

## Article

# Deep Convolutional Neural Network Regularization for Alcoholism Detection Using EEG Signals

Hamid Mukhtar<sup>1,\*</sup>, Saeed Mian Qaisar<sup>2</sup> and Atef Zaguia<sup>1</sup>

<sup>1</sup> Department of Computer Science, College of CIT, Taif University,  
P. O. Box 11099, Taif 21944, Saudi Arabia; {h.mukhtar, zaguia.atef}@tu.edu.sa

<sup>2</sup> College of Engineering, Effat University, Jeddah 22332, Saudi Arabia; sqaisar@effatuniversity.edu.sa

\* Correspondance: h.mukhtar@tu.edu.sa;

**Abstract:** Alcoholism is attributed to regular or excessive drinking of alcohol and leads to the disturbance of the neuronal system in the human brain. This results in certain malfunctioning of neurons that can be detected by an electroencephalogram (EEG) using several electrodes on a human skull at appropriate positions. It is of great interest to be able to classify an EEG activity as that of a normal person or an alcoholic person using data from the minimum possible electrodes (or channels). Due to the complex nature of EEG signals, accurate classification of alcoholism using only a small data is a challenging task. Artificial neural networks, specifically convolutional neural networks (CNN), provide efficient and accurate results in various pattern-based classification problems. In this work, we apply CNN on raw EEG data, and demonstrate how we achieved 98% average accuracy by optimizing a baseline CNN model and outperforming its results in a range of performance evaluation metrics on the UCI-KDD EGG dataset. This article explains the step-wise improvement of the baseline model using the dropout, batch normalization, and kernel regularization techniques, and provides a comparison of the two models that can be beneficial for aspiring practitioners who aim to develop similar classification models in CNN. A performance comparison is also provided with other approaches using the same dataset.

**Keywords:** classification; optimization; batch normalization; kernel regularization; convolution; pooling; dropout layer; learning rate.

## 1. Introduction

One of the most prevalent psychiatric disorders and addiction difficulties is alcohol consumption disorder. In this situation, a person is unable to control his or her urge to use alcohol, despite the fact that it has negative health consequences such as liver failure, cardiovascular disease, neurological damage, and some categories of cancer [1]. It has the potential to cause significant harm and even death. More than 3 million people die each year because of alcoholism across the world. According to the World Health Organization (WHO), it accounts for more than 5% of all fatalities globally [2]. As a result, it is one of the leading causes of mortality and disability [3]. It also leads to family disintegration and broader societal dysfunction by causing additional damages such as traffic and other accidents, domestic and public violence, and criminality [4].

An early diagnosis of alcoholism will aid individual subjects in becoming aware of their situation and preventing permanent damage. Traditional techniques of measuring the effect of alcohol on a specific person, such as questionnaires, blood tests, and physiological tests, are used in various ways to identify alcoholism. Each approach has advantages and disadvantages. Because of personal and social reasons [5], the questionnaire-based analysis is less accurate. The blood test is not only invasive and unpleasant, but it is often unreliable [4]. In this scenario, the Electroencephalogram (EEG) based non-invasive approaches have been presented for alcoholism identification [6-8].

The EEG is a non-invasive diagnostic technique that can be used to examine the brain's neuronal activities. It measures the potentials that replicate the electrical activity of the human brain. EEG signals are a good source of obtaining information about neurological conditions of a person [9].

Thanks to its low cost and non-invasiveness, automatic EEG signal classification is an important step towards their use in practical applications and less reliance on trained professionals [10]. This approach is based on the hypothesis that brain signals fluctuate as a function of the brain's functioning condition. Multiple electrodes are placed in various locations on the subject's scalp during the EEG test to capture the electrical activity of the brain [11]. The electrodes are positioned following norms [12]. The EEG pulses are converted into signals by each electrode, which are then amplified and conditioned via a bandlimited filter [13]. A similar analog processing chain is implemented for each EEG channel. It prepares the EEG signal for proper digitization. Next, the processed signal from each channel is digitized by using appropriate Analog to Digital Converters (ADCs). The digitized signals are conveyed to an electronic EEG recording machine, where they are converted into waves and analyzed on a computer screen. EEG is widely used in cognitive psychology, neuroscience, and neurophysiological research for a variety of reasons.

Biometric signals, such as EEG signals, are known to contain various kinds of noises and other artifacts added to the main signals from physiological, environmental, and technical sources as part of acquisition and processing [14]. These influence the quality of the actual signal and, thus, improving the quality of EEG signals has been a major but challenging task in EEG classification, and a variety of sophisticated preprocessing methods have been proposed in this context [15-18]. In [15], the authors used Discrete Wavelet Transform (DWT) for EEG signal enhancement. A tactful preprocessing can enhance the performance of post-processing stages [11]. The preprocessed EEG signals are analyzed for pertinent features extraction. Various techniques including autoregressive berg, wavelet decomposition, sub-bands entropy, chaos theory, empirical mode decomposition (EMD) and optimum allocation-based sampling (OAS) have been used for features extraction [6-8, 11-13, 19]. Subsequently, these mined features are conveyed to the classifiers for categorization. In this context, robust machine learning algorithms namely K-Nearest Neighbor (KNN), Support Vector Machine (SVM), decision trees, logistic regression, Least Square – Support Vector Machine (LS-SVM), Random Forest (RF), Rotation Forest (ROF) and Bagging have been used [6-8, 11, 13]. Considering the stages involved from preprocessing to feature extraction and classification, some approaches use domain-specific processing pipelines to streamline the development process [20]. However, these methods further reduce the flexibility and generalization of the approaches.

Deep neural networks (DNN) are known for automatic feature extraction [21] and they facilitate the end-to-end development of classifiers without the need for processing pipeline. Due to their ability to classify non-linear relations between the input and output, they have been a powerful tool for EEG classification in recent years. It has been shown that features extracted through convolutional neural networks (CNN) have more representative potential than the traditional feature extraction methods [22]. Another popular concept in DNNs is transfer learning. In transfer learning, a machine learning model learned for one task is reused in another, similar task. Thanks to the availability of several successful deep learning models in the domain of image analysis and computer vision, transfer learning in EEG signals involves converting a 1-dimensional EEG signal into a 2-dimensional image and applying an existing portion from a learned model from the images-domain for classification of the 2-dimensional EEG images.

One of the limitations of applying deep learning models is that they require a large amount of data for training the network. In case of limited data, a number of data augmentation techniques are applied to generate synthetic data [23, 24]. Deep learning methods are also known to require high computation and suffer from slow convergence [25]. But the availability of advance processing units such as the Graphical Processing Units (GPU) and the Tensor Processing Units (TPU) have solved the computation issues to a large extent [26]. To solve the convergence issue, various methods have been proposed recently. This article details the usage of such methods to achieve state-of-the-art performance on EEG signal classification.

In a review of the studies for classification of EEG signals for a wide range of tasks, by Craik et al. [10], it was found that the average accuracy of deep learning methods that used signal values as input performed better than the methods that used images or calculated features as input. It was a motivating factor for us to start our approach with the raw EEG signals as input. The review also

mentioned that no previous study had analyzed if “deep learning can achieve comparable results without any artifact cleaning or removal process.” Comparatively, as a novel contribution, the current approach does not perform any artifact cleaning and achieves high performance on various evaluation metrics.

