DETECTION OF AUTISM SPECTRUM DISORDER BASED ON TIME DOMAIN ELECTROENCEPHALOGRAM SIGNAL USING ENHANCED BILSTM MODELS

NUR ALISA BINTI ALI

DOCTOR OF PHILOSOPHY

2024

Faculty of Electronics and Computer Technology and Engineering

DETECTION OF AUTISM SPECTRUM DISORDER BASED ON TIME DOMAIN ELECTROENCEPHALOGRAM SIGNAL USING ENHANCED BILSTM MODELS

NUR ALISA BINTI ALI

Doctor of Philosophy

2024

DETECTION OF AUTISM SPECTRUM DISORDER BASED ON TIME DOMAIN ELECTROENCEPHALOGRAM SIGNAL USING ENHANCED BILSTM MODELS

NUR ALISA BINTI ALI

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

Faculty of Electronics and Computer Technology and Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2024

DEDICATION

Dedicated to my beloved parents, mother in-law, husband and children.

ABSTRACT

The global prevalence of Autism Spectrum Disorder (ASD) has driven researchers to develop well-defined automated approaches for early detection, surpassing standard behavioral assessments. Behavioral assessment methods face challenges to detect ASD at early stage because pronounced symptoms of autism are often observed between the ages of two and three years old, leading to delayed or missed diagnoses in some individuals. The electroencephalogram (EEG) has emerged as a promising quantifiable tool for identifying ASD biomarkers earlier than standard behavioral assessments. Its integration with deep learning methodologies has advanced ASD diagnosis through computer-aided diagnosis (CAD) systems. This research intends to classify the time-series EEG data of ASD and typical development (TD) samples from the SFARI dataset, which comprises 53 subjects (14 TD and 39 ASD) ranging in age from 10 months to 21 years. Two deep learning methods are particularly suitable for handling time-series data, namely Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) family. While most researchers utilize CNN-based approaches that require conversion from the time domain to the time-frequency domain, this study explores the potential of a Long Short-Term Memory (LSTM)-based models from RNN family to classify EEG data directly in the time domain. Specifically, this research examines the efficacy of LSTM and Bidirectional Long Short-Term Memory (BiLSTM) networks in distinguishing between ASD and TD individuals without relying on prior demographic knowledge or the requirement of data conversion. The optimized BiLSTM model achieved an accuracy of 99.68%, outperforming the LSTM in classifying ASD and TD using 117 multichannel EEG recordings. However, managing multichannel EEG data presents challenges, particularly with unpredictable ASD individuals. To address this, a hybrid model incorporating Autoregressive (AR) feature extraction, General Learning Equilibrium Optimizer (GLEO) feature selection, and optimized BiLSTM was developed to perform channel selection. This hybrid method achieved 99.89% accuracy using only 29 EEG channels, thereby reducing the complexity of the experimental setup by 75%. The discriminative ability of each channel in distinguishing between ASD and TD EEG data was supported through a one-way Analysis of Variance (ANOVA) method. This analysis revealed that 27 channels produced significantly different outputs, while the remaining 2 channels yielded p-values slightly higher than 0.05. These findings underscore the reliability of the proposed AR-GLEO-BiLSTM method for diagnosing ASD and lay a foundation for detecting ASD biomarkers in individuals before behavioral diagnosis is typically possible, or when behavioral features are not apparent until two years of age or later.

PENGESANAN GANGGUAN SPEKTRUM AUTISME BERDASARKAN ISYARAT ELECTROENCEPHALOGRAM DOMAIN MASA MENGGUNAKAN MODEL BiLSTM DIPERTINGKATKAN

