



**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

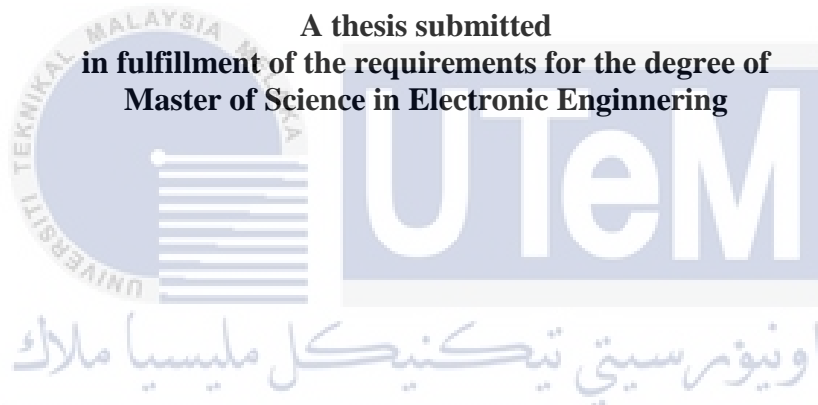
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

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



Faculty of Electronics and Computer Technology and Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2024

DEDICATION

To my beloved ibu and ayah.



اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

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 Kecerdasan Buatan (AI), telah membuka penyelesaian inovatif dalam pemulihan pertuturan. Kajian AI telah menumpukan pada pengesanan 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 pengesanan 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 diperolehi 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

	PAGES
DECLARATION	
APPROVAL	
DEDICATION	
ABSTRACT	i
ABSTRAK	ii
ACKNOWLEDGEMENT	iii
TABLE OF CONTENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xvi
LIST OF PUBLICATIONS	xviii
 CHAPTER	
1. INTRODUCTION	1
1.1 Research Background	1
1.2 Research Problems	5
1.3 Research Question	8
1.4 Research Objective	9
1.5 Scope of Research	9
1.6 Significance	11
1.7 Thesis Outline	11
 2. LITERATURE REVIEW	13
2.1 Introduction	13
2.2 Medical Terminology-related Speech Impairments and The Development of Speech Rehabilitation	14
2.2.1 Impact of Stroke on Speech Disorders	14
2.2.2 Aphasia and Related Speech Disorder	15
2.2.3 The Development of Speech Rehabilitation	19
2.2.4 Language Testing	20
2.2.5 Standardized Assessment Tools	21
2.2.6 Cranial Nerve Examination	22
2.3 Before Deep Learning and Speech Classification Methods	23
2.3.1 Motor Learning and Articulation Drill Therapy	24
2.3.2 Phonological/Lexical Interventions	25
2.4 Deep Learning and Approaches to Speech Classification	26
2.4.1 Image Profile in Speech Classification	26
2.4.2 Mel Spectrogram	27
2.4.3 Mel Frequency Cepstral Coefficient (MFCC)	29
2.4.4 Pre-trained Network Model in Deep Learning	31
2.4.5 VGG Network Model	32
2.4.6 Inception Network Model	34
2.4.7 AlexNet Network Model	35
2.4.8 Medical Related to Deep Learning Approaches	36

2.4.9 Speech Classification by Deep Learning	41
2.5 Summary	46
3. METHODOLOGY	47
3.1 Introduction	47
3.2 Development of Dataset Collection	48
3.2.1 Datasets Partition for Analytical Use	48
3.2.2 Recording Procedure and Specifications	50
3.2.3 Conversion Audio to the Image Profile	52
3.2.4 Image Profile via Mel Spectrogram	54
3.2.5 Image Profile via Mel Frequency Cepstral Coefficient (MFCC)	56
3.2.6 Cropping Process	57
3.3 Convolutional Neural Network in Deep Learning Approach	60
3.3.1 Network Models	61
3.3.2 Proposed Network Model	62
3.3.3 Images Distribution	66
3.4 Quantitative Comparative Analysis	73
3.4.1 Characterization of Errors	73
3.5 Qualitative Analysis	74
3.6 Summary	76
4. RESULT AND DISCUSSION	77
4.1 Introduction	77
4.2 Hyper-Parameter Variables	78
4.3 Performance Accuracy of the Mel Spectrogram Image Profile Sound	80
4.3.1 Analysis 1(a): Classification Performance Accuracy of the Mel Spectrogram Image Profile Sound for the Normal Person Group	81
4.3.2 Analysis 1(b): Classification Performance Accuracy of the Mel Spectrogram Image Profile for the Post-stroke Patients Group	85
4.3.3 Analysis 1(c): Classification Performance Accuracy of the Mel Spectrogram Image Profile Sound for the Combination Group	89
4.4 Performance Accuracy of the Mel Frequency Cepstral Coefficient (MFCC)	93
4.4.1 Analysis 2(a): Classification Performance Accuracy of the MFCC Image Profile Sound for the Normal Person Group	93
4.4.2 Analysis 2(b): Classification Performance Accuracy of the MFCC Image Profile Sound for the Post-stroke Patients Group	97
4.4.3 Analysis 2(c): Classification Performance Accuracy of the MFCC Image Profile Sound for the Combination Group	101
4.5 Performance Accuracy of Proposed Network models between Six and Twelve Vowel Classes	104
4.5.1 Analysis 3: The Comparison of Comparative network models by using Mel Spectrogram Image Profile between Six and Twelve Vowels Classes	104
4.5.2 Analysis 4: The Comparison of Comparative network models by using MFCC Image Profile between Six and Twelve Vowels Classes	109

