Real Time Bangla Sign Language Detection

A project report submitted to the Department of Information and Communication Technology at Comilla University in partial fulfillment of the requirements for the degree of Bachelor of Science (Engineering) in Information and Communication Technology

Submitted By

Team Member

ID

11909018

11909041

11909049

Session: 2018-2019

Department of Information and Communication Technology

Comilla University,Bangladesh

Date of Submission: January,2024

CERTIFICATE OF APPROVAL

This is to certify that the project titled "**Real Time Bangla Sign Language Detection**" carried out by Jasmine Akter, Shanta Islam, Akash Hasnat for the partial fulfillment of the requirements of B.Sc engineering final year project. The dissertation has been carried out under our guidance and is record of the authentic work carried out successfully.

.....

Alimul Rajee

Lecturer

Dept. of ICT

Comilla University.

ACKNOWLEDGEMENT

We take this opportunity to thank a few persons that made our initiative possible. Also, we would like to thank the department leader. We would like to sincerely thank our mentor Alimul Rajee for his unwavering support, assistance, and advice, without which this project would not have been finished.

Throughout the course of the project, we would like to express our profound gratitude to our supervisor for their unwavering support and wise counsel. Finally, but certainly not least, we would like to express our profound gratitude to our parents, friends, and coworkers who have helped us in some way or another to complete this project.

ABSTRACT

In the present almost 440 million people which is over 5% over the total population around the world is affected by hearing loss according to the World Health Organization. And this novel system provides an effective way to communicate with them properly. It argues for an advanced method for recognizing sign languages. The algorithm is designed such that it can copy the signs from the image sequence using hand gesture recognition. It is helpful in finding the difference between static and dynamic expression and copying the suitable feature vector. Vector Support Machines are used to classify them. On the Basis of standard module Speech identification is constract-Sphinx. For Speech and hearing disabled person SL is the basic means for communication . However, the majority of healthy people don't have the basic knowledge about these SLs . While multiple systems exist for various sign language recognition tasks, no superior system exists for Bangla sign language. In order to translate bangla sign language movements into bangla text that the average Bangladeshi citizen can comprehend, that is the goal of our effort. With two ways to indicate meanings—one handed and two handed—Bangla, the fifth most spoken language in the world, has its own unique sign language. Here an SL interpreter is developed which takes input using sign gestures and gives the output on display. The system is trained with CNNs with our own novel dataset. In this project we made our own novel dataset for the both training and also for testing purpose .We take 50000+ samples of the 38 symbols—nine for vowels and twenty-seven for consonants—visualized with one hand in Bengali sign language are defined to represent the alphabet, while over 4000 double-hand symbols are available for frequently used Bengali words. Datasets are structured into classes and for training and testing CNN is applied. Experimental results give high accuracy and in diverse backgrounds it gives desired segmentation. After the experimental results we have found 99.98% testing accuracy and 99.85% validation accuracy .

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASL	American sign language
D&D	Deft & dump
BdSL	Bangla Sign Language
SLR	Sign Language Recognition
SL	Sign Language
NUI	Natural User Interface
HMMs	Hidden Markov Models
HGR	Hand Gesture Recognition
VR	Virtual Reality
CNN	Convolutional Neural Network
BSL	British Sign Language
ZSL	Zero Shot Learning
CRF	Condition Random Field
KNN	K-Nearest Neighbour
ANN	Artificial Neural Networks
SVM	Support Vector Machine
SOM	Self Organizing Map

CHAPTER-ONE

INTRODUCTION

1.1 Background

Sign language is a visual language that uses hands, shapes, movements, and facial expressions to communicate words and sentences. Technologies that recognize sign language can be used to recognize hand gestures used for traffic signals or other signs that contain numbers, alphabets, sentences, or other symbols. While some scientists are concentrating on still images, others are working on real-time sign identification. Recently, methods based on artificial neural networks have been employed for real-time ASL phrase [1] and alphabet [2] recognition. The study uses Microsoft Sensor System hand sign recognition [3] to identify signals for two practical uses: arithmetic computation and the rock, paper, scissors game, with a mean accuracy of more than 90% for ASL.

Sign language is an immensely significant system of communication for special beings like deaf individuals. In sign expression, every expression has an earmarked meaning. Depth meaning can be expounded by the collaboration of amalgamation of different basic aspects. Sign expression involves the many uses of various parts of the body like fingers, arm, hand, facial expression to deliver messages and parameters. Majority sign expression involves the upper portion of the body from chest level upward. Therefore, the same expression has greatly huge changes in shapes. Hand gestures are categorized into different types like communicative gestures, manipulative gestures, controlling gestures, and conversational gestures. Sign language is a type of communicative system. Since sign expressions are deeply structural, it is used for computerized vision algorithms, parameters [8].

However, Studies of sign language detection are greatly impacted by detection of hand gesture probing, sign expressions are communicative courtesy. While analyzing literature in sign expression recognition, gesture recognition is also required to study. Sign and expression language detection includes the entire procedure of tracking and recognizing the signs performed and transforming into semantically purposeful expression. Gesture recognition was tried early in

1993, speech identification is adapted from gesture identification techniques and parameters. Sign recognition recognizes a particular structure that differentiates the object from the residual shapes [9].

The D&D people primarily communicate through sign language. Each language has its own collection of graphic symbols that are used to represent various hand gestures, positions, and movements. Some hand gestures require the use of both hands. For the alphabet and words, respectively, there are typically a set of static and dynamic symbols. A set of 38 symbols—nine for vowels and twenty-seven for consonants—visualized with one hand in Bengali sign language are defined to represent the alphabet, while over 4000 double-hand symbols are available for frequently used Bengali words [4].

The technical part copied the features of sign language contacts is the social indication and meaning, technological amenities and these are easy for application. Webcam is used to capture the image and sign will be pre-processed by using Microsoft Visual Studio as an IDE and Open-CV library. On the receiving end, the input image is captured by a web-cam using a distance based algorithm. By using convexity hull algorithm Fingertip is recognized .Then in last, Starner et al suggest HMMs to categorize exposure, flow of information and shape the output sign language. HMMs are used for recognition of speech. In a complete count of 262 signs are gathered from many individuals, and HMMs classifiers give 94% average accuracy . It is seen that accuracy is decreased when the dataset has a minimum number of data, accuracy decreased to 47.6%.And 91.3% accuracy can be achieved by training the database[10].

In this project paper, we will try to explore the technologies applied for bangla language detection using sign, including also computer vision, natural expression processing, and machine learning. sign language gestures. And we also study the recent literature on recognizing sign language and highlighting some approaches. Finally, we will have to discuss the application of detecting bangla sign language. Examples are public spaces increasing by improving accessibility, increasing education and training for bangla sign language interpreters, and the active communication between hearing-impaired and non-hearing-impaired beings. Our only aim is to provide a comprehensive overview detecting bangla sign language and its impact on society.

The centre of attraction of this work is to recognize bangla sign language. However, research in bangla sign language recognition is highly performed by gesture of hand research, as sign language is the type of the communicative gestures. It is greatly required to study the literature on recognizing gestures before reviewing the literature.

In recent times, besides mouse and keyboard people use different types of control devices .Different systems are available for specific tasks or applications, Interactions between humans and machines is one of the famous interactions .The advantages of natural interactions are taken by Web browsers, video games, VR environments e.g., voice control devices , touchpads, webcam etc. The motive of this research is to involve the individuals into application in an authentic way. Without using any hardware, natural interactions use the user's body gestures. This is known as NUI. It understands commands and performs specific tasks in the application by using various sensors.

