Hybrid Feature in Typing Biometrics for User Authentication Using EEG Signals

by

Intan Amalina binti Abdul Rahim
(1931312918)

A thesis submitted in fulfillment of the requirements for the degree of Master of Science in Biomedical Electronic Engineering

Faculty of Electronic Engineering Technology
UNIVERSITI MALAYSIA PERLIS

2022

ACKNOWLEDGEMENT

First and foremost, praises and thanks to the God, the Almighty, for His showers of blessings throughout my research work and complete my thesis successfully.

I would like to express my deep and sincere gratitude to my main supervisor, Dr. Saidatul Ardeenawatie bt Awang for the continuous support in this research as well as for her patience, motivation and valuable knowledge. Her guidance helped me in all time of this research and writing of this thesis. My sincere gratitude also goes to my co-supervisor, Dr. Chong Yen Fook for his advice especially in technical work and thesis writing throughout this research. I also would like to thank the technicians of the Biomechanics Laboratory for their assistance throughout the completion of my research, Mr. Anuar b. Ahmad. My thanks also to all lecturers in Biomedical Electronic Engineering programme too, for the teachings and knowledge throughout my study in UniMAP.

Furthermore, I would like to convey my heartfelt appreciation to my mother, my late father, siblings, in-laws and especially my husband, Muhammad Syamiel bin Azman for their constant prayers, love and support along my master journey. Also, I would like to say thanks to my friends and research colleagues, for their encouragement and help throughout this research.

Lastly, I would like to express my profound gratefulness to Malaysia Government too that provide financial support under Fundamental Research Grant Scheme (FRGS). Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
BCI	Brain Computer Interface
BLSTM-NN	Bidirectional Long Short Term Memory Neural Network
BP	Back Propagation
CCA	Canonical Correlation Analysis
CNN	Convolutional Neural Network
CNS	Central Nervous System
CSP	Common Spatial Pattern
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EC	Eyes Close
ECG	Electrocardiogram
EEG	Electroencephalogram
EER	Equal Error Rate
EO	Eyes Open
EOG	Eletro-oculagram
ERP	Event-Related Potential
ErrP	Error Potential
ERR	Equivalent Error Rate
FAR	False Acceptance Rate
FIR	Finite Impulse Response
FLD	Fisher Linear Discriminant
FN	Functional Network
FRR	False Rejection Rate
GA	Genetic Algorithm
HDCA	Hierarchical Discriminant Component Analysis
HOG	Histogram Oriented Gradient
HTER	Half Total Error Rate
ICA	Independent Component Analysis
k-NN	*k*-Nearest Neighbour
LDA	Linear Discriminant Analysis

LSM	Least Square Method
LSTM	Long Short Term Memory
PIN	Personal Identification Number
PSD	Power Spectral Density
RBF	Radial Basis Function
RSVP	Rapid Serial Visual Presentation
SSVEP	Steady State Visually Evoked Potential
SVM	Support Vector Machine

LIST OF SYMBOLS

f_0	Power supply frequency
x	Column vector
p	Order of autoregressive (AR) model
Pxx	Power spectral density estimate
m	Embedding dimension
τ	Time delay
r	Tolerance
M	Number of concatenated signals
T	Transposition
x_i	Column vector
U	Mann-Whitney U test
n_1	Sample size one
n_2	Sample size two
R_i	Rank of the sample size
X_i	Training data
X_j	Testing data
d	Euclidean distance
Θk	Random vectors that independent and distributed

Pencampuran Ciri dalam Biometrik Menaip untuk Pengesahan Pengguna Menggunakan Isyarat EEG

