

Faculty of Electrical and Electronic Engineering Technology



MUHAMMAD MU'IZZ BIN ZAKARIA

Bachelor of Electronics Engineering Technology with Honours

2022

DEVELOPMENT OF MALAYSIAN SIGN LANGUAGE (MSL) TRANSLATOR USING DEEP LEARNING APPROACH

MUHAMMAD MU'IZZ BIN ZAKARIA

A project report submitted in partial fulfillment of the requirements for the degree of Bachelor of Electronics Engineering Technology with Honours



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2022

DECLARATION

I declare that this project report entitled "Development of Malaysian Sign Language (MSL) Translator Using Deep Learning Approach" is the result of my own research except as cited in the references. The project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	ou WALAYSIA	
Student Name	: Muhammad Mu'izz Bin Zakaria	
Date	: 16/1/2023 اونيۇىرسىتى تىكنىكل مليسيا ملال	
U	NIVERSITI TEKNIKAL MALAYSIA MELAKA	

APPROVAL

I hereby declare that I have checked this project report and in my opinion, this project report is adequate in terms of scope and quality for the award of the degree of Bachelor of Electronics Engineering Technology with Honours.

Signature	WALAYSIA ME
Supervisor	Name : Dr. Haslinah Binti Mohd Nasir
Date	: 16/1/2023
	AINO
Signature	اونيۆم سيتي تيڪنيڪل مليسيا ملاك
Co-Supervi	UNIVERSITI TEKNIKAL MALAYSIA MELAKA sor :
Name (if a	ny)
Date	:

DEDICATION

To my dear family, To my loyal friends, To my supervisor and mentors, And to my dearest self.



ABSTRACT

Malaysian Sign Language (MSL) has been used as a means of communication in Malaysia between the deaf and the hearing people. Many people still could not take a grasp on what is being conveyed by the deaf which may cause misunderstanding. Thus, the aim of this project is to develop a MSL translator interface which able to detect the type of gestures when signing and a speech to text conversion. A deep learning approach is used focusing on the image classification that are known as "Hai", "Tak", "Terima Kasih" by utilizing TensorFlow and Mediapipe software where the model build will be trained using labeled images and identify its classes. This project has been develop succesfully of a MSL translator interface that can assist both the deaf and hearing people for a better two way communication in the future with the rate of confidence of more than 95 percent. In essence, it is to help in bridging the gap between the communication of the deaf and hearing people.



ABSTRAK

Bahasa Isyarat Malaysia (BIM) telah digunakan sebagai satu cara untuk komunikasi di Malaysia antara komuniti pekak dan yang mampu. Ramai orang yang masih lagi tidak dapat memahami apa yang cuba disampaikan oleh komuniti pekak yang boleh menyebabkan salah faham. Oleh itu, matlamat utama projek ini adalah untuk membina satu sistem penterjemah dimana ianya mampu untuk mengesan jenis pergerakan apabila sedang memberi isyarat tangan dan penukaran kata kepada ayat. Satu pendekatan pembelajaran dalam yang tertumpu kepada pengelasan imej yang dikenali sebagai "Hai", "Tak", "Terima Kasih" dengan memanfaatkan penggunaan perisian TensorFlow dan Mediapipe di mana binaan model akan di latih menggunakan label imej dan mengenalpasti pengelasan tersebut. Projek ini telah berjaya dibangunkan laitu satu antara muka penterjemah BIM yang dapat membantu kedua-dua komuniti pekak dan mampu untuk komunikasi dua hala yang lebih baik pada masa yang akan datang dengan kadar keyakinan melebihi 95 peratus. Asasnya, ianya adalah untuk mengurangkan jurang komunikasi antara komuniti pekak dan mampu.

اونيۈم سيتي تيڪنيڪل مليسيا ملاك UNIVERSITI TEKNIKAL MALAYSIA MELAKA

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all those who have helped and supported me throughout the completion of this project.

First and foremost, I would like to thank my supervisor, Dr. Haslinah Binti Mohd Nasir, for her guidance, encouragement and invaluable insights. Her expertise and experience have been instrumental in shaping this project.

I would also like to thank my mentors, for their invaluable contributions, feedback and suggestions. Their input has helped to make this project more robust and complete.

WALAYS /

I would like to extend my appreciation to Universiti Teknikal Malaysia Melaka, for providing the necessary resources and facilities to carry out this project.

Finally, I would like to acknowledge my family and friends, for their unwavering support and encouragement throughout this journey. Without their love and understanding, this project would not have been possible.

I am truly grateful to everyone who has been a part of this project and made it a success.

TABLE OF CONTENTS

		PAGE
DEC	LARATION	
APP	ROVAL	
DED	ICATIONS	
ABS	TRACT	i
ABS	TRAK	ii
ACK	NOWLEDGEMENTS	iii
ТАВ	LE OF CONTENTS	i
LIST	T OF TABLES	iii
LIST	T OF FIGURES	iv
LIST	T OF ABBREVIATIONS	v
LIST	T OF APPENDICES	vi
CHA 1.1 1.2 1.3 1.4	PTER 1 INTRODUCTION Background Problem Statement Project Objective Scope of Project SITI TEKNIKAL MALAYSIA MELAKA	1 1 2 3
CHA	APTER 2 LITERATURE REVIEW	4
2.1	Introduction	4
2.2	Dearness Sign Longue an	4
2.3	Sign Language	4
2 1	2.3.1 Static and Dynamic Sign Language	07
2.4	2 A 1 Sensory Glove-Based	7
	2.4.2 Computer-Based	8
	2.4.3 Smartphone-Based	9
	2.4.4 Summary	12
CHA	APTER 3 METHODOLOGY	13
3.1	Introduction	13
5.2 2.2	FlowChart	15
3.3	Equipment/Technology	15 15
	3.3.2 Visual Studio Code	1J 16
	3.3.3 Google's Speech-to-Text	10
		± /

