1-2022

DETECTION OF EPILEPSY USING MACHINE LEARNING

Balamurugan Murugesan
California State University - San Bernardino

DETECTION OF EPILEPSY USING

MACHINE LEARNING

A Project

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Computer Science

---

by

Balamurugan Murugesan

January 2022

DETECTION OF EPILEPSY USING

MACHINE LEARNING

A Project

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Balamurugan Murugesan

January 2022

Approved by:

Dr. Ronald Salloum, Advisor, Computer Science and Engineering

Dr. Ernesto Gomez, Committee Member

Dr. Yunfei Hou, Committee Member

ABSTRACT

Epilepsy is a complex neurological disorder characterized by recurrent seizures. An electroencephalogram (EEG) is typically used in the diagnosis of Epilepsy. Normally, EEGs are reviewed and analyzed by trained neurologists, but this can be time-consuming and error-prone. In this paper, we propose combining multiple classifiers in a multi-level fashion using stacked generalization to develop an effective solution for the detection of epilepsy using EEG data. Different classifiers such as Random Forest (RF), Recurrent Neural Networks (RNN), and XGBoost (XGB) were tested. The method was evaluated using Children's Hospital Boston and Massachusetts Institute of Technology (CHB-MIT) dataset. The experimental results demonstrated that the proposed method outperforms existing methods, and achieved an accuracy of 96.166%.

ACKNOWLEDGEMENTS

I take this opportunity to express my sincere thanks to my project advisor Dr. Ronald Salloum for his continuous support and guidance, which helped me in the successful completion of this project. I would also like to thank Dr. Ernesto Gomez and Dr. Yunfei Hou for serving on my committee and for providing their continuous support.

Finally, I would like to thank my family and friends for their appreciation and guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Epilepsy is a complex neurological disorder characterized by recurrent seizures that affects millions of people across the world [1]. A seizure is a sudden surge of electrical activity in the brain. The leading causes of epilepsy are stroke, tumors, and head trauma. An electroencephalogram (EEG), which records electrical activity in the brain by placing electrodes along the scalp, is often used in the diagnosis of epilepsy. Typically, neurologists will manually examine the EEG recordings to determine if there are any patterns of the disease, but this is a time-consuming and error-prone process [3]. Thus, there is significant interest in developing techniques that can detect epilepsy in an automated manner.

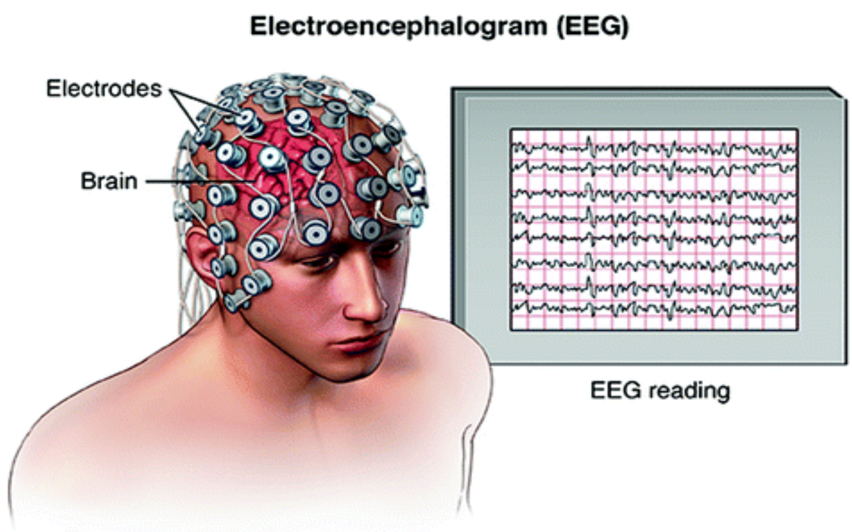

Figure 1. *An illustration of an electroencephalogram (EEG) [2].*

In this work, we propose a technique that combines multiple classifiers in a multi-level fashion using stacked generalization to develop an effective solution for the detection of epilepsy using EEG data. Different classifiers such as Random Forest (RF), Recurrent Neural Networks (RNN), and XGBoost (XGB) were tested. The method was evaluated using Children's Hospital Boston and Massachusetts Institute of Technology (CHB-MIT) dataset [6,7].

