A DEEP LEARNING APPROACH FOR RECOGNIZING BENGALI CHARACTER SIGN LANGUAGE

BY

DEVJOYTI AICH
ID: 161-15-7486

ABDULLA AL ZUBAIR
ID: 161-15-6774

AND

ANTORA DEB NATH
ID: 161-15-7304

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

Supervised By

Md. Zahid Hasan
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2019

APPROVAL

This Project titled **"A DEEP LEARNING APPROACH FOR RECOGNIZING BENGALI CHARACTER SIGN LANGUAGE"**, submitted by Devjoyti Aich, ID No: 161-15-7486, Abdulla Al Zubair, ID No: 161-15-6774, Antora Deb Nath, ID No: 161-15-7304 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment 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 5th December, 2019.

BOARD OF EXAMINERS

Dr. Syed Akhter Hossain — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Saiful Islam — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

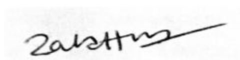
Shaon Bhatta Shuvo — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Dewan Md. Farid — **External Examiner**
Associate Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md. Zahid Hasan, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:



Md. Zahid Hasan
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Devjoyti Aich
ID: 161-15-7486
Department of CSE
Daffodil International University

Abdulla Al Zubair
ID: 161-15-6774
Department of CSE
Daffodil International University

Antora Deb Nath
ID: 161-15-7304
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to Almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Md. Zahid Hasan, Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Deep Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Syed Akhter Hossain, Professor and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate at Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

For many years, researchers are trying to recognize Bengali sign language for helping deaf-mute people which is very challenging task on the perspective of our country. Every research has its own margins and is still incapable to be used commercially. For that reason, de-vice interpreter is obligate to accommodate that deaf and hard-of-hearing community to communicate with normal people. In this paper, the main target to con-struct a model to recognize Bengali Character Sign Language using deep leaning approach. For that reason, we use Convolutional Neural Network (CNN) to train individual signs. For those individual signs, we construct a data set called Bengali Ishara-Lipi to achieve our goal. This model is trained by 5760 preprocessed images and tested by 1440 pictures. The quantitative relation of the trained and test-ed pictures was 80% and 20% severely. Finally, our model gained 92.7% accuracy to recognize Bengali alphabetical sign language. Our model will avail for commencing to make Bengali sign language device interpreter.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Sign language is a universal language for all human beings. By born humans communicate with one another by speaking or writing in an exceedingly structured standard way victimization their native languages or different languages' alphabets. However, deaf and mute those that cannot hear or speak don't have any diverse selection; however, to measure their daily lives with none various possible means that. So, language becomes the sole feasible mean for his or her communication. What is more, there's no common sign language internationally di modern nations use their sort of gestures.

Bangla Sign Language collection is that the customer once it involves Bangladesh. Bangla Sign Language (features a structure that doesn't match with the sign languages of different countries. In BSL, for expressing alphabets usually, each hands area unit used. This is why AN interpreter is essential to know what they're attempting to convey. Sign Language Recognition (SLR) has become one among the foremost distinguished fields of analysis in current times [8]. With the help of innovative technologies, it will facilitate bridge the communication gap that deaf folks have with traditional people.

Different countries have different languages. Hence, signing varies from country to state. Yank signing is taken into account as International signing among all of the sign languages. In our country, Bengali signing is often utilized by the utmost hearing and speaking impaired community. Bengali signing is formed from many-core one-handed static sign alphabet and variety signs. Several researchers are engaged in signing recognition systems for numerous sign Languages. A system developed by Pavel analyzes video clips of various gestures of sign languages taken as input and offers audio output [24].

Angles of various components of the hand with the body were calculated manually by analyzing captured pictures from input video frames and keep manually in a piece of very information with corresponding audio meanings.

1.2 Motivation

The work is done to contribute to automatic sign language recognition. We tend to area unit specializing in detecting the alphabets of Bangladeshi Sign Language. There are unit 2 main steps in our framework. Initial is that the method of extracting options from the photographs that will end in one or a lot of feature vectors that also are known as descriptors.

These feature vectors are each native, or it may be world yet. In our project, we are mostly taking into consideration the vectors, and also the second step is the classic fiction based on the feature vectors between event signs. For general image classification tasks, many kinds of Convolutional Neural Networks (CNNs) area units used as they are thought of to be a study tool once it involves supervised feature learning and classification, thus we tend to area unit mistreatment CNN to implement our system.

1.3 Future Scope

Digitizing linguistic communication and mistreatment it as a program to assist create the communication gap between deaf and traditional individuals narrower. Alternative countries have already created their linguistic communication digital; however, BSLA has not been digitized.

Few others before us have worked on sign recognition on the total Bangla Sign Language; however, we want to develop this. By creating this program, we tend to guarantee that as few as potential kids feel as outcasts by society, thanks to communication difficulties.

This analysis is focusing on assisting the hearing and speaking impaired individuals by developing a system by that signing is understood to other individuals each regular and trial and speaking impaired people.

This project focuses explicitly on Bengali signing. This project can produce a straightforward communication scope between a standard person and the person who cannot speak or hear; however, each has to be compelled to perceive Bengali Language and have to be forced to be ready to acknowledge Bengali alphabets and numerals.

We hope our system can cut back the communication gap of the hearing or speaking impaired individuals with others. Hopefully, it'll produce additional scopes for special youngsters to specific themselves. By victimization, our system, hearing, or speaking impaired youngsters might raise further queries, so learn others. Victimization of this technique, we hope, negligence towards special youngsters or those that cannot speak or hear can cut back in our society.

1.4 Overview

We have researched different sorts of linguistic communication interpreters or devices for the linguistic communication interpretation system. Currently, we tend to are attempting to determine a system that interprets Bangla linguistic communication with people.

This has been exhausted 2 phases. First, from our dataset, it takes the pictures and applies skin detection formula and detects the complexion pixels from it, then makes it binary. From this binary image, we'll extract options employing a bag of options technique and train the system victimization of the support vector machine classifier.

This paper is organized as follows. In chapter-02, the background study for this technique has been delineated. The projected model, features, method square measure enclosed in chapter-03. The Result analysis, applied math analysis with graphical pictures square measure shown in Chapter-04 and that we over the paper in Chapter-05.

CHAPTER 2
RELATED WORKS

2.1 Sign Language History

In western society, linguistic communication started developing within the seventeenth century for a visible language. It's a style with a mix of standard gestures, hand signs, signing conjointly includes the position of the hand positions to represent a meaning line. The primary Yankee faculty for the deaf was supported in 1817 by Laurent Clerc and Thomas Hopkins educator [1]. They're aforesaid to be the principal founding father of Yankee linguistic communication. This is often part genuine. Laurent Clerc was from Europe and educated French linguistic communication. Thomas Hopkins educator brought Clerc back to America to begin the primary Yankee faculty for the deaf. Like archimandrite Charles Michel American state L'Epee's faculty, kids from everywhere the country traveled to attend this faculty, transfer their home-signs with them. These home-signs, combined with French linguistic communication, became Yankee linguistic communication. Before the 19th-century, sign languages were confined solely to fastened words employing a signing system [29]. It evolved step by step, and currently, several forms of analysis square measure happening for decoding period sign languages to create its understanding easier for everybody. Some inventors acclaim grouping because of the originators of the primary linguistic communication. This is often, in all probability, true. Early man, before auditory communication, in all likelihood, used gestures. They presumably patterned and created signs for those things they couldn't speak.

2.2 Various Sign Language Recognition

There have been loads of analysis on language interpretation. Most of the works are in severe trouble sign language (American Sign Language or ASL) [30]. Differing kinds of techniques were accustomed to creating the model. From GE's R.H. Sapat faculty of Engineering, Bharat, there was a paper printed on "Intelligent language interpretation." They planned to form the machine to perceive human language and that they additionally build AN HCI (Human-Computer Interface). The Human-Computer interface will recognize the voice, facial expression, and, therefore, the hand gesture of humans. They in the main targeted on machine learning programming and example matching for higher output generation [17].