Using body and hand gestures is known as NUI. Therefore, The important element of this system is HGR. By touching and performing gestures we can control the smartphones and tablets . NUIs use different cameras like the Xbox Kinect camera to discover the user's distinctive body parts. The hands of the user can provide a collection Different kinds of gestures are collected from the users ,the system allows the control of application if the sample is perfectly matched. Segmenting the hands from Gestures are collected from the cameras and performing the segmentation of SLR. Usually mute people face communication problems so the SL system is developed [11].

Usually, we don't learn sign languages if it is not involved in our job sectors or if there is no disabled person in our surroundings. Whenever we want to communicate with mute person it becomes very hard and difficult. It's required to hire a translator whenever a mute individual tries to communicate with a person who doesn't know sign languages . Hiring a translator is difficult and expensive as it creates various problems depending on the consequences. And here the image recognition techniques provide a vital role of identifying signs. This type of communication problem is solved though this system where there is no need of a human translator as it uses different algorithms for communication .

1.2 Necessity Of The Project

Bangladesh, with its own sign language and two ways of representation (one-handed and two-handed), is the fifth most spoken language in the world. On the other hand, BdSL still has to be recognized by a standard automatic system. The right channel is needed to communicate with the general public. Not everyone can understand sign language. There are several system available for American sign language recognition but for Bangla sign language there is no such system. Therefore, our project aims to convert bangla sign language gestures into bangla text that bangladeshi ordinary people can understand.

1.3 Problem Statement

A lot of people in our country go through hearing and speech difficulties. This system designs a model to detect the bangla sign language from a web camera .Using OpenCv webcam captures the hand gestures. Then the images are labelled after the images captured procedure is done, The images labelling is an important step and for sign recognition SSD mobile net version2 which is a pre trained model is required. Thus, effortless communication can be established between deaf and normal individuals.

To solve these problem 3 steps required:

1.Taking input from the users.

2. Images are classify to sign

3.Form classification displays the output.

System Difficulties Including:

1. Problem in detecting Sign boundary .
2. Some fingers can be out of the view area causing problems.
3. Environmental problems e.g. sensitivity of light, camera position.

1.4 Objective

1. We will Develop system that will detect signs and convert them into bangla text.
2. The system used CNN to recognize multiple gesture and

   by webcam gestures are capture and converted into sequence.
3. Image segmentation is done .
4. TensorFlow and OpenCV images. are used to identify hand gestures.

1.5 Summary

CNN is used to recognize the hand gesture and convert it into bangla text . Here in this system we study the letters and symbols of the Bangladeshi Sign Language system. We used 50 alphabets which consists of 11 stand-alone vowels and 39 consonant + vowel for the dataset. Here all the gestures are executed in a static manner. We used a predefined scene through which the image is captured consistently. This process helps in capturing images errorlessly. From an ideal scenario, If we have a large dataset then we will have more gestures and features which will make our system more robust.

In this research work we potentially focused on classification models for providing dataset and recognizing hand gestures. The system determines the hand images from the dataset and recognizes the gesture from images. Hand normalize values in the dataset is fixed size. Since all the images are pre-processed and classified to detect the sign language.

CHAPTER-2

BACKGROUND STUDY

2.1 Literature Review

Sign expression is not universal for all, it changes according to the country and region. The language of sign is categorized into 300 sign languages worldwide, As ASL, ISL, BSL etc. In accordance with Ethnologue2014 in the US, for about 2,50,000-5,00,000 people the ASL becomes the native language. Approximately 1 M to 20M people with hearing problems in China use the Chinese Sign Language. BSL is used by around 1,50,000 people living in the UK (United Kingdom). In Brazil, around 3 million people use Brazilian Sign Language for communication, like French Sign Language or the Portugal Sign Language. According to Ethnologue2008 in India, Indo-Pakistani Sign Language is used by around 1.5 million people.

SLR (Sign Language Recognition) is not only required for deaf and physically challenged people. Ordinary or normal people can also communicate in discordant areas like public places and even in libraries without annoying the other person. There are two mediums used in sign language like Manuel (hands are the means of communication) and the non-manual (body posture or facial expression becomes the means for communication). People also use fingers through which words are divided into letters and then letters are spelled using fingers. Detailed discussion about the Manual and non-manual SLR in the following subsections.

2.1.1.Manual Sign Language Recognition:

Components of manual SLR are hand gesture, shape of hand, Hand motion, and hand location etc.With the help of hands , the person with disability can easily interact with another individual. The manual SLR is categorized into two types:

Isolated Manual SLR:

The BdSLR system was proposed by BC Karmokar in order to identify BdSL. D&D users who are acclimated to BdSL can take advantage of BdSLR's interpretation feature when participating in socially significant activities. We have used feature extraction in conjunction with the NCL

[17] method for training, which is capable of performing a good recognition, in order to develop BdSLR. In NCL, BdSLR has been applied with varying numbers of individual NNs. The outcomes of the experiment demonstrate that NCL combined with feature extraction produces a strong 93% recognition accuracy. Other sign languages can be learned with BdSLR. One potential avenue for future development is to use BdSLR to detect BdSL in real time..[12]

For the automatic recognition of Bangla Sign Language (numerals and alphabets), Sunanda Das suggested a hybrid model that combines a random forest classifier with a convolutional neural network based on deep transfer learning. This system's overall performance is confirmed using the "Ishara-Bochon" and "Ishara-Lipi" datasets. These two datasets, "Ishara-Bochon" and "Ishara-Lipi," are the first comprehensive multifunctional open-access datasets for BSL. They contain isolated numerals and alphabets, respectively. Additionally, we suggested a method for backdrop reduction from the sign images, which eliminates extraneous characteristics. The system achieves accuracy, precision, recall, and f1-score values of 91.67%, 93.64%, 91.67%, and 91.47% for character recognition and 97.33%, 97.89%, 97.33%, and 97.37% for digit recognition, respectively, when the suggested background elimination strategy is used.[13]

Sadiah Ahmed Moon presents a method that uses binary masking, SVM classification, and hand gesture segmentation to identify BdSL. To increase accuracy over previous efforts, the system's characteristics of various notations have been painstakingly identified and extracted. The system generates a binary picture by converting it into the YCbCr color space and using Fuzzy C-means clustering. This binary image is then subjected to binary masks that are specifically created for different skin segmentations. The YCbCr color space has been utilized to pre-process the photos since it offers an efficient way to employ chrominance information for simulating the color of human skin. The dataset that is utilized in this system was created and selected to function well throughout the manual feature extraction procedure. A multiclass SVM classifier is used to compare, classify, and train the hand gestures of ten alphabets after around eighteen distinct characteristics were extracted. High dimensional spaces are frequently classified using multiclass SVM classifiers, which are equally useful for smaller datasets. With this technique, the system's accuracy for static hand positions is 99.8%.[14]

Zaid Bin Mahmud explains how to understand all 37 static signs in the one-handed Bangla sign alphabet using a deep transfer learning method based on CNNs and ResNet18. To

reduce sign gesture similarities between classes, a color-coded fingertip pattern was used. As part of this effort, a massive dataset has been developed and made publicly available, consisting of 45,958 photographs of all the sign letters (varying from 821 to 1999 images for each sign), gathered from 5 distinct subjects. The performance of the transfer learning model in identifying BdSL indicators was assessed after it had been trained and validated for 37 classes using five-fold cross validation. According to experimental data, the method's overall accuracy, sensitivity, and specificity for recognizing Bangla signs are 99.97, 99.49, and 99.99 percent, respectively. Millions of people use sign language for daily communication, and the suggested approach of automatic BdSL identification with extremely high accuracy can help youngsters who are deaf or hard of hearing acquire the language.[15]

