SPEECH-BASED DEPRESSION RECOGNITION FOR BAHASA MALAYSIA SPEAKERS USING DEEP LEARNING MODELS

BY

MUGAHED AL EZZI AHMED EZZI

A thesis submitted in fulfillment of the requirement for the degree of Master of Science in Engineering

Kulliyyah of Engineering
International Islamic University Malaysia

OCTOBER 2021

ABSTRACT

Depression is a mental disorder of high prevalence, leading to a negative effect on individuals, family members, society, and the economy. Traditional clinical diagnosis methods are subjective, complicated, and require extensive participation of experts. Furthermore, the severe shortage in psychiatrists' ratio per population in Malaysia imposes patients' delay in seeking treatment and poor compliance to follow-up. On the other side, the social stigma of visiting psychiatric clinics also prevents patients from seeking early treatment. Automatic depression detection using speech signals is a promising depression biometric because it is fast, convenient, and non-invasive. However, current machine learning algorithms could not achieve high accuracy and robust results yet. Moreover, the existing researches and approaches have minimal support to Bahasa Malaysia. This research attempts to develop an end-to-end deep learning model to classify depression from Bahasa Malaysia speech using our dataset collected from clinically depressed and healthy Bahasa Malaysia speakers. The dataset was collected via an online platform using participants' mobile phones to record their read and spontaneous speech and depression status. Depression status is identified by the Patient Health Questionnaire (PHQ-9), the Malay Beck Depression Inventory-II (Malay BDI-II), and subjects' declaration of Major Depressive Disorder diagnosis by a trained clinician. The dataset consists of 42 and 11 depressed female and male participants, respectively, and 68 and 9 healthy female and male participants. However, this research study focuses on female data only due to data insufficient. We provided a detailed implementation of the deep learning model using two approaches: raw audio input and acoustic features input. Multiple combinations of speech types were analyzed using various deep neural network models. Additionally, an analysis of robust feature selection was carried out on the acoustic features input before proceeding to the deep learning models. After performing hyperparameters tuning, raw audio input from female read and female spontaneous speech combination using AttCRNN model achieved an accuracy of 91%. In comparison, robust acoustic features input from female spontaneous speech using RNN model achieved an accuracy of 85%. These results could be improved by providing a larger dataset. Besides, male and gender-independent models could be further studied.

خلاصة البحث

الاكتئاب هو اضطراب عقلي ذو انتشار واسع، يؤدي إلى تأثير سلبي على الفرد والأسرة والمجتمع وأيضا على الاقتصاد. طرق التشخيص السريري التقليدية غير موضوعية ومعقدة وتتطلب مشاركة واسعة من الخبراء. علاوة على ذلك، فإن النقص الحاد في نسبة الأطباء النفسيين للسكان في ماليزيا سبب في تأخر تلقي العلاج وصعوبة امتثال المرضى للمتابعة الطبية. من ناحية أخرى، فإن النظرة المجتمعية -الخاطئة- حول زيارة عيادات الطب النفسي تمنع المرضى من تطلب العلاج المبكر. ومع ذلك، لم تتمكن خوارزميات التعلم الآلي الحالية من تحقيق دقة عالية ونتائج قوية حتى الآن. علاوة على ذلك، فإن الأبحاث والأساليب الحالية تحظى بدعم ضئيل للغة البهاسا الماليزية. يعد اكتشاف الاكتئاب باستخدام مؤشر الكلام مقياسًا حيويًا واعدًا لأنه سريع، وعملي، ولا يتطلب التدخل الجراحي. يهدف هذا البحث إلى تطوير نموذج تعلم عميق، شامل لتصنيف الاكتئاب عن طريق التخاطب بلغة البهاسا الماليزية باستخدام مجموعة من البيانات جمعت من متحدثين أصحاء وآخرين مصابين بالاكتئاب. تم جمع مجموعة البيانات عبر منصة إلكترونية باستخدام هواتف المشاركين المحمولة لتسجيل خطاب صوتي مقروء وعفوي، لرصد حالات الاكتئاب. يتم تحديد حالة الاكتئاب من خلال استبيان لتحديد صحة المريض (PHQ-9)، ومن خلال استبيان قائمة بيك للاكتئاب المترجمة للملايو (Malay BDI-II)، ثم يتم توضيح تشخيص الاكتئاب للمشتركين من قبل طبيب مختص. تتكون مجموعة البيانات من ٤٢ امرأة مصابة بالاكتئاب، ٦٨ امرأة سليمة إضافة الى ١١ رجلا مصابا، ٩ رجال أصحاء. ومع ذلك، تركز هذه الدراسة البحثية على بيانات الإناث فقط بسبب وفرة البيانات وقدرة الباحث على الاعتماد عليها. قدمنا نموذج تنفيذي مفصل للتعلم العميق باستخدام طريقتين: مدخلات الصوت الخام ومدخلات الملامح/السمات الصوتية. تم تحليل مجموعات متعددة من أساليب الكلام باستخدام نماذج مختلفة للشبكات العصبية العميقة. بالإضافة إلى ذلك، تم إجراء تحليل على إدخال السمات الصوتية لاختيار السمات الراسخة قبل الانتقال إلى نماذج التعلم العميق. بعد إجراء ضبط المعلمات الفائقة، حقق إدخال الصوت الخام من دمج الخطاب العفوي والمقروء للإناث باستخدام نموذج AttCRNN دقة ٩١٪. في المقارنة. بينما حققت السمات الصوتية التي تم إدخالها عن طريق الكلام العفوي للإناث باستخدام نموذج RNN دقة ٨٥٪. يمكن تحسين هذه النتائج من خلال توفير مجموعة بيانات أكبر. إلى جانب ذلك، يمكن دراسة نماذج للذكور ونماذج غير معتمدة على جنس المتكلم.

