SPEECH REHABILITATION FOR DISORDER PEOPLE

NUR SYAHMINA BINTI AHMAD AZHAR

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

SPEECH REHABILITATION FOR DISORDER PEOPLE

NUR SYAHMINA BINTI AHMAD AZHAR

This report is submitted in partial fulfilment of the requirements for the degree of
Bachelor of Electronic Engineering with Honours

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

2022

DECLARATION

I declare that "Speech Rehabilitation for Disorder People" is the result of my own work except for quotes as cited in the references.

Signature : .....

Author : NUR SYAHMINA BINTI AHMAD AZHAR

Date : 11/1/2022

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Bachelor of Electronic Engineering with Honours.

Signature : .....

Supervisor Name : NIK MOHD ZARIFIE HASHIM

Date : JANUARY 11, 2022

DEDICATION

I humbly dedicated this thesis to be submitted in partial fulfillment of for the degree of Bachelor of Electronic Engineering with Honours. The project aim's is to develop a network that can detect vowels of stroke patients. Hence, the content of this study is to provide a *Malay* dataset as the process for building the neural network.

This thesis is dedicated to my supervisor, Ir. Ts. Nik Mohd Zarifie Bin Hashim for his continued support and counsel during the time I developed the system. The dedication also for my parents, who have never failed to give me financial support, and moral support which have made possible the success of this project. I dedicate this project to *Pusat Rehabilitasi Perkeso Melaka* staff, who work so hard to provide the best hospitality while completing the recording session.

ABSTRACT

Communication is one of the crucial elements in life. It helps humans make decisions, solve problems, share ideas, and develop relationships. Effective communication will help people in many aspects, especially in the medical field. However, this effective communication is a limitation for some people who have problems communicating, like stroke patients, children with speech delay, Aphasia patients, and more. For gaining their capability in speech, rehabilitation is one method to retrain back this lost capability. Here, the major problems for the speech capability among the patient have commonly affected patients' understanding of speeches and have disorganized speech. Although manual rehabilitation is the most common way to train them, there are still problems of housekeeping and human power in ensuring the training and rehabilitation activities are on track. Lack of human strength and the effect of pandemics infers the delay time of patient rehabilitation programs. This project showed an outperform vowel recognition result via Convolution Neural Network (CNN) via vowel sound dataset. Compared to the comparative and the baseline method, VGG16, we manage to recognize the vowel using not directly sound but from an image file. Our proposed work could contribute to the medical and rehabilitation field helping the disorder people by gaining this outperform accuracy. At the end of this project, a network model, which will be utilized as Malay Language vowel detection with good performance accuracy, can be achieved.

ABSTRAK

Komunikasi adalah salah satu elemen penting dalam kehidupan. Ia membantu manusia membuat keputusan, menyelesaikan masalah, berkongsi idea, dan membina hubungan. Komunikasi yang berkesan akan membantu manusia dalam pelbagai aspek terutamanya dalam bidang perubatan. Walau bagaimanapun, komunikasi berkesan ini terhad untuk sesetengah orang yang mempunyai masalah berkomunikasi, seperti pesakit strok, kanak-kanak yang mengalami kelewatan pertuturan, pesakit Aphasia dan banyak lagi. Untuk mendapatkan keupayaan mereka dalam pertuturan, pemulihan adalah satu kaedah untuk melatih semula keupayaan yang hilang ini. Di sini, masalah utama untuk keupayaan pertuturan di kalangan pesakit lazimnya menjejaskan pemahaman pesakit tentang pertuturan dan pertuturan yang tidak teratur. Walaupun pemulihan manual adalah cara yang paling biasa untuk melatih mereka, masih terdapat masalah pengemasan dan tenaga manusia dalam memastikan latihan dan aktiviti pemulihan berada di landasan yang betul. Kekurangan kekuatan manusia dan kesan pandemik menyimpulkan masa kelewatan program pemulihan pesakit. Projek ini menunjukkan hasil pengecaman vokal yang lebih baik melalui 'Convolutional neural Network' (CNN) melalui data set berbunyi vokal. Berbanding kaedah perbandingan dan VGG16, kami berjaya mengecam vokal menggunakan bukan bunyi secara langsung tetapi daripada fail imej. Kerja yang dicadangkan oleh kami boleh menyumbang kepada bidang perubatan dan pemulihan membantu orang yang mengalami gangguan dengan memperoleh ketepatan prestasi ini. Pada akhir projek ini, model rangkaian,

yang akan digunakan sebagai pengesanan vokal Bahasa Melayu dengan ketepatan prestasi yang baik, boleh dicapai.

ACKNOWLEDGEMENTS

I want to express my gratitude and admiration to my supervisor, Ir. Ts. Nik Mohd Zarifie Bin Hashim for the countless ways by his wise counsel, thoughtful criticisms, and patient support for me to learn new knowledge and give the proper way to solve the problem at hand. Thanks to the final year project coordinator, Dr. Mas Haslinda Binti Mohamad, for helping me in many aspects, especially writing a great and good thesis. While preparing this report, I had to go through too many trials and challenges. However, there were valuable lessons and experiences as the hard work finally paid off when this report was completed perfectly and successfully.

Apart from that, a word of appreciation to all the lecturers of Universiti Teknikal Malaysia Melaka (UTeM) for giving me guidance on this study. This speech is also addressed to both of my parents as they are very supportive and helpful to me financially, spiritually, and morally. Not forgetting my classmates as well and for those, directly and indirectly, involved in helping me complete this final report successfully.

Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES


UNIVERSITI TEKNIKAL MALAYSIA MELAKA
UTeM
اونيۏرسيتي تيکنيکل مليسيا ملاك
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

CHAPTER 1

INTRODUCTION

1.1 Introduction

This section focuses on the history and issue statement of the project, followed by the project objective and scope. Finally, we shall discuss the recommended solution.

1.2 Background

In medical terms, a disorder is a disturbance of the normal functioning of the mind or body. Genetic factors, disease or trauma may cause the disorder including mental illness or mental health disorder. Disorder patients have health conditions that affect their mood, thinking, and behaviour. Examples of mental illness include depression, anxiety disorders, schizophrenia, eating disorders, and addictive behaviours. Stroke can also cause communication problems as this illness damages the parts of the brain that responsible for language. A stroke can affect people in speaking, reading, writing, and understanding speeches.

Some of the disorders people like Aphasia have communication problems in real life. Aphasia is a complex language and communication disorder resulting from damage to the language centres of the brain. This damage may be caused by stroke, a head injury, a brain tumour, and any neurological illness [4]. Around a third of people who have a stroke will experience Aphasia. This deficiency will make them unable to express their will and needed as they have difficulty communicating.

Next, Dyspraxia is a brain-based motor disorder and neurological disorder that impacts an individual's ability to plan, and it affects movement and coordination. According to the Children's Hospital at Westmead, Australia, children with Dyspraxia disorder are always hard to make sounds, repeat sequences of sounds or words, and have difficulty sustaining normal intonation patterns. So, the purpose of this project is to develop an algorithm to detect vowels for disordered people [5].

All these conditions significantly impede communication, such as disorganized speech. The range of communication of a disordered patient is slightly different compared to normal people. It can affect their communication in many ways, and usually, all practicing physicians have learned about communicating effectively with these patients [6].

However, speech rehabilitation for disorder person project will help them communicate as this project will focus on vowel recognition by collecting sound data from disorder patients and normal people. It will evaluate and test the recorded data by comparing the sound recording of disorder people to normal people using Neural Network.

1.3 Problem Statement

Disorder people especially stroke patient have problems in communicate. This condition will make them unable to express their will, help or needs and this communication problems is a big problems to their family or friends. In addition, there is no 'malay' data set available. From the source from Wikipedia of the list of data set for machine-learning research, there are data set for Japanese Vowels Data set, Arabic Speech Corpus, Spoken Arabic Digits and Parkinson Speech Dataset. So, in order to design a network for vowel's recognition for disorder people in Malaysia, the important things that need to have before design a network is create for 'malay' vowels data set.

1.4 Objective

The objective of this project is:

- To create vowel's data set of normal person and disorder people.
- To design a neural network system that can help disorder patients to communicate.

1.5 Scope of the Project

This project will be focused on vowel recognition for disordered people. The data collection of sound/voice recording for this project is within two group disorder patient, and normal people. The project recording sounds, voices and next, it will be stored in the data collection. Then, it will evaluate and test the recorded data by comparing the sound recording of disorder people to normal people by using the proposed network model, Neural Network. Journals and paper applicable to the project is used as references. Software is used to analyze the output, and later on, graphs are plotted to show the signal strength based on the simulated environment.

1.6 Thesis Outline

- **Chapter 1** : This chapter focuses on this project's background problem statement, followed by the objective and scope identified. Finally, the proposed solution will be discussed.

- **Chapter 2** : This chapter was related to the literature review's reading material and precious work discussed about the speech rehabilitation and some related research. Besides that, this chapter discusses about the related works and methods that have been proposed before. The information obtained from the literature review will be used to assist improved the implementation of the next process.

- **Chapter 3** : This chapter discusses how to analyse the development of the neural network of vowel's recognition for stroke patient and normal person based on chapter 2 literature review and what method to do to get a better accuracy and performance.

CHAPTER 2

LITERATURE REVIEW

This chapter discusses the background study, literature of several related journals, articles, and previous studies on the speech rehabilitation and the progress of speech rehabilitation technology for disorder people.

2.1 History of the Disorder

Since time immemorial, some unfortunate people have various deficiencies exist in our society. As a person with concerning humanity, it has been our responsibilities to protect them and give them the best care. To build up a society with a huge concern, we need to put aside all the lousy nature as the basic emotional responses of disable people are anxiety, shame, and hostility [7] towards those attitudes. All the mistreatment will lead them to feel rejected and it must be avoided at all cost as this will affect their self-healing and treatment.

Figure 2.1: Stroke Rehabilitation [1]

Some disabilities are structural, while others are internal, which are not as apparent or visible as others [1]. There are various type of disorder classification. In the medical field, there are personality disorder, mental disorder, psychological disorder, speech disorder and more kind of disorder involved. For this project, the focus of classification is on speech disorders.

2.1.1 Aphasia and Related Disorder

When a person suddenly loses the easy use of words, it must be a devastating experience for them to handle [7]. The term Aphasia refers to a family of clinically diverse disorders that affect the ability to communicate by oral or written language, or both [7]. The causes of Aphasia are brain-damaged typically occurs after a stroke or a head injury. Most Aphasic patients have difficulties in expressing, communicating, and formulating. The incidence and prevalence of Aphasia is unknown to be exact, but some of the studies and research show that the estimated incidence and prevalence of Aphasia is stroke but it have contain of different methodologies and criteria to determine the presence of Aphasia in people with stroke [8].

In medical terms, there are two types of Aphasia. An aphasic condition can be fluent and non-fluent. A non-fluent Aphasia has a lack of articulatory precision, prosody in speech and they like hesitant and slow with many pauses while communicate [1]. Mostly the causes of Aphasia is from stroke and brain damaged. A stroke occurs when the blood supply to the brain is interrupted or reduced. It will prevent the brain tissue from getting enough oxygen and nutrients. The symptoms of someone having a stroke are headache, trouble walking and problems seeing with in one or both eyes. In addition, the signs of stroke are like having some trouble in speaking and understanding what others try to saying and it can make someone paralysis and having numbness on the face, arm and leg [9].

2.2 History of Speech Rehabilitation

In medical terms, speech rehabilitation is a service that focused on communication problem which mainly occurred in disorder people who have lost the ability to communicate normally. Nowadays, many therapy centers and rehab provide the rehabilitation's services to overcome the communication's deficiency. As technology evolves, there are a lot of research and techniques involved in the speech therapy and rehabilitation field. Speech rehabilitation for disorder people project focused on vowel's recognition of disorder people to communicate better in their life [2].

