**ABSTRACT**

In the whole world, around 7% of people have hearing and speech impairment prob­lems. They use sign language as their communication method. People from various countries use a variety of sign languages. As an example, there are ASL, CSL, JSL, etc. Even in our country, there are lots of people born with hearing and speech impairment problems. So, our primary focus is to work for those people by convert­ing Bangla sign language into text. There are already various projects on Bangla sign language done by other people. However, they focused more on the separate alphabets and numerical numbers. That is why we want to concentrate on Bangla word signs since these people prefer to communicate using words or phrases rather than alphabets. There is not any proper database for Bangla word sign language, so we are making a database for Bangla word sign language for our work. In recogni­tion of sign language (SLR), there usually are two types of scenarios: isolated SLR, which takes words by word and completes recognize action, and the other one is continuous SLR, which completes action by translating the whole sentence at once. We are working on isolated SLR. We introduce a method where we are going to use PyTorch and YOLOv5 for our model to convert Bangla sign language into the text in real time.We are working to build a system that will make it easier for hearing and speech-disabled people to interact with the general public.

**Keywords:** Bangla Sign Language, Convolutional Neural Network (CNN), Video Classification, PyTorch, YOLOv5, Image Processing.

**ACKNOWLEDGMENTS**

By the grace of Allah who is the most merciful and the most graceful, it’s our pleasure to take this occasion to thank a few people, who have assisted, encouraged, directed and supported us throughout our Thesis program. We, Afsana Parveen Anta and Sadia Afrin want to thank our family, who has endowed their immeasurable-innumerable support and encouragement to attain this exquisite event of our life. We are very appreciative to Dr. Muhammad Hasibur Rashid Chayon, Associate Professor & Coordinator of Department of Computer Science and Engineering, IUBAT- International University of Business Agriculture and Technology for his unrelenting direction and sustain throughout the semester. We would like to pay our gratitude to our thesis advisor Md. Hasibul Islam, Lecturer of Computer Science and Engineering Department, who has given us the opportunity to make such a report not only in this semester but also throughout our education life at IUBAT by giving his valuable suggestions and advice at any time, at any situation. We would be able to make this report effectively and properly only for his right direction.

**TABLE OF CONTENTS**

[**Letter of Transmittal**……………………………………………………………………………..…iii](#_TOC_250035)

[**Student’s Declaration**………………………………………………………….……………………iv](#_TOC_250034)

[**Supervisor’s Certification**……………………………………………………….…………………..v](#_TOC_250037)

[**Abstract**…………………………………………………………………………….……………......vi](#_TOC_250037)

[**Acknowledgement**……………………………………………………………………….................vii](#_TOC_250035)

[**Table of Contents**…………………………………………………………………….…………….viii](#_TOC_250033)

[**List of Figures**...........................................................................................………………………….ix](#_TOC_250032)

[**List of Tables**……………………………………………………………………………..................x](#_TOC_250031)

**Chapter I. Introduction**…………………………………………………………………………….1

1.1 [Background and Context …………………………………………………………...2](#_TOC_250028)

1.2  [Objectives …………………………………………………………………………...2](#_TOC_250028)

1.3 [About Bangla Sign language ………………………………………………………3](#_TOC_250028)

1.4 [Bengali Sign Language Recognition ……………………………………………...4](#_TOC_250024)

* 1. [Research Objective…….……………………………………………………...…….5](#_TOC_250026)

Chapter II. Literature Review…………………………………………………………………….6

**Chapter III. Research Methodology**:…………………………………………….………………..9

3.1 You Only Look Once(YOLO)………………………………………………………. ...9

3.2 [PyTorch…………………………………………………………… ………10](#_TOC_250016)

3.3 [Sign Language Detection With Video Classification:………………… ………11](#_TOC_250013)

3.4 [Proposed Methedology………………………………………………….……….12](#_TOC_250011)

3.4.1 Dataset Preparation……………………………………………………………..13

3.4.2 Input Data […………………………………………………… ………...14](#_TOC_250016)

3.4.3 [Data Preprocessing…………… ………………………………………………14](#_TOC_250013)

3.4.4 [Data Labeling……………………………………………………………….14](#_TOC_250011)

Chapter IV. Result and Discussion………………………………………………………………16

4.1[Training Dataset……………………………………………………………….…...16](#_TOC_250008)

* 1. [Sign Detection and Text Generation………………………………………………………20](#_TOC_250006)

4.3[Accuracy ……………………………………………………………………………....22](#_TOC_250005)

ChapterV.Conclusion......................................................................................................................25

5.1 [Limitation ……………………………………………………………………………..25](#_TOC_250002)

5.2[Future Work……………………………………………………………….…………25](#_TOC_250001)

**Reference**…………………………………………………………………………………………..27

# 

**LIST OF FIGURES**

**Figure 1.1** Some of the basic numerical signs from Bangla sign language dictionary…….3

**Figure 1.2** Some of the vowels and consonants signed by two hands from Bangla sign

language dictionary………………………..…………………….…………….4

**Figure 3.1** Variation in YOLOv5 models…………………………………...……………12

**Figure 3.2** Flowchart of the working process…………………………………………….13

**Figure 3.3** Flowchart for Dataset creation……………………………………………….14

**Figure 4.1** Training process………………………………….…………………………………17

**Figure 4.2**  Mosaics and labels of train batch 0 and 1………………………….……………..……..18

**Figure 4.3** Model summary of YOLO………………………………………………………..……….18

**Figure 4.4** Prediction result of classes……………………..…………………………………..……19

**Figure 4.5** Word Classification after model training………………………………………….…..20

**Figure 4.6** Different accuracy of the words …………………………………………………………..21

**Figure 4.7** Accuracy, loss score charts for train and test sets…………………………….………..22

**Figure 4.8** Comparison between train-test average accuracy…………………………………………24

**LIST OF TABLES**

**Table 2.1** Summary of related works………………………………………………………………..8

**Table 6.1** Test Accuracy Table (5 Instance)……………………………………………………….23

**Table 6.2** Percentage of accuracy and loss in train and test data…………………………………..24

**Chapter I. Introduction**

Language is the method of communication among people which plays a very signifi­cant role in our daily life. Everyone cannot hear or talk like normal people, so they use sign language. Sign language is used by speech and hearing-impaired people to communicate with non-sign people and in between themselves. Sign language means vision-based communication through bodily movement: hand and finger movement, lip movement, and facial expression. In our society, deaf, dumb, and sign language users are neglected. It is not easy for them to communicate with normal language-speaking people. That is why they cannot keep pace with other people in any field. The recognition of continuous and real-time sign language in BdSL is challenging. Many scientists are working on developing sign language recognition systems to make them more feasible. Sign language is divided into two types: isolated, in which people communicate using single-word gestures, and continuous, in which a series of movements create a meaningful sentence. Again recognition of signs can be done either by using gloves and a sensor or by using some hardware devices like a leap motion sensor or Microsoft Kinect or by image processing from videos or images. But the first two methods are expensive, so we will approach the last method.

To recognize signs, we first need to train our algorithm with the dataset so that it can analyze and then classify and identify the signs. We will make our own dataset. Then our model inference and predict sign frm vedio sequences to the text of the normal Bangla language.

