

**Real-Time Bengali Sign Language Recognition on Word and Read Out Loud The
Speech Based on Computer Vision**

BY

**Mehedi Hasan
ID: 191-15-2766**

**Sirajus Sakib
ID: 191-15-12510**

This Report Presented in Partial Fulfillment of the Requirements for
the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By
Md. Sanzidul Islam
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By
Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH**

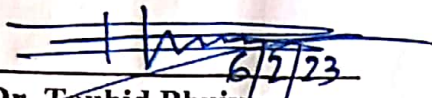
February 2023

APPROVAL

This Project titled "Real-Time Bengali Sign Language Recognition on Word and Read Out Loud The Speech Based on Computer Vision", submitted by Mehedi Hasan, ID No: 191-15-2766, Sirajus Sakib, ID No: 191-15-12510 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 02/02/2023.

BOARD OF EXAMINERS

Chairman


6/2/23

Dr. Fouhid Bhuiyan
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Sheak Rashed Haider Noori
Professor and Associate Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Md. Sazzadur Ahamed
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Md. Sazzadur Rahman
Associate Professor
Institute of Information Technology
Jahangirnagar University

Internal Examiner

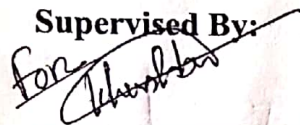
Internal Examiner

External Examiner

DECLARATION

We hereby declare that this research work has been done by us under the supervision of **Md. Sanzidul Islam, Senior Lecturer, Department of CSE, Daffodil International University**. Also, we want to make it clear that neither this research nor any aspect of it has been submitted for a degree elsewhere.

Supervised By:



Md. Sanzidul Islam
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

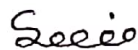


Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University

Submitted by:



Mehedi Hasan
ID: 191-15-2766
Department of CSE
Daffodil International University



Sirajus Sakib
ID: 191-15-12510
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First of all, We begin by expressing our sincere gratitude to Almighty Allah for the divine blessing which enabled us to successfully finish the research.

We would like to express sincere gratitude to our supervisor **Md. Sanzidul Islam, Senior Lecturer**, Department of CSE, Daffodil International University, for providing the necessary guidance in completing this study on the "**Real-Time Bengali Sign Language Recognition on Word and Read Out Loud The Speech Based on Computer Vision**". His support and direction given us with the certainty to complete the research accurately. He given us with all the related assets and significant resource to do this research possible.

We would like to thank **Sharun Akter Khushbu**, our co-supervisor, for helping us in completing this research. We are earnestly thankful to our department head for his important offer assistance in doing this kind of research and also thank to the other workforce individuals and our department's representatives for their support.

We would like to thank all of our classmates, friends at Daffodil International University who participated in this discussion while also attending classes.

Finally, we must respectfully appreciate our parents' unwavering assistance and endurance and well-wisher for their support and inspiration for completing this research work.

ABSTRACT

What if you are a foreigner among your own people due to not speaking in the same language as the community you're born in, living in? And even it's not your choice to make. It's just a glimpse of the harshness of an average deaf person has to go through from the very moment he was born. He is a foreigner among his own people and it's not by choice, but because of his disability. There are sign languages for such person to communicate with the world, but it's also not helping him that much to avoid the communication isolation, due to the lack of interest in learning another language by his surrounding people. From this scenario the idea of this research arose, if we could mitigate the hassle of learning sign language for non-deaf person through an AI system that could interpret sign language for him, even if it means one way, at least the life of a deaf would be much easier than now. But the first problem we faced was lack of dataset to train state of art models like YOLO family, detectron2, 3d CNN etc. There is Ishara-Lipi but it's only hand wrist images, to build a robust model we might need more variation and quantity. So, we managed to gather around 2000 dataset for Bengali alphabet and train and build basic model like VGG16, VGG19 with accuracy of 55% and 52% respectively, to more advance model like 3D CNN to detect signs in real-time and generate words to sentence and also read out loud for people who could not read.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgement	iii
Abstract	iv
List of Figures	viii-ix
List of Tables	x
CHAPTER	
CHAPTER 1: INTRODUCTION	1-5
1.1 Introduction	1
1.2 Motivation	2-3
1.3 Rational of the study	3-4
1.4 Research questions	4
1.5 Expected output	5
CHAPTER 2: BACKGROUND STUDIES	6-11
2.1 Introduction	6
2.2 Related Work	7-9
2.3 Research summary	10
2.4 Challenges	10-11

CHAPTER 3: RESEARCH METHODOLOGY	12-22
3.1 Introduction	12
3.2 Research Subject and Instrumentation	12
3.3 Data Collection	13
3.4 Statistical Analysis of Data	16
3.5 Implementation Requirements	16
3.5.1 Proposed Model	16
3.5.1.1 VGG16, VGG19, ResNet50	16
3.5.1.2 3D CNN	16
3.5.1.3 YOLO v7	17
3.5.2 Train, Test, Validation Split	17
3.5.3 Train Data Augmentation	18
3.5.4 VGG16	18
3.5.5 VGG19	19
3.5.6 ResNet	20
3.5.7 3D CNN	21
3.5.8 YOLO v7	22
3.5.9 Optimizer	22
3.5.10 Classification Report	22
CHAPTER 4: RESULTS AND DISCUSSION	23-27
4.1 Introduction	23
4.2 Experimental Result and discussion	23-24
4.3 Descriptive Analysis	24-27
CHAPTER 5: Summary, CONCLUSION AND FUTURE WORK	28-29
5.1 Summary	28
5.2 Conclusion	28-29

5.3 Future Work

29

REFERENCES

30-32

LIST OF FIGURES

FIGURES	PAGE
Fig 1.2.1: Deaf people using hand sign	3
Fig 2.1.1: Bangla Sign Language Preview	6
Fig 2.4.1: Similarity between signs	11
Fig 3.3.1: Cropping one image to create two	14
Fig 3.3.2: Workflow of data	15
Fig 3.5.1.1.1: Overview of the System for CNN models	16
Fig 3.5.1.2.1: Overview of the System for 3D CNN	16
Fig 3.5.1.3.1: Overview of the System for YOLO v7	17
Fig 3.5.2.1: Data Distribution	18
Fig 3.5.4.1: Architecture of VGG16	19
Fig 3.5.5.1: Architecture of VGG19	20
Fig 3.5.6.1: Architecture of ResNet	21
Fig 3.5.7.1: Architecture of 3D CNN	21
Fig 3.5.8.1: Architecture of YOLO v7	22
Fig 4.3.1: VGG16 Loss	25

Fig 4.3.2: VGG16 Accuracy	25
Fig 4.3.3: VGG19 Loss	25
Fig 4.3.4: VGG19 Accuracy	25
Fig 4.3.5: Training Accuracy Comparison	25
Fig 4.3.6: Validation Accuracy Comparison	25
Fig 4.3.7: Test Accuracy Comparison	26
Fig 4.3.8: Confusion matrix for YOLOv7	26
Fig 4.3.9: Precision Recall Curve	27
Fig 4.3.10: Detection Sample	27

LIST OF TABLES	PAGE
Table. 2.2.1: Comparative analysis of various models	9
Table. 4.2.1: Experimental Result Summary	24
Table. 4.2.2: Experimental Result Summary	24
Table. 4.3.1: Dataset Distinguishing for Test and Training	24

CHAPTER 1

Introduction

1.1 Introduction

Language is a systematic, traditional system of using words or signs to convey ideas to one another, whether orally, in writing, or symbolically. Over 5% of the world's population, or over 466 million individuals, have hearing impairments, according to the WHO. According to the National Census 2011, speech and hearing impairments affect 0.38% of Bangladesh's overall population. The development of an individual's identity is greatly influenced by sign language, which is thought of as the primary language for the deaf and hard of hearing. Sign language employs manual communication and body language to transmit expressions. Recognizing sign language is a developing area of study today to enhance communication with the deaf community.