	3.3.4	Long-Short Term Memory	18
	3.3.5	TensorFlow	19
	3.3.6	Mediapipe	20
	3.3.7	OpenCV	20
	3.3.8	Data Collection	21
		3.3.8.1 Keypoints	21
		3.3.8.2 Array Folders	22
		3.3.8.3 .npy Files	23
	3.3.9	Summary	23
CHAI	PTER 4	RESULTS AND DISCUSSIONS	24
4.1	Introdu	uction	24
4.2	Result	s and Analysis	24
	4.2.1	Interface	24
	4.2.2	Translator	25
	4.2.3	Analysis	28
4.3	Summ	ary	30
CHAI	PTER 5	CONCLUSION AND RECOMMENDATIONS	31
5.1	Conch	ision	31
5.2	Future	Works	32
5.3	Potent	ial Commercialization	33
REFE	CRENC		35
APPE	NDICH	S AINO	40
		اونيۈمرسيتي تيڪنيڪل مليسيا ملاك	
	L	JNIVERSITI TEKNIKAL MALAYSIA MELAKA	

LIST OF TABLES

TABLE

TITLE

PAGE

10

Table 2.1	Comparisons between the past research technologies
-----------	--



LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 1.1	Scope of project	3
Figure 2.1	Differences between BSL and ASL	5
Figure 3.1	Training Data	13
Figure 3.2	Project Flowchart	14
Figure 3.3	Hand and Face Keypoints	21
Figure 3.4	3 sets of Sign Language Folder	22
Figure 3.5	30 sets of Videos	22
Figure 3.6	30 sets of Frame in 1 Video	23
Figure 4.1	Interface of the Computer Vision	24
Figure 4.2	Interface before Sign Language Translator	25
Figure 4.3	Detection of "Hai"	26
Figure 4.4	ويبوم سيتي تيڪنيڪ "Detection of "Tak	26
Figure 4.5	U Detection of "Terima Kasih" MALAYSIA MELAKA	27
Figure 4.6	Speech-to-Text option interface	27
Figure 4.7	Example of "hello" translation to text	28
Figure 4.8	Graph of Accuracy of the Trained Data	29
Figure 4.9	Graph of Loss accuracy of the Trained Data	29
Figure 4.10	Accuracy of the Trained Data at 2000 Epochs	29
Figure 4.11	Rate of confidence of "system testing"	30

LIST OF ABBREVIATIONS

MSL	-	Malaysian Sign Language
ASL	-	American Sign Language
API	-	Application Programming Interface
BSL	-	British Sign Language
BIM	-	Bahasa Isyarat Malaysia
MFD	-	Malaysian Federation of the Deaf
kNN	-	K-Nearest Neighbour
PCA	-	Principal Component Analysis
HMM-SVM	-	Hidden Markov Model – Support Vector Machine
SLR	-	Sign Language Recognition
IDE	-	Integrated Development Environment
LSTM	-	Long Short-Term Memory
NLP	-	Natural Language Processing
		ALAYSI



ONIVERSITI TERNINAL MALATSIA MELANA

LIST OF APPENDICES

APPENDIX		TITLE	PAGE
Appendix A	PSM 1 Gantt Chart		40
Appendix B	PSM 2 Gantt Chart		41



CHAPTER 1

INTRODUCTION

1.1 Background

Over 5% of the world's population, require rehabilitation to address their 'disabling' hearing loss [1]. Clear communication are keys to better outcomes. However, creating a responsible community wholly to better improve communication between all groups with a disability, especially the deaf, also plays an important role. In Malaysia, only a small number of people can converse in sign language known as the Malaysian Sign Language (MSL), which began in 1954 that started its course with the enrollment of seven students [2]. Miscommunication is a major issue in the deaf and the hearing people since it has been neglected as it is not seen as important and interactions with them are a little to none. The project is intended to aid in communication between deaf and hearing people by using a mobile application. By applying this project, sign language can be directly understood and learned the essence of the sign language itself. The sign language used is focused on MSL which originally had some similarity to that of its origin which is the American Sign Language (ASL) [3]. In order to achieve this, images of hand gestures are fed into the system to further train the system focusing on the MSL. The image can be patterned in the form of classes depending on the context.

1.2 Problem Statement

As the year progresses, communication medium has been fast-growing and accessible, but not so much for the deaf. Engaging with other people, attending interviews,

finding employment, or even asking for help can be much work when it should be done much more effortless. The main issue is when the outside world especially hearing people have a hard time communicating with the deaf [4]. There are several ways to bridge this gap by referring to the internet and books, taking class on sign language, or possibly hiring a translator. All this method has one thing in common: taking time and money. There is no denying that referring to the internet or books can help, but translating a conversation correctly can be time-consuming, even through flipping book pages. Even worse for the internet, the information put up online could be misleading and inaccurate.

1.3 Project Objective

The main objective of this project is to develop a MSL translator interface that will assist in the communication between the deaf and hearing people. Specifically, the objectives are as follows;

- a) To study the Malaysian Sign Language (MSL) and its structure.
- b) To develop a machine learning artificial intelligence model to detect types of gestures of sign language.
- c) To validate the functionality of developed MSL translator through a Computer Vision user interface.

1.4 Scope of Project



Figure 1.1 Scope of project

From the Figure 1.1, under the detection, the laptop's camera is used to detect the signing gesture. Followed by the process, the gesture detected will be processed by the TensorFlow and Mediapipe library for the image classification. Lastly, the translation is displayed through the laptop screen. Similar for the speech-to-text, the microphone from the laptop will detect the voice input from the user and send it to Google's server for processing and return it to the system for the program to display on the MSL translator interface.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

Generally, the scope of this project are as follows:

- a) Focuses on communication between the deaf and hearing people.
- b) Involves image classification that is run through TensorFlow and Mediapipe
- c) Using Computer Vision algorithm for the MSL translator user interface.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter explain the researches of literature related to this project. It contains of some work that already done by other researches or institutes. They have also described several concepts of this project in this chapter. This is because the understanding between theory and work will help much in preparing this project. Some of the past researches includes the usage of hardware and as the year progresses, the system are upgraded to better

portability.

