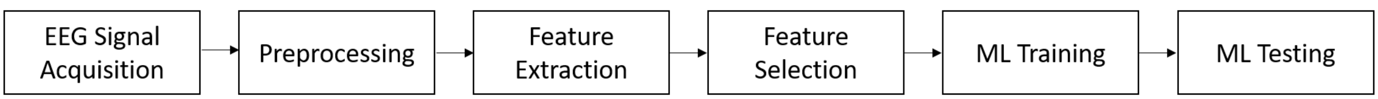

Fig. 3. The overall steps for EEG analysis by machine learning include preprocessing, feature extraction, feature selection, model training, model testing.

TABLE I
REGRESSION MODELS APPLIED FOR EEG ANALYSIS

Author(s)	Machine Learning Method	Application	Data Set	Results
Rajaguru et al., 2017 [95]	Logistic Regression	Epilepsy Classification	20 patients	Performance Index 91.39% and Accuracy 95.88%
Kim et al., 2014 [68]	Non-Linear Regression	Reconstruction of hand movements from EEG signals	4 subjects	not listed
Dora et al., 2016 [27]	Linear Regression	Robust ECG Artifact Removal	not listed	Accuracy 98.11%
Rajaguru et al., 2017 [94]	LRGMM	Epilepsy Classification	not listed	Accuracy 97.91%
Murakami et al., 2015 [82]	Logistic Regression	Motion Discrimination	3 subjects	Single Threshold Processing Accuracy 77.0%
Li et al., 2015 [75]	Logistic Regression	Ocular Artefacts Correction Method for Discriminative EEG Analysis	68 subjects	not listed
Dong et al., 2013 [26]	Linear Regression	Visual Attention Modeling	6 subjects	not listed
Jain et al., 2016 [60]	Auto-Regression	Fatigue Detection and Estimation	14 subjects	not listed
Hu et al., 2016 [56]	Auto-Regression	EEG Authentication System	not listed	Accuracy 92.93%
Hamilton et al., 2015 [35]	EBMAL Regression	Offline EEG-Based Driver Drowsiness Estimation	16 subjects	not listed
Struck et al., 2017 [108]	Logistic Regression	Seizure probability in hospitalized patients	4772 Participants	area under the curve of 0.819 and average calibration error of 2.7% (95% CI, 2.0%-3.6%)
Roy et al., 2018 [98]	Logistic Regression, neural networks, CNN, RNN	Automatic Abnormal EEG Identification	1488 abnormal, 1529 normal EEG	Deep gated RNN achieve 3.47% better performance than previously reported results

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

$$\sum_{i=1}^{l} y_i \alpha_i = 0 \quad \underset{0 \le \alpha_i \le C}{} \tag{4}$$

D. KNN K-Nearset Neighbours

KNN is one of the supervised machine learning algorithms. In supervised learning, the relationship between the input and output is already established for the training data set, i.e for a given input the output is already known. Supervised learning is categorized into regression and classification. KNN can be used for both classification and regression. The input for both classification and regression is the same but the output differs respectively. Example input-output pairs are used for predicting the output for untrained data set. KNN classifies the input based on the classification of its K neighbors. To find the nearest neighbors, Euclidean distance or Mahalanobis distance is calculated from the input to all known data points. After the distance is calculated, K nearest neighbors are selected. It then classifies the input based on similarities between input and its K- neighbors. The selection of K is based on the size of the data set. The square root of the size of the data set is taken and if the result is an even number then 1 is added or subtracted from it. The result is then established as K for that data set. K is selected to be an odd number to avoid bias in the prediction of input.

E. ANN

Neural networks, commonly called the Artificial neural networks in the computing world, is a mathematical model very similar to the structure of neural networks seen in a human brain. To understand how the model works, researches have put forth several theories and examples showing the interaction between different layers of the neural networks to convert the given input into the desired output [5].

Imagine you are at a bar, and looking at the menu to order a nice beer. Your favorite is IPA and as soon as you see that on the list, you order it. So what happened in your brain is that you provided multiple inputs for beer choice to your brain's neural network, IPA choice had a preferable weight as that being your most favorite beer; the brain made a decision and gave you the output. This is a basic example of how neural networks operate. The architecture of the model shows the decision-making process which involves much deeper layers of interaction that lie between the input and the output layer. Fig. 5 shows the classification of different layers, of ANN.

From the Mathematical model of neural networks, it is evident that any given input at the first layer undergoes a function put forth by the algorithm which will narrow down multiple combinations and options, to depict the desired output. This can be observed in Fig. 6. For ANN, the classification technique can be brought about by:

Summation of input-weight product and Bias

$$\sum_{i=1}^{n} (w_i x_i) + bias \tag{5}$$

Activation Layer

$$Output = f(x) = \begin{cases} 1 if \sum wx + b \ge 0 \\ 0 if \sum wx + b < 0 \end{cases} \tag{6}$$

Compared to other supervised and unsupervised learning techniques, neural networks deliver the most optimum results, by segregating data to the deepest level intended. The modern generation has been successful in identifying and documenting many techniques involving the use of Artificial neural networks for analyzing and classifying the EEG signals and simulate the results for serving different medical conditions [85]. EEG signals or commonly called as Electroencephalogram are set of highly complex signals which are studied to find abnormalities in a human brain activity [101]. The latter is measured when neurons in a brain start firing which produces an electric current within the dendrites. This electric current, therefore, produces a magnetic field all over the scalp which is recorded through a signal generator that gives the activity at any given location in a brain. These signals need to be studied and segregated for the type of abnormality we are focusing on, and for that, we are discussing different Machine learning models that can be used for different EEG applications. The need for proper classification and analysis gets itself the priority when we compare the traditional method of EEG analysis. The standard procedure involves time-frequency analysis and spectrogram image processing for analyzing EEG signals. The characteristic waveform of EEG fall on specific frequency bands namely: alpha (8 to 15Hz), beta (14 to 30HZ), theta (4 to 8Hz) and delta (< 4Hz), and previous classification methods like FFT (Fast Fourier transforms) have been characterized to have very high noise sensitivity which limits the ability to analyze the signals effectively. This brings the spotlight to implement Neural networks or ANN for measuring the activity of EEGs effectively. The current section discusses how ANN can be applied for the synthesis of EEG data, and classify different signals based on their attributes and features. As we know by now, ANN works by analyzing the data set provided against multiple possibilities and connections to deliver an optimum output as required [31]. The main characteristics of the network architecture depend on a few highlighted factors show below:

1. Type of input signal (signal dimensionality and behavior)
2. Topology of connections
3. Interactions between different network layers
4. Mode of operation
5. Output interpretation

As each application varies and has to have a specific approach – Long term or short term EEG segment analysis, real-time process or time-delayed process, type of EEG channel analysis (single or multiple) – which can be easily targeted and synthesized using ANN. Once the EEG signals are converted to waveforms in user-friendly GUIs, the classification of these signals happens with ANN, with the selection of a particular type of network for a specific use case – Feedforward back-propagation, Radial basis function, Recurrent Neural networks. It is important to see how different types of ANN operate and the architecture which facilitates that operation.

1. Feedforward Neural Networks: This is a type of network where data flows in only one direction, starting from the input nodes, passing through the hidden nodes and arriving at the output nodes. This network ensures no loop or cycle formation, making the information flow in a specific direction

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

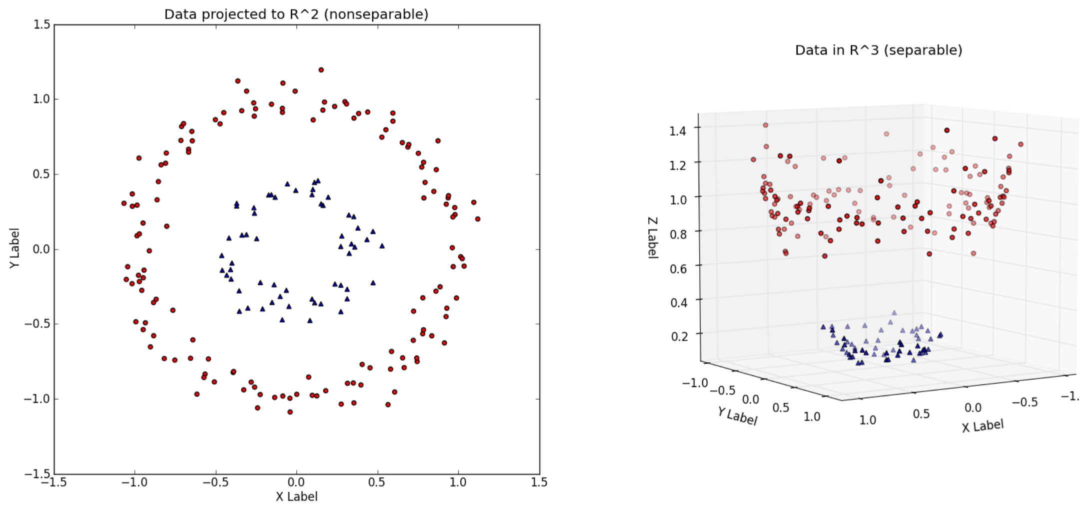

Fig. 4. Higher dimension kernel separation. The kernel trick involves the transformation of the existing algorithm from a lower dimensional data set to a higher one.

TABLE II
SUPPORT VECTOR MACHINE APPLICATIONS WITH EEG

Author(s)	Machine Learning Method	Application	Data Set	Results
Jalilifard et al., 2016 [64]	SVM	Emotion Classification	19 patients	Accuracy 96.83%
Sai et al., 2018 [100]	SVM	EEG artifact removal	11 patients	Accuracy 99.1%
Zhang et al., 2017 [122]	SVM	Seizure Detection of EEG	Not Reported	Accuracy 98.1%
Torabi et al., 2017 [111]	SVM, KNN	Multiple Sclerosis Detection	Not Reported	Accuracy 93.08%
Aghajani et al., 2017 [3]	SVM	EEG Measuring Mental Workload	17 patients	Accuracy 90.0%
Amin et al., 2017 [9]	SVM	EEG Classification	48 practice patterns	Accuracy 98.57%
Jaiswal et al., 2017 [62]	SVM	Epilepsy Detection of EEG	7 patients	Accuracy 100%
Huang et al., 2017 [57]	SVM	Multimodal Facial Recognition	40 patients	Accuracy 82.75%
Ahani et al., 2014 [4]	SVM	Meditation EEG Detection	34 patients	Accuracy 85%
Mumtaz et al., 2018 [81]	SVM	Alcohol Use Disorder Detection	30 patients and 30 normal	Accuracy 98.0%
Dian et al., 2015 [25]	SVM	Identification of brain regions of interest for epilepsy surgery planning	6 patients	Accuracy Proposed method is scalable across multiple patients exhibiting Engel Class I outcomes
Zhuang et al., 2018 [123]	SVM	Emotion recognition	30 participants	research lays a substantial foundation for real-time recognition of comprehensive endogenous emotion%
Beganovic et al., 2018 [13]	SVM, KNN	epileptic seizure occurrence	20 patients, dimensionality reduction/selection %	

only. Fig. II-E shows the architecture for Feedforward network mechanism

2. Radial basis function: In the field of artificial neural networks and mathematical modeling, RBF is a type of ANN which makes use of radial basis functions (An arbitrary real-valued function, the value of which is determined by functions location from the origin). Thus, the network determines the output by a linear combination of RBF of the inputs and parameters given for the neurons. As shown in Fig. 8 the structure operates by summing the centers/widths of the points with the associated weights to get us the final output.

