DISTRACTION DESCRIPTOR FOR BRAINPRINT AUTHENTICATION MODELLING USING PROBABILITY-BASED INCREMENTAL FUZZY-ROUGH NEAREST NEIGHBOUR TECHNIQUE

LIEW SIAW HONG

DOCTOR OF PHILOSOPHY

2021

Faculty of Information and Communication Technology

DISTRACTION DESCRIPTOR FOR BRAINPRINT AUTHENTICATION MODELLING USING PROBABILITY-BASED INCREMENTAL FUZZY-ROUGH NEAREST NEIGHBOUR TECHNIQUE

Liew Siaw Hong

Doctor of Philosophy

2021

DISTRACTION DESCRIPTOR FOR BRAINPRINT AUTHENTICATION MODELLING USING PROBABILITY-BASED INCREMENTAL FUZZY-ROUGH NEAREST NEIGHBOUR TECHNIQUE

LIEW SIAW HONG

A thesis submitted
in fulfillment of the requirements for the degree of Doctor of Philosophy

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

DECLARATION

I declare that this thesis entitled "Distraction Descriptor for Brainprint Authentication Modelling using Probability-based 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 the candidature of any other degree.

Signature : ……………………………………

Name : Liew Siaw Hong

Date : 29/09/2021

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature : ....

Supervisor Name : Associate Professor Ts. Dr. Choo Yun Huoy

Date : 29/09/2021

DEDICATION

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

To my dearest supervisors, Associate Professor Ts. Dr. Choo Yun Huoy, Associate Professor Dr. Low Yin Fen and Associate Professor Dr. Chong Shin Horng for being responsible, receptive, and always by my side to encourage and motivate me.

To my dear friend, especially Sam Weng Yik, for your support and motivation throughout this study.

ABSTRACT

The characteristics of uniqueness and proof of aliveness have driven the research in Brainprint as a biometric modality. Brainprint measuring by Electroencephalogram (EEG) suffers from low signal-to-noise ratio and are varied across time. Most of the brainprint authentication models were tested in well-controlled environments to minimize the influence of ambient disturbance on the EEG signals. These settings significantly contradict the real-world situations. Thus, making use of the distraction is wiser than eliminating it. This research aims to design a distraction descriptor, elicited through the object variation, to refine the granular knowledge incrementally, using the proposed probability-based update strategy in Incremental Fuzzy-Rough Nearest Neighbor (IncFRNN) technique. The research follows the experimental methodology, starting from data acquisition to data imputation, EEG distraction descriptor, probability-based IncFRNN and model analysis. The EEG of 45 volunteer human subjects were collected using visual stimuli in three levels of auditory ambient distraction, which are in quiet, low, and high distraction conditions. An artefact rejection with amplitude greater than 100 µV was applied for data cleaning. Occasionally, missing values occurred after removing the noisy trials. A similarity matching imputation method is proposed for EEG data imputation. The power spectral density, wavelet phase stability, and coherence were used as feature extraction methods. The probability-based IncFRNN technique was used to construct the learning model. The proposed probability-based incremental update strategy is benchmarked with the ground truth (actual class) incremental update strategy. Besides, the proposed technique is also benchmarked with First-In-First-Out (FIFO) incremental update strategy in K-Nearest Neighbour (KNN). The authentication accuracy, area under receiver operating characteristic curve, recall, precision, and the F-measure were used to evaluate the proposed technique. The experimental results have shown equivalence discriminatory performance in both high distraction and quiet conditions. This has proven that the proposed distraction descriptor is able to utilize the unique EEG response towards ambient distraction to complement person authentication modelling in the uncontrolled environment. However, the authentication results in low distraction condition are significantly worse than both the quiet and high distraction conditions. This might because the distraction is too mild to elicit the cognitive measures representing individual characteristics. The probability-based IncFRNN technique has significantly outperformed the KNN technique for both with and without defining the window size threshold. Nevertheless, its performance is slightly worse than the actual class incremental update strategy since the ground truth represents the gold standard. In overall, this study demonstrated a more practical brainprint authentication model with the proposed distraction descriptor and the probability-based incremental update strategy. However, the data acquisition was carried out in a single session. The EEG distraction descriptor may vary due to intersession variability. Future research should focus on the intersession variability to improve the robustness of the brainprint authentication model.

