Faculty of Cognitive Sciences and Human Development

A COMPARATIVE STUDY ON MACHINE LEARNING APPROACH TOWARDS EPILEPTIFORM EEG SIGNALS DETECTION

PEARLY OH BEI QING

Bachelor of Science with Honours
(Cognitive Science)
2017

UNIVERSITI MALAYSIA SARAWAK

Grade: A-

Please tick one
Final Year Project Report ☒

Masters ☐

PhD ☐

DECLARATION OF ORIGINAL WORK

This declaration is made on the 09 of JUNE 2017.

Student's Declaration:

I, PEARLY OH BEI QING , 48700, FACULTY OF COGNITIVE SCIENCES AND HUMAN DEVELOPMENT, hereby declare that the work entitled, A COMPARATIVE STUDY ON MACHINE LEARNING APPROACH TOWARDS EPILEPTIFORM EEG SIGNALS DETECTION is my original work. I have not copied from any other students' work or from any other sources with the exception where due reference or acknowledgement is made explicitly in the text, nor has any part of the work been written for me by another person.

09 JUNE 2017

PEARLY OH BEI QING (48700)

Supervisor's Declaration:

I, DR LEE NUNG KION, hereby certify that the work entitled, A COMPARATIVE STUDY ON MACHINE LEARNING APPROACH TOWARDS EPILEPTIFORM EEG SIGNALS DETECTION was prepared by the aforementioned or above mentioned student, and was submitted to the "FACULTY" as a *partial/full fulfillment for the conferment of BACHELOR OF SCIENCE WITH HONOURS (COGNITIVE SCIENCE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by: ______________________ Date: 09 JUNE 2017

(DR. LEE NUNG KION)

I declare this Project/Thesis is classified as (Please tick (√)):

☐ CONFIDENTIAL (Contains confidential information under the Official Secret Act 1972)*
☐ RESTRICTED (Contains restricted information as specified by the organisation where research was done)*
☒ OPEN ACCESS

I declare this Project/Thesis is to be submitted to the Centre for Academic Information Services (CAIS) and uploaded into UNIMAS Institutional Repository (UNIMAS IR) (Please tick (√)):

☒ YES
☐ NO

Validation of Project/Thesis

I hereby duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic and research purposes only and not for other purposes.
- The Centre for Academic Information Services has the lawful right to digitize the content to be uploaded into Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis if required for use by other parties for academic purposes or by other Higher Learning Institutes.
- No dispute or any claim shall arise from the student himself / herself neither a third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student himself/herself without first obtaining approval from UNIMAS.

Student's signature: ____________ Supervisor's signature: ____________
Date: 09 JUNE 2017 Date: 09 JUNE 2017

Current Address:
Jalan Datuk Mohammad Musa,
94300 Kota Samarahan, Sarawak, Malaysia.

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the date of restriction indicated, and the reasons for the confidentiality and restriction.

A COMPARATIVE STUDY ON MACHINE LEARNING APPROACH TOWARDS EPILEPTIFORM EEG SIGNALS DETECTION

PEARLY OH BEI QING

This project is submitted
in partial fulfilment of the requirements for a
Bachelor of Science with Honours
(Cognitive Science)

Faculty of Cognitive Sciences and Human Development
UNIVERSITI MALAYSIA SARAWAK
(2017)

The project entitled 'A comparative study on machine learning approach towards epileptiform EEG signals detection' was prepared by Pearly Oh Bei Qing and submitted to the Faculty of Cognitive Sciences and Human Development in partial fulfillment of the requirements for a Bachelor of Science with Honours (Cognitive Science)

Received for examination by:

(DR LEE NUNG KION)

Date:

15/June/17

Grade
A

ACKNOWLEDGMENTS

First and foremost, praises and sincere thanks to the God, for His almightiness and abundant blessings on cherishing me throughout the thick and thin on my way to complete this final year project smoothly.