APPROVAL PAGE

I certify that I have supervised and read this study and that, in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Engineering

…………………………………..
Nik Nur Wahidah Nik Hashim
Supervisor

…………………………………..
Hasan Firdaus bin Mohd Zaki
Co-Supervisor

I certify that I have read this study and that, in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Engineering

…………………………………..
Azhar Mohd Ibrahim
Internal Examiner

…………………………………..
Norashikin Yahya
External Examiner

This thesis was submitted to the Department of Mechatronics Engineering and is accepted as a fulfillment of the requirement for the degree of Master of Science in Engineering

…………………………………..
Ali Sophian
Head, Department of
Mechatronics Engineering

This thesis was submitted to the Kulliyyah of Engineering and is accepted as a fulfillment of the requirement for the degree of Master of Science in Engineering

…………………………………..
Sany Izan Ihsan
Dean, Kulliyyah of Engineering

DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Mugahed Al Ezzi Ahmed Ezzi

Signature .... Date ......28/10/2021......

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

SPEECH-BASED DEPRESSION RECOGNITION FOR BAHASA MALAYSIA SPEAKERS USING DEEP LEARNING MODELS

I declare that the copyright holders of this thesis are jointly owned by the student and IIUM.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may be used by others in their writing with due acknowledgment.

2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purposes.

3. The IIUM library will have the right to make, store in a retrieved system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understood the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Mugahed Al Ezzi Ahmed Ezzi

……..…………………….. Signature

…… 28/10/2021 …….. Date

ACKNOWLEDGMENTS

It is my utmost pleasure to dedicate this work to my dear parents and my family, who granted me the gift of their unwavering belief in my ability to accomplish this goal: thank you for your support and patience.

I wish to express my appreciation and thanks to those who provided their time, effort, and support for this project. Special thanks to Asst. Prof Dr. Nik Nur Wahidah Nik Hashim as a supervisor for her continuous support, encouragement, and leadership, and for that, I will be forever grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AttCRNN	Attention Convolutional Recurrent Neural Network
AGC	Automatic Gain Control
ANS	Automatic Nervous System
BN	Batch Normalization
BDI	Beck Depression Index
CV	Coefficient of Variance
CONV	Convolutional
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
DFT	Discrete Fourier Transform
DNN	Deep Neural Network
ECG	Electrocardiogram
EEG	Electroencephalogram
FFT	Fast Fourier Transform
FR	Female Read
FS	Female Spontaneous
FC	Fully Connected
GMM	Gaussian mixture model
GPU	Graphical Processing Unit
HAMD	Hamilton Rating Scale for Depression
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
LLD	Low-Level Descriptors

MDD	Major Depressive Disorder
MAE	Mean Absolute Error
MFCCs	Mel-Frequency Cepstral Coefficients
MADRS	Montgomery Asberg Depression Rating Scale
MIDI	Musical Instrument Digital Interface
QIDS	Quick Inventory of Depressive Symptomology
RNN	Recurrent Neural Network
SNS	Somatic Nervous System
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
AVEC	The Audio/Visual Emotion Challenge
Tonnetz	Tonal centroid features
VQal	Voice Quality
WHO	World Health Organization