2.2 Deafness



2.3 Sign Language

American Sign Language (ASL) is a fundamentally different language from English. It has its own rules for pronunciation, word creation, and word order, as well as all of the essential properties of language. Languages differ in how they signal distinct functions, such as asking a question rather than delivering a statement. For example, English speakers may pose a question by raising their voices and modifying word order; ASL users raise their eyebrows, broaden their eyes, and lean their bodies forward to ask a question. ASL is a complete, natural language with similar linguistic features to spoken languages but different grammar than English. Hand and face movements are used to express ASL [8].

Just like other languages, there are also other sign languages such as British Sign Language (BSL) and Malaysian Sign Language (MSL) just to name a few. BSL is a spatial and visual language that many beginners mistake for mime. The first crucial thing to understand is that BSL grammar differs significantly from that of everyday English. Despite the fact that both the United Kingdom and the United States of America speak English as their first language, British Sign Language differs from American Sign Language [2]. This indicates that, despite the fact that English is the primary language in various countries, the sign language used varies.

Supposed that ASL and BSL are similar to each other but they are in fact completely different which takes about 30% of similarities between the signage even both languages are of one the same. Compared to ASL, BSL uses two hand to sign while the former uses one [9]. As per below Figure 2.3, the difference between the two sign languages is the used of hand to sign.

British Sign Language



American Sign Language

ABBARD DO DO DO DO ONCO PPROBRESETOUS ax my az de 10203 44 566 17 Ha to Po the B CLANED TEENT OF P

Figure 2.1 Differences between BSL and ASL

As for MSL, although it is based on American Sign Language (ASL), the two languages are considered separate. In turn, Bahasa Isyarat Malaysia (BIM) served as the foundation for Indonesian Sign Language. Malaysian deaf people use MSL, or Bahasa Isyarat Malaysia (BIM), for everyday communication. Despite the fact that MSL has been localised in specific regions or communities, most deaf persons can understand the standard MSL. As a result, the focus of this research is solely on the standard MSL specified by the Malaysian Federation of the Deaf (MFD), a competent body that monitors and maintains Malaysia's MSL [3].

2.3.1 Static and Dynamic Sign Language

In terms of static sign language recognition, studies have explored the use of deep learning techniques such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to classify static sign language gestures from video data. A study by Buckley et al. (2021) used a deep CNN model to developed real-time web-camera based of British Sign Language (BSL) recognition system and achieved an accuracy of 89% [10]. Another study by Joshi et al. (2022) proposed an action recognition sign language certification platform by using LSTM for ASL learning platform and achieved an accuracy of more than 98% [11]. These methods have shown promising results, with some studies achieving high accuracy rates in gesture recognition.

Research on dynamic sign language recognition has also been conducted, with a focus on developing algorithms that can analyze the motion and dynamics of sign language gestures. Methods such as hidden Markov models (HMMs) have been used to model the temporal dynamics of sign language and improve recognition performance. A study by Lynn

et al. (2022) proposed a HMM-based approach to recognize dynamic hand gesture and achieved an accuracy of 92.86% [12].

2.4 Related Past Research

Related past research has shown that the project related to the sign language has been evolving throughout the year from the usage of the hardware and to fully software based. It is clear that communication is a two-way process which in this case the sign language should not only help in healthy people understands, but also for the deaf to understand the conveyed information as well.

2.4.1 Sensory Glove-Based

One of the research is using portable sign language translator by using glove. The glove is integrated with a mobile phone application which uses both left and right hand. The device collects gesture posture with internal special sensors, realizes sign language posture using gloves, broadcasts sign language recognition results via mobile application, and allows non-deaf-mute individuals to accurately understand sign language [13].

Other research intends to integrate the glove with computer system using a technology of a flex sensor. The technology works by having the signer wear a sensor-based hand glove that is connected to a computer that automatically interprets the sign mode and translates it into the appropriate English alphabet or phrase. The sensor data for the signs are recorded by an Arduino and analyzed using the K-nearest neighbor (kNN) machine learning algorithm once the sign gestures have been taken using the glove [14].

Abdullah et. al develop a project where it improves the usefulness and mobility of Malaysian Sign Language (MSL) gestures by optimizing an inertial measurement unit (IMU) sensor-based technique. The optimization itself shows that arranging the sensors in specific location that has a complementary behavior to each other increases the accuracy by more than 98% [15].

2.4.2 Computer-Based

Apart from that, a computer visual-based using camera that acts as the device input has also taken the researches before to another stepping stone since that there are some of the micro-gestures that could not be detected by glove hence, the desired output could not be achieved. This method were designed with multiple attributes including authentication of user, recognition of the sign language, generation of the sign language, and services of remote communication. To be more specific, the system itself used PyQT user interface where once the user has logged into the system, option can be made to choose between sign language recognition and sign language generation. The former will capture the handmovement of the user where the output is English letters along with number from zero until nine that focuses on the ASL and Arabic numerical gestures. Whereas for the latter, detecting the user's voice or text and converting it to English, then again, converting the identified sign language pictures into a video display for the user [16].

Promila Haque et al. created a two-handed Bangla Sign Language Recognition system that recognises 26 different sign motions. The system is divided into three phases: data formation, training, and classification. It makes no use of instrumented wiring devices such as "data gloves" or supplementary hardware and markers. The proposed system can be used by anyone because it does not require any additional components. The system used were Principal Component Analysis (PCA) and K-Nearest Neighbors as their classification algorithm [17].

2.4.3 Smartphone-Based

Smartphone is undeniably an important medium for communication as it is more portable then a computer. Hence, the technology of the sign language recognition has migrated into a better mobility to carry everywhere. According to a research paper by Wang et al. (2018), speech to sign language using animation model were developed for Android phone that focuses on the Chinese Language. It is designed as a support to answer calls. In addition, it improves the interrogative expression of the system by using interrogative pronoun checking and an HMM-SVM model [18].

Another research involved a real-time sign language translation from several Indian languages and English with the use of authentication. The system called 'Sign Translator' that will detect speech as input where it will match the voice with the string and the suitable image connected with the string. Thus, it is then translated to Sign Language where healthy people will be able to convey their thoughts and translated into sign language where deaf individuals will be able to grasp the output and easily understand in their comfortable sign language [19].