A typical RBF is a Gaussian distribution, in case of a scalar input, and is given by:

$$h(x) = exp(\frac{-(x-c)^2}{r^2}) \tag{7}$$

Where c is the center, and r is the radius parameters. A Gaussian RBF distribution decreases as the distance from the center increases.

For a multiquadric RBF with a scalar input can be shown as:

$$h(x) = \frac{\sqrt{r^2+(x-c)^2}}{r} \tag{8}$$

In this case the Gaussian RBF increases with increase in the distance from the center.

3. Recurrent Neural Networks: As the name suggests, RNN is a type of Artificial Neural Network which has connections between different nodes, with a specific assigned direction for output flow to a specific node. Here, the flow of data can form loops and cycles to feed the data back to a specific node as intended. This technique is illustrated in Fig. 9 which shows the backpropagation of information from one layer to another and to a specifically intended node.

To understand the working of RNN it is important to define the transitions from one previous state to a new state. Let Xt be the input vector, Ht be the new state, and Ht-1 be the previous state. RNN is observed to be a function of the input vector and the previous state, which will land us to the new

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

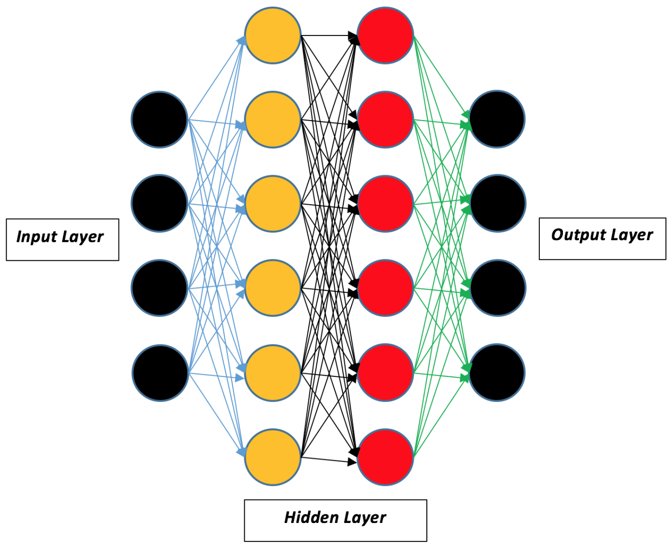

Fig. 5. A feedforward neural network with input layer, hidden layers, and output layer.

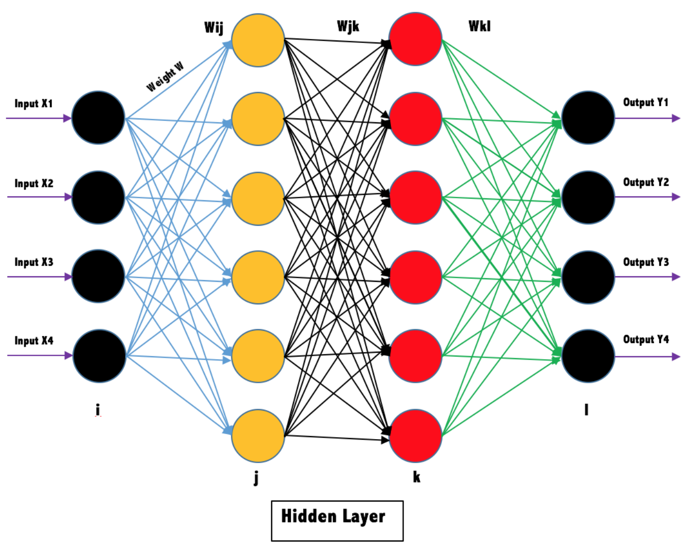

Fig. 6. Structure breakdown for ANN with three types of nodes, input nodes, hidden nodes, and output nodes.

state Ht. We can represent a simple Vanilla version of the RNN by obtaining the weight function Fw and implementing that to find the output function Yt. This can be represented as follows:

$$h_t = f_w(h_{t-1}, x_t) \tag{9}$$

$$h_t = tanh(W_{hh}.h_{t-1} + W_{xh}.x_t) \tag{10}$$

By applying the tan hyperbolic function the dot product of associated weights from previous states and the dot product of associated weights and input state, we shall have the value of the new state. We can have the final output function as:

$$y_t = W_{hy}.h_t \tag{11}$$

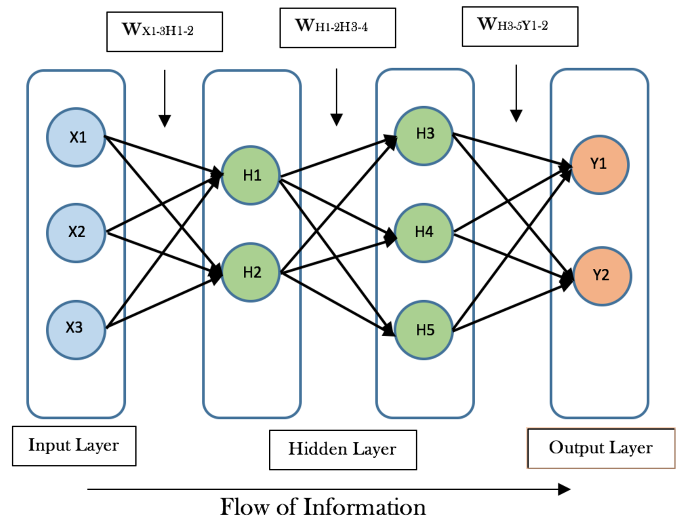

Fig. 7. Feedforward Neural Network. There are two directions for information flows, forward propagation and backpropagation. Forward propagation is used in the prediction time while backpropagation is used for adjusting the weights to minimize loss.

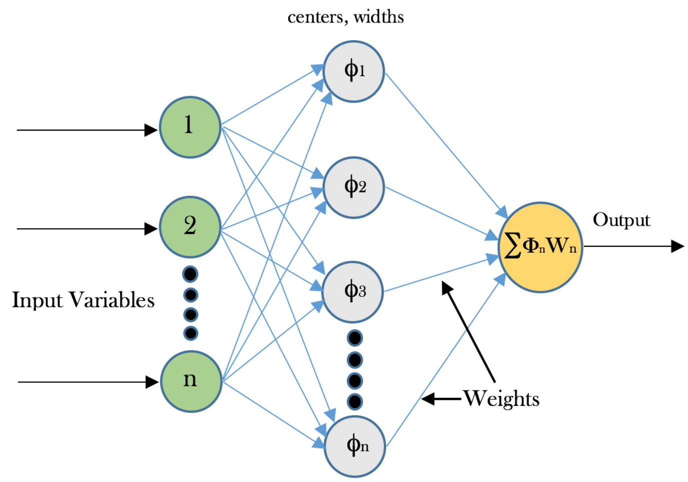

Fig. 8. A radial basis function network is an ANN which uses radial basis functions as activation functions. A linear combination of radial basis functions of the inputs and the parameters of neurons is used for the output of the network. These structures have many applications such as time series prediction, classification, and function approximation.

F. Naive Bayes

Naive Bayes classifier is a popular text categorization method that applies Bayes' theorem to separate data based on simple trained features. Essentially, the model assigns labels as feature vectors within a finite set. While simple in nature, with adequate pre-processing it can match more advanced methods such as SVM discussed above. The one disadvantage of the naive Bayes method is that it considers all of the feature vectors as independent from one another regardless of any real correlation. The main advantage of it is that it only needs a small number of training data sets to begin correctly estimating the parameters necessary for classification. Several models can be implemented for the Bayes method. The most common of which is the probabilistic model. In this model, the features are

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

TABLE III
ARTIFICIAL NEURAL NETWORKS APPLICATION FOR EEG ANALYSIS

Author(s)	Machine Learning Method	Application	Data Set	Results
Hramov, Alexander E., et al., 2017 [53]	ANN	Perceptual Interpretations of a Bistable Image	Not Reported	Accuracy 95%
Tzallas, A. T., Tsipouras, M. G. et al. 2007 [114]	ANN	Automatic seizure detection .	Not Reported	Accuracy 97.72%
Sharma, A.,Tewari R. P., et al. 2018 [103]	ANN	Epileptic seizure anticipation	Not Reported	Accuracy 92.3%
Saini, J. and Dutta, M., et al. 2018 [101]	ANN	Epilepsy classification using optimized artificial neural network.	100 Samples	Accuracy 99.3%
Chiarelli, Antonio Maria, et al. 2018 [20]	ANN	Deep learning for hybrid EEG.	15 Participants	Accuracy 99.3%
Lee, Y. H., Hsieh, Y. J., Shiah, Y. J., et al. 2017 [73]	ANN, SVM	EEG Cross section Analysis	10 Samples	Accuracy 98%
Guo, L., Rivero, D., Pazos, A., et al. 2010 [33]	ANN	Epileptic seizure detection	Public	Not Reported
Ogulata, S. N., Şahin, C., Erol, R., et al. 2009 [85]	ANN	Classification of primary generalized epilepsy by EEG signals..	4 groups	Accuracy 78-98%
Srinivasan, V., Eswaran, C., Sriraam, N., et al. 2007 [107]	ANN	Approximate entropy-based epileptic EEG detection.	Public	Accuracy 100%
Ghosh-Dastidar, S., Adeli, H., et al. 2009 [31]	ANN	Multiple spiking neural networks for application in epilepsy and seizure detection.	Not Reported	Accuracy 90-94%

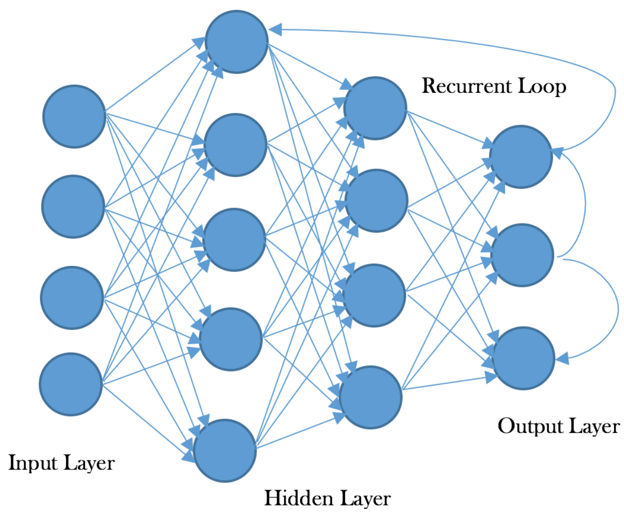

Fig. 9. Recurrent Neural Network where connections between nodes form a directed graph along a temporal sequence. It makes previous outputs to be used as inputs.