Due to a lack of resources, the majority lack an education and are unable to compete in the employment market. As a result, they employ sign language, a nonverbal form of communication that relies on hand gestures and body motions to assist them communicate with others. The literature has documented excellent accuracy automated identification of American, British, and French sign languages. Despite being one of the most widely spoken languages in the world, there is little study on the recognition of Bangla sign language in the literature. The lack of a dataset for Bangla sign language might be the major cause of this lag.

This research suggests a method for BdSL recognition that can translate BdSL from a series of photos and instantly provide both written phrases. Bangla sentence and speech creation has received less attention recently than Bangla Sign Language(BdSL) categorization. We developed a dataset with 1.9k BdSL images of 39 different classes, including 39 Bangla alphabets used in standard BdSL, along with three proposed signs for generating sentences: compound characters, space, and end of sentence.

1.2 Motivation

Sign language is the basic way of communicating between listening and hearing impaired people using hand and symbolic gesture instead of sound or spoken language, while the general people do not use it. It is very hard for the general people to communicate with the speaking and listening impaired people because the sign language is not understandable for them.

Unfortunately, some persons in our society are handicapped as a result of their poor speech and listening skills. Due to their limited ability to speak, individuals must express themselves using unique signals. The general public has a very difficult time communicating with those who have speech and hearing impairments since sign language is not clear to them.

There is no international version of sign language. Even though English is the official language of both nations, it is interesting to note that the American and British sign languages (BSL) are not identical. In Bangladesh, there are 2.4 million persons who are deaf. The origins of each sign language are distinct and separate. By offering alternate means of communication, education, culture, and sports, individuals in contemporary society are attempting to improve the quality of life for the community of hearing-impaired persons. A few studies on the identification of BdSL have recently been published in the literature.

Students with hearing, speech, and visual impairments are more numerous than ever in Bangladesh's educational institutions. They can converse with one another and with other people thanks to automatic BdSL identification, sentence production, and voice synthesis. Computer vision has advanced to the point where it is being utilized to help the deaf population by enabling the sign language identification process because of the enormous advancements in technology and artificial intelligence.

For our work, we have used the finger-spelling method of sign language. Using various hand formations, fingerspelling is a method of representing each alphabet in a writing system.

These finger alphabets, commonly referred to as the manual alphabet, are widely used to educate pupils who are deaf or have any kind of hearing loss. In this situation, the primary form of communication for the deaf is word finger-spelling. Finger-spelling may also be used to compose papers using the method we've suggested.

With the device, mute and blind persons may converse. Additionally, it can aid the deaf and mute in regular conversation, which contributes to the development of a BdSL translator.

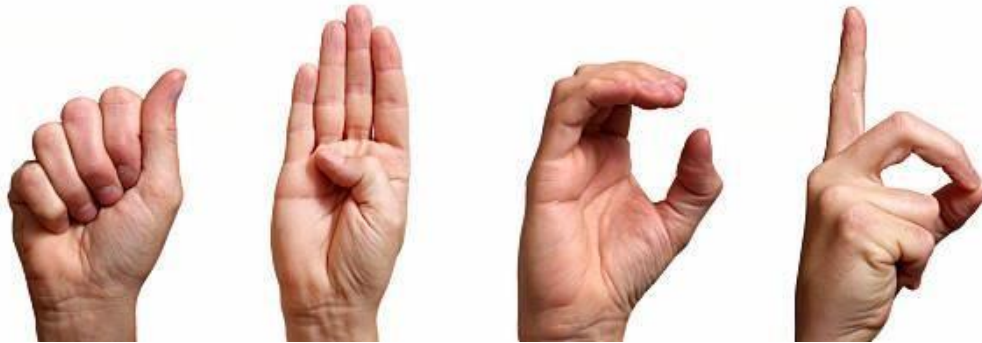


Fig 1.2.1: Deaf people using hand sign.

1.3 Rational of the study

Sign language is a global tool for communication and there is often a communication gap between normal and disabled persons, sign language is very significant. Additionally, we are aware that communication is an important component of our daily lives. However, disabled persons are not given the same opportunities (education, communication, health, etc.) as those without disabilities in our society. It will be difficult for them to engage with others, particularly from the standpoint of Bangladesh. The suggested work may be used as an interpreter for communication between normal and disabled individuals utilizing computer vision-based continuous recognition of hand-signs that spell BdSL to tackle this

issue. Additionally, this technique will assist common people in converting sign language into spoken language and disabled persons will also benefit. They can translate spoken language into sign language.

Without the involvement of every individual of the nation, Bangla Sign Language cannot flourish. This kind of concept is essential for Digital Bangladesh. The Bangla sign language recognizer model will be useful for both regular people and disabled persons. Therefore, a model to identify Bangla sign language may also assist persons with disabilities in participating in national development.

Again, the suggested approach may be utilized to provide hand-signs-based spelling recognition for distance learning teaching support. The technology may be used in the medical industry to track patients' health, particularly that of the elderly and handicapped, by identifying their gestures. With the suggested approach for simple and quick communication, the care of elderly patients may be improved by regular observation of patient behavior. People with disabilities may readily engage in any task, job, or other activity by employing this paradigm.

1.4 Research Questions

- What is Bangla Sign Language?
- What is the difference between Bangla Sign Language (BSL) and American Sign Language (ASL)?
- How Bangla Sign Language works?
- How do disable people learn Bangla Sign Language?
- How disable people can interact with the model?
- Which types of steps should be followed to help the disable people through the model?
- How can normal people interact with deaf and dumb people?

1.5 Expected Output

Basically, we offer a real-time model based on computer vision with continuous hand-spelled Bangla sign language recognition system as part of our research-based project. This system makes use of a model with 39 different classes and three suggested signals for creating sentences.

The system is trained and tested for several signers in various settings, demonstrating its signer independence and Bangla gesture-based communication. Therefore, anybody who want to create anything pertaining to Bangla sign language must go to work straight once. It will help developers in their future work. On the other hand, a system that can translate sign language into text or a phrase will be created for persons with disabilities. As a result, persons with disabilities may work together.

Although the simplicity, cheap computational cost, and high identification accuracy of the proposed system make it appealing, the signer will have trouble executing certain signs in various orientations since the camera used to record the picture is a basic, fixed camera. Therefore, developers will continue to add new features. Communication will be simpler as a result. It will facilitate communication between normal people and dumb and deaf persons. To deaf or dumb persons, in general, people may readily convey their sentiments and emotions.

CHAPTER 2

Background Studies

2.1 Introduction

In Bangladesh perspective, sign language, speech training and lip reading are the methods that have been used in Bangladesh to instruct hearing-impaired students. Despite the fact that sign language's fundamentals are universal, each nation has created its own methods in accordance with its own language and culture. There are several sign languages used in various nations, each of which has its own alphabet, phrases, and gestures. People who work in this sector get their training in several nations, each of which has its unique ways and procedures, particularly with regard to sign language.