ABSTRAK

Isyarat elektroencephalogram (EEG) adalah aktiviti rakaman elektrik otak yang disebabkan oleh pengaktifan sinaptik neuron otak yang direkodkan di sepanjang permukaan kulit kepala. Selama bertahun-tahun, terdapat pelbagai penyelidikan mengenai pelaksanaan isyarat EEG sebagai pengesahan biometrik. Ini disebabkan oleh potensi otak manusia yang unik dan daya tahan yang tinggi terhadap pemalsuan. Kajian ini menganalisis eksperimen pengesahan berdasarkan EEG semasa menjalankan tugas menaip yang dikenali dan tidak dikenali. Sebanyak 30 subjek (mahir dengan tangan kanan) dengan usia antara 19 hingga 23 tahun dipilih untuk melakukan tugas menaip dua kali (dikenali dan tidak dikenali) selama 3 minit dengan rehat di antaranya selama 1 minit. Subjek menaip nama depan dan belakang mereka diikuti dengan menaip nama rawak masing-masing selama 3 minit. Mereka diminta untuk berehat sebelum dan sesudah melakukan tugas menaip selama satu minit. Alat sukan EEGO™ (ANT Neuro, Enschede, Belanda) dengan persampelan frekuensi 512 Hz dan 32 saluran telah digunakan. Kajian ini menggunakan 'Independent Component Analysis' untuk menghilangkan kedipan mata, penapis takuk untuk menghilangkan gangguan talian kuasa 50 Hz, dan penapis jalur lebar untuk memisahkan isyarat ke dalam sub frekuensi seperti delta, theta, alfa, beta dan gamma. Ciri-ciri tersebut diekstraksi melalui pengekstrakan ciri linier (kaedah 'Welch', kaedah 'Burg', kaedah 'Yule-Walk') dan ciri tidak lelurus ('Fuzzy Entropy') juga diekstrak. Ciri-ciri yang diekstrak dikelaskan melalui pengklasifikasi tidak lelurus (pengelas 'k-Nearest Neighbour', 'Random Forest' dan 'Ensemble Bagged Tree') untuk mendapatkan prestasi data eksperimen. Ciri-ciri pengekstrakan yang memperoleh ketepatan prestasi tinggi kemudian disatukan berdasarkan jalur frekuensi dan bahagian otak melalui penggabungan cirian. Teknik pemilihan ciri seperti ujian-t statistik dan ujian-U Mann-Whitney juga digunakan pada ciri yang diekstrak dengan mengurangkan bilangan pembolehubah ciri masukan yang meningkatkan prestasi pengelas. Hasil klasifikasi ciri yang diekstrak dan disatukan oleh pengklasifikasi menunjukkan peningkatan peratusan prestasi tugas menaip antara dikenali dan tidak dikenali. Secara kesimpulannya, ketepatan peratusan tertinggi untuk biometrik menaip ialah dengan menggunakan kaedah Burg untuk penggabungan ciri bahagian depan dan parietal iaitu 95.94% dengan menggunakan pengelas 'Ensemble Bagged Tree'.

Hybrid Feature in Typing Biometrics for User Authentication Using EEG Signals

ABSTRACT

Electroencephalogram (EEG) signals are the electrical activities of brain recording caused by the synaptic activations of the brain's neurons which are recorded along the scalp surface. Throughout the years, there have been various researches on the implementation of EEG signals as biometric authentication. This is due to the potential of the human brain that is unique and high resistant to forgery. This research analysed the experiment on the EEG based authentication during the performed familiar and unfamiliar typing tasks. A total of 30 subjects (right-handed) with the age of between 19 to 23 years old were chosen to perform two times typing tasks (familiar and unfamiliar) for 3 minutes with rest in between for 1 minute. The subjects typed their first and last name followed by typing random names for 3 minutes each. They were required to rest before and after performing the typing tasks for one minute. The EEGO™ sports device (ANT Neuro, Enschede, The Netherlands) with frequency sampling of 512 Hz and 32 channels were used. This research applied Independent Component Analysis to remove eye blinks, notch filter to remove 50 Hz powerline artefacts, and bandpass filter to separate the signals into sub frequency bands such as delta, theta, alpha, beta and gamma. The features were extracted through linear feature extraction (Welch's method, Burg's method, Yule–Walk's method) and non-linear features (Fuzzy Entropy) were also extracted. The extracted features were classified through non-linear classifiers (*k*-Nearest Neighbour, Random Forest and Ensemble Bagged Tree classifiers) to obtain the performance of the experimental data. The extracted features that obtained high performance accuracy were then hybrid among them based on frequency bands and brain lobes through concatenation. The feature selection techniques such as statistical t-test and Mann-Whitney U test were also applied to the extracted features by reducing the number of input variables which improve the classifier performance. The classification results of the features extracted and hybrid by the classifiers showed an improvement of performance accuracy of familiar and unfamiliar typing tasks. In conclusion, the highest percentage accuracy for typing biometrics by using Burg's method for frontal and parietal lobes hybrid feature which is 95.94% by using Ensemble Bagged Tree classifier.