Figure 2.2: Speech Rehabilitation [2]

Figure 2.2 shows the speech rehabilitation conducted with children who have communication difficulties. For this project, the speech rehabilitation focused on stroke patient who have lost their ability to speak. This project will detect the vowel's of stroke patient in order to help them to communicate.

2.2.1 Early Method of Speech Rehabilitation

2.2.1.1 Traditional Articulation Therapy Approach

The Traditional Articulation Therapy Approach is a traditional articulation approach focused on students with articulation disorders. Articulation disorder is a speech disorder involving difficulties in articulating specific types of sounds, the substitution of sound for another, slurring of speech, or indistinct speech. This approach focused on the phonetic placement of the sound in error and teaching the motor skills to produce the sound correctly. This intervention method uses a hierarchy to help children establish the correct sound and learn about the motor movements to use the sound in conversational contexts. At the end, this approach aims to establish the target sound and help the student be able to correctly produce the sound in conversation with a variety of speaking partners across many different settings [7].

2.2.1.2 Speech Production Accuracy for Individual Sounds

The second research is about the severity of speech disorders. The severity of speech disorder can range from mild to quite severe and a complete inability to speak. Mild to moderate speech disorders like speaking with a lisp, deleting or distorting the final sounds in words and consonant clusters, and substituting one sound e.g, 'w' for 'r'. The disorders are treated with a focus on speech production accuracy for individual sounds. In more severe speech disorders, the focus of intervention may be on improving global speech intelligibility [10]. This traditional approach teaches them how to pronounce accurately every individual sound. This is one of early method in speech rehabilitation method conducted by the professional in this field to fix the severity in speech disorder.

2.2.2 Classification before Deep Learning

2.2.2.1 Articulation Drill and Motor Learning

Articulation drill approaches focused on motor placement and production of individual speech sounds. The traditional methods taught directly how to move and coordinate the articulatory mechanism for producing individual speech sound. One key component of articulation drills is repeated motor practice of the tongue movements and coordination of the articulators, such as lips and jaw which required accurate pronunciation. In Phonological/Lexical Approaches' interventions are designed to improve word-level production rather than starting at the motor learning or individual phoneme level.

Based on the Speech Correction by Van Riper in his book, there is connection between motor skills and speech. Table 2.1 shows the studies of children that show a parallel course of development between key motor skills and the acquisition of speech [7]. The delay in the acquisition of motor skills can affect slow speech development, especially among children. So, the articulation drill and the motor learning method focused on motor practice of the tongue movement and coordination of the other articulators such as lips and jaw, and the result of this proposed method is to make the people who have a deficiency in communicating have accurate pronunciation [10].

Table 2.1: A parallel course of development between key motor skills and the acquisition of speech

Age	Motor Skill	Speech
6 months	Sits alone	Pre-speech : Babbling
12 months	Stands and takes first step	First word
18 to 22 months	Walks alone	Two-word phrases

2.2.2.2 Phonological/Lexical Interventions

Besides of Articulation Drill and Motor Learning approach, there is a method named Phonological/Lexical interventions, which are applied to words and phrases. This method considering speech sound within the context of word production. These interventions are designed to improve word-level production [10]. This method are different compared to the Articulation Drill and Motor Learning methods which focused more on motor placement and production of individual speech sound.

2.2.3 Recent Method

2.2.3.1 Machine Learning Algorithm to Predict the Persistence and Severity of Major Depressive Disorder from Baseline Self-Reports

One related project about disorder people is about testing a machine learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. This project focused on Heterogeneity of major disorder (MDD) illness. Machine learning (ML) models developed from self-reports about incident occurs among MDD's respondents. ML model prediction accuracy was also compared to the conventional logistic regression model. The materials and methods used in this project are sample/survey and the baseline assessment of DSM-III-R disorders. ML methods improve on conventional methods, and these results confirm that clinically useful MDD risk-stratification models can be generated from baseline patient self-reports [11]. Although the methods are not generally involved in speech rehabilitation, the method of using machine learning can be as guidance while performing for this project.

2.3 Deep Learning

In general, transfer learning is some of the convolution neural network designed to give the state-of-art algorithm to classify different images. There are various transfer learning concepts and various algorithms that have come up with better accuracy with respect to this image rate classification. Some of the algorithms that are designed for image classification are like 'Xception', 'VGG16', 'VGG19', 'Resnet', 'ResnetV2', 'ResNeXt', 'MobileNet', 'MobileNetV2', 'InceptionV3', 'InceptionResNetV2', 'DenseNet' and 'NASNet'.

These all are the various image classification models that have come up with better accuracy when classifying with thousand or different categories of images. For this project, the transfer learning approach model is VGG16. The model VGG is like Convolutional Neural Network (CNN) architectures which operate on spectrograms at lower layers and use many convolution layers and fully connected layers as we go upper layers.

2.3.0.1 Convolutional Neural Network

This articles about the comparison between Convolutional Neural Network (CNN), Recurrent neural network (RNN), and Recurrent Multilayer Perceptron (RMLP) like shown in 2.3.

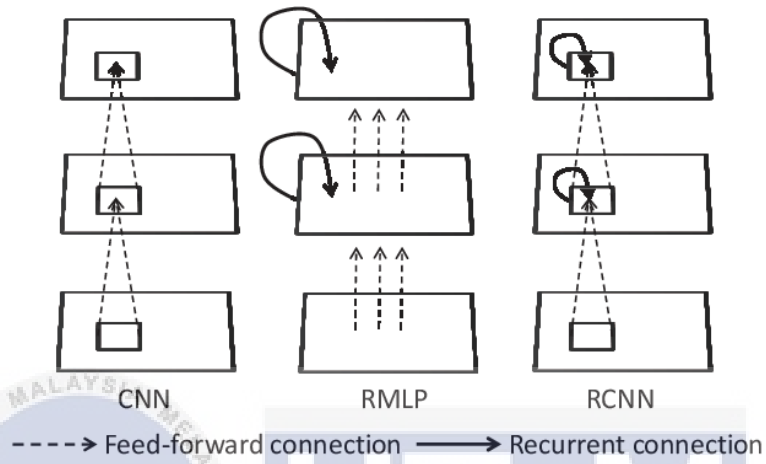

Figure 2.3: Illustration of the architectures of CNN, RMLP and RCNN [7]

The advantage of CNN compared to another algorithm is CNN can detect the important features without any human supervision. CNN is useful in a lot of application especially in image related tasks including the image classification, image semantic segmentation, object detection in images, and more. A CNN usually takes an order 3 tensor as its input and the input then sequentially goes through a series of processing. One processing step is usually called a layer, which could be a convolution layer, a pooling layer, a normalization layer, a fully connected layer, a loss layer, and more [12].

CNN are primarily used in the field of pattern recognition within images. The basic functionality of the example CNN above can be broken down into four key areas which are the input layer that will hold the pixel values of the image. Then, the rectified linear unit (ReLu) aims to apply an 'elementwise' activation function such as sigmoid to the output of the activation produced by the previous layer. The third layer is the pooling layer and lastly is fully-connected layers which will then perform the same duties found in standard ANN [13] like the figure 2.4 shown.

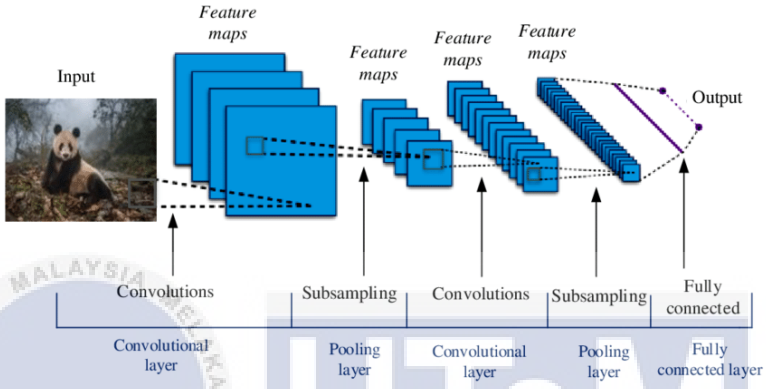

Figure 2.4: A simple CNN architecture [8]

2.3.1 Related Work of Convolutional Neural Network

2.3.1.1 Fruit Classification based on Convolution Neural Network

In the Development of Control System for Fruit Classification based on Convolution Neural Network project published in 2018, this project used Convolution Neural Network (CNN) to develop a fruits detection and recognition based on CNN. The project's accuracy is closed to 94 percent for 30 classes of 971 images. The data set consists of 971 images, classified into 30 different fruit classes and every fruit class contains about 32 different images.

Next, the implementation of CNN for fruit classification is applied and there is a control system design for vision based automated decision-making system to develop computer vision based control system [14]. This project is about to develop a control system by using 'Alexnet'. The implementation by using graphic processing unit in Matlab to perform the classification and simulation process. At the end, the accuracy result of the project is closed to 94 percent [14]. This research paper helps to indicate the implementation of CNN in speech rehabilitation for disorder people project as this project also implement CNN in order to evaluate the accuracy in normal and disorder people .

2.3.1.2 Convolution Neural Network (CNN) in Fruit Image Processing

This research paper is about a review of Convolution Neural Network (CNN) applied to fruit image processing. There are Convolution Neural Network (CNN) based approaches for fruit classification tasks contained in this research paper which includes various of data set. The summary of state-of-the-art CNN-based approaches applied for fruit classification tasks include data set, data type, CNN model, and the performance results. The summary of state-of-the-art CNN-based approaches applied for fruit quality control tasks can be viewed in the types of fruit, data type, CNN model, and performance results [15].

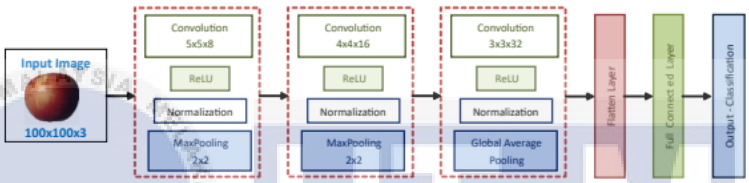

Figure 2.5: Designed CNN architecture for fruit classification [10]

Summary of state-of-the-art CNN-based approaches applied to the detection of fruits for automatic harvest. The discussion on the review of CNN-Based Approaches for Fruit Image Processing can be observe that in 70 percent. It is good to review this research paper as this site have various of data set in every aspect, and category of fruit, the data type, the CNN model they used to implement the project. Lastly, the performance result is showed in this research paper, and we can observed the accuracy in every aspect [15].

CHAPTER 3

METHODOLOGY

This chapter presents implementation to achieve this project's goals. The chapter will present a flow chart to illustrate and clarify when completing the project.

3.1 Research Methodology Flow Chart

The flow chart in figure 3.1 shows the beginning of the process initiated with the project research based on the reading from published journals and other reading resources related based on the literature review to the case study on speech rehabilitation and deep learning method to disorder people. The reading helps to introduce the concept and operation of vowel's recognition for speech rehabilitation. Afterwards, the process continued with the development of the project based on research and conclusions on the project and recommendations for the future works to be suggested would be carried out in the last stage.

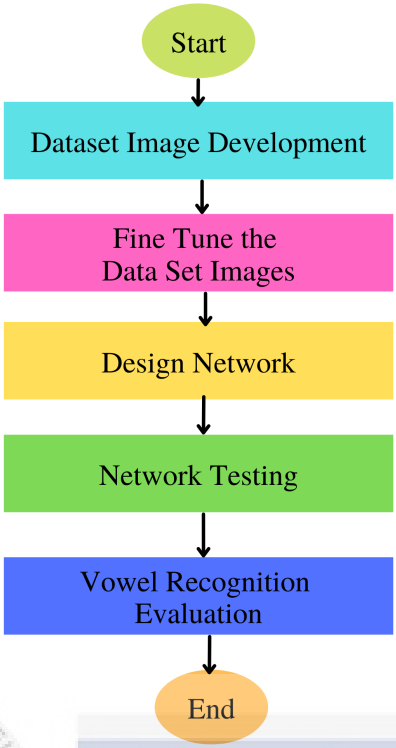

Figure 3.1: Project Flow Chart

Figure 3.1 consists of 3 main phases in completing the project. The first phase is data set image development, the second phase is process of fine tune the data set images and the final phase will be process of designing the neural network from the data set collected. In the beginning, this project start with literature review research regarding the speech rehabilitation and many journals and articles examined to gain a better understanding of this project.

