Stochastic Modelling and Computational Sciences

INDIAN SIGN LANGUAGE RECOGNITION SYSTEM: EFFECTIVE COMMUNICATION FOR DEAF AND DUMB

Dr Latha N R[1], Dr Shyamala G[2], Dr. Pallavi GB[3] and Sumit Gupta[4]
[1,2,3]Associate Professor, Department of Computer Science and Engineering, BMS College of Engineering
[4]Research Head, R&D Department, Xsys Software Technologies, Bangalore, India
[1]latha.cse@bmsce.ac.in, [2]shyamala.cse@bmsce.ac.in, [3]pallavi.cse@bmsce.ac.in and
[4]sumit@xsyssoftwaretechnologies.com

ABSTRACT
In India, where sign language is not widely used, communication without an interpreter can be challenging. Therefore, there is a need for a device that can transcribe sign language symbols into text, enabling real-time communication and providing immersive training for learning sign language. Sign language relies on hand gestures to convey meaning. While extensive research has been conducted on American Sign Language (ASL), Indian Sign Language (ISL) remains underexplored. The primary challenges hindering ISL research include the lack of standardized datasets, occluded features, and regional variations in the language. This work aims to address these challenges by focusing on ISL character classification, compiling a dataset, and analyzing various features. The project seeks to develop a machine learning model capable of recognizing hand gestures used in sign language for finger spelling. A user-independent model is trained using classification-based machine learning algorithms on a dataset of images, and performance is evaluated through testing. Convolutional Neural Network (CNN) algorithms are applied to improve accuracy in recognizing ISL gestures.

Keywords: *gesture, pre-processing, feature extraction, Communication, Convolutional Neural Network*

I. INTRODUCTION
Humans are naturally gifted with speech, enabling them to connect and communicate with one another. Communication is a fundamental necessity for social life, making spoken language a crucial aspect of human interaction. However, individuals with hearing and speech impairments lack this ability. While deaf and mute individuals use sign language to communicate, it remains challenging for others to understand without an interpreter. Indian Sign Language (ISL) conveys meaning through visually transmitted sign patterns and serves as an effective communication method for the deaf community in India. Despite significant research efforts on various sign languages worldwide, progress in ISL research has been limited. The lack of standardized datasets has hindered advancements, making ISL less developed compared to American Sign Language (ASL). Additionally, ISL requires the use of both hands for gestures, adding complexity to gesture recognition compared to ASL. The goal of an ISL recognition system is to develop efficient software that translates sign language gestures into text and speech, bridging the communication gap between the deaf and those who do not understand sign language.

Our proposed approach focuses on creating an application that facilitates seamless communication between deaf-mute individuals and the general public. This application leverages deep learning techniques, including image classification, pre-processing, feature extraction, and gesture recognition, to enhance accuracy and usability.

Several researchers have focused on sign language recognition in other countries, but ISL is different. A very small amount of research has been done in this area; hence it is viable to explore further and achieve feasible results which motivated us to take up this topic. Two hands are used by ISL for communicating whereas a single hand is used by ASL. Due to overlapping of hands, the use of both hands also contributes to obscurity of features. There are also very few datasets and a lot of variation in sign language with locality have resulted in restricted attempts to detect ISL gestures. Our work can be used as communication bridge between humans who can speak and the speech and hearing impaired. In this project we are adding words and common expressions which make the speech and hearing-impaired people communicate faster with each other and easier with the outer world.

Thus, the recognition system converts gestures in sign languages into letters, words and sentences which can be further converted into audio that helps in real-time communication.

II. LITERATURE SURVEY

Sign language is a language in which communication takes place to convey the message by visually communicating the sign patterns. It is a vital tool for communication among the hearing and speaking impaired. Several researchers have focused on various sign languages, such as American Sign Language, British Sign Language, Taiwanese Sign Language, etc., but few studies on the Indian Sign Language have made progress. The deaf community in India employs Indian Sign Language (ISL). ISL is a common communication form in India for people with hearing impairments. ISL lags behind American Sign Language because there is a shortage of standard datasets for study in this area. It utilizes both hands to make gestures, unlike American Sign Language, which increases the difficulty of understanding gestures. The aim of the ISL Recognition system is to provide effective software that translates sign language into text and speech in the form of gestures.