### *1.1 Objectives and Research Contribution*

The objective of this study was to investigate a simple, but effective, deep learning model – comprising CNN architecture – for the analysis of brain EEG signals for binary classification of alcoholic and non-alcoholic persons. Unlike existing studies, which focus on the manual tasks of noise removal and feature extraction from EEG signals, the current work is novel as it uses CNN for automatic feature selection in the temporal domain of signals. To deal with the scarcity of data, data segmentation is performed without adding any artificially generated data. Furthermore, the neural network architecture utilized is very compact and consists of only four convolutional layers with associated pooling and normalization operations. The number of parameters used by our classification model is also very small ensuing efficient data processing. The article not only compares the optimal model with a baseline model but also compares with existing approaches that employ neural networks for EEG classification. Because the majority of the papers are hard or impossible to reproduce [20], we make available our code and explain in detail the stages in the development of the complete neural network architecture so that our work can be reproduced by other interested researchers. We also explain various optimization tasks to improve the accuracy of our model in the hope that the research community can benefit from applying similar techniques with better knowledge in their experiments.

The remainder of the article is organized as follows. Section 2 details the background on EEG signal-based classification and identifies some relevant work as well as a comparison of some approaches with ours. Section 3 is about the materials and methods used in this study. Section 4 describes a baseline CNN architecture for classification followed by the description of methods and their results for improving the baseline architecture. Section 5 discusses the results and discussions of the proposed result. Section 6 concludes the article with future work directions.

## **2. Background and Related Work**

Identification, analysis, and treatment of alcoholism is an important societal issue, and there have been a few studies dealing with the analysis of EEG for alcoholism detection. Several recent reviews in the last couple of years [10, 20, 25, 27] covering the state-of-the-art in EEG classification of brain-related activities have identified about a dozen domains of application. However, it seems that more efforts have been spent in EEG analyses in the other domains of neuroscience and neuronal activity classification as compared to alcoholism. Only one [27] of these studies has reported alcoholism classification using EEG and none of the cited studies have applied any form of deep learning for classification.

The absence of alcoholism-related studies in these reviews may indicate a lack of coordinated efforts but we think it is mainly because of the availability of limited data for this task. When compared to tasks of EEG analysis such as a brain-computer interface or sleep analysis, the researchers have a significant choice in terms of the number of subjects, age and gender of the subjects, duration of the recordings, etc. [20]. In the case of alcoholism, however, the UCI-KDD dataset (explained later) is the only mainstream publicly available dataset that has been used in dozens of studies. A possible reason for the lack of an alcoholism dataset may be the stricter conditions in the preparation and curation of the dataset. For example, the UCI-KDD alcoholism EEG dataset required both the alcoholic and non-alcoholic participants to have abstained from drinking alcohol in the past 30 days. Finding subjects that meet such a criterion is a difficult task. Also, people may not feel comfortable when they are being analyzed for alcoholism due to certain barriers and social stigmas such as embarrassment, fear of losing the job, or concerns about what other might think about them [5]. Despite these issues, analysis and treatment of alcoholism is an important issue for an individual as well as for society and any effort in this regard will get us one step closer to better analysis and

treatment of this disorder. The use of EEG is one important tool in advancing the research in this direction rapidly.

Biomedical Signal processing has been an active research area for long and many signal processing techniques have been applied traditionally for EEG signal processing. Then, with the advancement of computational statistics, some statistical signal processing was performed. After the dawn of machine learning the statistical signal processing emerged in machine-learning-based classification. More recently, as DNNs have shown significant potential in various pattern recognition and classification tasks, there has been a surge in the application of deep learning for EEG signal processing. Gong et al. [25] provide a detailed journey of the evolution of EEG signal classification from simple statistical methods to deep learning in the last decade.

We can divide the existing approaches for EEG-based classification into those that use feature extraction with traditional machine learning classifiers and those that use deep learning methods.

### *2.1 Feature-Extraction and Machine Learning*

Researchers have proposed several techniques for feature extraction from EEG signals, such as time-domain, frequency domain, time-frequency analyses, wavelet analyses, entropy analyses, and energy distribution [9]; and most of the good results come from the combination of two or more of such methods [28]. A major limitation of the feature extraction approaches is that these methods are specific to the dataset and cannot be generalized to other datasets with different characteristics [22] without reworking the feature extraction techniques. Another problem with feature extraction is that not only it is computation-costly, but it is also laborious and time-consuming. Furthermore, as manual data processing is highly subjective, it is unlikely that other researchers may reproduce the results. Despite these limitations, more than a quarter of the studies reviewed by [10] performed manual removal of artifacts.

More recently, the traditional techniques are augmented by the application of machine learning for signal classification. The initial machine learning approaches for EEG classification relied on manually-extracted features. These methods depend on engineering techniques such as EEG rhythm feature, analytic wavelet transform, functional connectivity, graph and spectral entropies, and empirical mode decomposition (EMD) [6, 29, 30]. The common spatial patterns (CSP) algorithm is another method to extract discriminative features of raw EEG signals [31]. Jiajie et al. [32] used approximate entropy (AE) and sample entropy (SE) as feature extractors and SVM, KNN, and ensemble bagged tree as classification methods in the development of a clinical decision support system for alcoholism classification. With this approach, they could achieve sensitivity and accuracy of up to 95% on the UCI-KDD dataset with Quadratic SVM. [8] provides a summary of some methods of feature extraction with different classifiers and their performance comparison on the UCI-KDD dataset.

Automated techniques like Independent Component Analysis (ICA) for artifact removal and Principal Component Analysis (PCA) and Local Fisher's Discriminant Analysis (LFDA) are some of the methods that are applied as preprocessing before any classification methods are applied to the EEG signals [33]. Ren and Han [34] combined linear methods (autoregression, wavelet transform, and wavelet packet decomposition) with non-linear feature extraction method (sample entropy) on the EEG signal and then remove the redundant features using class separability methods. The classification is then performed using ensemble extreme learning machines based on linear discriminant analysis (LDA). Rahman et al. [35] have shown that ICA performed better in the instance-based learning method, KNN, while PCA had better results when used with a deep learning (bi-directional long-short term memory) model. Thus, one must carefully choose a feature selection method with the type of classification method adapted. Finally, image features of the signals, such as Fourier feature maps [36] or 3D grids [37] are some feature-based methods. Saminu et al. [9] provide a summary of techniques that combine traditional feature extraction methods with machine learning classifiers for EEG signal classification.

Bavkar et al. [38] used the gamma-band power as a feature in EEG signal on the entire dataset. They compared a total of 13 machine learning classifiers such as linear discriminant, linear SVM,

quadratic SVM, etc. With 61 channels, and using the ensemble subspace KNN classifier, they got the maximum accuracy of 95.1%. To carry out a classification using reduced number of channels, they applied various optimization algorithms like genetic algorithm, binary gravitational search algorithm (BGSA), etc. but none of the optimization methods could give an accuracy of more than 94%.

Feature extraction is a time-consuming task and requires specific knowledge of the domain and expertise [25]. Also, when using traditional machine-learning methods, one must experiment with a significant subset of machine learning classifiers before being able to identify the important features and the best-performing classifier. To counter these limitations, deep learning-based approaches have emerged as a way out. Deep learning does not require manual feature selection or extraction, but important features are learned by the deep learning model automatically.

