CLASSIFICATION OF EEG SIGNALS BY USING MACHINE LEARNING AND DEEP LEARNING METHODS FOR EPILEPTIC SEIZURE RECOGNITION

Aruzhan Zhaubassarova
Computer Science
Suleyman Demirel University
Kaskelen, Almaty
aruzhan.zhaubassar@gmail.com

Darkhan Suleimenov
Computer Science
Suleyman Demirel University
Kaskelen, Almaty
darkhan.amangeldievic@gmail.com

Temir Yerlan
Computer Science
Suleyman Demirel University
Kaskelen, Almaty
170103024@stu.sdu.edu.kz

ABSTRACT

Epilepsy is a neurological disorder, characterized by the occurrence of epileptic seizures. According to the World Health Organization epilepsy is neurological condition that affects about fifty million people according. [1]. More than 170 thousand [2] people in Kazakhstan have epilepsy. Machine learning can give a better life to persons with epilepsy by automatic detection of epileptic seizures. The aim of this paper at investigating the use of Convolutional Neural Network (CNN), k-nearest neighbors algorithm (k-NN), Support Vector Machine (SVM) models also efficiency and interpretation of this methods to extract features and classification from human electroencephalogram (EEG). Identification are evaluated for accuracy, specificity, and sensitivity on classification of each patient into three classes EEG signals: epileptic seizure, seizure free and pre-seizure. Average accuracy are 78,33% (GaussianNB), 73.33%(KNeighborsClassifier) and 75.00-76.6%(LinearSVC), respectively.. It is observed that, **Random Forest** classification of EEG signals gives better results and these results can also be used for detecting epileptic seizure

KEYWORDS

EEG, SVM, Linear SVM, k-NN, Neural Network, XGBoost

1. INTRODUCTION

1.1 Overview

Epilepsy is one of the most popular neurological diseases and second most common conditions after the stroke, according to the World Health Organization (WHO). Seizures may occur inspite of circumstances. People with epilepsy cannot feel themselves safely, because they always can get an sudden attack of epilepsy and cannot protect from suffocation or death.

Today, this disease is mainly treated with medications and surgeries; There is no treatment, and anticonvulsant treatment is not fully effective for all types of epilepsy. Electroencephalography (EEG) plays an important role in detecting epilepsy, because it measures the differences in voltage changes between electrodes along the scalp with sensitive ionic currents flowing inside the brain's neurons, and provides temporal and spatial information about the brain.

Detection using EEG requires a direct examination of the doctor, as well as a considerable amount of time and effort. In addition, experts with different levels of

diagnostic experience sometimes report discrepancies in diagnostic results. Therefore, the development of an automated automated method for the diagnosis of epilepsy is urgently needed. [1].

1.2 Previous papers review

In previous studies, classification of EEG signals for epileptic seizure recognition were topic for many research works before. *Pei et al., 2018* have been proposed use hand-engineered techniques for feature extraction from EEG signals, such as time domain, frequency domain, time-frequency domain, and nonlinear signal analyses *Swapna et al., 2013; Yan et al., 2017b. Hamad et al., 2017* used the discrete wavelet transform method to extract a feature set and then trained the support vector machine (SVM) with a radial basis function, showing that the proposed gray wolf optimizer SVM approach is capable of detecting epilepsy and thus further enhancing diagnosis. *Abdullah et al. (2012)* suggested using standard artificial neural network for classification. Fast Fourier Transform (FFT) and Discrete Wavelet Transform (DWT) are used as feature extraction methods. Test shows that FFT features give satisfactory recognition accuracy, however in combination with DWT, it generated more powerful input for classifier. Subasi et al. established a hybrid model to optimize the SVM parameters based on the genetic algorithm and particle swarm optimization, showing that the proposed hybrid SVM is an efficient tool for neuroscientists to detect epileptic seizures using EEG *(Subasi et al., 2017).* [3]

Based on previous papers we concluded to use neural network, xgboost, knn, svm as classifiers in this project. After comparing this classifiers we choose the best one according to accuracy for our project. Input data for classifier derived by combination time and frequency domain features.

2. DATA/FEATURE:

2.1 FEATURE EXTRACTION

Transformation of input data into a set of **feature**s. **Feature**s are distinctive properties of input patterns that help in differentiating between the categories of input patterns. Three main types of the analysis of signal feature are time domain, frequency domain.