Sanil jain and K.V. Sameer Raja [1] proposed a work which aimed towards identifying alphabets in Indian Sign Language from the corresponding gestures. They collected the dataset with the support of Jyothi BAdhir Vidyalaya, a school for deaf located in a remote section of Bithoor, Kanpur. Their approach consisted of three stages:

- The initial stage was to segment the skin part from the image.
- The next stage was to extract relevant features from the skin segmented images.
- In the third stage extracted features were used as input for training and learning models and use the trained models for classification.

Due to bad quality of images and lack of dataset they couldn’t achieve feasible results.

Muthu Mariappan H and Dr Gomathi V [2] proposed a FCM based system to recognize the real-time signs. A web camera was for capturing the gestures. The real-time sign language recognition system was designed as a portable unit. The raw videos taken in a dynamic background was given as an input to the system. The image frames were resized to maintain the equality among all the videos. OpenCV was used for feature extraction and video classification. The data samples were collected for 80 words and 50 sentences. The videos were recorded using a digital camera. The proposed system achieved 75% accuracy. Though FCM is efficient, it requires more computation time than the others. The algorithm also suffered for high dimensional datasets.

Brian Toomajian and Youngwook Kim [4] proposed Hand Gesture Recognition Using Micro-Doppler Signatures with Convolutional Neural Network. The feasibility of recognizing human hand gestures was investigated by using microDoppler signatures measured by Doppler radar with a deep convolutional neural network (DCNN). The classification accuracy of the proposed method was found to be 85.6% for ten gestures .The accuracy increased to 93.1% with seven gestures. As practical applications require high classification accuracy, using seven gestures was suggested when Doppler radar is used.

A. *Supervised Algorithms:*

The SVM supports both dense and sparse sample vectors as inputs. SVM finds the suitable plane that maximally separates the two classes. However, using an SVM to make predictions for sparse data would have been adequate for such information.

RANDOM FOREST: Random forest or random decision forests is a machine learning algorithm that uses supervised learning. "The forest" it constructs is an ensemble of decision trees, usually trained in the method of bagging. The basic concept of the method of bagging is that the cumulative outcome is improved by a mixture of learning models. Random forest constructs and merges several decision trees to achieve a more precise and stable prediction.

B. ***Hierarchical Classification:***

A hierarchical classifier is a classifier that maps input information into specified categories of subsumptive output. This output may be one of a collection of pre-defined outputs, one of a set of on-line learned outputs, or even a new novel classification that has not been seen before, depending on application-specific specifics. Sanil jain and K.V. Sameer Raja [1] proposed a project aimed at distinguishing alphabets from the corresponding gestures in Indian Sign Language. With the help of Jyothi Badhir Vidhyalaya, a deaf school located in a remote part of Bithoor, Kanpur, they compiled the dataset. Their method consists of three phases:

- The first stage was the segmentation of the skin section from the picture.
- The second stage was to extract relevant characteristics from the images segmented by the skin .
- The third stage was to use extracted features for training and use the trained models for classification as input into various supervised learning models. They could not produce feasible results due to low picture quality and lack of data collection.

C. ***Unsupervised Algorithms:***

The Fuzzy c-means (FCM) algorithm is an efficient algorithm for clustering and is commonly used in image segmentation. In FCM, however, we need to provide both the number of clusters and the initial membership matrix parameters in advance, and they have a strong effect on the efficiency of the clustering. Muthu Mariappan H and Dr Gomathi V [2] proposed a FCM-based framework to classify the real-time signals. To record the movements, a web camera was used. As a portable device, the real-time sign language recognition scheme was developed. The raw videos that were taken in a complex sense were provided as an input to the system. To ensure equality between all the videos, the picture frames were resized. For feature extraction and video classification, OpenCV was used. Data samples for 80 words and 50 phrases were collected. The videos were filmed using a digital camera. While FCM is successful, it takes more computing time than the others. For high-dimensional datasets, the proposed method achieved 75 percent precision. The algorithm also suffered for high dimensional datasets.