Ragib Amil Nihal suggests two methods for automatic BdSL alphabet identification of both visible and unseen data, based on traditional transfer learning and modern ZSL. The effectiveness of the suggested method is assessed using a sizable dataset of 35,149 photos from more than 350 persons, differing in terms of backdrops, camera angles, light contrast, skin tones, hand sizes, and orientations, in addition to both forms of Bangla sign representations. For the ZSL method, a suggested division of the dataset into seen and unseen classes is made, along with the creation of a new semantic descriptor specifically for BdSL. With six unseen classes, our model's harmonic mean accuracy, seen accuracy, and zero-shot accuracy were, respectively, 68.21%, 91.57%, and 54.34%. After conducting quantitative experimentation on 18 CNN architectures and 21 classifiers, we found that the pre-trained DenseNet201 architecture was the best performing feature extractor and that Linear Discriminant Analysis was the best classifier with an overall accuracy of 93.68% on the large dataset for the transfer learning-based approach.[16]

In an attempt to identify the optimal model for BdS Alphabets and Numericals interpretation, Kanchon Kanti Podder compared classification with background and classification without background approaches. It was discovered that the CNN model trained with background-containing images performed better than the model trained without background. To increase the overall accuracy of sign recognition, the segmentation approach's hand detection element has to be more precise. ResNet18 fared the best in the first method, with 99.99% accuracy, precision, F1 score, sensitivity, and 100% specificity, using several visualization techniques and

performance measures. When BdS Alphabets and Numericals recognition is compared to existing research, the accuracy of the model in this study was shown to be significantly greater.. In order to decrease inter-class similarity while working with different image data that includes a range of backdrops and skin tones, the largest accessible dataset for BdSL Alphabets and Numericals is being supplied in this study.[17]

Sheikh Abujr presented a system that attempts to identify BdSL digits by building a model using a deep learning technique. In order to achieve our goal, our method trained specific indicators using the appropriate training dataset (Eshara-Lipi) using a CNN. 860 training photos and 215 test images (20%) of various tent classes of digits were used to train and test the model, respectively. In the end, the training model was able to recognize Bangla sign language digits with an accuracy of almost 95%. This model will help forward the process of creating a machine translation for BdSL.[18]

Tazkia Mim Angona proposes a model and develops a computer system that uses a deep CNN to convert Bangla Sign Language alphabets to matching Bangla letters. This model incorporates CNN in the form of a pre-trained model named "MobileNet," which generated an average accuracy of 95.71% in the recognition of 36 alphabets in Bangla Sign Language.[19]

This research project, which was proposed by Sabbie Ejaz, aims to develop a suitable deep learning model that can accurately detect the alphabets used in Bangla Sign Language, therefore reducing communication challenges between the speech and hearing impaired and the general population of Bangladesh. This study introduces a novel CNN architecture to detect Bengali sign alphabets with corresponding Ishara-Lipi databases. This design achieved a 99.86% general precision in Bengali sign alphabet recognition, surpassing all previous studies.[20]

2.2 Literature Summary

Table -2.1: Summary (Brief Overview) Of Different Research Paper

Data Set	**Algorithm**	**Result/Finding**
A sizable dataset of 35,149 photos	3D Convolutional Neural Network.	Got overall accuracy of 93.68% on the large dataset [16]
860 training photos and 215 test images	3D Convolutional Neural Network	95% training accuracy and0.059% training loss[18]
300 images for each letter space and delete	CNNs and Deep Neural Network	Approximately 76% accuracy rate[22].
MNIST image dataset	With feed forward and back propagation Artificial Neural Networks are used.	We got approximately 88% of Accuracy rate[23].
821 to 1999 images for each sign	CNN	Got optimal 99.97, 99.49, and 99.99 percent accuracy[15].
BdSL-binary dataset	CNN	Accuracy Rate we got is 99.8%[14].

We created our own dataset with 2500 different images.	CNN VGG-16 model	99.65% Training Accuracy and 0.02595%training loss and 99.62% test accuracy[10].
Total 200 Dataset is used for 38 Bangla alphabets.	CNN	Got 86% of Accuracy rate.[21]
Two datasets are used each containing 2400 samples to train the system.	(KNN)K-nearest neighbors.	Success rate 97.7% for recognizing Bangla signs[25].
Dataset is customized with 630 samples almost 11 statics signs these are hi, hello,good morning something like that	LSTM-GRU and RNN	88% accuracy rate[26]
MNIST image dataset	Used the K-means algorithm.	Got 89% of Accuracy rate[9].
KArSL (Arabic Sign Language Dataset)	CNN and KNN	Vision Transformer for sign language recognition attains 93.72% accuracy.[24].
Uses Microsoft Kinectto collect info from hand motions and hand signs are recognized from hierarchical conditional random field (CRF)	CNNs and Deep Neural Network	94.65% Training Accuracy and 0.01595%training loss and 94.62% test accuracy[27].

CHAPTER-THREE

METHODOLOGY & MODELING

3.1 Details Explanation

This chapter will cover the detailed explanation of methodology that is being used to make this project complete and working well. Many methodology or findings from this field mainly generated into journals for others to take advantage of and improve as upcoming research on projects. The methodology refers to the overall approach that our project requires. The methods are the tools of data collection, the procedure of our project. The procedures or strategies used to find, select, process, and analyse information about a topic are referred to as methodology. A particular procedure or set of procedures demonstrating the issue is a massive revision of teaching methodology. In a report or article, the methodology section allows the reader to critically evaluate a study's overall validity and reliability. So, this methodology chapter explains what we did and how we did it.

3.2 Data Set

In this project we made our novel dataset for both training and also for testing purpose. We take 50000+ samples of the 50 alphabets which consists of 11 stand-alone vowels and 39 consonant + vowel for the dataset. Using OpenCV these samples were collected and also observed the depth of the corresponding data. We collected each sample for the dataset using WebCam in which we take samples from each individual from different background. In our system we used 50000+ samples of data for training purposes and 500+ samples of data for testing purposes. Here we visualize raw input for the dataset in the following.

Here we collected Data samples by a camera and also spatially aligned the data which are in RGB form into data. We use the Mediapipe and OpenCV for offline data alignment