## 2.2 Deep Learning-based Approaches

DNNs are a powerful tool for classification of complex non-linear systems. Of the various deep learning architectures, CNN have been found to be the most suitable ones in situations like analysis of EEG signals and 40% of the deep neural architectures are CNN and have been the architecture type of choice since 2015 [20]. We identify some approaches that have used CNN for the classification of EEG signals. Chaabene et al. applied CNN for drowsiness detection [39]. For EEG signals, they chose 14 channels followed by preprocessing for noise removal and band annotations in the refined signals. They also used data augmentation to artificially create new training instances from the previous ones. By utilizing four convolution layers, one max-pooling layer, and two fully-connected layers, they had more than 14 million parameters in the network to classify a person either in the drowsy or awake state. Without network optimization they achieved the best test accuracy of 79%. After doing some optimization in their network architecture, by adding various normalization and dropout layers, their test accuracy improved to a maximum of 90% with 7 channels.

Qazi et al. [40] applied a multi-channel pyramidal convolutional neural network (MP-CNN) for EEG signal classification of alcoholism. They started with a total of 61 channels from five different brain regions and by assessing the performance of each channel one-by-one. Of the five developed models tested with varying architectures and several parameters, the best model uses 19 best-performing channels as input and gives an accuracy of 100% with 14,066 parameters. A major limitation of their work is the lengthy job of preprocessing the process of trial-and-error involved in the selection of the best channels.

A slightly different approach is taken by [41] as they developed two new activation functions to speed up and improve the performance of CNN in EEG classification. With one of the activations functions, they could achieve an accuracy of 92.3% on the alcohol EEG dataset, which is an improvement over the usage of the default activation functions of softmax or sigmoid.

In the case of transfer learning, Xu et al. [31] used the VGG-16 CNN model, originally developed for the general image classification task, for the classification of motor imagery (MI) EEG signals. The newer model consists of the same initial layers as used in VGG-16, except for the final output layer, which is fine-tuned in the target model using the EEG dataset. The EEG signals are converted in time-frequency spectral images using Short Time Fourier Transform (STFT) before applying them as input into the target model. The classification is done by applying 2D CNN on these images. In total, their CNN model contains 13 convolutional layers, five max-pooling layers and 2 fully-connected layers. The average reported accuracy for all subjects is 74.2%. This is 2.8% better than their designed CNN.

Transfer learning was also performed by Zhang et al. [22] that combined the low- and high-frequency signals from 11 high-variance channels from the UCI-KDD dataset. This translates into a visual heatmap of brain activity in different parts of the brain. The authors used a combination of three traditional feature extraction techniques (Gray-Level Co-Occurrence Matrix (GLCM), Hu moments, and Local Binary Patterns (LBP)) and 12 deep learning feature extraction models with seven machine learning classifiers (KNN, SVM linear/polynomial/RBF, RF, MLP, and NB). After experimenting with a total of 105 possible combinations, the model with MobileNet CNN



architecture as feature extractor and SVM-RBF as classifier achieved about 95% score in accuracy, precision, and recall metrics on the UCI-KDD EEG dataset.

The potency of deep networks is because of their depth and the activation properties of the hidden layers. As an example, [42] carried out the classification of alcoholism using EEG signals by five machine learning algorithms and an MLP (multilayer perceptron, a shallow neural network with one hidden layer only). With several feature selection approaches experimented on, the MLP did not perform as well as the rest of the algorithms. It performed worst in most of the cases.

From the above analysis, we can conclude that there is a diversity of methods for the classification of EEG signals and the competition to find the best method is still ongoing. In our opinion, the best method should be selected based on flexibility, efficiency, and overall performance on various metrics. Unfortunately, many of the mentioned approaches report only a few evaluation criteria such as accuracy, while ignoring other metrics like precision and recall, which can be possibly lower, hence, diluting the overall performance. Thus, we intend to report our performance in terms of several metrics in this article.

As identified at the start of the section, due to lack of EEG data in the domain of alcoholism, many of the approaches cited above use the UCI-KDD EEG dataset [22, 32, 34, 40, 42]. Table 1 provides a comparison of these approaches with ours.

**Table 1.** Comparison of approaches for UCI-KDD EGG classification dataset.

Approach	Feature Extractor	Classifiers	Performance
Transfer learning [22]	GLM, Hu moment, LBP + 12 CNN models	KNN, SVM linear/poly/RBF, RF, MLP and NB Best: SVM RBF	Accuracy: 95.33 Precision: 95.68 Recall: 95.00 F1-score: 95.24
[32] Machine learning	AE, SE, mean, std	SVM cubic/Quadratic, KNN, ensemble tree Best: Quadratic SVM	Accuracy: 95 Sensitivity: 95 AUC: 98
Hybrid Features + EELM [34]	AR, WT, WPD, SE, and class separability	ELM, Bagging, Boosting Best: LDA + EELM	Accuracy: 91.17
ML + MLP [39]	Min/max, mean, std, power value, Daubechies, Coiflets, Symlets and Biorthogonal wavelets	SVM, OPF, KNN, NB, MLP Best: NB	Accuracy: 99.6 Specificity: 99.6 Sensitivity: 99.6 PPV: 99.6
MP-CNN [41]	5 MP-CNN models	Best: 19 best channels in CNN with 3 convolution layers and softmax classifier	Accuracy: 100 Specificity: 100 Sensitivity: 100 F1-score: 100
Our approach - CNN	CNN	CNN with 3 convolution layers, dropout, batch normalization, and kernel regularization and softmax classifier	Accuracy: 98 Precision: 100 Recall: 96.8 F1-Score: 98.4 AUC: 98.4

### 3. Materials and Methods

#### 3.1 EEG Dataset

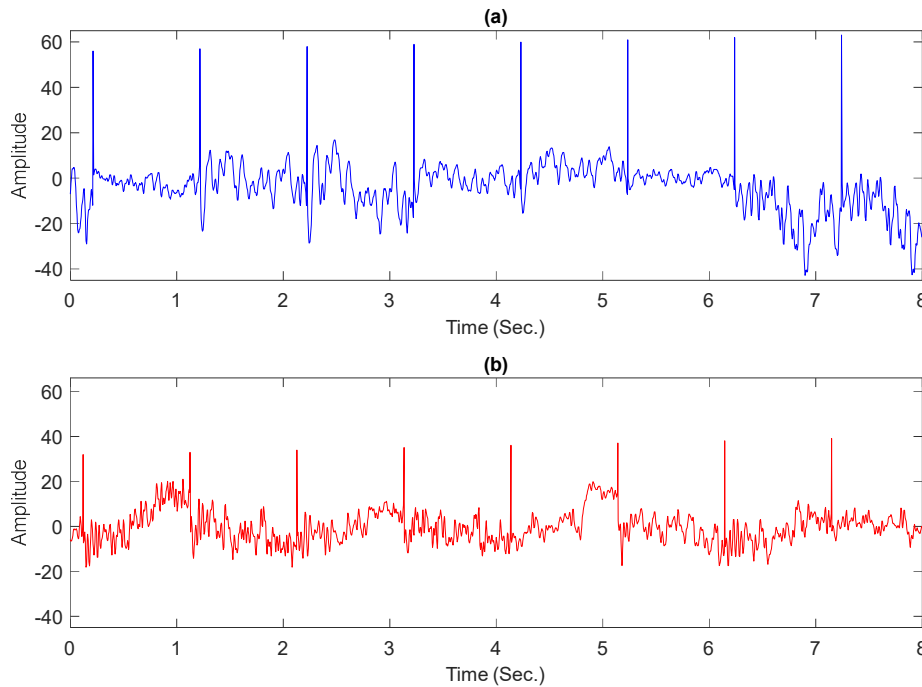
The dataset for EEG signals was obtained from the UCI-KDD<sup>1</sup> online repository that is made available for research [43]. The complete dataset consists of 122 subjects, each with 120 trials with two different stimuli. The subjects were divided into alcoholic and control groups. Each subject was exposed to either a single stimulus or two stimuli in either a matched condition where S1 was

<sup>1</sup> <https://kdd.ics.uci.edu/databases/eeg/eeg.data.html>

identical to S2 or in a non-matched condition where S1 differed from S2 [43]. After removing trails containing unwanted eye and body movements, the EEG recordings of each class were retrieved.