D. ***Neural Network CNN:***

CNNs are similar to conventional ANNs in that they are composed of neurons that achieve self-optimization through learning. An input will still be received by each neuron. CNNs consist of three types of layers. Those layers are convolutional layers, pooling layers and fully-connected layers. ANN: Artificial neural networks are computer structures loosely inspired by the biological neural networks that make up animal brains, typically simply called neural networks. An ANN is based on a set of artificial neurons called linked units or nodes that loosely model the neurons in the biological brain. Brian Toomajian and Youngwook Kim [4] proposed Hand Gesture Recognition Using Micro-Doppler Signatures with Convolutional Neural Network. Using microDoppler signatures measured by Doppler radar with a deep convolutional neural network (DCNN), the feasibility of identifying human hand movements was investigated. For ten gestures, the classification accuracy of the proposed method was found to be 85.6 percent. With seven gestures, the precision increased to 93.1 per cent. As practical applications involve high accuracy of classification, when Doppler radar is used, the use of seven gestures was proposed. Yogeshwar I Rokade, Prashant M. Jadav [14] suggested a way for the automated recognition of the finger spelling in the Indian Sign Language. They provided the sign as an input to the machine in the form of gestures. The suggested approach includes segmenting the hand on the basis of skin color statistics, then converting the segmented image to binary, applying binary image feature extraction, feature extraction, and classification and recognition. To accomplish this, they used two separate approaches: the Artificial Neural Network (ANN) system and SVM. They concluded in this experiment that the use of ANN gives a high accuracy of 94.37 percent over 92.12 percent SVM. They have used publicly accessible databases for training and research in this scheme. This model was unable to predict sentences in real time and also had a lack of data since only existing datasets were used. An IEEE paper “Hybrid SIFT Feature Extraction Approach for Indian Sign Language Recognition System Based on CNN” [16] published by Abhishek Dudhaland his peers: Glove-based systems incorporated were expensive and difficult to use; in contrast, the image classification-based system proposed in this paper is much cheaper and easier to use. This paper proposed a method for Indian sign language recognition using CNN classification and feature extraction by hybrid SIFT.CNN is strong and stable in such a way that very little image preprocessing is necessary. The proposed solution obtained a 92.78 percent validation precision for CNN with a hybrid SIFT approach. With CNN's adaptive thresholding strategy, 91.84 percent accuracy was achieved. The framework suggested in this paper can function only with a laptop and webcam and can therefore be used by the hearing-impaired population with mobility. For learning Indian sign language, the scheme may also be used. Neel Kamal Bhagat, Vishnusai Y and Rathna.G N [13] proposed a real time hand gesture recognition system based on the data captured by the Microsoft Kinect RGB-D camera. For the training of 36 static gestures relating to Indian sign language, they used Convolutional Neural Networks (CNN). For preparation, they generated their own dataset. The dataset was split into two parts, one for gestures based on static and the other for gestures based on dynamic. The dataset was obtained from five subjects, including male and female genders, with an average age of 25.

The static dataset comprised 45,000 camera depth-based images and 45,000 RGB camera-based images. For dynamic, they captured videos pertaining to 10 commonly used words of ISL. Custom datasets were generated and different models were used for training. The static model achieved accuracy of 98.81% and dynamic model achieved 99.08 % accuracy on the training set. This model was not able to achieve real time prediction of more words associated with ISL and on sentence formation. Neel Kamal Bhagat, Vishnusai Y and Rathna.G N [13] proposed a real time hand gesture recognition system based on the data captured by the Microsoft Kinect RGB-D