There are many country all over the world who uses sign language like American, Arabic, French, Spanish, Chinese, Indian, and many more. A universal and more widely accepted sign language for the hearing impaired in Bangladesh has not been developed, despite the fact that the hearing impaired in each region of the nation have their own sign language specific to their requirements and regional culture. Deaf and mute persons also find it really challenging for them to learn those signs. Therefore, this translated sign language into text or a phrase for communication features, including this model, to help the handicapped communicate with non-disabled persons.

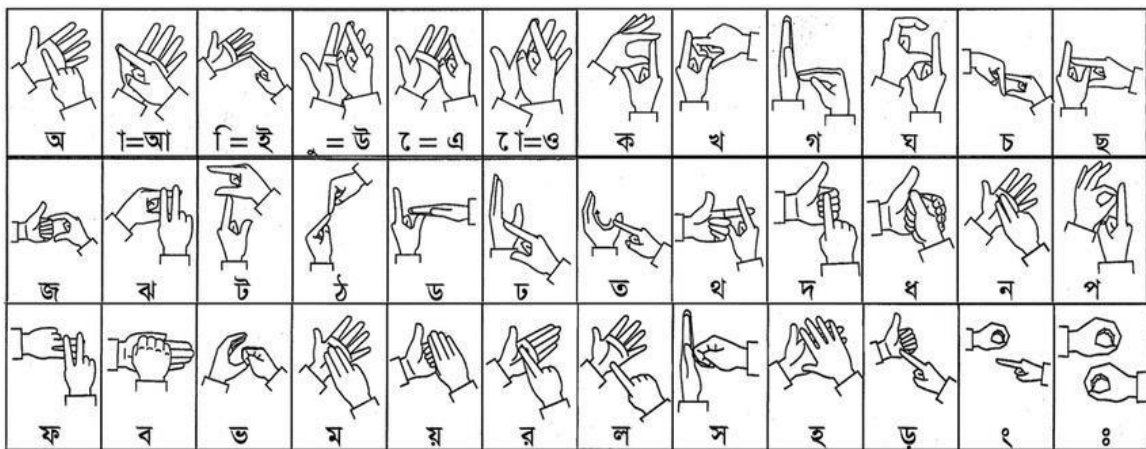


Fig 2.1.1: Bangla Sign Language Preview

2.2 Related Work

Hossen et al. [9] proposed a framework for a system that exploits the advances in deep learning to build a Bengali Sign Language Recognition system [9]. Here, Using the concept of transfer learning and fine tuning, The VGG16 network that was pre-trained (on a large dataset called ImageNet). The model has achieved the training accuracy of around 0.96 after 50 epochs with the proposed recognition system and the following results – validation loss of 0.3523 (categorical cross-entropy) and validation accuracy of 84.68%. So, This accuracy is very high considering the small size of the dataset.

Thasin et al. [10] proposed a novel architecture for BdSL recognition by combining an image network with a pose estimation network [10]. This paper refers to the proposed architecture as “Concatenated BdSL Network”. In this paper a novel model, a combination of Image Network (CNN) and Pose Estimation Network is proposed. This Concatenated BdSL Network has been trained for 30 epochs due to lack of computational resources which achieved validation score of 98.28% and overall test score of 91.51%.

Alam et al. [11] presented a convolutional neural network (CNN) architecture to recognize Bangla Sign Language (BdSL) characters [11]. The proposed model provided the directed dataset by capturing new images of variant persons in real-time. Here, 2D CNN architecture with Keras was used to build the proposed model to get better performance. This model successfully detects 36 letters and 10 numbers of BdSL with significant accuracy.

Finally, This model achieved accuracy of 99.57%.

Angona et al. [12] introduced a quick and compact method for translation of BSL alphabets into Bangla Language [12]. This paper also propose a model with a computer system that can recognize BdSL alphabets and translate them to deep convolutional neural network (CNN). A CNN method has been introduced in this model in form of a pre-trained model called “MobileNet” in recognizing 36 Bangla Sign Language alphabets.

The transfer learning approach was taken to fine tune the model to our specific classification problem. The proposed model is trained using stratified k fold cross validation scheme. The model has been trained for 50 epochs. This CNN model recognize 36 alphabets with an average accuracy of 95.71%.

Hoque et al. [13] developed a system that would recognize Bangla Sign Letters and also present a technique to detect BdSL from images that performs in real time [13]. This method introduced on CNN based object detection leads us to focus on using Faster R-CNN to detect the presence of signs in the image region and to recognize its class. Then, The training was completed with loss of 0.07538 and accuracy was about 98.2 percent in average. and the detection time was about 90.03 milliseconds. Overall, the average accuracy rate of 98.2 percent.

Yasir et al. [14] presented a learning based approach to Bangla Sign Language (BdSL) recognition using the convolutional neural network [14]. They also introduced Hidden Markov Model (HMM) to segment the sign from the time series which divided the continuous frames into the specific state. The method considered as the hand and finger specific features in out input layer of the convolutional neural network. Finally, the significant result from our basic sign expressions in a 3% rate of error where without distortion the rate reduced to 2% which is really satisfying result.

Hossain et al. [15] developed a computer vision-based model to detect Bangla Sign Language using the Convolution Neural Network (CNN) [15]. CNN is used here to recognize and classify hand image in the screen and categorize the hand skeletal features extracted from the image. To detect various sign gestures, machine learning method is used which contains training classifier with HOG structures as well as the selected CNN algorithm for forecasting targeted signs. The proposed model generated 100% testing accuracy on digits or numbers, and 97.5% testing accuracy for the alphabets. The overall accuracy gain to 98.75%.

Uddin et al. [16] presented a framework for Bangla hang sign recognition method using support vector machine [16]. In proposed system, hand signs are first converted to HSV color space from RGB image. In our work, we build two big SVM classifiers for vowels and consonants respectively with a large dataset for training and testing purpose. Above all, The Bangla Sign Language recognition accuracy is 97.7%.

Urmee et al. [17] proposed an approach with a system which works in real-time by using Xception which is trained with the “BdSLInfinite” dataset [17]. Using this dataset, a convolutional neural network (CNN) based model is trained using Xception architecture. By using the augmented dataset for the training and completed with 99.97% training accuracy and 0.0065 loss. Among all the model, the overall accuracy rate is 98.93%.

Nazmul et al. [18] developed a computer vision and machine approach for learning skills to develop a Bangla sign language gesture detection system [18]. This paper presents a simple, low-cost (BdSL) recognition model with maximum accuracy to convert signs into Bangla text. This model is based on neural networks as a deep learning method to train individual signs to achieve our goals in the suggested model. Using a CNN, they separated the dataset into two parts: training data and test data as well. By using image processing techniques, and neural networks, so raw images or videos are turned into text that can be read and comprehended with accuracy near 92%.

Table. 2.2.1: Comparative analysis of various models

Related Works	Dataset Source	Dataset Size	Status	Method	Accuracy
Shahinur et al.	Internet	4600	Real Time	CNN	99.57%
Oishee et al.	Self	3600	Real Time	R-CNN	90.03%
Progya et al.	Self	2000	Real Time	CNN	98.93%(rs 48.53)
Azer et al.	Internet	4800	Not Real Time	SVM	97.7%
Shanta et al.	Self	7600	Not Real Time	CNN	90.63%
Angona et al.	Ishara-Lipi	1800	Not Real Time	CNN	95.71%
Our proposed method	Self	1939	Real-Time	YOLOv7	0.777(mAP@.5)

2.3 Research Summary

In Bangladesh, Two hands are often used to symbolize the Bangla and English alphabets. Because of their higher level of uniformity, the Bangla sign language letters created by CDD are used as an example in this study. We illustrated a BdSL recognition model built from photos from our own dataset. Our dataset includes three newly suggested special signs that include space, compound characters, and the conclusion of a sentence in addition to 50 sets of the 36 basic sign for Bangla Sign Language.

