APRIL 19, 2018

SEIZURE DETECTION WITH EEG SIGNALS USING THE CLASSIFICATION LEARNER APPROACH

BY YOU LONG WU

DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

MCGILL UNIVERSITY, MONTRÉAL

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Electrical Engineering

Abstract

Epilepsy is characterized by unpredictable seizures secondary to electrical abnormality in the brain. Electrical activity in the brain can be monitored by electroencephalogram (EEG). This is currently the most effective and convenient tool for seizure detection. A needed tool in this disease is a model that can detect disease processes. Classification is one of the most used supervised machine learning approaches. In order to train models that are able to "learn" how to classify new observations from examples of labeled input; this research focuses on evaluating the performance of multiple classifiers for seizure detection, by applying their corresponding prediction models to labeled inputs using MATLAB's classification learner application. Many types of classifiers are used in this research such as: decision trees, support vector machines, and logistic regression, amongst others. The result has demonstrated that bagged trees of the ensemble classifiers had the highest prediction accuracy among all classifiers, which could be helpful to other researchers who wish to investigate seizure detection from EEG signals using classification methods. Potentially this could be a useful clinical tool in the future.

Résumé

L'épilepsie est caractérisée par des crises imprévisibles secondaires à une anomalie électrique dans le cerveau. L'activité électrique dans le cerveau peut être surveillée par électroencéphalogramme (EEG). C'est actuellement l'outil le plus efficace et le plus pratique pour la détection des crises. Un outil nécessaire dans cette maladie est un modèle qui peut détecter les processus de la maladie. La classification est l'une des approches d'apprentissage automatique supervisé les plus utilisées. Afin de former des modèles capables "d'apprendre" comment classer de nouvelles observations à partir d'exemples d'entrées étiquetées; cette recherche se concentre sur l'évaluation de la performance de plusieurs classificateurs pour la détection des crises, en appliquant leurs modèles de prédiction correspondants aux intrants étiquetés à l'aide de l'application d'apprentissage de la classification de MATLAB. De nombreux types de classificateurs sont utilisés dans cette recherche: arbres de décision, machines vectorielles de support et régression logistique, entre autres. Le résultat a démontré que les arbres ensachés des classificateurs d'ensemble avaient la précision de prédiction la plus élevée parmi tous les classificateurs, ce qui pourrait être utile à d'autres chercheurs qui souhaitent étudier la détection des crises à partir des signaux EEG en utilisant l'application de classification des apprenants. Potentiellement, cela pourrait être un outil clinique utile à l'avenir.

Acknowledgements

I would like to express my special thanks of gratitude to my supervisor Prof. Odile Liboiron-Ladouceur as well as my co-supervisor Dr. Edward J. Harvey, for guidance and support during my research.

Furthermore, I would also like to acknowledge the crucial role of my former supervisor and teacher Dr. Vamsy Chodavarapu, who give me the opportunity to start this research.

Finally, I would like to thank the Temple University Hospital for providing the TUH EEG Corpus that is freely available which provides all the EEG recordings needed to complete this research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABSZ	Absence seizure
AR	Averaged reference
AVG	Average
BCKG	Background, no seizure occurs
CPSZ	Complex partial seizure
CR	Correct Rate
CT	Computed tomography
DWT	Discrete wavelet transform
EEG	Electroencephalography
FFT	Fast Fourier transform
FNSZ	Focal non-specific seizure
GNSZ	Generalized non-specific seizure
LE	Linked ear reference
MRI	Magnetic resonance imaging
PSD	Power spectral density
SEIZ	Seizure events
SD	Standard deviation
SPSZ	Simple partial seizure
SVM	Support vector machines
TCSZ	Tonic clonic seizure
TNSZ	Tonic seizure

1. INTRODUCTION

1.1. Background

Epileptic seizure is a relatively common neurological disorder. It is caused by an excessive electrical activity in the brain that can affect awareness, behavior and cause abnormal movement. Epilepsy is an intrinsic electrical conduit abnormality that can occur for reasons such as low levels of oxygen, low blood sugar, brain trauma, and other metabolic reasons. A seizure event can manifest for a variable amount of time - usually from a couple of seconds to more than five minutes. Its symptoms and signs can vary and are dependent on the seizure's origin. There are different types of seizure defined as focal, tonic, partial, absence or other depending on location and extent of brain matter involved. Usually, to appropriately diagnose seizure events, an electroencephalography is used (mapping the electrical signals in major brain areas-Figure 1), along with diagnostic imaging by CT scan and MRI. In some cases depending on the presumed cause, blood tests and lumbar puncture may also be useful.