The rest of this paper is organized as follows. Related previous work is reviewed in Section 2. The proposed approach is described in Section 3. Experimental results are presented in Section 4. Finally, concluding remarks are given in Section 5.

CHAPTER TWO

RELATED WORK

A review on the existing methods for the detection of epilepsy using EEG data are discussed in this section.

In [4], the authors proposed an architecture that consists of IndependentRNN (IndRNN) blocks, one average pooling layer and two fully connected layers for the classification of seizures. The dataset used was CHB-MIT. The EEG recording were split into various segments and different segment lengths were tested (ranging from 23 to 110 seconds) and an accuracy of 87% was achieved using a segment length of 23 seconds and 15 IndRNN layers.

In [5], the authors proposed an architecture that consists of 8 convolutional layers (including both 1D and 2D convolutional layers) followed by 2 fully connected layers. The architecture was tested using two datasets: CHB-MIT and Seoul National University Hospital (SNUH-HYU) dataset. Different segment lengths were tested (10, 20, 30 seconds) and an accuracy of 85.6% was achieved using a segment length of 30 seconds for the CHB-MIT dataset.

In these existing methods, a single classifier was used. In contrast, in our work, we propose combining multiple classifiers in a multi-level fashion using stacked generalization. The proposed method is described in detail in the next section.

CHAPTER THREE

PROPOSED METHOD

3.1 Dataset

CHB-MIT dataset is an open-source dataset that is recorded from 22 patients at Children's Hospital Boston [6,7]. Data were recorded from 5 males (ages 3-22) and 17 females (ages 1.5-19). The chb21 recordings were obtained 1.5 years after the chb01 recordings from the same female patient. Each patient contains between 9 and 42 recordings, each of which includes data over 23 channels. In most cases, each recording has a duration of 1 hour, although recordings belonging to chb10 are two hours long and recordings belonging to chb04, chb06, chb07, chb09, and chb23 are four hours long.

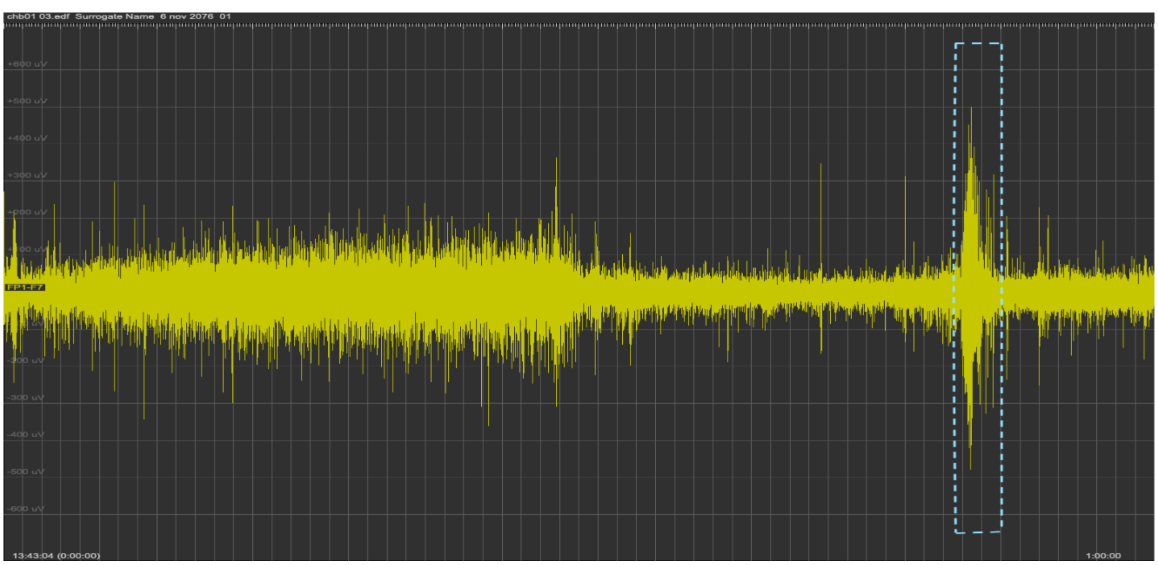

Figure 2. *One hour EEG recording of a patient where seizure occurs between 49:56 to 50:36 (indicated by the blue bounding box).*