# 

* 1. **Background and Context :**

Language is the method of communication among people which plays a very signifi­cant role in our daily life. Everyone cannot hear or talk like normal people, so they use sign language. Sign language is used by speech and hearing-impaired people to communicate with non-sign people and in between themselves. Sign language means vision-based communication through bodily movement: hand and finger movement, lip movement, and facial expression. In our society, deaf, dumb, and sign language users are neglected. It is not easy for them to communicate with normal language-speaking people. That is why they cannot keep pace with other people in any field. The recognition of continuous and real-time sign language in BdSL is challenging. Many scientists are working on developing sign language recognition systems to make them more feasible. Sign language is divided into two types: isolated, in which people communicate using single-word gestures, and continuous, in which a series of movements create a meaningful sentence. Again recognition of signs can be done either by using gloves and a sensor or by using some hardware devices like a leap motion sensor or Microsoft Kinect or by image processing from videos or images. But the first two methods are expensive, so we will approach the last method.To recognize signs, we first need to train our algorithm with the dataset so that it can analyze and then classify and identify the signs. We will make our own dataset. Then we will apply video classification to convert images to the text of the normal Bangla language.

## Objectives

## Signs or alphabets have been employed by humans to communicate, open up and share their ideas and feelings with each other since the dawn of time. Different cultures use different gestures to communicate themselves, and sign language is one of them. Again, there are persons who have physical limitations, such as being deaf, mute, or visually impaired, and they must communicate in certain ways, such as by using hand signals. It is difficult for a deaf child to learn language since they are deprived of voices pronouncing meaningful words from birth.So,for making easy communication with this group of people, our main objective is to generate text from particular signs.

## 

## 1.2 About Bangla Sign Language:

Sign language is a collection of gestures that can interpret information from one to another. Around 7% of the world’s population speaks SL as their first language. About 8% of the population in Bangladesh is deaf or mute and Bangla sign language (BdSL) is used by them for communication.[12]Though the development of sign language in Western civilization began in the 17th century as a visual language, Bangladesh lacks appropriate possibilities for learning sign language with almost 30 lakh hearing-impaired people. In 1994, the Ministry of Social Welfare and the Bangladesh National Federation of the Deaf collaborated to create a Bangla sign language dictionary in order to make sign language more ac­cessible to everyone. It is considered as a language of Bangladesh by Prime Minister Sheikh Hasina in 2009, it has no legal protection and no actions have been done to institutionalize the language.[2]As sign language is needed worldwide, different sign languages are used, for exam­ple – Arabian, American(ASL), Parisian, Chinese(CSL), etc. Like any other normal language, sign language also has some rules. The structure of Bangla Sign Language (BdSL) differs from that of other sign languages based on phonetics or basic units of SL, and it is based on our country’s cultural, social, historical, and religious tradi­tions. BdSL is a modified form of British, American, and Australian sign language, as well as several indigenous signs from around the world. [4] Bangla sign language dictionary employs 36 two-handed Bangla sign alphabets: 6 vowels and 30 consonants as shown in Figure 2 and 10 basic numerals as shown in Figure 1, from 51 Bangla written alphabets. [1] According to the Bangla Sign Language manual, there are over 5000 sets of signs for the alphabet, numbers, and basic words that are used to communicate by 30 lakh speech and/or hearing-impaired persons. [13]

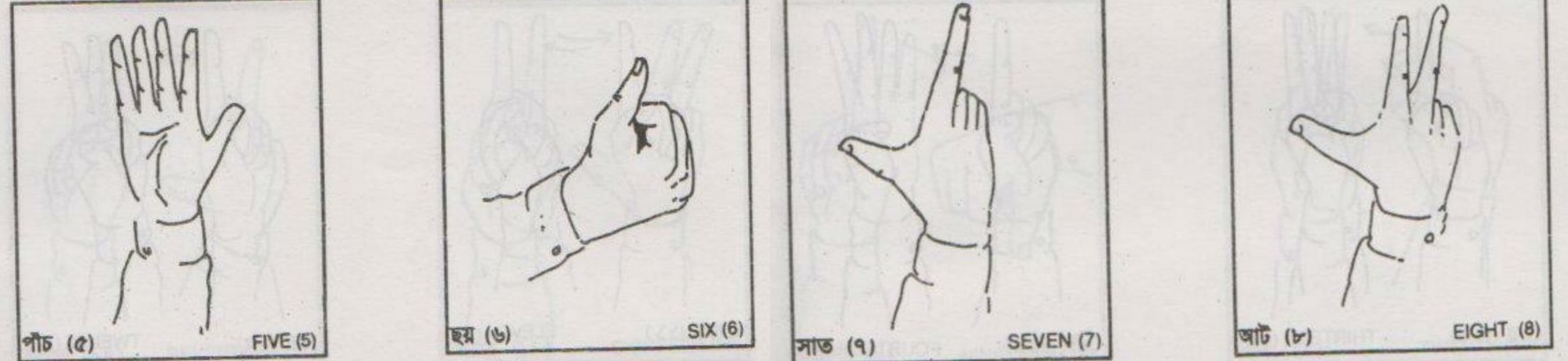


Figure 1.1: Some of the basic numerical signs from Bangla sign language dictionary.

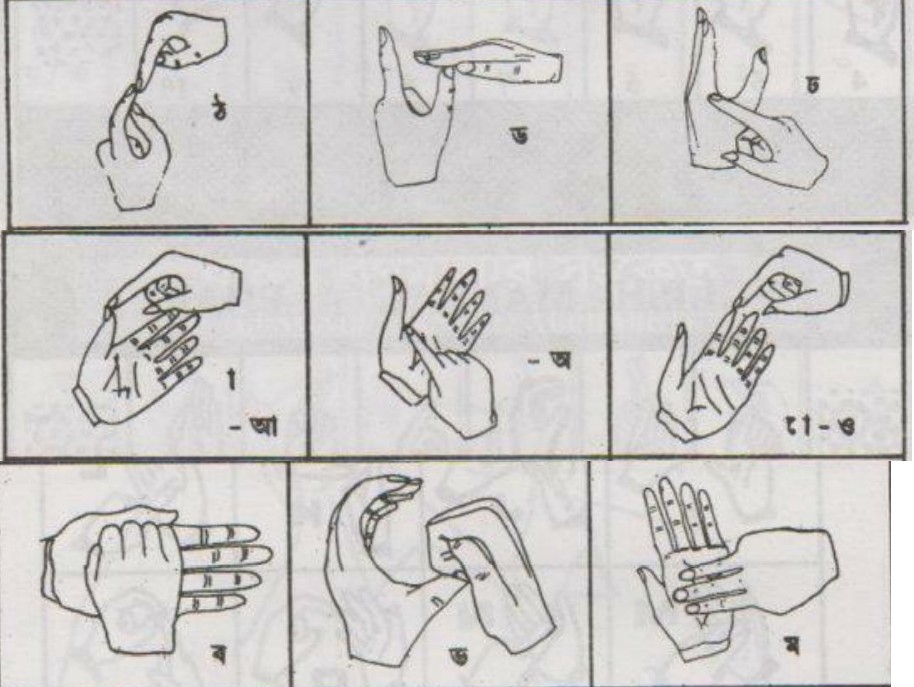


Figure 1.2: Some of the vowels and consonants signed by two hands from Bangla sign language dictionary.

* 1. **Bengali Sign Language Recognition:**

Sign Language recognition has recently gained popularity. There are numerous works on sign language in Western nations, the most of them are on ASL (American Sign Language). Despite the fact that there has been little study on Bangla Sign Language Recognition, there is currently no full Bangla Sign Language Recognition system. As a result, there is still a lot of room for advancement in the field of Bangla Sign language recognition study. The reason for this is that most BdSLR systems were designed to recognize just sign alphabets,single characters and digits. However, there are few studies on sign words and phrases, which are essential for communicating.The sentence structure Bangla sign languages must be done utilizing a word se­quence, hand forms, directions, motions, orientations, and placements relating to the human body. In this context, a dynamic sign is a moving gesture defined by a series of hand motions and formations, which may be complemented by body and face expression. As a result, implementing a recognition system for word or sen­tence level sign language is challenging. For SLR, there are several works on various algorithms and models. These are described in the section on literature review.