In Bangladesh, many scholars are engaged in this area of study and are attempting to further it. They use several techniques for doing this.

First, we gathered and analyzed photographs from a variety of sources, including male and female subjects. Then, using VGG16, VGG19, ResNet, 3D CNN and YOLOv7 respectively, we separated the dataset into three parts: training, testing, and validation data.

All Bangla signals are dynamic motions at the word level. However, Bangladesh currently lacks the grammar and syntax necessary to create signs at the sentence level. The sign language of Bangladesh is very complex. Most of these signals are gestures at the word and sentence levels. BdSL uses motions and specific signals to represent sign phrases and sentences that are often hard for a person to remember. Therefore, it is necessary to create a hand-sign-spelled system using the Bangla alphabet's 36 letters and special signs (space, compound character, and end of sentence) that will work with the suggested computer vision-based recognition system to establish communication between normal and disabled people. So, we believe that this model will be quite useful.

2.4 Challenges

We are aware of the communication challenges faced by those who are physically deaf or hard of hearing. We must analyze a sizable quantity of picture data for this. Certain characters contain duplicate type signs and some data that are not static images.

Therefore, it might be difficult for humans to think about sign languages in the sense that multiple signs can be expressed using the same hand shapes or motions.

- Handling a large number of people to help.
- Manually processing a large volume of picture data.
- Determining the precise image or character size.
- Hand sign movement angle.
- Hardware (GPU) restriction.
- Customize the standard sign with the three suggested signs.

For example, although the hand forms for the two or three signals vary, several letters in Bangla sign language have the same or a very similar hand movement. Therefore, it is quite difficult for any researcher to identify a hand gesture and get categorized data for Bangla Sign. Furthermore, the hardware restriction of not having a high-configuration GPU poses a significant problem.



Fig – 2.4.1(i) - Vowel "অ"



Fig - 2.4.1 (ii) - Consonant "র"



Fig - 2.4.1 (iii) Consonant "ল"

Fig 2.4.1: Similarity between signs

CHAPTER 3

Research Methodology

3.1 Introduction

Detecting sign language is difficult but detecting it in real-time is a lot harder. So far, researchers have been working in detection of sign from images, specially in Bengali. In our approach we are trying to detect the Bengali sign language in real-time using deep learning models.

3.2 Research Subject and Instrumentation

Our prime aim was to detect signs in real-time and generate words by joining those sign alphabets to make the non-deaf community understand the language of a deaf person without learning the sign language. So far, we have been discussing the proposed method and concepts and now we will discuss the list of instruments we will be needed to get those concepts in application.

Hardware instruments:

As we have been implementing all of our system on a virtual machine, we needed to have 2 types of hardware resources , one to run the virtual machine, the hardware used inside virtual machine.

1. Local machine hardware:

- 8th gen core i3 2.2GHz with 8 GB RAM
- 256gb SSD
- Canon IXY 220F

2. Virtual machine hardware:

- 15GB Tesla T4 GPU
- 12.64 GB RAM
- 78GB storage

Software and development tools:

- Google Chrome to run Google Colab
- Google Colab as virtual machine
- Python3
- Keras and Tensorflow
- pyTorch
- Numpy
- Pandas
- Matplotlib for visualization
- Sklearn for Evaluating the models

3.3 Data Collection

We created this dataset from field level. To create this we followed some steps, which are:

A. Capturing Image

At first, we went to CDD to get the standard data sample for sign language. After getting the standard sign data we recreated those signs by our volunteers and captured their photos. Volunteers are aged from 19 to 25 and 8 female, 9 male.

B. Manually Cropping Images:

We have to crop captured images manually like 16:9 ratio images to 1:1 image. Some pictures were taken for blurry; we also have to filter out those pictures. Some of the pictures have two models in it, so we have to separate them by cropping and creating two separate images from one.

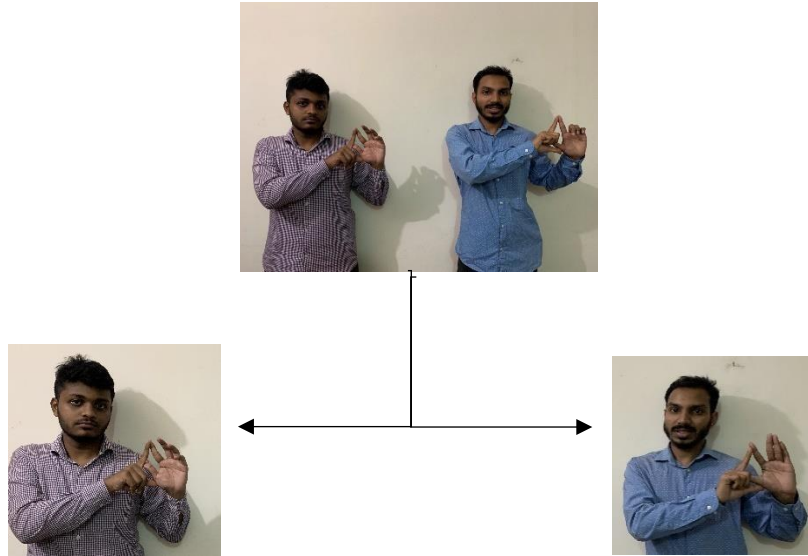


Fig 3.3.1: Cropping one image to create two

C. Data Labeling:

I. Labeling for CNN Algorithm :

We had to store each class of image in separate folders. As we have 39 classes including 36 alphabets and 3 special signs, we created 39 directories. And name the directories after the respective class.

II. Labeling for 3D CNN:

For 3D CNN we had to store 5 continuous images of same sign and same model inside a folder to treat them as 5 frames of a video clip. This way every 5 image becomes a video or single data for 3d CNN. Each such folder is stored under class labeled folders.

III. Labeling for YoloV7:

For yoloV7 we had to label image using LabelImg Software and store images and LabelImg generated corresponding txt files in respective Labels and Images folders.

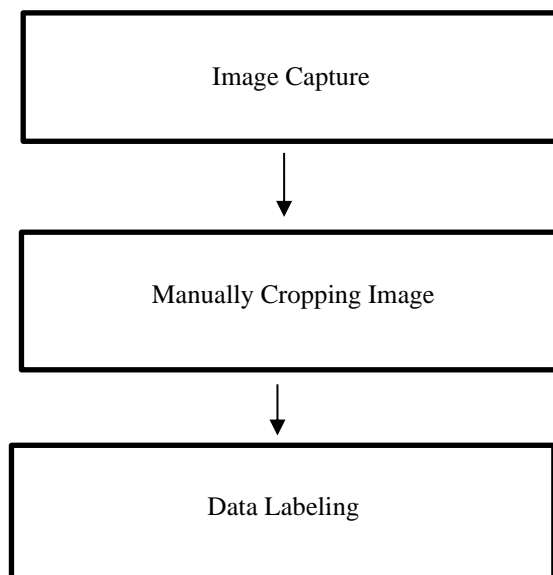


Fig 3.3.2: Workflow of data

3.4 Statistical Analysis of Data

1. All data is stored in 39 different folders
2. Total number of data 1939 for CNN models and YoloV7 and 386 for 3D CNN
3. Among 39 characters 6 Bangla vowels and 30 consonants and 3 special signs.