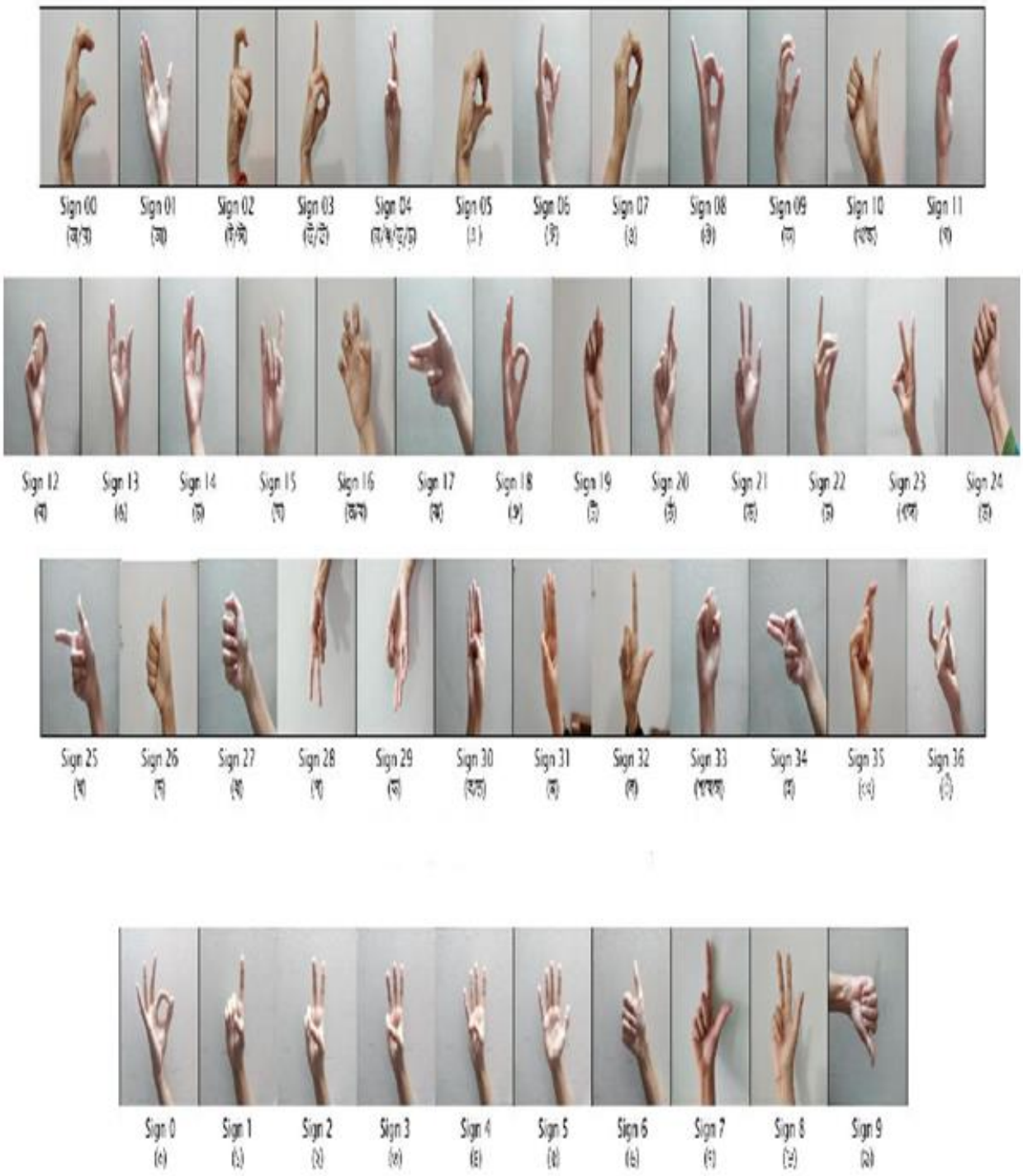

Fig:3.1 Bangladeshi Sign Language [28]

3.3 Flowchart Of Implementation

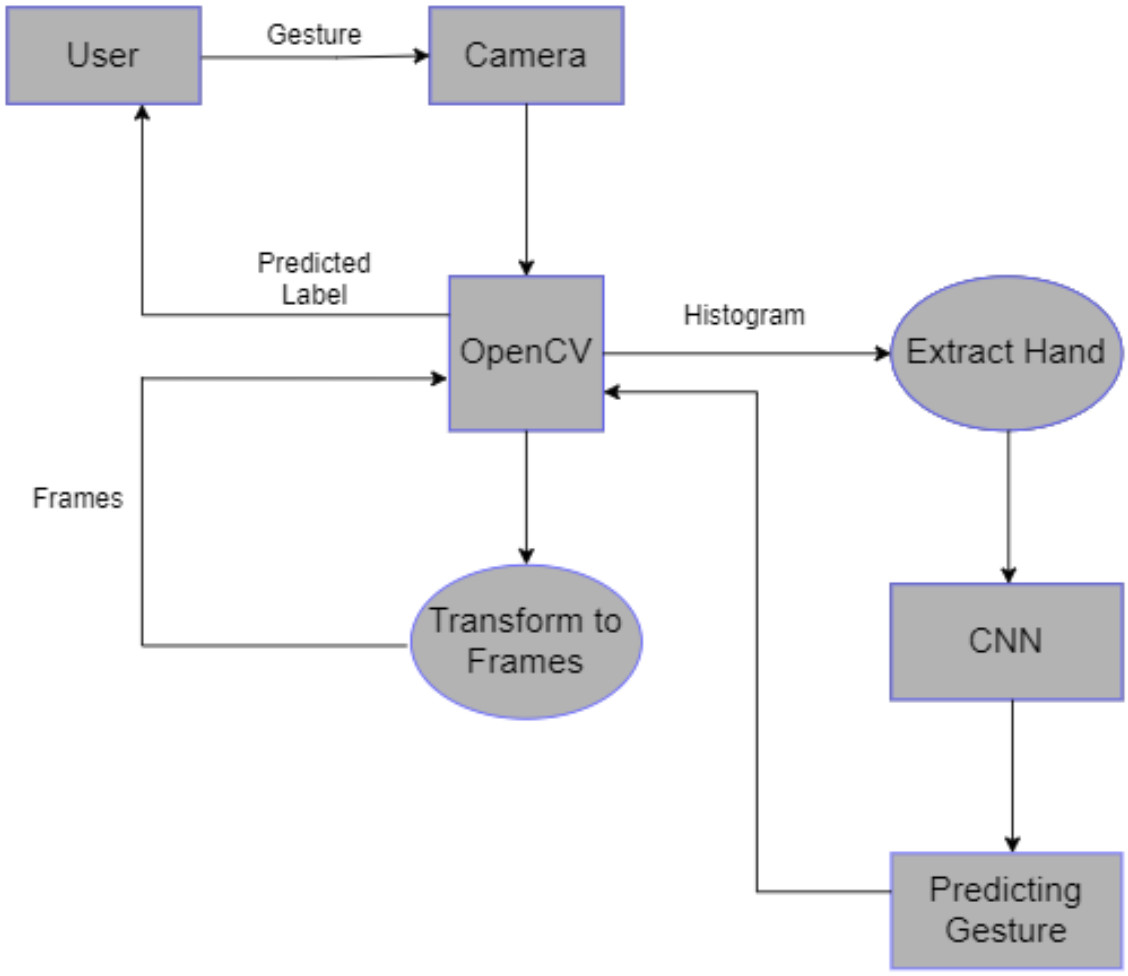

Fig:3.2 Working Flowchart Of Sign Language Detection

Here Fig 3.3 shows the proposed system detailed idea. We have a four module system. Using a webcam first we capture the image. For a good background we place the webcam in the top of the proposed system .Firstly we have to convert the image which are coloured into images of grayscale. And latter we convert grayscale image into the binary scale. Coordinates are calculated which correspond to X coordinates. Then we compare the coordinates which are recently created with the existing coordinates. Whenever this comparison gives us the success, then the a 9udio input will be converted into text form. our proposed system has two modes

i.e.one is the mode of training and another is the mode of operation. Alphabet Recognition is the part of a training mode which is used in machine learning. These technique is used in our system.

3.3.1 Project Implementation Phases

Phase 1:

Data Collection Phase: A model is being built and a sequence of images is being fed to the model. This phase is useful to further train the model based on the type of symbol.

Phase2:

Training And Testing Phase: This phase involves a set of inputs to the model and a particular output is being expected. Based on the outputs, the accuracy of the model is being identified.

Phase 3:

Recognition Of Output: In this phase, an image is being given as input on our system. And the images are trained with the model, it matches the image with a particular output and generates a message. The sign being identified in the above phase has a particular meaning and a message is being assigned in the training phase. The meaning is displayed to the user.

3.4 System Requirement

3.4.1 Specification Of Software Requirement:

The prerequisites software libraries for the sign language project are:

· Python (Version 3.7.13)

· IDE (PyCharm)

· NumPy (version 1.19.2)

· Cv2 (OpenCV) (version 3.4.2)

· Keras (version 2.3.1)

· For preprocessing the image we use TensorFlow

· Media Pipe (Version 22.0.4)

· MYSQL WORKBENCH (Version 8.0.29)

· MYSQL CONNECTOR (Version 5.7.24)

· PYCHARM (Version 2022.1)

3.4.2 Hardware Requirement Specification

Our Required Hardware are as follow:

· Camera: best quality, which has 3MP

· Ram: Whose Minimum storage should be 8GB

· GPU: must be 4GB

· Processor: processor should be Intel Pentium4

· HDD/SSD: minimum storage 10GB

· Monitor: Should have 15” or 17” color display monitor

· Mouse: Touch Pad

· Keyboard

3.4.3 Limitations

· Accuracy is low.