### Research Objective:

### While most people in Bangladesh are unaware of sign language, there are several works on the subject. Numerous publications on Bangla sign language exist, includ­ing alphabets and numerical numbers. However, we want to concentrate on Bangla individual signs since people with disabilities prefer to use words or phrases rather than alphabets to communicate for potential work. Considering that,

### We will generate a dataset of Bangla sign words using images and train It with a neural network algorithm.

### Our approach would take image inputs and extract meaningful features using our proposed method.

### We will then use our own dataset to train the model and then use the model, software, and algorithms to classify it. Therefore, as a result, a system will be developed that will make it easier for disabled persons to interact with the general public.

**Chapter II. Literature Review**

[16] This paper by Li et al.(2020)focused on a large-scale dataset. The dataset contains a total of 2000 words that are signed by more than 100 signers. They worked on American Sign Language. They worked on processing and comparing two distinct models: 2D human pose based approach and holistic visual appearance based approach. Their findings show that both models achieve relatable results on their dataset, with top-10 accuracy of up to 62.63 percent. A CNN and an RNN construct their first baseline in a visual appearance-based approach to extract the Spatio-temporal features from their given dataset that takes input from video frames. They extracted spatial features using VGG16 pretrained on ImageNet and then fed the found features to a GRU that is stacked type. In pose based approach, they mainly utilized RNN.[6] This paper written by Huang et al.(2018) stated that most of the isolated sen­sors are the basic building elements of SLR models,with data preprocessing and postprocessing layers on top of that. Temporal segmentation is a difficult task that inevitably leads to errors in subsequent steps. As far as we know, all isolated SLR models need manual tagging of each word in a sentence, which severely limits the amount of data that can be used in training. Instead of segmentation preprocessing, they employed a Hierarchical Attention Network in conjunction with Latent Space (LS-HAN). Two-stream CNN is used to extract video features, followed by LS and HAN for connecting the semantic gap. The new architecture has three elements.The paper by Tunga et al. (2021), by utilizing the data from the human skele­tal movement, they used an innovative approach that separates spatial and temporal data from the dataset and accomplishes merging in the end. Using a Graph Con­volutional Network (GCN) and BERT, their proposed architecture clearly captures the spatial interactions in the video. By modeling each found feature detail sep­arately, GCN and BERT get a wide efficiency rate of present newly formed pose based architecture.In the paper by Yin et al. (2020), he expanded on the structured approach for SLT, which may be divided into two parts: tokenization and translation, in this paper. The CSTL method analyzes sign language videos in order to create sign language glosses that capture the meaning of a variety of different signs, as well as to compare and contrast different Neural Machine Translation (NMT) architectures. They achieved superior performance in an end-to-end system that used an STMC network to extract glosses from videos.

## Table of related works:

## Table:2.1 Summary of related works

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Author Name | Dataset | Methods | Accuracy | Real-time  Detection |
| Li. et al. [16] | Total of 2000 words, signed by more than  100 signers for ASL | CNN | 62.63% | No |
| Rao. et al. [4] | 18 single shot signs for  ISL | ANN | 90% | Yes |
| Mariappan. et al. [10] | 40 Indian gestures | A machine learning method, fuzzy c-  means clustering | 75% | Yes |
| Talukdar. et al. [17] | 1250 images of dis­tinct classes, taken by  8 different people for BdSL | CNN,YOLOv4  model, CSP-Darknet53 and PANet architec-  ture | 97-  97.5% | Yes |
| Ahmed et al.[9] | 320 images of 10 sets  for BdSL | CNN | 89%,  92% | No |
| Steve D. et al.[20] | 4,547 pictures in 24 different forms for In­donesian Sign Lan-  guage | YOLO | 72.97% | Yes |
| Zaaba A. et al. [21] | Kaggle dataset with 3000 images of  Malaysian SL | Darknet frame­work, YOLO al-  gorithm. | 63.06% | Yes |
| Bhavadharshini  .et al.[22] | 500 images of Ameri-  can Sign Language | CNN, YOLO | 92.5% | Yes |

**Chapter III.Research Methodology**

This section is focused on describing our collected dataset for sign language recog­nition and the suitable algorithms and techniques which are used for training the dataset and classification process are included.

**3.1You Only Look Once(YOLO):**

YOLO algorithm is utilized in diverse applications to come across various signals, people, and specific signs by using CNN. The algorithm applies a single neural net­work to the complete image, and next separates the image into parts and predicts bounding boxes and probabilities for every part that are calculated by the predicted probabilities. YOLO can be trained with complete images and after that it increases the sign detection result. In the paper [20] they worked on both CNN and Yolo Architecture to identify In­donesian Sign Language by giving video data as input. They retrained the YOLOv3 pre-trained model by using Transfer learning which is useful for the number of chan­nels and classes. They used YOLOv3 to assist predictions on three different scales. They applied a detection layer for three feature maps of unequal sizes where each of them contained a stride of 32, 16, and 8. They claimed that the YOLO worked with accuracy on video as well as image. Again they stated that from one sign to another recognizing the transition frames on video data helped to improve the misrecognition error. Again in study[22] they stated that the system shows increased accuracy while us­ing YOLO calculation in 45 frames/second which is quite the speediest with high robustness and precision when detecting the ASL. They mentioned that by recog­nizing the occupancy of trained images, this system will arrange the input images. It can fulfill the predicted result of protest within bounding boxes. However in the research[15] they focused on observing how the YOLO algorithm can accurately detect some specific body language gestures and also whether they can utilize it in more depth to effectively find out how a person behaves. The YOLO algorithm generalizes both x and y between 0 and 1 and uses h,w,x,y as the data inputs, and they said there x and y portrayed the generalized center of the bounding box (the box around the specific part of the image used to identify) and w and h meant the width and height relative to the image.

### 3.2 PyTorch:

PyTorch is a deep learning framework which is used for imperative programming. It advances the path of production deployment from research prototyping. It is recognized as Python’s learning library, one of the broadly used Machine learning libraries. It works on CPU as well as GPU as an optimized tensor library to deep learn applications. In the paper [11] they discussed PyTorch which can do automatic differentiation and stimulate GPUs to work on instant implementation of dynamic tensor compu­tations. When compared to the other libraries for machine learning, it also balances the execution. According to them PyTorch provides an array-based programming model. They claimed that the goal of PyTorch is to maintain the interfaces of the design simple and consistent, ideally with one way of doing things by gener­ally declared design. It also models data loaders, and optimizers not complicated and effective by utilizing standard plotting,data processing tools and debugging. PyTorch library handles the basic machine learning complexities inwardly and un­expected performance cliffs. Researchers can manually execute and control the code by the tools pytorch provides that helps to identify their own execution improve­ments rather than the result the library provides automatically. They also stated that deep learning workflows are benefitted by internal executions of PyTorch as it executes extra features, observes different circumstances, and follows with perfor­mance. The features of Python are available all over the complete design execution. PyTorch balances a strict regulation between its control and data flow (i.e. tensors and the operations performed on them).