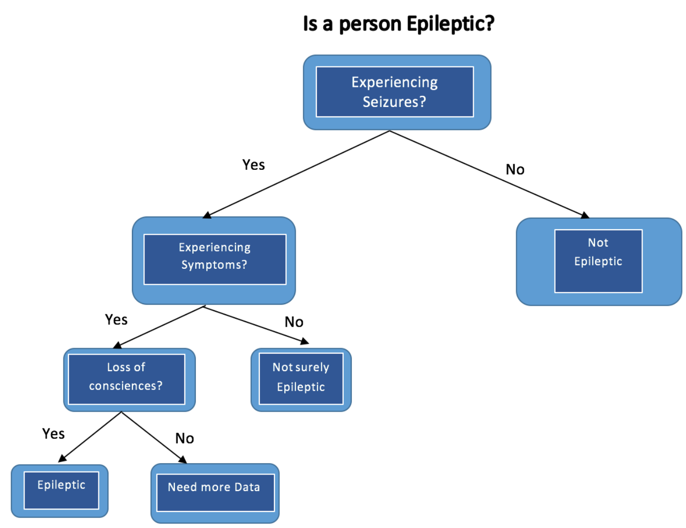

Fig. 10. Example for decision tree technique to determine a health condition

represented by vectors and it assigns probabilities to a given outcome or case. Event models can be separated into 2 main classes, Gaussian Naive Bayes and Multinomial Naive Bayes. In a data set with continuous values, a good assumption would be that it follows a Gaussian distribution. Using this method the Bayes method assigns probabilities based on the curve. A multinomial event model represents the frequencies of specific events spawned from multinomials, often as a histogram. A potential concern is when a feature does not occur in the data set at all. This causes the multiple of all the estimates to be zero. It can be corrected with a pseudocount to smooth out any outliers in the data set [97].

$$P(c|x) = \frac{P(x)|(c)}{P(x)} \tag{12}$$

The probabilistic Naive Bayes Model

$$P(x = v \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}} \tag{13}$$

The Gaussian Naive Bayes Model

G. Decision Tree and Random Forest

Decision trees use questions about the features of an item to classify data. Each question can be represented as a node, in which there is a child node for each answer to that question. This creates a hierarchy, in other words, a tree. The most basic tree would be a binary one in which each question results in a yes or no answer. Therefore there is a yes and no child node for each parent node question. Data is sorted through the tree by starting at the top-most node, also known as the root, and maneuvering its way down to the leaf, or the node that has no children. The path taken is dependent on the data's features. Once the data reaches the leaf, it can be classified under the class associated with that particular leaf [69].

The advantages of decision trees are that they are simplistic and can be easily combined with other techniques for decision making. The disadvantages of decision trees are that they are somewhat unstable as well as inaccurate, especially with varying level sizes which cause biases towards larger levels.

In the study of machine learning, and different classifying and distribution methods, we come across the Random Forest technique, which can be used for both data classification and

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

TABLE IV
NAIVE BAYES APPLICATIONS WITH EEG

Author(s)	Machine Learning Method	Application	Data Set	Results
Amin et al., 2017 [9]	Naive Bayes	EEG Classification	48 practice patterns	Accuracy 81.07-91.60%
Fan et al., 2015 [29]	Naive Bayes	EEG autism detection	16 patients	Accuracy 65-76%
Rytkönen et al., 2011 [99]	Naive Bayes	Sleep Scoring	2 humans and 30 animals	Accuracy 92%
Biswal et al., 2015 [16]	Naive Bayes	Automated Information Extraction	42,972 reports	Accuracy 97.53%
Mumtaz et al., 2018 [80]	Naive Bayes	Major Depressive Disorder	30 patients and 30 normal	Accuracy 93.6%
Fallani et al., 2011 [28]	Naive Bayes	Subject Recognition	50 subjects	Accuracy 78-89%
Laton et al., 2014 [71]	Naive Bayes	Schizophrenia Subject Recognition	54 patients and 50 normal	Accuracy 79.8%
Bigdely et al., 2008 [15]	Naive Bayes	Brain Activity Classification	7 subjects	Accuracy 87%
Sharmila et al., 2017 [105]	Naive Bayes	Epilepsy Detection	Not Reported	Accuracy 98.6%
Biswal et al., 2015 [16]	Naive Bayes	Automated information extraction from free-text EEG reports	3277 documents	Accuracy The average [95% CI] area under the receiver operating curve was 99.05 [98.79, 99.32]% for detecting reports with seizures, and 96.15 [92.31, 100.00]% for detecting reports with epileptiform discharges
Gao et al., 2016 [30]	KNN, naive Bayes, SVM	automatic sleep scoring in mice	16 mice	At 1% rejection rate, the algorithm matches the accuracy of a human scorer
Sharmila et al., 2018 [104]	Naive Bayes, SVM	detection of epileptic seizure	Not Reported	Accuracy 100%
Page et al., 2015 [88]	Naive Bayes, SVM, KNN	ultra-low power feature extraction and classification system for wearable seizure detection	10 Patients	Accuracy 100%
Combrisson et al., 2015 [21]	Naive Bayes, SVM	The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy	Not Reported	Accuracy 70%
Jirayucharoensak et al., 2014 [65]	Naive Bayes, SVM, Deep Learning Network	emotion recognition	32 Subjects	Accuracy 49.52%

regression operations. As the name suggests, Random Forest operates by producing a multitude of decision trees and trained by performing bagging operation to combine multiple decision trees or models to arrive at a more stable and accurate data prediction. Random Forest creates additional randomness to the data being structured; i.e. instead of finding the most important feature from the given set, it operates to find the best feature among a random set of a defined subset of features. This results in a more diverse and better result model.

One of the most recognized and well-known application for EEG analysis is for diagnosing epileptic seizures. Usually, the diagnosis is carried out during an epileptic attack, however, the current process involves doctors who diagnose the problem during a post-attack period which makes it a difficult process and often leads to wrong conclusions. This created an opening for analyzing the condition through the generation of EEG signals, by recording the activity of the brain, by measuring the discharged electric signals from the neurons, and calibrating those results in a waveform. Though the method may seem practical and computational, this is known to be one of the most sophisticated processes to arrive at accurate results. This has led to the development and implementation of Machine Learning models to analyze the EEG signals and classify them according to the most appropriate medical condition under study. Earlier we discussed how different models contribute towards this cause, and how those models interact with the given set of signals. But when it comes to applying the Random Forest model, the process is different, where the classification happens by a majority voting method, for the decision trees which have been formed. The process of randomization takes in not only the sample data provided, but also on the predictor variables, which leads to a collection of different sized classification trees. This highly depends on the number of random variables which will be randomly selected to form the classification tree. One of the many trees describes a series of questions pertaining to the condition under study, where the next set of questions depend on the answer from the query. The solution from all the trees is summed up and classification happens through a majority voting where the best suitable classification is chosen. However, if the trees are found to be unstable, where minor changes in the data set can change the whole decision tree, we might end up with a wrong classification. Combining the RF model with other Machine Learning models have shown an improvement in analyzing EEG signals more precisely and accurately.

H. Ensemble Learning

Ensemble learning is a supervised learning algorithm. As the name suggests, ensemble learning ensemble's many different algorithms to make a model that gives a better predictive performance. The general idea is to improve the overall performance by combining decisions received from different multiple models. It is based on the concept of diversity, more diverse models are considered for obtaining the results for the same problem in comparison to single models. This gives a set of hypotheses which can be combined to gain better performance. All the single models are called as base learners when combined are called as an ensemble. The ensemble is mostly better than the base learners from which the ensemble is made. Ensemble learning can be used in the fields of medicine, fraud detection, banking, malware and intrusion detection, face and emotion recognition, etc.

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

TABLE V
REVIEW ON DECISION TREE AND RANDOM FOREST

Author(s)	Machine Learning Method	Application	Data Set	Results
Rajaguru et al., 2017 [96]	SDT	Epilepsy Classification	not listed	Accuracy 96.83%
Ishfaque et al., 2013 [58]	DT	Brain Computer Interface	not listed	Accuracy 81.6%
Jakaite et al., 2010 [63]	DT	Newborn Brain Maturity	200 patients and 100 normal	Accuracy 86.5%
Anastasiadou et al., 2017 [10]	Random Forest	Scalp Recordings for Automatic Muscle Artifact Detection and Removal	not listed	not listed
Hu et al., 2018 [55]	GBDT	Automated Driver Fatigue Detection	22 subjects	Accuracy 94.0%
Vijayakumar et al., 2017 [116]	Random Forest	Quantify and Characterize Tonic Thermal Pain	not listed	Accuracy 89.45%
Le et al., 2017 [72]	Random Forest	Surface and intracranial EEG spike detection	17 scalp patients and 10 intracranial	Accuracy 62% recall and 26% precision for surface EEG subjects and 63% recall and 53% precision for intracranial EEG subjects
Bentlemsan et al., 2014 [14]	Random Forest	Motor Imagery Classification	9 subjects	Kappa 0.59
Wang et al., 2013 [117]	Random Forest	Classification of Neonatal Amplitude-Integrated EEG	209 normal and 73 abnormal infants	Accuracy 91.46%
Weichwald et al., 2014 [118]	Random Forest	Decoding Index Finger Position from EEG	12 subjects	Accuracy 12.29%
Hamilton et al., 2015 [35]	Boosted Rotational Forests	Eye State Prediction	not listed	Accuracy 97.4%
Bose et al., 2017 [18]	Random Forests	Seizure detection	5 patients	Accuracy 98%

TABLE VI
A REVIEW ON ENSEMBLE LEARNING STATE OF ARTS

Author(s)	Machine Learning Method	Application	Data Set	Results
Prabhakar et al.,2015 [89]	KNN	Diagnosis of multiple sclerosis detection	Not reported	Accuracy For the direction-based and the color-luminance-based tasks, maximum classification performances were 93.08 and 79.79% respectively
Gunay et al., 2018 [32]	KNN, Naïve Bayes	Epilepsy Disorder	500	Accuracy 73% for KNN and 92% for Naive Bayes
ozerdem et al., 2017 [87]	KNN,SVM	Emotion recognition	32 healthy subjects	72.92%
Sharmila et al.,2017 [105]	PCA, LDA with K-NN	Wavelet-based feature extraction for classification of epileptic seizure EEG	Not reported	Accuracy PCA, LDA with K-NN achieves 98.5% and 100%
Manjusha et al.,2016 [77]	K-means, KNN	Performance analysis of KNN classifier and K-means clustering for robust classification of epilepsy	20 Patients	Accuracy A high Quality value of 22.37 with K-means clustering and a low value of 18.02 are obtained with KNN classifier
Tuyisenge et al.,2018 [113]	ensemble learning	Automatic bad channel detection	206 patients	Accuracy 99.77% for 110 patients
Hosseini et al., 2018 [48]	ensemble learning	EEG classification	8 patients	Accuracy Using leave-one-out cross-validation, the accuracy, sensitivity, specificity, and both false positive and false negative ratios of the proposed method were found to be 0.97, 0.98, 0.96, 0.04, and 0.02, respectively.
Al Zoubi et al., 2018 [6]	stack-ensemble learning	Predicting Age From Brain EEG Signals	468 Healthy and 297 female patients from Tulsa-1000	Accuracy The stack-ensemble age prediction model achieved R2 = 0.37 (0.06), Mean Absolute Error (MAE) = 6.87(0.69) and RMSE = 8.46(0.59) in years
Antoniades et al., 2018 [11]	ensemble learning	Mapping Scalp	Intracranial EEG (iEEG) data	Accuracy classification accuracy of 68% an increase of 6% over the previously proposed linear regression mapping
Hassan et al., 2016 [37]	ensemble learning	Automatic identification of epileptic seizures	segments of EEG signals	Accuracy Proposed seizure detection scheme performs better than the existing works in terms of accuracy, sensitivity, specificity, and Cohen's Kappa coefficient
Meyer et al., 2014 [78]	random forest ensemble classifier	Predicting motor learning	6 patients	Accuracy learned models successfully generalized to novel subjects
Smart et al., 2015 [78]	Smart et al., 2015	Semi-automated patient-specific scalp EEG seizure detection	24 patients	classification via centroid-based clustering methods, such as k-means and k-mediod algorithms, or agglomerative clustering methods appear best suited for scalp EEG seizure-detection applications
Iturrate et al., 2010 [59]	reinforcement learning	Robot reinforcement learning using EEG-based reward signals	5 classes	it is possible to apply RL using EEG based reward signals