2.2.1 Real-Time Based Bare Hand Gesture Recognition

In 2013 Kashmera Khedkkar Safaya and faculty member. (DR.). J.W.Bakal of Bombay analyzed of the time primarily based on a clean hand gesture. During this system, they used a dynamic vision detector camera for recognizing pure hand gestures. DVS that's Dynamic detector Camera is utterly different from a traditional camera [12]. The distinction is DVS camera solely responds with pixels with temporal light distinction. That reduces the process price of consecutive scrutiny frames to trace a nonstop object. At first, they determined the delivery purpose track range for every frame and checked if the quantity of frame events is a smaller amount than the given threshold. Then this delivery purpose is employed to observe the structure and acknowledge the hand position [7]. Within the extraction section, they calculated the articulation plan purpose and calculable the breadth of a hand together with the horizontal axis and so update the modification in thickness from right to left with their own rule.

2.2.2 Glove Based Language Gesture Recognition

Glove based mostly gesture recognition model was projected by St. Christopher Lee and Yang sheng Xu. They were ready to acknowledge fourteen letters from hand gestures, and also, the machine may learn the new model and update the model for every gesture with a 10Hz rate. However, then there have been tons of glove devices was designed like Power glove, Sayre Glove, and deft Hand master, etc. In 1970, Zimmerman used the VPL information glove to see sign language [9]. This was the first winning out of their glove. It's engineered on original fiber sensors on the rear of the fingers. Star-near and Puntland established a glove-environment system which will determine solely forty yanks signing with a rate of 5Hz. Lam T. Phi, Hung D. Nguyen, and T.T. Quyen Bui of Vietnam Academy of Science and Technology, Hanoi, Vietnam, analyzed Vietnamese signing in 2015 [8]. They connected ten flex sensors and one measuring system. They used the flex sensors for choosing up the arch of fingers, and also the measuring system was used for the police work movement of the hand. Counting on the hand's poses, i.e., vertical, horizontal, and action, the signing of letters of the Vietnamese alphabets are often divided into classes one, 2, and 3, correspondingly. Firstly, the hand's posture is acknowledged. Next, if the hand's position fits to whichever class one or class a pair of, an identical rule is employed to find a letter. If the posture belongs to level three, a dynamic time deformation rule is applied.

2.2.3 Copycat: Interactive American Sign Language Game

Copycat could be a game that's designed for two functions, one is to gather gesture knowledge for yank signing (ASL) recognition system, and two is to make a scope of usage to assist deaf youngsters in acquiring language skills once they play the sport. The system is consisting of a video camera and wrist-mounted accelerometers because of the primary sensors. The theme of the game, Iris, the cat, communicates with the user with American sign language [30]. Mortal is meant with a restricted, age-appropriate phrase set. Whereas taking part in the sport, if a baby cannot give the proper sign or the child's sign is poor, Iris, the cat appearance nonplussed and told the kid to do once more. If the kid is ready to provide a robust indication, Iris, the cat acts consequently. The user may select choices (objects shown as icons) to instruct Iris. If the kid cannot keep in mind the proper phrase to direct Iris, the kid will click on the button bearing the image of any object. The system shows a brief video with a tutor demonstrating the proper American sign language phrases [22]. Then the kid will copy the teacher to speak with Iris. Mortal has used a "Wizard of Oz" approach wherever the interpreter simulates the pc recognizer. This technique permits analysis into the event of Associate in Nursing acceptable game interface and conjointly knowledge assortment to coach the Hidden Andre Markoff Model (HMM) based mostly American sign language recognition system.

2.2.4 Using PCA to recognize hand signs in real-time

Authors S. N. Sawant and M. S. Kumbhara mention in their analysis paper, Real-Time signing Recognition exploitation PCA, that they extracted hand sign options of Indian signing exploitation the Eigenvectors and Eigenvalues from the photographs and then exploitation Principal Element Analysis (PCA) algorithmic rule for gesture recognition and conversion to text [10]. The photographs area unit smitten a digital camera, whereas there's a white background behind the hand for ease in the process. Segmentation and morphological sterilization ways area unit applied for separating the hand object from the environment. For segmentation, the Otsu algorithmic rule is employed, whereas the morphological sterilization ways denies the image and provide a consistent contour. Here, the PCA algorithmic rule is additionally used to reduce spatial property and to extract options. Then the removed values area unit compared with the dataset of 260 pictures of 26 hand signs and corresponding matches are displayed as text [18].

2.2.5 Bracelet and Rings translate Sign Language (Leap Reader)

Here this LEAP READER may be a device that helps, those that don't perceive signing, to know them. This device is underneath development. The abstract method may be a combination of a hoop and a bracelet. Each bracelet and, therefore, the set of rings area unit detachable [33]. One must wear them along to sight the motion of the fingers. Ring captures the action of the fingers and sends a proof to the bracelet, and so the bracelet interprets the gestures into loud sound or text message through the inherent speaker and screen on the bracelet. By this device not solely people, however conjointly special people will perceive sign languages. This idea was fictional by Cao Zu-Wei, Hu Ya-Chun, Huang Ching-Lan, dynasty Po-Yang, Tsai Yu-Chi, and principle Yi-Hsien impressed by Buddhist rosary. This project may be a 2013 Red Dot style Award winner [3].

2.3 Bangladeshi Sign Language

Although in western society, this language was adjusted for an extended time along in the People's Republic of Bangladesh in 2000, CDD (Center for Disability in Development) took the step to order this communication method. About 2.6 million individual’s area unit is deaf in the People's Republic of Bangladesh [15]. Bangla linguistic communication user’s community is the largest community among the language-based minority communities in the People's Republic of Bangladesh. The Bengali Language is that the fifth most generally used orthography language in terms of population. There are a unit eleven vowels that area unit referred to as a "sôrôbôrnô" and thirty-six Consonants that is termed "bænjônbôrnô." For of these alphabets, their area unit separate signs and conjointly for meaningful words [15].

2.4 Bangla Sign Language Recognition

Sign Language recognition becoming very popular in the current research field. But most works are done in ASL (American Sign Language). In the research field, Bangla Sign language recognition is still not used widely [26].

2.4.1 Employing Neural Network Ensemble

By Bikash Chandra Karmakar, Kazi Md. Rokibul Alam and Md. Kibria Siddiquee of Khulna University of Engineering and Technology, Khulna, Bangladesh, did analysis using Neural Network Ensemble. In Pre-processing, they took input from a digital camera and detects the coloring of hand. The coloring of the side has been unbroken distinctive within the setting to make sure uniform detection. Then in the image process, the captured image has been reborn into its threshold worth. Then the reborn image has been normalized to 30x33 scale pixels by applying the standardization method. Then the feature extraction technique has been used, and also the NCL rule has been wont to train these pictures [21].

2.4.2 Normalized cross-correlation

By Kaushik debutante, state capital Parvin Mony & Sujan Chowdhury of port University of Engineering and Technology, Chittagong, People's Republic of Bangladesh did analysis victimization Normalized Cross-Correlation for two-handed language recognition [31]. They projected this model in 2 steps. Frist associate RGB (Red-Green-Blue) color model is enforced to work out heuristically threshold worth for adopting human regions. When the human regions area unit obtained by applying color segmentation, then actions for purification of the candidate region area unit followed by victimization two different color radiocarpal joint band regions and elucidative. Finally, a statistically-based model matching technique is employed for the popularity of hand sign regions. Numerous hand sign pictures area unit want to check the projected technique and results area unit given to supply its effectiveness [19].

2.4.3 Automatic Recognition of Bangla Sign Language

In that project, they understood Bangla signing, victimization Kinect for capturing pictures, and used a neural network for coaching. Their initial goal was to spot isolated signs from movements of hands. They failed to embrace any indication of the Bangla alphabet. For pre-processing, they used the OpenNI framework then used Artificial Neural Network in MATLAB for coaching [26].

CHAPTER 3
METHODS AND MATERIALS

3.1 Planned Model

The flowsheet of our planned model, shown in Figure: 3.1, at first checks, if relevant images area unit is accessible. If pictures area unit convenient, the system reads and separates the images from the video and crops out the world of interest. The world of interest is then resized to match the resolution of the coaching information set. Once cropping the image, it then sent to the origin v3 model. The picture with the best prediction accuracy is picked. The ensuing character is shown as output.