ABSTRAK

Kecelaruan Spektrum Autism (ASD) telah mendorong penyelidik untuk membangunkan pendekatan automatik yang terperinci untuk pengesanan awal, melebihi penilaian tingkah laku standard. Kaedah penilaian tingkah laku menghadapi cabaran untuk mengesan ASD pada peringkat awal kerana gejala autisme yang ketara sering diperhatikan antara umur dua dan tiga tahun, yang menyebabkan diagnosis tertangguh atau terlepas bagi sesetengah individu. Elektroensefalogram (EEG) telah muncul sebagai alat kuantitatif yang menjanjikan untuk mengenal pasti biomarker ASD lebih awal daripada penilaian tingkah laku standard. Integrasinya dengan metodologi pembelajaran mendalam telah memajukan diagnosis ASD melalui sistem diagnosis berbantu komputer (CAD). Penyelidikan ini adalah untuk mengklasifikasikan data EEG bersiri masa bagi sampel ASD dan perkembangan tipikal (TD) dari set data SFARI, yang terdiri daripada 53 subjek (14 TD dan 39 ASD) yang berumur antara 10 bulan hingga 21 tahun. Dua kaedah pembelajaran mendalam yang sesuai untuk menangani data bersiri masa adalah Rangkaian Neural Konvolusi (CNN) dan keluarga Rangkaian Neural Berulang (RNN). Walaupun kebanyakan penyelidik menggunakan pendekatan berasaskan CNN yang memerlukan penukaran dari domain masa ke domain masa-frekuensi, kajian ini meneroka potensi model berasaskan Memori Jangka Panjang Pendek (LSTM) untuk mengklasifikasikan data EEG secara langsung dalam domain masa. Secara khusus, penyelidikan ini meneliti keberkesanan rangkaian Memori Jangka Panjang Pendek (LSTM) dan Memori Jangka Panjang Pendek Dua Arah (BiLSTM) dalam membezakan antara individu ASD dan TD tanpa bergantung pada pengetahuan demografi sebelumnya atau penukaran data. Model BiLSTM yang dioptimumkan mencapai ketepatan 99.68%, mengatasi prestasi LSTM dalam mengklasifikasikan ASD dan TD menggunakan 117 rakaman EEG berbilang saluran. Walau bagaimanapun, pengurusan data EEG berbilang saluran menghadirkan cabaran, terutamanya dengan individu ASD yang tidak dapat diramal. Untuk menangani perkara ini, model hibrid yang menggabungkan pengekstrakan ciri Autoregresif (AR), pemilihan ciri Pengoptimum Keseimbangan Pembelajaran Umum (GLEO) dan BiLSTM yang dioptimumkan telah dibangunkan untuk melaksanakan pemilihan saluran. Kaedah hibrid ini mencapai ketepatan 99.89% menggunakan hanya 29 saluran EEG, sekaligus mengurangkan kerumitan susunan eksperimen sebanyak 75%. Keupayaan diskriminatif setiap saluran dalam membezakan antara data ASD dan TD EEG telah disahkan melalui kaedah Analisis Varians sehala (ANOVA). Analisis ini mendedahkan bahawa 27 saluran menghasilkan output yang berbeza dengan ketara, manakala baki 2 saluran menghasilkan nilai p lebih tinggi sedikit daripada 0.05. Penemuan ini menggariskan kebolehpercayaan kaedah AR-GLEO-BiLSTM yang dicadangkan untuk mendiagnosis ASD dan meletakkan asas untuk mengesan biomarker ASD dalam individu sebelum diagnosis tingkah laku biasanya boleh dilakukan, atau apabila ciri-ciri tingkah laku tidak jelas sehingga umur dua tahun atau lebih.

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful

Alhamdulillah, all praises and thanks to Allah SWT for His greatness and giving me the strength, knowledge, ability, opportunity and courage to complete this thesis. I would like to take this opportunity to extend my gratitude to number of individuals who have contributed either directly or indirectly to the successful completion of this thesis.

First and foremost, I would like to express my sincere appreciation to my supervisor, Associate Prof. Dr. Syafeeza Ahmad Radzi, for her valuable critics, guidance and advice during my PhD journey. Also, Associate Prof. Dr. Norazlin, a consultant in developmental paediatric of Child Development Centre, HUKM for sharing lots of ideas, opinions and her encouragement in carrying out this research in autism research area.

Most importantly, I would like to thank you my family, especially my loving parents, Haji Ali Abd Ghani and Hajjah Hamidah Omar, my dear husband, Ts. Dr. Abd Shukur and my precious children, Muhammad Eirfan, Aisyah Sofiyyah, Athiya Syahirah, and Amna Solehah. My deepest gratitude for their continuous ‘doa’, love, motivation, patience, boundless support and encouragement in all my endeavours.

My greatest pleasure to my research colleagues and friends, Dr Haslinah, Siti Huzaimah, Dr. Zarina, Dr. Azdiana, Dr. Lizawati, Dr. Norhidayah, Dr. Afifah, Dr. Noor Asyikin, Dr. Norhashimah, Dr. Norihan Najmiah Radiah, Siti Aisah, and those who directly and indirectly involved for their kind help and supporting me in completing this research study. I am very fortunate to have these people during my ups and downs throughout this journey.

Last but not least, I would like to thank and acknowledge Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Education (MOE) for providing research facilities, financial support, and the opportunity to pursue my studies up to this level. Their support is gratefully acknowledged.