The EEG signals were recorded using the placement of 64 electrodes on the head according to the International 10/20 system [12]. Frontal Polar (FP), Frontal (F), Temporal (T), Central (C), Parietal (P), ground (G) and Occipital (O) areas are represented by the 10/20 system elements. The outcomes of electrodes are very sensitive to noise. Therefore, each electrode outcome is amplified and then passed through a filter with a pass-band of [0.02 Hz, 50 Hz]. This band-pass filter not only limits the signal bandwidth but also avoids the low frequency baseline wander noise. In next step, the data is sampled at a frequency of 256Hz with an Analog-to-Digital Converter (ADC) of 12-Bit resolution.

Experiments are conducted by using a partial dataset in this study since the publicly accessible entire dataset is incomplete, with certain trails containing empty files or tagged as "err.". For both classes, normal and alcoholic, 60 EEG recordings are kept. For balanced representation 30-recordings are considered from each category. Following the work presented in [6] and [44], 61 electrodes are used to acquire the EEG signals and the remaining 3 electrodes are used as reference. The Cz electrode is utilized as a reference, while the X and Y electrodes are used to capture horizontal and vertical bipolar signals, respectively. The baseline filter removes artifacts such as eyes and muscle movements from intended records [19]. Each conditioned recording has 8192 samples and is 32-seconds long. Each preprocessed record is further divided in four portions. Each portion comprises 2048 samples and is 8-seconds long. Figure 1 (a) and (b) show sample plots from normal and alcoholic persons.



**Figure 1.** Examples of EEG signals (a) Normal person (b) Alcoholic person.

### 3.2 Data Segmentation

To increase the data size each 8-Seconds length EEG portion is divided in four 2-seconds length segments, each contains 512 samples. The process of segmentation is carried out by using the rectangular window function, with the windowing operation given by Eq. (1) and (2) [17]

$$zw_n = z_n \times w_n. \quad (1)$$

$$zw_n = \sum_{-\frac{\tau}{2}}^{\frac{\tau}{2}} z_n. \quad (2)$$

Here,  $z_n$  is the digitized version of EEG band limited signal, obtained from the considered dataset [16].  $z_{wn}$  is its segmented version.  $w_n$  is the window function coefficients vector. Its length is equal to  $\tau=2.0$  seconds and it contains 512 coefficients, each of magnitude 1. This operation of windowing breaks the longer EEG signal in smaller segments. Each segment is considered as an instance. In total 960 instances are studied, out of which 480 belong to the normal class and other 480 belong to the alcoholic class.

### 3.3 Data Normalization

An important aspect of the current work is that we did not perform any special preprocessing tasks to the raw input data, with the only exception of standard normalization. In general, the signals in the data can have arbitrary positive or negative values (as shown in Figure 1). To reduce any wide dynamic ranges in the signals, it is suggested that data normalization be performed before the training process [45]. The standard normalization process scales the data so that it has a mean ( $\mu$ ) of 0 and a standard deviation ( $\sigma$ ) of 1.

### 3.4 CNN for Feature Selection

CNN's are a special kind of artificial neural network for processing data that is usually in a series in 1 dimension, e.g., speech, or EEG/ECG signals or in 2 dimensions, such as images. The principal operation in CNN is that of a *convolution*, which is a specialized kind of linear operation on two functions of a real-valued argument [46]. The first function of the convolution operation is the input and the second is known as the *kernel* while the output is known as the *feature map*. Because of the smaller size of the kernel compared to the input, convolution requires less number of parameters due to *sparse connectivity*, which not only reduces the memory requirements of the model but also improves its statistical efficiency [46]. A kernel of the convolution function can extract only one kind of feature at different input locations. To extract various kinds of features, we apply more than one convolution function in a single CNN model. The convolution function is usually followed by a pooling function, which modifies the output of the convolution layer by downsampling. It does so by replacing the output at a certain location with the summary statistics of the nearby outputs using maximum or averaging functions, etc. [46]. The convolution and pooling operations are usually represented by a single convolution block.

CNN's have been used in problems like speech recognition, image classification, recommender systems, speech recognition and text-classification. More recently, CNN's have been shown to classify EEG brain signals for autism [47], epilepsy [47-50], seizure detection in children [51], schizophrenia [52], brain-computer interface (BCI) [53], alcoholism predisposition [22, 40], drowsiness detection [39, 54] and neurodegeneration and physiological aging [55] into normal and pathological groups of young and old people.

### 3.5 Fully-Connected Layer for Classification

The convolution block is followed by fully-connected or dense classification layers. In general, several such layers may be needed for improved discrimination. Depending upon the problem, each layer may have a specific activation function. For binary classification, the Sigmoid or the logistic regression activation function is used.

### 3.6 Hyperparameters Tuning

In addition to the parameters for model definition, e.g., the number of layers, their types, and their activation functions in a neural network, a set of hyperparameters also governs the performance of a model by controlling various aspects of the algorithm's behavior [46]. The hyperparameters can be tuned manually or automatically and the range of their values can affect the time and running cost of the algorithm. Examples of hyperparameters include the type of optimizer, the learning rate, the input batch size, dropout rate, and the convolution kernel width.



### 3.7 Performance Metrics and Evaluation

For binary classification, accuracy is the best measure. We mainly performed all the optimizations using accuracy as the main metric. For the final model, we also report the cross-validation results. In addition, the widely-used metrics of precision, recall, F1-score, Cohen's Kappa, and area under curve (AUC) values were also determined.

## 4. CNN Architecture for EEG Classification

The proposed work applies CNN for the classification of an EEG signal as belonging to an alcoholic or a normal person. The main components of the model are 1-dimensional convolutional layers and a dense (fully-connected) layer.

### 4.1 1D-Convolution and Pooling

We started with a relatively small number of layers where each layer had a small kernel of size 3 and the number of filters as 8 to get a modest accuracy of about 72%. The capacity of the model was kept increasing by adding more layers and increasing the size of the kernels and the number of filters progressively in various layers. This was done by constructing a different grouping of the same number of filters ranging from 8 and increasing by a multiple of 2 until 128 filters in each layer. Similarly, by mixing various other sizes of filters we kept measuring the change in the error loss and accuracy. The finally selected model had four convolution layers with 16 filters in the first, 32 in the second, and 64 in the last layer. The kernel size was fixed at 15 in all layers. The convolution stride is also fixed to two steps in every layer. ReLU (Rectified Linear Unit) activation function is used at each convolution layer for bringing non-linearity in the process. Reducing the value of any of these parameters resulted in decreased performance, while increasing the value did not achieve any performance gains.

