MALAY LANGUAGE VOWEL CLASSIFICATION USING AUDIO IMAGE PROFILE VIA DEEP LEARNING FOR SPEECH DISORDER REHABILITATION ASSESSMENT

NUR SYAHMINA BINTI AHMAD AZHAR

MASTER OF SCIENCE IN ELECTRONIC ENGINEERING

2024

Faculty of Electronics and Computer Technology and Engineering

MALAY LANGUAGE VOWEL CLASSIFICATION USING AUDIO IMAGE PROFILE VIA DEEP LEARNING FOR SPEECH DISORDER REHABILITATION ASSESSMENT

Nur Syahmina Binti Ahmad Azhar

Master of Science in Electronic Engineering

2024

MALAY LANGUAGE VOWEL CLASSIFICATION USING AUDIO IMAGE PROFILE VIA DEEP LEARNING FOR SPEECH DISORDER REHABILITATION ASSESSMENT

Nur Syahmina Binti Ahmad Azhar

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

Faculty of Electronics and Computer Technology and Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2024

DEDICATION

To my beloved ibu and ayah.

ABSTRACT

Communication impairments can result from various medical conditions, such as speech problems, hearing loss, brain injuries, strokes, and physical disabilities. These conditions can affect verbal and non-verbal communication and may require specific rehabilitation and therapy. Currently, speech rehabilitation and treatment are time-consuming and involve physical activity, with most facilities still manually performing the process. However, technological advancements, such as Artificial Intelligence (AI), have opened up innovative solutions in speech rehabilitation. AI studies have focused on speech classification for various human languages, with the potential to revolutionize speech rehabilitation and make it more accessible to individuals worldwide. Since computer vision has impacted this field, machine learning and deep learning have been applied to the medical and healthcare industries to enhance rehabilitation by utilizing the new technology. Convolutional Neural Network (CNN) network models have been proven in countless studies to be precise at classifying performance in object and speech classification. This research analyzed the performance accuracy of different deep learning comparative network models, proposed network models, VGG-Net, AlexNet, and Inception, and performed a complete comparative analysis to assess these network models' classification accuracy and suitability for rehabilitation purposes. This thesis aims to develop a reliable vowel classification system with high-performance accuracy that can successfully recognize the classification of vowels in the normal person group, the post-stroke patient group with speech disorders, and the combination of both groups using the two proposed image profiles: the Mel spectrogram and the Mel Frequency Cepstral Coefficients (MFCC). According to the experimental results, the proposed network network model, which used six batch sizes, 20 epochs, and ADAM as the optimizer, managed to outperform the performance accuracy of the other existing comparative network network models. The highest performance accuracy gained for the Mel spectrogram, and MFCC image profile in the analyses conducted was 96.30% and 98.77%, respectively.

KLASIFIKASI VOKAL BAHASA MELAYU MENGGUNAKAN PROFIL IMEJ AUDIO MELALUI PEMBELAJARAN MENDALAM BAGI PENILAIAN PEMULIHAN GANGGUAN PERTUTURAN