**3.3 Sign Language Detection With Video Classification Model:**

Since the dynamic nature of videos, it can be difficult to work with. Videos may be defined as a collection of frames or group of pictures ordered in a specified order. As a result, a video classification problem is similar to an image classification problem. We have to additionally extract frames from the video. In paper [21], the system re­ceives sign language image input which is then processed by the YOLOv3 algorithm. A labelled bounding box is generated when the system will find the trained image in the input. Also in [17],they used YOLOv4 that bring out the features from an input image, produces an object at various scales, and shows bounding-box regions and related classes. Pytorch model of the YOLO network was utilized in the paper [15],for hand detection. This framework will also be used for sign detection in our case. We will utilize an object detection technique, the YOLOv5s model from one of the well-known ML libraries-PyTorch Hub, in our video classification model. Transfer learning is utilized in the research [20] with the pretrained YOLOv3 model for Indonesian Sign Language demand where they used a 4:1 ratio of train set and validation set which will evaluate the trained model. We will manually generate the sets. In this case, we will choose with the YOLOv5s, the smallest and fastest

variant accessible.

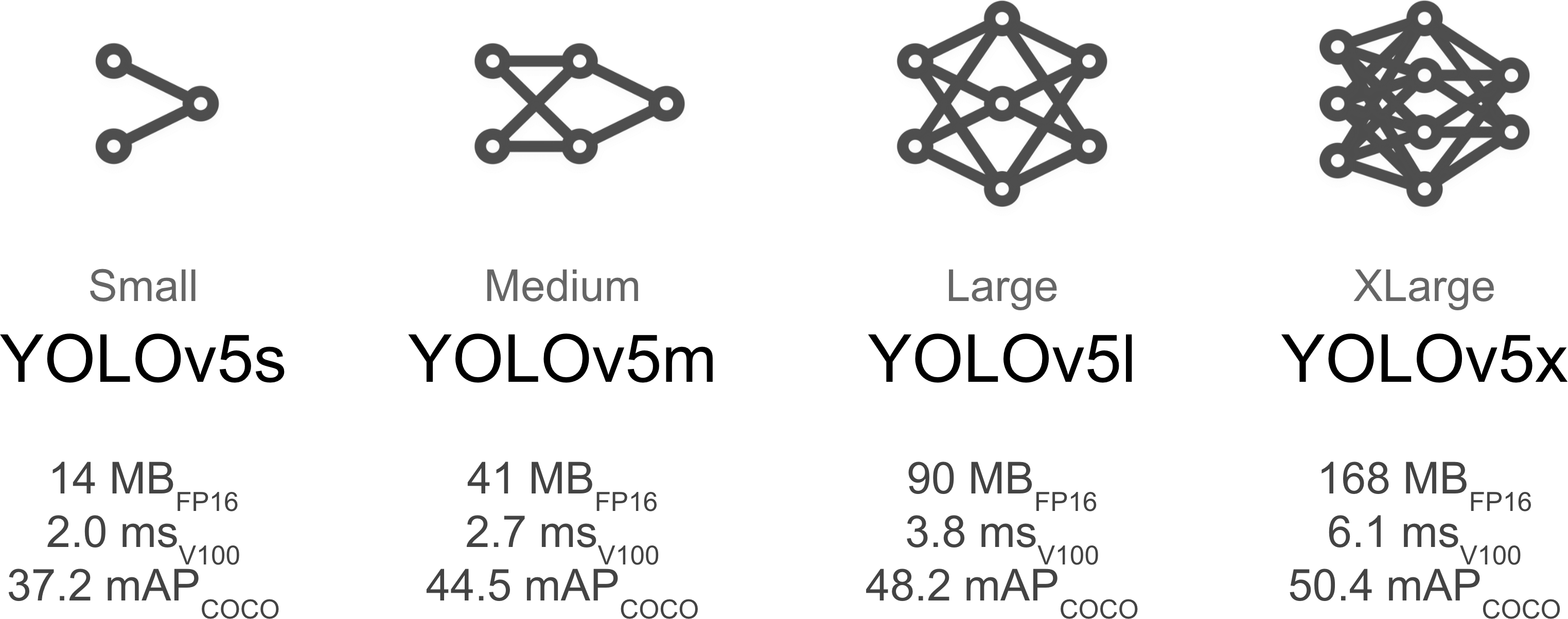


Figure 3.1: Variation in YOLOv5 models

**3.4 Proposed Methodology:**

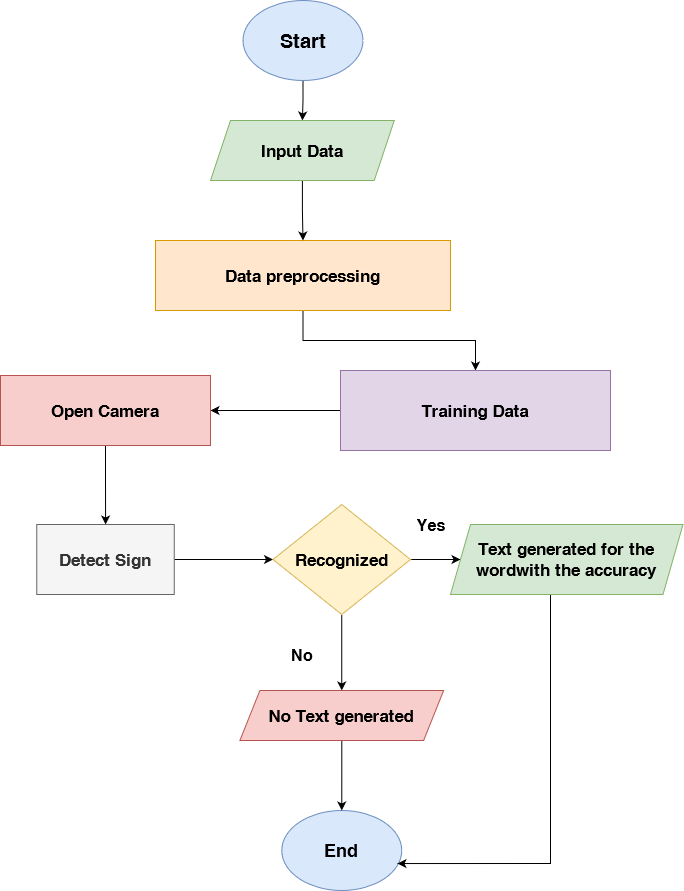
****

Figure:3.2 Flowchart of the working process.

**3.4.1 Dataset Preparation:**

The entire process of creating a dataset for our Bangla Sign Language Recognition is shown below using the flowchart:

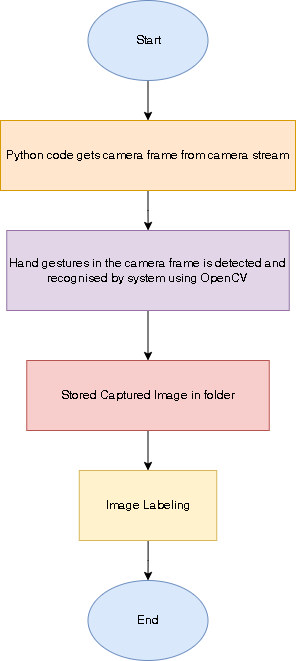
****

Figure:3.3 Flowchart for Dataset creation.

**3.4.2 Input Data:**

There are many datasets available for sign language recognition of words in various languages. However, there is no available dataset for Bangla sign language for word detection. That is why we created our own dataset in a similar approach that we read in the paper [16]. We applied openCV to capture our image of particular sign .Automatically it captured 20 pictures for a sign by frame and created a final dataset with selected frames of each word. We created a zip file to upload in the drive so that we can connect the whole. dataset with our collab file easily.