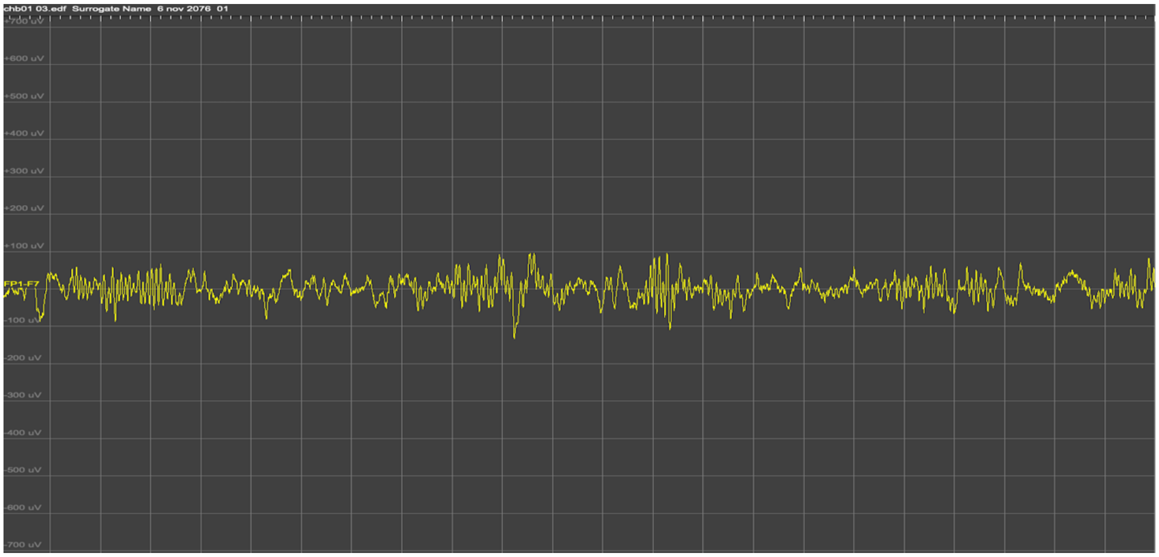

Figure 3. *EEG recording without seizures.*

All signals were sampled at 256 samples per second with 16-bit resolution. For each patient, only a subset of those recordings contains seizures. The seizure that occurs in each recording lasts between 10 and 120 seconds. Some recordings were discarded because they had channels with missing values.

Figure 2 shows an example of an EEG recording of a patient where seizure occurs between 49:56 to 50:36 (indicated by the blue bounding box), while Figure 3 shows an example of an EEG recording without any seizures.

3.2 Preprocessing and Segmentation

The EEG recordings used in this work are from the publicly available CHB-MIT dataset. Each data recording is split into non-overlapping segments by dividing the total duration of the record by 23 seconds (segment length). If the total duration is not divisible by the segment length and a seizure is happening in the remaining part, we need to make sure the remaining part with the seizure has the same segment length (23 seconds) by overlapping with the prior segment. If there is no seizure happening in the remaining part, then that part it is excluded.

