# EEG-BASED PERSON AUTHENTICATION MODELLING USING INCREMENTAL FUZZY-ROUGH NEAREST NEIGHBOUR TECHNIQUE

LIEW SIAW HONG

MASTER OF SCIENCE IN INFORMATION AND COMMUNICATION TECHNOLOGY

2016

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

# Faculty of Information and Communication Technology

## EEG-BASED PERSON AUTHENTICATION MODELLING USING INCREMENTAL FUZZY-ROUGH NEAREST NEIGHBOUR TECHNIQUE

**Liew Siaw Hong**

**Master of Science in Information and Communication Technology**

**2016**

# EEG-BASED PERSON AUTHENTICATION MODELLING USING INCREMENTAL FUZZY-ROUGH NEAREST NEIGHBOUR TECHNIQUE

LIEW SIAW HONG

A thesis submitted
in fulfillment of the requirements for the degree of Master of Science
in Information and Communication Technology

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2016

## DECLARATION

I declare that this thesis entitled “EEG-based Person Authentication Modelling using Incremental Fuzzy-Rough Nearest Neighbour Technique” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

| | | |
|---|---|---|
| Signature | : | …………………………………….. |
| Name | : | …………………………………….. |
| Date | : | ………………………….……… |

## APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in term of scope and quality for the award of Master of Science in Information and Communication Technology.

| | | |
|---|---|---|
| Signature | : | …………………………….. |
| Supervisor Name | : | …………………………….. |
| Date | : | ………………….………… |

## DEDICATION

To my beloved parents, Mr. Liew Ted Kion and Mrs. Lee Chiu Lin, your love and support are my greatest inspiration upon accomplish this study.

To my dearest supervisors, Associate Professor Dr. Choo Yun Huoy, Dr. Low Yin Fen and Dr. Zeratul Izzah Binti Mohd Yusoh for being responsible, receptive and always by my side to encourage and motivate me.

To my dear friend, especially Ong Phaik Ling for your support and motivation throughout this study.

# ABSTRACT

High level security has nurtured the arisen of Electroencephalograms (EEG) signals as a noteworthy biometrics modality for person authentication modelling. Modelling distinctive characteristics among individuals, especially in a dynamic environment involves incremental knowledge updates from time to time. K-Nearest Neighbour (KNN) is a well-known incremental learning method which applies First-In-First-Out (FIFO) knowledge update strategy. However, it is not suitable for person authentication modelling because it cannot preserve the representative EEG signals patterns when individual characteristics changes over time. Fuzzy-Rough Nearest Neighbours (FRNN) technique is an outstanding technique to model uncertainty under an imperfect data condition. The current implementation of FRNN technique is not designed for incremental learning problem because there is no update function to incrementally reshape and reform the existing knowledge granules. Thus, this research aims to design an Incremental FRNN (IncFRNN) technique for person authentication modelling using feature extracted EEG signals from VEP electrodes. The IncFRNN algorithm updates the training set by employing a heuristic update method to maintain representative objects and eliminate rarely used objects. The IncFRNN algorithm is able to control the size of training pool using predefined window size threshold. EEG signals such as visual evoked potential (VEP) is unique but highly uncertain and difficult to process. There exists no consistant agreement on suitable feature extraction methods and VEP electrodes in the past literature. The experimental comparison in this research has suggested eight significant electrodes set located at the occipital area. Similarly, six feature extraction methods, i.e. Wavelet Packet Decomposition (WPD), mean of amplitude, coherence, cross-correlation, hjorth parameter and mutual information were used construct the proposed person authentication model. The correlation-based feature selection (CFS) method was used to select representative WPD vector subset to eliminate redundancy before combining with other features. The electrodes, feature extraction, and feature selection analysis were tested using the benchmarking dataset from UCI repositories. The IncFRNN technique was evaluated using a collected EEG data from 37 subjects. The recorded datasets were designed in three different conditions of ambient noise influence to evaluate the performance of the proposed solution. The proposed IncFRNN technique was compared with its predecessor, the FRNN and IBk technique. Accuracy and area under ROC curve (AUC) were used to measure the authentication performance. The IncFRNN technique has achieved promising results. The results have been further validated and proven significant statistically using paired sample t-test and Wilcoxon sign-ranked test. The heuristic incremental update is able to preserve the core set of individual biometrics characteristics through representative EEG signals patterns in person authentication modelling. Future work should focus on the noise management in data acquisition and modelling process to improve the robustness of the proposed person authentication model.