I would like to express my heartfelt appreciation to my supervisor, Dr Lee Nung Kion, on nurturing me with invaluable guidance, patience and encouragement, for the sake of enhancing my endeavors to produce a good quality of work. You have taught me with your keen machine learning insights, especially the core methodology to conduct this project and to present the works as clearly as possible. It was indeed an honor to be under your guidance and I treasure it deeply. Thank you.

I wish to thank the Pompeu Fabra University for offering such extensive online EEG time series data sets, so that I can carry out this project smoothly and generate desirable results.

To my mother, Janice Lee, thank you for your unconditional love, prayers and sacrifices for educating and giving me unlimited strength to endure through this journey. I always know that you believed in me and supported me emotionally in all of my pursuits. Also, I am extending my sincere thanks to my family members who are always be there cheering me up and encouraging me with their best wishes.

Most significantly, this project is dedicated to my cousin brother, Patrick Ong who is one of the epileptic victim that exhibits seizures and is undergoing medication. He has been my source of inspiration. Thank you brother.

Last but not least, with a special mention to Chong Yuin Wei, Jessica Yang, Lim Sing, Ling Mee Yien, Ling Nie Hui and Wong Wan Ting, who deserve credits for your genuine support, general assistance and advices from multiple perspectives through this entire process, besides being there whenever I needed a friend.

Finally, my thanks go to all my surrounding people who have motivated me directly or indirectly to successfully complete this final year project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Electroencephalogram (EEG) signal is extensively used for the diagnosis of various kinds of neurological brain disorders. The classification of normal and abnormal electrical brain spikes through visual inspection is highly subjective and varying across medical experts. Hence, in this project, comparisons between multiple supervised learning approaches are presented, in order to discriminate those epileptiform EEG signals data from non-epileptiform with high generalizability and promising results. Furthermore, both Discrete Wavelet Transform (DWT) and Independent Component Analysis (ICA) are incorporated respectively as a preprocessing stage on reducing dimensionality, besides removing unnecessary noise adequately. Then, a set of statistical extracted features are served as input parameters to various machine learning classifiers, namely Multilayer Perceptron Neural Network (MLPNN), Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM) respectively with two discrete outputs (normal or epilepsy). As a result, the experimental outputs inferred that the wavelet coefficients which extracted by DWT, have demonstrated as the most well representation of EEG signals. Hence, the ANFIS classifier which trained on these salient features using combination of neural network learning capabilities and fuzzy logic decision making approach, has depicted the highest classification performance accuracy of 99.35% **(ICA+ANFIS)** and 99.67% **(DWT+ANFIS)** respectively. Meanwhile, others classifiers namely, MLPNN and SVM, have also proven the diagnostic results with potentially high accuracies, which are 94.39% **(ICA+MLPNN)**, 93.67% **(DWT+MLPNN)**, and 96.22% **(ICA+SVM)**, 94.39% **(DWT+SVM)** respectively, after tuning the system parameters. Therefore, these findings yield promising outcomes to be presented as a framework for training and testing epileptic prediction on EEG

data by configuring intelligent devices, so that every patient can be treated in an optimum manner, prior to surgical evaluation.

Keywords: Electroencephalogram (EEG), Epileptic seizure, Independent Component Analysis (ICA), Discrete Wavelet Transform (DWT), Multilayer Perceptron Neural Network (MLPNN), Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Machine (SVM)