CHAPTER 1 : INTRODUCTION

1.1 Research Background

In today's world, security is a critical issue in all industries, including banks, government applications, military organizations and educational institutions. Government bodies are establishing guidelines, passing rules, and requiring organizations and agencies to adhere to these standards, with non-compliance carrying severe consequences. Therefore, the security mainly requires an authentication system. Authentication refers to a method of verifying an individual's identity (Nakamura et al. 2018).

The authentication mechanism in a security system compares the information given by the user to the information stored in the database. If the data corresponds to the database, then the user is given access to the security system based on the information provided (Lal et al. 2016). It is an important medium for individuals to control access to physical and digital resources including buildings, houses, rooms and computers (Jalaly et al. 2020).

The application of the unique biological human features has been developed in authentication which is known as biometrics. Physiological features and behavioural features are the two major types of individual biometric-based authentication (Alsaadi, 2015). Physiological characteristics refer to the characteristics of a human body that do not change with age. Face recognition, hand geometry, palm print, fingerprint, iris recognition, DNA, retina, and recognition are examples of physiological traits.

Contrarily, the behavioural access to biometrics, is restricted to human behavioural patterns that deal with personal actions, such as voice recognition, signature recognition, gait, and keystroke dynamics.

Existing technologies mostly use fingerprints, facial features, gait, and touch gestures as a base for recognition (Bashar et al. 2016) according to Figure 1.1. There are three main factors of user authentication namely knowledge-based user authentication, biometric-based user authentication and object-based user authentication (Pahuja & Nagabhushan, 2015).

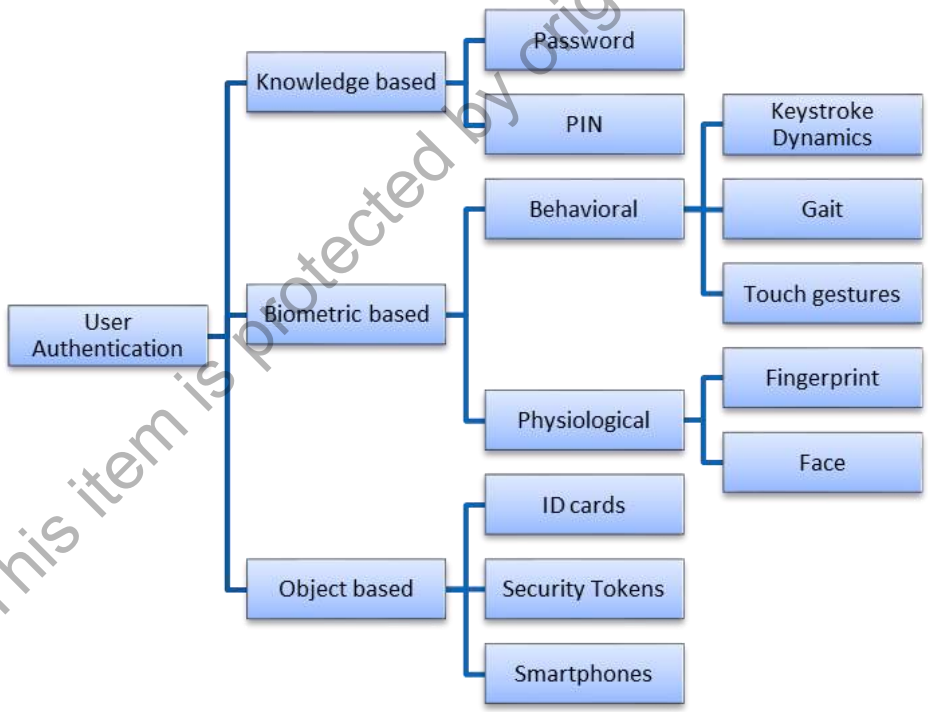

Figure 1.1 Factors of user authentication (Jayabalan, 2019)

The most common factor of user authentication is the knowledge-based user authentication such as password and Personal Identification Number (PIN) (Bultel et al. 2018) as shown in Figure 1.1. However, passwords are difficult to remember and frequently forgotten. Hence, some users tend to create short and easy to remember

passwords. Such weak passwords are vulnerable to dictionary, brute-force, and other authentication system attacks (Tsokkis & Stavrou, 2018). A hacker who is involved in unauthorised hacking activities by secretly accessing and manipulating the data through a network system is called a black hat hacker.