# ABSTRAK

*Ciri-ciri keselamatan di tahap yang tinggi telah menjadikan isyarat EEG sebagai modaliti biometrik yang baik untuk memodelkan pegesahan individu. Pemodelan ciri-ciri yang berlainan bagi setiap individu terutamanya dalam persekitaran yang dinamik melibatkan penambahan pengetahuan dari masa ke masa. Teknik "K-Nearest Neighbour" (KNN) adalah teknik "incremental learning" yang terkenal dengan menggunakan strategi "First-In-First-Out" (FIFO). Namun begitu, strategi ini tidak sesuai untuk pengesahan individu kerana ia tidak dapat mengekalkan isyarat perwakilan indvidu EEG yang berubah dari masa ke masa. Teknik "Fuzzy-rough nearest neighbour" (FRNN) adalah teknik yang terbaik untuk memodelkan ketidakpastian dalam keadaan data yang tidak sempurna. Pelaksanaan model FRNN tidak direka untuk "incremental learning" kerana teknik ini tidak mempunyai fungsi untuk menyesuaikan diri dengan pembentukan semula dan pembaharuan granul pengetahuan. Oleh itu, kajian ini bertujuan untuk mereka bentuk teknik "Incremental FRNN" (IncFRNN) untuk pengesahan individu dengan menggunakan ciri-ciri yang diekstrak daripada isyarat EEG menerusi elektrod VEP. Teknik IncFRNN mengemaskini set latihan dengan menggunakan kaedah heuristik untuk mengekalkan perwakilan objek dan menghapuskan objek yang jarang digunakan. Teknik IncFRNN boleh mengawal saiz kumpulan latihan dengan menentukan nilai ambang saiz tetingkap. Isyarat EEG seperti "visual evoked potential" (VEP) adalah unik tetapi sukar untuk diproses. Sehingga kini tiada lagi penetapan yang konsisten bagi kaedah pengekstrakan ciri dan elektrod VEP untuk pengesahan individu. Oleh itu, kajian ini akan membandingkan saluran elektrod yang biasa digunakan dalam penerbitan penyelidikan semasa, dan mencadangkan 8 set elektrod yang terletak di sekitar kawasan belakang. Enam kaedah pengestrakan iaitu "Wavelet Packet Decomposition (WPD), mean of amplitude, coherence, cross-correlation, hjorth parameter" dan "mutual information" digunakan untuk model pengesahan individu. Kaedah "correlation-based feature selection" (CFS) digunakan dalam pemilihan vektor perwakilan WPD untuk membuang lebihan vektor sebelum menggabungkannya dengan ciri-ciri yang lain. Analisis elektrod, pengekstraan ciri dan analisa pemilihan sifat dinilai dengan menggunakan dataset dari UCI. Teknik IncFRNN dinilai dengan menggunakan data EEG yang direkodkan daripada 37 subjek. Dataset yang direkodkan telah direka dalam tiga keadaan yang mempunyai tahap kebisingan yang berbeza untuk menilai prestasi pengelasan yang dicadangkan. Teknik IncFRNN dibandingkan dengan teknik terdahulu iaitu, FRNN dan KNN. Ketepatan dan luas kawasan di bawah lengkung ROC (AUC) telah digunakan untuk menilai prestasi pengelasan. Teknik IncFRNN mencapai hasil penilaian yang cemerlang. Hasil penilaian telah disahkan dan dibuktikan mempunyai perbezaan yang signifikan secara statistik dengan menggunakan "paired sample t-test" dan "Wilcoxon sign-ranked test". Pengubahsuaian heuristik dapat mengekalkan ciri-ciri biometrik individu melalui wakil isyarat EEG dalam pengesahan individu. Penambahbaikan kajian boleh dilakukan dengan memberi fokus kepada pengurusan bunyi dalam pemerolehan data dan proses pemodelan untuk meningkatkan keteguhan model pengesahan individu.*

**ACKNOWLEDGEMENTS**

First and foremost, I would like to take this opportunity to express my sincere acknowledgement to my main supervisor, Associate Professor Dr. Choo Yun Huoy from the Faculty of Information and Communication Technology Universiti Teknikal Malaysia Melaka (UTeM) for her essential supervision, support and encouragement towards the completion of this thesis. Special thanks to Ministry of Higher Education Malaysia and UTeM through the Fundamental Research Grant Scheme, FRGS/2/2013/ICT02/FTMK/ 02/6/F00189.