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

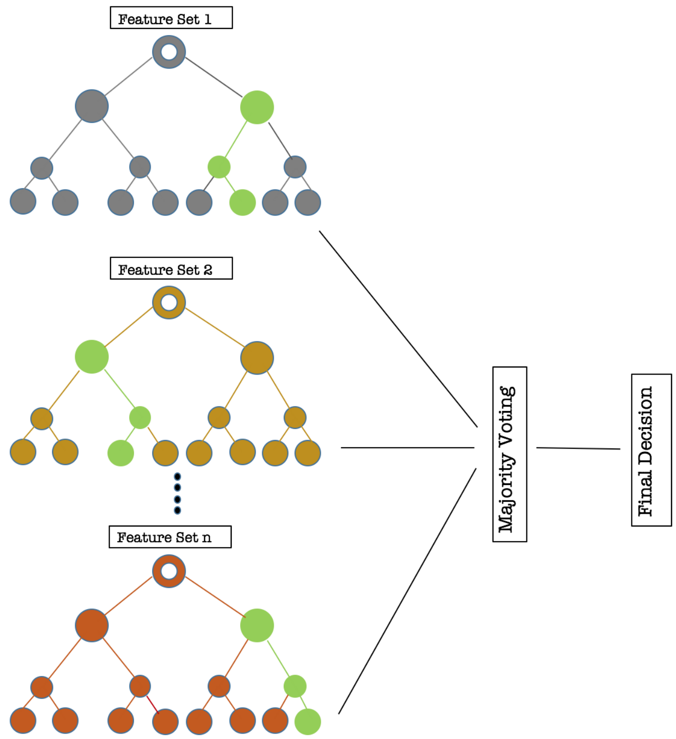

Fig. 11. Random Forest is an ensemble learning method which is used mostly for classification and regression. It operates by creating a multitude of decision trees on various sub-samples of the dataset and uses majority voting or averaging for finding output. This model improves the accuracy of prediction and can control over-fitting.

I. Fuzzy Logic

Fuzzy Logic is a machine learning technique that segregates the given data by assigning a range of truth values between 0.0 to 1.0. This method does not merely accept or reject data by assigning a definite truth or false value (binary value 0 or 1). Instead, it decides upon justification by assigning a degree of truth value, where the data is not completely true, but to an extent can be considered true by a certain value. Consider the following example: We have two types of known shades of blue color which are light blue and dark blue. If you are in a process of training your computer to determine which exact shade of blue a given input (sea blue) might be, we can assign a truth value for that input ranging from 0.0 to 1.0 which will state that the input can be 80 percent closer to the intended true color.

Almost every household machine or equipment (like the air conditioner, washing machine, etc.) operates on the concept of Fuzzy Logic. This logic is fed to a control system usually called the Fuzzy system control, where each component is designed to function and alter another physical operating system, to achieve the desired functionality. To understand how a fuzzy system works, it is necessary to analyze the system requirements and the intent for using a fuzzy system [22]. To make a system a knowledge-based functioning element with the capacity to apply the human cognitive processes, such as reasoning and thinking, has to have a stable component that can provide output on the perspective of the degree of truth for a given set of input variables. Fuzzy logic is observed to be implemented in many applications in the field of Machine learning, right from control theory to AI (Artificial Intelligence). This is an exact simulation of a human being's brain for thinking and reasoning abilities [115]. For a Fuzzy system to work effectively, the following features and components need to be assured of performance:

1. Fuzzy sets: A fuzzy set is considered to be correspondent with the member function, which is defined in a fuzzy space where the variables are set. The feature of a member function is to provide a degree of membership to any element within the well defined fuzzy sets. Then the member function assigns these elements a numerical value between 0 to 1, where 0 implies the corresponding element is not an element in the fuzzy set or 1 means the corresponding element is an element of the fuzzy set. One main aspect here is that the elements are allowed to be assigned a partial membership, stating that the corresponding element can be an element from the fuzzy set to a certain degree ranging from 0 to 1. However, this criterion is only satisfied according to the set of defined Fuzzy Rules. Fig. 12 shows the breakdown of a typical fuzzy system

2. Fuzzy Rules: The way a fuzzy logic is intended to function is defined by a set of applied fuzzy rules, which determines the output which will be specified by the IF-THEN rules. The IF-THEN rules are observed to create a conditional statement that will consist of fuzzy logic. For example, the IF-THEN assumes where X and Y are intended terms and are evaluated by the terms of fuzzy sets with the range being U and V. This divides the statement into two parts namely antecedent and consequent. If the antecedent is a preceding statement which specifies the terms X and U, then the consequent statement should conclude with Y and V. These combined makes a rule which states: if X is U, then Y is V. However, these rules are based on the natural language and model representation, based on the given fuzzy sets and logic.

3. Fuzzy Logic Inference or Fuzzy Inference System (FIS): Once the set of fuzzy rules and membership functions have been defined, the FIS is implemented for process simulation, and control, and is done by the type of data or knowledge provided. The FIS system usually operates on 3 stages: In the first stage, the numerical input variables which are provided to the system, are mapped for a degree of compatibility for the respective fuzzy sets. This is called the Fuzzification process. This process allows the system to express the input and output in fuzzy-readable linguistic terms. In the second stage, the system processes the rules according to the strengths of each input variable. in the third stage, the resulting fuzzy values are converted back to numerical values, by the process of Defuzzification. This process thereby maps the fuzzy domain output back to the crisp domain, which makes the output clear.

4. Fuzzy Score: The output from the FIS system is in the form of a fuzzy score, for all the individual input scores that are known to be generated by the system. The FIS system calculates the fuzzy score by taking into considerations all the defined fuzzy constraints and membership functions. The score is dependent on the type of rules applied and the type of input variables. Every input variable is assigned a score by the FIS based on the fuzzy rules criteria. Typical fuzzy system

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

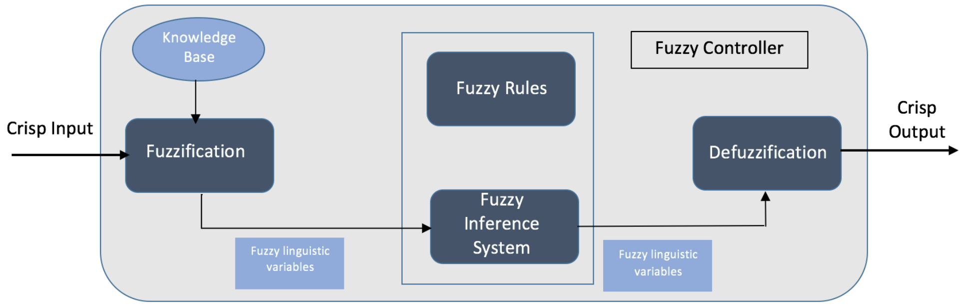

Fig. 12. Example for Fuzzy System. For a Fuzzy system to work effectively, the following features and components needs to be assured of performance: 1. Fuzzy sets, 2. Fuzzy Rules, 3. Fuzzy Logic Inference, 4. Fuzzy Score.

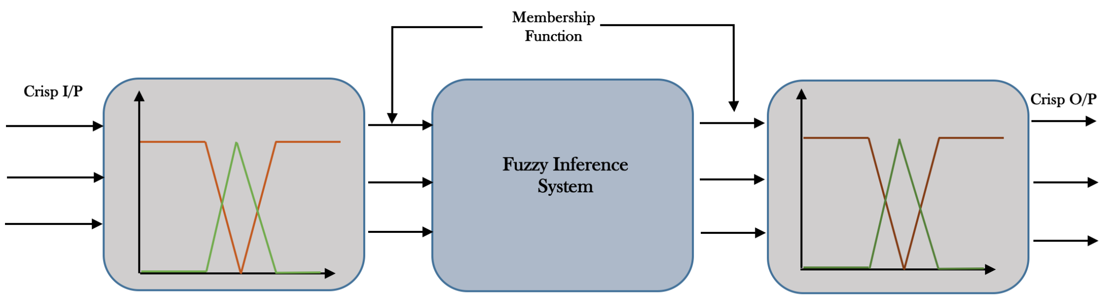

Fig. 13. Operation of Fuzzy System. The input variables are mapped by fuzzy sets which are sets of membership functions. This process of mapping a crisp input value to a fuzzy value is named fuzzification.

TABLE VII
FUZZY LOGIC FOR EEG ANALYSIS

Author(s)	Machine Learning Method	Application	Data Set	Results
Li Peng., et al., 2018 [74]	Fuzzy Logic	Detection of epileptic seizure	Public	Accuracy 93%
Abbasi Hamid., et al. 2016 [2]	Fuzzy Logic	Stereotypic evolving micro-scale seizures (SEMS) identification .	Not Reported	Accuracy 78.71%
Rabbi A. F., Azinfar, L., Fazel-Rezai, R, et al. 2013 [91]	Fuzzy Logic, ANN	Seizure prediction using adaptive neuro-fuzzy inference system	Not Reported	Accuracy 80%
Rabbi A. F., Fazel-Rezai, R., et al. 2012 [92]	Fuzzy Logic	Seizure onset detection in Intracranial EEG.	20 Patients	Accuracy 95.8%
Aarabi, A., Fazel-Rezai, R., Aghakhani, Y., et al. 2009 [1]	Fuzzy Logic	Seizure detection in Intracranial EEG using a fuzzy inference system	21 Patients	Accuracy 98.7%
Cosenza-Andraus, M. E., et al. 2006 [22]	Fuzzy Logic	Video-electroencephalography prolonged monitoring	22 Adult patients	Accuracy 91%
Sharif, B., Jafari, A. H., et al. 2017 [102]	Fuzzy Logic	Prediction of seizures from EEG sugnals	19 Patients	Sensitivity 91.8-96.6%
Hsu, Wei-Yen. , et al. 2015 [54]	Fuzzy Logic	Assembling multi-Feature EEG classifier	Not reported	Accuracy 88.2%
Ubeyli, E. D., et al. 2006 [115]	Fuzzy Logic	Fuzzy similarity index for discrimination of EEG signals.	5 Patients	Not Reported
Subasi, Abdulhamit., et al. 2007 [109]	Fuzzy Logic, ANN	Epileptic seizure detection using wavelet feature extraction	Public	Not Reported

operation is illustrated in Fig. 13.

As the main application of Machine Learning is found to be in pattern recognition of EEG signals, Fuzzy Logic can be used to determine the correct recognition rate of EEG classifications at different stages. However, a combination of Fuzzy logic with Neural networks often called the Neuro-Fuzzy system, is adopted, where the system can apply the fuzzy parameter (like fuzzy sets, fuzzy rules) and combine that with the neural network approximation techniques for extensive analysis. The Neuro-Fuzzy system [91] is found to be highly beneficial for medical condition diagnostics, density and regression estimation, pattern recognition, and data analytics [?]