· the dataset of gestures cannot be used by other frameworks.

· missing some gestures because of the classification algorithm's accuracy contrast.

· Sometimes gesture recognition fails due to background noise of the given input image/ video frame.

3.5 Algorithm Details

3.5.1 Convolutional Neural Network

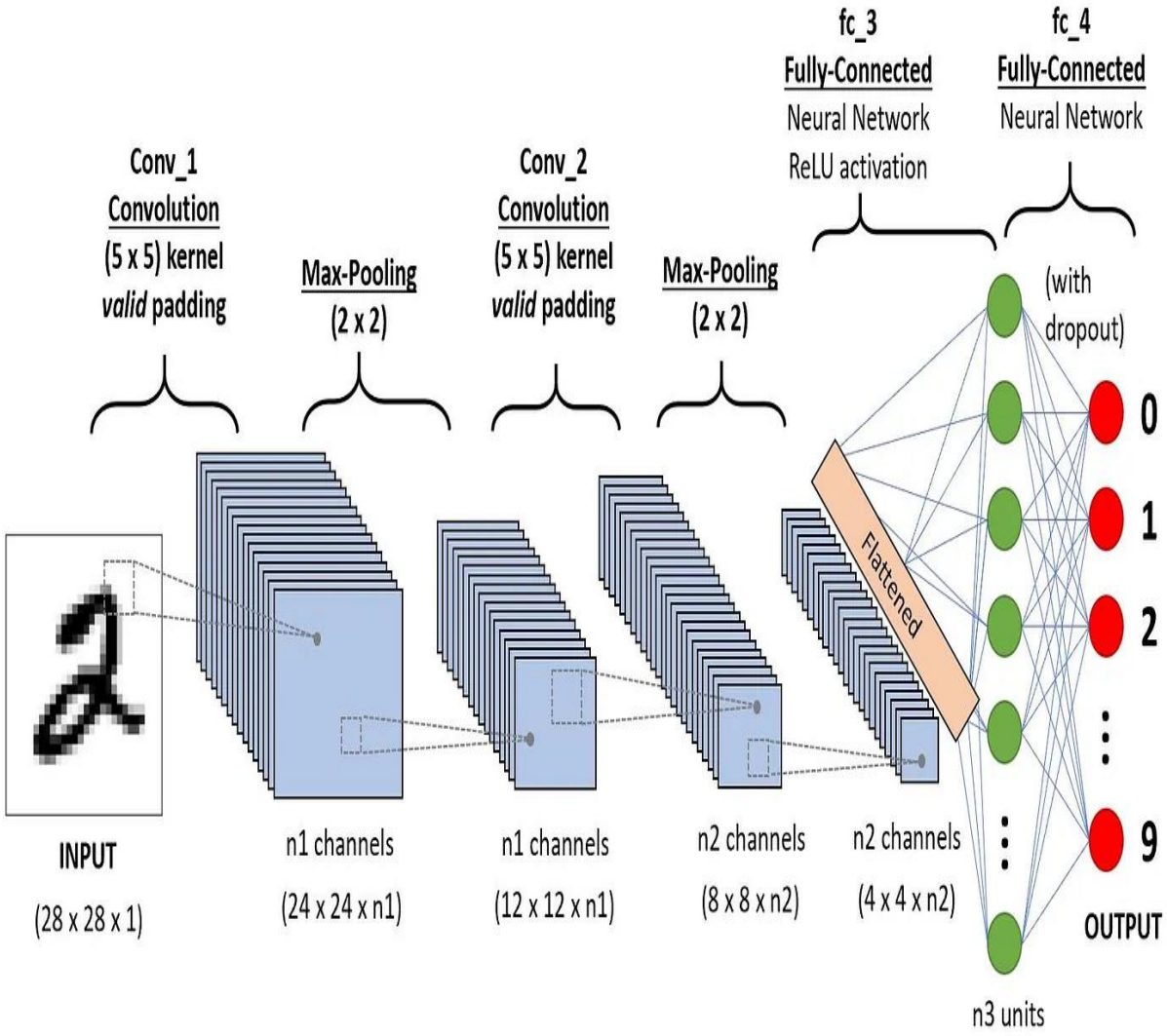

Fig:3.3 Convolutional Neural Network[29]

We employed a deep learning technique that makes use of the CNN to take the input image from the system, and in this process, different objects in the image are allocated in order to create

efficiency from one another. Comparatively speaking to other algorithms, the ConvNet needed the pre-processing stages. Due to the need for filters for effective training in the basic approaches.

A ConvNet's architecture was influenced by how the Visual Cortex is organized and is similar to the connection network of neurons in the human brain. Only in this constrained area of the visual field, known as the Receptive Field, do individual neurons react to stimuli. The whole visual field is covered by a series of such fields that overlap.

The ability of artificial intelligence to close the gap between human and computer skills has been growing dramatically. Both professionals and amateurs focus on many facets of the field to achieve great results. The field of computer vision is one of several such disciplines. The goal of this field is to make it possible for machines to view and understand the world similarly to humans, and to use that understanding for a variety of tasks like image and video recognition, image analysis and classification, media recreation, recommendation systems, natural language processing, etc. With time and a focus on one particular algorithm, the convolutional neural network, deep learning in computer vision has developed and been refined.

3.5.2 Analysis Using KNN Algorithm

A supervised machine-learning method for gesture identification is the KNN method. Each data point's classification is impacted by the classification of its neighbors when using KNN for classification. The Euclidean distance is used in KNN to identify the nearest neighbor [23]. The computation is done utilizing a variety of smaller distances with the aim of achieving the shortest Euclidean distance feasible. As the k value increases, accuracy also increases.

Typically, the Euclidean distance formula is employed. By averaging the k data points that are closest together, KNN does categorization using a threshold value that is determined. The closest neighbor's distance, a similarity test, and a threshold value constitute the basis of the performance entirely.

The size of the hidden layer is determined by the hidden layer's number of neurons, which yields a measurement of accuracy. The learning rate is determined and represented by leaning rate in double, and weight optimization is carried out using a solver. Scikit-learn is where the entire configuration is located.

3.5.3 Analysis Using CNN Algorithm

CNN is well-known for popularity and has finest effects over specific techniques, ordinarily due to the fact it may gain the necessary detail esteems from the records picture and come to be acquainted with the contrast between diverse examples with the aid of using limitless examples in its guidance. Anyways, formerly, its advancement has been confined because of the speed of system computing. lately, due to the progression of semiconductor fabrication, the computing tempo of illustrations getting ready gadgets is becoming faster, and the bottleneck of gadget handling speed has authorized the CNN community to develop fast.Applying this CNN involves the following steps: first, a picture is inputted (which is interpreted as an array of pixels); second, processing and filtering must be done; and third, the results are obtained after the classification. To be employed in a layered architecture with numerous convolutional layers using kernels (or filters) and a Pooling operation, each model must first be trained, followed by testing.