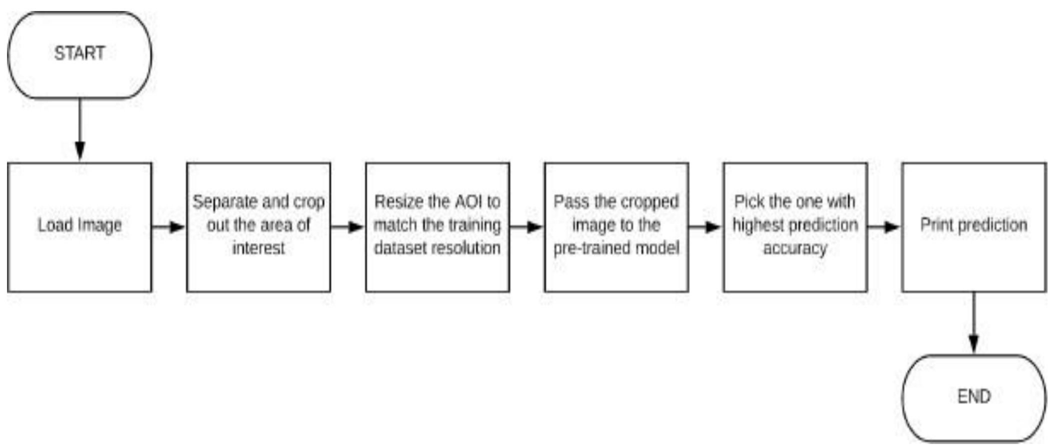


Figure 3.1: Working Flowchart.

3.2 CNN Method and Material

When the terms come to Machine Learning, Artificial Neural Network done very well, there is various use of Neural Network, most likely image, audio, word reorganization, etc. as well as. There are many types of Neural Networks, and these are used for various platforms. Now we are going to discuss the Convolution Neural Network that we have been using in our project.

3.3 Layers in CNN

There are four layers in CNN. Those are,

- Convolution layer
- Activation layer
- Pooling layer
- Fully connected layer

Now we are going to discuss the layers.

3.3.1 Convolution Layer

At first, we have to extract features from the input image. Then we have to learn image features with the help of small squares of input data. Convolution preserves the relationship between pixels — two inputs such as image matrix and a filter or kernel for mathematical operation.

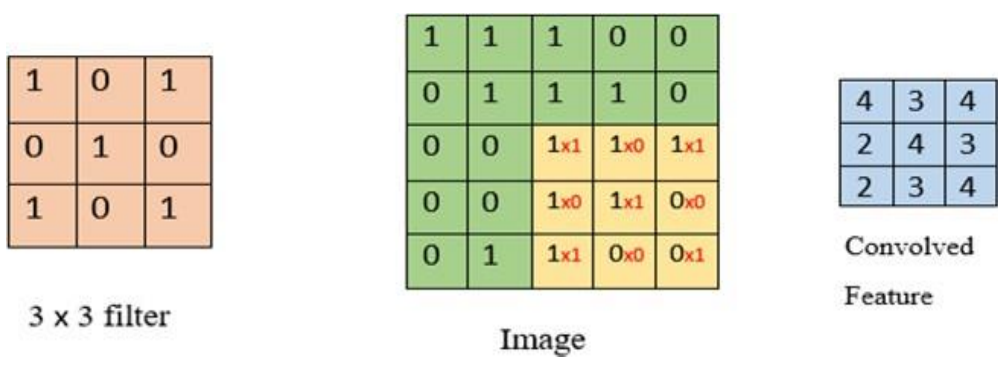


Figure 3.2: Convolution Layer

To understand this there are some equations what are we use in our project, For one-dimension convolution

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau).\, g(t - \tau)\, d\tau \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

$$(f * g)(t) = \int_{0}^{t} f(\tau).\, g(t - \tau)\, d\tau \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

This equation represents the percentage of the area of the g that overlaps the input f at a time t over all time T. It represents the future function value.

For multidimensional convolution

$$(I * h)(x, y) = \int_{0}^{x}\int_{0}^{y} I(i, j).\, h(x - i, y - j)\, di\, dj$$
$$= \int_{0}^{x}\int_{0}^{y} I(x - i, y - j).\, h(i, j)\, di\, dj \ldots\ldots\ldots \text{(3)}$$

Consider an image input I kernel h which convolution is divine to show. Fist equation performs by sliding the image over the kernel, and the second equation does its other way around. The result will be the scaler value [25]. We repeat the process for every point for x and y for store convolved matrix that represented by me and h. This is the mathematical expression.

3.3.2 Activation Layer

Nonlinear activation function for learning in this process.

$$W_1 * (W_2 * X) = (W_1 * W_2) * X = W * X \ldots\ldots\ldots\ldots\ldots\ldots \text{(4)}$$

Let's consider W_1 and W_2 to be the two sequences Convolution fixture applied on X without nonlinear activation. Because of the associate property of convolution, these two layers are useful as a single layer.

Several layers with typical ANN without activation function are affected as just having a single layer. Typically,

- read / leaky reLU used for activation function.
- Leaky rear avoids them during the problem.

We can say that,

□□□□ (□□□□□□□(□□□□(□))) = □□□□□□□ (□□□□(□□□□(□)))……. (5)

3.3.3 Pooling Layer

Down sampling of features. Involves the down sampling of features so that we need to learn fewer parameters during training.

Two hyper of parameter involving with pooling:

- The dimension of the spatial extent
- Stride

In Dimension of spatial scale, the value of n for which we can take N cross featured presentation and mapped to a single value and in the stride part how many features the sliding window skips along the width and height, which is similar to that what we saw in convolution.

A standard pooling layer uses two cross max filters with a stride of two non-overlapping screens with a pace of two non-overlapping filters. Max filter returns maximum value among the feature.

Average filter returns a common element in the reason also be used, but max pooling is the better practice [1].

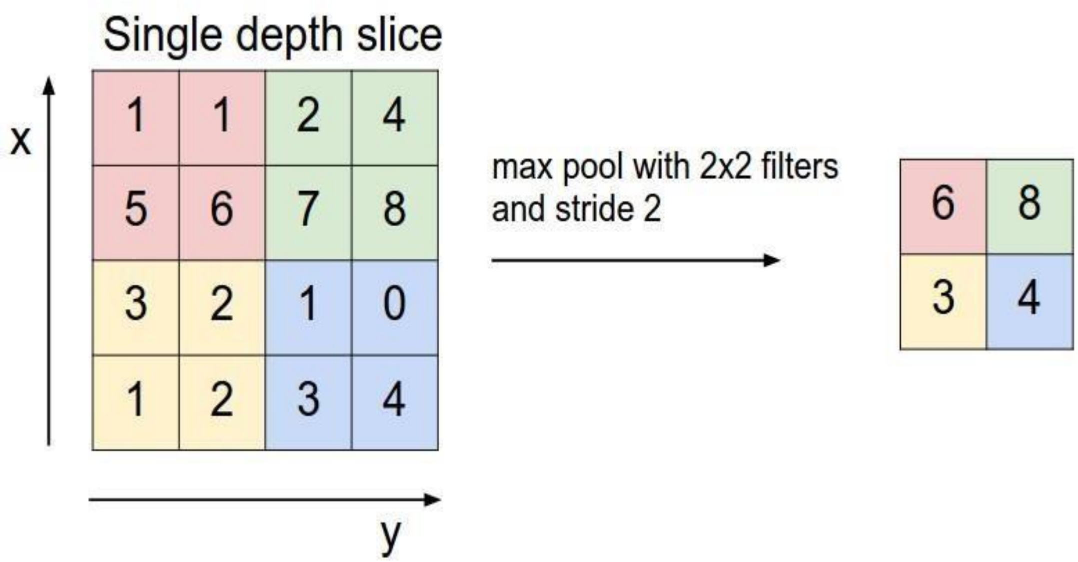


Figure 3.3: Pooling Layer

- Depth remains the same after the pooling.

Performing pooling reduces the chances of overfitting as their less parameter.