ABSTRAK

Isyarat electroencephalogram (EEG) digunakan secara meluas untuk diagnosis pelbagai jenis gangguan saraf otak. Pengkelasan tersebut normal dan tidak normal otak elektrik pakuan melalui pemeriksaan visual adalah sangat subjektif dan berbeza-beza merentasi pakar-pakar perubatan. Oleh yang demikian, dalam projek ini, perbandingan antara pelbagai pendekatan pembelajaran diselia terkenal dibentangkan, untuk memilih bulu epileptiform itu EEG isyarat data dari epileptiform dengan generalizability tinggi dan menjanjikan keputusan. Selain itu, kedua-dua mengubah ubahan diskret (DWT) dan analisis komponen bebas (ICA) digabungkan masing-masing sebagai peringkat preprocessing mengurangkan dimensionality, selain daripada mengeluarkan bunyi bising tidak perlu secukupnya. Kemudian, satu set ciri-ciri diekstrak statistik disajikan sebagai input parameter untuk mesin pelbagai pembelajaran classifiers, iaitu rangkaian Neural Multilayer Perceptron (MLPNN), sistem inferens Neuro-kabur mudah suai (ANFIS) dan Mesin Vektor Sokongan (SVM) masing-masing dengan dua output diskret (biasa atau sawan). Hasilnya, keluaran eksperimen disimpulkan babawa pekali ubahan dengan yang dipetik oleh mati, telah menunjukkan sebagai mewakili Tempahan serta isyarat EEG. Oleh yang demikian, Pengelas ANFIS yang terlatih mengenai ciri-ciri-ciri yang menggunakan gabungan Rangkaian neural pembelajaran logik kabur pendekatan membuat keputusan, dan keupayaan telah digambarkan klasifikasi prestasi ketepatan yang tinggi. Sementara itu, orang lain classifiers iaitu MLPNN dan SVM, juga telah membuktikan hasil diagnostik dengan ketepatan perakaman yang berpotensi tinggi, selepas penalaan parameter sistem. Oleh itu, penemuan ini menghasilkan menjanjikan hasil yang akan dibentangkan sebagai satu rangka kerja untuk latihan dan ujian ramalan epilepsi EEG data dengan mengkonfigurasi peranti pintar, supaya setiap pesakit boleh dirawat secara optimum, sebelum pembedahan penilaian.

Kata kunci: Electroencephalogram (EEG), Sawan, analisis komponen bebas (ICA), mengubah ubahan diskret (DWT), rangkaian Neural Multilayer Perceptron (MLPNN), sistem inferens Neuro-kabur mudah suai (ANFIS), Mesin Vektor Sokongan (SVM)

CHAPTER ONE

INTRODUCTION

Introduction

This chapter relates about the **background of study**, **research problem**, **research objectives**, **research questions**, **significance** and **scope of the study**. This chapter also describes the **definition** of relevant terms and finally it closes with a short **summary**.

Epilepsy is a kind of neurological disorder in which the underlying grey matter disturbance, that causes hypersynchronization of neuronal activity in the inner brain, occurs in a transient period of time, which is known as epileptic seizures (Fisher et al., 2005). In other words, according to Acharya, Sree, Chattopadhyay, and Ang (2011), epilepsy is a physical condition that occurs in the brain and affects the nervous system. The excessive and abnormal neural activities which take place in the brain, lead to brain disturbances in which the symptoms often occur in the form of seizures. As a result, epilepsy can also cause someone to become unconscious in a short period of time depending on the degree or level of chronic epilepsy experienced by the individuals.

Thus, in order to probe the spatiotemporal dynamics of human brain, a non-invasive method, electroencephalography (EEG) is implemented by directly measuring the embedded cortical activity in a millisecond resolution (Jahankhani, Revett, & Kodogiannis, 2007). According to Mohseni, Maghsoudi, and Shamsollahi (2006), EEG can be considered as a comprehensive signal which encompasses information of the electrical potentials produced by the cerebral cortex nerve cells. Thus, the diagnosis of epileptiform brain spikes apparently rely on the components of EEG signals. After the electroencephalogram (EEG) detects and records the brain signals, the signals are then being classified according to its patterns by observing, analyzing and identifying patients with epilepsy or without epilepsy later on.