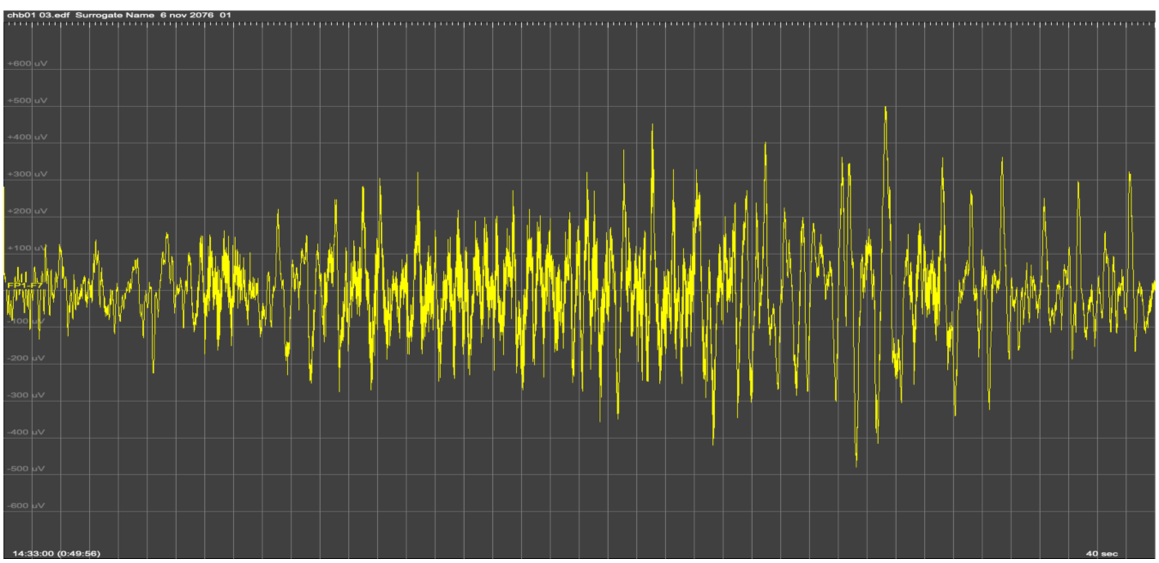

Figure 4. *Seizure segment from the recording in Fig 2.*

Using the non-overlapping technique, we obtained 506 seizure segments and 506 non-seizure segments (for a total of 1012 segments). Because the number of segments is very small, we decided to do data augmentation by preprocessing the data again, but this time split each recording into overlapping segments with a stride of 2. Using this technique, we obtained 11866 segments (5933 seizure segments and 5933 non-seizure segments). Each segment contains data from 23 channels, and each channel consists of 5888 features. The feature dimension is calculated by the product of segment duration (23 seconds) and sampling rate (256 Hz).

3.3 Model Architecture

Our proposed architecture consists of two levels or stages. The architecture is depicted in Figure 5. We evaluated two different approaches, explained as follows.