*E. **Rule-based Classifier:***

Rule-based classifiers are just another form of classifier which, depending on the use of different "if...else" rules, makes the class decision. These rules are easily interpretable and are therefore commonly used to create descriptive models for these classifiers. An IEEE paper has been published by Zhi-hua Chen Yuan [8] proposing the hand gesture recognition approach as follows: A new technique was developed for the detection of hand movements. The method of background subtraction is used to detect the hand area from the background. Then the palm and the fingertips are segmented. On the basis of segmentation, the fingers in the hand picture are discovered and recognized. To accept hand gestures, a simple rule classifier is used. A data collection of 1300 hand pictures was used to test the performance of this method. The experimental results have shown that this method has worked well and is suitable for applications in real time. Better results than the state-of-the-art FEMD on a picture set of hand movements were obtained by the proposed process. This method's efficiency depends on the outcome of hand detection. As a consequence of hand detection, moving objects with a color comparable to that of the skin occur and then degrade the efficiency of hand gesture recognition. A system has been developed in an IEEE paper published by Purva C. Badhe and Vaishali Kulkarni [9] that translates the gestures made into English in the Indian Sign language (ISL). The ISL translator framework helps in such a way that it inputs, decodes, interprets and outputs the gesture sense of the ISL sign in English. With the documented videos of deaf and mute signers, the database for the creation of this system is developed on its own. The database uses 78,000 separate videos from a total of 130,000 videos recorded to create the database. The input videos are transformed into frames to create a database and those frames are pre-processed in order to get the enhanced features. In a codebook, these features are then extracted and saved

*F. **B-Spline Cirve:***

A B-spline or base spline is a spline function that has limited support in the mathematical subfield of numerical analysis with respect to a given degree, smoothness, and domain partition. A linear combination of B-splines of that degree can be represented as any spline function of a given degree. Geetha M and Manjusha U C [12] suggested an approach using B-Spline approximation for a vision-based recognition of Indian Sign Language static signs. The approach consisted of the pre-processing stage, boundary tracing, finding the MCPs, boundary approximation to a B-Spline curve, resampling and smoothening, feature extraction, classification and recognition. An experiment was performed with 50 samples of each alphabet from A-Z and numbers from 0-

5.Full Curvature Points (MCPs) were taken as the control points in this algorithm and the boundary derived from the Region of Interest was approximated to a curve of B-Spline. After this step, the Main Maximum Curvature Points (KMCPs) were extracted and the B-Spline curve was then iterated for smoothening. The main contributors to the gesture form were KMCPs. For the number of samples, the proposed model achieved 92 percent accuracy, and 91 percent accuracy was achieved by the alphabets. LBPV Local Binary Pattern Variance (LBPV) is a texture characteristic where during local binary pattern (LBP) histogram computation, variance in contrast functions as an adaptive element. For texture analysis, the LBP incorporates both structural and statistical approaches. A system for recognition of sign language at the sentence level was introduced by H.S.Nagendraswamy and B.M.Chethan Kumara[15]. To extract the local contrast information from the frames of a video of signs, the LBPV definition was used. With the help of deaf people from the Mysore area, they carried out experiments on the UoM-ISL dataset that was developed by them. Compared to other classifier combinations, they achieved greater accuracy. However, on a broad dataset, the scalability of the proposed method has to be studied. To fix the issues of real-time situations, the issues related to continuous sentences in a sign video were considered. Published by Munir Oudah, the paper "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques"[12] focuses on a review of the literature on hand gesture techniques and addresses their merits and drawbacks in various circumstances. It focuses on the merits and drawbacks of the approaches that follow: Centered on Instrumented Glove Approach hand movements. Hand Movements Based on the methodology of computer vision.

- Recognition Based on Colors.
- Recognition Based on Appearance
- Recognition dependent on Motion
- Recognition Based on Skeleton
- Recognition Based on Depth
- 3D Recognition Based on Models
- Deep-Learning Recognition Based and on applications of each of the above methods.