J. Linear Discriminant Analysis

For a given data set with a wide selection of random variables, it is necessary to perform a dimensionality reduction to reduce the number of parameters to specific principle variables to reduce the dimensional space of the dataset. As there are many possible ways to classify the data, the dimensionality reduction technique is implemented by two techniques: The Principle component analysis, and linear discriminant analysis.

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

Both PCA and LDA have similar functionalities and applications. However, the LDA technique can handles situations where the within-class frequencies need not be equal and the standout factor is that it offers a high ratio and significant separation between the between-class variance, and the within-class variance. The main difference between the PCA and LDA being, PCA is more applicable for classification of features, and LDA is applicable for data classification.

The most common technique used for dimensionality reduction is Linear discriminant analysis (LDA). The main criteria behind this technique are to offer a good separability between different classes and to avoid overfitting of the curve. This will significantly reduce computational costs and provides better classification, by projecting the given feature space with n-dimensional samples onto a precise and smaller feature subspace. In a typical PCA analysis, the location, shape, and structure of the data set completely change. But for LDA, the technique maintains the location and shape of the data set when transformed into a different smaller space. This happens through defining a set of vectors on the transformed space to distinguish and separate. In an LDA technique this usually happens by two different approaches:

1. Class-independent transformation: This approach mainly focuses on increasing the ratio of overall variance to the within-class variance and it only uses one criterion to optimize the process of data set transformation. This transforms all the necessary data points irrespective of their class. So here, each class is observed to be separate from all other classes.

2.Class-dependent transformation: Here, the main objective is to increase the ratio between the class variables to that of the within-class variables, to offer a sufficient range of separability for classification

For the application of analysis of EEG signals and the Brain-computing interface, the exploration of advanced methods to separate and segregate the data sets with multiple variables, in an effective manner. A received EEG signal may be distorted by noise disturbance and may have to be separated effectively, to achieve accurate results. For this purpose, the technique of dimensionality reduction is being implemented, to reduce the data set and separate the unwanted signal frequencies from the ones in interest.

K. K-Means

K-means is an unsupervised learning method that is used for the clustering problem. The way it works is by using an algorithm to locate a partition to minimize the error between a cluster's empirical mean and points within. Using these K clusters, K-means tries to minimize the summation of the squared errors [61].

There are two commonly used methods for initialization: Forgy and Random Partition. With the Forgy method, K observations are chosen randomly from the data set. These observations are then used as the initial means. For the Random Partition method, each observation is first assigned a random cluster. This is then updated as the initial mean is computed such that it is at the center of the cluster.

One of the advantages of K-means is its easy implementation of high computational speed given that K is relatively small. Some of the disadvantages of K-means include the high significance of initial conditions on final outputs, sensitivity to scaling, and a correlation between data order with final results.

L. Reinforcement learning

Machine learning can be divided into three categories: Supervised learning, Unsupervised learning and Reinforcement learning. Reinforcement learning uses the given data to choose an action for an environment that yields the maximum expected long term reward. Reinforcement learning concentrates more on performance than the given data. Concerning the given data, Reinforcement learning can be described as a combination of supervised learning and unsupervised learning. Reinforcement learning may or may not have known input/output pairs of the result. Reinforcement learning tries to find a combination between exploration of unknown and exploitation of current knowledge which also poses exploration vs exploitation trade-off. The biggest problem in modern-day brain-computer interface (BCI) systems is that the performance factor of these systems in controlling a BCI can and will decrease significantly over the period. Due to this issue, the necessity of controlling a BCI has increased, and the motivation factor behind this is quite low. To eradicate this scenario and find a solution to the addressed problem, we must enable a continuous feedback system from the subject and feed that to a Reinforcement learning agent to train and support the case in finding an accurate solution. The purpose here is to use the RL agent to control the actions of the given task and as the process precedes, the supporting impact from the agent is decreased and the subject will take over the control mechanism. As the subject takes over the control, the criteria are to maintain the subject at the state and to measure the performance by implementing a reward system that assigns certain points to the subject on how well it controls the task without any agent present. The main objective of the reinforcement agent is to interact with the subject in uncertain conditions, and maximize the numerical long term reward for the subject, basically taking a subject from one state to another. For example, if in every state St, there exists an agent which can take up an action At to get to a new state St+1. The agent will gain the capacity to learn and interact in different states by increasing the numerical long term reward for the agent. Consider an example where if an athlete runs fast in a 100m race, he wins the medal. The athlete is the agent, the environment is the field or ground, the reward is the medal, the state is the 100 m race. This is shown in Fig. 15

A typical reinforcement learning setup consists of possible states denoted by S, A be the possible set of actions, a state transition function given by Delta: S*A which tends to state S, a reward function given by S*A*S which tends to R, and a policy pi which defines state which tends to action A. [106] The representation is shown below:

$$\delta : S \times A \to S \tag{14}$$

$$r : S \times A \times S \to \mathbb{R} \tag{15}$$

$$\pi : S \to A \tag{16}$$

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

TABLE VIII
LINEAR DISCRIMINANT ANALYSIS

Author(s)	Machine Learning Method	Application	Data Set
Kirar, J. S.,Agrawal, R. K. (2018 [70]	LDA	Feature selection and classification for EEG using LDA	Public
Yuan, S., Zhou, W., Chen, L. 2018 [121]	LDA	Siezure detection using Bayesian LDA.	21 patients
Liu, Y. H., Huang, S., Huang, Y. D. 2017 [76]	LDA	Motor Imagery EEG classification	Not Reported
Neto, E., Biessmann, F., Aurlien, H., Nordby, H., Eichele, T. 2016 [84]	LDA	Regularized LDA of EEG features	114 Patients
Treder, M. S., Porbadnigk, A. K., Avarvand, F. S., Müller, K. R., Blankertz, B. 2016 [112]	LDA	Optimal estimation of ERP source time	Public
Chen, W., Shen, C. P., Chiu, M. J., Zhao, Q., Cichocki, A., Lin, J. W., Lai, F. 2015 [19]	LDA	Epileptic EEG visualization based on LDA	2 normal, 4 seizure patients
Mirsadeghi, M., Behnam, H., Shalbaf, R., Moghadam, H. J. 2016 [79]	LDA	Characterizing awake and anesthetized states using LDA	25 Patients
Ying, X., Lin, H., Hui, G. 2015 [119]	LDA	Non-linear bistable dynamics model based on LDA	Public
Onishi, A. and Natsume, K., 2014 [86]	LDA	Multi-class ERP based BCI analysis	Not Reported
Onishi, A., Natsume, K. 2014, [86]	LDA	Epileptic seizure detection Bayesian LDA	Public

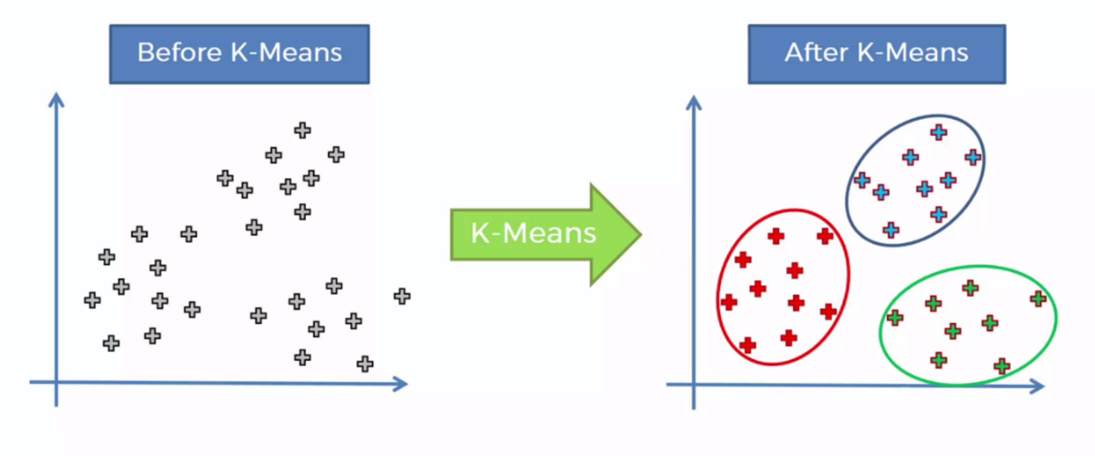

Fig. 14. General K-means classification. K-means works based on using an algorithm to locate a partition in order to minimize the error between a cluster's empirical mean and points within. Using these K clusters, K-means tries to minimize the summation of the squared errors

TABLE IX
K MEANS FOR EEG ANALYSIS

Author(s)	Application	Data Set	Results	
Manjusha, M., et al., 2016 [77]	Robust Epilepsy Classification	20 patients	Accuracy 93.02%	
Prabhakar, et al., 2015 [89]	Epilepsy Risk Level Classification	20 patients	Accuracy 71.09%	
Rai, et al., 2015 [93]	Novel Feature Identification	5 patients	Accuracy 99.00%	
Teramae, et al., 2010 [110]	estimation of feeling	patients not listed	discrimination ratio 84.2%	
Harikumar, et al., 2012 [36]	fuzzy outputs optimization	20 patients	Accuracy 95.88%	
Bizopoulos, et al., 2013 [17]	epileptic seizure detection	patients not listed	Accuracy 98%	
Asanza, et al., 2016 [12]	EEG occipital signal classification	patients not listed	Accuracy unknown	
ozerdem et al., 2017 [87]	KNN,SVM	Emotion recognition	32 healthy subjects	72.92%

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

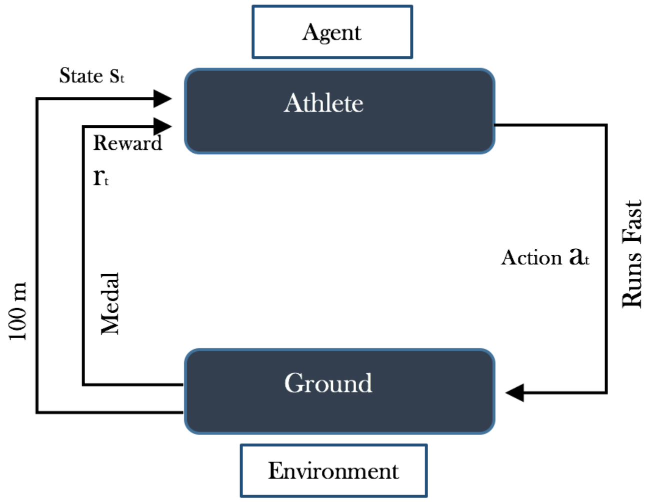

Fig. 15. Operation of Reinforcement Learning. Software agents must to take suitable actions in an environment to maximize reward in a particular situation.

One of the advantages of using the RL model is that it maintains the balance between Exploration and Exploitation. Other supervised algorithms cannot perform this balance. For EEG analysis applications, the RL model has shown constant progress towards the control mechanism of the brain-computer interface system, maintaining an equal balance between state transitions and reward mechanisms for optimum functioning.

M. Combination of Methods

A combination of methods involves the use of two or more of the machine learning algorithms to take advantage of the unique characteristics that each method possesses. This allows the multimodal algorithm to extract additional desired features. The significance of multimodal integration is that it allows high-resolution classification using primarily already existing methods [23], [50]. Additionally, this resolution will generally be higher than that of the individual methods separately. However, multimodal extraction is not without limitations. Due to the increased complexity of the algorithm, it may be difficult to determine the true accuracy as it is not directly comparable to existing methods. An example of this application in EEG is the diagnosis of multiple sclerosis patients. In the paper, the T-test [45], [46] and Bhattacharyya were used for feature extraction as part of the preprocessing. Following this a combination of KNN and SVM as the primary classification algorithm. This resulted in a total accuracy of 93% [111]. While other sections above have dedicated tables with reviewed literature, we wanted to bring attention to multimodal analysis as some literature above already demonstrated the application of the combination of methods [41].