2.1.1 Time and Frequency Domain Signals

There are number of methods for analysis of features in the time domain. According to the past research, the article (Phinyomark et al., 2012) presents 24 feature extraction, methods and analysis of EEG signals in time domain are straightforward and don't require conversion of raw data (Tkach et al., 2012). In this project even the time domain features are components of the feature vector and transformation of signals to frequency domain is required for make analysis of features. Frequency or spectral domain analysis of feature vector is more complex in computation than time domain features, but it has shown powerful ability for identification of classes in researches done before (Miller, 2008). [1]

2.2 DATA

The dataset that we use for this project was provided by our teacher Gulnur Tolebi.
EEG Signals were received from 8 patients with no epilepsy. We have three groups of signals and each group represents a different phase of seizure: free seizure, pre seizure and seizure. For each group, 100 samples of 2-second 19-channel EEG epoch were selected. Data is recorded on 19 channels(channel has 512 sample points) placed in columns.

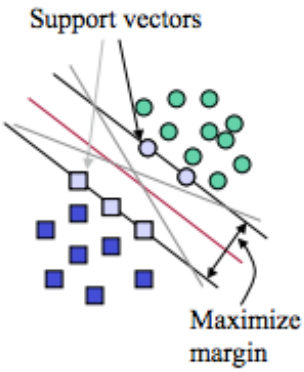
3. METHODS

3.1 MACHINE LEARNING METHODS

Machine learning methods are a popular choice recently applied for classifying of EEG. We decided to use decided to use svm, knn, xgboost, random forest as a machine learning methods.

3.1.1 Support Vector Machines (SVM)

Support vector machine **(SVM)**, one of the most commonly used classifiers, is applied to classify vectors of EEGs into classes, formally defined by a diving hyperplane.
The aim of SVM is to find a hyperplane in an N(number of features)-dimensional space that neatly classifies the data points.



SVM maximize margin over a separating hyperplane.
The decision function is completely determined (usually very small) by a subset of training samples, support vectors.

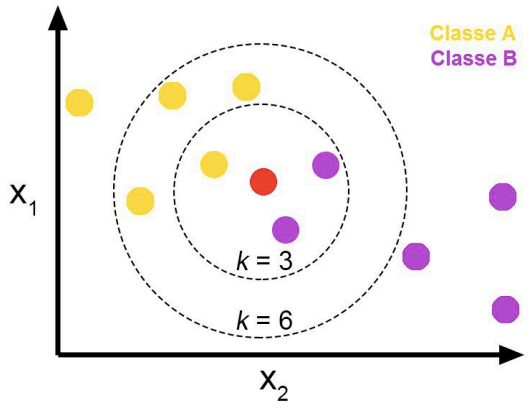
3.1.2 K Neighbors Classifier (knn)

KNN is supervised learning algorithm, which is used in machine learning and data mining. It is a classifier algorithm where the main idea is how similar or how near vectors from each other

The KNN's steps are:

1—Receive an unclassified data;

2—Measure the distance from the new data to all others data that is already classified;



3—Gets the K(parameter that you define) smaller distances;

4—Check the list of classes had the shortest distance and count the amount of each class that appears;

5—Takes as correct class the class that appeared the most times;

6 —Classifies the new data with the class that you took in step 5;

3.1.3 Naive Bayes Algorithm (GaussianNB)

A Naive Bayes classifier is machine learning model, which classifies on probabilistic idea based on Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

This theorem can find the probability of A, given that B occurred. Then, B is evidence and A is hypothesis. We suppose that features are independent, then one feature does not affect to another. That's why it is called naive.

3.1.4 XGBoost

Also we have used a Supervised Learning method - XGBoost. **XGBoost** is an algorithm that has recently been applied in machine learning

and Kaggle competitions for structured or tabular data. We think that this is a remarkably simple method that leads to results over respectable accuracy. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

Table 1	XGBoost our training parameters
colsample_bytree	0.7
learning_rate	0.05
max_depth	6
min_child_weigh	11
n_estimators	5
nthread	4
objective	multi:softprob
silent	1
subsample	0.8

3.1.4 Random Forest

The random forest classifier creates a forest(and is an ensemble of Decision Trees) makes it someway random.

Decision trees are extremely intuitive ways to classify or label objects: you simply ask a series of questions to focus on classification.