3.5 Implementation Requirements

3.5.1. Proposed Method

We have implanted three types of models based on input type, which are: 1. Input from static images, input from videos and real-time videos as input.

3.5.1.1 VGG16, VGG19, ResNet50:

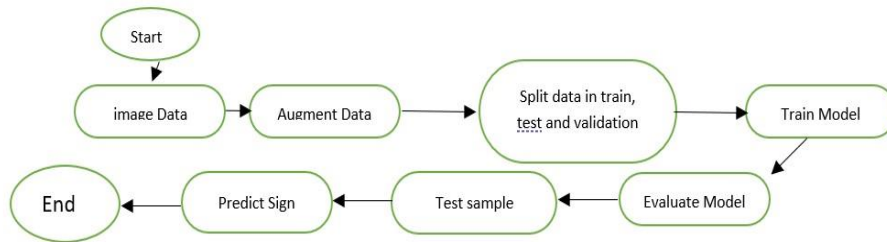


Fig 3.5.1.1.1: Overview of the System for CNN models

Our system is going take input as image and send it to CNN models with transfer learning and model detects the sign.

3.5.1.2 3D CNN:

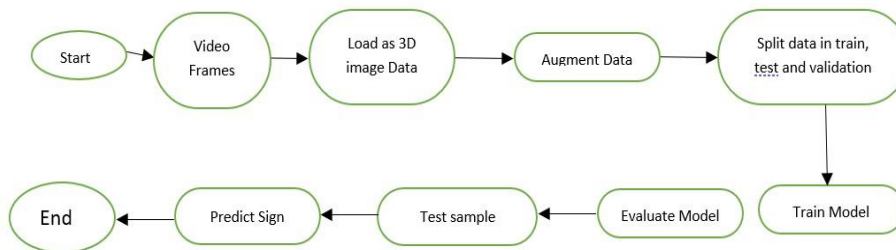


Fig 3.5.1.2.1: Overview of the System for 3D CNN

Our system is going to extract the 5 frames from video and gray scaled those and store them in 3d shape, images along Z-axis, from video input, then sent to 3D CNN model and model detects the sign.

3.5.1.3 YOLO v7:

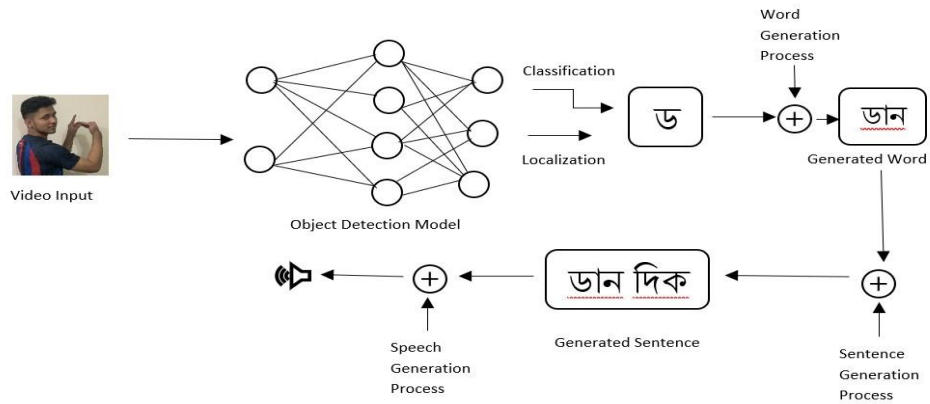


Fig 3.5.1.3.1: Overview of the System for YOLO v7.

Our system is going to extract the image from video input, then sent to Yolov7 model and model detects the sign and combined detected signs to generate word and then those generated words are combined to generate sentence and with the help of google text-to-speech API the sentence is read out loudly.

3.5.2 Train, Test, Validation Split

We've split the main data directory into 3 portion trainsets with 70% data, test set with 20% data and validation set with 10% data. We used python script to do so.

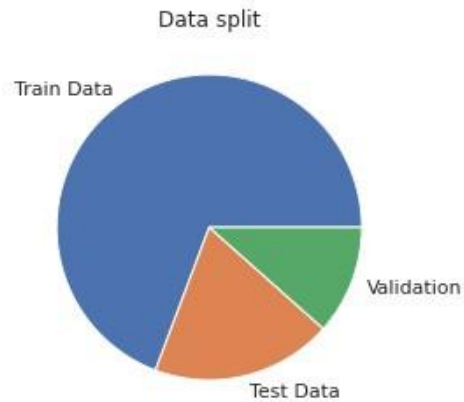


Fig 3.5.2.1: Data Distribution

3.5.3 Train Data Augmentation

We augmented the train data with shearing range=.2, zoom range=.2, horizontal flip before training to make the model more robust to new data.

3.5.4 VGG16

We have implemented VGG16 and got the best output so far. It's a CNN architecture with 16 layers deep. It's basic classification on ImageNet was 1000 classes. We trained with ImageNet weights and got the highest accuracy of 60%. Input size = 224X224.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 39)	978471
=====		
Total params: 15,693,159		
Trainable params: 978,471		
Non-trainable params: 14,714,688		
=====		

Fig 3.5.4.1: Architecture of VGG16.

3.5.5 VGG19

It's also another version of VGG16 with 19 layers deep. Trained on the same ImageNet dataset. It got us the second highest accuracy with 52%. But it has the same problem as VGG16, which is gradient vanishing. Input size = 224X224.

Layer (type)	Output Shape	Param #
input_3 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590880
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590880
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590880
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv4 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv4 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten_2 (Flatten)	(None, 25088)	0
dense_2 (Dense)	(None, 39)	978471

Total params: 21,002,855		
Trainable params: 978,471		
Non-trainable params: 20,024,384		

Fig 3.5.5.1: Architecture of VGG19

3.5.6 ResNet

To mitigate the problem of vanishing gradient we also implemented Resnet50, which is more complex network, with residual blocks. Unfortunately, we could not get more out of this algorithm. Input size = 224X224.

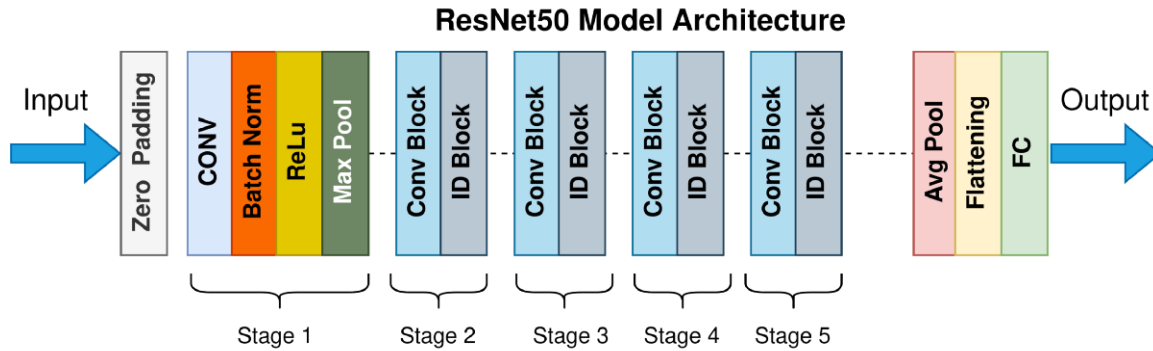


Fig 3.5.6.1: Architecture of ResNet

3.5.7 3D CNN

It's our first approach to detect signs from video, it takes input as 3d shape which is formed from frames of a video. It has only one convolution layer following a maxpooling layer and 2 dropout layers to avoid overfitting issues.