"Hand Gesture Recognition System For Translating Indian Sign Language Into Text And Speech" were intended to use sensor-based system that is integrated to an android phone that works as a display. It caters specifically to the Indian Sign Language which converts to English Language and Malayalam Language in the form of text and speech. The gesture will be displayed on the Android phone if it is genuine, otherwise it will be captured again. Arduino UNO is the microcontroller used where on the other hand, HC 05 bluetooth module were utisilized to communicate wirelessly between the microcontroller and the Android phone [20].

Ku et al. makes use of an adaptive ensemble framework to effectively translate sign language from smartphone footage. For efficient execution on smartphones, the framework implements a pose estimation technique. The proposed adaptation method is simple to adopt and delivers excellent results on smartphones. Unlike existing sign language translation technologies, their framework does not rely on special cameras and instead makes use of the standard camera on smartphones [21].

Types of Technology	Approach	Advantage(s)	Disadvantage(s)
Sensory Glove- Based [13] – [15]	 Collects gestures posture using internal special sensors, realizes sign language using gloves, displays sign language recognition via the mobile application [13]. Uses flex sensor hand glove with the sensor data for the sign recorded by an Arduino [14]. Optimization of IMU sensor, reducing number of other sensor [15]. 	 Experiments have proven the system's viability and precision, and it has a wide range of potential applications [13]. Abled to identify few phrases and English alphabet [14]. Abled to achieve higher accuracy; more than 98% [15]. 	• Bulky equipments.

Table 2.1 Comparisons between the past research technologies

Types of Technology	Approach	Advantage(s)	Disadvantage(s)
Computer- Based [16],[17]	 Logged into system, choose between sign language recognition and sign language generation [16]. Uses two-handed Sign Language Recognition (SLR) system on the static gesture [17]. 	 Accuracy at 95.52 for both Arabic numerals and American Sign Language (ASL)[16]. Recognizes sign with acceptable running time and accuracy with any background [17]. 	 Authentication is needed. Data needs to be standardized for better readability.
Smartphone- Based [18]- [21]	 Sign Language Animation [18] Speech to image conversion [19] Gesture to Speech and Text displayed on Android [20] Pose estimation through smartphone footage [21] 	 Good accuracy for interrogative recognition [18]. Does not have to type in conveyed information [19]. Consume low power and high accuracy [20] Three specific words in videos, with a final identification accuracy of up to 91% [21] 	 Needs update from time to time. Not portable. Needs a lot of facial expression. Pose analysis might be insufficient.

In Table 2.2, the three types of technology used in sign language systems are Sensory Glove-Based, Computer-Based, and Smartphone-Based. The Sensory Glove-Based system uses special sensors in gloves to collect gesture posture and displays recognition through a mobile app with high accuracy, but the bulky equipment is a disadvantage. Compared with the Computer-Based system, it uses a two-handed Sign Language Recognition system with acceptable running time and accuracy on any background, but authentication is needed and data standardization can improve readability. As for the Smartphone-Based system, it uses pose estimation through smartphone footage and consumes low power with good accuracy for interrogative recognition, but the system needs updates and might not be portable, and facial expression and pose analysis might be insufficient.

2.4.4 Summary

The literature review concludes the past researches related to this project with its technology used during the project development as well as the upgraded system for better flexibility. As it contains the previous researches, the understanding between theory and the work will assist in developing this project.



CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter provide an in-depth explanation of the system utilised in this project. This chapter is responsible for documenting the project's objectives and ensuring that the project operates well. The next chapter covers the method for advancing the bachelor's degree project and creating a framework for the project design. Additionally, this part of the method includes software explanations of the simulation and the issues encountered along



Figure 3.1 Training Data

From the Figure 3.1, as the gesture of the MSL is made, the system will capture the video frame by frame and made its classification. As it move to the next stage, the classified data are stored into each respective folder defined beforehand. Once the data and folder has been created, the system proceed to read the data images and by using the LSTM, the data images are trained using the LSTM.



Figure 3.2 Project Flowchart

From Figure 3.2, the process of using the MSL translator interface starts with choosing a translator option that are Sign Language Translator and Speech-to-Text. The program will prompt into Sign Language Translator when the user keys in "1". Another window will pop-up on the screen showing the sign language translator. The sign language translator will process the gesture made once the user starts making the sign language gestures. The sign language gesture will be captured and sent into the LSTM system where it compares the input to the MSL signs and translation in the system. The system will then determines the most accurate translation and finally display the translation on the same window. The program will keep on translating until the key button "Q" is pressed.

If the user keys in "2", the program will prompt into the Speech-to-Text option where the program will ask the user to speak. Once the user speaks, the system will transmit the speech to Google's servers where it processes the speech and convert it to text. Once it is done, the text is returned to the program and will be displayed on the screen. The program will keep on asking the user to speak until the close button is clicked.

3.3 Equipment/Technology_KNIKAL_MALAYSIA_MELAKA

ahun

3.3.1 Dynamic Sign Language

There are two main approaches to sign language translator that is the dynamic and static. In this development, the focus is mainly on the dynamic sign language. Dynamic sign language translator, is the process of analyzing continuous streams of sign language gestures and determining the corresponding word or phrase in real-time. This approach involves capturing a video of the signer and then analyzing the position and movement of their hands and fingers over time to identify the specific gesture being made. This can be done using computer vision techniques such as motion tracking, gesture recognition, and deep learning algorithms. In this case, the approach of deep learning specifically on LSTM is used.

3.3.2 Visual Studio Code

Visual Studio Code (VSCode) is a versatile code editor that can be used in conjunction with LSTM deep learning using TensorFlow and MediaPipe.

VSCode can be used to develop deep learning models using LSTM layers in TensorFlow and its high-level API tf.keras. With VSCode, Python code can be written to define and train LSTM models using TensorFlow. It can also run the code and debug it using the built-in terminal and the Python extension. The Python extension also provides features such as code completion, linting, and debugging, which can help to write and troubleshoot the code more efficiently. Additionally, the TensorFlow extension for VSCode to can be use train, evaluate, and debug TensorFlow models.