TABLE OF CONTENTS

PAGES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D - One - Dimensional

2D - Two - Dimensional

ADI-R - Autism Diagnostic Interview- Revised

ADOS - Autism Diagnosis Observation Schedule

ANN - Artificial Neural Network

ANOVA - Analysis of Variance

ANN - Artificial Neural Network

ASD - Autism Spectrum Disorder

ADHD - Attention Deficit Hyperactive Disorder

BCI - Brain Computer Interface

BiLSTM - Bidirectional LSTM

CAD - Computer-Aided Design

CARS - Childhood Autism Rating Scale

CNN - Convolutional Neural Network

CV - Cross Validation

CWT - Continuous Wavelet Transform

DL - Deep Learning

DSM 5 - Diagnostic and Statistical Manual of Manual Mental Disorder 5

DWT - Discrete Wavelet Transform

EEG - Electroencephalogram

EMG - Electromyogram

ERP	-	Event-Related Potential
FSA	-	Feature Selection Algorithm
fNIRS	-	functional Near-Infrared Spectroscopy
FT	-	Fourier Transform
GPU	-	Graphical Processing Unit
GRU	-	Gated Recurrent Unit
KNN	-	K-Nearest Neighbour
LOOCV	-	Leave-One-Out Cross Validation
LSTM	-	Long Short Term Memory
MEG	-	Magnetoencephalography
ML	-	Machine Learning
MRI	-	Magnetic Resonance Imaging
ODU	-	Old Dominion University
PCA	-	Principle Component Analysis
PET	-	Positron Emission Tomography
ReLU	-	Rectified Linear Unit
ReSNet	-	Residual Neural Network
RNN	-	Recurrent Neural Network
SNR	-	Signal-to-Noise Ratio
STFT	-	Short-Time Fourier Transform
SVM	-	Support Vector Machine
TD	-	Typical Development
VEP	-	Visual Evoked Potential

LIST OF SYMBOLS

s	-	seconds
σ	-	variance
Σ	-	summation
λ	-	lamda

LIST OF APPENDICES

LIST OF PUBLICATIONS

The followings are the list of publications related to the work on this thesis:

Journal Articles

N. A Ali, A.R Syafeeza, A. S Jaafar, M.K Mohd Fitri Alif, 2020. Autism Spectrum Disorder Classification on Electroencephalogram Signal using Deep Learning Algorithm. *International Journal of Artificial Intelligence (IJ-AI)* 9(1), pp. 91-99. ISSN: 2252-8938 (Scopus).

N. A. Ali, A. R Syafeeza, A. S. Jaafar, S. Shamsuddin, Norazlin Kamal Nor, 2021. LSTM-Based Electroencephalogram Classification on Autism Spectrum Disorder. *International Journal of Integrated Engineering (IJIE),* 13(6), pp. 321-329. ISSN: 2229-838X (Scopus).

N. A. Ali, A. R Syafeeza, A. S. Jaafar, Norazlin Kamal Nor, 2022. The ConVnet BiLSTM for ASD Classification on EEG Brain Signal. *International Journal of Online and Biomedical Engineering (iJOE).*

CHAPTER 1

INTRODUCTION

This chapter provides an introductory background of the neurodevelopmental disorder of Autism Spectrum Disorder (ASD), the quantifiable method using an electroencephalogram (EEG) data and the deep learning model for the classification. It also explains the research objectives, motivations of the presented research, and highlighting the research gap.

1.1 Research Background

The number of children diagnosed with Autism Spectrum Disorder (ASD) is prevalent worldwide. The studies conducted in Asia, Europe, and North America have identified individuals with ASD at an average prevalence rate changing between 1% and 2% (Talantseva et al., 2023). Recent statistics indicate that approximately 300,000 individuals with ASD reside in Malaysia, yet only 20,000 are registered with the ministry and issued the People with Disabilities (PWDS) card, as per 2018 data (Mohd Salleh et al., 2018). However, the exact number of individuals with autism for the year 2024 is unknown.

1.1.1 Autism Spectrum Disorder (ASD) and Standard Diagnostic Tool

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder marked by difficulties in social communication, restricted patterns of interest, and repetitive behaviors, with a broad spectrum of symptoms and varying degrees of severity (Hodges, Fealko and Soares, 2020). The current standard screening and diagnostic approach for ASD are wellknown as behavioral assessment methods. The widely utilized diagnostic tools include the Diagnostic and Statistical Manual of Mental Disorder-Fifth Edition (DSM-5), Autism Diagnostic Observation Schedule (ADOS), and Childhood Autism Rating Scale (CARS)(Randall et al., 2018a; Gulati et al., 2019). This type of assessment method imposes various limitations especially the long-winded process (Falkmer et al., 2013). It necessitates several, time-consuming patient visits and observations. The manual screening and diagnosis are susceptible to human errors, and they can be tedious and time-consuming due to changes in behavior during the assessment. According to (Catherine and Somer, 2015), the heterogeneity among ASD people and symptoms of ASD can change periodically. In fact, individuals with high-functioning ASD (mild severity) tend to exhibit symptoms at later developmental stages, which can impact the observations made by pediatricians and subsequently affect the accuracy of the diagnostic outcome (Ari et al., 2022). The challenges encountered for behavioral assessment method is due to pronounced symptoms of autism are often observed between the ages of two and three, while the average ASD diagnosis age is above 3 years as stated by (Van 't Hof et al., 2021). In some cases, it is possible to detect signs of autism as early as 6–12 months (Filipek et al., 1999). Nevertheless, the majority of specialized professionals involved in diagnosing the disorder refrain from providing a definitive diagnosis until the child reaches the age of 2 or 3 years (Baird, Cass and Slonims, 2003). The limitations also have caused some individuals go undiagnosed or the diagnosis is delayed (Pham et al., 2020).

Researchers studying autism encountered significant challenges as they strive to identify children at risk for this disorder as early as possible (Levin et al., 2017; Ramachandram, 2019; Tran, 2019; Haputhanthri et al., 2020). Early detection is crucial, as