Advantages:

- flexible
- one of the most used algorithms, simplicity
- can be used for both classification and regression
- default hyperparameters give a good prediction results

Disadvantages:

- a large number of trees can make the algorithm slow and inefficient for real-time predictions

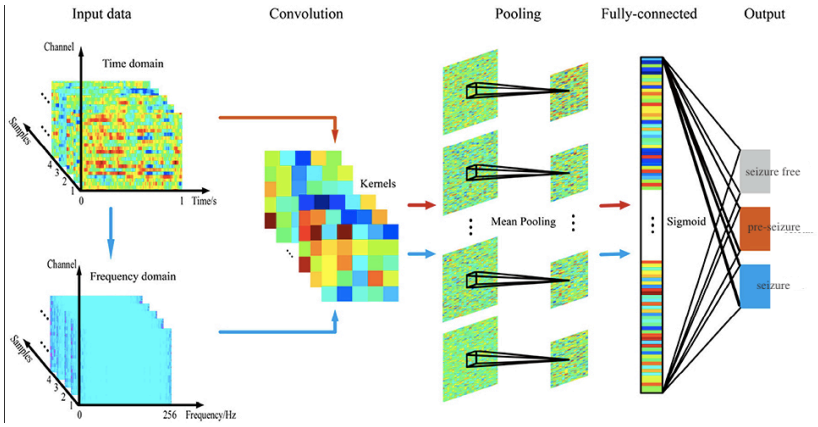
3.2 DEEP LEARNING METHODS

We decided to use decided to use CNN as a deep learning methods.

3.2.1 Convolutional Neural Network (CNN)

In deep learning, a convolutional neural network (CNN) is a class of deep neural networks, most commonly applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing They are also known as shift invariant or space invariant artificial neural networks, based on their shared-weights architecture and translation invariance characteristics. [7]

Example of our CNN shown in the Figure 3.2.1 below:

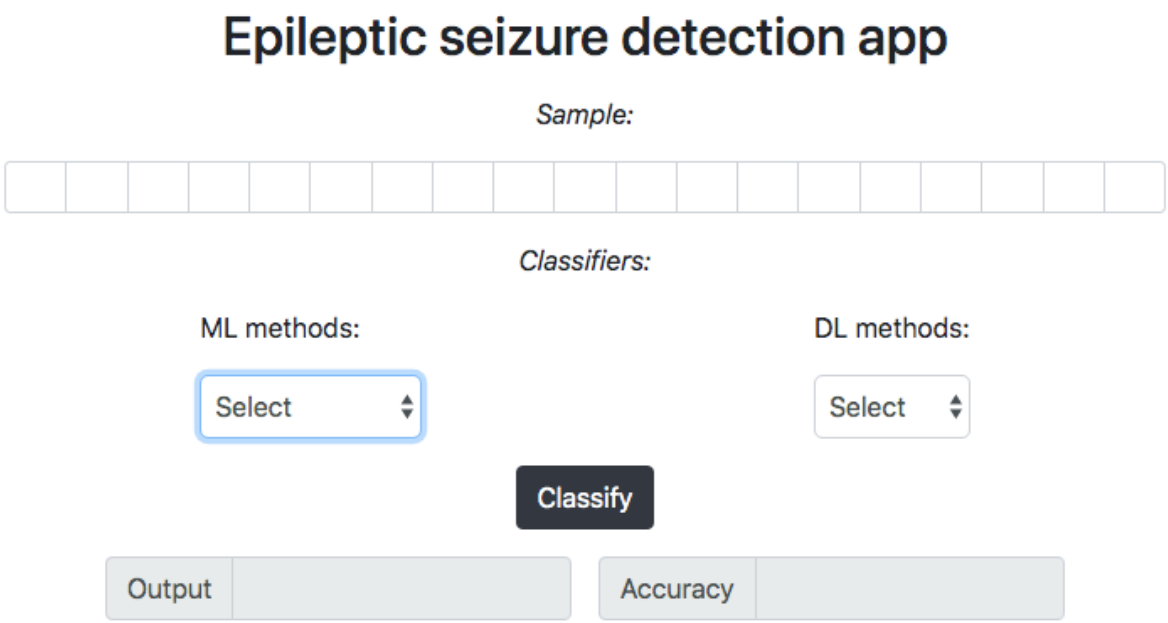


CNN advantages:

- Neural networks work best with more data.
- With any number of inputs and layers, CNN can train.
- Once trained, the predictions are pretty fast.

4. WEB APPLICATION IMPLEMENTATION

Also we have Desktop application as shown in the Figure below.



Application allows user to load EEG signal(19 inputs) sample, and classify the signal using 6 types(in select area) of classifiers.

Application was implemented on framework Electronjs. Electronjs is a framework for creating custom desktop applications with web technologies like JavaScript, HTML, and CSS.