VSCode also provides support for C++ and can be used to develop multimodal applied machine learning pipelines using MediaPipe. VSCode provides a C/C++ extension that can help to write and troubleshoot C++ code. The extension also to manage the build and run tasks, this is useful when working with MediaPipe, as it helps to build and run the MediaPipe examples and pipelines.

All in all, Visual Studio Code can be used for developing deep learning models using LSTM layers in TensorFlow and multimodal applied machine learning pipelines using MediaPipe, its support for Python, C++ and its extensions make it a great tool for working on deep learning and multimodal applied machine learning projects.

3.3.3 Google's Speech-to-Text

Google's Speech-to-Text technology uses a combination of machine learning algorithms and models to transcribe spoken words into written text. The technology has a wide range of applications, from dictation software to voice-controlled personal assistants, and has become an essential part of modern digital communication.

The process of speech-to-text begins with speech recognition, where the audio input is transformed into a digital signal and analyzed to identify the individual sounds, or phonemes, that make up the speech. This step uses advanced signal processing techniques to filter out background noise and other interference, and to isolate the speech signal for further analysis.

Once the individual sounds have been identified, the system uses a process called acoustic modeling to match the phonemes to a set of predefined models that represent the sounds of a particular language. This step uses machine learning algorithms to analyze the audio and identify the most likely phonemes. The system can be trained on a variety of different languages and dialects, allowing it to transcribe speech in many different languages where in this case is using an English Language.

The next step is language modeling, where the system uses a language model to analyze the context and grammar of the speech, and to determine the most likely word or phrase that corresponds to the sounds. This step is critical to the accuracy of the transcription, as it allows the system to account for variations in pronunciation, accent, and speaking style.

Finally, the system generates a written transcript of the speech, which is then returned as the output of the system. The output can be in the form of a simple text transcript, or it can include additional information such as speaker identification and confidence scores.

3.3.4 Long-Short Term Memory

Long Short-Term Memory (LSTM), is a type of recurrent neural network that is well-suited for dynamic sign language and modeling temporal sequences. It has been widely used in a variety of tasks, including sign language language detection and language translation. The LSTM process input frames of varying lengths and to learn and utilize the temporal structure of sign language image sequences for the recognition purposes [22].

The first step in using LSTM for sign language detection is to gather a dataset of sign language examples. This can be done by recording videos of individuals signing and transcribing the corresponding text. It is important to have a diverse set of signs and signers in the dataset, as this will improve the generalizability of the model.

Next, the dataset must be preprocessed and formatted in a way that is suitable for input into an LSTM model. This involve converting the videos into sequences of image frames, and then extracting features from these frames using techniques such as edge detection or texture analysis. The extracted features should be normalized and standardized to ensure that the model can learn effectively. Once the dataset has been preprocessed, it can be used to train the LSTM model. This involves specifying the architecture of the model, such as the number of layers and the size of the hidden states. The model is then trained using an optimization algorithm.

After training is complete, the model can be evaluated on a separate test set to assess its performance. Common evaluation metrics for sign language detection include accuracy, precision, and recall. If the model's performance is not satisfactory, the architecture of the model or the preprocessing steps may need to be modified and the model re-trained.

18

3.3.5 TensorFlow

Deep learning is a learning technique based on artificial neural networks that is becoming increasingly popular where it goes under the subfield of machine learning. While traditional machine learning algorithms assume classification methods to be either 0 or 1, deep learning can produce numerical results in the range of 0 to 1. Hence, the fundamental benefit of deep learning over typical machine learning approaches is that it performs better in a variety of situations, especially when learning from enormous datasets [23]. As a result, more accurate responses to the current problem can be obtained, and classification systems can attain faster and higher accuracy levels. One of the medium that can be used is known as TensorFlow.

TensorFlow is an open source machine learning platform that runs from start to end. It has a large, flexible ecosystem of tools, libraries, and public resources that allow academics to advance in machine learning and developers to quickly create and deploy machine learning applications. At its foundation, TensorFlow relies on dataflow graphs with mutable state [24]. The dataflow itself is a depiction of the flow of data in any process or system which is image classification.

Image classification is one of the algorithm used under deep learning where it can classify the features retrieved from the original photographs and differentiate between different types of images. A thematic maps would then be produced as the data has been categorized. Image classification is one of the highest accuracy and most stable results. This is because the mapping relationship are describe accurately by the regression neural network. The said mapping are of the image categories, image classification input vectors. In addition, it could differentiate between numerous image classification categories therefore, the ideal image classification results can be produced [25].

By making sure that the data feed into the system is producing optimal result, data augmentation is needed. Data augmentation is included as a data analysis where it is used when overfitted data models can benefit to make them more generalizable. By adding changed copies of previously existing data, it helps in managing and acts as a stabilizer for the overfitting data. In short, data augmentation helps in improving the accuracy models of the output result by creating varieties of the model [26].

3.3.6 Mediapipe

MediaPipe is a framework used for a real-time analysis of video and audio. A sizable number of human body detection and tracking models are available from MediaPipe, and they are all trained using Google's largest and most varied dataset. They serve as the skeleton of nodes, edges, or markers, tracking important places on various body sections [27]. In this case, it is possible for the suggested framework to identify a hand in the input image and to use machine learning to identify important points or landmarks in a hand from a single frame [28].

UNIVERSITI TEKNIKAL MALAYSIA MELAKA 3.3.7 OpenCV

The purpose of OpenCV in a sign language translator is to provide the tools necessary to capture and pre-process the video of sign language gestures. This includes cropping the video to focus on the hands and fingers, converting the video to grayscale, and performing other image processing operations to enhance the quality of the video.

OpenCV also provides a wide range of feature extraction and object detection algorithms that can be used to identify and locate the hands and fingers in the video. These algorithms can be used to extract relevant features from the detected hands and fingers, such as the position and shape of the fingers. Additionally, OpenCV can be used to perform real-time video processing, which is essential for a sign language translator that needs to provide translations in real-time. This allows the model to process the video as it is being recorded, and provide translations in near real-time.