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
conv3d (Conv3D)	(None, 32, 14, 14, 4)	608
max_pooling3d (MaxPooling3D)	(None, 32, 14, 14, 4)	0
dropout (Dropout)	(None, 32, 14, 14, 4)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 128)	3211392
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 39)	5031
activation (Activation)	(None, 39)	0

```

Total params: 3,217,031
Trainable params: 3,217,031
Non-trainable params: 0

```

Fig 3.5.7.1: Architecture of 3D CNN

3.5.8 YOLO v7:

We have gained our desired outcome with YoloV7. It is the main algorithm that gives us real time detection and word generation with mAP.5 0.758.

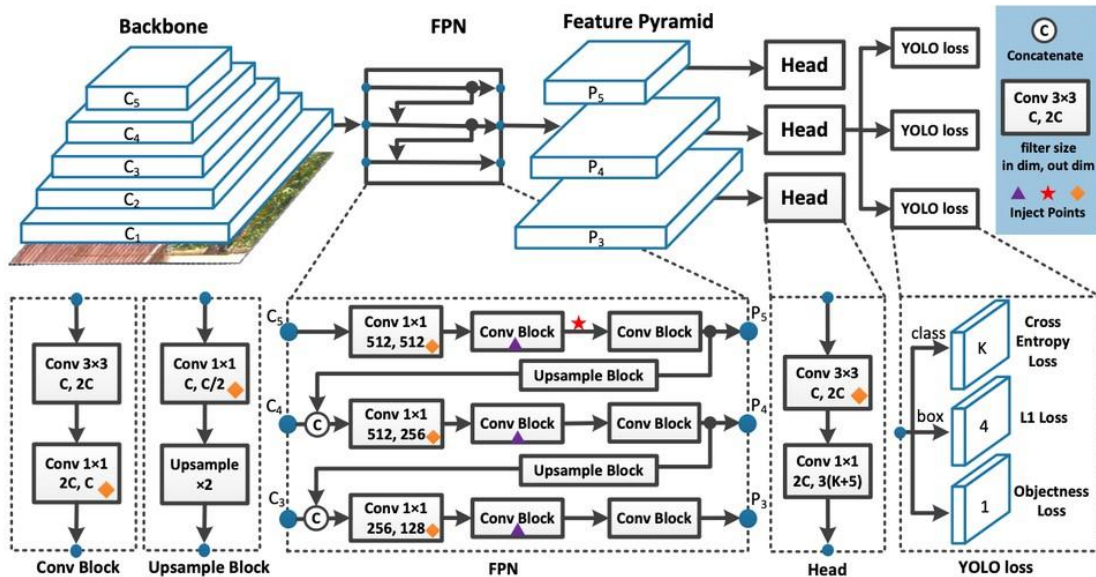


Fig 3.5.8.1: Architecture of YOLO v7.

3.5.9 Optimizer

We have used Adam as our optimizer with learning rate = 0.001 for all three CNN models and Yolo model and RMSprop for 3D CNN. Optimizer optimizes the loss function to reach the minimum loss.

3.5.10 Classification Report

We have used classification report to evaluate our models based on Precision, Recall, F1score, Specificity.

CHAPTER 4

Results and Discussion

4.1 Introduction

In this part we are going to discuss about the output of our application and how to move forward to our goal.

4.2 Experimental Result and discussion

We have evaluated our model based on some variables which are:

True Positive (TP): It means the prediction and ground truth are the same for the detected class.

True Negative (TN): It means the prediction and ground truth are the same for not detecting the class.

False Positive (FP): It means the prediction and ground truth are not the same for the detected class.

False Negative (FN): It means the prediction and ground truth are not the same for not detecting the class.

Based on above parameters precision, recall, specificity, f1-score is calculated.

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FN})$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FP})$$

$$\text{Specificity} = \text{TN}/(\text{TN}+\text{FP})$$

$$\text{F-score} = 2*((\text{Precision}*\text{Recall}) / (\text{Precision} + \text{Recall}))$$

$$\text{Accuracy} = (\text{TN} + \text{TP})/(\text{TN}+\text{FP}+\text{FN}+\text{TP})$$

Table. 4.2.1: Experimental Result Summary

Model	Training Accuracy	Validation Accuracy	Epoch
VGG16	0.96	0.51	20
VGG19	0.96	0.53	20
ResNet	0.19	0.13	22
3D CNN	0.97	.98	50

Table. 4.2.2: Experimental Result Summary

Model	mAP@.5	F1	Epoch
YoloV7	.777	0.71	50

4.3 Descriptive Analysis

As we have discussed previously, our total data has been divided into 3 portions: training, test and validation.

Table. 4.3.1: Dataset Distinguishing for Test and Training

Total Data	Training Data	Test Data	Validation Data
1939	1343	372	224

In our observation, we have found that, in VGG16 at 20 epoch training set hits 96% accuracy, while validation set only reaches 51% accuracy. The loss for both the curves become almost same at epoch 2 and later kept decreasing until epoch 20 despite of some minor spikes.

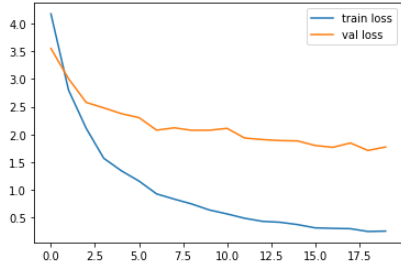


Fig 4.3.1: VGG16 Loss

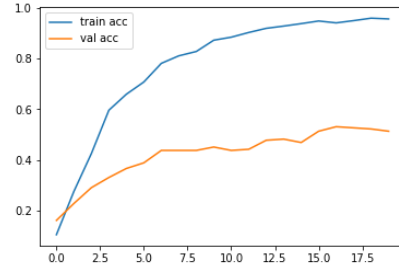


Fig 4.3.2: VGG16 Accuracy

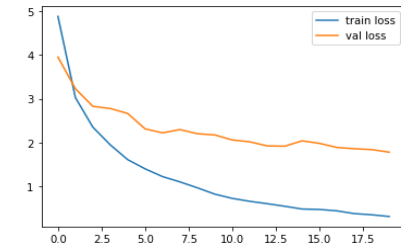


Fig 4.3.3: VGG19 Loss

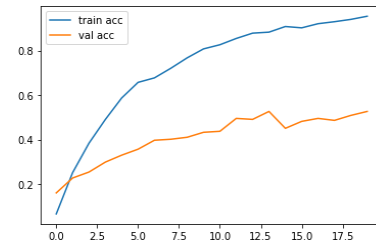


Fig 4.3.4: VGG19 Accuracy

In VGG19, we get the highest accuracy for both the curves at last epoch which are 96% and 53% respectively for training and validation data set. The lowest minimum for loss is .30 for training and 1.77 for validation set.