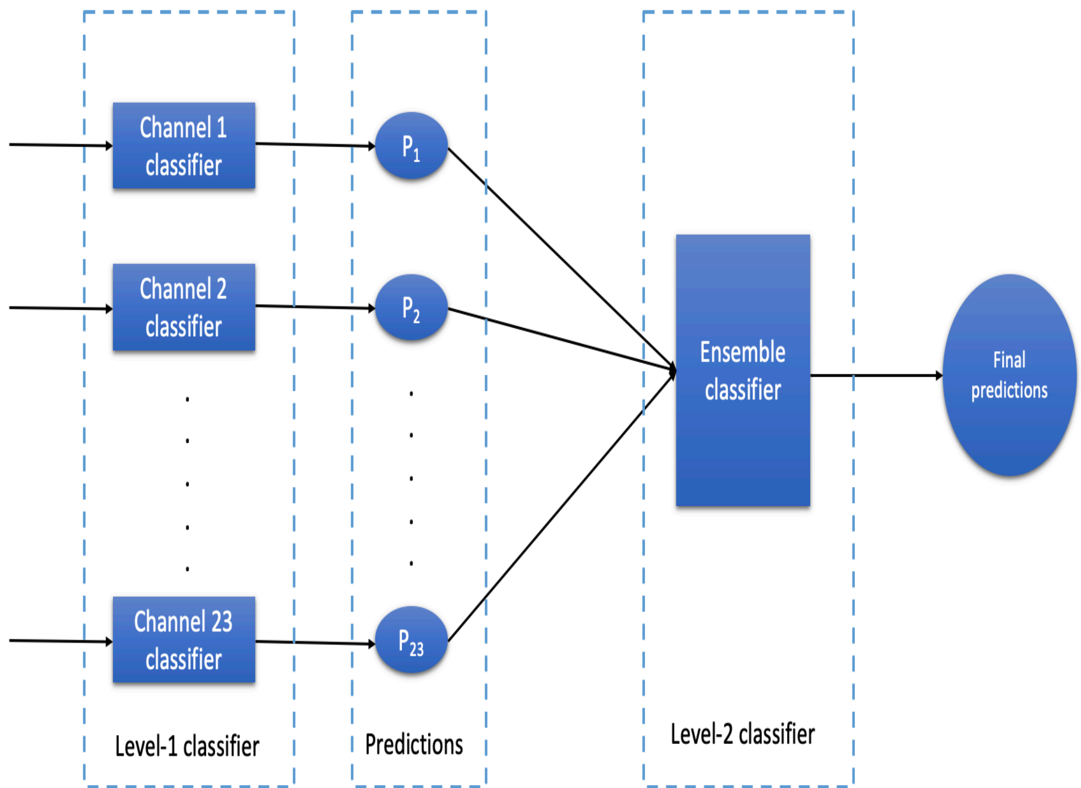

Figure 5. *Proposed Architecture.*

First Approach:

The level-1 architecture consists of 23 classifiers, one for each of the 23 channels. The available input data is divided into training and testing datasets. Each level-1 classifier is trained using the training data from the corresponding channel. The output from each level-1 classifier is then concatenated into a single new feature vector that is fed as input to train an ensemble classifier (level-2). Once the training stage is complete, we can get the predictions for the testing data by feeding it through the same multi-level pipeline. We evaluated different types of level-1 and level-2 classifiers, as discussed in Section 3.4.

Second Approach:

In the second approach, we use the same multi-level pipeline, but this time, we divide the available input data into training, validation and testing datasets. Again, each level-1 classifier is trained using the training data from the corresponding channel. Then, we determine the accuracy of each level-1 classifier using the validation data from the corresponding channel. A subset of the level-1 classifiers with the best validation accuracy is chosen and the output from those classifiers is then concatenated into a single new feature vector that is fed as input to train the ensemble classifier (level-2). Once the training stage is complete, we can get the predictions for the testing data by feeding it through the same pipeline.

3.4 Implementation

The preprocessing of the dataset as well as the implementation and training of the classifiers were performed in Jupyter notebook. The version of the Jupyter notebook used is 6.1.4. Python is used as the programming language and the version used is 3.8. We have used built-in libraries such as Sklearn, TensorFlow, and Matplotlib. We evaluated different types of level-1 classifiers, including random forest (RF), recurrent neural networks (RNNs), and XGBoost (XGB), while for the level-2 classifier, we evaluated Logistic Regression (LR), Support Vector Machine (SVM), XGB, and RF.

For the first approach, we are stacking 23 level-1 classifiers, one for each of the 23 channels. Again, we have a total of 11866 segments (5933 seizure segments and 5933 non-seizure segments). Each segment contains data from 23 channels, and each channel consists of 5888 features. The input data is split into 80% training and 20% testing. This means that for each channel, the size of the training and testing data is (9492, 5888) and (2374, 5888), respectively. Each level-1 classifier is trained using the training data from the corresponding channel and the output predictions from each level-1 classifier are concatenated into a single new feature vector, which is of the shape (9492, 23), that is used to train the ensemble classifier (level-2). Once the training stage is complete, we can get the predictions for the testing data by feeding it through the same pipeline.

In the second approach, the input data is split into 72% training, 8% validation, and 20% testing. The size of the training, validation, and testing data is (8452, 5888), (1040, 5888) and (2374, 5888), respectively. For training the ensemble classifier, only the top-k level-1 classifiers with the highest validation accuracy are chosen. We tried different values for k, as discussed in Chapter 4. The output predictions from those level-1 classifiers are concatenated into a single new feature vector, which is of the shape (8452, k), that is used to train the ensemble classifier (level-2). The testing data is fed through the same pipeline as the training data.

CHAPTER FOUR

EXPERIMENTAL RESULTS

As discussed in Chapter 2, [4] achieved an accuracy of 87% on the CHB-MIT dataset, while [5] achieved an accuracy of 85.6%. Our proposed method outperforms these existing methods. Table 1 shows the experimental results for the first approach. The table shows the accuracy of the various classifiers we have used for level-1 and level-2. It can be seen that a classification accuracy of 95.32% can be achieved when we use XGBoost classifiers for level-1 and level-2.