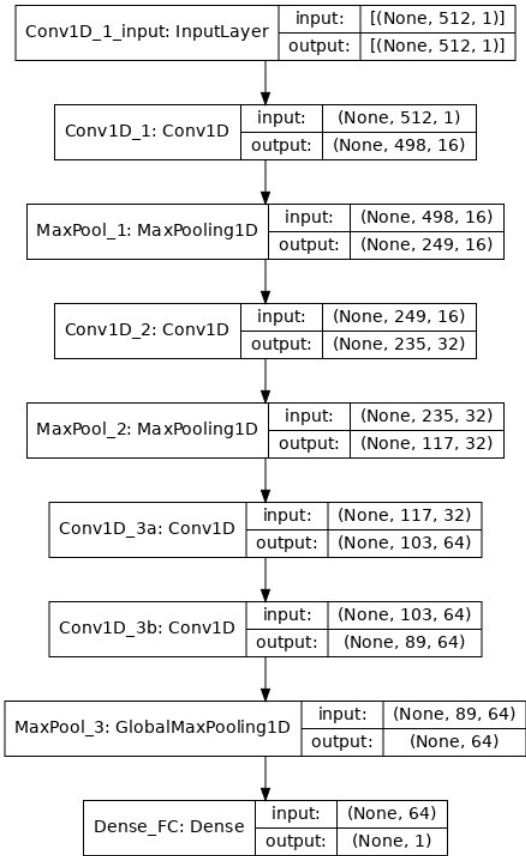
This is the general design pattern for CNN – the number of filters is increased in the latter layers, starting with a relatively small number at the start. Also, a convolution operation is mostly followed by a spatial pooling operation. Max-pooling is generally a preferred approach over other forms of pooling such as averaging giving better results. Also, the final convolution operation is generally followed by a global max-pooling operation.

The convolutional/pooling layers are responsible for feature selection. To classify the EEG signals, a single fully-connected layer is applied that uses the sigmoid activation function for binary classification and the binary cross entropy method for loss minimization. The architecture of the model is shown in Figure 2.

### 4.2 Training and Testing of the Model

The training of the model is performed by a variable size of training examples in the input dataset where the size of each example is 512 dimensions. The training is carried out using mini-batch gradient descent with a batch size of 64. Going through the successive convolution and pooling layers, the dimensions get reduced because of feature selection at each layer as shown in Figure 2.

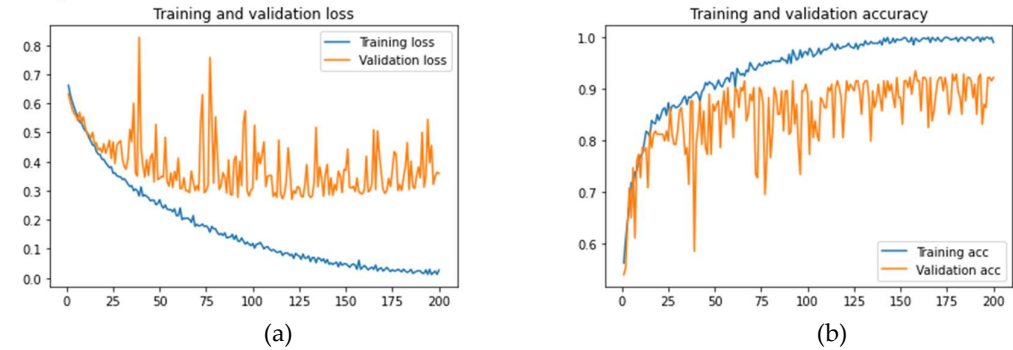
As the model starts learning, a small validation set is used after each epoch to adjust the weights. We fixed the validation set to be 20% of the size of the dataset; the model automatically chooses validation examples in each iteration. At test time, the model accepts input examples in the same dimension as the training data.



**Figure 2.** The architecture of the CNN model. There is an input layer, four convolution layers, three max-pooling layers, and a final fully-connected layer.

Figures 3 (a) and (b) show the loss and accuracy for the developed model. As can be observed, the training loss and accuracy continued to improve (low training error vs. high accuracy), but the validation loss and accuracy stalled after a few iterations. If the number of iterations was increased further, the training data would achieve an accuracy of 100% without any improvement in the validation accuracy which remained at 92% in the best case. This is a case when the model overfits the training data. By increasing the capacity of the layers in terms of bigger kernels or adding more filters cannot improve the validation accuracy as that will only help in memorizing the training data.

Thus, we needed to fine-tune the CNN model to maximize the results on the validation set. Once a desirable validation accuracy is achieved, we can then use the model on the test data to assess its final performance.



**Figure 3.** Training and validation loss and accuracy of the baseline CNN model. (a) loss (b) accuracy

### 4.3 Optimizing the Neural Network Model

We considered the previously developed model as a *baseline model* and brought improvement to it to achieve high accuracy on the validation set. To avoid overfitting of the network on training data, a technique called regularization was applied to the model. Regularization adds penalty to the model so that the model does not overfit the training data. Three regularization techniques were used: dropout, batch normalization, and L1/L2 regularization. Although regularization is an important technique for obtaining better generalization in DNNs [46], more than half of the papers in EEG classification did not mention using any regularization techniques [20]. Since regularization methods helped us in achieving good results, it would be essential to give details of the different regularization techniques we have applied in our approach.

#### 4.3.1 Batch Normalization

The purpose of normalization is “to make different samples seen by a machine-learning model more similar to each other” [56]. Normalization of the input was performed in the preprocessing step. However, during input processing, data is processed by the layers and transformed into a wide range of values that need to be normalized. Batch normalization [57] tries to achieve the same effect in the DNN. In our case, we applied batch normalization to the output of all max-pooling layers. The normalization of layers improved the accuracy by a few percent.

#### 4.3.2 Dropout Layers

Dropout layers were proposed by Srivastana et al. as a simple way to prevent the neural network from overfitting [58]. Dropout consists of dropping out some units along with their connections from the neural network, significantly reducing overfitting. Dropout regularization was used in all the layers with a dropout ratio set to 0.1. An increase in dropout ratio to various other values resulted in a decreased performance.

#### 4.3.3 L1/L2 Regularization

One way to reduce the complexity of the model is to put constraints on the model weights. This is done by adding a cost with the large weight in the loss function. This cost can come in two ways: L1 and L2 regularization. In our case, the training was regularized by applying both L1 and L2 regularizers simultaneously in the fully-connected layer. For both, the penalty weight was set to 0.01. Using these regularizers helped in the early convergence of the network and improving the accuracy as well.

#### 4.3.4 Optimizer, Learning Rate, and Early Stopping

After evaluation of various commonly used optimizers, e.g., Adam, Adagrad, Nadam, etc., the best overall results were provided by the RMSprop optimizer, which we opted to choose for all our experiments. We experimented with various learning rates. The Keras deep learning library helped decide on the optimal learning rate. Using the built-in *LearningRateScheduler* callback function and starting with a learning rate of  $1 \times 10^{-8}$  we measured the performance improvement in various training epochs while increasing the learning rate by a small factor finally reaching the value of 0.1. The best performance (in this case the minimum training loss) was achieved with a learning rate of  $6 \times 10^{-2}$ . Early stopping is a hyperparameter that allows the learning process to stop when there is no improvement in the performance beyond a certain threshold. We configured the training process to be stopped when there was no decrease in the training loss in the last 15 iterations.