3.6 System Design

3.6.1 System Architecture

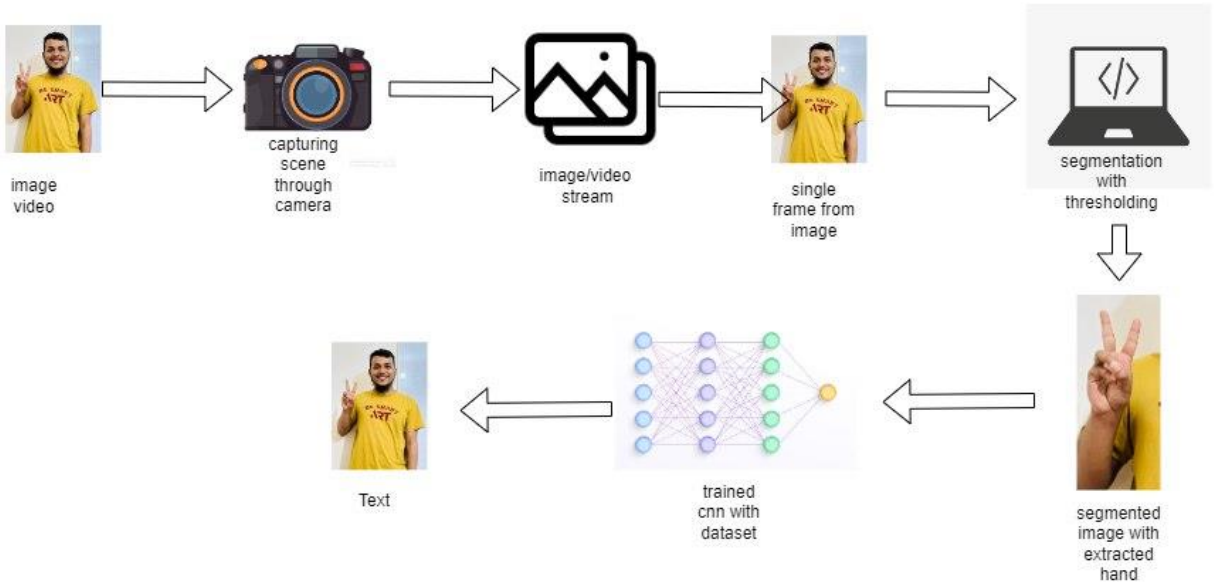

Fig:3.4 System Architecture Of Bangla Sign Language Recognition System

3.6.2 Neural Network Model

Artificial Neural Networks with feed onwards and back extension algorithms had been used. A specific input pattern is used to calculate the output by using Feed forward algorithm . At the same time, Back propagation algorithms are being used just for learning of the network. The network has been used by three layers of nodes. First layer being the input layer which takes 7 sensor values to the sensors over the glove. So 7 sensors of this layer. This layer just passes the values forward and it does not do any processing. Then the next layer being the secret layer, that carries the values to the input layer and applies the load on them.52 nodes of this layer. The third layer of this output. Actually, the third layer being the output layer in the networks, that takes it is input from secret layer and applies load to them. This layers have 26 nodes. Every node

explains at least one alphabet of the sign language subset. The final output passed by this layer. The final output is applied by a threshold. Just only the values upon this threshold and deliberated. No one of the nodes gives an output upon the value of threshold and no letter being outputted. If more than one node sends a value upon the threshold, no letter can not be outputted. The sigmoid function used by the activation function. When at both of the processing layers after the load have been applied then this activation function applied. We used this function in processing and instruction of the network. 4 times a second is done by sampling. In general, users must keep the sign accomplished for 3 or 4 and of a second to get it avowed. For faster performers, we see that this limit can be lowered.

3.6.3 Image Enhancement And Segmentation

In this webcam are RGB images captured , but RGB images are immensely sensory for different light conditions so RGB information converts into Y Cb Cr. Where Y is the luminance component which explains luminance fact of image, and Cb, Cr are chromo components that give colour information of images of red variety & blue difference. Luminance components may create problems therefore mere chrominance components get processed further. After that the YCbCr image was converted to binary.

3.6.4 Colour Filtering And Skin Segmentation

In reality, images are captured by a web camera containing a gathering of frames. There is a need to convert HSV image frames from RGB images, because it's connected to human colour savvy. Materially the colour spaces are differentiated into three components: (1) hue (H), (2) saturation(S), (3) value (V). The hand object and boundaries of images are typically performed to locate in image segmentation, users to specify boundary of skin colour in terms of hue and saturation value for this HSV features helps. It is easy to classify skin colour and non-skin colour information in images when Value gives brightness information. In this access, adjust the value of HSV within range 0 to 255 to the substance and get an accurate boundary of object [24].

3.6.5 Noise Removal: Erosion And Dilation

Morphological operations are the set of operations which perform on the image based on shapes . Morphological operations have two basic operations: (1) Erosion & (2) Dilation, it is used for Removing noise, individual elements and joining misaligned elements in an image are separated, even Finding of severity percussion or den in an image. Erosion curtails boundaries of an image and expands holes; Erosion must be used to dispel noises from an image. Dilation is used to add pixels at regions of boundaries or to fill in a den that is generated during the erosion process. Dilation may also be used to join disjoint pixels and affix pixels at edges.

3.6.6 Thresholding

The simplest method of image segmentation is Thresholding. Fold every pixel in the image into black pixel, if image severity is less than some actual and a white pixel if severity is greater than constant value. An initial property which pixels in an image can partition is its severity. Therefore in thresholding images are separated into regions which are depending on Light and dark regions.

3.7 Summary

The mentioned methodology for Sign language recognition using hand gestures being narrated in this chapter. Bengali Sign language is detected by utilizing different processes and is thoroughly presented by the help of an accurate diagram and explanation.

CHAPTER-FOUR

RESULTS AND DISCUSSION

4.1 Experimental Result

In earlier sections, we have discussed our execution of the hand segmentation, unaccountable feature learning and classification sub-blocks. In this segment, we report our system's efficiency using tables and figures. Our principle appraisal metric is developed on the exactness of the classification. Provided an undivided picture of a person signing a letter, we expect to see the correctness in predicting the signed letter in the image of our model. The precision exactness is about 90.04% the same as many MNIST digit identification is expected. As an initial diagnostic, we scattered a learning inclination presenting the training and test inaccuracy as a function of training set size.

Our system has managed to achieve 96.01% training exactness, with 90.04% test correctness in letter detection, and number detection of 93.44% . Achieved an average of 93.667% based on limited time sign recognition. The training of each system involved utilizing 5,000 images with a size of 240 x 240 pixels for each of the letter, number, or word gestures.

[30] With 3D-CNN has an accuracy of 94.2%, while another dataset has the accuracy of 78.8% and [36] has an accuracy of 84% in contest dataset. [31] EFD and ANN has an accuracy of 91.50% and [32] has an accuracy of 80%. 98.90% accuracy is achieved by using SimpSVM method[33]. and 97.5% accuracy achieved by using SVM[34]. Table 4.1 shows the comparison result of the Sign Language Recognition system proposed by researchers.

Table- 4.1: Comparison Result of Various Method in Sign Language Recognition

Authors	Method	Accuracy Result
[30]	3D-CNN	94.20%
[31]	EFD and ANN	91.50%
[36]	CNN	84%
[32]	SOM	80%
[33]	SimpSVM	98.90%
[35]	SVM and HMM	85.14%
[34]	SVM	97.50%