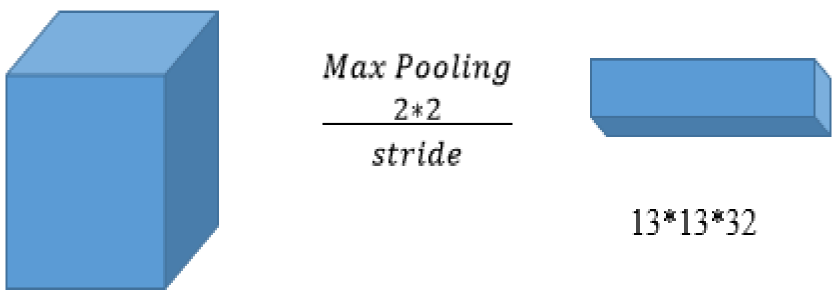


Figure 3.4: Pooling Layer

26*26*32 volume we have. Then using pool layer filature and stride of two of this volume now a reduce to 13*13*32 feature map. We cut the number of features to 23% of the original amount. This is a significant decrease in the name of parameters.

3.3.4 Fully Connected Layer

The convolution layer is providing a meaningful low dimension, and a fully connected layer is learning a possible nonlinear function in that space.

- Pooling output: 3D feature map
- Fc input: 1D feature vector

For 3D volume, they are usually very deep at this point because the increasing number of kernels is introduced every convolution layer. To convert the 3D amount into one dimension.

We want the output width and height to be 1.

Conversion: flatten 3D volume

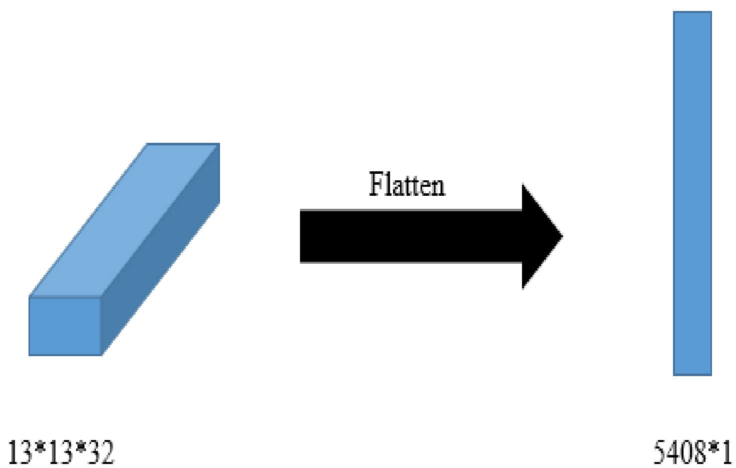


Figure 3.5: Fully Connected Layer

By flattening (13*13*32) into one on equal 5408*1. This is a 5408-dimension vector, which is a fully connected layer [7].

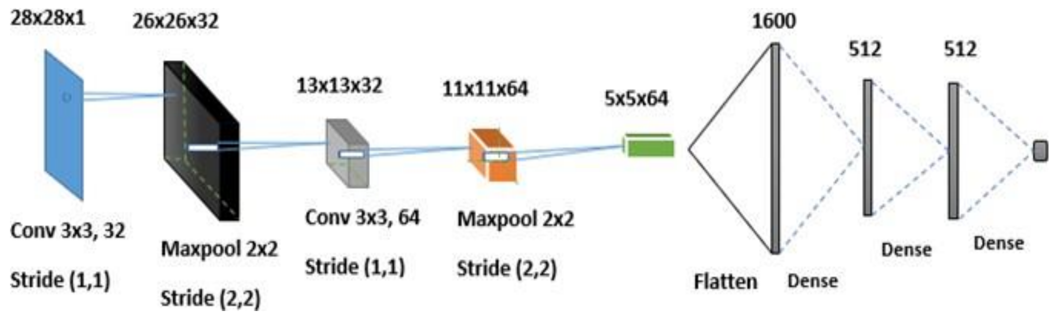


Figure 3.6: Fully Connected Layer

Initially start 28*28*1 image (greyscale). The images pass through a convolution layer. Here we apply 32:3*3 filter. The depth of the same as the input. The output will be like that.

$$\square\square\square_{\square\square\square\square h} = \square_{\square\square\square\square h} - h_{\square\square\square\square h} + 1 \text{ ....(6)}$$

After activation, we can pass into max-pooling 2*2 stride (2,2) in Keras: padding parameter in max pool 2D takes two values.

- Valid: No padding for image bounded
- Same: Padding for image border

The first is either valid which means that there is no padding, so we don't slide the kernel of the border of the image and secondly cover the image

$$\square\square\square_{\square\square\square\square h} = \frac{\square_{\square\square\square\square h} - h_{\square\square\square\square h} + 1}{\square\square\square\square\square\square} \text{ ....(7)}$$

Here, the output height is the same, and the depth remains unchanged. We can now feet the output of this pulling layer to a fully contently layer by flattening layer it to one dimension.

3.4 Data-set Collection

We started the data-set assortment by receiving an information set of all 36 categories whose image was of one individual with a black background. Some hand signs from this knowledge set are shown in Fig. 3.7, 3.8, and 3.9. From this initial knowledge set, we tend to distend upon it by adding two additional forty-five image set of 2 different people. With these three sets, do we manage applied augmentation on that through scripts and different ways to extra expand our data-set? Within the augmentation method, we tend to have a variety of parameters. These parameters were rotation angle, and dimension shift varies, height shift varies, shear varies, zoom range, horizontal flip, and all mode. Once the augmentation was done, we tend to have for every category eight hundred pictures for coaching, two hundred images for testing. For a total of one, 1000 photos for every one of the 36 groups that put our entire range of pictures at 36000.