**3.4.3 Data Preprocessing:**

We processed the dataset for image labeling. Image labeling is useful when automat­ing the process of generating metadata or making recommendations to users based on details in their images. It helps in object detection. In our case, as we have to detect our hands and so we used image labeling to label the hand’s sign with the particular word that the signs represent. We also applied data argumentation on our dataset.

### 3.4.4 Data Labeling:

For image labeling, we used YOLO version 5. We created a folder to put all our selected images for our dataset. After this, we created another folder named “label” Then we opened our labeling window from the anaconda prompt and opened one frame. By clicking on the ”create RectBox ” button, we created our bounding box around our hands in the frame. In the labeling box, we named the signs that we were showing on the frame. We changed the format to YOLO. Then we saved the annotations in our ”label” folder. In the folder, two .txt files are going to be created, among which, one is a class file where the names that we gave in labeling will be saved, and another one shows various points of the frame. Those points represent the location, the center point ( x and y ), the width, and the height of our hands, respectively. We renamed our annotation file the same name as our frame (.jpg file). That is how we labeled each frame of our database one by one.

In our dataset, we have a total of 14 words of sign language in Bangla. For each word, we have 20 images, and for one word we have created five instances. In total, for one word, we have 280 frames. In our whole dataset, we have 560 frames for 14 words. Before labeling we manually had to rename all our image files from 1 to 560 for better results. We kept the image files of the same word side by side while renaming.We labeled each of them one by one. In the class file, we had the name of the frames that is what sign is shown in the frames, and in the other .txt file, we have all points that represent the height, width, location, center point x, and center point y. The label file name starts from 1 till 560. The first frame’s label file is 1, and the last frame’s label file is 560. As for the location, it starts from 0 and ends at 13. From 0 to 13 we serially label our words. One word consists of 20 frames which means in each location 20 image files exist. The location automatically increases by 1 when we add labels for new words. We used labeling for our labeling and used the default data list that is why our labeling starts from 0 instead of words it ends at 13.

**Chapter IV. Implementation and Results**

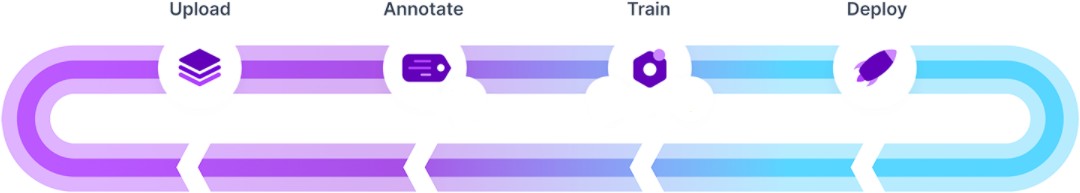
For our video classification model, we first installed all the dependencies and set up our PyTorch and GPU. Here, we extracted frames from the videos, then utilized feature extractors to take out features from pictures, and then categorized the image according to the recovered features. This model will recognize signs in an iterative process that includes gathering and organizing frames, classifying signs, training our model, and deploying it to make predictions.

Figure 4.1: Training process.

**4.1Training Dataset:**

To begin training, among our 560 images, we manually generate all of them for the train­ing and organize the directories accordingly Our SL dataset will then be used to train this pre-trained model, YOLOv5. The batch size is set to 20, the image size is set to 640 pixels, and the pretrained weights ‘yolov5s.pt’ are downloaded automatically from the most recent YOLOv5. We be­gan training with 100 epochs and increased the number progressively. The value remained steady after 1000 epochs. As a result, we trained our model for 1000 epochs which took 2 hours to complete. We employ the Stochastic Gradient Descent (SGD) optimizer with parameter groups, and momentum is set at 0.97 as a hyper-parameter. For training, we utilize Mosaic Dataloader, which merges four pictures into one mosaic where we can see train and val jpgs, labels, predictions, and aug­mentation effects.

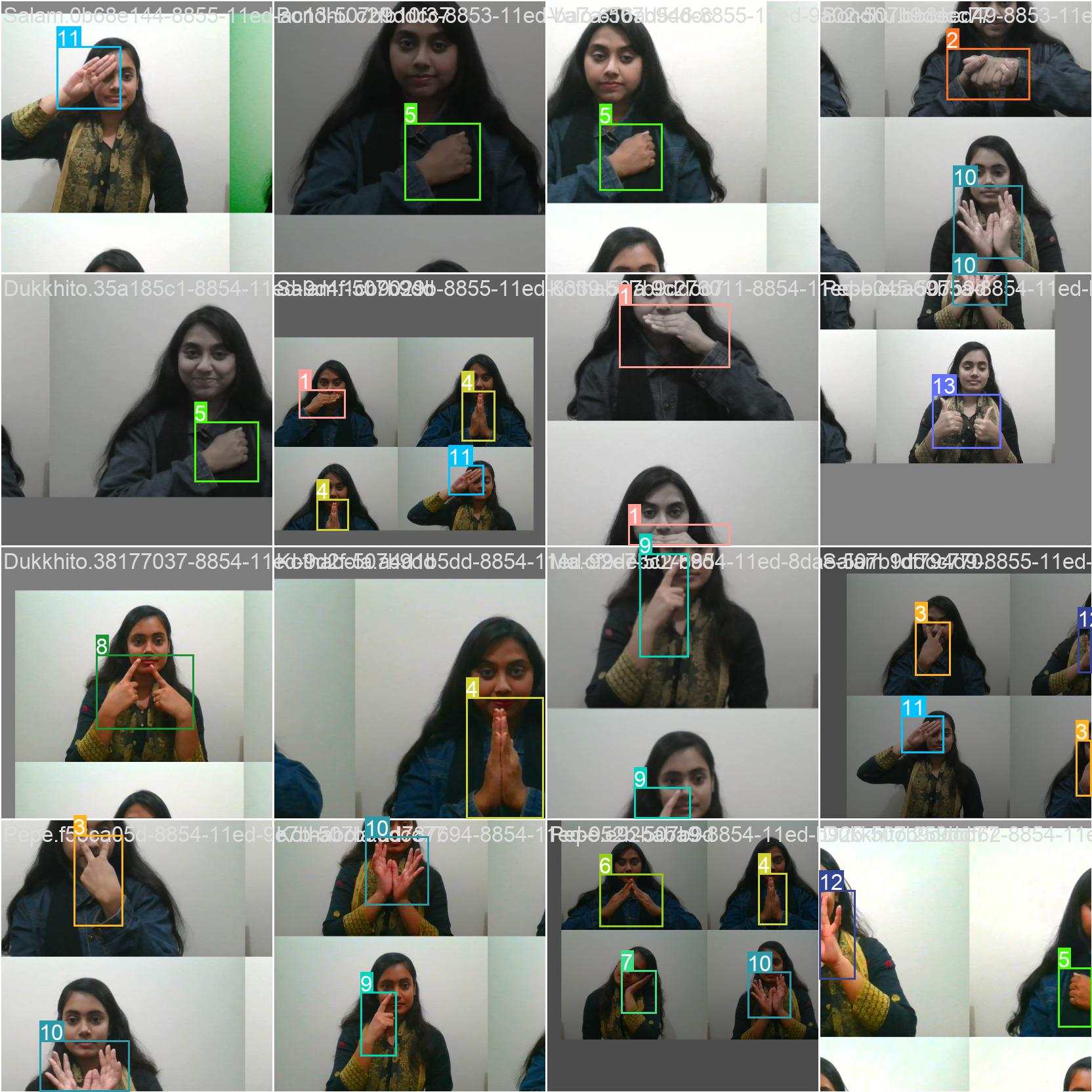


Figure 4.2: Mosaics and labels of train batch 0 and 1

Following training, we receive a custom 14-class trained YOLOv5s model with Py-Torch Hub in the runs/train/exp directory that contains all of the weights that we will utilize to generate predictions.