***Table 1**. Comparison of the results of various classifiers.*

Level-1 classifier	Level-2 classifier			
	LR	XGB	SVM	RF
XGB	95.02%	95.32%	95.02%	94.14%
RF	79.94%	73.39%	79.94%	79.14%
RNN	86.22%	83.44%	84.28%	83.1%

For the second approach, we fixed XGBoost as the level-1 classifier, since it significantly outperformed RF and RNN. Again, we tried LR, XGB, SVM, and RF for the level-2 (ensemble) classifier. Figure 6 shows the accuracy when using

XGB as the level-2 classifier, as we vary the parameter k (where k refers to the top-k performing level-1 classifiers). We see that an accuracy of 95.74% was achieved when we choose the top-five level-1 classifiers with highest validation accuracy and XGB as the level-2 classifier.

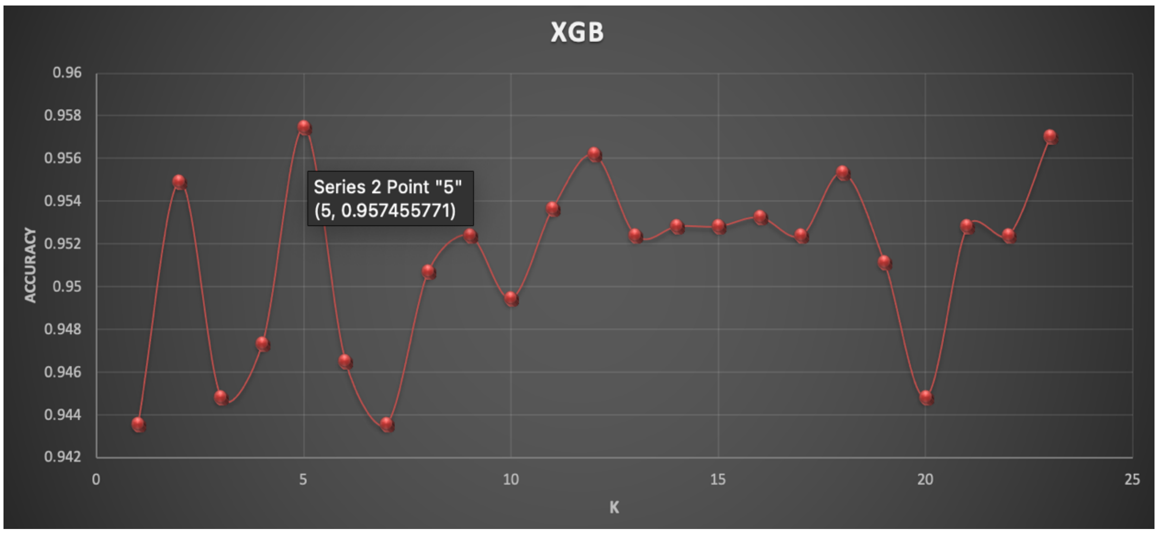

Figure 6. *The accuracy when using XGB as the level-2 classifier, as we vary the parameter k (where k refers to the top-k performing level-1 classifiers).*

Figure 7 shows that an accuracy of 96.164% was achieved when we choose the top-five level-1 classifiers with highest validation accuracy and LR as the level-2 classifier.