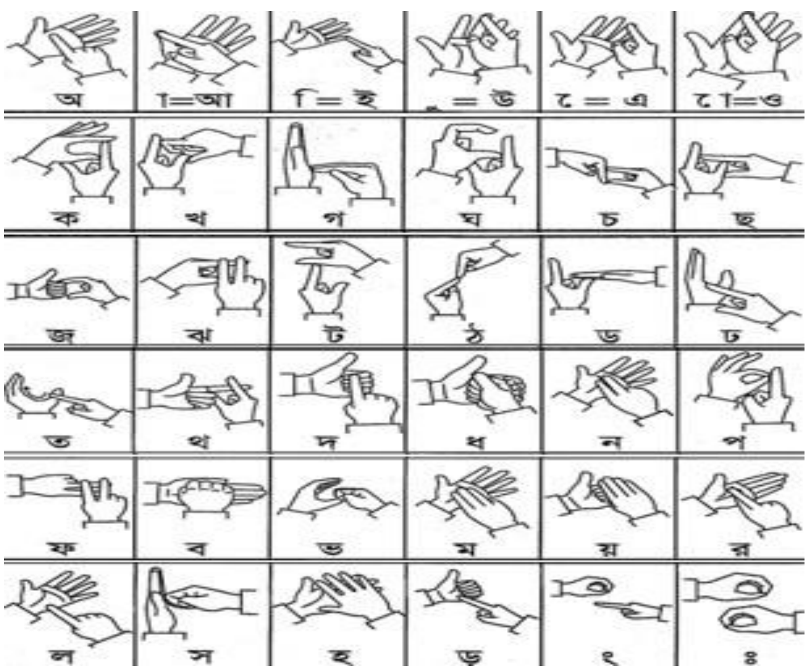


Figure 3.7: Sample Data Image

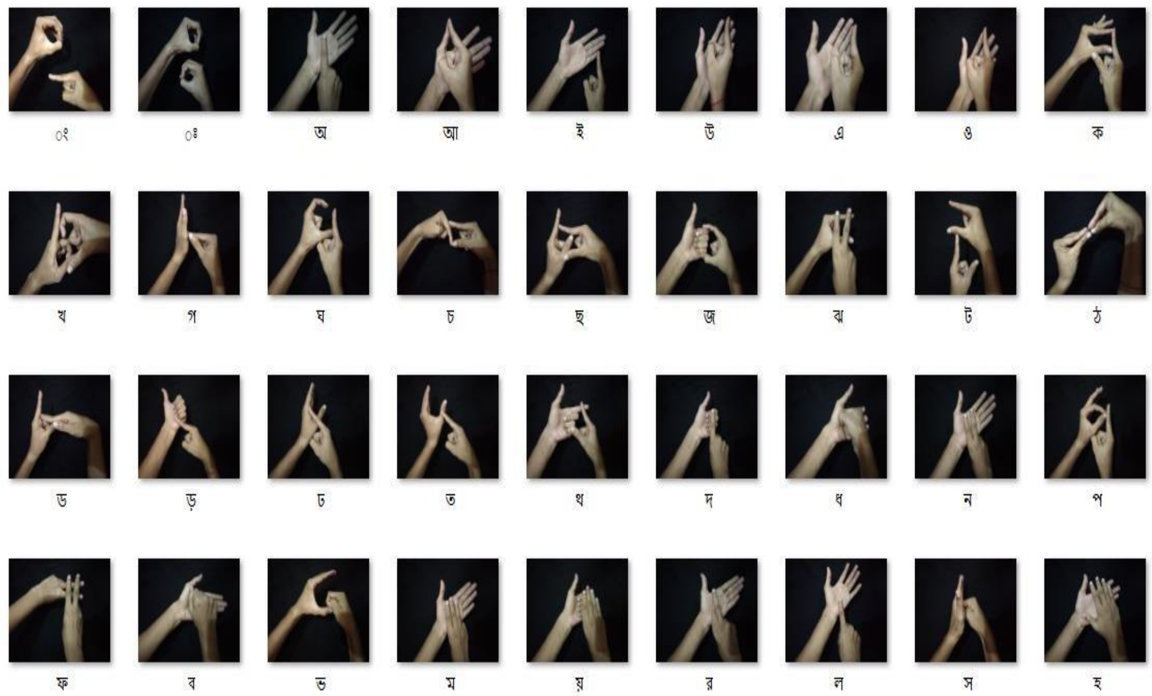


Figure 3.8: Raw Data Image

3.5 Converting Raw Image to Binary Image

First, we have to import cv2 and glob. Cv2 is mainly using for editing this image likely binary conversion and glob for choosing the directory or file name. First, we have to read the image and convert it to a grey image. With the help of the threshold function, then we have to save the binary image. Here we have to wait for an essential role and destroy All Window to clear the previous image record. Then with the help of cv2, we can rename the image. For multiple conversion image, we have used a for loop onto the folder that we have using.

Figure 3.9: Binary Data Image

3.6 Error Metrics

We can calculate the final loss by combining for each digit value loss of the digit caps layer. The equation is given below,

□□=□□ max (0, □+-|□□|)2+ □□(1-□□) max (0, |□□|− □−)2 ....(8)

The value of □□ will be one within the digit has the proper level from digit caps layer if do it isn't, then it will be 0. For the other nine digits, if Ta value is one for one numbering digit caps for losing the first digit, we subtract it from m+. The costs fixed at 0.9. Predicting the correct loss will be zero.

3.7 Classification Performance Assessment

To recognize the sign language, we need Two distinct supervised machine learning algorithms. This performs of Classification to this assessment evaluated by the five-evaluation metrics.

Accuracy = (True Positive + True Negative) / (True Positive + False Negative + False Positive + True Negative)

Sensitivity = True Positive / (True Positive + False Negative)

Specificity = True Negative / (True Negative + False Positive)

Precision = True Positive / (True Positive + False Positive)

F1-Score =2 × (Sensitivity × Precision) / (Sensitivity + Specificity)

CHAPTER 4

RESULT AND DISCUSSION

4.1 The Confusion Matrix of Convolutional Neural Network (CNN)

CNN has implemented the classification of Bangladeshi Sign Language. The CNN models have been trained by 5760 pre-processed binary images tested by 1440 images. The ratio of trained and verified images is 80% and 20% of total images.

Table 4.1: MODEL CLASSIFICATION OF BANGLA WORD

Word	Classified	Word	Classified	Word	Classified
অ	C l		C13		C25
আ	C2	ঝ	C14	ব	C26
ই	CS	ট	C15	ভ	C27
উ	C4	ঠ	C l6	ম	C28
এ	CS	ড	C17	য়	C29
ও	Ca	ঢ	C18	র	C30
ক	C7	ত	C I'9	ল	C3l
খ	C6	থ	C20	স	C32
গ	C9	দ	C2l	হ	C33
ঘ	C1I)	ধ	C22	ড়	C34
চ	C I I	ন	C23		C35
ছ	C12	প	C24		C36

Here in table 4.2, 4.3 and 4.4 show the confusion matrix of CNN. Here we calculated the value of True Positive, False Positive, True Negative, False Negative to build the confusion matrix. The diagonal values of this matrix significant as the True Positive of a specific class. False Negative is the summation of all values without True Positive of every row. The summation of all values is different from True Positive from the significant column in False Positive. Real Negative also the sum of all values in the row and columns.

Table 4.2: CONFUSION MATRIX

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
C1	41	0	0	59	0	0	0	0	0	0	0	0	0
C2	0	100	0	0	0	0	0	0	0	0	0	0	0
C3	0	0	100	0	0	0	0	0	0	0	0	0	0
C4	41	0	0	59	0	0	0	0	0	0	0	0	0
C5	0	0	0	0	100	0	0	0	0	0	0	0	0
C6	0	0	0	0	0	100	0	0	0	0	0	0	0
C7	0	0	0	0	0	0	100	0	0	0	0	0	0
C8	0	0	0	0	0	0	0	100	0	0	0	0	0
C9	0	0	0	0	0	0	0	0	100	0	0	0	0
C10	0	0	0	0	0	0	0	0	0	100	0	0	0
C11	0	0	0	0	0	0	0	0	0	0	100	0	0
C12	0	0	0	0	0	0	0	0	0	0	0	100	0
C13	0	0	0	0	0	0	0	0	0	0	0	0	100

Table 4.3: CONFUSION MATRIX

	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26
C14	100	0	0	0	0	0	0	0	0	0	0	0	0
C15	0	100	0	0	0	0	0	0	0	0	0	0	0
C16	0	0	100	0	0	0	0	0	0	0	0	0	0
C17	0	0	0	100	0	0	0	0	0	0	0	0	0
C18	46	0	0	0	33	0	0	0	0	7	0	0	14
C19	0	0	0	0	0	100	0	0	0	0	0	0	0
C20	0	0	0	0	0	0	100	0	0	0	0	0	0
C21	0	31	0	0	0	3	0	56	0	0	0	0	10
C22	0	0	0	0	0	0	0	0	100	0	0	0	0
C23	0	0	0	0	0	0	0	0	0	100	0	0	0
C24	0	0	0	0	0	0	0	0	0	0	100	0	0
C25	0	0	0	0	0	0	0	0	0	0	0	100	0
C26	28	0	22	9	0	5	7	0	4	14	0	0	11

Table 4.4: CONFUSION MATRIX

	C27	C28	C29	C30	C31	C32	C33	C34	C35	C36
C27	41	0	0	59	0	0	0	0	0	0
C28	0	100	0	0	0	0	0	0	0	0
C29	0	0	100	0	0	0	0	0	0	0
C30	41	0	0	59	0	0	0	0	0	0
C31	0	0	0	0	100	0	0	0	0	0
C32	0	0	0	0	0	100	0	0	0	0
C33	0	0	0	0	0	0	100	0	0	0
C34	0	0	0	0	0	0	0	100	0	0
C35	0	0	0	0	0	0	0	0	100	0
C36	0	0	0	0	0	0	0	0	0	100

4.2 The Classification Result of Convolutional Neural Network (CNN)

We represent the True Positive, True Negative, False Positive, False Negative values what is generated by CNN. Here CNN predicted the amount of True Positive 3200 and True Negative 39300. Even CNN predicted the amount of False Positive 300 and False Negative 400.