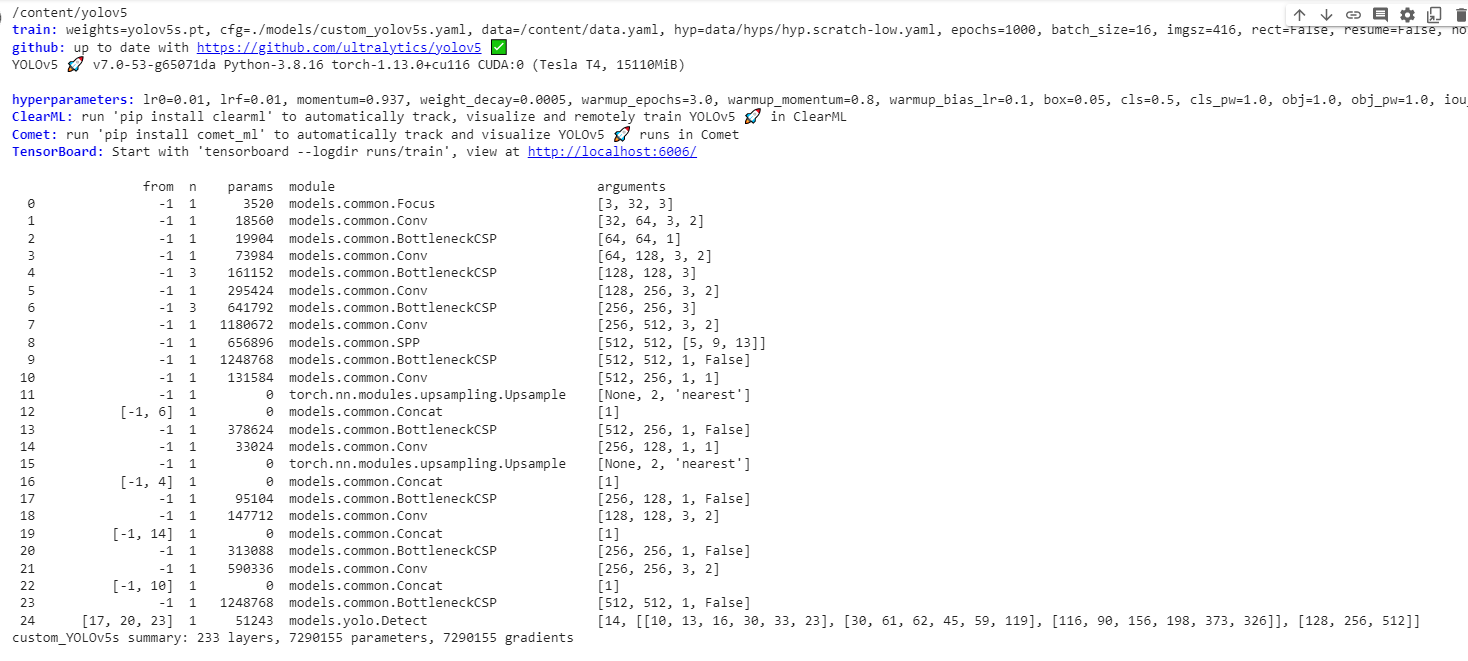


Figure 4.3: Model summary of YOLO.

As we take vedio in real time frame , each word has a varied level of precision. We can observe this in the below figures.





Figure 4.4: Prediction result of classes.

This phase is evaluated using Mean Average Precision (mAP). After this we achieve the following results for each class shown in bellow figure:

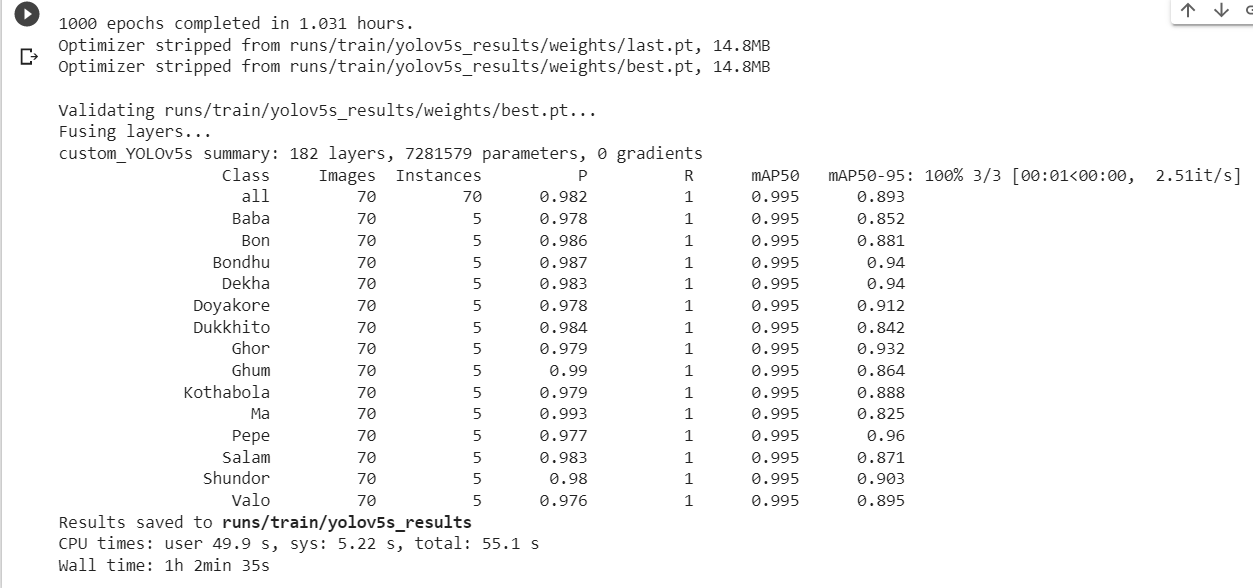


Figure 4.5: Word Classification after model training.

**4.2** **Sign Detection and Text Generation:**

After the classification is complete, detection will be performed on each frame of video data during video testing. Our model will match each frame with our dataset.These detection findings will be compared to the ground truth, and a Confusion Matrix will be created to determine accuracy .We will set the confidence threshold to 0.25.As we take the videos frame by frame, a bounding box with the name of the sign and the accuracy of recognizing that sign will be displayed in each frame of the testing video. As a result, each word has a varied level of precision. We can observe this in the below.