LIST OF SYMBOLS

x	Audio Signal
μ	Mean
σ	Standard Deviation
t_n	Tonal Centroid Vector
n	Time Frame
c_n	Chroma Vector
T	Transformation Matrix

CHAPTER ONE

INTRODUCTION

1.1 OVERVIEW

Depression is one of the most common mental disorders. Globally, it is estimated that more than 300 million people of all ages suffer from depression (World Health Organization, 2017). Consequently, it can cause the affected person to function poorly at work, school, and within the family. At its worst, depression can lead to suicide. The case of Malaysia is no exception; in fact, depression is the most common mental illness reported in Malaysia (Mukhtar & P. S. Oei, 2011). In 2015, it was reported that 29.2% of Malaysians had suffered from mental illnesses, including depression and anxiety disorder (Institute for Public Health, 2015). An early intervention aimed at preventing the onset of clinical depression can provide an effective means for reducing the disease's burden. However, currently, the range of diagnostic tools for identifying depression is quite limited. Assessment methods rely almost exclusively on patient self-reporting and clinical opinion, risking a variety of subjective biases. Consequently, it is essential to look for new objective measures that help clinicians diagnose and monitor clinical depression (H. Jiang et al., 2017).

Non-verbal information processing of speech for creating various IT tools is an important area for cognitive info-communication (Baranyi, Csapo, & Sallai, 2015). The fascinating field of using speech as a bio-signal and developing different non-invasive diagnostic tools makes an automatic assessment of the people's cognitive and psychological state possible. Automatic depression detection based on speech processing belongs to this research field

1.2 STATEMENT OF THE PROBLEM

Nowadays, the number of people who have a mental illness is increasing dramatically all over the world. In regions with a relatively large population, such as Western-Pacific and South-East Asia, around half of the people suffer from depressing live (World Health Organization, 2017). In Malaysia, the prevalence of mental health issues has been steadily increasing from 10.7% in 1996, to 11.2% in 2006, to 29.2% in 2015 (Institute for Public Health, 2015). World Health Organization (WHO) 's Global Health Observatory data repository reported that there were only 1.05 psychiatrists per 100,000 population for Malaysia in 2016 (World Health Organization, 2019). As of the year 2018, a more recent study showed 410 registered psychiatrists in Malaysia, representing 1.27 psychiatrists per 100,000 population. WHO has recommended a ratio of psychiatrists to the Malaysian population of 1:10,000. However, the current ratio is only 1:80,000 (Guan, Lee, Francis, & Yen, 2018). This severe shortage of psychiatrists in the country may pose several problems for those facing mental health issues. These problems include delay in seeking treatment, long waiting time for psychiatric consultation, low-quality outpatient mental health care, poor compliance to follow-up and treatment, increased drug abuse and addiction cases, a surge in suicide rates, unemployment, and homelessness.

Another contributing factor to preventing people with mental health issues from seeking clinical treatment is the social stigma, especially in Malaysia. People are afraid to be called crazy or psycho; hence, they will not attempt to see a psychiatrist once they feel they have a mental issue during the early stage.

Early detection and treatment of depression can provide effective means for minimizing the negative impacts of the illness. However, the available techniques and tools for depression diagnosis are pretty limited. In psychiatric clinics, diagnosis procedures need extensive participation of experts, and it can be subjective. Other techniques such as electroencephalogram (EEG) and electrocardiogram (ECG) signals are time-consuming, complicated, and expensive. They require a skilled and experienced clinician to perform the device setup and proper wire connection to get correct data, then analyze them and draw a diagnosis. Additionally, it restricts body movement during the test, causing lots of inconvenience to the patient.

The utilization of speech as a bio-marker tool to assess the state of depression is non-invasive, fast, and easy to access by a broader range of people. Speech-based detection of depression has attracted increasing attention from researchers in psychology, computer science, linguistics, and related disciplines. However, current machine learning algorithms could not achieve high accuracy and robust results yet. Moreover, the existing researches and approaches have minimal support to Bahasa Malaysia.

1.3 RESEARCH OBJECTIVES

The research aims to achieve the following objectives:

1. To develop a speech depression dataset from Bahasa Malaysia female speakers.
2. To develop an end-to-end deep learning models that classifies the depression state from Bahasa Malaysia female speech using raw audio input and acoustic feature input and compare between read and spontaneous speech.
3. To analyze microphone independent features for robust speech depression detection and evaluate the performance of the developed models.