The most frequent attacks used by the hacker to crack password hashes are dictionary attacks and brute force attacks (Chester, 2015). Hackers prefer the dictionary attack because it produces fast and effective results while cracking hashes. It provides a very large dictionary file containing millions of potential passwords or dictionary word combinations in this attack. Each password in the dictionary file has its hash value determined and compared to the hash value of an unknown user password. When a match is successfully created, the plaintext in the dictionary that corresponds to the hash value, which is the same as the input password, is revealed (Jose et al. 2016).

On the other hand, a brute-force attack is a type of attack in which the attackers guess every possible password in a logical password space. Although the brute-force attack can eventually crack passwords, the size of the search space, time, and processing power required to crack strong passwords can make it impossible to be cracked (Zhang et al. 2013). Compared to dictionary attack, it is faster and yields high success rates based on the number of recorded attacks and common passwords (Brumen & Makari, 2017). Hence now, most of the attackers use this dictionary attack to crack a large number of passwords.

Another factor of user authentication is the object-based user authentication which consists of physical objects to authenticate to a system through items such as ID cards, security tokens, smartphones and other physical objects (Rusdan & Manurung, 2020).

Recently, the previous researchers also explored the potential of using electroencephalogram (EEG), electrocardiogram (ECG) and electromyogram (EMG) as new types of biometrics in user authentication. The most widely used signal is the EEG signal which is obtained from the brain surface (Jayarathne & Cohen, 2017).

Basically, the EEG signals acquire special attributes that make it more acceptable to be used in authentication. They are affected by mental responses and stress which alters the natural brain wave patterns in any action (Jalaly et al. 2020). Since the brain response is significantly altered depending on the current mental state, it is more robust and stable (Jayarathne & Cohen, 2016). As a result, EEG biometrics cannot be utilized in situations when the user is threatened or intimidated, but other authentication methods, such as threat response methods, allow an attacker to get access to equipment by employing force or threatening the user (Rui & Yan, 2019).

Previous researchers mostly developed bimodal user authentication system by fusing two single-modality recognition schemes such as EEG signals and voice signals (Moreno-rodriguez et al. 2021), EEG signals and gait signals (X. Zhang et al. 2020), EEG signals and eye tracking data (Krishna et al. 2019) and EEG signals and face modality (W. Rahman et al. 2017). There were few types of hybrid methods that were developed to combine the modalities for user authentication.

In this research, an EEG based authentication by using typing task is performed. Typing is considered as behavioural biometrics. By employing the electroencephalogram (EEG) device, the signals extracted from the brains were recorded during user typing. The typing task consists of two different parts; familiar and unfamiliar typing task. Given typing tasks were designated to check the familiarity of the users to their password and others' password.

The brain (EEG) signals collected were analysed to verify that the password should be something that the genuine user could easily type, but other users should not be equally fluent in typing it. This research requires the use of brain signals as the brain signals are very unique and cannot be copied. Furthermore, hybrid feature has only been successfully explored previously between EEG data signals and other recognition schemes through various types of hybrid thus, the concatenation hybrid of the features only by using EEG signals features was developed in this research.

1.2 Problem Statement

Every year, businesses and the government spend millions of dollars on the most up-to-date security and access control systems. As the data volumes increase, the access control also becomes more convenient. With the advancement of computer hardware and the development of the internet, breaking protection systems has become much simpler, requiring them to develop a more reliable framework to protect the data (Khalifa et al. 2013). Data protection is highly related to the protection of information's safety, credibility, and accessibility in all ways. Several techniques and procedures can be developed to maintain security management using knowledge-based, biometric-based and object-based user authentication. Even though the techniques have been successfully developed, they may also have their restrictions.

For instance, an attacker may monitor when the user withdraws money from an ATM, or by using a camera that records the PIN code, or even using a thermal camera to know the touch used to enter the PIN code (Bultel et al., 2018). A thermal camera captures the thermal traces left on the surface of a mobile device after authentication during a thermal attack. These details have been recovered and are being used to restructure the password. As a result, the order in which PINs are entered and their patterns will be exposed (Abdelrahman et al. 2017). Some users also choose passwords that are short, easy to remember, and can be found in a database (Taneski et al. 2016). In this condition, the attackers will be able to perform a range of hacking techniques against password protection, including dictionary attacks, brute-force attacks, and others (Shetty et al. 2018).