3.1.1 Data set Image Development

In the first phase, the data set image development is conducted. Speech rehabilitation for disorder people project contains two type of data set which are data set for normal people and data set for disorder people. For disorder people, the data set collected is from stroke patient and the collection of data set from the both category contains vowel's 'a', 'e', 'E', 'i', 'o', and 'u'.

Figure 3.2: Vowel's recording for normal person

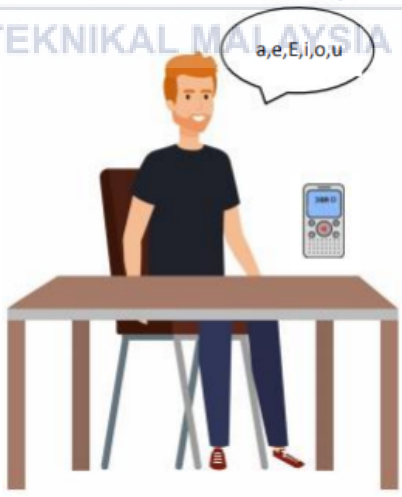

Figure 3.3: Vowel's recording for disorder people

From the Figure 3.2 and Figure 3.3, the process of data set image development is conducted. The target for this project is vowels sound and it will be recorded in wav. file. In the development of data set image, the data set will be collected from two group which is disorder people and normal people and it will be split into two directory. For the time being, the focused of speech rehabilitation in this project is vowel's recognition of disorder people which is from stroke patient. There is no 'malay' language vowel data set was made until now. So, this project will develop a new data set of vowel's recognition of stroke patient.

3.1.1.1 Method of Vowel's Recording

There are three section of every vowel's recording. The table 3.1, table 3.2, table 3.3, table 3.4, table 3.5 and table 3.6 shows the method used for this project during the data set development.

Table 3.1: The 'a' sound for vowel's recording

No.	Type of Vowel	Section	Quantity
1.	'a' sound	Short	30 sound
2.	'a' sound	Moderate	30 sound
3.	'a' sound	Long	30 sound

Table 3.2: The 'e' sound for vowel's recording

No.	Type of Vowel	Section	Quantity
1.	'e' sound	Short	30 sound
2.	'e' sound	Moderate	30 sound
3.	'e' sound	Long	30 sound

Table 3.3: The 'E' sound for vowel's recording

No.	Type of Vowel	Section	Quantity
1.	'E' sound	Short	30 sound
2.	'E' sound	Moderate	30 sound
3.	'E' sound	Long	30 sound

Table 3.4: The 'i' sound for vowel's recording

No.	Type of Vowel	Section	Quantity
1.	'i' sound	Short	30 sound
2.	'i' sound	Moderate	30 sound
3.	'i' sound	Long	30 sound

Table 3.5: The 'o' sound for vowel's recording

No.	Type of Vowel	Section	Quantity
1.	'o' sound	Short	30 sound
2.	'o' sound	Moderate	30 sound
3.	'o' sound	Long	30 sound

Table 3.6: The 'u' sound for vowel's recording

No.	Type of Vowel	Section	Quantity
1.	'u' sound	Short	30 sound
2.	'u' sound	Moderate	30 sound
3.	'u' sound	Long	30 sound

3.1.1.2 Total of Data set Collected

At phase 1, the recording was recorded by 20 normal person and 6 stroke patient. Phase 1 must be completed and validate the functionality to proceed to second phase. Table 3.7 shows the total sound recorded by normal person and stroke patient while 3.8 shows the total data set collected from by normal person and stroke patient.

Table 3.7: Total Sound for Normal Person and Stroke Patient

No.	Category of sound	Classes of Vowel	Quantity	Total Sound
1.	Normal People	6	90 sound	540 Sound
2.	Stroke Patient	6	30 sound	180 sound

Table 3.8: Total Data Set for Normal Person and Stroke Patient

No.	Category of sound	Total Sound	Quantity of Person	Total Data Set
1.	Normal People	540	20 normal person	10800 Sound
2.	Stroke Patient	180	6 stroke patient	1080 sound

The total data set collected from normal person and stroke patient is 11880 sound for vowel a,e,E,i,o and u. The quantity of sound recorded for stroke patient is not much as normal people because there are some obstacle during the Malaysia Government Movement Control Order (MCO). Due to the pandemic and health reasons, the total of stroke patient that successfully recorded is 6 patient and the total data set collected from stroke patient is 1080 of 6 classes of vowels.

3.1.1.3 Method of Vowel's Recording

The process of vowel's recording required 20 normal person and 6 stroke patients. Every process of recording have the same method in measurement between sound recorder and person. The distance between the voice recorder and people's mouth is 15cm apart. The measurement is fixed and all the external noises is blocked during the recording session in order to get a good quality and accurate of vowel's sound.

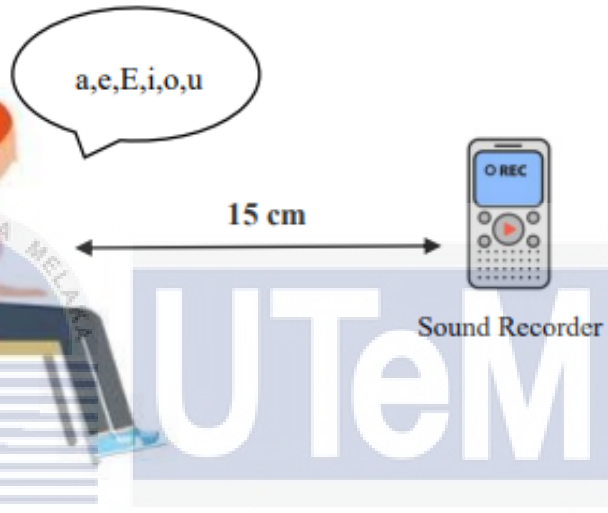

Figure 3.4: Method of Vowel's Recording

3.1.2 Fine Tune of Data set Images

3.1.2.1 Conversion from Audio To Image

Proceed to second phase of this project which is the process of fine tune of the data set images, all the recording collected by normal person and stroke patient will convert into spectrogram image. A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time. In this stage, the process of cropping and splitting the data set into two category will be occurred. It will converted the sound into image, and the image will be cropped into its vowel's category by using manual cropping.

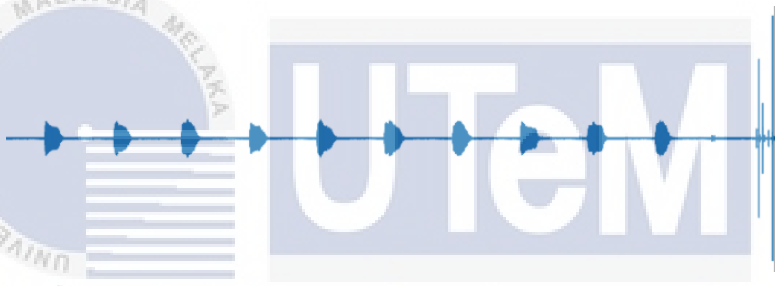

Figure 3.5: Audio Track in wav. file

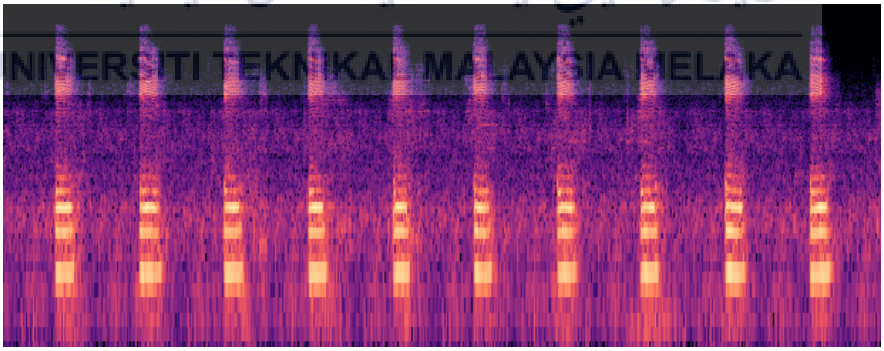

Figure 3.6: Spectrogram Image

3.1.2.2 Data set Manual Cropping

The next process is cropping process after the conversion from wav.file audio to spectrogram image. This process is conducted manually one by one and every cropped image have the dimension of 55 pixels x 240 pixels. The height of the spectrogram is 240 pixels while the width is 55pixels. The bit depth of every image is fixed which is 32 bit depth. Figure 3.8 shows the cropping process of the spectrogram images. All the cropping images have been saved properly into its directories.

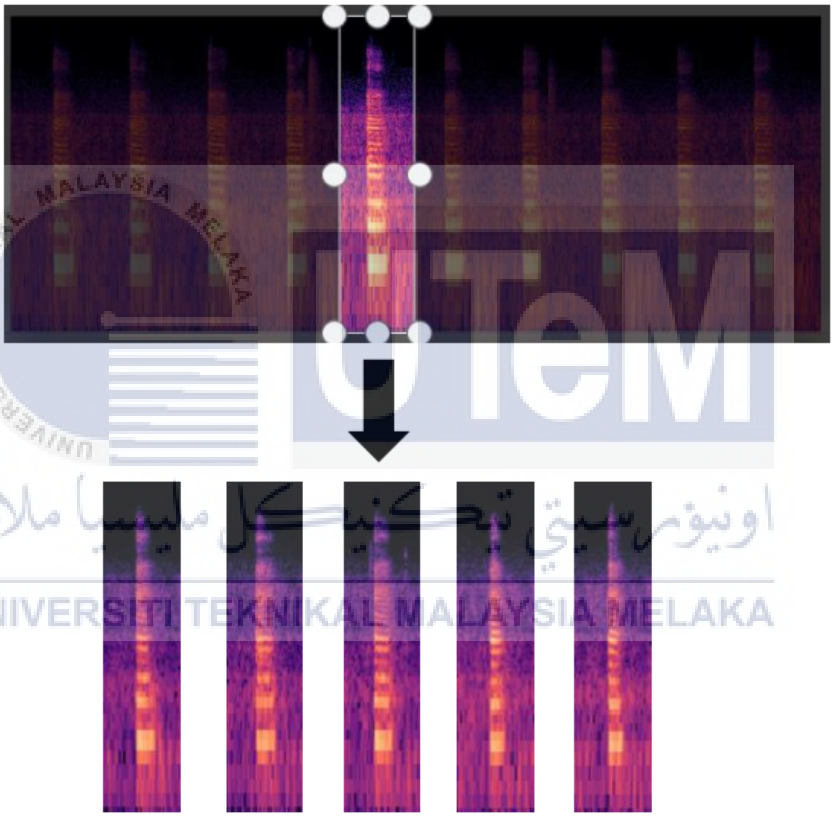

Figure 3.7: The process of cropping data

3.1.3 Design Nueral Network

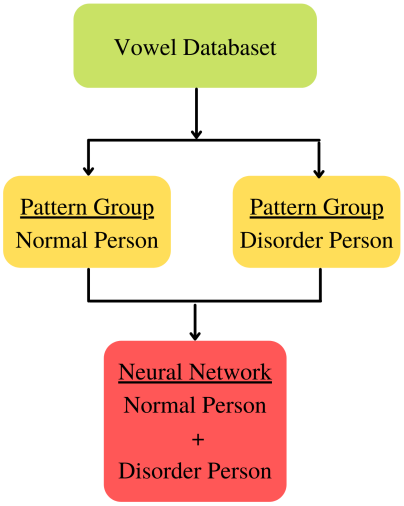

Figure 3.8: The Methodology of Neural Network

In the third phase, a neural network for normal person and disorder person will be design by using Convolutional Neural network (CNN). Particularly, CNN is the main deep learning (DL) architecture used for image processing and deep learning currently is one of the most used machine-learning (ML) based method. For the CNN frameworks, it is known that CNN are one of deep learning method and model and in turn, deep learning is a component of machine learning [16]. For this project, the neural network for every directory will be designed.