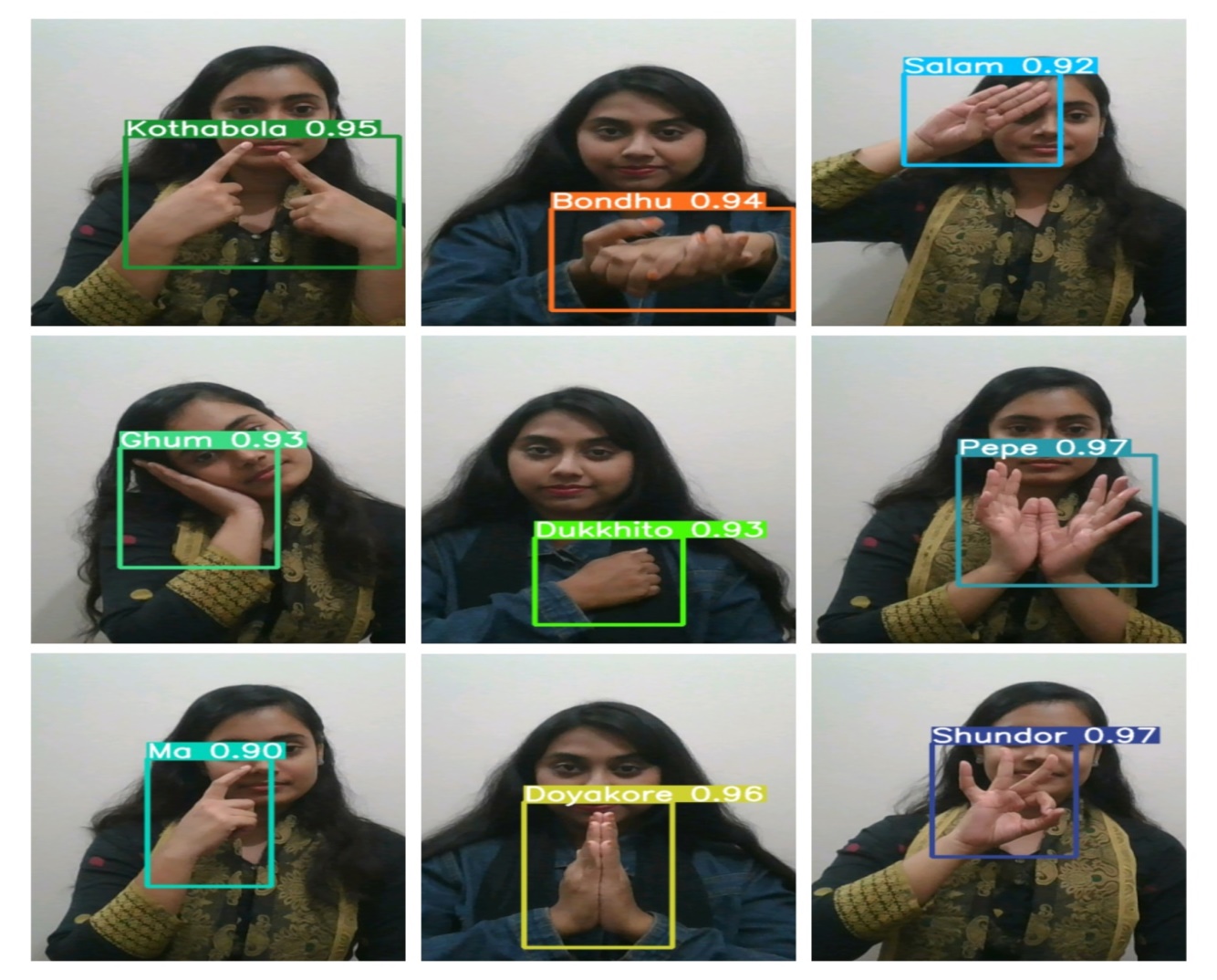


Figure 4.6: Different accuracy of the words

**4.3Accuracy:**

# Our approach recognizes signs in videos in one frame that can be identified in one second. Our training model’s overall accuracy is 90 %, while its testing accuracy is 87 %. Object detection tasks are evaluated using Mean Average Precision (mAP). The object classification task will be evaluated using the AP(Average Pre­cision) and AR(Average Recall) evaluation metrics. These are used to calculate the accuracy score and loss score. We can observe this for both train and test sets from the given charts below:

# C:\Users\ANTU\Downloads\1.png

# Figure:4.7 Accuracy, Precision, Recall, mAP and loss score charts for train and test sets.

# In both training, we achieve varying accuracy for 14 words. We take the highest ac­curacies as well as the average of these accuracies in our calculation. The outcomes of the experiment for all the words are detailed in table 3 below with 4 instances for training:

0 = no value of accuracy, data lost.

Table : 6.1Test Accuracy Table (5 Instance)

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Name of Words | Highest Accu-  racy (%) | Average Accu-  racy (%) |
| 1 | baba | 98 | 96 |
| 2 | bhalo | 97 | 96 |
| 3 | bon | 98 | 98 |
| 4 | bondhu | 98 | 96 |
| 5 | dekha | 98 | 97 |
| 6 | doyakore | 97 | 96 |
| 7 | dukkhito | 98 | 87 |
| 8 | ghor | 97 | 95 |
| 9 | ghum | 99 | 98 |
| 10 | Kotha bola | 93 | 86 |
| 11 | maa | 97 | 93 |
| 12 | pepe | 97.7 | 95 |
| 13 | salam | 98 | 96 |
| 14 | shundor | 98 | 95 |

The bar chart representing the above training data is shown below in figure

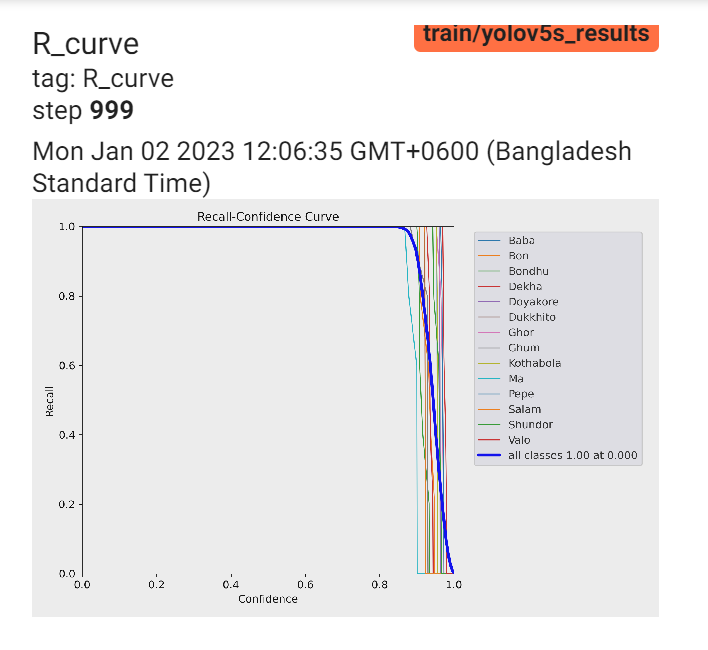
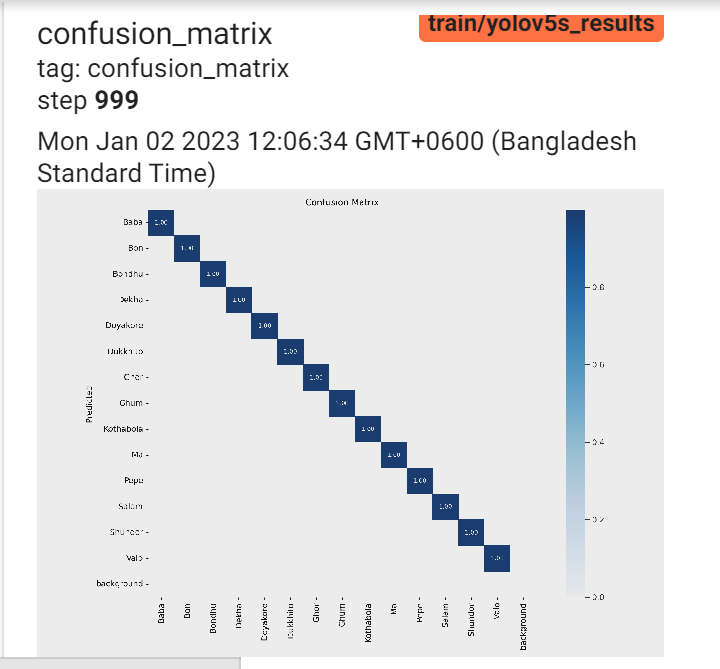


Figure 4.8: Comparison between train-test average accuracy

All of the detections provided the correct class or the classes that our model could not identify, In both the training and testing phases, we determined the percentage of classes our model was able to recognize and the percentage of classes our model failed to detect.