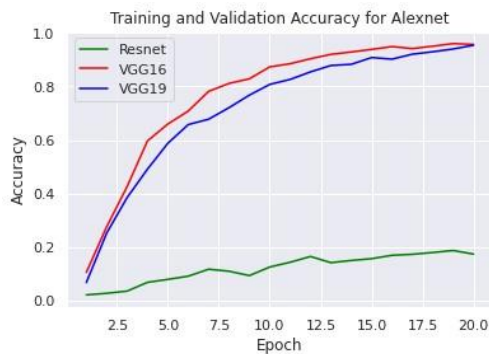


Fig 4.3.5: Training Accuracy Comparison

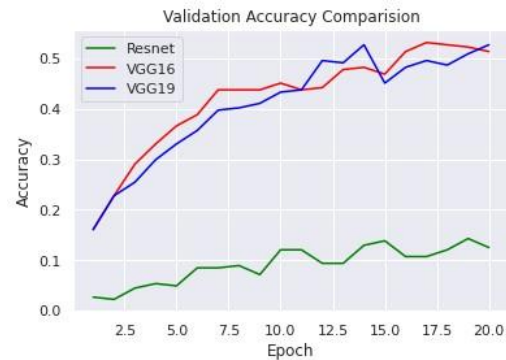


Fig 4.3.6: Validation Accuracy Comparison

Though in comparison between 3 implemented model VGG19 seems to be the accurate one based on validation and training accuracy, the test accuracy is lesser than the VGG16.

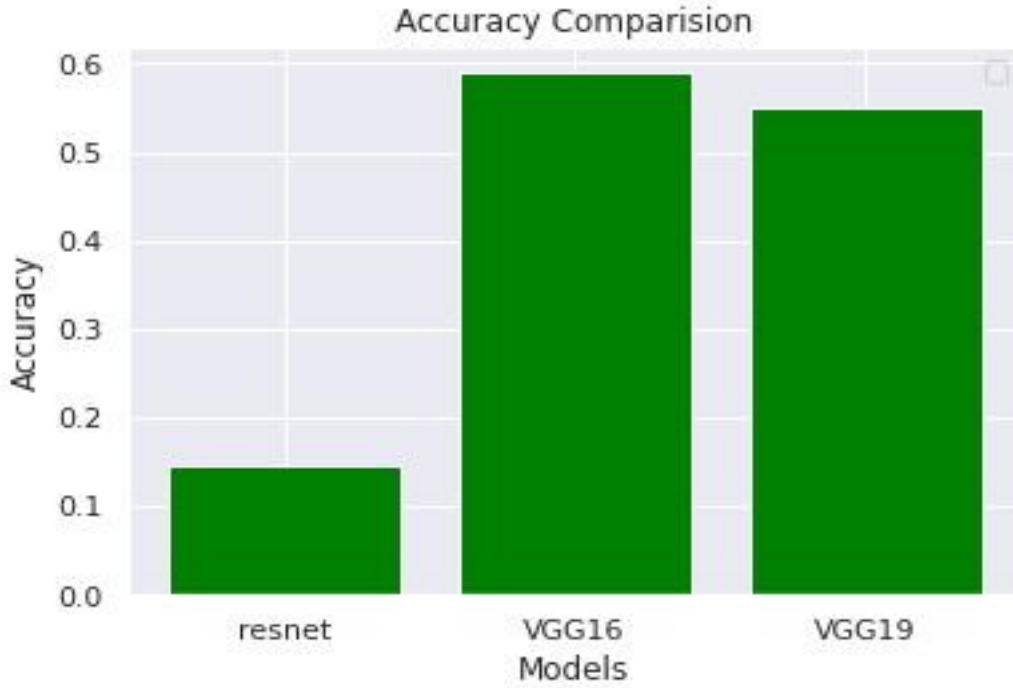


Fig 4.3.7: Test Accuracy Comparison

For Yolov7 we got this confusion matrix. Which showed us all the classes were detected accurately almost.

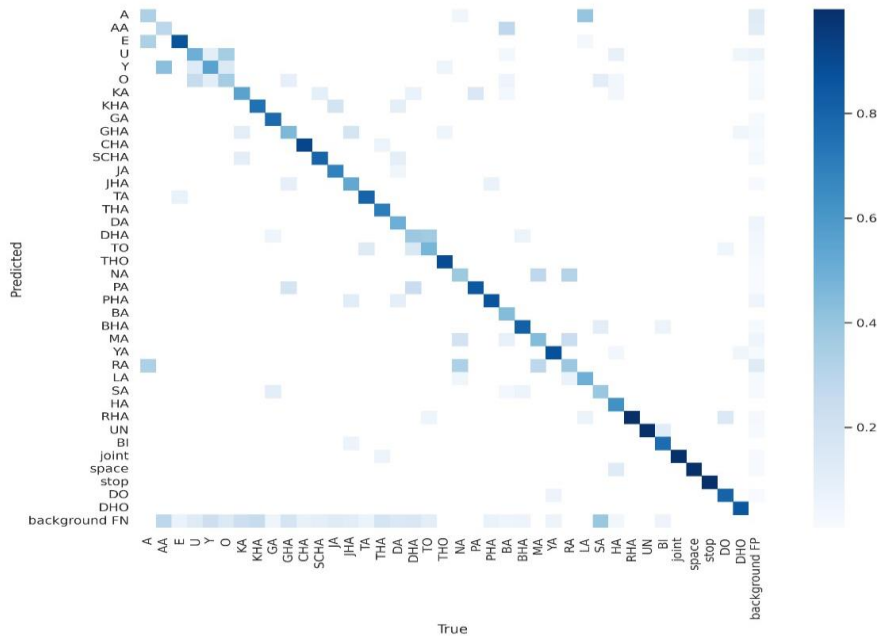


Fig 4.3.8: Confusion matrix for YOLOv7

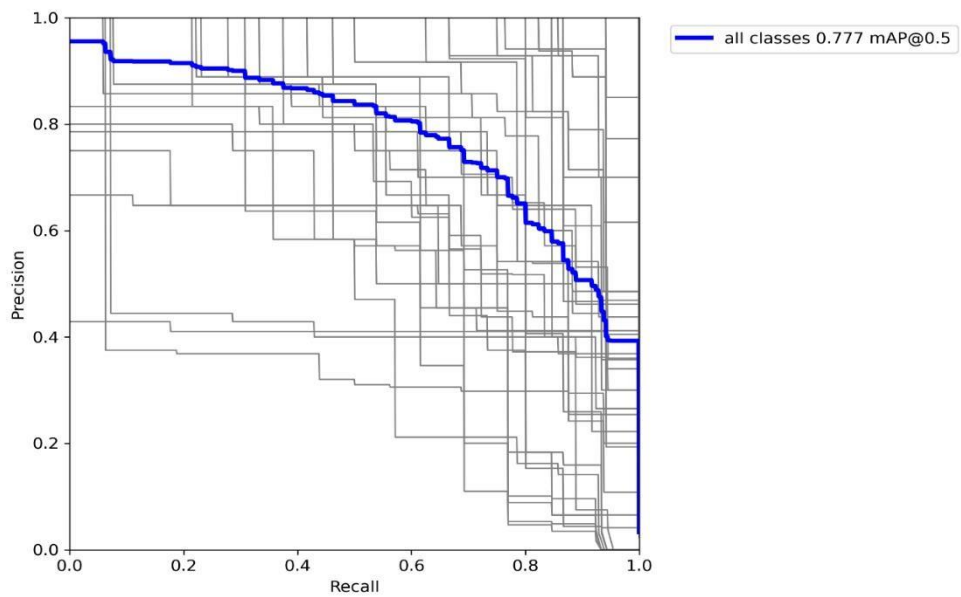


Fig 4.3.9: Precision Recall Curve



Fig 4.3.10: Detection Sample

CHAPTER 5

5.1 Summary

Our implementation of real-time Bengali sign language detection is partially successful considering the amount of data and as data collection is still going on we are hoping to achieve great accuracy in near future. Below is a brief summary of our work so far in this project :

Phase-1: Data collection, Data Pre-processing.