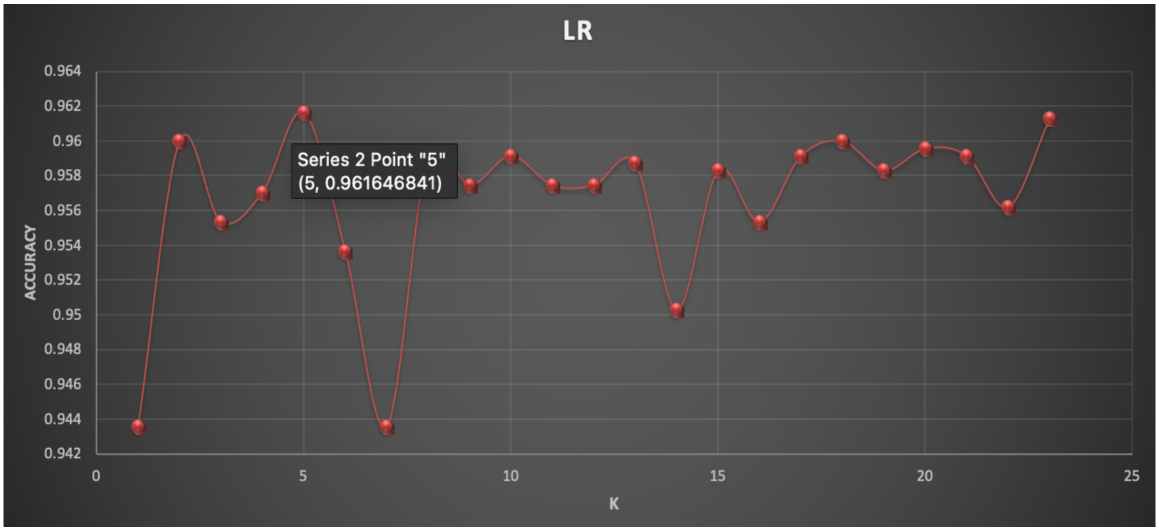

Figure 7. *The accuracy when using LR as the level-2 classifier, as we vary the parameter k (where k refers to the top-k performing level-1 classifiers).*

Figure 8 shows that an accuracy of 96.04% was achieved when we choose the top-five level-1 classifiers with highest validation accuracy and RF as the level-2 classifier.

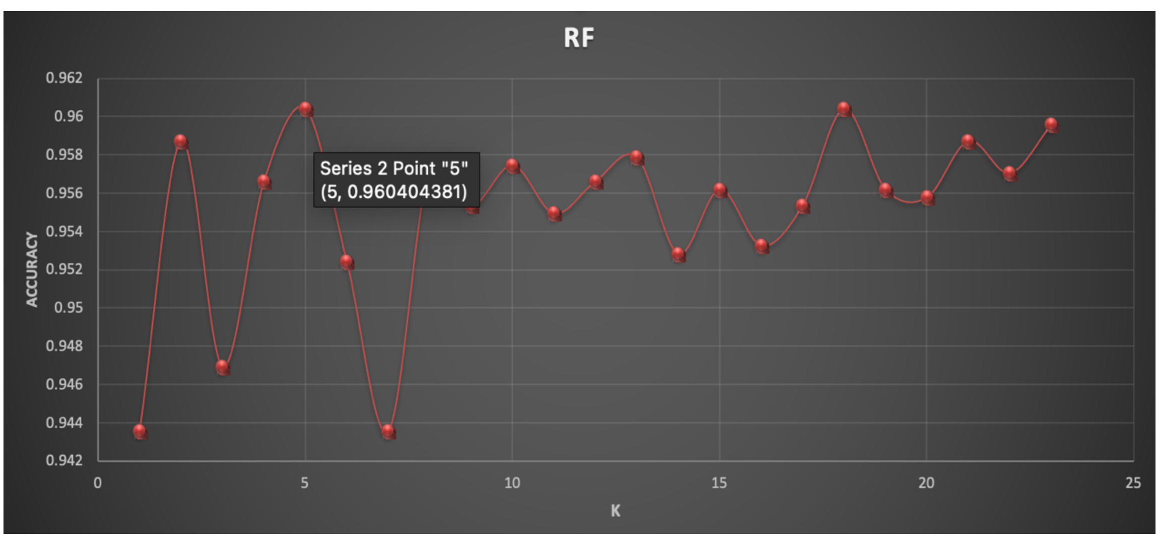

Figure 8. *The accuracy when using RF as the level-2 classifier, as we vary the parameter k (where k refers to the top-k performing level-1 classifiers).*

Figure 9 shows that an accuracy of 96.166% was achieved when we choose the top-twelve level-1 classifiers with highest validation accuracy and SVM as level-2 classifier.