In order to achieve this goal, this project aims to analyze the EEG signals using different reliable computational methods and comparative study will be carried out towards these superiorly intelligent models. Basically, in order to solve this biomedical problem, the recognition of epileptic seizure is divided into two phases, which are feature extraction, followed by feature classification. Firstly, the EEG signals are decomposed into discrete tiny sub-bands in order to extract statistical features for the classification work, using Discrete Wavelet Transform (DWT). Besides, another comprehensive feature extractions method in which independent components analysis (ICA) is also implemented as an alternative to solve the multi-dimensionality problem of signals, in order to serve as an input for the classifiers to produce two discrete outputs (epileptic or non-epileptic seizures). Next, different kinds of feature classification models are adopted such as Multilayer Perceptron Neural Network(MLPNN), Support Vector Machine (SVM) and Adaptive Neuro-Fuzzy Inference System (ANFIS), by acting as a transportation towards the approach of machine learning for the detection of seizures activities via EEG signals, besides comparing their classification performance to obtain excellent output. All in all, the classification performance of three machine learning algorithms are compared for two feature extraction methods in terms of effectiveness.

Background of study

The interconnections and synchronizations of millions of neurons embedded in the human brain are very complicated, and resulted in non-stationary and non-linear EEG signals in which the statistical features such as mean, standard deviation and variance of EEG signals fluctuate dynamically over time (Juárez-Guerra, Alarcon-Aquino, & Gomez-Gil, 2015). In addition, in-depth research is necessary to be conducted for the sake of exploring the mechanisms which lead to epileptic disorder, besides gaining valuable insights into this unpredictable brain malfunction, via the aid of spectral analysis of EEG signals (Subasi, 2007). Undeniably, according to Meier, Dittrich, Schulze-Bonhage, and Aertsen (2008), the characteristics of epileptiform discharges portrayed on EEG are compatible or even preceding the beginning observable, an anomaly behavior. Thus, Juárez-Guerra et al. (2015) indicated that the analysis, diagnosis, detection and classification works towards the multivariate EEG signals are crucial by implementing multiple superior machine learning techniques in order to differentiate accurately the epileptic patients from normal, healthy subjects.

In addition, Darvishi and Al-Ani (2007) indicated that EEG feature extraction and feature classification techniques are undeniably a highlighted signal processing method in Brain Computer Interface (BCI) which offers a brand new dimension in human computer interface, in which it directly correlates a computer with human common sense, decision making and reasoning processes. As a result, automated epileptic seizure identification with the aid of EEG signals has become a significant research. EEG signals extraction is a non-invasive process in which it certainly opens the door to emerging possibilities for the application of effective signal processing and data analysis approaches in the detection of epileptic seizures. With the aid of huge datasets of EEG signals today, a computationally intelligent system is capable to adapt the incoming new EEG signals through learning process and take over the role of medical observers.

Hence, Smith (2005) postulated that an accurate detection system of epileptic seizures or non-seizures subgroups is vital to implement in order to clarify and ease the choice of appropriate drug treatments, besides differentiating electroencephalography (EEG) signals that indicates the brain waves activity from neurons firing in the brain, further assessment on prognosis in anti-epileptic drugs (AED) can be predicted.

Therefore, effective machine learning tools are necessary in doing the feature extractions and feature classifications work by discriminating the EEG signals of epileptic seizures or not. According to Sahin, Ogulata, Aslan, Bozdemir, and Erol (2008), these researchers found that the capability of machine learning tools in generalizing those qualitative information obtained from EEG signals to distinguish seizures events for epileptic patients, achieves a high accuracy of exceeding 90% , by applying mathematical formalization.

Various approaches of Detection Algorithms in detecting Epileptiform Activity via EEG signals

In accordance to this research, Sahin, Ogulata, Aslan, and Bozdemir (as cited in Sahin et al., 2008, p.974) have developed a study on classification of epileptic syndrome into two major subgroups which are partial seizures and generalized seizures respectively, via MLPNN as a classifier and resulted in a high accuracy of 89.2%.

Furthermore, some researchers have applied the method of statistical algorithm in machine learning in order to predict the preictal and interictal events of epileptic seizures from the EEG datasets, with the aid of support vector machine (SVM) as a classifier. As result, a significantly high sensitivity of 90% is obtained by adopting the least mean square algorithm on EEG feature extractions (Chisci, 2010). However, this study is only constrained on the focal seizures detection, whereby it is not comprehensive enough to generalize all types of seizures.