Table 4.5: THE CLASSIFICATION RESULT OF CONVOLUTIONAL NEURAL NETWORK (CNN)

	True Positive	False Negative	False Positive	True Negative
C1	41	59	41	1059
C2	100	0	0	1100
C3	100	0	0	1100
C4	59	41	59	1041
C5	100	0	0	1100
C6	100	0	0	1100
C7	100	0	0	1100
C8	100	0	0	1100
C9	100	0	0	1100
C10	100	0	0	1100
C11	100	0	0	1100
C12	100	0	0	1100
C13	100	0	0	1100
C14	100	0	0	1100
C15	100	0	0	1100

C16	100	0	0	1100
C17	100	0	0	1100
C18	33	67	33	1067
C19	100	0	0	1100
C20	100	0	0	1100
C21	56	44	56	1044
C22	100	0	0	1100
C23	100	0	0	1100
C24	100	0	0	1100
C25	100	0	0	1100
C26	11	89	11	1089
C27	41	59	41	1059
C28	100	0	0	1100
C29	100	0	0	1100
C30	59	41	59	1041
C31	100	0	0	1100
C32	100	0	0	1100
C33	100	0	0	1100
C34	100	0	0	1100
C35	100	0	0	1100
C36	100	0	0	1100

4.3 The Classification Performance of Convolutional Neural Network (CNN)

Here we describe the classification performance of CNN in table 4.6. The average result of Sensitivity, Specificity, Precision, and F1-Score is calculated for every word class of the CNN.

Table 4.6: THE CLASSIFICATION PERFORMANCE OF CONVOLUTIONAL NEURAL NETWORK

	Sensitivity	Specificity	Precision	F1-Score
C1	0.41	0.88	0.50	0.45
C2	1.0	1.0	1.0	1.0
C3	1.0	1.0	1.0	1.0
C4	0.59	0.88	0.50	0.54
C5	1.0	1.0	1.0	1.0
C6	1.0	1.0	1.0	1.0
C7	1.0	1.0	1.0	1.0
C8	1.0	1.0	1.0	1.0
C9	1.0	1.0	1.0	1.0
C10	1.0	1.0	1.0	1.0
C11	1.0	1.0	1.0	1.0
C12	1.0	1.0	1.0	1.0
C13	1.0	1.0	1.0	1.0
C14	0.41	0.88	0.50	0.45
C15	1.0	1.0	1.0	1.0
C16	1.0	1.0	1.0	1.0
C17	0.59	0.88	0.50	0.54
C18	0.33	0.97	0.5	0.26
C19	1.0	1.0	1.0	1.0
C20	1.0	1.0	1.0	1.0
C21	0.56	0.95	0.5	0.38
C22	1.0	1.0	1.0	1.0
C23	1.0	1.0	1.0	1.0
C24	1.0	1.0	1.0	1.0
C25	1.0	1.0	1.0	1.0
C26	0.11	0.99	0.5	0.1
C27	0.41	0.88	0.50	0.45
C28	1.0	1.0	1.0	1.0
C29	1.0	1.0	1.0	1.0
C30	0.59	0.88	0.50	0.54
C31	1.0	1.0	1.0	1.0

C32	1.0	1.0	1.0	1.0
C33	1.0	1.0	1.0	1.0
C34	1.0	1.0	1.0	1.0
C35	1.0	1.0	1.0	1.0
C36	1.0	1.0	1.0	1.0

In this graph, we denoted the average result of Accuracy, Sensitivity, Specificity, Precision, and F1-Score.

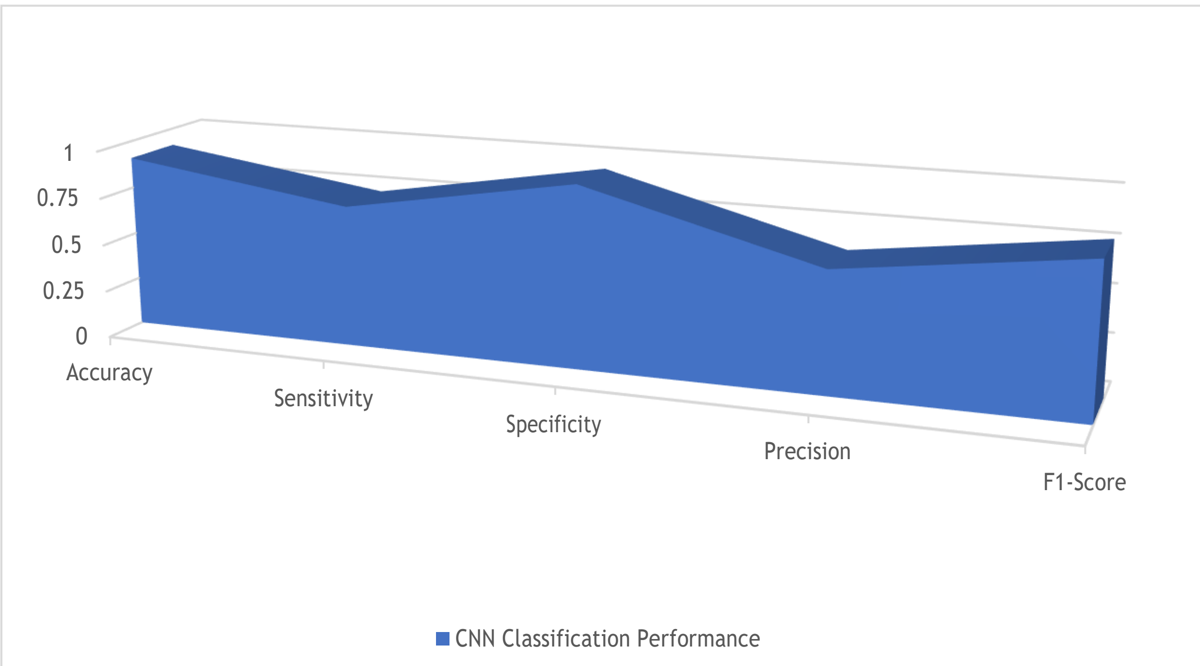


Figure 4.1: CNN Classification Performance

4.4 Epochs Information of Convolutional Neural Network (CNN)

Here we describe the epochs information (table 4.7) of CNN during the training and testing period for learning to recognize and classify the Bangla Sign Language.

Table 4.7: EPOCHS INFORMATION OF CONVOLUTIONAL NEURAL NETWORK (CNN)

	Training Accuracy	Training Loss	Testing Accuracy	Testing Loss
Epoch 1	0.6734	0.9314	0.7188	0.8181
Epoch 2	0.8466	0.3910	0.9083	0.2062
Epoch 3	0.8907	0.2484	0.9125	0.1388
Epoch 4	0.9025	0.1895	0.9042	0.1483
Epoch 5	0.9055	0.1679	0.9062	0.1375
Epoch 6	0.9109	0.1530	0.9062	0.1371
Epoch 7	0.9093	0.1458	0.8958	0.1345
Epoch 8	0.9100	0.1365	0.9125	0.1314
Epoch 9	0.9124	0.1375	0.9125	0.1292
Epoch 10	0.9168	0.1281	0.9125	0.1268
Epoch 11	0.6734	0.9314	0.7188	0.8181
Epoch 12	0.8466	0.3910	0.9083	0.2062
Epoch 13	0.8907	0.2484	0.9125	0.1388
Epoch 14	0.9025	0.1895	0.9042	0.1483
Epoch 15	0.9055	0.1679	0.9062	0.1375
Epoch 16	0.9109	0.1530	0.9062	0.1371
Epoch 17	0.9093	0.1458	0.8958	0.1345
Epoch 18	0.9100	0.1365	0.9125	0.1314
Epoch 19	0.9124	0.1375	0.9125	0.1292
Epoch 20	0.9168	0.1281	0.9125	0.1268
Epoch 21	0.6734	0.9314	0.7188	0.8181
Epoch 22	0.8466	0.3910	0.9083	0.2062
Epoch 23	0.8907	0.2484	0.9125	0.1388
Epoch 24	0.9025	0.1895	0.9042	0.1483
Epoch 25	0.9055	0.1679	0.9062	0.1375
Epoch 26	0.9109	0.1530	0.9062	0.1371
Epoch 27	0.9093	0.1458	0.8958	0.1345
Epoch 28	0.9100	0.1365	0.9125	0.1314
Epoch 29	0.9124	0.1375	0.9125	0.1292
Epoch 30	0.9168	0.1281	0.9125	0.1268
Epoch 31	0.9055	0.1679	0.9062	0.1375
Epoch 32	0.9109	0.1530	0.9062	0.1371

Epoch 33	0.9093	0.1458	0.8958	0.1345
Epoch 34	0.9100	0.1365	0.9125	0.1314
Epoch 35	0.9124	0.1375	0.9125	0.1292
Epoch 36	0.9168	0.1281	0.9125	0.1268

Final loss: 0.126842

Final accuracy: 0.927500

4.5 The Accuracy and Loss of Convolutional Neural Network (CNN)

The training accuracy and validation accuracy comparison are plotted in the figure: 4.2.