III. Conclusion

EEG is a noninvasive electrophysiological device that records the electrical activity of the brain by placing electrodes around the scalp. Based on this activity and oscillating electric potentials, EEG can be used to diagnose neurological disorders, such as epilepsy, or for emotion recognition. Of methods for emotion recognition, EEG is one of the most reliable because its signals are highly accurate and more objective than other external appearance approaches.

The art of machine learning has led to the development and application of different techniques, which has made it possible for the computer to analyze and learn information from a given set of data, and make the desired prediction accordingly. By providing a sufficient amount of data, along with the help of precision coding and algorithms, it has been possible to train a computer in determining the exact output as required [49]. The proposed machine learning models are being used in interpreting the type of epileptic disorder being treated, and evaluate the treatment methods by analyzing EEG signals. EEG signals play a major role in the accurate segregation of different forms of epilepsy. The EEG signals are extracted by the application of discrete wavelet transforms and auto-regressive mathematical models and are supplied as input with the ML algorithm. In this paper, we presented a broad and comprehensive review of state-of-art EEG analysis methods with medical applications. We tried to familiarize the readers with machine learning methods which have been implemented for different application.

As the process of epileptic seizure detection is a bit complicated biomedical situation, it has generated a substantial amount of concerns towards the utilization of machine learning processes as a solution [83]. Most of the recent literature surveys regarding EEG signal analysis have proposed multiple learning models and different artificial neural network algorithms like radial bias function, recurrent neural networks, and vector quantizations to interpret epileptic seizure patterns in a given set of EEG signals. The problem is also being targeted and solved using other models like Support Vector Machines (SVM), adaptive neuro-fuzzy interference system (ANFIS), adaptive learning, and time-frequency analysis.

Reviewing the published papers in EEG analysis for epilepsy the following points are considerable. Dimensionality reduction and selection have been identified as an interesting topic for EEG analysis using machine learning methods [13]. Also, wavelet transform and Auto-regressive methods have played a pivotal role in machine learning for EEG such as the following studies [7]. Subasi used wavelet feature extraction for epileptic seizure detection with an adaptive neuro-fuzzy inference system in‘ [109]. Linear programming boosting has been used for EEG signals analysis in [37]. The effect of de-noising such as multiscale PCA in EEG analysis is shown in [67]. Data preparation methods such as PCA, ICA and, LDA can be used to increase the classification accuracy [34]. Ensemble methods and combining classifiers have shown good performance in EEG analysis such as the following studies [66].

The incorporation of deep learning models in neuroimaging and electrodiagnostic analytics has allowed for large amounts of data to be correlated from multiple modalities [44]. These models have been shown to perform better and faster than current state-of-the-art analysis techniques through both supervised and unsupervised learning tasks. Recent advancements and advantages in using deep learning in EEG analysis can provide more accurate and faster analysis for a large amount

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

of data. Hosseini et al. [39], [42], [51] proposed a cloud-based method for EEG analysis. In [52] convolutional neural networks (CNN) have been developed for EEG analysis. In [47] optimization modules consisted of PCA, ICA and DSA analysis are developed for CNN and stacked auto-encoder deep learning structures in EEG analysis.

We reviewed different applications of machine learning in EEG such as sleep stage analysis. Alickovic et al. [8] and Hassan et al. [38] proposed automated identification of sleep stages from single-channel EEG signals. Kevric et al. presented signal decomposition methods in the classification of EEG signals for the motor-imagery BCI system.

The coefficients of the wavelet transform and the numerical autoregressive model are used in recognizing the changes and behaviors in EEG signals. These coefficients are taken as inputs and combined with different machine learning algorithms like multiple layered neural networks, K-means, Support vector machines, K-nearest neighbors, and Naive Bayesian; to break the EEG signal into machine recognizable components, for extracting and determining the power points which are responsible for triggering seizures.

As there are multiple techniques involving machine learning for analyzing a given set of EEG signals, it is required to evaluate the best-suited technique for a given application. Each model has a specific use case about the type of application and subject data set. As to our topic of study here, we were concerned about the analysis of waveforms to determine an output. Here we will see how each different ML models can be used for the intended use case:

K-NN classifiers can be used for both regression and classification of data, which for our purpose can be used for identifying and classifying different acquired EEG signals and finding the nearest possible output point to the desired classification line for possible detection of abnormality. ANN, on the other hand, has the capability of segregating the physical shape of the EEG waveform and dividing it into segments. These segments are each given a specific weight value accordingly by analyzing the waveform, and the output is determined by subjecting the final equation to a bias. The final chosen bias brings down the output to a desired expected range. As more data is being involved, the number of interactions in the hidden layer will increase. So depending on the type of problem and the amount of data being considered, a suitable selection has to be made while selecting an appropriate process.

IV. Acknowledgement

We would like to thank students at Santa Clara University, Thi-Vu Huynh, Pradnya Patel, Elissa Yang, and Haygreev Venkatesh for their contributions to this work.

References

[1] A. Aarabi, R. Fazel-Rezai, and Y. Aghakhani. Seizure detection in intracranial eeg using a fuzzy inference system. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 1860–1863. IEEE, 2009.

[2] H. Abbasi, L. Bennet, A. J. Gunn, and C. P. Unsworth. Identifying stereotypic evolving micro-scale seizures (sems) in the hypoxic-ischemic eeg of the pre-term fetal sheep with a wavelet type-ii fuzzy classifier. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 973–976. IEEE, 2016.

[3] H. Aghajani, M. Garbey, and A. Omurtag. Measuring mental workload with eeg+ fnirs. *Frontiers in human neuroscience*, 11:359, 2017.

[4] A. Ahani, H. Wahbeh, H. Nezamfar, M. Miller, D. Erdogmus, and B. Oken. Quantitative change of eeg and respiration signals during mindfulness meditation. *Journal of neuroengineering and rehabilitation*, 11(1):87, 2014.

[5] S. B. Akben, A. Subasi, and D. Tuncel. Analysis of repetitive flash stimulation frequencies and record periods to detect migraine using artificial neural network. *Journal of medical systems*, 36(2):925–931, 2012.

[6] O. Al Zoubi, C. Ki Wong, R. T. Kuplicki, H.-w. Yeh, A. Mayeli, H. Refai, M. Paulus, and J. Bodurka. Predicting age from brain eeg signals–a machine learning approach. *Frontiers in aging neuroscience*, 10:184, 2018.

[7] E. Alickovic, J. Kevric, and A. Subasi. Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomedical signal processing and control*, 39:94–102, 2018.

[8] E. Alickovic and A. Subasi. Ensemble svm method for automatic sleep stage classification. *IEEE Transactions on Instrumentation and Measurement*, 67(6):1258–1265, 2018.

[9] H. U. Amin, W. Mumtaz, A. R. Subhani, M. N. M. Saad, and A. S. Malik. Classification of eeg signals based on pattern recognition approach. *Frontiers in computational neuroscience*, 11:103, 2017.

[10] M. N. Anastasiadou, M. Christodoulakis, E. S. Papathanasiou, S. S. Papacostas, and G. D. Mitsis. Unsupervised detection and removal of muscle artifacts from scalp eeg recordings using canonical correlation analysis, wavelets and random forests. *Clinical Neurophysiology*, 128(9):1755–1769, 2017.

[11] A. Antoniades, L. Spyrou, D. Martin-Lopez, A. Valentin, G. Alarcon, S. Sanei, and C. C. Took. Deep neural architectures for mapping scalp to intracranial eeg. *International journal of neural systems*, page 1850009, 2018.

[12] V. Asanza, K. Ochoa, C. Sacarelo, C. Salazar, F. Loayza, C. Vaca, and E. Peláez. Clustering of eeg occipital signals using k-means. In *Ecuador Technical Chapters Meeting (ETCM), IEEE*, pages 1–5. IEEE, 2016.

[13] N. Beganovic, J. Kevric, and D. Jokic. Identification of diagnostic-related features applicable to eeg signal analysis. In *Proceedings of the Annual Conference of the PHM Society*, volume 10, 2018.

[14] M. Bentlemsan, E.-T. Zemouri, D. Bouchaffra, B. Yahya-Zoubir, and K. Ferroudji. Random forest and filter bank common spatial patterns for eeg-based motor imagery classification. In *Intelligent Systems, Modelling and Simulation (ISMS), 2014 5th International Conference on*, pages 235–238. IEEE, 2014.

[15] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig. Brain activity-based image classification from rapid serial visual presentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(5):432–441, 2008.

[16] S. Biswal, Z. Nip, V. M. Junior, M. T. Bianchi, E. S. Rosenthal, and M. B. Westover. Automated information extraction from free-text eeg reports. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 6804–6807. IEEE, 2015.

[17] P. A. Bizopoulos, D. G. Tsalikakis, A. T. Tzallas, D. D. Koutsouris, and D. I. Fotiadis. Eeg epileptic seizure detection using k-means clustering and marginal spectrum based on ensemble empirical mode decomposition. In *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, pages 1–4. IEEE, 2013.

[18] S. Bose, V. Rama, and C. R. Rao. Eeg signal analysis for seizure detection using discrete wavelet transform and random forest. In *Computer and Applications (ICCA), 2017 International Conference on*, pages 369–378. IEEE, 2017.

[19] W. Chen, C.-P. Shen, M.-J. Chiu, Q. Zhao, A. Cichocki, J.-W. Lin, and F. Lai. Epileptic eeg visualization and sonification based on linear discriminate analysis. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 4466–4469. IEEE, 2015.

[20] A. M. Chiarelli, P. Croce, A. Merla, and F. Zappasodi. Deep learning for hybrid eeg-fnirs brain–computer interface: application to motor

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

imagery classification. *Journal of neural engineering*, 15(3):036028, 2018.

[21] E. Combrisson and K. Jerbi. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of neuroscience methods*, 250:126–136, 2015.

[22] M. Cosenza-Andraus, C. Nunes-Cosenza, R. Gomes-Nunes, C. Fantezia-Andraus, and S. Alves-Leon. Video-electroencephalography prolonged monitoring in patients with ambulatory diagnosis of medically refractory temporal lobe epilepsy: application of fuzzy logic's model. *Revista de neurologia*, 43(1):7–14, 2006.

[23] S. Dähne, F. Bießmann, F. C. Meinecke, J. Mehnert, S. Fazli, and K.-R. Mtüller. Multimodal integration of electrophysiological and hemodynamic signals. In *Brain-Computer Interface (BCI), 2014 International Winter Workshop on*, pages 1–4. IEEE, 2014.

[24] A. Delorme and S. Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.

[25] J. A. Dian, S. Colic, Y. Chinvarun, P. L. Carlen, and B. L. Bardakjian. Identification of brain regions of interest for epilepsy surgery planning using support vector machines. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 6590–6593. IEEE, 2015.

[26] Q. Dong, B. Hu, J. Zhang, X. Li, and M. Ratcliffe. A study on visual attention modeling—a linear regression method based on eeg. In *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pages 1–6. IEEE, 2013.