Apart from that, Pradhan, Sadasivan, and Arunodaya (1996) found that it is feasible to adopt learning vector quantization (LVQ) neural network in quantifying the epileptiform discharges (ED) as vector inputs and moreover, generalizing the "unseen" cases in which adapting supervised learning using rule-based system, with the guidance of desired training datasets. In fact, this study owns a drawback whereby false positive epoches existed as diagnosed by LVQ neural networks, due to the interference waveforms recorded in the EEG data (Pradhan et al., 1996).

Thus, these research results revealed that various applications of machine learning, be it statistical algorithms or neural network algorithms, have justified some significance on this field of study and motivated other researchers to explore more in depth.

Research Problem

Challenges on seizure detection solely based on visual inspection of EEG signals

Sadati, Mohseni, and Maghsoudi (2006) indicated that visual scanning of EEG signals on highly subjective seizures spikes might bring unreliable outcomes due to dissentience of different medical observers and cause time consumption. This is because the use of traditional methods in which the detection of seizures spikes on certain feature characteristics using naked eyes, has led to inaccurate decisions made by different medical personnels. Moreover, there exist drawbacks and inaccuracy of signal EEG when the detection of epileptic seizures via epileptiform activity within the brain is specific, but does not achieve sensitivity, due to the loss of consciousness in a transient period of time that prone to the seizures onset (Smith, 2005).

However, according to Guerrero-Mosquera, Trigueros, Franco, and Navia-Vázquez (2010), there is still lack of automated detection system available on demonstrating competitive specificity and sensitivity values, besides capable in eliminating Electrooculogram (EOG) artifacts and noises. Thus, it is vital to implement good machine learning algorithms in order to extract essential signals by discriminating real epileptic seizures from artifacts during non-epileptic events to achieve high specificity, besides classify accurately epileptic and normal subjects using effective machine learning tools (Guerrero-Mosquera et al., 2010).

Challenges on seizure detection due to inadequacy of appropriate combination of effective machine learning tools

Apart from that, basically, the scalp EEG plays the role of reflecting the rhythmic activities of the hyperactive neurons on the brain surface; meanwhile, it fails to record the hypersynchronization of neuronal activities in which the underlying grey matter disturbance located in the deep brain (Shoeb, 2009). As a result, the scalp EEG may convey a result of delayed onset activities, coupled with some background noise such as muscle contractions, sleep deprivations and eye blinks, which consequently impede the performance of EEG signals and result in high probability of false alarm (Shoeb, 2009).

However, the constraints of computational and machine learning are known to be an existing hard-to-solve problem when encounter with vast amount of throughput EEG recording data whereby a fast learning tool is required to extract and normalize those datasets for the sake of detecting epileptic seizures (Saulnier-Comte, 2013). Besides that, He and Garcia (2009) postulated that the events of seizures onset normally only last for less than 5 minutes, thus this situation has subsequently impeded the performance of conventional machine learning in which no modification on the combination of architecture of neural network to improve classification accuracy. As a result, the datasets will become highly imbalanced due to the period of seizures onset will be measured much lower than the period of interictal.

Despite there are multiple of current studies conducted involving the approach of detection of epileptic seizures from EEG signals using different intelligent detection algorithms, there is still a gap in which inadequate of previous studies which attempt to do the comparative study between multiple machine learning algorithms in the detection of presence of epilepsy and yet, comparison between various types of extracted features in effectively discriminating various input vectors. Hence, this is the first approach towards the diagnosis of epilepsy using multiple advanced machine learning algorithms based on EEG signals, which will be further decomposed into multiple extracted features that act as inputs into different classifiers for effective diagnosis.

Research Objectives

General Objective

The purpose of this research is to evaluate the feature extractions and classifications performance of different machine learning tools for classification of epileptiform EEG signals.

Specific Objectives

1. To evaluate the effectiveness of two different feature extractions models which are Independent Component Analysis (ICA) and Discrete Wavelet Transform (DWT).
2. To evaluate the effectiveness of three different classifiers which are Multilayer Perceptron Neural Network (MLPNN), Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM), in classifying those extracted features.