ABSTRAK

Kecacatan komunikasi boleh disebabkan oleh pelbagai keadaan di dalam bidang perubatan, seperti masalah pertuturan, kehilangan pendengaran, kecederaan otak, strok dan kecacatan fizikal. Keadaan ini boleh menjejaskan komunikasi lisan dan bukan lisan dan mungkin memerlukan pemulihan dan terapi khusus. Pada masa ini, pemulihan dan rawatan pertuturan memakan masa dan melibatkan aktiviti fizikal, dengan kebanyakan kemudahan masih melakukan proses secara manual. Walau bagaimanapun, kemajuan teknologi, seperti Kepintaran Buatan (AI), telah membuka penyelesaian inovatif dalam pemulihan pertuturan. Kajian AI telah menumpukan pada pengecaman pertuturan untuk pelbagai bahasa manusia, dengan potensi untuk merevolusikan pemulihan pertuturan dan menjadikannya lebih mudah diakses oleh individu di seluruh dunia. Memandangkan penglihatan komputer telah memberi kesan kepada bidang ini, pembelajaran mesin dan pembelajaran mendalam telah digunakan pada industri perubatan dan kesihatan untuk meningkatkan pemulihan dengan menggunakan teknologi baharu. Model Rangkaian Neural Convolutional (CNN) telah terbukti dalam banyak kajian tentang ketepatannya dalam mengklasifikasi prestasi dalam pengecaman objek dan pertuturan. Penyelidikan ini menganalisis ketepatan prestasi cadangan model rangkaian dengan model-model pra-latihan pembelajaran mendalam yang berbeza model yang direka bentuk, VGG-Net, AlexNet dan Inception, serta melakukan analisis perbandingan lengkap untuk menilai ketepatan klasifikasi dan kesesuaian model ini untuk tujuan pemulihan. Projek ini bertujuan untuk membangunkan sistem yang boleh diandalkan dengan ketepatan prestasi tinggi, dan ia berjaya mengenali klasifikasi vokal terhadap kumpulan orang normal, kumpulan pesakit selepas diserang strok dengan gangguan pertuturan, dan gabungan kedua-dua kumpulan ini menggunakan dua profil imej, Spektrogram Mel dan 'Mel Frequency Cepstral Coefficients' (MFCC). Mengikut keputusan percubaan, rangkaian model yang direka, yang menggunakan saiz kelompok enam, 20 kitaran, dan ADAM sebagai pengoptimum, berjaya mengatasi ketepatan prestasi rangkaian model-model pra-latihan sedia ada yang lain. Ketepatan prestasi tertinggi yang diperoleh untuk spektrogram Mel, dan profil imej MFCC dalam analisis yang dijalankan masing-masing ialah 96.30% dan 98.77%.

ACKNOWLEDGEMENT

In the Name of Allah, the Most Gracious, the Most Merciful. First and foremost, I would like to take this opportunity to express my sincere gratitude, acknowledgment and admiration to my supervisor, Ir. Ts. Dr. Nik Mohd Zarifie bin Hashim for his invaluable guidance and support throughout my research journey. His expertise and dedication have shaped my understanding of this research matter and pushed me to achieve my best. Additionally, his constructive feedback and encouragement have contributed significantly to the quality of my work. His attentive guidance and approachable nature have made it easy for me to seek clarification and discuss any challenges I encountered during the research process. I am genuinely grateful for his mentorship, as it has enhanced my research skills and instilled in me a sense of confidence and resilience. Without his unwavering support, this research would not have been possible.

I also want to dedicate my gratitude to my co-supervisor, Professor Madya Dr. Masrullizam bin Mat Ibrahim, for sharing his knowledge, expertise, and experiences. His support throughout my research journey has been invaluable. I am truly grateful for the opportunity to learn from a knowledgeable and dedicated co-supervisor. I would like to express my appreciation for my lecturers at UTeM, who have significantly shaped my academic journey. Their passion for teaching and dedication to their students has inspired me to strive for excellence. I am thankful for their guidance and the valuable lessons they have imparted, contributing greatly to my growth as a researcher. I sincerely thank my fellow friends who provided valuable insights and feedback during discussions, assignments, and presentations. Their contributions have greatly enriched my understanding of the subject matter and helped me refine my research.

Finally, I want to dedicate my gratitude to my parents for all the sacrifices they have made for me. They have always supported me emotionally and financially, and I am truly grateful for their unwavering love and guidance. With their constant encouragement, I can complete this research comprehensively. Their belief in my abilities and willingness to provide me with the needed resources have been instrumental in my success. Their unwavering support has given me the confidence to overcome any obstacles that came my way, and I am forever grateful to them for their selflessness and dedication. I would also like to express our gratitude to the PERKESO Rehab Center in Ayer Keroh, Melaka, for their assistance in providing the source of the dataset for the post-stroke patient and other resources to make this research successful. Last but not least, I owe a great deal of gratitude to everyone who helped make this research possible.

Thank you, ALLAH.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

$UTeM$	-	Universiti Teknikal Malaysia Melaka
AI	-	Artificial Intelligence
ML	-	Machine learning
DL	-	Deep Learning
CT	-	Computed Tomography
CNN	-	Convolutional Neural Network
EEG	-	Electroencephalography
$MFCC$	-	Mel Frequency Cepstral Coefficients
CVA	-	Cerebral Vascular Accident
$CETI$	-	Communicative Effectiveness Index
$ALPS$	-	Aphasia Language Performance Scales
$SPICA$	-	Porch Index of Communicative Ability
WAB	-	Western Aphasia Battery
$BDAE$	-	Boston Diagnostic Aphasia Examination
$MTDDA$	-	Minnesota Test of Differential Diagnosis of Aphasia
$CADL$	-	Communicative Ability in Daily Living
$ASHA - FACS$	-	Functional Assessment of Communicative Skills for Adults
EMG	-	Electromyography
ASD	-	Autism Spectrum Disorder
VF	-	Visual Field
ASR	-	Automatic Speech Classification
DRL	-	Deep Reinforcement Learning