3.1.3.1 Convolutional Neural Network

Convolutional Neural Network (CNN) is widely used in images recognition, image classification, object detection and recognition of faces. Figure 3.9 shows the CNN architecture which consists of convolutional layer, pooling layer, fully connected layer, dropout and activation functions [3]. The pooling layer will perform down the sampling along the spatial dimensionality of the given input. Then, the fully connected layers will perform the same duties found in standard ANNs and attempt to produce class scores from the activation. The dropout layer is utilized wherein a few neurons are dropped from neural network during training process and it will reduced size of the model. The function of activation functions is to learn and approximate any kind of continuous and complex relationship between variables of the network [17].

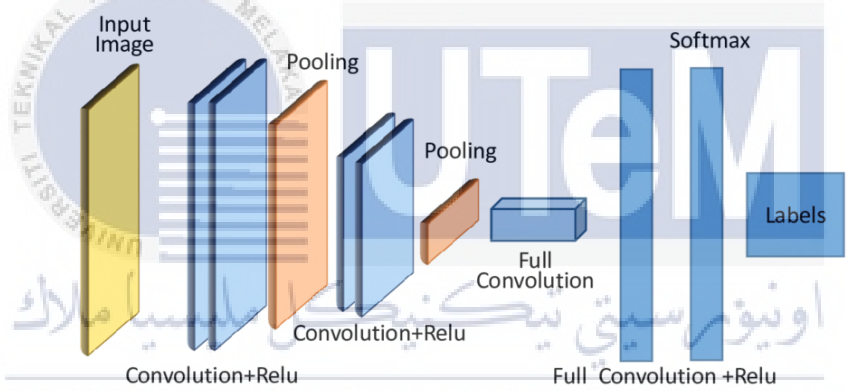

Figure 3.9: CNN Architecture [3]

3.1.3.2 VGG-16

In order to design the neural network, software VGG-16 will be used. VGG-16 is a deep convolutional neural network consist of 16 layers it has combination of 3×3 convolutional layers and 2×2 pooling layers repeatedly. VGG-16 has a better feature learning ability than AlexNet because it is deeper than AlexNet because of the combination of 3×3 convolutional layers and 2×2 pooling layers repeatedly. It is also simpler that InceptionNet and it an obtain a good effect in image classification. In addition, VGG-16 has a better generalization's ability, and it can adapt to a variety of data sets including tumor images [18].

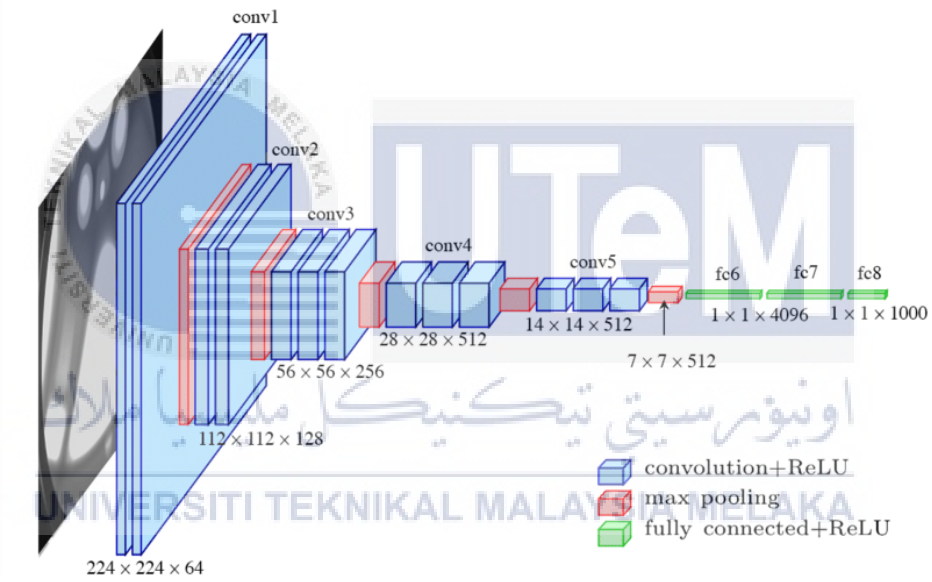

Figure 3.10: VGG-16 Network Architecture [12]

Table 3.9 shows the architecture of VGG-16. The input is a image of dimension (240,55,3). The first two layer in block 1 have 64 channels of 3x3 and same padding before a max pooling layer. Then, two layers which have convolution layers of 128 filter size and followed by max pooling layer same as previous block [19]. After that, there are three sets of 3 convolution layer and a max pool layer. The third block have convolution layers of 256 filter size and a max pooling layer and the fourth and fifth block have convolution layers of 512 filter size with same padding and a max pooling layer [20].

Table 3.9: Architecture of VGG-16

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 240, 55, 3)]	0
block1_conv1 (Conv2D)	(None, 240, 55, 64)	1792
block1_conv2 (Conv2D)	(None, 240, 55, 64)	36928
block1_pool (MaxPooling2D)	(None, 120, 27, 64)	0
block2_conv1 (Conv2D)	(None, 120, 27, 128)	73856
block2_conv2 (Conv2D)	(None, 120, 27, 128)	147584
block2_pool (MaxPooling2D)	(None, 60, 13, 128)	0
block3_conv1 (Conv2D)	(None, 60, 13, 256)	295168
block3_conv2 (Conv2D)	(None, 60, 13, 256)	590080
block3_conv3 (Conv2D)	(None, 60, 13, 256)	590080
block3_pool (MaxPooling2D)	(None, 30, 6, 256)	0
block4_conv1 (Conv2D)	(None, 30, 6, 512)	1180160
block4_conv2 (Conv2D)	(None, 30, 6, 512)	2359808
block4_conv3 (Conv2D)	(None, 30, 6, 512)	2359808
block4_pool (MaxPooling2D)	(None, 15, 3, 512)	0
block5_conv1 (Conv2D)	(None, 15, 3, 512)	259808
block5_conv2 (Conv2D)	(None, 15, 3, 512)	2359808
block5_conv3 (Conv2D)	(None, 15, 3, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 1, 512)	0
flatten (Flatten)	(None, 3584)	0
dense (Dense)	(None, 6)	0

3.1.3.3 Python

Python is an interpreted, high-level programming language for the general purpose and it is a language and object oriented approach aim to enable programmers to write clear, logical code for large and small projects [21]. Figure 3.11 shows the version of the Phyton use for this project. Python 3.0, released in 2008, was an extensive revision of the language which are not fully backward-compatible and Python 2 code does not run unmodified on Python 3 [22].

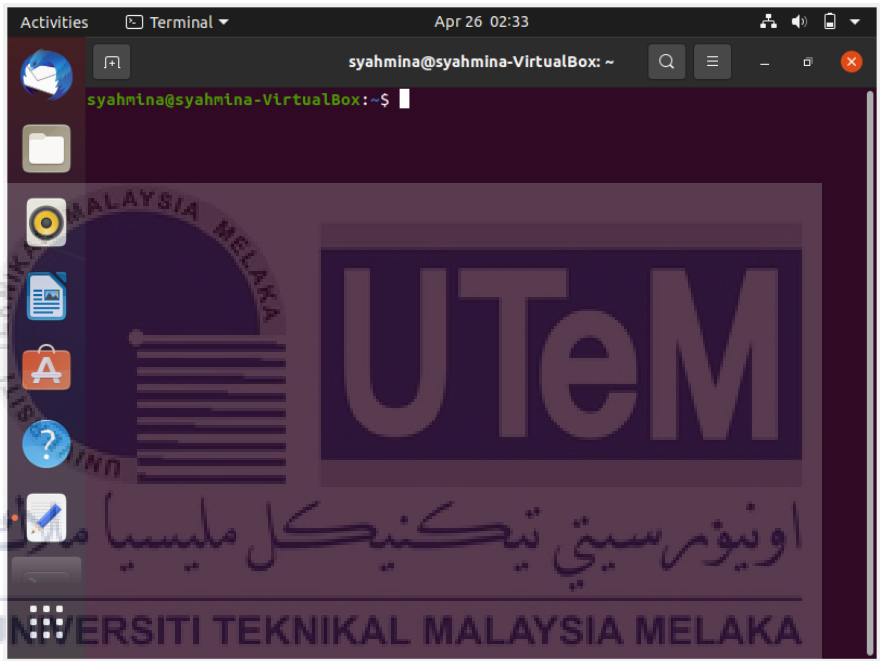

Figure 3.11: Python in Ubuntu Terminal

3.1.4 Vowel Recognition Evaluation

When the data set has been split into two directories and the neural for each directory has been designed, the training and testing process will be conducted. For this process, the framework that will be used is Keras and TensorFlow. There are a lot of machine learning libraries available as the functions of machine learning are like recognizing speech, determining objects and more. Some popular machine learning libraries are 'TensorFlow', 'Keras', 'sciKit learn', 'Theano', and 'Microsoft Cognitive toolkit(CNTK)' [23].

3.1.4.1 TensorFlow

TensorFlow is included among the most popular machine learning libraries and it is an open-source ML library that uses symbolic math for data flow and differentiable programming. The advantages of using TensorFlow is that it can increase the functionality as it is more advanced in high-level operations. It also increases control because some of neural networks may require a lot of control. TensorFlow was developed by Google and has been released in 2015.

Figure 3.12: TensorFlow Logo

3.1.4.2 Keras

Another machine learning libraries that have been used by many users is Keras. Like TensorFlow, Keras is also an open-source machine learning library. Keras builds and trains neural networks, and it is more friendly and modular than TensorFlow. The advantages of Keras are it is highly flexible and extendable, easy of use and Keras models can connect configurable building blocks with few restrictions.

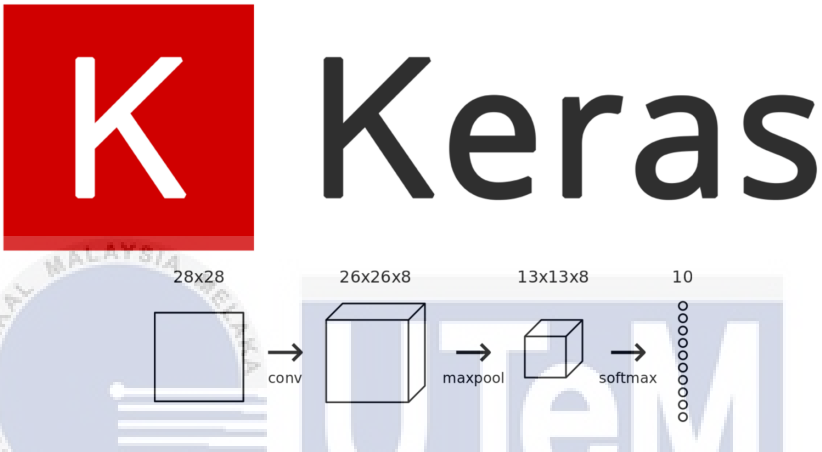

Figure 3.13: Keras Framework

The differences of TensorFlow and Keras framework are Keras put the functionalities of machine learning and deep learning libraries including TensorFlow, Theano and Microsoft Cognitive Toolkit. Most of users like using Keras framework as it makes TensorFlow mush easier to navigate.

3.2 Hardware Requirement

Each of the hardware used in this project is listed in Table 3.10.