## 5. Experiments and Results

### 5.1 Experimental Setup

We implemented our model in the Python language using Keras API (<https://keras.io/>) for deep learning with Tensorflow (<https://www.tensorflow.org/>) as the backend. For program development, we used the GPU environment in the Kaggle platform ([www.kaggle.com](http://www.kaggle.com)), which can execute the model several factors faster than a CPU.

The training was carried out for a total of only 100 epochs (or thousand iterations considering that each epoch comprised of 10 iterations as the input data of size 614 training examples was divided into batches of 64). It should be pointed out the our network converged very quickly as compared to many deep learning models where the number of iterations can reach up to several million [59].

### 5.2. Results

Table 2 shows the comparative results of the two models on the test set: the baseline, unregularized model versus the optimized, regularized, model for a single run. The regularized model outperforms the baseline model in every aspect, but its value is starkly high for Cohen's Kappa. The Kappa value shows the inter-rater agreement. In the case of the baseline, we can see that lower precision resulted in lower Kappa; and lower precision was the result of misclassifying a normal person as alcoholic (false positive). In contrast, the regularized CNN model had much better Kappa due to high precision and recall.

**Table 2.** Comparing performance of the baseline and regularized CNN model using various metrics.

CNN Model	Accuracy	Precision	Recall	F1-Score	AUC	Kappa
Baseline	91.15%	92.22%	89.24%	90.71%	91.08%	82.25%
Regularized	98.43%	100%	96.77%	98.36%	98.38%	96.87%

To evaluate the performance of the final, regularized model more objectively, we carried out K-fold cross validation [46] using  $K=\{3, 5, 10\}$ . K-fold cross validation is the preferred approach for model evaluation when the available data is not very big. Moreover, we also experimented on varying batch sizes ranging from  $2^2$  to  $2^8$ .

The total data ( $n=960$ ) was divided into 20% test set ( $n_{test}=192$ ), while the remaining data was split into training and validation according to the value of K. For  $K=3$ , 2/3 of data was used for training ( $n_{tr}=512$ ), while 1/3 of data was used for validation ( $n_{val}=256$ ). As the number of folds increased to 5 and 10, the training data got increased (4/5 and 9/10), while the validation data got reduced (1/5 and 1/10).

Table 3 shows the results of K-fold cross validation and variation in the batch size. The validation columns show the average accuracy of K-runs and their standard deviation for each value of K. The test columns show the accuracy of the test set. As can be seen, the average performance across the folds kept increasing and the best results on test data are obtained for 10-fold cross validation. The final column shows the best run of all the folds in the case of 10-fold cross validation, which achieves an accuracy of as high as 100%. On the dimension of batch size, we see varying results for different folds, without any specific trend, except that larger batch sizes had worsening performance both for validation and test datasets.

We can draw the following conclusion from Table 3. As the number of training samples increases from  $K=3$  ( $n_{tr}=512$ ) to  $K=10$  ( $n_{tr}=692$ ), the accuracy also increases. This is in agreement with the general principle of machine learning: increasing the number of samples increases the accuracy of the model [46].

**Table 3.** Result of K-fold cross validation (K={3, 5, 10}) accuracy with varying batch size ( $2^2$  to  $2^8$ ).

	3-fold		5-fold		10-fold		
	Validation	Test	Validation	Test	Validation	Test	Best Run
# Samples	256	192	153	192	76	192	
Batch size	$\mu$ ( $\sigma$ )		$\mu$ ( $\sigma$ )		$\mu$ ( $\sigma$ )		
4	0.92 (0.01)	0.96	0.92 (0.01)	0.93	0.95 (0.01)	0.97	0.97
8	0.93 (0.01)	0.94	0.94 (0.02)	0.95	0.94 (0.02)	0.96	0.97
16	0.92 (0.01)	0.94	0.94 (0.01)	0.96	0.94 (0.01)	<u>0.98</u>	0.96
32	0.93 (0.02)	0.95	0.93 (0.03)	0.95	0.94 (0.03)	0.96	0.97
64	0.92 (0.01)	0.95	0.94 (0.01)	0.95	0.94 (0.03)	0.95	1.0
128	0.92 (0.01)	0.90	0.92 (0.02)	0.95	0.95 (0.03)	0.95	0.99
256	0.88 (0.02)	0.89	0.90 (0.03)	0.90	0.90 (0.04)	0.89	0.97

### 5.3 Discussion

Convolutional neural networks (CNN) have been used in many application areas for classification of audio, video, or text data, while very limited work can be found in use of CNN in EEG classification. CNN's are efficient in processing inputs of large dimensions thanks to their properties of weight-sharing and sparse connections [46]. These properties not only reduce the number of parameters but also reduce training time and enhance training effectiveness [25]. Data scarcity is one issue when it comes to using deep learning models for medical analysis. On the other hand, training on large data means longer times for model fitting and evaluation. The best solutions are those that try to obtain best results on smaller datasets. As such, most of existing approaches that used the referenced dataset of UCI-KDD worked on the smaller dataset containing only 120 trials of alcoholic and 120 trials of normal subjects' diagnosis [6-8]. In this work, we also employ a total of 240 EEG segments, 120 from alcoholic and 120 from non-alcoholic subjects, respectively, without introducing any data through augmentation. It needs further experiments to see the results if the data is divided into even smaller segments.

As EEG signals are non-stationary, and their statistics vary over time, it is said that a classifier trained on one part of a user-data might not generalize to some other part of the same person [20]. However, we have seen that segmenting a signal into four parts, in our case, yields impressive results. Thus, classification using different parts of the signal may be feasible in some situations. Some approaches work on outlier detection to improve their performance, e.g., [22]. However, this results in a model that loses the quality of being robust and tolerant in the view of generalization. Thus, we do not apply any outlier detection in our methodology.

Simonyan and Zisserman [59] have shown that very deep CNN architectures with small convolution filters in all layers perform very well in areas like image recognition and natural language processing. However, in the current work, use of small convolution filters did not achieve good results. The optimal results were obtained with 1D convolution filters of size 15 in all the convolution layers. This discrepancy may be due to the reason that our architecture is not *very* deep as mentioned in [59]. Alternatively, it could be that the convolution filter size is reduced at the expense of increasing the depth of the current network. We leave the comparative analysis for a future work, as it might involve optimizing other parameters as well.

Regularization has played an important role in achieving high performance in the classification task. Batch normalization, dropout layers, kernel regularization, and learning rate were the hyperparameters that were tuned in this work. This caused only a very small change in the number of parameters to be learned by the network. Table 4 shows the comparison of the architecture of the baseline CNN model with the regularized model.