DNN	-	Deep Neural Network
$CALL$	-	Computer-Assisted Language Learning
$CAPT$		Computer-Assisted Pronunciation Training
SVM	-	Support Vector Machine

LIST OF PUBLICATIONS

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

Journal Publication

i. **N. S. A. Azhar**, N. M. Z. Hashim, A. I. Kamaruddin, N. A. H. Zahri and M. D. Sulistiyo, 2023. Vowel's Classification for Stroke Patients through Rehabilitation Performance via Image-profiled Sound Data. *ARPN Journal of Engineering and Applied Sciences.*, vol. 18, pp. 1411–1424, 2023. (ISI indexed, Q4, IF = 0.252 (2023)

ii. **N. S. A. Azhar**, N. M. Z. Hashim, M. M. Ibrahim, M. D. Sulistiyo, 2023. Vowel Classification for Speech Disorder Patient via Analysis on Mel Frequency Cepstral Coefficient (MFCC) Images. *Journal of Theoretical and Applied Information Technology*, vol. 101, pp. 5418–5431, 2023. (ISI indexed, Q4, IF = 0.195 (2023)

iii. **N. S. A. Azhar**, N. M. Z. Hashim, M. M. Ibrahim, M. D. Sulistiyo, 2023. Vowel Classification for Rehabilitation Assessment of Speech Disorder Patients via Multi-source Frequency Spectrum Images. *Baghdad Science Journal*. (ISI indexed, Q3, IF = 0.198 (2023)

Proceeding Publication

i. **N. S. A. Azhar**, N. M. Z. Hashim, A. I. Kamaruddin, M. D. Sulistiyo, 2022. A Deep Learning-based Smart Application for Malay Vowel Classification Toward Rehabilitation, in *2022 Engineering Technology International Conference (ETIC), 2022*, pp. 1–7 (SCOPUS indexed).

ii. **N. S. A. Azhar**, N. M. Z. Hashim, A. I. Kamaruddin, M. D. Sulistiyo, 2022. Desktop Application-based Malay Language Vowel Classification for Stroke Patient Rehabilitation Assessment, in *2022 International Borneo Innovation Exhibit and Competition (IBIEC), 2022*, pp. 106–110.

Innovation and Invention Competition Awards

i. Gold Winner Award in the JEJAK INOVASI UTeM 2022 for the innovation project of Vowel Classification System (VORECS).

ii. Silver Winner Award in the International Berneo Innovation Exhibition & Competition (IBIEC) 2022 for the innovation project of Desktop Application-based Malay Language Vowel Classification for Stroke Patient Rehabilitation Assessment.

iii. Bronze Winner Award in the Malaysia Technology Expo (MTE) 2023 for the innovation project of Vowel classification System (VoReCS).

Intellectual Property

Copyright title: Human Audio to Spectrogram Image for Malay Language Vowel Dataset (HASIM)

Application number: AR2022M03701

CHAPTER 1

INTRODUCTION

1.1 Research Background

Artificial intelligence (AI) is a complex creation of humans. AI is everywhere, moulding modern society in many ways, and continuing to advance, which nowadays its impact on various industries is becoming increasingly evident. While AI has made remarkable progress in tasks like language processing and pattern classification, it still needs to understand the humanly context and emotions. These questions are not just philosophical musings but pressing inquiries that will shape the future of our world.

The seeds of AI were sewn in the mid-20th century, a time of siginificant technological advancement. One of the key figures who laid the groundwork for AI was Alan Turing, known as the father of modern computing. Turing proposed an experiment that would become the benchmark for artificial intelligence (Varol Akman, 2000). Others, like John McCarthy, came up with the phrase "artificial intelligence" and planned the first AI conference at Dartmouth College in 1956. The idea put forth by Turing and his associates was that each aspect of learning and every other aspect of intelligence could be so thoroughly specified that a computer could replicate it (Cordeschi, 2007). It is the first formal classification of AI as a field of research, and it sent ripples through the scientific community.

Artificial intelligence began to permeate various sectors, subtly influencing the way humans live, work and play in healthcare. AI started to lend a helping hand to doctors and