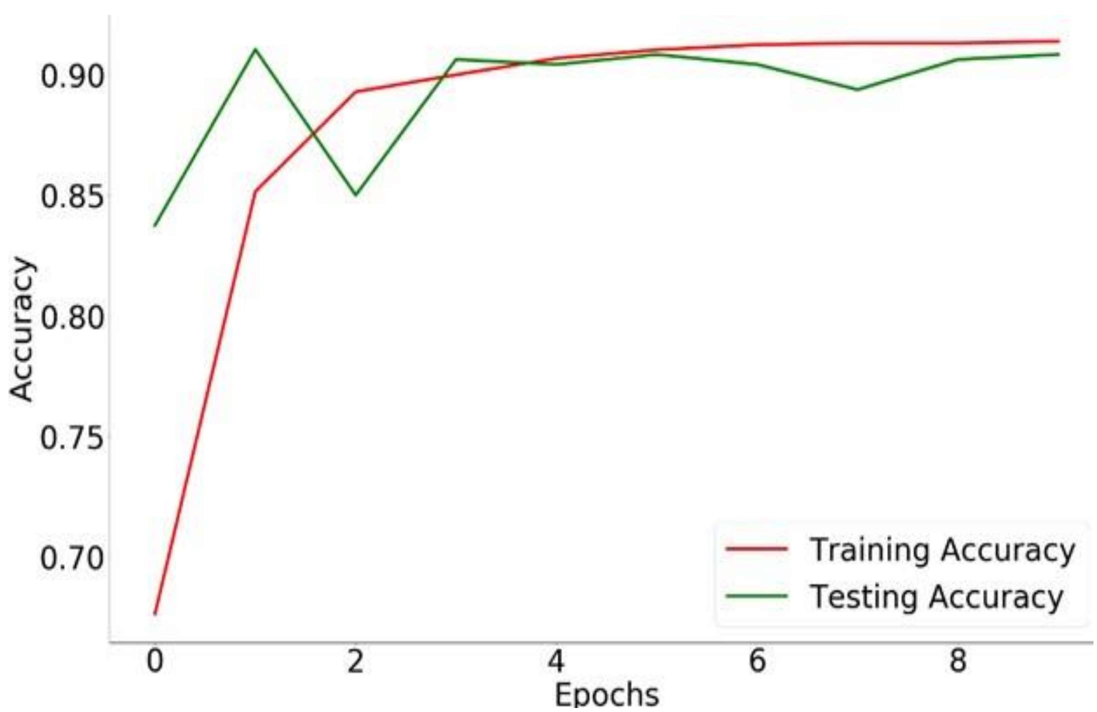


Figure 4.2: Accuracy of CNN

The training loss and validation loss comparison are plotted in the figure: 4.3.

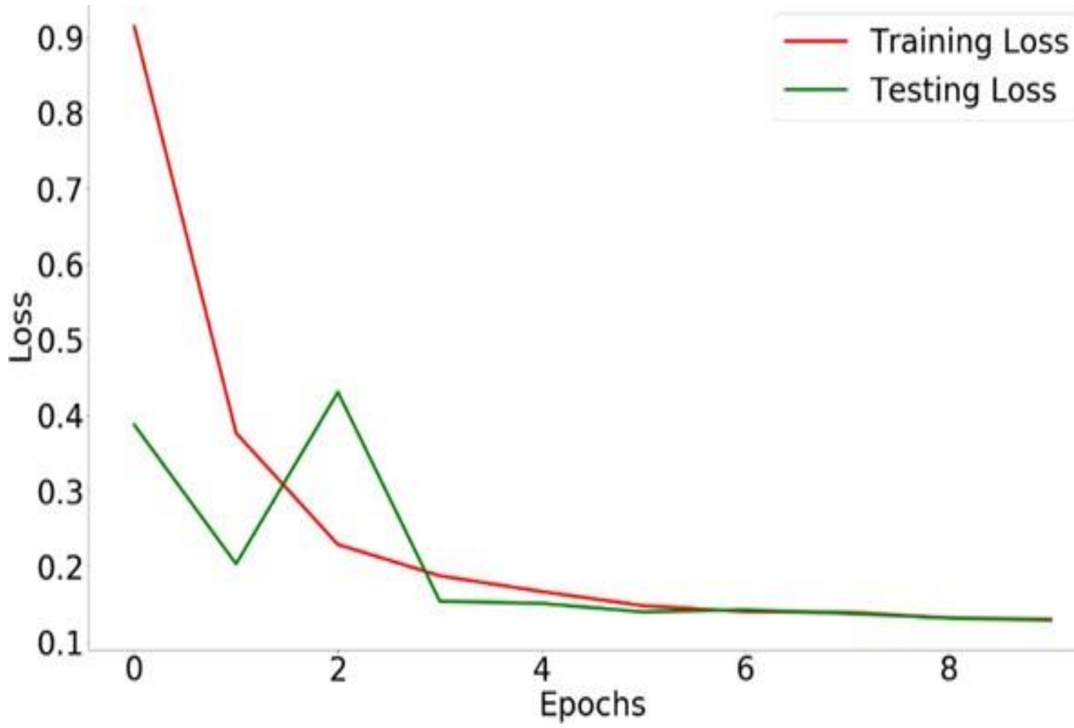


Figure 4.3: Loss of CNN

CHAPTER 5
CONCLUSIONS AND FUTURE SCOPES

5.1 Conclusion

Without sign-language communication, deaf and dumb peoples would have whole been neglected worldwide. As our country may be thickly settled, the proportion of those reasonable folks is over average. Therefore, for the event of our country, the involvement of those special folks is incredibly a lot of necessary.

Our country is additionally a developing country. Still, several of our folks don't seem to be well educated, and a few of them are illiterate. An immense proportion of individuals in our country don't understand English expeditiously. As Bangla is our maternal language, it's easier for everybody to interconnect in Bangla in our country. Hence, Bangla linguistic communication is essential for special folks to speak with the peoples of all spheres of life in our country. Whereas developing the system, we've got targeted on the unique people of all categories of our country. Through our network, those hearing and listening impaired folks will communicate with all reasonable people. This may build their life as like as traditional folks. And this may play a crucial role in the development of our country.

5.2 Future Scopes

In the future, we are going to build our system a lot of economical. We tend to solely tested for fifteen Bengali alphabets as we tend to fail to have enough resources for our dataset. We have a tendency to had to make all the datasets by ourselves that took loads of your time. In the future, we are going to live the accuracy for all the Bengali alphabets and even for binary numbers too. We tend to plan this model supported still pictures; however, up next, we are going to attempt to build it in a period.

In language, there are individual signs. There are different signs for chair, table, book, eat, so more. If we tend to attempt to embody all, it'll be an enormous dataset. We are going to try to organize them too.