**Table 4.** A comparison of the baseline and the regularized CNN model architectures.

Baseline CNN Model			Regularized CNN Model		
Layer (type)	Output Shape	Param #	Layer (type)	Output Shape	Param #
<b>Conv1D</b>	(None, 498, 16)	256	<b>Conv1D</b>	(None, 498, 16)	256
MaxPooling1D	(None, 249, 16)	0	MaxPooling1D	(None, 249, 16)	0
<b>Conv1D</b>	(None, 235, 32)	7712	BatchNormal	(None, 249, 16)	64
MaxPooling1D	(None, 117, 32)	0	Dropout	(None, 249, 16)	0
<b>Conv1D</b>	(None, 103, 64)	30784	<b>Conv1D</b>	(None, 235, 32)	7712
<b>Conv1D</b>	(None, 89, 64)	61504	MaxPooling1D	(None, 117, 32)	0
GlobalMaxPooling	(None, 64)	0	BatchNormal	(None, 117, 32)	128
Dense	(None, 1)	65	Dropout	(None, 117, 32)	0
Total params: 100,321			<b>Conv1D</b>	(None, 103, 64)	30784
Trainable params: 100,321			<b>Conv1D</b>	(None, 89, 64)	61504
Non-trainable params: 0			GlobalMaxPooling	(None, 64)	0
			BatchNormal	(None, 64)	256
			Dropout	(None, 64)	0
			Dense	(None, 1)	65
			Total params: 100,769		
			Trainable params: 100,545		
			Non-trainable params: 224		

It is well-known that initialization of network weights is important and bad initialization can result in instability of gradients, thus, affecting the learning process. This is considered as a regularization technique [60]. However, in the current work we did not focus any weight initialization technique and random weights were chosen by default by the network using small Gaussian value with mean 0, so it is quite possible that the model performance could be affected if better/worse initial weights are chosen [61]. Other techniques such as the input perturbation technique can be explored for understanding the causal relation between the input and the decision of the model [20]. We leave this aspect of network optimization for a future work.

A weakness of the current work is that for achieving the best model, some fine-tuning of the neural network model was required. However, this tuning of parameters and hyperparameters was limited to only one specific CNN model. In comparison, other similar approaches went through a laborious job of tuning several candidate models before choosing a final one. Zhang et al. [22] experimented on 12 different classifiers, adjusting various hyperparameters of each one before settling on a final model. In total, there were 105 combinations for feature extraction methods and classifiers.

## 6. Conclusions

Detection of alcoholism is an important social issue. EEG is an important tool for identification of alcoholism. This article explains the use of CNN in classification of EEG signals for alcoholism

identification. Various regularization techniques were explained so that other researchers and practitioners can build on the present knowledge to create more efficient and better performing CNN architectures. The methods described in the article do not apply to EEG classification only and can be applied in a wide range of applications of CNN.

With the current methodology, we achieved an accuracy of up to 98% on the UCI-KDD dataset while also obtaining good results in precision, recall, AUC, etc. Further investigation may be needed in the form of weight initialization, input perturbation, and data segmentation choices, which have been identified as the near future work.

**Author Contributions:** Conceptualization, S.M.Q.; methodology, H.M.; software, H.M. validation, H.M. and S.M.Q.; investigation, H.M.; resources, S.M.Q.; data curation, S.M.Q.; writing—original draft preparation, HM and SMQ; writing—review and editing, H.M., S.M.Q. and A.Z.; visualization, H.M.; funding acquisition, A.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by Taif university Researchers Supporting Project Number (TURSP-2020/114), Taif University, Taif, Saudi Arabia.

**Acknowledgments:** We acknowledge Taif University Deanship of Scientific Research for their administrative and financial support as required in this project through the funded project TURSP-2020/114.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

ADC	Analog to Digital Converters
AE	Approximate Entropy
AUC	Area Under Curve
CNN	Convolutional Neural Networks
DWT	Discrete Wavelet Transform
EEG	Electroencephalogram
ICA	Independent Component Analysis
LDA	Linear Discriminant Analysis
LFDA	Local Fisher's Discriminant Analysis
PCA	Principal Component Analysis
NB	Naïve Bayes
ReLU	Rectified Linear Unit
SE	Sample Entropy
SVM	Support Vector Machines
KNN	K-Nearest Neighbor

## References

1. Khemiri, L., et al., *Family History of Alcohol Abuse Associated with Higher Impulsivity in Patients with Alcohol Use Disorder: A Multisite Study*. European addiction research, 2020. **26**(2): p. 85-95.
2. WHO, *Global status report on alcohol and health 2018*. 2019: World Health Organization.
3. Mehta, G. and N. Sheron, *No safe level of alcohol consumption—Implications for global health*. Journal of hepatology, 2019. **70**(4): p. 587-589.
4. Le Pham, T.T., S. Callinan, and M. Livingston, *Patterns of alcohol consumption among people with major chronic diseases*. Australian Journal of Primary Health, 2019. **25**(2): p. 163-167.
5. Schuler, M.S., et al., *Perceived barriers to treatment for alcohol problems: a latent class analysis*. Psychiatric Services, 2015. **66**(11): p. 1221-1228.
6. Anuragi, A. and D.S. Sisodia, *Alcohol use disorder detection using EEG Signal features and flexible analytical wavelet transform*. Biomedical Signal Processing and Control, 2019. **52**: p. 384-393.

7. Anuragi, A., D.S. Sisodia, and R.B. Pachori, *Automated alcoholism detection using fourier-bessel series expansion based empirical wavelet transform*. IEEE Sensors Journal, 2020. **20**(9): p. 4914-4924.
8. Siuly, S., et al., *An advanced analysis system for identifying alcoholic brain state through EEG signals*. International Journal of Automation and Computing, 2019. **16**(6): p. 737-747.
9. Saminu, S., et al., *A Recent Investigation on Detection and Classification of Epileptic Seizure Techniques Using EEG Signal*. Brain Sciences, 2021. **11**(5): p. 668.
10. Craik, A., Y. He, and J.L. Contreras-Vidal, *Deep learning for electroencephalogram (EEG) classification tasks: a review*. Journal of neural engineering, 2019. **16**(3): p. 031001.
11. Qaisar, S.M. and A. Subasi, *Effective epileptic seizure detection based on the event-driven processing and machine learning for mobile healthcare*. Journal of Ambient Intelligence and Humanized Computing, 2020: p. 1-13.
12. Acharya, J.N., et al., *American clinical neurophysiology society guideline 2: guidelines for standard electrode position nomenclature*. The Neurodiagnostic Journal, 2016. **56**(4): p. 245-252.
13. Qaisar, S.M. and S.F. Hussain, *Effective epileptic seizure detection by using level-crossing EEG sampling sub-bands statistical features selection and machine learning for mobile healthcare*. Computer Methods and Programs in Biomedicine, 2021. **203**: p. 106034.
14. Urigüen, J.A. and B. Garcia-Zapirain, *EEG artifact removal—state-of-the-art and guidelines*. Journal of neural engineering, 2015. **12**(3): p. 031001.
15. Thenappan, S., *Performance Improvement in Electroencephalogram Signal by Using DWT*. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021. **12**(10): p. 2770-2775.
16. Mishra, A., et al., *Noise removal in EEG signals using SWT–ICA combinational approach*, in *Smart Intelligent Computing and Applications*. 2019, Springer. p. 217-224.
17. Yakoubi, M., R. Hamdi, and M.B. Salah, *EEG enhancement using extended Kalman filter to train multi-layer perceptron*. Biomedical Engineering: Applications, Basis and Communications, 2019. **31**(01): p. 1950005.
18. Ghanem, N.H., A.S. Eltrass, and N.H. Ismail. *Investigation of EEG noise and artifact removal by patch-based and kernel adaptive filtering techniques*. in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 2018. IEEE.
19. Kornilov, V., M. Ostroukhov, and A. Dmitriev. *Optimization of pathology diagnosis by applying chaos theory and fractal analysis to EEG-signal processing*. in *Journal of Physics: Conference Series*. 2019. IOP Publishing.
20. Roy, Y., et al., *Deep learning-based electroencephalography analysis: a systematic review*. Journal of neural engineering, 2019. **16**(5): p. 051001.
21. Shaheen, F., B. Verma, and M. Asafuddoula. *Impact of automatic feature extraction in deep learning architecture*. in *2016 International conference on digital image computing: techniques and applications (DICTA)*. 2016. IEEE.
22. Zhang, H., et al., *Bi-Dimensional Approach based on Transfer Learning for Alcoholism Predisposition Classification via EEG Signals*. Frontiers in Human Neuroscience, 2020. **14**: p. 365.
23. Taylor, L. and G. Nitschke. *Improving deep learning with generic data augmentation*. in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2018. IEEE.
24. Wen, Q., et al., *Time series data augmentation for deep learning: A survey*. arXiv preprint arXiv:2002.12478, 2020.
25. Gong, S., et al., *Deep Learning in EEG: Advance of the Last Ten-Year Critical Period*. IEEE Transactions on Cognitive and Developmental Systems, 2021: p. 1-1.