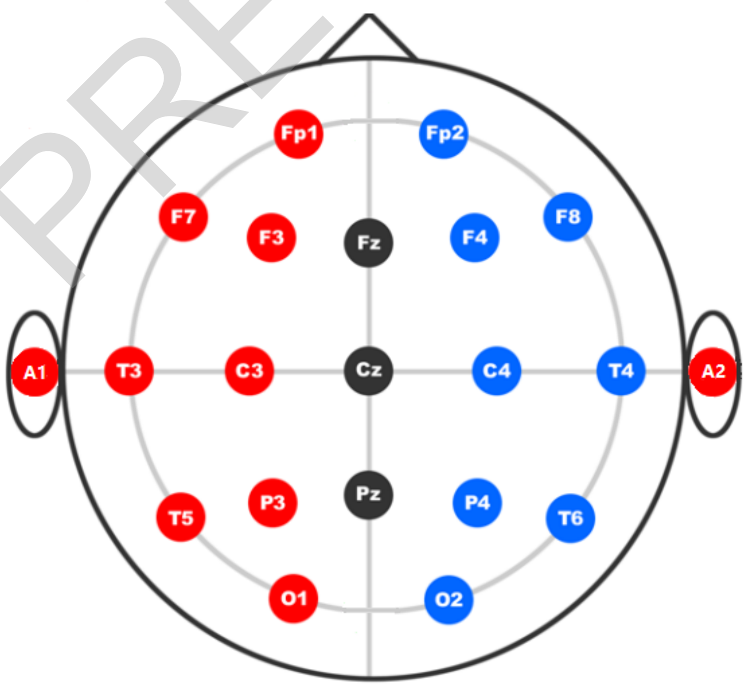

Figure 1 International 10-20 System [1]

The international 10-20 system (shown in Figure 1) is a validated method of electrode placement for EEG test. The *10* and *20* refer to the distance between adjacent electrodes spread over the scalp, meaning the distances between them are either 10% or 20% of the total front-back or right-left distance of the skull. Each electrode starts with a letter that identifies the lobe, F for frontal, T for temporal, C for central, P for parietal and O for occipital lobe. All electrodes end with a number, where even numbers refer to a position on the right hemisphere, and odd numbers refer to the left hemisphere.

Research has been performed that centered on seizure detection from EEG recordings using different classifiers for the supervised machine learning approach. The article "*A Machine Learning System for Automated Whole-Brain Seizure Detection*" [2] used the *k*-NN classifier, "*A High-Performance Seizure Detection Algorithm Based on DWT and EEG"* [3] used the support vector machine, and "*Automatic Seizure Detection in EEG using Logistic Regression and ANN*" [4] used both logistic regression and multilayer neural networks. No-one has compared the classification result of different classifiers and checked which is more accurate for seizure detection. This research reported here focuses on comparing the stratified results of multiple classifiers, in order to determine the classifier with the best prediction accuracy on seizure detection from EEG recordings.

1.2. Dataset

Data used in this research was retrieved from the TUH EEG Seizure Corpus [5], the version number is v1.1.1. Data are separated in two main directories: evaluation or training sets. The training set is used to train different prediction models using different classifiers; the evaluation set validates the performance of the classifiers on new input data. Both sets are separated into two subdirectories containing two type of EEGs: averaged reference (AR) and linked ear reference (LE). For this research, only AR EEG was used, in which outputs of all amplifiers are summed and averaged to produce a signal that is used as a common reference for all input channels. The advantage of subtracting the average signal from each channel is that the model error is averaged