[27] C. Dora and P. K. Biswal. Robust ecg artifact removal from eeg using continuous wavelet transformation and linear regression. In *Signal Processing and Communications (SPCOM), 2016 International Conference on*, pages 1–5. IEEE, 2016.

[28] F. D. V. Fallani, G. Vecchiato, J. Toppi, L. Astolfi, and F. Babiloni. Subject identification through standard eeg signals during resting states. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2331–2333. IEEE, 2011.

[29] J. Fan, J. W. Wade, D. Bian, A. P. Key, Z. E. Warren, L. C. Mion, and N. Sarkar. A step towards eeg-based brain computer interface for autism intervention. In *Conference proceedings:... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, volume 2015, page 3767. NIH Public Access, 2015.

[30] V. Gao, F. Turek, and M. Vitaterna. Multiple classifier systems for automatic sleep scoring in mice. *Journal of neuroscience methods*, 264:33–39, 2016.

[31] S. Ghosh-Dastidar and H. Adeli. A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection. *Neural networks*, 22(10):1419–1431, 2009.

[32] M. Günay and T. Ensari. Eeg signal analysis of patients with epilepsy disorder using machine learning techniques. In *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, pages 1–4. IEEE, 2018.

[33] L. Guo, D. Rivero, and A. Pazos. Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. *Journal of neuroscience methods*, 193(1):156–163, 2010.

[34] M. I. Gursoy and A. Subast. A comparison of pca, ica and lda in eeg signal classification using svm. In *2008 Ieee 16th Signal Processing, Communication and Applications Conference*, pages 1–4. IEEE, 2008.

[35] C. R. Hamilton, S. Shahryari, and K. M. Rasheed. Eye state prediction from eeg data using boosted rotational forests. In *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on*, pages 429–432. IEEE, 2015.

[36] R. Harikumar, T. Vijayakumar, and M. Sreejith. Performance analysis of svd and k-means clustering for optimization of fuzzy outputs in classification of epilepsy risk level from eeg signals. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2012 9th International Conference on*, pages 1–4. IEEE, 2012.

[37] A. R. Hassan and A. Subasi. Automatic identification of epileptic seizures from eeg signals using linear programming boosting. *Computer methods and programs in biomedicine*, 136:65–77, 2016.

[38] A. R. Hassan and A. Subasi. A decision support system for automated identification of sleep stages from single-channel eeg signals. *Knowledge-Based Systems*, 128:115–124, 2017.

[39] M. P. Hosseini. *Brain-computer interface for analyzing epileptic big data*. PhD thesis, Rutgers University-School of Graduate Studies, 2018.

[40] M.-P. Hosseini, A. Hajisami, and D. Pompili. Real-time epileptic seizure detection from eeg signals via random subspace ensemble learning. In *2016 IEEE international conference on autonomic computing (ICAC)*, pages 209–218. IEEE, 2016.

[41] M.-P. Hosseini, A. Lau, K. Elisevich, and H. Soltanian-Zadeh. Multimodal analysis in biomedicine. In *Big Data in Multimodal Medical Imaging*, pages 193–203. Chapman and Hall/CRC, 2019.

[42] M.-P. Hosseini, S. Lu, K. Kamaraj, A. Slowikowski, and H. C. Venkatesh. Deep learning architectures. In *Deep Learning: Concepts and Architectures*, pages 1–24. Springer, 2020.

[43] M.-P. Hosseini, M. R. Nazem-Zadeh, F. Mahmoudi, H. Ying, and H. Soltanian-Zadeh. Support vector machine with nonlinear-kernel optimization for lateralization of epileptogenic hippocampus in mr images. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1047–1050. IEEE, 2014.

[44] M.-P. Hosseini, M.-R. Nazem-Zadeh, D. Pompili, K. Jafari-Khouzani, K. Elisevich, and H. Soltanian-Zadeh. Automatic and manual segmentation of hippocampus in epileptic patients mri. *arXiv preprint arXiv:1610.07557*, 2016.

[45] M.-P. Hosseini, M.-R. Nazem-Zadeh, D. Pompili, K. Jafari-Khouzani, K. Elisevich, and H. Soltanian-Zadeh. Comparative performance evaluation of automated segmentation methods of hippocampus from magnetic resonance images of temporal lobe epilepsy patients. *Medical physics*, 43(1):538–553, 2016.

[46] M.-P. Hosseini, M. R. Nazem-Zadeh, D. Pompili, and H. Soltanian-Zadeh. Statistical validation of automatic methods for hippocampus segmentation in mr images of epileptic patients. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4707–4710. IEEE, 2014.

[47] M.-P. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh. Optimized deep learning for eeg big data and seizure prediction bci via internet of things. *IEEE Transactions on Big Data*, 3(4):392–404, 2017.

[48] M.-P. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh. Random ensemble learning for eeg classification. *Artificial intelligence in medicine*, 84:146–158, 2018.

[49] M. P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor. Computer-aided diagnosis system for the evaluation of chronic obstructive pulmonary disease on ct images. *Tehran University Medical Journal*, 68(12), 2011.

[50] M. P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor. Three cuts method for identification of copd. *Acta Medica Iranica*, pages 771–778, 2013.

[51] M.-P. Hosseini, H. Soltanian-Zadeh, K. Elisevich, and 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)*, pages 1151–1155. IEEE, 2016.

[52] M.-P. Hosseini, T. X. Tran, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh. Deep learning with edge computing for localization of epileptogenicity using multimodal rs-fmri and eeg big data. In *2017 IEEE International Conference on Autonomic Computing (ICAC)*, pages 83–92. IEEE, 2017.

[53] A. E. Hramov, V. A. Maksimenko, S. V. Pchelintseva, A. E. Runnova, V. V. Grubov, V. Y. Musatov, M. O. Zhuravlev, A. A. Koronovskii, and A. N. Pisarchik. Classifying the perceptual interpretations of a bistable image using eeg and artificial neural networks. *Frontiers in neuroscience*, 11:674, 2017.

[54] W.-Y. Hsu. Assembling a multi-feature eeg classifier for left–right motor imagery data using wavelet-based fuzzy approximate entropy for improved accuracy. *International journal of neural systems*, 25(08):1550037, 2015.

[55] J. Hu and J. Min. Automated detection of driver fatigue based on eeg signals using gradient boosting decision tree model. *Cognitive Neurodynamics*, pages 1–10, 2018.

[56] J. Hu and Z. Mu. Eeg authentication system based on auto-regression coefficients. In *Intelligent Systems and Control (ISCO), 2016 10th International Conference on*, pages 1–5. IEEE, 2016.

[57] Y. Huang, J. Yang, P. Liao, and J. Pan. Fusion of facial expressions and eeg for multimodal emotion recognition. *Computational intelligence and neuroscience*, 2017, 2017.

[58] A. Ishfaque, A. J. Awan, N. Rashid, and J. Iqbal. Evaluation of ann, lda and decision trees for eeg based brain computer interface. In *Proceedings of the IEEE 9th International Conference on Emerging Technologies (ICET'13)*, pages 1–6, 2013.

[59] I. Iturrate, L. Montesano, and J. Minguez. Robot reinforcement learning using eeg-based reward signals. In *Proceedings of the IEEE*

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

International Conference on Robotics and Automation (ICRA), number EPFL-CONF-205134, pages 4822–4829, 2010.

[60] A. Jain, B. Abbas, O. Farooq, and S. K. Garg. Fatigue detection and estimation using auto-regression analysis in eeg. In *Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on*, pages 1092–1095. IEEE, 2016.

[61] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666, 2010.

[62] A. K. Jaiswal and H. Banka. Epileptic seizure detection in eeg signal with gmodpca and support vector machine. *Bio-medical materials and engineering*, 28(2):141–157, 2017.

[63] L. Jakaite, V. Schetinin, C. Maple, and J. Schult. Bayesian decision trees for eeg assessment of newborn brain maturity. In *Computational Intelligence (UKCI), 2010 UK Workshop on*, pages 1–6. IEEE, 2010.

[64] A. Jalilifard, E. B. Pizzolato, and M. K. Islam. Emotion classification using single-channel scalp-eeg recording. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 845–849. IEEE, 2016.

[65] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena. Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *The Scientific World Journal*, 2014, 2014.

[66] S. Jukić and J. Kevrić. Majority vote of ensemble machine learning methods for real-time epilepsy prediction applied on eeg pediatric data. *TEM Journal*, 7(2):313, 2018.

[67] J. Kevric and A. Subasi. The effect of multiscale pca de-noising in epileptic seizure detection. *Journal of medical systems*, 38(10):131, 2014.

[68] J.-H. Kim, F. Bießmann, and S.-W. Lee. Reconstruction of hand movements from eeg signals based on non-linear regression. In *Brain-Computer Interface (BCI), 2014 International Winter Workshop on*, pages 1–3. IEEE, 2014.

[69] C. Kingsford and S. L. Salzberg. What are decision trees? *Nature biotechnology*, 26(9):1011, 2008.

[70] J. S. Kirar and R. Agrawal. Relevant feature selection from a combination of spectral-temporal and spatial features for classification of motor imagery eeg. *Journal of medical systems*, 42(5):78, 2018.

[71] J. Laton, J. Van Schependom, J. Gielen, J. Decoster, T. Moons, J. De Keyser, M. De Hert, and G. Nagels. Single-subject classification of schizophrenia patients based on a combination of oddball and mismatch evoked potential paradigms. *Journal of the neurological sciences*, 347(1-2):262–267, 2014.

[72] J. Le Douget, A. Fouad, M. M. Filali, J. Pyrzowski, and M. Le Van Quyen. Surface and intracranial eeg spike detection based on discrete wavelet decomposition and random forest classification. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 475–478. IEEE, 2017.

[73] Y.-H. Lee, Y.-J. Hsieh, Y.-J. Shiah, Y.-H. Lin, C.-Y. Chen, Y.-C. Tyan, J. GengQiu, C.-Y. Hsu, and S. C.-J. Chen. A cross-sectional evaluation of meditation experience on electroencephalography data by artificial neural network and support vector machine classifiers. *Medicine*, 96(16), 2017.

[74] P. Li, C. Karmakar, J. Yearwood, S. Venkatesh, M. Palaniswami, and C. Liu. Detection of epileptic seizure based on entropy analysis of short-term eeg. *PloS one*, 13(3):e0193691, 2018.

[75] X. Li, C. Guan, K. K. Aug, C. Wang, Z. Y. Chin, H. Zhang, C. G. Lim, and T. S. Lee. An ocular artefacts correction method for discriminative eeg analysis based on logistic regression. In *Signal Processing Conference (EUSIPCO), 2015 23rd European*, pages 2731–2735. IEEE, 2015.

[76] Y.-H. Liu, S. Huang, and Y.-D. Huang. Motor imagery eeg classification for patients with amyotrophic lateral sclerosis using fractal dimension and fisher's criterion-based channel selection. *Sensors*, 17(7):1557, 2017.

[77] M. Manjusha and R. Harikumar. Performance analysis of knn classifier and k-means clustering for robust classification of epilepsy from eeg signals. In *Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on*, pages 2412–2416. IEEE, 2016.

[78] T. Meyer, J. Peters, T. O. Zander, B. Schölkopf, and M. Grosse-Wentrup. Predicting motor learning performance from electroencephalographic data. *Journal of neuroengineering and rehabilitation*, 11(1):24, 2014.