Phase-2: Divided data into train, test, validation set and Transfer learning.

Phase-3: Detecting the sign using VGG16, VGG19, ResNet, 3D CNN, YOLOv7.

Phase-4: Evaluating our models.

5.2 Conclusion

The more fascinating the goal of minimizing the community gap between deaf community and non-deaf community feels like, the more harder the task is as there are not much resources and scope for working for that cause in our country. The lesser the data, harder to train model and get the expected outcome. Many researchers are trying to enrich the database so that the goal could be achieved in near future and encourage people to know learn and co-operate with deaf communities without any language barriers. For our project, we have collected around 2000 static sign language data and trying to get a motion detection system so non-static sign could be recognized. Due to lack of data, we are trying to getting limited resource performances. And we would like to overcome this in near future with larger dataset.

We are hoping to reach people through this paper so could understand the scarcity of data in this sector and contribute building an enriched database on Bengali sign language. So, one day we could train more advanced model and solve this particular problem.

5.3 Future Work

We are focusing on these aspects for future:

1. Collecting more data samples
2. Deploying this model in a web app interface
3. Generating sentence
4. Create a system to both way communication

References:

- [1] Islalm, Md Shafiqul, et al. "Recognition bangla sign language using convolutional neural network." *2019 international conference on innovation and intelligence for informatics, computing, and technologies (3ICT)*. IEEE, 2019.
- [2] Das, Sunanda, et al. "A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier." *Expert Systems with Applications* 213 (2023): 118914.
- [3] Islam, Md Sanzidul, et al. "Ishara-lipi: The first complete multipurposeopen access dataset of isolated characters for bangla sign language." *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*. IEEE, 2018.
- [4] Garcia, Brandon, and Sigberto Alarcon Viesca. "Real-time American sign language recognition with convolutional neural networks." *Convolutional Neural Networks for Visual Recognition 2* (2016): 225-232.
- [5] Bencherif, Mohamed A., et al. "Arabic sign language recognition system using 2D hands and body skeleton data." *IEEE Access* 9 (2021): 59612-59627.
- [6] Martinez-Martin, Ester, and Francisco Morillas-Espejo. "Deep learning techniques for Spanish sign language interpretation." *Computational Intelligence and Neuroscience 2021* (2021).
- [7] Brock, Heike, Iva Farag, and Kazuhiro Nakadai. "Recognition of non-manual content in continuous japanese sign language." *Sensors* 20.19 (2020): 5621.
- [8] Rahaman, Muhammad Aminur, et al. "Real-time computer vision-based Bengali sign language recognition." *2014 17th international conference on computer and information technology (ICCIT)*. IEEE, 2014.
- [9] Hossen, M. A., et al. "Bengali sign language recognition using deep convolutional neural network." *2018 joint 7th international conference on informatics, electronics & vision (iciev) and 2018 2nd international conference on imaging, vision & pattern recognition (icIVPR)*. IEEE, 2018.
- [10] Abedin, Thasin, et al. "Bangla sign language recognition using concatenated BdSL network." *arXiv preprint arXiv:2107.11818* (2021).
- [11] Alam, Md Shahinur, et al. "Two Dimensional Convolutional Neural Network Approach for Real-Time Bangla Sign Language Characters Recognition and Translation." *SN Computer Science* 2 (2021): 1-13.

- [12] Angona, Tazkia Mim, et al. "Automated Bangla sign language translation system for alphabets by means of MobileNet." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 18.3 (2020): 1292-1301.
- [13] Lipi, Kulsum Ara, et al. "Static-gesture word recognition in Bangla sign language using convolutional neural network." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 20.5 (2022): 1109-1116.
- [14] Rahman, Md Touhidur, et al. "A Computer Vision-Based Real Time Bangla Numerical Sign Language Recognition using Convolutional Neural Networks (CNNs)." *ResearchGate*. <https://www.researchgate.net/publication/335292263> (Accessed Apr. 18, 2021) (2021).
- [15] Shanta, Shirin Sultana, Saif Taifur Anwar, and Md Rayhanul Kabir. "Bangla sign language detection using sift and cnn." *2018 9th international conference on computing, communication and networking technologies (ICCCNT)*. IEEE, 2018.
- [16] Hoque, Oishee Binte, et al. "Real time bangladeshi sign language detection using faster r-cnn." *2018 international conference on innovation in engineering and technology (ICIET)*. IEEE, 2018.
- [17] Yasir, Farhad, et al. "Bangla Sign Language recognition using convolutional neural network." *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*. IEEE, 2017.
- [18] Hossain, Sohrab, et al. "Bengali hand sign gestures recognition using convolutional neural network." *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2020.
- [19] Ahammad, Khalil, et al. "Recognizing Bengali sign language gestures for digits in real time using convolutional neural network." *Int J Comput Sci Information Security (IJCSIS)* 19.1 (2021).
- [20] Uddin, Md Azher, and Shayhan Ameen Chowdhury. "Hand sign language recognition for bangla alphabet using support vector machine." *2016 International Conference on Innovations in Science, Engineering and Technology (ICISSET)*. IEEE, 2016.
- [21] Urme, Progya Paromita, et al. "Real-time bangla sign language detection using xception model with augmented dataset." *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*. IEEE, 2019.
- [22] Dima, Tasnim Ferdous, and Md Eleas Ahmed. "Using YOLOv5 algorithm to detect and recognize American sign language." *2021 International Conference on Information Technology (ICIT)*. IEEE, 2021.

- [23] Wangchuk, Karma, Panomkhawn Riyamongkol, and Rattapoom Waranusast. "Realtime Bhutanese sign language digits recognition system using convolutional neural network." *Ict Express* 7.2 (2021): 215-220.
- [24] Asri, M. A. M. M., et al. "A real time Malaysian sign language detection algorithm based on YOLOv3." *International Journal of Recent Technology and Engineering* 8.2 (2019): 651-656.
- [25] Hayani, Salma, et al. "Arab sign language recognition with convolutional neural networks." *2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*. IEEE, 2019.
- [26] Katoch, Shagun, Varsha Singh, and Uma Shanker Tiwary. "Indian Sign Language recognition system using SURF with SVM and CNN." *Array* 14 (2022): 100141.
- [27] Tharwat, Gamal, Abdelmoty M. Ahmed, and Belgacem Bouallegue. "Arabic sign language recognition system for alphabets using machine learning techniques." *Journal of Electrical and Computer Engineering* 2021 (2021): 1-17.
- [28] Miah, Abu Saleh Musa, et al. "BenSignNet: Bengali Sign Language Alphabet Recognition Using Concatenated Segmentation and Convolutional Neural Network." *Applied Sciences* 12.8 (2022): 3933.

ORIGINALITY REPORT

20%

SIMILARITY INDEX

10%

INTERNET SOURCES

16%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

1

arxiv.org

Internet Source

2%

2

Md Shafiqul Islalm, Md Moklesur Rahman, Md. Hafizur Rahman, Md Arifuzzaman, Roberto Sassi, Md Aktaruzzaman.

"Recognition Bangla Sign Language using Convolutional Neural Network", 2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2019

Publication

2%

3

Nazmul Hassan. "Bangla Sign Language Gesture Recognition System", ScienceOpen, 2022

Publication

2%

4

M.A Hossen, Arun Govindaiah, Sadia Sultana, Alauddin Bhuiyan. "Bengali Sign Language Recognition Using Deep Convolutional Neural Network", 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International

2%