Table 3.10: List of Hardware

No.	Hardware	Model
1.	Laptop	Dell 3000 Series
2.	Webcam	Full HD 1080P Webcam
3.	Sound Recorder	Remax RP1 Voice Recorder

3.3 Software Requirement

Each of the software used in this project is listed in Table 3.11.

Table 3.11: List of Software

No.	Software	Version
1.	VGG-16	-
2.	Python (Programming Language)	3.7.5
3.	Keras (Framework)	2.3.0
4.	TensorFlow (Framework)	2.0

CHAPTER 4

RESULTS AND DISCUSSION

This chapter includes all the results acquired at each stage of the methodology. Analysis and discussion is also done on the results.

4.1 Analysis of the Project

In the Speech Rehabilitation for disorder people project, there are three analysis and every analysis contained different classes of data set . Analysis 1, 2 and 3 with different specification have been tested by using Convolutional neural Network method and the accuracy and validation accuracy obtained from the network have been observed.

Table 4.1 shows the three different epoch used in this analysis. Epoch means the passes over the data set forward and backward and it will pass only once [24]. Every epoch shows different results in this analysis. Table 4.2 shows three different batch size used in this analysis. Batch size is a hyperparameter that defines the number of samples that can tuning during the neural network process [25]. The larger the batch size, the faster the model per epoch during training as the batch size is a number of samples processed before the model is updated.

Table 4.1: Epoch used in Analysis 1

No.	Epoch
1.	20
2.	50
3.	80

Table 4.2: Batch size used in Analysis 1

No.	Batch Size
1.	6
2.	9
3.	15

4.1.1 Analysis 1

Analysis 1 contained data set from 20 normal people of spectrogram images. This analysis did not include any data set from disorder patient as this analysis want to observe the accuracy obtained from the data set of normal people. Table 4.3 shows the total data set of spectrogram images for this analysis 1.

Table 4.3: Total data set for Analysis 1

No.	Type	Classes of vowels	Total Spectrogram Images
1.	Train	1440 images x 6	8640 images
2.	Validation	180 images x 6	1080 images
3.	Testing	180 images x 6	1080 images

Total Images (Train + Validation + Testing) = 10800 images

4.1.1.1 Result Designed Model of Analysis 1

Figure 4.1 shows the model accuracy and model loss of designed model for normal people data set with epoch 20 and batch size 6 while figure 4.2 shows the model accuracy for epoch 20 and batch size 15. Designed model is a model created by someone else to solve a similar problem.

Epoch = 20 , Batch size = 6

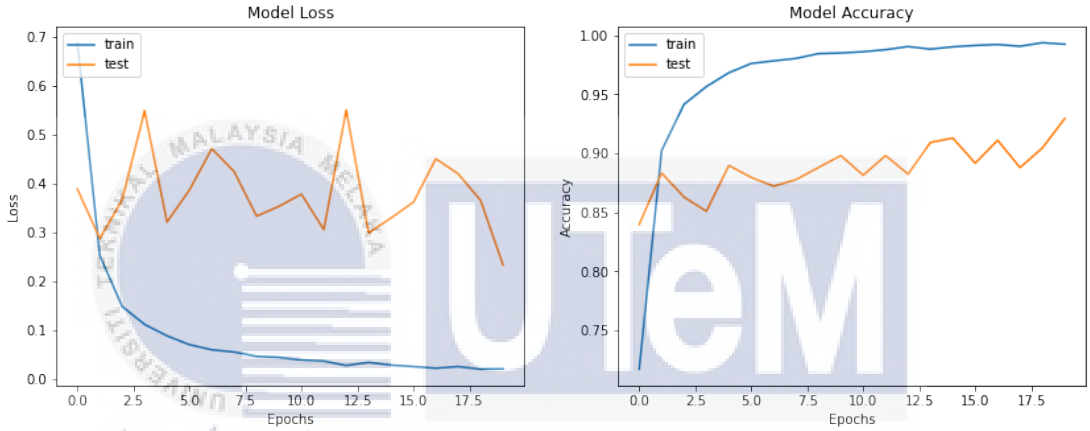

Figure 4.1: Graph of Model Loss and Model Accuracy for Analysis 1

Epoch = 20 , Batch size = 15

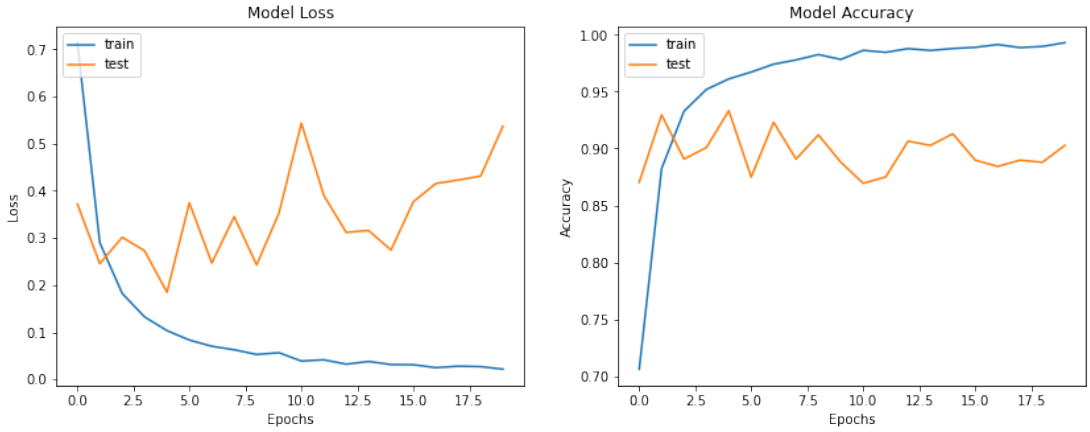

Figure 4.2: Graph of Model Loss and Model Accuracy for Analysis 1

Table 4.4 shows the result of model accuracy and validation accuracy for the designed model for every different epoch and batch size. All the accuracy have been recorded and compared for every training. From the table 4.4, the highest validation accuracy is 92.96% with epoch 20 and batch size 6. All the validation accuracy passed up above 85% and it proves that the data set of spectrogram images for 20 normal people have higher accuracy in every classes of vowels.

Table 4.4: Accuracy and Validation Accuracy for Analysis 1

No.	Epoch	Batch Size	Training Accuracy (%)	Validation Accuracy (%)
1.	20	6	99.95	92.96
2.	50	6	100.00	87.96
3.	80	6	100.00	88.80
4.	20	9	99.99	90.19
5.	50	9	99.97	85.65
6.	80	9	100.00	87.59
7.	20	15	99.73	90.28
8.	50	15	100.00	91.52
9.	80	15	100.00	85.74

4.1.1.2 Result VGG-16 of Analysis 1

Figure 4.3 shows the model accuracy and model loss of VGG-16 model for normal people data set with epoch 20 and batch size 6. VGG-16 is a Convolutional Neural Network that have 16 layers deep.

Epoch = 20 , Batch size = 6

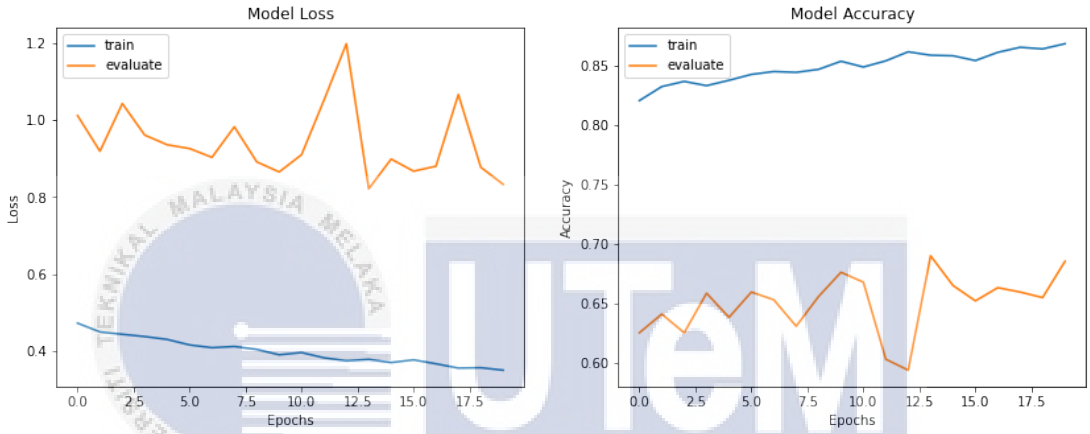

Figure 4.3: Graph of Model Loss and Accuracy of VGG-16 model for Analysis 1

Table 4.5 shows the result of model accuracy and validation accuracy for VGG-16 model for epoch 20 and 80. The validation accuracy for VGG-16 model is decreased compared to the designed model because VGG-16 model is a convolutional neural network that have 16 layers deep.

Table 4.5: Accuracy and Validation Accuracy of VGG-16 model for Analysis 1

No.	Epoch	Batch Size	Training Accuracy (%)	Validation Accuracy (%)
1.	20	6	87.88	68.52
2.	50	6	85.46	66.34

4.1.2 Analysis 2

Analysis 2 contained data set from 20 normal people and 6 disorder patient of spectrogram images. This analysis have include the data set from patient and from this analysis, the accuracy obtained from the data set of normal people and patient have been observed. This analysis have been tested by using 3 different epochs and batch sizes. Table 4.6 shows the total data set of spectrogram images for this analysis 2.

Table 4.6: Total data set for Analysis 1

No.	Type	Classes of vowels	Total Spectrogram Images
1.	Train	1584 images x 6	9504 images
2.	Validation	198 images x 6	1188 images
3.	Testing	198 images x 6	1188 images

Total Images (Train + Validation + Testing) = 11880 images

4.1.2.1 Result Designed Model of Analysis 2

Figure 4.4 shows the model accuracy and model loss for normal people data set with epoch 20 and batch size 6 while figure 4.5 shows the model accuracy for epoch 20 and batch size 15.

Epoch = 20 , Batch size = 6

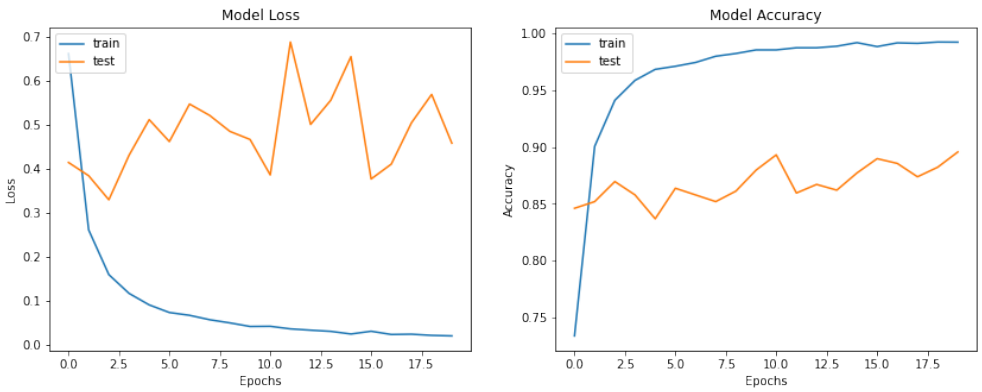

Figure 4.4: Graph of Model Loss and Model Accuracy for Analysis 2

Epoch = 20 , Batch size = 15

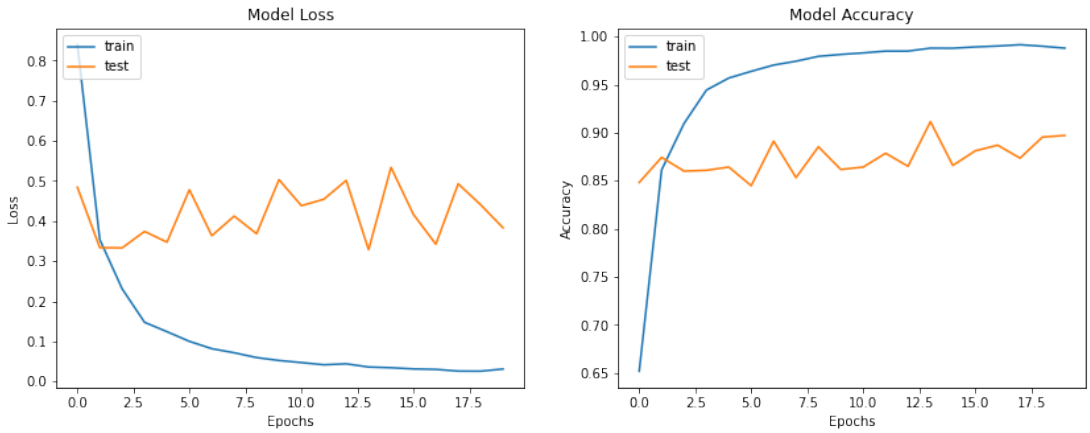

Figure 4.5: Graph of Model Loss and Model Accuracy for Analysis 2

Table 4.7 shows the result of model accuracy and validation accuracy for every different epoch and batch size. All the accuracy have been recorded and compared for every training. From the table 4.7, the highest validation accuracy is 89.73% with epoch 20 and batch size 15. The validation accuracy for epoch 20 and batch size 6 have a slight difference which is 89.56% compared to the batch size 15. All the validation accuracy passed up above 85% and it proves that the data set of spectrogram images for 20 normal people and 6 stroke patient have higher accuracy in every classes of vowels.