[79] M. Mirsadeghi, H. Behnam, R. Shalbaf, and H. J. Moghadam. Characterizing awake and anesthetized states using a dimensionality reduction method. *Journal of medical systems*, 40(1):13, 2016.

[80] W. Mumtaz, S. S. A. Ali, M. A. M. Yasin, and A. S. Malik. A machine learning framework involving eeg-based functional connectivity to diagnose major depressive disorder (mdd). *Medical & biological engineering & computing*, 56(2):233–246, 2018.

[81] W. Mumtaz, N. Kamel, S. S. A. Ali, A. S. Malik, et al. An eeg-based functional connectivity measure for automatic detection of alcohol use disorder. *Artificial intelligence in medicine*, 84:79–89, 2018.

[82] M. Murakami, S. Nakatani, N. Araki, Y. Konishi, and K. Mabuchi. Motion discrimination from eeg using logistic regression and schmitt-trigger-type threshold. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, pages 2338–2342. IEEE, 2015.

[83] M.-R. Nazem-Zadeh, J. M. Schwalb, H. Bagher-Ebadian, F. Mahmoudi, M.-P. Hosseini, K. Jafari-Khouzani, K. V. Elisevich, and H. Soltanian-Zadeh. Lateralization of temporal lobe epilepsy by imaging-based response-driven multinomial multivariate models. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5595–5598. IEEE, 2014.

[84] E. Neto, F. Biessmann, H. Aurlien, H. Nordby, and T. Eichele. Regularized linear discriminant analysis of eeg features in dementia patients. *Frontiers in aging neuroscience*, 8:273, 2016.

[85] S. N. Oğulata, C. Şahin, and R. Erol. Neural network-based computer-aided diagnosis in classification of primary generalized epilepsy by eeg signals. *Journal of medical systems*, 33(2):107–112, 2009.

[86] A. Onishi and K. Natsume. Multi-class erp-based bci data analysis using a discriminant space self-organizing map. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 26–29. IEEE, 2014.

[87] M. S. Özerdem and H. Polat. Emotion recognition based on eeg features in movie clips with channel selection. *Brain informatics*, 4(4):241, 2017.

[88] A. Page, S. P. T. Oates, and T. Mohsenin. An ultra low power feature extraction and classification system for wearable seizure detection. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 7111–7114. IEEE, 2015.

[89] S. K. Prabhakar and H. Rajaguru. Pca and k-means clustering for classification of epilepsy risk levels from eeg signals? a comparitive study between them. In *2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, pages 83–86. IEEE, 2015.

[90] S. Puntanen. Regression analysis by example, by samprit chatterjee, ali s. hadi. *International Statistical Review*, 81(2):308–308, 2013.

[91] A. F. Rabbi, L. Azinfar, and R. Fazel-Rezai. Seizure prediction using adaptive neuro-fuzzy inference system. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 2100–2103. IEEE, 2013.

[92] A. F. Rabbi and R. Fazel-Rezai. A fuzzy logic system for seizure onset detection in intracranial eeg. *Computational intelligence and neuroscience*, 2012:1, 2012.

[93] K. Rai, V. Bajaj, and A. Kumar. Novel feature for identification of focal eeg signals with k-means and fuzzy c-means algorithms. In *Digital Signal Processing (DSP), 2015 IEEE International Conference on*, pages 412–416. IEEE, 2015.

[94] H. Rajaguru and S. K. Prabhakar. Logistic regression gaussian mixture model and softmax discriminant classifier for epilepsy classification from eeg signals. In *Computing Methodologies and Communication (ICCMC), 2017 International Conference on*, pages 985–988. IEEE, 2017.

[95] H. Rajaguru and S. K. Prabhakar. Non linear ica and logistic regression for classification of epilepsy from eeg signals. In *Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of*, volume 1, pages 577–580. IEEE, 2017.

[96] H. Rajaguru and S. K. Prabhakar. Sparse pca and soft decision tree classifiers for epilepsy classification from eeg signals. In *Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of*, volume 1, pages 581–584. IEEE, 2017.

[97] I. Rish et al. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46. IBM New York, 2001.

[98] S. Roy, I. Kiral-Kornek, and S. Harrer. Deep learning enabled automatic abnormal eeg identification. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2756–2759. IEEE, 2018.

[99] K.-M. Rytkönen, J. Zitting, and T. Porkka-Heiskanen. Automated sleep scoring in rats and mice using the naive bayes classifier. *Journal of neuroscience methods*, 202(1):60–64, 2011.

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/RBME.2020.2969915, IEEE Reviews in Biomedical Engineering

[100] C. Y. Sai, N. Mokhtar, H. Arof, P. Cumming, and M. Iwahashi. Automated classification and removal of eeg artifacts with svm and wavelet-ica. *IEEE journal of biomedical and health informatics*, 22(3):664–670, 2018.

[101] J. Saini and M. Dutta. Epilepsy classification using optimized artificial neural network. *Neurological research*, pages 1–13, 2018.

[102] B. Sharif and A. H. Jafari. Prediction of epileptic seizures from eeg using analysis of ictal rules on poincaré plane. *Computer methods and programs in biomedicine*, 145:11–22, 2017.

[103] A. Sharma, J. Rai, and R. Tewari. Epileptic seizure anticipation and localisation of epileptogenic region using eeg signals. *Journal of Medical Engineering & Technology*, pages 1–14, 2018.

[104] A. Sharmila and P. Geethanjali. Effect of filtering with time domain features for the detection of epileptic seizure from eeg signals. *Journal of Medical Engineering & Technology*, pages 1–11, 2018.

[105] A. Sharmila and P. Mahalakshmi. Wavelet-based feature extraction for classification of epileptic seizure eeg signal. *Journal of medical engineering & technology*, 41(8):670–680, 2017.

[106] E. Shephard, G. M. Jackson, and M. J. Groom. Electrophysiological correlates of reinforcement learning in young people with tourette syndrome with and without co-occurring adhd symptoms. *International Journal of Developmental Neuroscience*, 51:17–27, 2016.

[107] V. Srinivasan, C. Eswaran, and N. Sriraam. Approximate entropy-based epileptic eeg detection using artificial neural networks. *IEEE Transactions on information Technology in Biomedicine*, 11(3):288–295, 2007.

[108] A. F. Struck, B. Ustun, A. R. Ruiz, J. W. Lee, S. M. LaRoche, L. J. Hirsch, E. J. Gilmore, J. Vlachy, H. A. Haider, C. Rudin, et al. Association of an electroencephalography-based risk score with seizure probability in hospitalized patients. *JAMA neurology*, 74(12):1419–1424, 2017.

[109] A. Subasi. Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction. *Computers in biology and medicine*, 37(2):227–244, 2007.

[110] T. Teramae, D. Kushida, F. Takemori, and A. Kitamura. Estimation of feeling based on eeg by using nn and k-means algorithm for massage system. In *SICE Annual Conference 2010, Proceedings of*, pages 1542–1547. IEEE, 2010.

[111] A. Torabi, M. R. Daliri, and S. H. Sabzposhan. Diagnosis of multiple sclerosis from eeg signals using nonlinear methods. *Australasian physical & engineering sciences in medicine*, 40(4):785–797, 2017.

[112] M. S. Treder, A. K. Porbadnigk, F. S. Avarvand, K.-R. Müller, and B. Blankertz. The lda beamformer: optimal estimation of erp source time series using linear discriminant analysis. *Neuroimage*, 129:279–291, 2016.

[113] V. Tuyisenge, L. Trebaul, M. Bhattacharjee, B. Chanteloup-Forêt, C. Saubat-Guigui, I. Mîndruţă, S. Rheims, L. Maillard, P. Kahane, D. Taussig, et al. Automatic bad channel detection in intracranial electroencephalographic recordings using ensemble machine learning. *Clinical Neurophysiology*, 129(3):548–554, 2018.

[114] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis. Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Computational Intelligence and Neuroscience*, 2007, 2007.

[115] E. D. Ubeyli. Fuzzy similarity index for discrimination of eeg signals. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 5346–5349. IEEE, 2006.

[116] V. Vijayakumar, M. Case, S. Shirinpour, and B. He. Quantifying and characterizing tonic thermal pain across subjects from eeg data using random forest models. *IEEE Transactions on Biomedical Engineering*, 64(12):2988–2996, 2017.

[117] Y. Wang, W. Chen, K. Huang, and Q. Gu. Classification of neonatal amplitude-integrated eeg using random forest model with combined feature. In *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference On*, pages 285–290. IEEE, 2013.

[118] S. Weichwald, T. Meyer, B. Scholkopf, T. Ball, and M. Grosse-Wentrup. Decoding index finger position from eeg using random forests. In *Cognitive Information Processing (CIP), 2014 4th International Workshop on*, pages 1–6. IEEE, 2014.

[119] X. Ying, H. Lin, and G. Hui. Study on non-linear bistable dynamics model based eeg signal discrimination analysis method. *Bioengineered*, 6(5):297–298, 2015.

[120] W. Yu, T. Liu, R. Valdez, M. Gwinn, and M. J. Khoury. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Medical Informatics and Decision Making*, 10(1):16, Mar 2010.

[121] S. Yuan, W. Zhou, and L. Chen. Epileptic seizure prediction using diffusion distance and bayesian linear discriminate analysis on intracranial eeg. *International journal of neural systems*, 28(01):1750043, 2018.

[122] T. Zhang and W. Chen. Lmd based features for the automatic seizure detection of eeg signals using svm. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(8):1100–1108, 2017.

[123] N. Zhuang, Y. Zeng, K. Yang, C. Zhang, L. Tong, and B. Yan. Investigating patterns for self-induced emotion recognition from eeg signals. *Sensors*, 18(3):841, 2018.

Mohammad-Parsa Hosseini is a senior member at IEEE. He received his PhD in Electrical and Computer Engineering (ECE) with research in Computer Science from Rutgers University, NJ, USA in 2018. He had previously received his BSc degree in Electrical and Electronic Engineering in 2006, a graduate study in MSc of Biomedical Engineering in 2008, and a second MSc program in Electrical and Communication Engineering in 2010. He is collaborating with Medical Image Analysis Lab at Henry Ford Health System and with the Clinical Neurosciences Department, Spectrum Health, MI, USA. He is working as a data scientist and machine learning researcher at Silicon Valley, CA, USA since 2017. He has been teaching as an adjunct lecture and faculty member at several universities since 2009 and currently is with Santa Clara University. His current research focus in machine learning, deep learning, signal and image processing. He has served on the scientific committees and review boards of several national and international conferences and journals.

Amin Hosseini is an IEEE member in signal processing and machine learning societies. He is with the Electrical and Computer Engineering with minor in computer science at Azad University, Tehran Central Branch. His research interests include digital signal and image processing, machine learning, artificial intelligence and biomedical engineering.

Kiarash Ahi received his M.Sc. degree in Electrical and Information Engineering from the Leibniz University of Hannover, Germany in 2012, and his Ph.D. degree in Electrical and Computer Engineering from the University of Connecticut, USA in 2017. He is currently a Senior Researcher and Lead Product Development Engineer in the advanced Semiconductor Industry. He architects automated systems, empowered by Machine Learning and Image Processing, and leads multinational RD teams. His research and scientific interests include digital image and signal processing, semiconductor devices, Machine Learning, and intelligent software development.

1937-3333 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Northwestern University. Downloaded on May 03,2020 at 19:09:51 UTC from IEEE Xplore. Restrictions apply.