Table 6.2: Percentage of accuracy and loss in train and test data

|  |  |  |
| --- | --- | --- |
|  | Train Data | Test Data |
| Total words | 14 | 14 |
| Total Instance | 5 | 5 |
| Accurately Recognized | 560 | 280 |
| Recognition Percent-  age | 97.79% | 85.29% |
| Loss | 0 | 0 |
| Loss Percentage | 0% | 0% |

**Chapter V.Conclusion**

In the end, we would like to mention that all the factual assessments given by us are based on the paper we read correlates with the Bangla sign to text generation system are all different types of neural networks. The reviewed literature and working flow, we wrote, are completely based on the paper we have studied connected to our field to understand how we should approach and get a clear view of those fields which we will use for our further research. Not only that, but we have also mentioned the dataset which we used to generate a clear idea about those fields and how it works and that’s how we implemented it in our research. Basically, we built our own model where we made our own dataset for training . We have completed all the processes like taking input of the Data, labeling the data, augmentation, training and classification of the data etc. Finally, we’d like to point out that the techniques and models we discuss in this study are based on a publication we read for our research.It is a preliminary idea and needs to be more developed to get the highest accuracy output.

**5.1 Limitation:**

1.If background is not clear then accuracy might be low.

2.Sometimes low accuracy might be happened because of changing hand gestures or position.

**5.2 Future Work:**

In the future, we will endeavor to improve the efficiency of our system. As previ­ously indicated, we constructed our own dataset for word recognition utilizing 14 words. We aim to use it for sentences in the future. In other languages, such as American Sign Language, sentences can be recognized in real time, however in the case of Bangla, not much work has been done on sentence recognition in real time. As a consequence, for future work, our main motive will be to work on sentence recognition in real time. Nonetheless, our main objective always will be to facilitate the persons who have speech and hearing problems.

# Reference

* + 1. *Bangla sign language dictionary [6lkvezd3e204]*, Mar. 1997. [Online]. Available: https://doku.pub/documents/bangla-sign-language-dictionary-6lkvezd3e204.
    2. F. Z. Siddiqua, *The silent conversation*, Mar. 2015. [Online]. Available: [https:](https://www.thedailystar.net/the-silent-conversation-37929)

[//www.thedailystar.net/the-silent-conversation-37929](https://www.thedailystar.net/the-silent-conversation-37929).

* + 1. F. N. Arko, N. Tabassum, T. R. Trisha, and F. Ahmed, Apr. 2017. [Online]. Available: <http://dspace.bracu.ac.bd/xmlui/handle/10361/8216>.
    2. G. A. Rao, P. V. V. Kishore, D. A. Kumar, and A. S. C. Sastry, “Neural network classifier for continuous sign language recognition with selfie video,” *Far East Journal of Electronics and Communications*, vol. 17, pp. 49–71, 2017.
    3. C. Andrew and R. Fiona, “A survey on video classification using action recog-nition,” *International Journal of Engineering and Technology(UAE)*, vol. 7, pp. 89–93, May 2018. doi: [10.14419/ijet.v7i2.31.13404](https://doi.org/10.14419/ijet.v7i2.31.13404).
    4. J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li, *Video-based sign language recognition without temporal segmentation*, 2018. arXiv: [1801.10111 [cs.CV]](https://arxiv.org/abs/1801.10111).
    5. A. P. Jana, A. Biswas, and Mohana, “Yolo based detection and classification of objects in video records,” in *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTE-ICT)*, 2018, pp. 2448–2452. doi: [10.1109/RTEICT42901.2018.9012375](https://doi.org/10.1109/RTEICT42901.2018.9012375).
    6. S. Masood, A. Srivastava, H. C. Thuwal, and M. Ahmad, “Real-time sign language gesture (word) recognition from video sequences using cnn and rnn,” 2018.
    7. S. Ahmed, M. R. Islam, J. Hassan, *et al.*, *Hand sign to bangla speech: A deep learning in vision based system for recognizing hand sign digits and generating bangla speech*, 2019. arXiv: [1901.05613 [cs.CV]](https://arxiv.org/abs/1901.05613).
    8. H. M. Mariappan and V. V. Gomathi, “Real-time recognition of indian sign language,” *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, pp. 1–6, 2019.
    9. A. Paszke, S. Gross, F. Massa, *et al.*, *Pytorch: An imperative style, high-performance deep learning library*, 2019. arXiv: [1912.01703 [cs.LG]](https://arxiv.org/abs/1912.01703).
    10. M. A. Rahaman, M. Jasim, M. Ali, and M. Hasanuzzaman, “Bangla language modeling algorithm for automatic recognition of hand-sign-spelled bangla sign language,” *Frontiers of Computer Science*, vol. 14, 2019.
    11. P. Roy, S. M. Uddin, M. A. Rahman, M. M. Rahman, M. S. Alam, and M. S. Rashid Mahin, “Bangla sign language conversation interpreter using image processing,” in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2019, pp. 1–5. doi: [10 .](https://doi.org/10.1109/ICASERT.2019.8934614) [1109/ICASERT.2019.8934614](https://doi.org/10.1109/ICASERT.2019.8934614).
    12. T. M. Tan, A. M. Mondol, N. Nawal, and S. Ahmed, “Bangla sign language recognition using leap motion sensor,” 2019.
    13. A. Ananthanarayan, “Hand detection and body language recognition using yolo,” Ph.D. dissertation, 2020.
    14. D. Li, C. R. Opazo, X. Yu, and H. Li, *Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison*, 2020. arXiv: [1910.11006 [cs.CV]](https://arxiv.org/abs/1910.11006).
    15. D. Talukder and F. Jahara, “Real-time bangla sign language detection with sentence and speech generation,” in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, 2020, pp. 1–6. doi: [10.1109/](https://doi.org/10.1109/ICCIT51783.2020.9392693) [ICCIT51783.2020.9392693](https://doi.org/10.1109/ICCIT51783.2020.9392693).
    16. A. Tunga, S. V. Nuthalapati, and J. Wachs, *Pose-based sign language recog­nition using gcn and bert*, 2020. arXiv: [2012.00781 [cs.CV]](https://arxiv.org/abs/2012.00781).
    17. K. Yin, “Sign language translation with transformers,” *ArXiv*, vol. abs/2004.00588, 2020.
    18. S. Daniels, N. Suciati, and C. Fathichah, “Indonesian sign language recognition using yolo method,” *IOP Conference Series: Materials Science and Engineer­ing*, vol. 1077, p. 012 029, Feb. 2021. doi: [10.1088/1757-899X/1077/1/012029](https://doi.org/10.1088/1757-899X/1077/1/012029).
    19. S. Ibrahim, I. Mohtar, and Z. Ahmad, “A real time malaysian sign language detection algorithm based on yolov3,” *International Journal of Recent Tech­nology and Engineering*, vol. 8, p. 651, Apr. 2021. doi: [10.35940/ijrte.B1102.](https://doi.org/10.35940/ijrte.B1102.0982S1119) [0982S1119](https://doi.org/10.35940/ijrte.B1102.0982S1119).
    20. B. M, J. J, K. M, and S. S, “Sign language translator using yolo algorithm,” in Nov. 2021, isbn: 9781643682181. doi: [10.3233/APC210136](https://doi.org/10.3233/APC210136)