PEMERIHAL GANGGUAN UNTUK PEMODELAN PENGESAHAN CETAKAN OTAK MENGGUNAKAN TEKNIK PENAMBAHAN JIRAN TERDEKAT KABUR-KASAR BERASASKAN KEBARANGKALIAN

ABSTRAK

Ciri-ciri keunikan dan bukti keaktifan Elektroensefalogram (EEG) telah mendorong kajian cetakan otak sebagai modaliti biometrik. EEG mengalami nisbah isyarat-hinggar yang rendah dan berubah sepanjang masa apabila ia digunakan untuk mengukur cetakan otak. Kebanyakan pemodelan pengesahan cetakan otak telah diuji dalam persekitaran yang terkawal untuk mengurangkan gangguan ambien pada isyarat EEG. Penetapan ini adalah bertentangan dengan alam nyata. Oleh itu, mempergunakan tindak balas otak ke atas gangguan persekitaran adalah lebih baik daripada membuat penapisan isyarat hinggar. Kajian ini bertujuan untuk mereka bentuk pemerihal gangguan bagi membolehkan pengkemaskinian granul pengetahuan yang tertimbul melalui variasi objek, dengan menggunakan strategi penambahan berasaskan kebarangkalian dalam teknik Penambahan Jiran Terdekat Kabur-Kasar (IncFRNN) yang dicadangkan. Kajian ini mengikuti metodologi ujikaji, bermula dari pemerolehan data, imputasi data, pemerihal gangguan EEG, IncFRNN berasaskan kebarangkalian dan analisis modal. Signal EEG daripada 45 subjek manusia sukarela telah dikumpulkan melalui rangsangan visual yang berada di tiga tahap gangguan auditori, iaitu sunyi, rendah, dan tinggi. Penolakan artifak yang mempunyai amplitud lebih tinggi daripda 100 µV digunakan untuk pembersihan data. Kadang-kala, kehilangan nilai berlaku selepas penghapusan sampel yang bising. Kaedah imputasi pemadanan keserupaan dicadangkan untuk mengimputasikan data EEG. "Power Spectral Density, Wavelet Phase Stability" dan "coherence" digunakan sebagai kaedah pengekstrakan. Teknik IncFRNN berasaskan Kebarangkalian digunakan untuk membina modal pembelajaran. Perbandingan telah dilakukan antara strategi penambahan berasaskan kebarangkalian dengan strategi asas (kelas sebenar). Selain itu, perbandingan juga dilakukan antara teknik yang dicadangkan dan strategi "First-In-First-Out (FIFO)" dalam Jiran Terdekat k (KNN). Ketepatan pengesahan, luas kawasan di bawah lengkung "receiver operating characteristic", "recall", kepersisan, dan sukatan F digunakan untuk menilai prestasi teknik yang dicadangkan. Hasil eksperimen telah menunjukkan prestasi persamaan diskriminasi dalam keadaan gangguan yang tinggi dan sunyi. Ini telah membuktikan bahawa pemerihal gangguan yang dicadangkan dapat menggunakan tindak balas EEG terhadap gangguan ambien untuk melengkapi pemodelan pengesahan dalam persekitaran yang tidak terkawal. Hasil penilaian pengesahan dalam keadaan gangguan yang rendah berprestasi lebih teruk daripada keadaan sunyi dan gangguan yang tinggi. Kemungkinan ini adalah kerana gangguan yang kurang tidak dapat mencungkil ukuran kognitif untuk menunjukkan ciri-ciri individu. Teknik IncFRNN berasaskan kebarangkalian berprestasi lebih baik daripada teknik KNN dengan dan tanpa menentukan nilai ambang saiz tetingkap. Namun begitu, penilaian ini agak teruk daripada strategi penambahan kelas sebenar kerana strategi asas mewakili standard emas. Keseluruhannya, kajian ini telah mendemotrasi modal pengesahan cetakan otak yang lebih praktikal dengan menggunakan pemerihal gangguan dan strategi penambahan berasaskan kebarangkalian. Namun begitu, pemerolehan data dilakukan dalam satu sesi. Pemerihal gangguan mungkin berbeza dalam sesi yang berlainan. Penambahbaikan kajian boleh fokus dalam pembolehubahan sesi yang berlainan untuk meningkatkan keteguhan modal pengesahan cetakan otak.

ACKNOWLEDGEMENTS

First and foremost, I would like to take this opportunity to express my sincere acknowledgement to my main supervisor, Associate Professor Ts. 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 UTeM Zamalah Scheme for providing the scholarship. Besides, this work also funded by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme, FRGS/1/2017/ICT/05/FTMK-CACT/F00346.