Research Qnestions

1. How good can those extracted features perform in discriminating different input objects?
2. Does different machine algorithms bring different outcomes on the epilepsy detection?

Significance of Study

This study aims to significantly reduce human errors in interpreting visually the EEG signals by enhancing the automated diagnosis system of epilepsy to avoid the happening of False Negative (FN) of sub-band in which any epileptic sub-band is being diagnosed as non-epileptic result. As a result, physicians are able to spot the precursory signals of ictal seizures event before a severe one manifests.

This study is also a prototype which targets to implement effective algorithms and be capable in generalizing a large proportion of epileptic seizures without the need of inputting any prior experiences or knowledge of input objects. As a result, the intelligent detection algorithms are capable in classifying those new features which never come across before and thus, a positive outcome of early detection can improve the quality of patient care.

Scope of Study

In this study, the area of concern will be narrowed down into modeling an intelligent classifier with the aids of training datasets, in order to generate high generalizability towards the testing datasets. In fact, the real-time problem is tentatively not applicable to this specific system.

Definition of Terms

In this study, relevant of terms are defined conceptually.

Feature extraction. According to Guyon and Elisseeff (2006), feature extraction is an approach on picking a good data representation by using appropriate measurements and specific domain, where effort on converting raw data sources into a set of useful features can be done. In order words, it is a kind of feature selection which involves in multiple functions such as to choose significant and relevant features, reduce multi-dimensionality data, improve learning performance to obtain a high predictive accuracy, as well as to understand the incoming data by gaining knowledge through the process of generation of data (Guyon & Elisseeff, 2006).

Feature Classification. Tang, Alelyani, and Liu (2014) indicated that classification can be defined as the identification of problem in which a set of training datasets containing instances is categorized according to its membership in a certain domain. However, in the prediction phase of feature classification, those extracted features which signify data, will then perform the classification work by mapping those feature represented data to labels (Tang et al., 2014). In this project, the real word problem can be modelled as a classification issue, in which diagnosis of epileptic or non-epileptic seizures through the discrimination on epileptiform or non-epileptiform EEG signals.

Summary

Preliminary study on comparison between multiple learning algorithms on the ability of discriminating various kinds of new input vectors to motivate us in early diagnosis of epileptic seizures before the manifestation of severe condition. Besides that, the comparison approach towards the multiple extracted features will be conducted so that efficient classification work can be carried out rapidly without always tuning backward to update the weight values.

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter discusses about those painstaking efforts which have been carried out by different researchers previously on the application of multiple intelligent machine learning algorithms in the detection of presence of epileptic seizures and also various alternatives in interpreting EEG signals so that an appropriate and accurate outcome can be achieved.

Phase One: Feature Extraction Models (Different Researchers' Work)

Discrete wavelet transform (DWT). DWT is a spectral analysis approach in analyzing fluctuated signals such as EEG signals which possess the characteristics of non-stationary by normalizing the signals with time-frequency representation (Orhan, Hekim, & Ozer, 2011). In addition, Adeli, Zhou, and Dadmehr (2003) postulated that transient features are possible to obtain and allocate in accurate time-frequency context via decomposition of EEG signals into discrete wavelets. Basically, signal analysis of EEG signals using DWT is necessary to consider two important aspects, in which ideal wavelet choice selection and number of signal decomposition levels (Subasi, 2007). Thus, Alarcon Aquino and Barria (2009) indicated that Daubechies wavelets of order 4 (db4) is the most ideal in computing the wavelet coefficients in existing studies as it comprises smoothing features which facilitates it in diagnosing changes in EEG signals more accurately. Meanwhile, Subasi (2007) asserted that the ideal number of decomposition levels was selected to be 4 as there is no desired frequency component embedded in EEG signals above 30Hz. Hence, DWT acts as a decomposer whereby it decomposes the EEG signals into different sub-bands in order to obtain the wavelet coefficients for each sub-bands, with different statistical features such as Minimum (MIN), Maximum (MAX), MEAN, and