I would also like to express my greatest gratitude to Dr. Low Yin Fen from the Faculty of Electronics and Computer Engineering, co-supervisor of this project for her advice and suggestions in Electroencephalograms (EEG) data acquisition and pre-processing of raw data.

Particularly, I would also like to express my deepest gratitude to Dr. Zeratul Izzah Binti Mohd Yusoh from the Faculty of Information and Communication Technology, co-supervisor of this project for her advice and comments to make this project more perfect.

I would like to express my sincere thanks to my beloved parents and siblings for their moral support in completing this degree. Lastly, thank you to everyone who had been to the crucial parts of realization of this project.
.

**TABLE OF CONTENTS**

# LIST OF TABLES

## LIST OF FIGURES

## LIST OF APPENDICES

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| AEP | - | Auditory Evoked Potential |
| AI | - | Artificial Intelligence |
| ANN | - | Artificial Neural Network |
| AR | - | Auto-Regression |
| ATM | - | Automated Teller Machine |
| AUC | - | Area Under ROC Curve |
| BCI | - | Brain Computer Interface |
| BNC | - | Bayonet Neil-Concelman |
| CFS | - | Correlation-based Feature Selection |
| CIT | - | Computational Intelligence and Technology |
| CPU | - | Central Processing Unit |
| DB4 | - | Daubechies with order 4 |
| DRSA | - | Dominance-based Rough Sets Approach |
| DWT | - | Discrete Wavelet Transform |
| ECG | - | Electrocardiogram |
| EE | - | Energy Entropy |
| EEG | - | Electroencephalograms |
| EMG | - | Electromyograms |
| EOG | - | Electrooculograms |

| | | |
|---|---|---|
| EP | - | Evoked Potential |
| E-Prime | - | Experimenter’s Prime |
| ERP | - | Event Related Potential |
| ESD | - | Energy Spectrum Density |
| FAR | - | False Acceptance Rate |
| FFT | - | Fast Fourier Transform |
| FIFO | - | First-In-First-Out |
| FIS | - | Fuzzy Inference Systems |
| FN | - | False Negative |
| FP | - | False Positive |
| FPR | - | False Positive Rate |
| FRNN | - | Fuzzy-Rough Nearest Neighbour |
| FS | - | Feature Selection |
| FTMK | - | Faculty Information and Communication Technology |
| GUI | - | Graphical User Interface |
| IBk | - | Instance-based Learning with parameter k |
| IncFRNN | - | Incremental Fuzzy-Rough Nearest Neighbour |
| ISI | - | Inter-Stimulus Interval |
| ITR | - | Instance-based Template Reconstruction |
| KNN | - | K-Nearest Neighbour |
| K-S | - | Kolmogorov-Smirnov |
| LC | - | Linear Complexity |
| MATLAB | - | Matrix Laboratory |
| MLP | - | Multilayer Perceptron |

| | | |
|---|---|---|
| MREC | - | Medical Research and Ethics Committee |
| NN | - | Nearest Neighbour |
| PIN | - | Personal Identification Number |
| PLV | - | Phase Locking Value |
| Psychtoolbox | - | Psychophysics Toolbox |
| RAM | - | Random Access Memory |
| ROC | - | Receiver Operating Characteristic |
| SSVEP | - | Steady State Visual Evoked Potential |
| SVM | - | Support Vector Machine |
| $TAS^2$ | - | True Active Signal Shielding |
| TMSi | - | Twente Medical Systems International |
| t-norm | - | Triangular Norm |
| TNR | - | True Negative Rate |
| TPR | - | True Positive Rate |
| TTL | - | Transistor-Transistor Logic |
| UCI | - | University of California-Irvine |
| UTeM | - | Universiti Teknikal Malaysia Melaka |
| VEP | - | Visual Evoked Potential |
| VPRS | - | Variable Precision Rough Set |
| VQRS | - | Vaguely Quantified Rough Set |
| WD | - | Wavelet Decomposition |
| WEKA | - | Waikato Environment for Knowledge Analysis |
| WPD | - | Wavelet Packet Decomposition |

## LIST OF PUBLICATIONS

Liew, S. H., Choo, Y. H., Low, Y. F., Mohd Yusoh, Z. I., Yap, T. B., & Muda, A. K. (2015). Comparing Features Extraction Methods for Person Authentication Using EEG Signals. In *Pattern Analysis, Intelligent Security and the Internet of Things* (pp. 225-235). Springer International Publishing.

Liew, S.H., Choo, Y.H., Low, Y.F., & Mohd Yusoh, Z. I. (2015). Identifying Visual Evoked Potential (VEP) Electrodes Setting for Person Authentication. In *Int. J. Advance Soft Compu. Appl*, 7(3), 85-99.