I would also like to express my greatest gratitude to Associate Professor 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 Associate Professor Dr. Chong Shin Horng from the Faculty of Electrical Engineering, 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 the realization of this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AI	-	Artificial Intelligence
ANN	-	Artificial Neural Network
AR	-	Autoregressive
AUC	-	Area Under ROC Curve
BCIs	-	Brain Computer Interfaces
CFS	-	Correlation-based Feature Selection
CIELM	-	Class Incremental Extreme Learning Machine
CIRF	-	Class Incremental Random Forests
DNA	-	DeoxyriboNucleic Acid
Dq-DTRS	-	Double-quantitative Decision-Theoretic Rough Sets
DRSA	-	Dominance-based Rough Set Approach
DWT	-	Discrete Wavelet Transform
EEG	-	Electroencephalogram
EER	-	Equal Error Rate
ERT	-	Extremely Randomized Trees
FAR	-	False Acceptance Rate
FFT	-	Fast Fourier Transform
FIFO	-	First-In-First-Out
FIR	-	Finite-duration Impulse Response

fMRI	-	functional Magnetic Resonance Imaging
FN	-	False Negative
fNIRS	-	functional Near Infrared Spectroscopy
FP	-	False Positive
HTER	-	Half Total Error Rate
IBk	-	Instance-based Learning with parameter *k*
IDRA	-	Incremental Discernibility Relation Algorithm
iEMPCA	-	incremental Expectation Maximization Principal Component Analysis
iKSSVM	-	Incremental Kernel Subclass SVM
IncFRNN	-	Incremental Fuzzy-Rough Nearest Neighbour
IRLSC	-	Incremental RLSC with class recording
ISI	-	Inter-Stimulus Interval
ISVM	-	Incremental Support Vector Machine
ITR	-	Instance-based Template Reconstruction
KNN	-	K-Nearest Neighbour
LDA	-	Linear Discriminant Analysis
MAP	-	Maximum A Posteriori
MEG	-	Magnetoencephalography
MLP	-	Multilayer Perceptron
PASAT	-	Paced Auditory Serial Addition Task
PCA	-	Principal Component Analysis
PLD	-	P-generalized decision Lower Domain
PSD	-	Power Spectral Density

PUD	-	P-generalized decision Upper Domain
ROC	-	Receiver Operating Characteristic
SVM	-	Support Vector Machine
TMSi	-	Twente Medical Systems International
TN	-	True Negative
t-norm	-	triangular norm
TP	-	True Positive
TPR	-	True Positive Rate
VEP	-	Visual Evoked Potential
VPRS	-	Variable Precision Rough Set
WEKA	-	Waikato Environment for Knowledge Analysis
WPD	-	Wavelet Packet Decomposition
WPS	-	Wavelet Phase Stability

LIST OF SYMBOLS

α	-	Alpha
β	-	Beta
δ	-	Delta
θ	-	Theta
γ	-	Gamma
μ	-	Micro

LIST OF PUBLICATIONS

Journals

1. Liew, S.H., Choo, Y.H., Low, Y.F., and Chong, S.H., 2020. Investigation of Alpha and Beta Band for Brainprint Authentication with Auditory Distractor. *International Journal of Computer Information Systems and Industrial Management Applications*, 12, pp. 14-22.
2. Liew, S.H., Choo, Y.H., and Low, Y.F., 2018. Missing Values Imputation using Similarity Matching Method for Brainprint Authentication. *International Journal of Advanced Computer Science and Applications*, 9(10), pp. 364-370.
3. Liew, S.H., Choo, Y.H., Low, Y.F., and Mohd Yusoh, Z.I., 2018. Selection of Performance Measures for Brainprint Authentication. *International Journal of Computer Information Systems and Industrial Management Applications,* 10, pp. 164-173.

Conference paper

1. Liew, S.H., Choo, Y.H., and Low, Y.F., 2019. Data Imputation in EEG Signals for Brainprint Identification. *Frontier Computing. FC 2018. Lecture Notes in Electrical Engineering*, pp. 278-286. Springer International Publishing.

CHAPTER 1

INTRODUCTION

1.1 Overview

This chapter portrays the briefing of the research. The description encompasses the background study, problem statement, research question, research objective, research scope, research significance, and expected output. At the end of this chapter, a brief discussion of the thesis organization is provided to give an overview picture of this thesis.

1.2 Background study

Brain signals are being investigated within the medical field for more than a century to study on brain diseases such as Alzheimer, schizophrenia, spinal cord injuries, epilepsy, and stroke among the others. Furthermore, they are also applied in assistive, rehabilitative, and entertainment applications as the basis for brain computer interface and brain machine interface. Despite widespread interest in clinical applications, the use of brain signals such as Electroencephalogram (EEG) is used as a biometric modality for person authentication or person identification (Poulos et al., 1999).

Biometric authentication is a security process of verifying an individual identity with the unique biological characteristics to grant accessibility permission. Common biometric modalities in real-world practice are the fingerprint, iris, and facial recognition. However, these modalities pose different drawbacks in practical implementation, crucially because they appear on the body surface with no obligatory liveness evidence. Impostor is able to forge access using fake fingers, printed iris images, or printed facial images since these