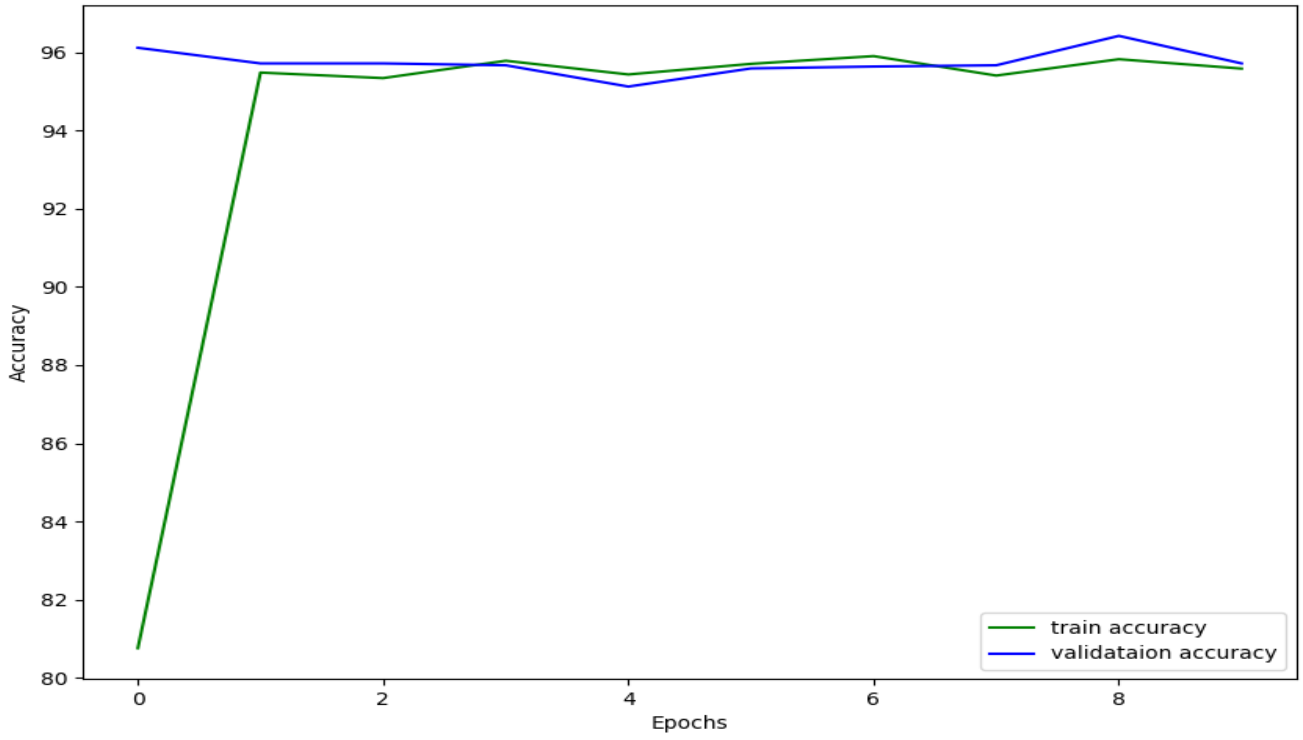

Fig 4.1 : Epochs Vs Accuracy

By collecting a dataset of 5,000 images from 3 students and implementing the CNN algorithm on Bengali Sign Language detection. From distinct angles, directions, backgrounds and lighting conditions pictures and videos were taken. The size of the images was fixed to 240x240. As a deep learning library, TensorFlow was used. algorithms.

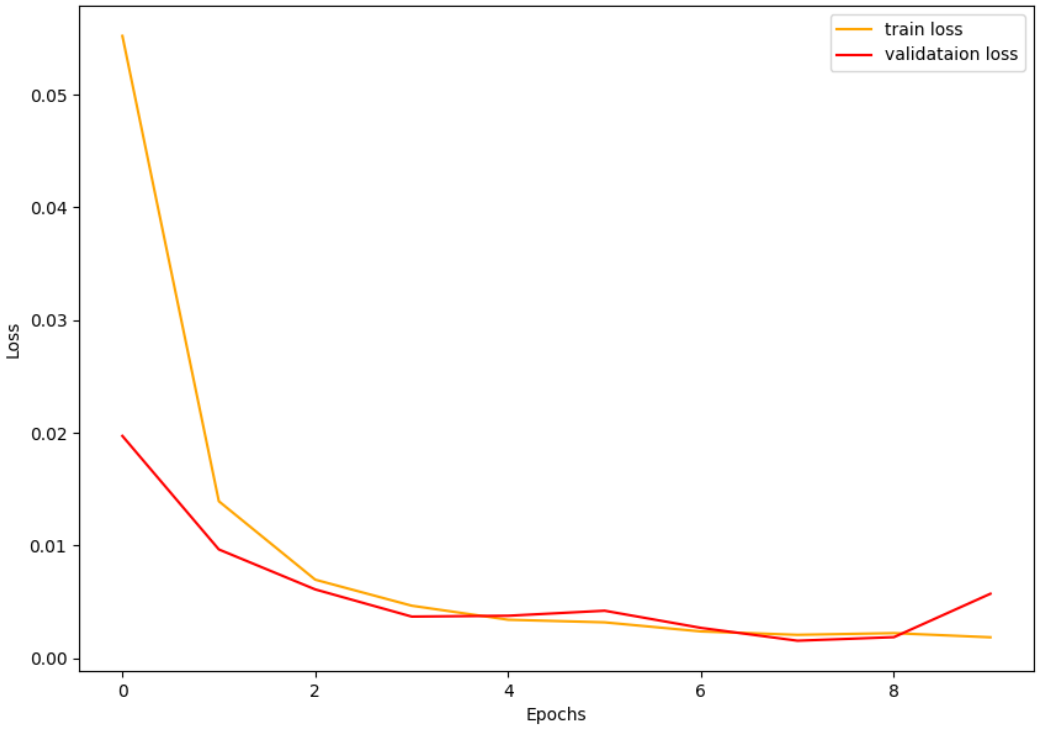

Fig 4.2: Epochs Vs Training Loss

4.2 Summary

In our progress report, we used 50 concealed units for our autoencoder. We observe by surveying our learning curve that despite one aberration at training set size 3000, the training error and the test error are close to each other, and even if the training inaccuracy is greater than 1.5%, that is somewhat great. As a result, we suspected that we were in a region of high bias, so we decided to

increase the size of our features by increasing the number of hidden units of our autoencoder to 100. Using this 100-length feature vector, we achieved 98% classification accuracy on the test set.

CHAPTER-FIVE

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Images are now widely used as a source of information for interpretation and analysis in applications. The conversion of an image from one form to another includes digitization, scanning, and communication, storage etc. In order to enhance the appearance of the output image, it undergoes a process called image correction, which includes a set of methods that aim to enhance its appearance. Enhancing images means making images clearer or more understandable to humans, so other image processing systems can get better input. The picture function is extracted using different forms to make the picture of the computer more readable. Sign language recognition system is a powerful tool for preparing expert knowledge, on-board detection and combining incorrect information .Sign language recognition system is a powerful tool for preparing expert knowledge, on-board detection and combining incorrect information from various origins. The purpose of the neuronal convolution network is to obtain the proper classification.

This project was carried out with appropriate support and consultation from experts in sign language. One of the aims of the study was to reach the developmental training stage or even surpass the precision of studies submitted through deep learning. Our system has managed to achieve 99% training exactness, with 90.04% test correctness in letter detection, number detection of 93.44% and static word recognition is 97.52% , Achieved an average of 93.667%

based on limited time sign recognition. The training of each system involved utilizing 5,000 images with a size of 240 x 240 pixels for each of the letter, number, or word gestures.

The compilation of Bangladeshi Sign Language letters and symbols dataset will be distributed among the scholarly circles. The work's objective is to cater to deaf individuals who have fluency in sign language. Therefore, it would be beneficial to extend the service to more skilled and capable signers. However, the amount of collected images from non-fluent users exceeded that of the fluent users, which suggests that the images may not meet the needs of all users.

5.2 Future Work

The sign language recognition system designed for identifying letters in sign language has the potential to go beyond that and identify facial expressions and gestures as well. To provide a more precise interpretation of a language, it is preferable to present sentences instead of letter tags. This can also enhance the overall legibility of sign language, as it increases the range of sign language options available. Additional training data can be incorporated to enhance the accuracy of identifying the letter. This endeavor can also be broadened to transcribe symbols into spoken language.