# CHAPTER 1

# INTRODUCTION

## 1.0 Overview

This chapter portrays the briefing of the research. The description encompasses the background of study, problem statement, research questions, research objectives, research scope, research significance and the contribution of the research. At the end of this chapter, there is a brief discussion of the organization of the following chapters in order to give an overall picture of this thesis.

## 1.1 Project Background

An authentication or verification system involves accepting or rejecting the identity that claimed by a particular individual, which is one-to-one matching. In contrast, an identification system attempts to establish the identity of a given person out of a closed pool of N people, which is one-to-N matching (Marcel and Millán, 2007). There are several types of methods that can be used for person authentication such as knowledge-based, token-based and biometrics. Password and Personal Identification Number (PIN) are the examples of knowledge-based authentication method while signature is the example of token-based authentication method. Most of the people still prefer the use of signature and password as authentication methods because it is easier and do not require maintenance work. Unfortunately, password and signature are considered the weakest authentication methods because password can be stolen and guessed easily and signature can be forged easily. Therefore, biometric authentication systems are introduced to overcome traditional authentication methods. Biometric is any measurable feature(s), in terms of physical, physiological, behavioural trait or their combinations, that can be used to authenticate the claimed identity of an individual (Neela and Kahlon, 2012). Fingerprint authentication system is one of the biometric authentication systems, which is widely used in many applications. However, possibility of counterfeit fingerprint brings down the uniqueness of this modality. Forgery issue encourages alternative authentication modality using brainwaves through EEG signals for person authentication in the recent studies.

Electroencephalogram (EEG) signals are brain activities recorded from electrodes mounted on the scalp. EEG signals are the product of ionic current flows that happens in the brain's neurons. EEG is the most practical capturing method that can be used in biometrics due to the advances in its hardware devices. It is unique, confidential and cannot be duplicated. Besides that, EEG signals are biodynamic and also as a proof of aliveness for a particular individual. Thus, it cannot be duplicated like most of the other static physical biometric techniques.

EEG signals are usually nonlinear, non-stationary and hard to recreate because these may have been influenced by some sources of noise such as environmental noise and physiological noise. Therefore, the classification of EEG signals is not a trivial task. Various approaches like those classified with uncertainty methods are developed with fuzzy set theory, rough set theory and the combination of fuzzy and rough set theory. These approaches are proposed for EEG signals classification to deal with the vagueness and uncertainty in the signals (Hu and Knapp, 1991; Jahankhani et al., 2008; Lee et al., 2013; Liew et al., 2013). However, the current researches and applications mainly focus on static information systems. Many real-life applications are observed facing dynamically changing data patterns and the data volume growth in dimensions. Thus, incremental learning is becoming a key area of data mining research as the trend is towards dynamic data sources. Incremental learning does not require sufficient training set but it can update the training pool from time to time when the new test object arrives. It does not need to be retrained from scratch and hence the computational resources can be reduced (Geng and Smith-Miles, 2009; Charles Ditzler, 2011; Read et al., 2012). In addition, incremental learning can be learnt continuously for improvement when the system is running. It is adaptable to the changes of the target concept.

Fuzzy-Rough Nearest Neighbour (FRNN) is a hybrid technique combining the strengths of two natural computing design, i.e. the fuzzy sets and rough sets complement each other in defining the uncertainty knowledge granules. FRNN is an extension to the K-Nearest Neighbour (KNN) which employs fuzzy-rough set theory (Qu et al., 2013). Instead of using Euclidean distance, the FRNN calculates the nearest neighbours by using fuzzy similarity. The nearest neighbours are then used to construct the fuzzy lower and upper approximations to quantify the membership value of a test object to determine its decision class. FRNN performs promisingly in various domains (Hu et al., 2008; Hassanien et al., 2009; Boongoen and Shen, 2010; Parthaláin and Jensen, 2010; Maji, 2011) as it aims to mimic human decision making in solving real world problems.

KNN is a well-known classification technique. It is a type of instance-based learning, or lazy learning technique, where the technique builds no general model until a new test object arrives. KNN applies First-In-First-Out (FIFO) strategy to update the knowledge granules incrementally. The function is approximated locally and all computation is deferred until classification. They have a small training set so that it can update the training set and store the test object as a new training object in the information system. Thus, small training set makes the lazy algorithm particularly well suited for incremental learning tasks where a data stream updates continuously the set of input data (Aha, 1998).