Furthermore, the current user authentication of biometric authentication such as fingerprint identification has its limitations in the aspect of its image quality. The acceptable fingerprint image is determined by some factors, for example, skin condition, sensor condition and poor user cooperation (Borra et al. 2016). Also in the research done by Patel et al. (2016) shows that the unimodal continuous authentication depending on the modalities such as touch gestures was influenced by poor lighting on the user's face that generates noisy data. In addition, gait modality may produce non-universality data due to incorrect leg patterns caused by unexpected leg injuries (Ennaama et al. 2019).

To sum up everything that has been mentioned, it is very important to have an EEG based authentication method by using typing tasks since the EEG signals from the brain can distinguish between a user and intruder in which the brain reads the pattern of individual typing. User's identity and information are authenticated through familiar and unfamiliar typing task that shows the brain response during typing. Since the EEG signals cannot be forged by intruders, it is considered a safe method to detect between a user and another.

1.3 Research Objectives

The purposes of this research are:

a) To design and develop a database of EEG signals based on typing tasks for biometric authentication system.

b) To extract the features of EEG signals (linear and non-linear features) to differentiate familiar and unfamiliar typing tasks.

c) To develop hybrid feature based on brain lobes using linear and non-linear features for biometric authentication.

The data acquisition protocol was developed by obtaining EEG signal from 30 subjects using the EEGO™ sports device (ANT Neuro, Enschede, The Netherlands) with frequency sampling of 512 Hz and 32 channels. They were required to perform two typing tasks (familiar and unfamiliar) using a laptop provided for three minutes each, with one minute of rest in between. The features were extracted through linear feature extraction (Welch's method, Burg's method, Yule–Walk's method) and non-linear feature (Fuzzy Entropy) was also extracted. The extracted features were classified through non-linear classifiers (k-Nearest Neighbour, Random Forest and Ensemble Bagged Tree) to obtain the performance of the familiar and unfamiliar typing tasks. Hybrid features were developed based on frequency bands and brain lobes through concatenation method by using the extracted features that obtained high performance accuracy.

1.4 Scope of Research

In order to achieve the objectives, this study is performed based on a few aspects. They are:

i) This research involved a total of 30 healthy and right-handed students from Universiti Malaysia Perlis (UniMAP) with ages ranging from 19 to 23 years old in the data acquisition protocol.

ii) The research was designed to investigate the brain wave pattern during typing tasks.

iii) The data acquisition protocol was carried out by using an EEG device (EEGO^{TM} sports 32 pro by ANT-Neuro, The Netherlands) with 32 electrodes with a sampling rate is 512 Hz. The collected EEG signals data from the EEGO^{TM} sports were analysed by using MATLAB R2015a.

iv) The features were extracted from the EEG typing signals using linear features (mean, median, standard deviation and variance) through Welch's method, Burg's method and Yule–Walk's method and also non-linear feature such as Fuzzy Entropy.

v) The experimental research was focusing on the hybrid feature of the EEG signals to achieve a better improvement of a performance rate. All the extracted features were hybrid through concatenation to improve the performance accuracy when

classified by using k-Nearest Neighbour, Random Forest and Ensemble Bagged Tree classifiers.

1.5 Thesis Organization

This thesis is structured according to five chapters.

Chapter 1: This chapter begins with the introduction of the implementation of EEG devices in user authentication. This chapter will also cover the problem statement, objectives, and scopes of this research.

Chapter 2: This chapter covers the background of the human brain and EEG device that related to user authentication as well as past researches on a similar topic of this research in terms of experimental protocol and signal processing. The chapter also presented the research gap to identify the direction of the potentially new idea of research based on previous studies of EEG based authentication.

Chapter 3: This chapter elaborates the methodology and approaches by using an EEG device during typing tasks. This chapter also covers the criteria of subject selection, equipment used, experimental protocol and data analysis such as signal processing techniques being used in this research. The signal processing techniques include pre-processing, feature extraction, hybrid feature, feature selection and classification of the EEG signals.