Table 4.7: Accuracy and Validation Accuracy for Analysis 2

No.	Epoch	Batch Size	Training Accuracy (%)	Validation Accuracy (%)
1.	20	6	99.97	89.56
2.	50	6	99.98	87.46
3.	80	6	100.00	86.87
4.	20	9	99.88	88.80
5.	50	9	100.00	88.55
6.	80	9	100.00	89.39
7.	20	15	99.80	89.73
8.	50	15	100.00	89.06
9.	80	15	100.00	88.38

4.1.2.2 Result VGG-16 of Analysis 2

Figure 4.6 shows the model accuracy and model loss of VGG-16 model for 20 normal people and 6 stroke patient data set with epoch 20 and batch size 6.

Epoch = 20 , Batch size = 6

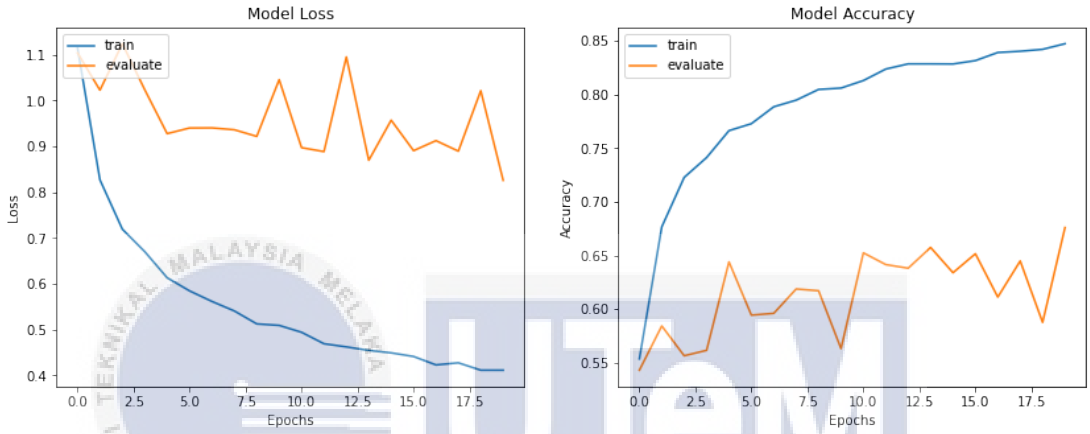

Figure 4.6: Graph of Model Loss and Accuracy of VGG-16 model for Analysis 2

Table 4.8 shows the result of model accuracy and validation accuracy for VGG-16 model for epoch 20 and 80. Compared to the designed model of the second analysis, the validation accuracy for VGG-16 model is decrease same as the first analysis.

Table 4.8: Accuracy and Validation Accuracy of VGG 16 model for Analysis 2

No.	Epoch	Batch Size	Training Accuracy (%)	Validation Accuracy (%)
1.	20	6	86.21	67.59
2.	80	6	89.33	67.34

4.1.3 Analysis 3

Table 4.9 shows the third analysis that contained data set from 6 normal people and 6 disorder patient of spectrogram images. This analysis have the same and balance amount of data set from normal people and disorder patient. The accuracy obtained from the data set of this analysis have been observed. This analysis have been tested by using 3 different epochs and batch sizes.

Table 4.9: Total data set for Analysis 3

No.	Type	Classes of vowels	Total Spectrogram Images
1.	Train	432 images x 6	2592 images
2.	Validation	54 images x 6	324 images
3.	Testing	54 images x 6	324 images

Total Images (Train + Validation + Testing) = 3240 images

4.1.3.1 Result Designed Model of Analysis 3

Figure 4.7 shows the model accuracy and model loss for normal people data set with epoch 20 and batch size 6 while figure 4.10 shows the model accuracy for epoch 20 and batch size 15.

Epoch = 20 , Batch size = 6

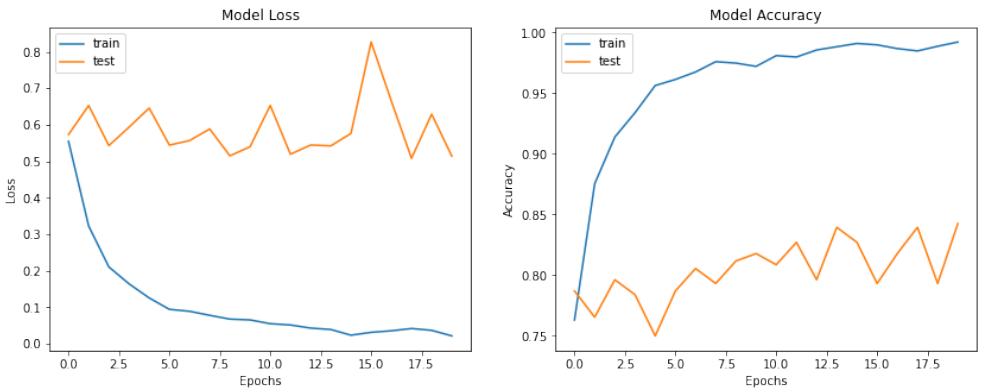

Figure 4.7: Graph of Model Loss and Model Accuracy for Analysis 3

Epoch = 20 , Batch size = 9

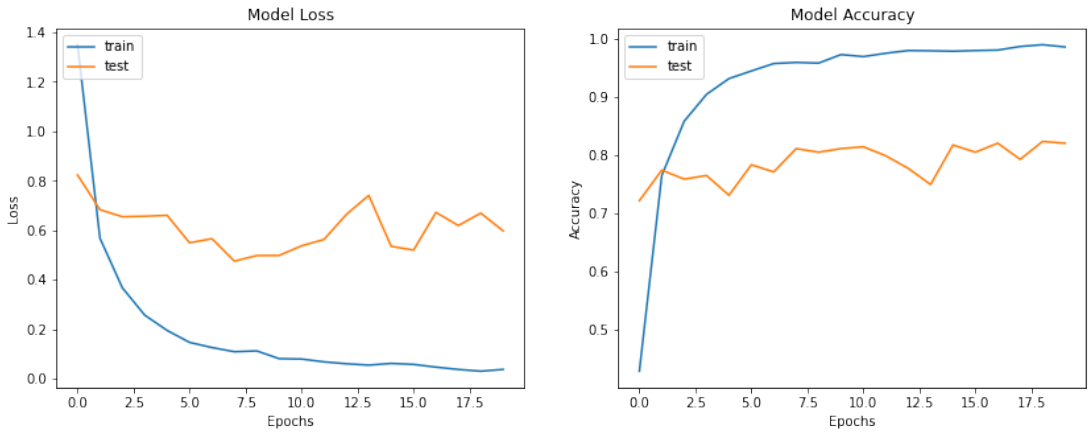

Figure 4.8: Graph of Model Loss and Model Accuracy for Analysis 3

Table 4.10 shows the result of model accuracy and validation accuracy for every different epoch and batch size. All the accuracy have been recorded and compared for every training. From the table 4.10, the highest validation accuracy is 84.57% with epoch 80 and batch size 9. The validation accuracy for epoch 20 and batch size 6 have a slight difference which is 84.26%. All the validation accuracy passed up above 75% and it proves that the data set of spectrogram images for 6 normal people and 6 stroke patient have a high accuracy in every classes of vowels.

Table 4.10: Accuracy and Validation Accuracy for Analysis 3

No.	Epoch	Batch Size	Training Accuracy (%)	Validation Accuracy (%)
1.	20	6	99.85	84.26
2.	50	6	100.00	77.47
3.	80	6	100.00	80.25
4.	20	9	99.85	82.10
5.	50	9	100.00	81.79
6.	80	9	100.00	84.57
7.	20	15	100.00	80.86
8.	50	15	100.00	76.85
9.	80	15	100.00	81.79

4.1.3.2 Result VGG-16 of Analysis 3

Figure 4.9 shows the model accuracy and model loss of VGG-16 model for 6 normal people and 6 stroke patient data set with epoch 80 and batch size 6.

Epoch = 80 , Batch size = 6

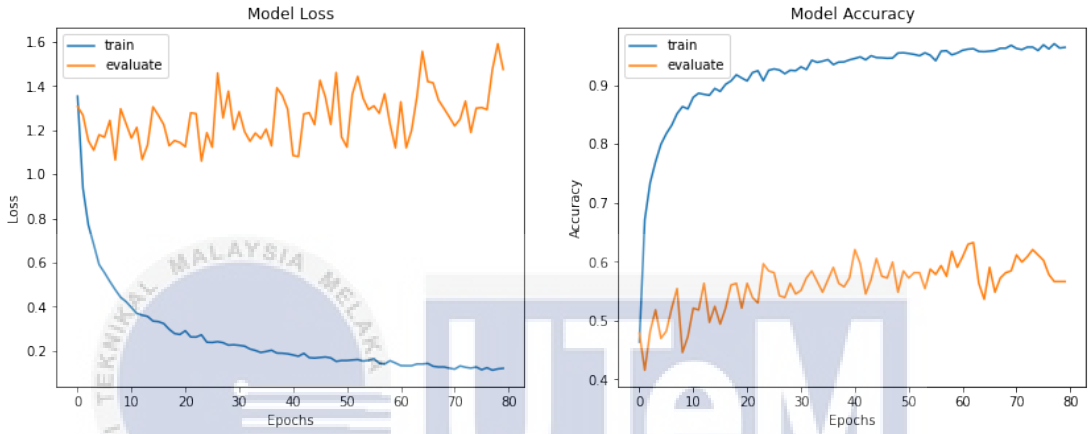

Figure 4.9: Graph of Model Loss and Accuracy of VGG-16 model for Analysis 3

Table 4.11 shows the result of model accuracy and validation accuracy for VGG-16 model for epoch 20 and 80. Compared to the designed model of the first and second analysis, the validation accuracy for VGG-16 model is decrease compared to the designed model.

Table 4.11: Accuracy and Validation Accuracy of VGG-16 model for Analysis 3

No.	Epoch	Batch Size	Training Accuracy (%)	Validation Accuracy (%)
1.	20	6	96.80	59.04
2.	80	6	96.26	56.63

4.1.4 Analysis Discussion

From the analysis, the result is observed and studied to understand the flow of each analysis. From the explanation of each function from previous chapter, the program flow is easily translated into sentences.