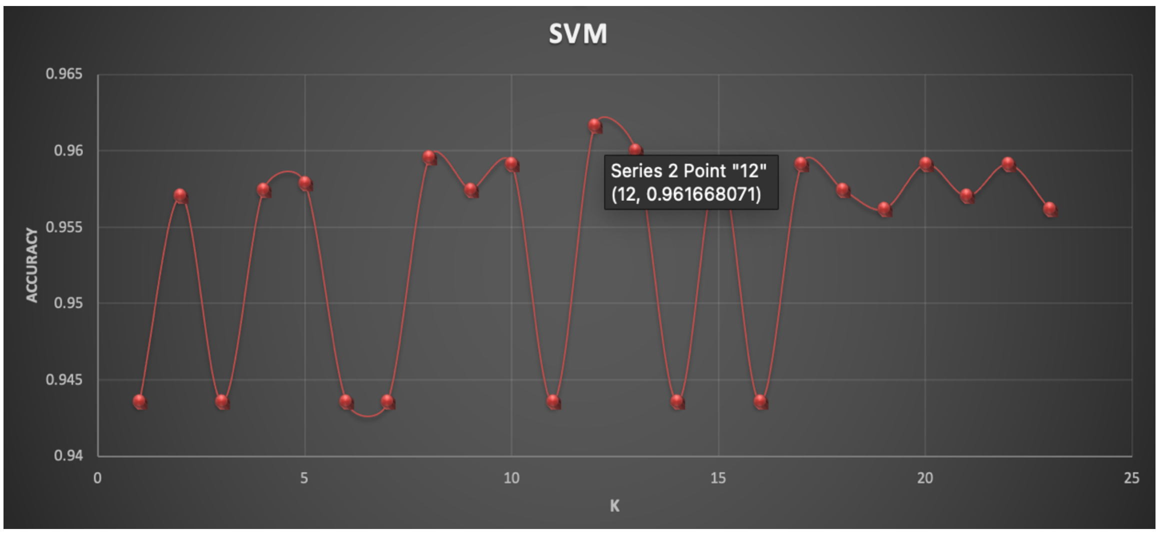

Figure 9. *The accuracy when using SVM as the level-2 classifier, as we vary the parameter k (where k refers to the top-k performing level-1 classifiers).*

For the second approach, we see that the highest accuracy we achieved is 96.166%, which is better than the highest accuracy achieved with the first approach (95.32%).

CHATPER FIVE

CONCLUSION

Automated detection of epilepsy using EEG data has many advantages such as faster diagnosis and continuous monitoring. This paper has demonstrated that combining multiple classifiers in a multi-level fashion using stacked generalization is a more effective tool for the automated detection of epilepsy, as compared to the existing methods that used the CHB-MIT dataset. In this method, we train 23 level-1 classifiers, one for each channel, and the accuracy of each level-1 classifier is determined using the validation dataset. A subset of the level-1 classifiers with the best validation accuracy is chosen and the output predictions from those level-1 classifiers are concatenated to form a single new feature vector that is used to train the ensemble (level-2) classifier. We tried various level-1 and level-2 classifiers, and the experimental results show that the proposed method achieves a maximum accuracy of 96.166%, indicating that an ensemble of classifiers works better as compared to using a single classifier. As future work, we would like to extend the proposed method to other applications, such as the detection of other diseases.

REFERENCES

[1] Duncan, John S., et al. "Adult epilepsy." *The Lancet* 367.9516 (2006): 1087-1100.

[2] Siuly, Siuly, Yan Li, and Yanchun Zhang. "Electroencephalogram (eeg) and its background." *EEG Signal Analysis and Classification*. Springer, Cham, 2016. 3-21.

[3] Smith, Shelagh JM. "EEG in the diagnosis, classification, and management of patients with epilepsy." *Journal of Neurology, Neurosurgery & Psychiatry* 76.suppl 2 (2005): ii2-ii7.

[4] Yao, Xinghua, Qiang Cheng, and Guo-Qiang Zhang. "A novel independent RNN approach to classification of seizures against non-seizures." *arXiv preprint arXiv:1903.09326* (2019).

[5] Park, Chulkyun, et al. "Epileptic seizure detection for multi-channel EEG with deep convolutional neural network." *2018 International Conference on Electronics, Information, and Communication (ICEIC)*. IEEE, 2018.

[6] Shoeb, Ali H., and John V. Guttag. "Application of machine learning to epileptic seizure detection." *ICML*. 2010.

[7] Goldberger, Ary L., et al. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals." *circulation* 101.23 (2000): e215-e220.