REFERENCES

1. Hsiao, Ko-Jen, Tse-Wei Chen, and Shao-Yi Chien. "Fast fingertip positioning by combining particle filtering with random particle diffusion." Multimedia and Expo, IEEE, 2008.
2. Darroch, Kathy, and Liza Marshall. "Interpreting. NETAC Teacher Tipsheet."
3. D. Shahriar Hossain Pavel, Tanvir Mustafiz, Asif Iqbal Sarkar, M. Rokonuzzaman, "Geometrical Model-Based Hand Gesture Recognition for Interpreting Bengali Sign Language Using Computer Vision," ICCIT, 2003.
4. Lee, Christopher, and Yangsheng Xu. "Online, interactive learning of gestures for human/robot interfaces." Robotics and Automation, 1996. Proceedings, 1996 IEEE International Conference on. Vol. 4. IEEE, 1996.
5. Safaya, Kashmera Khedkkar, and Asst Prof Rekha Lathi. "Bare Hand Gesture Recognition-A study." (2012).
6. Bellugi, Ursula. The signs of language. Harvard University Press, 1979.
7. Phi, Lam T., et al. "A glove-based gesture recognition system for Vietnamese sign language." Control, Automation, and Systems (ICCAS), 2015 15th International Conference on. IEEE, 2015.
8. Ahmed, Syed Tauhid, and M. A. H. Akhand. "Bangladeshi Sign Language Recognition using fingertip position." Medical Engineering, Health Informatics, and Technology (Meditec), 2016 International Conference on. IEEE, 2016.
9. Kaushik Deb, Dr, et al. "Two-handed sign language recognition for Bangla character using normalized cross-correlation." Global Journal of Computer Science and Technology 12.3 (2012).
10. Nowak, Eric, Frédéric Jurie, and Bill Triggs. "Sampling strategies for bag-of-features image classification." Computer Vision–ECCV 2006 (2006): 490-503.
11. "INRIA Summer School 2010 Cordelia Search Large | Information Retrieval". Scribd. Web. 2017.
12. Bouchard, Guillaume, and Bill Triggs. "Hierarchical part-based visual object categorization." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.
13. Alexander, Voskresensky, and Ivanova Marina. “A Methodical and Didactical Complex Sign Language”.
14. Darroch, Kathy, and Liza Marshall. "Interpreting. NETAC Teacher Tipsheet." (1998).
15. Pavel, Dewan Shahriar Hossain, et al. "Modeling of Bengali sign language expression as dynamic 28 3d polygons for developing a vision based intelligent system for dumb people." National Conference on Computer Processing of Bangla. 2004.
16. O'Hara, Stephen, and Bruce A. Draper. "Introduction to the bag of features paradigm for image classification and retrieval." arXiv preprint arXiv:1101.3354 (2011).
17. Suykens, Johan AK, and Joos Vandewalle. "Least squares support vector machine classifiers." Neural processing letters 9.3 (1999): 293-300.

18. Zisserman, A. "Lecture 2: The SVM classifier." C19 Machine Learning (2011): 12.
19. Mahmoud, Tarek M. "A new fast skin color detection technique." World Academy of Science, Engineering and Technology 43 (2008): 501-505.
20. Chai, Douglas, and King N. Ngan. "Face segmentation using skin-color map in videophone applications." IEEE Transactions on circuits and systems for video technology 9.4 (1999): 551-564.
21. Cheong, Sungmoon, Sang Hoon Oh, and Soo-Young Lee. "Support vector machines with binary tree architecture for multi-class classification." Neural Information Processing-Letters and Reviews 2.3 (2004): 47-51.
22. Lee, Seungyon, et al. "A gesture-based American sign language game for deaf children." CHI'05 Extended Abstracts on Human Factors in Computing Systems. ACM, 2005.
23. Guyon, Isabelle, et al., eds. Feature extraction: foundations and applications. Vol. 207. Springer, 2008.
24. Choudhury, Najeefa Nikhat, and Golam Kayas. Automatic recognition of Bangla sign language. Diss. BRAC University, 2012.
25. Guyon, Isabelle, and André Elisseeff. "An introduction to feature extraction." Feature extraction. Springer Berlin Heidelberg, 2006. 1-25.
26. Chaque, Adnan, et al. "A novel concept of 3d animation based'intelligent assistant' for deaf people: for understanding Bengali expressions." International Conference on Computer and Information Technology. 2002.
27. Rahman, Md, Md Abdullah, and S. Mondal. "Recognition of Static Hand Gesture of Alphabet in Bangla Sign Language‖." Conference (IOSRJCE). Vol. 8. No. 1. 2012.
28. Maraqa, Manar, and Raed Abu-Zaiter. "Recognition of Arabic Sign Language (Arce) using recurrent neural networks." Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the. IEEE, 2008.
29. Pramada, Sawant, et al. "Intelligent sign language recognition using image processing." IOSR Journal of Engineering (IOSRJEN) 3.2 (2013): 45-51.29
30. Baker-Shenk, Charlotte Lee, and Dennis Cokely. American Sign Language: A teacher's resource text on grammar and culture. Gallaudet University Press, 1991.
31. Alauddin, Mohammad, and Abul Hasnat Joarder. "Deafness in Bangladesh." Hearing Impairment. Springer Japan, 2004. 64-69.
32. Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." Computer vision–ECCV 2006 (2006).
33. Hiriart-Urruty, Jean-Baptiste, Jean-Jacques Strodiot, and V. Hien Nguyen. "Generalized Hessian matrix and second-order optimality conditions for problems with C 1, 1 data." Applied Mathematics & Optimization 11.1 (1984).

Bangla Sign Language Fall 2019

ORIGINALITY REPORT

28%	22%	13%	17%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.bracu.ac.bd Internet Source	7%
2	studentsrepo.um.edu.my Internet Source	4%
3	ltpwww.gsfc.nasa.gov Internet Source	3%
4	repository.maranatha.edu Internet Source	3%
5	Submitted to Daffodil International University Student Paper	2%
6	globaljournals.org Internet Source	1%
7	Lam T. Phi, Hung D. Nguyen, T.T. Quyen Bui, Thang T. Vu. "A glove-based gesture recognition system for Vietnamese sign language", 2015 15th International Conference on Control, Automation and Systems (ICCAS), 2015 Publication	1%

8	cennser.org Internet Source	1%
9	Submitted to University of Birmingham Student Paper	1%
10	Said, Joseph I, Zhongxu Lin, Xianlong Zhang, Mingzhou Song, and Jinfa Zhang. "A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton", BMC Genomics, 2013. Publication	1%
11	ijcem.com Internet Source	<1%
12	www.kuet.ac.bd Internet Source	<1%
13	www.cnlopb.ca Internet Source	<1%
14	Submitted to Miami University of Ohio Student Paper	<1%
15	idfresearch.org Internet Source	<1%
16	Submitted to University of Westminster Student Paper	<1%
17	Submitted to University of Surrey Student Paper	<1%

18	"Proceedings of the International Conference on Data Engineering and Communication Technology", Springer Science and Business Media LLC, 2017 Publication	<1%
19	Submitted to Universiti Teknologi Malaysia Student Paper	<1%
20	www.infitt.org Internet Source	<1%
21	Submitted to Marquette University Student Paper	<1%
22	www.unknowncheats.me Internet Source	<1%
23	"Pattern Recognition", Springer Nature, 2012 Publication	<1%
24	Submitted to Dhirubhai Ambani Institute of Information and Communication Student Paper	<1%
25	docplayer.net Internet Source	<1%
26	Submitted to University of Florida Student Paper	<1%
27	learneconometrics.com Internet Source	<1%

28	academic.oup.com Internet Source	<1%
29	Shujun Ou, Ning Jiang. "LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons", Plant Physiology, 2018 Publication	<1%
30	Submitted to King's College Student Paper	<1%
31	www.dbpia.co.kr Internet Source	<1%
32	Submitted to University of Southampton Student Paper	<1%
33	Jwu-Sheng Hu, Ming-Tang Lee, Ting-Chao Wang. "Wake-up-word detection for robots using spatial eigenspace consistency and resonant curve similarity", 2011 IEEE International Conference on Robotics and Automation, 2011 Publication	<1%
34	ntcuir.ntcu.edu.tw Internet Source	<1%
35	dspace.bracu.ac.bd:8080 Internet Source	<1%
36	Lecture Notes in Computer Science, 2004. Publication	<1%

37	spinepain.seas.upenn.edu Internet Source	<1%
38	www.ebmt.org Internet Source	<1%
39	www.acsij.org Internet Source	<1%
40	Submitted to University of Technology Student Paper	<1%
41	link.springer.com Internet Source	<1%
42	Submitted to University of Queensland Student Paper	<1%
43	Submitted to University of Glasgow Student Paper	<1%
44	Submitted to Multimedia University Student Paper	<1%
45	Hyun Sik Sim, Hae In Kim, Jae Joon Ahn. "Is Deep Learning for Image Recognition Applicable to Stock Market Prediction?", Complexity, 2019 Publication	<1%
46	Submitted to University of South Alabama Student Paper	<1%