4.6	Performance of the Classification Models	113
4.6.1	Proposed network Model Classification of Mel Spectrogram Image Profile	115
4.6.2	Proposed Network Model Classification of MFCC Image Profile	119
4.6.3	Proposed network Model Classification between Six and Twelve Vowel Classes	122
4.7	Qualitative Assessment of Vowel Classification by Human Assessment	131
4.7.1	Qualitative Comparative of Human Assessment for Vowel Class /a/	134
4.7.2	Qualitative Comparative of Human Assessment for Vowel Class /e/	136
4.7.3	Qualitative Comparative of Human Assessment for Vowel Class /E/	138
4.7.4	Qualitative Comparative of Human Assessment for Vowel Class /i/	140
4.7.5	Qualitative Comparative of Human Assessment for Vowel Class /o/	142
4.7.6	Qualitative Comparative of Human Assessment for Vowel Class /u/	144
4.7.7	Qualitative Comparative of Accuracy for Six and Twelve Vowel Classes	147
4.8	Summary	149
5.	CONCLUSION & RECOMMENDATIONS FOR FUTURE RESEARCH	150
5.1	Introduction	150
5.2	Summary of the Research Objectives	150
5.3	Future Prospects and remaining Constraints	153
5.4	Closing Remarks	156
6.	REFERENCES	158

LIST OF TABLES

TABLE	TITLE	PAGE
Table 2.1	Comparison of Fluent and Non-Fluent Aphasia Affecting Speech of Stroke Patients	16
Table 2.2	An Evaluation of Three Primary Fluent Aphasias for Speech Rehabilitation	17
Table 2.3	An Evaluation of Three Primary Non-Fluent Aphasias for Speech Rehabilitation	18
Table 2.4	The Cranial Nerves Examination Related to Speech Impairments	23
Table 2.5	The Cranial Nerves Examination Related to Speech Impairments	23
Table 2.6	Medical Studies Related to Deep learning Approaches	38
Table 2.7	CNN model specifications of the CNNeeg1-1 Architecture for EEG Signal Detection Research	39
Table 2.8	Studies Related to Speech Classification by Deep Learning	45
Table 3.1	Dataset Partition for Six and Twelve Vowel Classes	50
Table 3.2	Dimension of Image Profile Dataset	57
Table 3.3	The Proposed Network Model Layers	64
Table 3.4	Total of Mel Spectrogram Image Profile Collected in Every Vowel Class	71
Table 3.5	Total of MFCC Image Profile Collected in Every Vowel Class	72
Table 4.1	Training and Validation Accuracy for the Classification of the Different Models in Analysis 1(a)	81
Table 4.2	Training and Validation Accuracy for the Classification of the Different Models in Analysis 1(b)	86
Table 4.3	Training and Validation Accuracy for the Classification of the Different Models in Analysis 1(c)	90
Table 4.4	Training and Validation for the Classification Accuracy of the Different Models in Analysis 2(a)	94

Table 4.5	Training and Validation for the Classification Accuracy of the Different Models in Analysis 2(b)	98
Table 4.6	Training and Validation for the Classification Accuracy of the Different Models in Analysis 2(c)	101
Table 4.7	Performance Accuracy of Comparative network models using Mel Spectrogram Image Profile between Six and Twelve Vowel Classes	108
Table 4.8	Performance Accuracy of Comparative network models using MFCC Image Profile between Six and Twelve Vowel Classes	111
Table 4.9	Performance Accuracy of Proposed network Models using Mel Spectrogram and MFCC Image Profile between Six and Twelve Vowel Classes	112
Table 4.10	Comparison of the Precision, Recall and F1 for Mel Spectrogram Image Profile of Six Vowel Classes	118
Table 4.11	Comparison of the Precision, Recall and F1 for Mel Spectrogram Image Profile of Six Vowel Classes	119
Table 4.12	Twelve Vowel Classes Label for Confusion Matrix	122
Table 4.13	Comparison of the Precision, Recall and F1 for Mel Spectrogram Image Profile of Twelve Vowel Classes	124
Table 4.14	Comparison of the Precision, Recall and F1 for MFCC Image Profile of Twelve Vowel Classes	128

LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 2.1	Vowels, Diphthongs and Consonant Production in Human Larynx (Story, 2015)	19
Figure 2.2	The Flow to Generate Mel Spectrogram	27
Figure 2.3	The Flow to Generate MFCC Images	30
Figure 2.4	VGG Network Architecture (Prerepa Gayathri, 2023)	33
Figure 2.5	Inception Network Model Architecture	34
Figure 2.6	AlexNet Network Model Architecture (P.Haripriya, 2019)	35
Figure 3.1	Research Work Flow	48
Figure 3.2	Method of Vowel Recording	51
Figure 3.3	Recording Session for Normal Person Group	52
Figure 3.4	Signal Audio in wav.file	53
Figure 3.5	Audio Image of Mel Spectrogram Image Profile	53
Figure 3.6	Audio Image of Mel Frequency Cepstral Coefficient Image Profile	54
Figure 3.7	Mel Spectrogram images of part (a)-(f) for vowel /a/, /e/, /E/, /i/, /o/, and /u/ for normal person group and part (g)-(l) for vowel /a/, /e/, /E/, /i/, /o/, and /u/ for post-stroke patient group	55
Figure 3.8	MFCC images of part (a)-(f) for vowel /a/, /e/, /E/, /i/, /o/, and /u/ for normal person group and part (g)-(l) for vowel /a/, /e/, /E/, /i/, /o/, and /u/ for post-stroke patient group	57
Figure 3.9	The Image Cropping Process	59
Figure 3.10	Optimal Split Images Percentage for CNN Network model in Analysis 1(a)	67
Figure 3.11	Optimal Split Images Percentage for CNN Network model in Analysis 1(b)	67

Figure 3.12	Optimal Split Images Percentage for CNN Network model in Analysis 1(c)	68
Figure 3.13	Optimal Split Images Percentage for CNN Network model in Analysis 2(a)	69
Figure 3.14	Optimal Split Images Percentage for CNN Network model in Analysis 2(b)	69
Figure 3.15	Optimal Split Images Percentage for CNN Network model in Analysis 2(c)	70
Figure 4.1	Proposed network Model Performance Accuracy Graph using Epoch 50 and Batch Size 6	79
Figure 4.2	Proposed network Model Performance Accuracy Graph using Epoch 50 and Batch Size 15	79
Figure 4.3	Proposed network Model Performance Accuracy Graph using Epoch 80 and Batch Size 15	80
Figure 4.4	Performance Accuracy Graph for Proposed Network Model in Analysis 1(a)	82
Figure 4.5	Performance Accuracy Graph for VGG16 Model in Analysis 1(a)	83
Figure 4.6	Performance Accuracy Graph for VGG19 Model in Analysis 1(a)	83
Figure 4.7	Performance Accuracy Graph for Inception Model in Analysis 1(a)	84
Figure 4.8	Performance Accuracy Graph for AlexNet Model in Analysis 1(a)	84
Figure 4.9	Performance Accuracy Graph for Proposed network Model in Analysis 1(b)	85
Figure 4.10	Performance Accuracy Graph for VGG16 Model in Analysis 1(b)	87
Figure 4.11	Performance Accuracy Graph for VGG19 Model in Analysis 1(b)	87
Figure 4.12	Performance Accuracy Graph for Inception Model in Analysis 1(b)	88

Figure 4.13	Performance Accuracy Graph for AlexNet Model in Analysis 1(b)	88
Figure 4.14	Performance Accuracy Graph for Proposed network Model in Analysis 1(c)	89
Figure 4.15	Performance Accuracy Graph for VGG16 Model in Analysis 1(c)	90
Figure 4.16	Performance Accuracy Graph for VGG19 Model in Analysis 1(c)	92
Figure 4.17	Performance Accuracy Graph for Inception Model in Analysis 1(c)	92
Figure 4.18	Performance Accuracy Graph for AlexNet Model in Analysis 1(c)	92
Figure 4.19	Performance Accuracy Graph for Proposed network Model in Analysis 2(a)	95
Figure 4.20	Performance Accuracy Graph for VGG16 Model in Analysis 2(a)	95
Figure 4.21	Performance Accuracy Graph for VGG19 Model in Analysis 2(a)	95
Figure 4.22	Performance Accuracy Graph for Inception Model in Analysis 2(a)	96
Figure 4.23	Performance Accuracy Graph for AlexNet Model in Analysis 2(a)	97
Figure 4.24	Performance Accuracy Graph for Proposed network Model in Analysis 2(b)	98
Figure 4.25	Performance Accuracy Graph for VGG16 Model in Analysis 2(b)	99
Figure 4.26	Performance Accuracy Graph for VGG19 Model in Analysis 2(b)	99
Figure 4.27	Performance Accuracy Graph for Inception Model in Analysis 2(b)	99
Figure 4.28	Performance Accuracy Graph for AlexNet Model in Analysis 2(b)	100