3.3.8 Data Collection

3.3.8.1 Keypoints

part.

Keypoints are created using MediaPipe holistic to draw the landmarks which utilizes the pose of the face and hands. It generates a total of in this case, a 21 hand landmarks per hand. The point of creating the landmarks is to help for holistic tracking in the testing



Figure 3.3 Hand and Face Keypoints

Figure 4.1 shows the screenshot of using OpenCV to open the camera and start detecting the hands and face using the holistics of MediaPipe. All the landmarks data is then saved into an extracted keypoint values and saved into a .npy file format. The format itself preserves every landmarks and keypoints information required to accurately reconstruct the array.

3.3.8.2 Array Folders

The data collected will be saved into 3 sets of sign language folder in the Figure 3.4 which within each folder contains 30 data sets of video files in Figure 3.5. Each video files will also contains 30 sets of frame in length counting from 0 until 29.



Figure 3.4 3 sets of Sign Language Folder



Figure 3.5 30 sets of Videos

Through this part, a data collection is made by again using the openCV function and calling back the keypoints made beforehand in the previous part of creating landmarks and keypoints and recreating the array. It is then captured the movement made according to each sign language in a loop until the condition is satisfied which in this case is a total of 30 datasets of video in each sign, with 30 frame per video. It will save the data in the form of .npy format.



Figure 3.6 30 sets of Frame in 1 Video

3.3.9 Summary

This chapter concludes the proposed the in-depth methodology to develop a new approach in translating Malaysian Sign Language (MSL) and Speech-to-Text. The main focus of the proposed methodology is in accomplishing the objective in such a way that it would be less complicated. The endmost intend of the method is not to obtain many classes of data sets, instead, it is for easy and functionality MSL translator interface.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter presents the results and analysis on the development of a Malaysian Sign Language (MSL) interface translator and Speech-to-Text. The outcome of the project are aligned with what is stated in the objective.

4.2 **Results and Analysis**

AALAYSI

4.2.1 Interface

The interface of the program is shown as Figure 4.1 below. The interface will have a user option to choose from. The options are the Sign Language Translator, and Speech-to-Text. Each of the option has a number to them that is 1 and 2 for the user to enter and make a selection.



Figure 4.1 Interface of the Computer Vision

Once the option is typed as "1", the sign language translator is prompted and from the Figure 4.2, the program will show an interface with a blue line at the top to show the recent prediction when the sign language gesture is predicted.



Figure 4.2 Interface before Sign Language Translator

4.2.2 Translator

The program will then predict accordingly to each of the sign tallied in the array of data collected earlier as of Figure 4.3, Figure 4.4, and Figure 4.5. The gesture made can be predicted accurately as per the figure below, where the three arrays are of "Hai", "Tak", and "Terima Kasih".



Figure 4.3 Detection of "Hai"



Figure 4.4 Detection of "Tak"



Figure 4.5 Detection of "Terima Kasih"

If the option chosen is "2", the user will be prompted to the Speech-to-Text option. The Speech-to-Text will first display "Please speak, Recognizing...". The user then may input their voice and the Speech-to-Text will display the speech in the form of text. The initial Speech-to-Text interface is as below figure.



Figure 4.6 Speech-to-Text option interface



Figure 4.7 Example of "hello" translation to text

From the Figure 4.7 above, an example of speech "hello" is input into the Speechto-Text and will display the translation in the form of text with its confidence rate. Once the speech has been translated into text, it will further prompt the program to ask the user to input the next speech.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

4.2.3 Analysis

Data is of the probability over a 2000 epochs on each of the category of classification that is the "Hai", "Tak", and "Terima Kasih". It is seen that the accuracy has shown an increase of higher accuracy as the number of data increases. The high accuracy indicates a better prediction of the Malaysian Sign Language (MSL) in the MSL translator interface. At about 500 epochs, the training data leads to a better fit of the model to the training data from the Figure 4.8. Similarly the loss of accuracy decreases as the number of epochs increases from the Figure 4.9. The last epoch shows the accuracy of the trained data as of Figure 4.10.



Figure 4.8 Graph of Accuracy of the Trained Data



Figure 4.9 Graph of Loss accuracy of the Trained Data



Figure 4.10 Accuracy of the Trained Data at 2000 Epochs

The rate of confidence of the Speech-to-Text can also be seen to from the interface of the Speech-to-Text. From the Figure 4.11, the rate of confidence shows 95.6% for the example of sentence used that is "system testing".



4.3 Summary UNIVERSITI TEKNIKAL MALAYSIA MELAKA

The goal of a sign language translator is to provide a means of communication for individuals who are deaf or hard of hearing, allowing them to communicate with individuals who do not understand sign language. The results of sign language translator show that it is able to translate the three sign languages. The model is able to translate signs even when the signing is performed at two different lighting conditions. Additionally, the system is able to handle multiple signs at once and provide translations in real-time. It is important to continually monitor the model's performance to ensure its continued accuracy and relevance. Speech-to-text is a technology that converts spoken words into written text. The results of speech to text technology show that it can transcribe and provide real-time transcriptions.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

In conclusion, the development of a Malaysian Sign Language (MSL) translator using Long Short-Term Memory (LSTM), a type of deep learning algorithm, to model the relationships between MSL signs with their corresponding translations and Speech-to-Text to assist in the understanding of Malaysian Sign Language (MSL) is successfully developed along with the Speech-to-Text. It also has the potential to significantly improve the way of receiving information by enabling the system to help in the two way communication between the deaf and hearing people.

However, there is still room for improvement and further research in this area. For example, the current MSL translator may still struggle with ambiguous or complex signs, or may not be able to handle variations in MSL usage among different individuals or communities.

One potential solution is to improve the quality and diversity of the training dataset, which can help the LSTM model learn to handle a wider range of signs and translations. Additionally, there may be opportunities to incorporate additional information, such as context or facial expressions, to assist the translation process.