5. EXPERIMENTAL RESULT AND DISCUSSION

Classification was done by six different classifiers. In table below we compare all our results.

Name of algorithms	Correctly Classified Instances % (value)	Incorrectl Classified Instances % (value)	Time Taken (second)
KNN	73,3 %	26,7 %	0,067
SVM	75,1 %	25,9 %	0,09
NB	78,3 %	22,7 %	0,02
XGBoost	81,67 %	19,33 %	23
RF	83,3 %	26,7 %	0,075

Name of algorithms	Correctly Classified Instances % (value)	Incorrectl Classified Instances % (value)	Time Taken (second)
CNN	79,58 %	21,42 %	73

Also we have considered model accuracy(CNN algorithm) and model loss tables and made a comparison between training and testing data(Figure 2, Figure 3) .

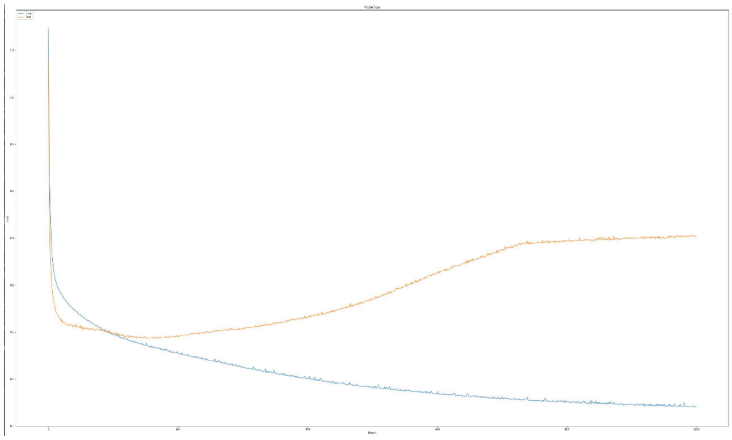


Figure 2. Model loss

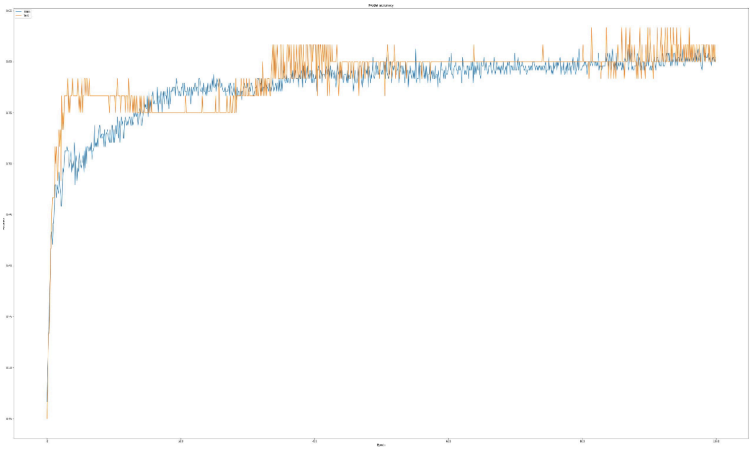


Figure 3. Model accuracy

CONSLUSION

The main objective of this work, a complete description of the method for classifying epilepsy based on EEG signals is proposed. In this paper used six different machine learning algorithms for classification. In any classification process, the important problem is to address the feature extraction and correct classification approaches. The overall highest accuracy 83.3% is achieved in the process by Random Forest classifier. In conclusion, results was good but it needs improvement.

FUTURE WORK

In future we would like to add another training data and make our model more wider and better. Also we are planning to make an implementation of our application by using other algorithms in order to get higher accuracy.

REFERENCES

[1] Tolebi, G. Classification of epilepsy using computational intelligence techniques
[2] tengrinews.kz 16 March 2018

URL: https://tengrinews.kz/kazakhstan_news/v-prezidentskoy-klinike-budut-lechit-bolnyih-epilepsiey-339970/

[3] Mengni, Zh., Cheng, T., Rui C., Bin, W., Yan, N., Ting, H., Hao, G., and Jie, X. Epileptic Seizure Detection Based on EEG Signals and CNN *Front. Neuroinform* 12:95.
URL: https://www.frontiersin.org/articles/10.3389/fninf.2018.00095/full

[4] Prabhu, medium.com, Mar 4, 2018 ,Understanding of Convolutional Neural Network (CNN)— Deep Learning
URL: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148