1.4 SIGNIFICANCE OF RESEARCH

The significance of contributing to depression detection research study is not providing better or safer treatment to the illness. Further, it does not mean dispensing with the psychiatrist role altogether but rather providing an alternative. Nevertheless, the study’s goal is to provide a convenient and accessible tool to predict depression. Thus, encouraging them to seek treatment at an early stage to prevent the disorder’s consequences before it worsens, including suicide as its worst-case scenario.

1.5 RESEARCH SCOPE AND LIMITATION

The research aims to focus on developing a speech-based depression detection. The scope of the research covers the Bahasa Malaysia language. Detection of depression consists of classifying the input audio signal as depressed or healthy (not depressed).

Collecting speech dataset is one of the biggest challenges in this research. Reaching out to people diagnosed with depression, recording their voices, and collecting their data are not easy tasks. Ideally, on-site recording by collaborating with a psychiatric clinic should be performed. However, due to the COVID-19 pandemic, this was totally out of the question. Alternatively, we performed online data collection utilizing social media to find our targeted participants. Due to the limited access to speech dataset in general, and male data in particular, we have decided to focus in this study on female participants’ data only.

1.6 RESEARCH METHODOLOGY

The following methodology will be adopted to achieve the objectives of the research.

1. Extensive literature review on speech-based depression detection.
2. Data collection

3. Dataset pre-processing and filtering
4. Acoustic feature extraction and analysis
5. Raw audio pre-processing and normalizing
6. Develop and train different DNNs models for classification
7. Model testing, validation, and analysis

Figure 1.1 shows the flowchart of the research methodology to be adopted in this research.