4.1.5 Comparison of Epoch in each Analysis

All the results in every analysis were collected and the comparison of the result have been displayed in table 4.12 and 4.13. The result is obtained from the table 4.4, 4.7 and 4.10.

Table 4.12: Comparison of the best Epoch in Analysis

Analysis	Epoch	Batch Size	Training Accuracy (%)	Validation Accuracy (%)
1	20	6	99.95	92.96
		9	99.99	90.19
		15	99.73	90.28
2	20	6	99.97	89.56
		9	99.88	88.80
		15	99.80	89.73
3	20	6	99.85	84.26
		9	99.85	82.10
		15	100.00	80.86

From table 4.12, all the result have been collected and compared in this table. The best epoch between epoch 20,50 and 80 is epoch 20 with the highest validation accuracy is 92.96% in analysis 1. In analysis 2, the highest percentage of epoch 20 is 89.73% while in the third analysis, the highest validation percentage is 84.26%.

In the first, second and third analysis, the best epoch is 20 because the training process for all analysis is enough with epoch 20 like figure 4.10. The validation accuracy for every analysis is decrease when the value of epoch is increased from epoch 20 to epoch 50. When the epoch value is increased to 80, the validation accuracy also increased, but the result for the graph is overfitting like figure 4.12 . This means, the data set for every analysis is sufficiently enough for epoch 20 without having an overfitting graph with a better accuracy.

While conducting this training data set for every analysis, there is no optimal number for epoch. Every data set has differ number of epoch and by using the graph, it shows the training process for when the training and validation data set are being learnt by the model.

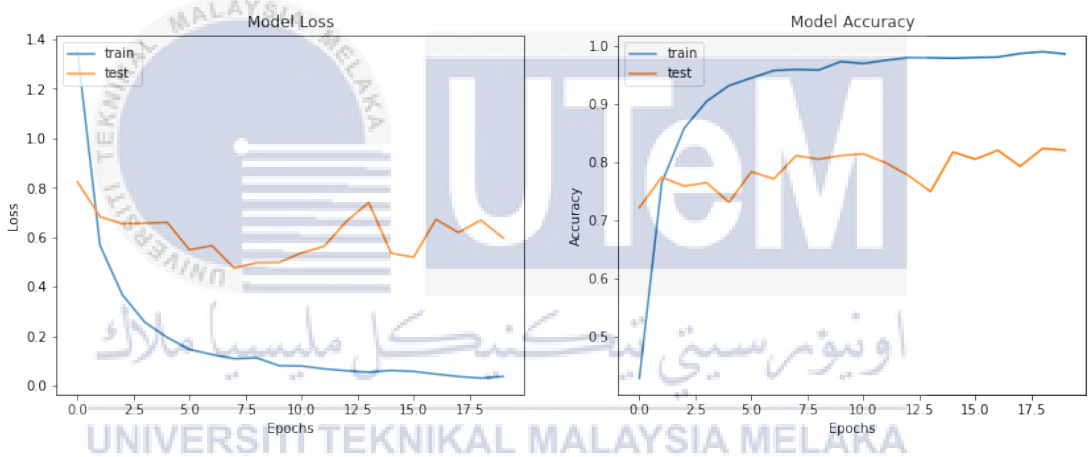

Figure 4.10: Graph of Model Loss and Accuracy for epoch 20

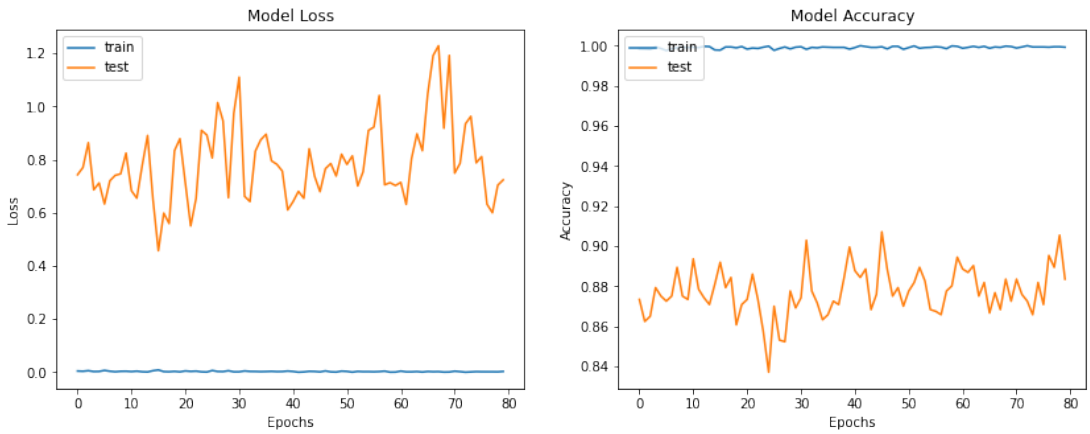

Figure 4.11: Graph of Model Loss and Accuracy for epoch 80

4.1.6 Comparison of Batch Size in each Analysis

Table 4.13: Comparison of the best Batch Size in Analysis

Analysis	Epoch	Batch Size	Training Accuracy (%)	Validation Accuracy (%)
1	20	15	99.73	90.28
	50		100.00	91.52
	80		100.00	85.74
2	20	15	99.80	89.73
	50		100.00	89.06
	80		100.00	88.38
3	20	15	100.00	80.86
	50		100.00	76.85
	80		100.00	81.79

Table 4.13 shows the comparison of the best batch size which is 15 in every analysis with the highest validation accuracy is 91.52% in analysis 1. In the second analysis, the highest validation accuracy is 98.73% while the highest accuracy for batch size 15 in analysis 3 is 81.79%.

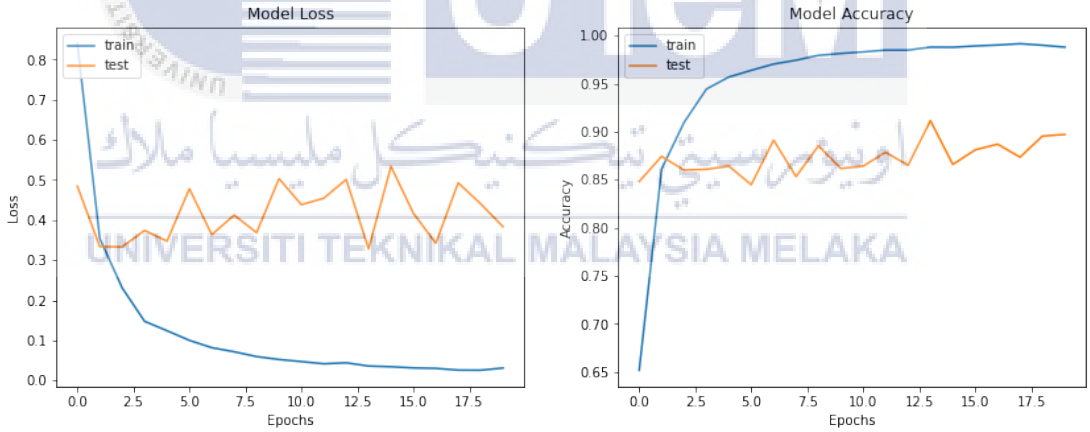

Figure 4.12: Graph of Model Loss and Accuracy for epoch 20

The comparison of each batch size give different validation accuracy in each analysis. Every analysis used three batch size which are 6, 9 and 15. The batch size is a number of samples processes before the model is updated. At the end of the batch, the predictions are compared to the expected output variables and an error is calculated [26]. The best batch size for every analysis is 15. The larger the batch size, the faster the model per epoch during training [27].

4.1.7 Comparison of Model in each Analysis

All the results in every analysis were collected and the comparison of the result have been displayed in table 4.14. The result is obtained from the table 4.4, 4.7 and 4.10.

Table 4.14: Comparison of Model in Analysis

No.	Model	Analysis	Training Accuracy (%)	Validation Accuracy (%)
1.	Designed	1	99.95	92.96
		2	99.97	89.56
		3	99.85	84.26
2.	VGG 16	1	87.88	68.52
		2	86.21	67.59
		3	96.80	59.04

For every analysis, the designed model have the highest training accuracy and validation accuracy compared to the VGG-16 model. Figure 4.14 shows the designed model layers and it have less layers compared to VGG-16 layers. So the process training required less layers through every layer and the accuracy is more higher compared to the VGG-16. Figure 4.13 and figure 4.15 shows the comparison of graph between designed model and VGG-16 model of training accuracy and validation accuracy.

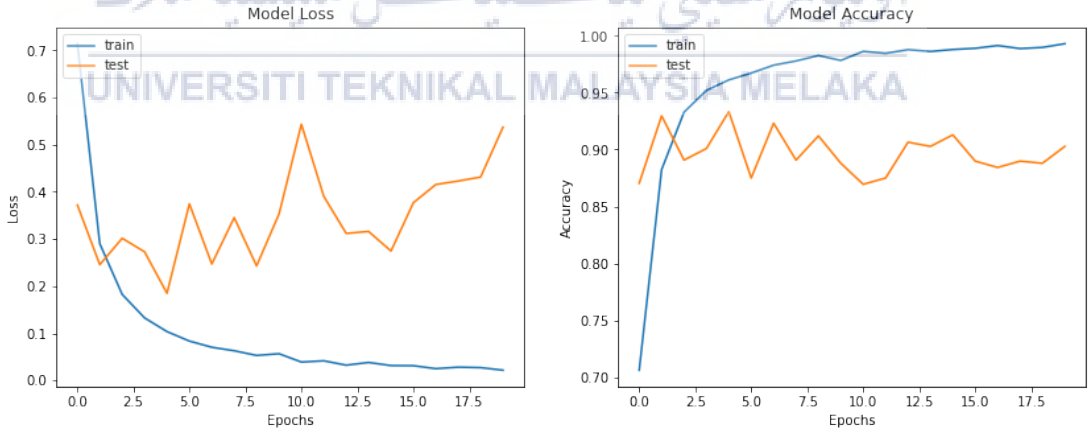

Figure 4.13: Designed Model

```
vowel_model = Sequential()
vowel_model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(240,55,3)))
vowel_model.add(Conv2D(32, kernel_size=(3, 3), activation='relu'))
vowel_model.add(MaxPooling2D(pool_size=(2, 2)))
vowel_model.add(Dropout(0.25))

vowel_model.add(Conv2D(32, kernel_size=(3, 3), activation='relu' ))
vowel_model.add(Conv2D(32, kernel_size=(3, 3), activation='relu'))
vowel_model.add(MaxPooling2D(pool_size=(2, 2)))
vowel_model.add(Dropout(0.25))

vowel_model.add(Conv2D(64, kernel_size=(3, 3), activation='relu'))
vowel_model.add(Conv2D(64, kernel_size=(3, 3), activation='relu'))
vowel_model.add(MaxPooling2D(pool_size=(2, 2)))
vowel_model.add(Dropout(0.25))

vowel_model.add(Flatten())
vowel_model.add(Dense(1024, activation='relu'))
vowel_model.add(Dropout(0.5))

vowel_model.add(Dense(6, activation='softmax'))
```

Figure 4.14: Designed Model Layers

The designed model as shown in Figure 4.14 shows the layer used during the training process. The input is a image of dimension (240,55,3). The first two layer in block 1 have 32 channels of 3x3 and same padding before a max pooling layer. Then, two layers which have convolution layers of 32 filter size in block 2 and followed by max pooling layer same as previous block. After that, there are 2 convolution layer of 256 filter size and a max pooling layer same padding and a max pooling layer. [20].