In the future, it would also be interesting to explore other approaches. This could potentially make the application more user-friendly and accessible to a wider audience. Overall, the development of an MSL translator using LSTM and speech-to-text has the potential to greatly improve the communication and accessibility for both the hearing and deaf individuals who use MSL as their primary means of communication in Malaysia.

5.2 Future Works

One potential future improvement for this MSL Sign Language Translator and Speech to Text Application would be to incorporate natural language processing (NLP) techniques to improve the accuracy and fluency of the translations. NLP is a field of artificial intelligence that focuses on the interaction between computers and human languages, and it can be used to analyze and understand the meaning and structure of human language. By incorporating NLP techniques, a deeper understanding of the context can improve the analysis and mining of the data. NLP allows for this by using the communication structures and patterns of humans [29].

Apart from that, A mobile application can also be implemented to bring several benefits to the translator. One of the main benefits of a mobile application is convenience. With a mobile application, users can access the sign language translator at any time and place. This means that users can use the translator whenever they need to communicate with others, even when they are on the go. This is particularly useful for users who need to communicate in real-time, such as when they are out in public or in a meeting. Additionally, a mobile application can be designed to work offline, which means it can still provide translations even when there is no internet connection. This is particularly useful for users in areas with poor internet connectivity, as they can still use the translator even when they are offline.

Another potential improvement would be to incorporate multi-modal learning, which involves training the model on multiple types of input data, such as face emotion, text, and audio data. Sometimes, it is not clear what is the conciseness of ones gesture to give certain meanings if it only relies on one dimension only. The use of multi-modal data for deep machine learning has shown promise when compared to uni-modal approaches with fusion of multi-modal features resulting in improved performance in several applications [30]. This could allow the model to make use of additional information, such as the emotion, tone, and inflection of speech to improve its translations. The Speech-To-Text can be improved by incorporating a diverse language library such as Alexa Library, and Siri Library.

Other than that, the MSL sign language translator could also implement a cloudbase deployment to allow for easy scaling of the system to accommodate more users and more data. This means that the system can handle an increased number of requests and provide a better service for users. It can also be an easier maintenance and updates of the system where the providers typically handle the maintenance and updates.

Finally, it would be useful to develop an app with other assistive technologies such as augmented reality to further improve accessibility for deaf and hard of hearing individuals.

5.3 Potential Commercialization

One potential commercialization opportunity is mobile applications. A mobile application version of the translator can be developed and made available for download on app stores such as Google Play Store or Apple App Store. The application can be free or offered as a paid version with additional features. This can make the translator more convenient for users, as they can access it from anywhere and at any time.

Another potential commercialization opportunity is educational institutions. The translator can be used in educational institutions to help students who are deaf or hard of hearing to better understand lectures and class discussions. It can also be used in special education classes. This can help to improve the education experience for deaf students and make it more inclusive for them.

Other than that, businesses is also a great opportunity for the translator. It can be used in businesses to help employees who are deaf or hard of hearing to better communicate with their colleagues and clients. It can also be used in customer service centers to help deaf customers. This can improve the efficiency and accessibility of businesses and help to improve customer service.

Government organizations could also be a key area where the translator can be used. It can be used in government organizations such as hospitals, courts, and government offices to help deaf individuals better access services. This can help to improve the accessibility of government services for the deaf community.

Last but not least, video conferencing is an opportunity for the translator where it can be integrated into video conferencing software, such as Zoom or Skype, to help deaf participants in online meetings or classes. This can make online meetings and classes more inclusive for the deaf community.



REFERENCES

- [1] 'Deafness and hearing loss', World Health Organization, Apr. 01, 2021.
- [2] V. Y. Chong, 'Development of Malaysian Sign Language in Malaysia', JSNE Volume 8, vol. 8, no. 8, pp. 15–24, 2018.
- [3] A. L. S. Lien and L. K. Yin, 'Gesture Recognition-Malaysian Sign Language Recognition using Convolutional Neural Network', Malaysia, 2020.
- [4] P. Agrawal, 'Design Issue and Proposed Implementation of Communication Aid for Deaf & Dumb People', International Journal on Recent and Innovation Trends in Computing and Communication, vol. 3, no. 5, May 2021.
- [5] A. Felman, 'What's to know about deafness and hearing loss?', *Medical News Today*, Jun. 27, 2018.
- [6] 'Hearing Loss: How It Affects Communication', Cincinnati Children's, Apr. 2019.
- [7] A. Parthasarathy, K. E. Hancock, K. Bennett, V. DeGruttola, and D. B. Polley, 'Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing', *Elife*, vol. 9, Jan. 2020, doi: 10.7554/eLife.51419.
- [8] 'American Sign Language', NIH Publication, Oct. 29, 2019.
- [9] 'American Sign Language and British Sign Language: How Are They Different?', *Akorbi*, Oct. 13, 2021.
- [10] N. Buckley, L. Sherrett, and E. Lindo Secco, 'A CNN sign language recognition system with single & amp; double-handed gestures', in 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Jul. 2021, pp. 1250–1253. doi: 10.1109/COMPSAC51774.2021.00173.

- [11] J. Joshi, P. Gandhi, and R. Sawant, 'Sign Language Certification Platform with Action Recognition using LSTM Neural Networks', in 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), Jun. 2022, pp. 1–6. doi: 10.1109/IC3SIS54991.2022.9885321.
- K. Y. Lynn and F. Wong, 'Recognition of Dynamic Hand Gesture using Hidden Markov Model', in 2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST), Oct. 2022, pp. 419–422. doi: 10.1109/GECOST55694.2022.10010517.
- [13] G. Liqing, L. Wenwen, S. Yong, W. Yanyan, and L. Guoming, 'Research on Portable Sign Language Translation System Based on Embedded System', in 2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE), Dec. 2018. doi: 10.1109/ICSCSE.2018.00136.
- [14] H. S. Anupama, B. A. Usha, S. Madhushankar, V. Vivek, and Y. Kulkarni,
 'Automated Sign Language Interpreter Using Data Gloves', in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Mar. 2021. doi: 10.1109/ICAIS50930.2021.9395749.
- [15] A. Abdullah, N. A. Abdul-Kadir, and F. K. Che Harun, 'An Optimization of IMU Sensors-Based Approach for Malaysian Sign Language Recognition', in 2020 6th International Conference on Computing Engineering and Design (ICCED), Oct.
 2020, pp. 1–4. doi: 10.1109/ICCED51276.2020.9415838.
- W. Li, H. Pu, and R. Wang, 'Sign Language Recognition Based on Computer Vision', in 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Jun. 2021, pp. 919–922. doi: 10.1109/ICAICA52286.2021.9498024.