Currently, the outcomes indicate that the dataset comprising characters and digits have remarkable precision, but it is important to note that the CNN design was influenced by my past experiences with certain tests and tutorials, as well as the structure which was introduced. When incorporating additional symbol classes to recognize, the precision of the model may decrease, which could result in reduced system performance. To enhance the system's framework, our suggestion is to begin with a set of gestures and evaluate different configurations of the CNN to identify the most optimal one for the given scenario. The group of gestures will be accelerated with extra classes and, using the assessment strategies, pick out the excellent configuration. From the relation number of training and best configurations, our goal is to discover a relation between the quantity of instructions and layout of the CNN so the gadget can assemble an effective version for every case. Once the letter recognition device has done excessive recognition, a logical step in the direction of the SL transcription is the concatenation of expected letters to

assemble words.Errors can occur during the recognition stage of the transcribing process since the CNN for letter recognition is not fully accurate.

We can also enhance our work so that we can accommodate different communities' languages datasets like Chakma and Monipuri sign languages.

REFERENCES

[1] N Sarawate, M Chan LEU, and Cemil OZ, “A real-time American Sign Language word recognition system based on neural networks and a probabilistic model,” Turkish Journal of Electrical Engineering & Computer Sciences, vol. 23, no. 1, pp 2107-2123, 2013, doi: 10.3906/elk-1303-167.

[2] B Garcia and S Alarcon Viesca, “Real-time American Sign Language Recognition with Convolutional Neural Networks,” Convolutional Neural Networks for Visual Recognition, vol. 2, pp. 225-232, 2012.

[3] Z Z. Ren, J. Yuan, J. Meng and Z. Zhang, "Robust Part-Based Hand Gesture Recognition Using Kinect Sensor," in IEEE Transactions on Multimedia, vol. 15, no. 5, pp. 1110-1120, Aug. 2013, doi: 10.1109/TMM.2013.2246148.

[4] Bengali Sign Language Dictionary, National Centre for Special Edu-cation, Ministry of Social Welfare in cooperation with the Norwegian Association of The Deaf and Bangladesh National Federation of The Deaf, 1997

[5] National Federation of The Deaf Annual Report by Bangladesh National Federation of The Deaf
National Federation of The Deaf Annual Report by Bangladesh National Federation of The Deaf.

[6] Foez M. Rahim, Tamnun E Mursalin, Nasrin Sultana, “Intelligent Sign Language Verification System – Using Image Processing,Clustering and Neural Network Concepts” International Journal of Engineering Computer Science and Mathematics, ISSN:0976-6146, Vol.1 No.1 January-June2010.

[7] 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.

[8] Geetha M, Manjusha C, Unnikrishnan P, Harikrishnan R (2013) A vision based dynamic gesture recognition of Indian sign language on Kinect based depth images. In: 2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (c2SPCA). IEEE, pp 1–7

[9] Agarwal, S.R.; Agrawal, S.B.; Latif, A.M. Article: Sentence Formation in NLP Engine on the Basis of Indian Sign Language using Hand Gestures. Int. J. Comput. Appl. 2015, 116, 18–22.

[10] David Forney G (1973) The Viterbi algorithm. Proc IEEE 61(3):268–278

[11] Geoffrey Hinton, "Training products of experts by minimizing contrastive divergence," Neural computation,vol. 14, no. 8, pp. 1771–1800, 2002.

[12] Karmokar, Bikash Chandra, Kazi Md Rokibul Alam, and Md Kibria Siddiquee. "Bangladeshi sign language recognition employing neural network ensemble." International journal of computer applications 58.16 (2012): 43-46.

[13] Das, S., Imtiaz, M. S., Neom, N. H., Siddique, N., & Wang, H. (2023). A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier. Expert Systems with Applications, 213, 118914.

[14] Sadik, Farhan, et al. "Bangla sign language recognition with skin segmentation and binary masking." 2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE). IEEE, 2019.

[15] Podder, Kanchon Kanti, et al. "Bangla sign language alphabet recognition using transfer learning based convolutional neural network." Bangladesh J. Sci. Res (2020): 31-33.

[16] Nihal, Ragib Amin, et al. "Bangla sign alphabet recognition with zero-shot and transfer learning." Pattern Recognition Letters 150 (2021): 84-93.

[17] Podder, Kanchon Kanti, et al. "Bangla sign language (bdsl) alphabets and numerals classification using a deep learning model." Sensors 22.2 (2022): 574.

[18] Islam, Sanzidul, et al. "A potent model to recognize bangla sign language digits using

convolutional neural network." Procedia computer science 143 (2018): 611-618.

[19] Angona, Tazkia Mim, et al. "Automated Bangla sign language translation system for alphabets by means of MobileNet." TELKOMNIKA (Telecommunication Computing Electronics and Control) 18.3 (2020): 1292-1301.

[20] Hossein, Md Jahangir, and Md Sabbir Ejaz. "Recognition of Bengali sign language using novel deep convolutional neural network." 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI). IEEE, 2020.

[21] Forster J, Schmidt C, Hoyoux T, Koller O, Zelle U, Piater JH, Ney H (2012) Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus. In: LREC, pp 3785–3789

[22] Gonzalez M, Collet C, Dubot R (2010) Head tracking and hand segmentation during hand over face occlusion in sign language. In: European Conference on Computer Vision. Springer, pp 234–243

[23] website: https://www.kaggle.com/datasets/grassknoted/asl-alphabet .asl-dataset image Last Access Date:May 20,2030

[24] Aliaa A. A. Youssif, AmalElsayedAboutabl, HebaHamdy Ali, "Arabic Sign Language (ArSL) Recognition System Using HMM ", (lJACSA) International Journal of Advanced Computer Science and Applications, Vo1.2, Issue. 11,2011

[25] P. Y. Simard. Best practices for convolutional neural networks applied to visual document analysis, August 2003. Seventh International Conference on Document Analysis and Recognition.

[26] X. Teng. A hand gesture recognition system based on local linear embedding, April 2005. Journal of Visual Languages and Computing.

[27] Geetha M, Manjusha C, Unnikrishnan P, Harikrishnan R (2013) A vision based dynamic gesture recognition of Indian sign language on Kinect based depth images. In: 2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (c2SPCA). IEEE, pp 1–7

[28] Hasan, Md Mehedi, Azmain Yakin Srizon, and Md Al Mehedi Hasan. "Classification of Bengali sign language characters by applying a novel deep convolutional neural

network." 2020 IEEE Region 10 Symposium (TENSYMP). IEEE, 2020.

[29] Lahamy, H.; Lichti, D.D. Towards real-time and rotation-invariant American Sign Language alphabet recognition using a range camera. Sensors 2012, 12, 14416–14441.

[30] Huang J, Zhou W, Li H, Li W. Sign language recognition using 3D convolutional neural networks. In Multimedia and Expo (ICME), 2015 IEEE International Conference on; 2015: IEEE. p. 1-6.

[31] Kishore PVV, Prasad MVD, Prasad CR, Rahul R. 4 Camera model for sign language recognition using elliptical fourier descriptors and ANN.In International Conference on Signal Processing and Communication Engineering Systems - Proceedings of SPACES 2015, in Associationwith IEEE; 2015. p. 34-38

[32] Tewari D, Srivastava S. A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self Organizing Map Algorithm. International Journal of Engineering and Advanced Technology (IJEAT). 2012; 2(2): p. 165-170

[33] Thang PQ, Dung ND, Thuy NT. A Comparison of SimpSVM and RVM for Sign Language Recognition. Proceedings of the 2017 International Conference on Machine Learning and Soft Computing - ICMLSC '17. 2017;(1): p. 98-104.

[34] Raheja JL, Mishra A, Chaudary A. Indian Sign Language Recognition Using SVM 1. Pattern Recognition and Image Analysis. 2016 September; 26

[35] Greg C. Lee & Fu-Hao Yeh & Yi-Han Hsiao. Kinect-based Taiwanese sign-language recognition system. Multimed Tools Appl. 2014 October

[36] Krizhevsky A, Skutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Advances In Neural Information Processing Systems. 2012;: p. 1-9