Figure 4.29	Performance Accuracy Graph for Proposed network Model in Analysis 2(c)	102
Figure 4.30	Performance Accuracy Graph for VGG16 Model in Analysis 2(c)	102
Figure 4.31	Performance Accuracy Graph for VGG19 Model in Analysis (c)	103
Figure 4.32	Performance Accuracy Graph for Inception Model in Analysis 2(c)	103
Figure 4.33	Performance Accuracy Graph for AlexNet Model in Analysis 2(c)	103
Figure 4.34	Performance Accuracy Graph of the Proposed Network Model for Normal Persons Group of Six Vowel Classes	105
Figure 4.35	Performance Accuracy Graph of the Proposed Network Model for Normal Persons Group of Twelve Vowel Classes	105
Figure 4.36	Sample of Mel Spectrogram of Vowel Class /a/ and /i/	106
Figure 4.37	Performance Accuracy Graph of the Proposed Network Model for Normal Persons Group of Six Vowel Classes	110
Figure 4.38	Performance Accuracy Graph of the Proposed Network Model for Normal Persons Group of Twelve Vowel Classes	110
Figure 4.39	Confusion Matrix of Mel Spectrogram Image Profile for Six Vowel Class of Normal Persons Group Dataset	116
Figure 4.40	Confusion Matrix of Mel spectrogram Image Profile for Six Vowel Class of Post-stroke Patients Group Dataset	117
Figure 4.41	Confusion Matrix of Mel Spectrogram Image Profile for Six Vowel Class of Combination Group Dataset	117
Figure 4.42	Confusion Matrix of MFCC Image Profile for Six Vowel Class of Normal Persons Group Dataset	120
Figure 4.43	Confusion Matrix of MFCC Image Profile for Six Vowel Class of Post-stroke Patients Group Dataset	121
Figure 4.44	Confusion Matrix of MFCC Image Profile for Six Vowel Class of Combination Group Dataset	121

Figure 4.45	Confusion Matrix of Mel Spectrogram Image Profile for Twelve Vowel Class of Normal Persons Group Dataset	125
Figure 4.46	Confusion Matrix of Mel Spectrogram Image Profile for Twelve Vowel Class of Post-stroke Patient Group Dataset	125
Figure 4.47	Confusion Matrix of Mel Spectrogram Image Profile for Twelve Vowel Class of Combination Group Dataset	126
Figure 4.48	Confusion Matrix of MFCC Image Profile for Twelve Vowel Class of Normal Persons Group Dataset	129
Figure 4.49	Confusion Matrix of MFCC Image Profile for Twelve Vowel Class of Post-stroke Group Dataset	129
Figure 4.50	Confusion Matrix of MFCC Image Profile for Twelve Vowel Class of Combination Group Dataset	130
Figure 4.51	Pie Chart for Respondent's Gender	131
Figure 4.52	Pie Chart for Respondent's Age	132
Figure 4.53	Pie Chart for Respondent's Stroke History	133
Figure 4.54	Pie Chart for Respondent's Rehabilitation Related	133
Figure 4.55	Pie Chart 1 for Human Assessment of Vowel Class /a/	134
Figure 4.56	Pie Chart 2 for Human Assessment of Vowel Class /a/	135
Figure 4.57	Pie Chart 1 for Human Assessment of Vowel Class /e/	136
Figure 4.58	Pie Chart 2 for Human Assessment of Vowel Class /e/	137
Figure 4.59	Pie Chart 1 for Human Assessment of Vowel Class /E/	138
Figure 4.60	Pie Chart 2 for Human Assessment of Vowel Class /E/	139
Figure 4.61	Pie Chart 1 for Human Assessment of Vowel Class /i/	140
Figure 4.62	Pie Chart 2 for Human Assessment of Vowel Class /i/	141
Figure 4.63	Pie Chart 1 for Human Assessment of Vowel Class /o/	142
Figure 4.64	Pie Chart 2 for Human Assessment of Vowel Class /o/	143
Figure 4.65	Pie Chart 1 for Human Assessment of Vowel Class /u/	144
Figure 4.66	Pie Chart 2 for Human Assessment of Vowel Class /u/	145

Figure 4.67 Bar Graph 1 of the Quantitative and Qualitative Comparative 147

Figure 4.68 Bar Graph 2 of the Quantitative and Qualitative Comparative 147



LIST OF ABBREVIATIONS

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

<i>DNN</i>	-	Deep Neural Network
<i>CALL</i>	-	Computer-Assisted Language Learning
<i>CAPT</i>		Computer-Assisted Pronunciation Training
<i>SVM</i>	-	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 significant 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