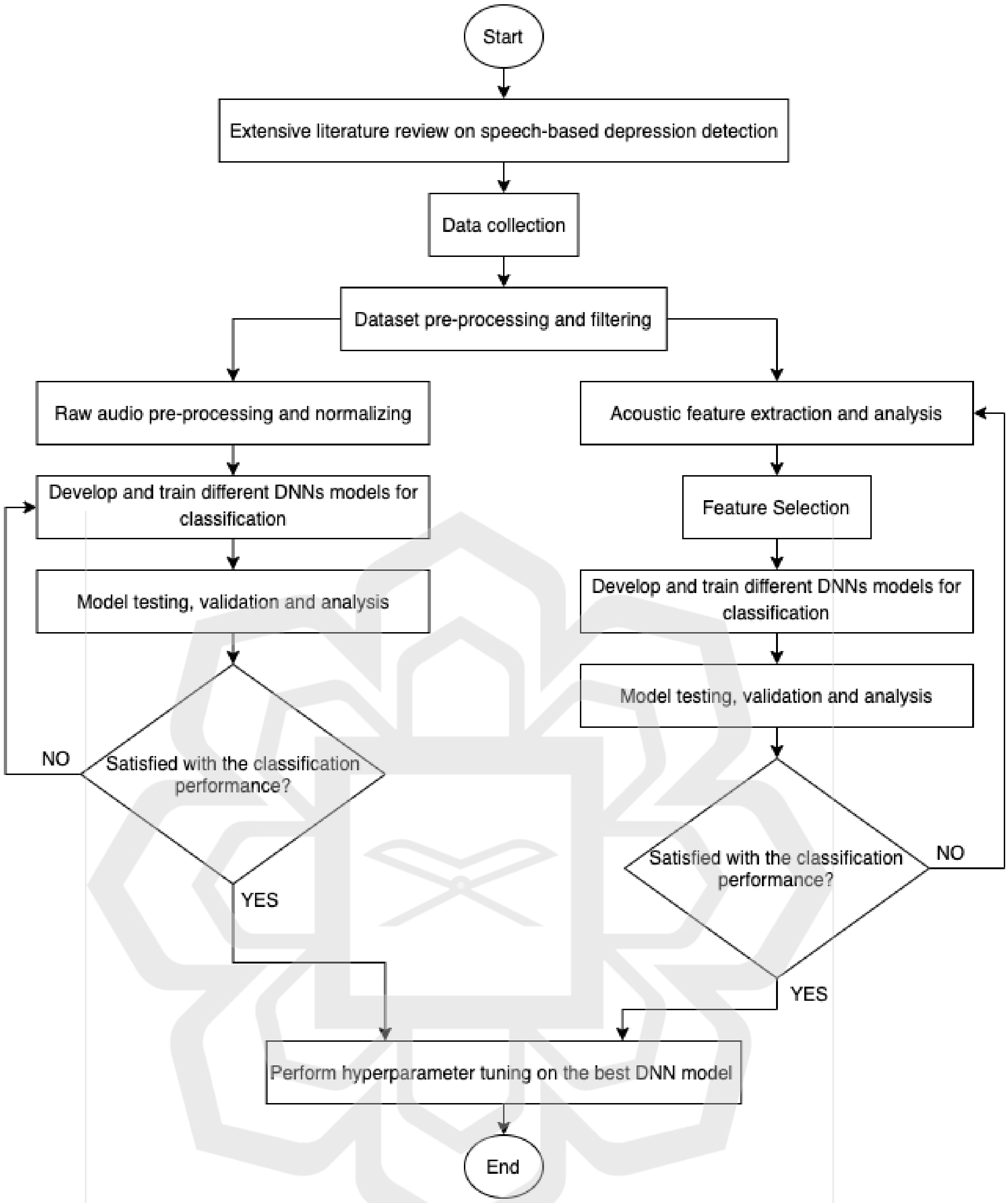

Figure 1.1: Research methodology flow chart

1.7 THESIS ORGANIZATION

This research thesis is divided into several chapters.

Chapter 1: Introduction

This chapter discusses the research overview, including research objectives, problem statements, research scope, research significance, and research workflow.

Chapter 2: Literature Review

This chapter reviews the literature on speech-based depression detection. The review covers an overview of depression, some of its assessment tools, and how it is related to speech signals. Furthermore, it reviews the previous works and researches on speech depression detection. This review will help us extract the important concepts and get the general concept to finally develop our model.

Chapter 3: Methodology

This chapter discusses the process of data collection and its pre-processing, DNNs model design and architecture, and audio features extraction.

Chapter 4: Results and Discussion

Results of different input formats such as raw input and feature input are compared and discussed in this chapter. This includes all training performance and classification results.

Chapter 5: Conclusion

This chapter summarizes what was achieved in this research study. Moreover, this chapter discusses the limitations and recommendations.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

Professionals across numerous fields are trying to develop helpful tools for depression assessment. Each subfield tends to approach the job from a unique point of view, with somewhat different objectives and totally different database sources. Due to these observed distinctions, it is difficult to compare approaches (Morales, Scherer, & Levitan, 2017).

Despite the existing differences, every system and research study shares the typical goal of discovering a method to utilize innovation to assist evaluate depression (Morales et al., 2017). This chapter aims to serve as a bridge between the subfields by providing a comprehensive review of depression detection systems throughout subfields. In this chapter, we will address how depression has been defined and annotated in depression detection systems, what are the conventional assessment tools, how are speech signals affected by the mental state of depression, what are the existing detection systems available, and how well do they perform.

Major depressive disorder (MDD) is a common and serious medical illness that negatively affects how you feel, think, and act. It can lead to various emotional and physical problems and decrease a person's ability to function at work and home (American Psychiatric Association, 2013).

Depression symptoms can differ from moderate to serious and can consist of the feeling of unhappiness, loss of interest, weight gained or loss unassociated to dieting, changes in appetite, insomnia or hypersomnia, increased fatigue or loss of energy, increase in purposeless physical activity, for example, hand-wringing or pacing or slowed movements and speech (actions observable by others), feeling blameworthy or worthless, bradyphrenia, difficulty concentrating or making decisions, thoughts of death or suicide. Symptoms must last for at least two weeks for a medical diagnosis of depression (American Psychiatric Association, 2013)

2.2 DEPRESSION ASSESSMENT

2.2.1 Gold-Standard Questionnaires

Frequently used assessment tools consist of meeting style assessments such as the Hamilton Score Scale for Depression (HAMD) (Hamilton, 1960) or self-assessments such as the Beck Depression Index (BDI) originally published in 1961 as well as changed in 1996 (Beck, Steer, & Brown, 1996). Both analysis methodologies rate the extent of 21 signs and symptoms observed in anxiety to offer a client a score related to their level of anxiety. The significant distinctions in between both scores are that HAMD is a clinician-rated survey that can be finished in 20 to 30 minuets, while BDI is a self-reported questionnaire that can be completed in just 5 to 10 minutes. Both ranges utilize different things; the HAMD prefers neuro-vegetative signs (signs and symptoms that affect an individual's everyday functioning such as weight, sleep, psychomotor retardation, and tiredness), while the BDI favors negative self-evaluation symptoms and also various weighting systems when generating their total score. Both evaluations have actually been revealed to have anticipating legitimacy as well as uniformity when setting apart dispirited from non-depressed individuals (Cummins et al., 2015).

The HAMD has actually long been considered as the gold standard assessment tool for depression for both diagnosis and research functions, although this status constantly comes into question (Maust et al., 2012). The HAMD evaluation rates the