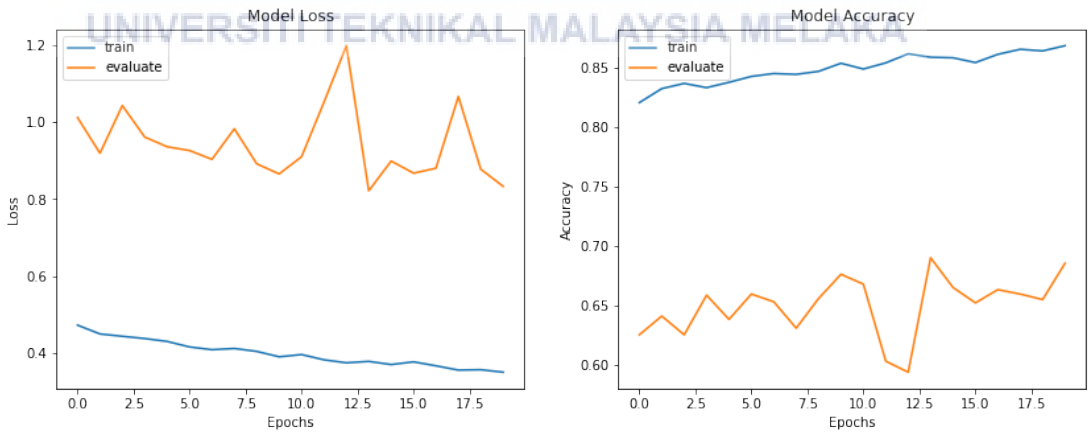

Figure 4.15: VGG-16 Model

4.1.8 Comparison of Analysis

Analysis 1,2 and 3 have different amount of data set and fixed dimension of spectrogram images. The width of spectrogram images is 240 pixels and the height is 55 pixels. The bit depth is same for every images which are 32 bit depth. Between the three analysis, analysis 1 have the highest validation accuracy for epoch 20 and batch size 6 with 92.96%. This is because analysis 1 only contained the data set from normal people and the data set from the first analysis have a balanced amount of spectrogram images between the male and female gender.

Mostly, normal and healthy people didn't have any problem to pronounce all the classes of vowels so, the quality of the spectrogram images after the conversion from the audio is better than stroke patient. Figure 4.16 shows the model accuracy which have higher accuracy than the stroke patient model accuracy in Figure 4.17.

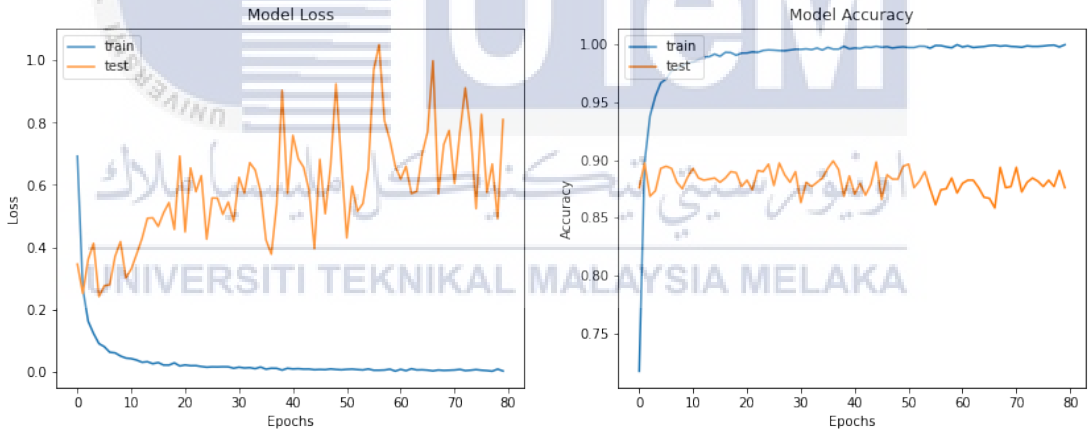

Figure 4.16: Designed Model for Normal People

The validation accuracy for stroke patient data set is lower than normal people because they have problem to pronounce the vowels during the recording session. Some of them have problems to pronounce some classes of vowel. Mostly, the vowel's pronunciation of e and i sounds similar and it will affect the data set training. In the second and third analysis, the data set have combination of normal person and stroke patient. So, from the training process, we can observe the model accuracy of the designed model and VGG-16 model.

From the observation, we can conclude that the validation accuracy of analysis 1 which contained 20 normal person data set have decrease in the range 1% to 5% when the data set have the combination with stroke patient. The validation accuracy didn't have a big difference from the first analysis to the third analysis because we have a large amount of normal people data set. It will give advantages during the training process for every classes of vowels as it can learn more for every vowel's features.

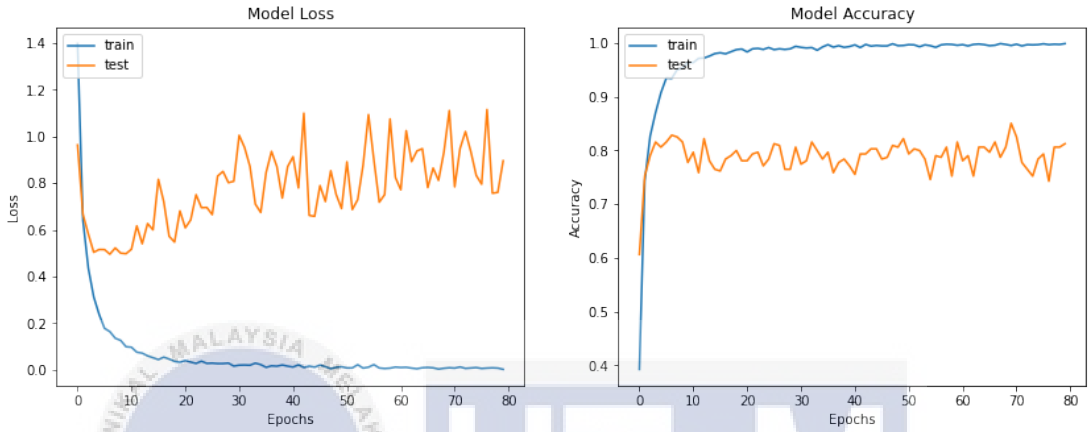

Figure 4.17: Designed Model for Stroke Patient

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

As the conclusion, choosing the proper deep models is important because the models is to fit in every small data set and it will determine the performance in every analysis. The data augmentation, dropout and fine- tune need to apply into the designed model to avoid the overfitting results. The deep models can be used to fit a very small data set as long as the good model is chosen and proper modifications are applied.

As a summary, the designed model has been chosen as the model to be used in this project. It give the best validation accuracy and it proves the accuracy in all the data set that have been tested. All the results of analysis 1, analysis 2 and analysis 3 have been recorded and as a result of comparison between all the analysis, it shows that analysis 1 has a better accuracy between analysis 1 and 2. However, every analysis shows a good accuracy and the data set of normal person and stroke patient is accurate. On overall, the objective has been achieved.

5.2 Future Work

In future, experiments can be conducted in different models with proper modifications in the future. This project can be modified more models to better fit in very small data sets and comparing them are meaningful to work in the future.

In addition, instead of spectrogram images, this project can be conducted by using Mel-frequency cepstral coefficients (MFCC) images and this images can be tested using Convolutional Neural Network and tested the training and validation accuracy. So, we can tested the accuracy between the spectrogram images and MFCC images and observe for the better accuracy.Last but not least, this project can come out with a application prototype which can help users to identify vowels especially to the disorder people which are need in communications.

REFERENCES

[1] S. M. C. N. S. P. K. J. R. T. A. W. J. Llewellyn, Carrie D. Ayers, "Cambridge handbook of psychology, health and medicine: Third edition," *Cambridge Handbook of Psychology, Health and Medicine: Third Edition*, pp. 1–682, 2019.

[2] C. W. Peipei Chen, Jun MA, "Assessment of outcomes of hearing and speech rehabilitation in children with cochlear implantation," pp. 57–62, 6 2019.

[3] M. Véstias, *Convolutional Neural Network*, 01 2022, pp. 1559–1575.

[4] D. E. Susan Wortman-Jutt, "Poststroke aphasia rehabilitation: Why all talk and no action?" 3 2019.

[5] Y. A. P. C. S. K. B. J. Waber DP, Boiselle EC, "Developmental dyspraxia in children with learning disorders: Four-year experience in a referred sample," pp. 210–221, 4 2021.

[6] A. E. Donna C. Tippett, John K. Niparko, "Aphasia: Current concepts in theory and practice," 1 2015.

[7] C. Van Riper, *Speech correction : an introduction to speech pathology and audiology / Charles Van Riper, Robert L. Erickson. — 9th ed. p. cm.* Needham Heights, MA: A Simon Schuster Company, 1995.

[8] T. J. Karten, "Historical background of disabilities," *Embracing Disabilities in the Classroom: Strategies to Maximize Students Assets*, pp. 2–32, 2015.

[9] M. Clinic, "Stroke." [Online]. Available: https://www.mayoclinic.org/diseases-conditions/stroke/symptoms-causes/syc-20350113

[10] S. P. Rosenbaum S, *Speech and Language Disorders in Children: Implications for the Social Security Administration's Supplemental Security Income Program*, Washington (DC): National Academies Press (US), 2016.

[11] A. Einstein, "Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports," *Molecular Psychiatry*, vol. 21, no. 10, pp. 1366–1371, 2016.

[12] N. Milosevic, "Introduction to convolutional neural networks," *Introduction to Convolutional Neural Networks*, pp. 1–31, 2020.

[13] X. Liang, Ming Hu, "Recurrent convolutional neural network for object recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, no. 1, pp. 3367–3375, 2015.

[14] Y. H. P. H. Khaing, Zaw Min Naung, "Development of control system for fruit classification based on convolutional neural network," *Proceedings of the 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2018*, vol. 2018-January, no. 10, pp. 1805–1807, 2018.

[15] M. H.-G. R. B. R. J. F. C. V. A. Naranjo-Torres, José Mora, "A review of convolutional neural network applied to fruit image processing," *Applied Sciences (Switzerland)*, vol. 10, no. 10, 2020.

[16] S. Hershey, Shawn Chaudhuri, "Cnn architectures for large-scale audio classification," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 131–135, 2017.

[17] A. K. G. Sakshi Indolia, "Conceptual understanding of convolutional neural network- a deep learning approach," *Journal of Advances in Technology and Engineering Studies*, pp. 679–688, 2018.

[18] Y. P. B. L. D. D. J. Guan, Qing Wang, "Deep convolutional neural network vgg-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study," *Journal of Cancer*, vol. 10, no. 20, pp. 4876–4882, 2019.

[19] S. Tammina, "Transfer learning using vgg-16 with deep convolutional neural network for classifying images," *International Journal of Scientific and Research Publication*, pp. 143–150, 2019.

[20] M. L. Ouiza Nait Belaid, "Classification of brain tumor by combination of pretrained vgg16 cnn," *Journal of Advances in Technology and Engineering Studies*, vol. 12, no. 2, pp. 13–25, 2020.

[21] A. Pajankar, *Introduction to Python*, 01 2022, pp. 1–23.

[22] D. Kuhlman, "A python book," *A Python Book*, pp. 1–227, 2013.

[23] J. Hillenbrand, M. Clark, and R. Houde, "Some effects of duration on vowel recognition," *The Journal of the Acoustical Society of America*, vol. 108, pp. 3013–22, 01 2001.

[24] D. S. R. Saahil Afaq, "A significance of epochs on training a neural network," *INTERNATIONAL JOURNAL OF SCIENTIFIC TECHNOLOGY RESEARCH*, vol. 9, no. 20, pp. 485–488, 2020.

[25] O. A. Nabeel Zuhair Tawfeeq Abdulnabi, "Batch size for training convolutional neural networks for sentence classification," *Journal of Advances in Technology and Engineering Studies*, vol. 2, no. 5, pp. 156–163, 2016.

[26] O. Altun and N. Abdulnabi, "Batch size for training convolutional neural networks for sentence classiiication," 10 2016.

[27] N. V. Athapol Ruangkanjanase, "Solving for an optimal batch size for a single machine using the closed-form equations to minimize inventory cost," pp. 211–221, 4 2019.