26. Dean, J., D. Patterson, and C. Young, *A new golden age in computer architecture: Empowering the machine-learning revolution*. IEEE Micro, 2018. **38**(2): p. 21-29.
27. Khosla, A., P. Khandnor, and T. Chand, *A comparative analysis of signal processing and classification methods for different applications based on EEG signals*. Biocybernetics and Biomedical Engineering, 2020. **40**(2): p. 649-690.
28. Orosco, L., A.G. Correa, and E. Laciari, *A survey of performance and techniques for automatic epilepsy detection*. Journal of Medical and Biological Engineering, 2013. **33**(6): p. 526-537.
29. Zhu, G., et al., *Analysis of alcoholic EEG signals based on horizontal visibility graph entropy*. Brain informatics, 2014. **1**(1-4): p. 19-25.
30. Shri, T.P. and N. Sriraam, *Pattern recognition of spectral entropy features for detection of alcoholic and control visual ERP's in multichannel EEGs*. Brain informatics, 2017. **4**(2): p. 147-158.
31. Xu, G., et al., *A deep transfer convolutional neural network framework for EEG signal classification*. IEEE Access, 2019. **7**: p. 112767-112776.
32. Jiajie, L., et al., *Clinical decision support system for alcoholism detection using the analysis of EEG signals*. IEEE Access, 2018. **6**: p. 61457-61461.
33. Velu, P. and V.R. de Sa, *Single-trial classification of gait and point movement preparation from human EEG*. Frontiers in neuroscience, 2013. **7**: p. 84.
34. Ren, W. and M. Han, *Classification of EEG Signals Using Hybrid Feature Extraction and Ensemble Extreme Learning Machine*. Neural Processing Letters, 2019. **50**(2): p. 1281-1301.
35. Rahman, S., T. Sharma, and M. Mahmud. *Improving alcoholism diagnosis: comparing instance-based classifiers against neural networks for classifying EEG signal*. in *International Conference on Brain Informatics*. 2020. Springer.
36. Abbas, W. and N.A. Khan. *DeepMI: deep learning for multiclass motor imagery classification*. in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2018. IEEE.
37. Wei, X., et al., *Automatic seizure detection using three-dimensional CNN based on multi-channel EEG*. BMC medical informatics and decision making, 2018. **18**(5): p. 71-80.
38. Bavkar, S., B. Iyer, and S. Deosarkar, *Rapid screening of alcoholism: an EEG based optimal channel selection approach*. IEEE Access, 2019. **7**: p. 99670-99682.
39. Chaabene, S., et al., *Convolutional Neural Network for Drowsiness Detection Using EEG Signals*. Sensors, 2021. **21**(5): p. 1734.
40. Qazi, E.-u.-H., M. Hussain, and H.A. AboAlsamh, *Electroencephalogram (EEG) Brain Signals to Detect Alcoholism Based on Deep Learning*. CMC-COMPUTERS MATERIALS & CONTINUA, 2021. **67**(3): p. 3329-3348.
41. Bhuvaneshwari, M. and E.G.M. Kanaga. *Convolutional Neural Network for Addiction Detection using Improved Activation Function*. in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. 2021.
42. Rodrigues, J.d.C., et al., *Classification of EEG signals to detect alcoholism using machine learning techniques*. Pattern Recognition Letters, 2019. **125**: p. 140-149.
43. Zhang, X.L., et al., *Event related potentials during object recognition tasks*. Brain research bulletin, 1995. **38**(6): p. 531-538.
44. Mehla, V.K., A. Singhal, and P. Singh, *A novel approach for automated alcoholism detection using Fourier decomposition method*. Journal of Neuroscience Methods, 2020. **346**: p. 108945.

45. Sola, J. and J. Sevilla, *Importance of input data normalization for the application of neural networks to complex industrial problems*. IEEE Transactions on nuclear science, 1997. **44**(3): p. 1464-1468.
46. Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. Vol. 1. 2016: MIT press Cambridge.
47. Ranjani, M. and P. Supraja. *Classifying the Autism and Epilepsy Disorder Based on EEG Signal Using Deep Convolutional Neural Network (DCNN)*. in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. 2021.
48. Thomas, J., et al. *EEG classification via convolutional neural network-based interictal epileptiform event detection*. in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2018. IEEE.
49. Gao, Y., et al., *Deep convolutional neural network-based epileptic electroencephalogram (EEG) signal classification*. Frontiers in neurology, 2020. **11**.
50. Mao, W., et al. *EEG dataset classification using CNN method*. in *Journal of Physics: Conference Series*. 2020. IOP Publishing.
51. Abdelhameed, A. and M. Bayoumi, *A deep learning approach for automatic seizure detection in children with epilepsy*. Frontiers in Computational Neuroscience, 2021. **15**: p. 29.
52. Sun, J., et al., *A hybrid deep neural network for classification of schizophrenia using EEG Data*. Scientific Reports, 2021. **11**(1): p. 4706.
53. Lun, X., et al., *A simplified CNN classification method for MI-EEG via the electrode pairs signals*. Frontiers in Human Neuroscience, 2020. **14**.
54. Stancin, I., M. Cifrek, and A. Jovic, *A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems*. Sensors, 2021. **21**(11): p. 3786.
55. Chriskos, P., et al., *Applications of Convolutional Neural Networks in neurodegeneration and physiological aging*. International Journal of Psychophysiology, 2021. **159**: p. 1-10.
56. Chollet, F., *Deep learning with Python*. 2017: Simon and Schuster.
57. Ioffe, S. and C. Szegedy. *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. in *International conference on machine learning*. 2015. PMLR.
58. Srivastava, N., et al., *Dropout: a simple way to prevent neural networks from overfitting*. The journal of machine learning research, 2014. **15**(1): p. 1929-1958.
59. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, in *arXiv preprint arXiv:1409.1556*. 2014.
60. Li, X., et al., *EEG based emotion identification using unsupervised deep feature learning*. 2015.
61. Thimm, G. and E. Fiesler. *Neural network initialization*. in *International Workshop on Artificial Neural Networks*. 1995. Springer.