- P. Haque, B. Das, and N. N. Kaspy, 'Two-Handed Bangla Sign Language
 Recognition Using Principal Component Analysis (PCA) And KNN Algorithm', in
 2019 International Conference on Electrical, Computer and Communication
 Engineering (ECCE), Feb. 2019, pp. 1–4. doi: 10.1109/ECACE.2019.8679185.
- [18] W. Wang, W. Zhang, and B. Sun, 'A sign language smart phone based on realistic animation', *MATEC Web of Conferences*, vol. 189, Aug. 2018, doi: 10.1051/matecconf/201818903024.
- [19] L. Priya, A. Sathya, and S. K. S. Raja, 'Indian and English Language to Sign Language Translator- an Automated Portable Two Way Communicator for Bridging Normal and Deprived Ones', in 2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Dec. 2020, pp. 1–6. doi: 10.1109/ICPECTS49113.2020.9336983.
- [20] J. Kunjumon and R. K. Megalingam, 'Hand Gesture Recognition System For Translating Indian Sign Language Into Text And Speech', in 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Nov. 2019, pp. 14–18. doi: 10.1109/ICSSIT46314.2019.8987762.
- [21] Y.-J. Ku, M.-J. Chen, and C.-T. King, 'A Virtual Sign Language Translator on Smartphones', in 2019 Seventh International Symposium on Computing and Networking Workshops (CANDARW), Nov. 2019, pp. 445–449. doi: 10.1109/CANDARW.2019.00084.
- [22] S. He, 'Research of a Sign Language Translation System Based on Deep Learning', in 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), Oct. 2019, pp. 392–396. doi: 10.1109/AIAM48774.2019.00083.

- [23] I. H. Sarker, 'Machine Learning: Algorithms, Real-World Applications and Research Directions', *SN Comput Sci*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.
- [24] M. Abadi, M. Isard, and D. G. Murray, 'A computational model for TensorFlow: an introduction', in *Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Languages*, Jun. 2017, pp. 1–7. doi: 10.1145/3088525.3088527.
- [25] Z. Wang, C. Tang, X. Sima, and L. Zhang, 'Research on Application of Deep Learning Algorithm in Image Classification', in 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Apr. 2021, pp. 1122– 1125. doi: 10.1109/IPEC51340.2021.9421185.
- [26] P. Soni, 'Data augmentation: Techniques, Benefits and Applications', *analyticsteps*, Jan. 09, 2022.
- [27] A. Halder and Tayade Akshit, 'Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning', *International Journal of Research Publication and Reviews*, vol. 2, no. 5, pp. 9–17, 2021.
- [28] S. Salim, M. M. A. Jamil, R. Ambar, R. Roslan, and M. G. Kamardan, 'Sign Language Digit Detection with MediaPipe and Machine Learning Algorithm', in 2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE), Oct. 2022, pp. 180–184. doi: 10.1109/ICCSCE54767.2022.9935659.
- [29] A. Torfi, R. A. Shirvani, Y. Keneshloo, N. Tavaf, and E. A. Fox, 'Natural Language Processing Advancements By Deep Learning: A Survey', Mar. 2020.

[30] D. Priyasad, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, 'Memory based fusion for multi-modal deep learning', *Information Fusion*, vol. 67, pp. 136–146, Mar. 2021, doi: 10.1016/j.inffus.2020.10.005.



APPENDICES

Appendix A PSM 1 Gantt Chart

ALAYS,	PR	OJECT	' PLAN	INING	PSM 1	l										
Mine and							2	022								
No.	10	MAF	RCH			API	RIL			MA	Y			JU	NE	
	1	2	_3	4	5	6	7	8	9	10	11	12	13	14	15	16
PSM1	NY.															
Proposed Project	Þ					1										
• Decide project title (Change in the 4 th week)						1			V.							
Meeting with supervisor, discussion on project content		1							K							
 Identify objective, problem statement & scope of project. 	G			/					BREA							
Report Writing	Z								R							
Chapter 1 writing	EF								E .							
Chapter 2 writing	2	-		- (10 A			S							
Chapter 3 writing	- E					2		No. 1	2	9						
Project Research:	SIV)	100	6	SE							
Finding related past research	Ч								Á							
Finding software to use INUMEDRITIE	TEV	CMI	CAL	M	AL /	NVS	1A	ME	M	CΔ	-					
Data acquisition					"L Basel		1.0			1.0						
Project deliverable (PSM1):																
Completing report until Chapter 5																
Submission PSM1 report																
Presentation PSM1																

PROJECT PLANNING PSM 2																	
	2022/2023																
		OCT	OBER		NOVEMBER					DECEMBER				JANUARY			
	1	2	3	4	5	6 7		8	9	10	11	12	13	14	15	16	
PSM2																	
Project Development	1																
• Meeting with supervisor, discussion on project	82																
content	2																
• Development of the project (Data collection,	>						-										
Data Training, Data Testing, Software							-										
Development, Software Testing)	Ċ																
Report Writing	Ž						1										
Chapter 3 alteration	Ξ		1														
Chapter 4 writing	RI																
Chapter 5 alteration	1 B																
Project Research:	S	1		1													
Finding related past research			Š.	-	N	3.7		U	â	0 6							
Finding software alternatives	0		1.0			C	2	0	10	1							
Project deliverable (PSM2):																	
Completing all chapters in report	IC M	MIL	(AL	BA/	1.1	ve			A	V N							
Submission PSM2 report	En	IAIL	CAL.	. WIP		0	IA I		-	-CP							
Presentation PSM2																	