G. ***Distance Metrics in ML: Euclidean Distance:***

The shortest distance between the two points is the Euclidean distance. The Euclidean is also the "default" distance used to find the "k closest points" of a given sample point, e.g., K-nearest neighbors (classification) or K-means (clustering). Another common example is hierarchical clustering, full and single linkage agglomerative clustering, where you want to find the gap between clusters. MANHATTAN DISTANCE: The distance between two real-valued vectors is determined by the Manhattan distance, also called the taxicab distance or the City Block distance. Vectors that represent objects on a uniform grid, such as a chessboard or city blocks, are perhaps more useful. Kumud Tripathi, Neha Baranwal and G.C.Nandi [5] worked on “Continuous recognition of Indian Sign Language Gesture and Sentence Formation”. They solved this issue using the method of gradient-based key frame extraction. To divide continuous sign language gestures into a series of signs and to delete frames that did not provide useful information, these key frames were used. In extracting pre-processed gesture characteristics, histogram orientation (OH) was used. In order to minimize the size of features obtained after OH, Principal Component Analysis (PCA) was applied. In the Robotics and Artificial Intelligence Laboratory (IIIT-A), tests were conducted using a canon EOS camera on their own continuous ISL dataset. For the testing of Probes, different forms of classifiers such as Euclidean distance, Correlation, Manhattan distance, city block distance etc. were used. In order to conduct a comparative analysis, different kinds of distance classifiers were used. The accuracy of the classification was calculated by the maximum number of matched frames. They found from this study that the results obtained from the distance from Correlation and Euclidean provided better precision than other classifiers. An approach where continuous video sequences of the signs were considered was used by Joyeeta Singha and Karen Das [10]. The proposed method consisted of the preprocessing stage, extraction of features and classification. The preprocessing steps involved skin filtering and matching histograms. For the Feature Extraction Point, Eigenvalues and Eigen Vectors were examined. The weighted Euclidean distance of the Eigen value was used to recognize the symbol. 24 separate alphabets to be considered in the video sequences were included in the dataset and the data was obtained from 20 individuals. In the archive, a total of 480 images were stored. The system was tested with 20 images. The proposed system obtained a 96.25 percent success rate. Yogeshwar I Rokade, Prashant M. Jadav [14] proposed a way for the automated recognition of the finger spelling in the Indian Sign Language. They gave the sign in the form of gestures as an input to the system. The proposed approach involves segmenting the hand based on the skin colour statistics, then convert that segmented image into binary, apply feature extraction on the binary image, extraction of features and classification and recognition. They used two different approaches: the Artificial neural network (ANN) method and SVM to achieve this. By this experiment they concluded that using ANN gives a high accuracy of 94.37 % over SVM with 92.12%. In this system they have used publicly available databases for training and testing. This model was not able to achieve real time prediction on sentences and it also had a lack of data as only existing dataset was used. H.S.Nagendraswamy and B.M.Chethan Kumara [15] proposed a method for sign language recognition at the sentence level . The concept of LBPV was used to extract the local contrast information from the frames of a video of signs. They did experiments on the UoM-ISL dataset which was created by them, with the support of deaf people from the Mysore region. They achieved better accuracy compared to other classifier combinations. However, the scalability of the proposed system has to be studied on a large dataset. The problems related to continuous sentences in a sign video were considered to address the issues of real time scenarios. An IEEE paper "Hybrid SIFT Feature Extraction Approach for Indian Sign Language Recognition System Based on CNN" [16] published by Abhishek Dudhaland his peers. Glove-based systems incorporated were expensive and difficult to use; in contrast, the image classification-based system proposed in this paper is much cheaper and easier to use. This paper proposed a method for Indian sign language recognition using CNN classification and feature extraction by hybrid SIFT. CNN is robust and stable such that it requires very little image preprocessing. The proposed approach achieved a validation accuracy of 92.78% for CNN with hybrid SIFT approach. 91.84% accuracy was achieved for CNN with adaptive thresholding approach. The system proposed in this paper can work just with a laptop Indian Sign Language Recognition was created by Purva A. Nanivadekar and Dr. Vaishali Kulkarni [6]: Database Creation, Segmentation and Hand Monitoring. The primary step of this scheme was to create an Indian Sign Language database. When they performed the hand movements, this was achieved by gathering the videos from the signers. The next step was Hand and Segmentation Monitoring. This was accomplished in order to derive features from a single gesture. Using a three-step algorithm, better quality hand tracking and segmentation was achieved. Motion tracking, edge detection and skin color detection worked on this algorithm.

H. Rule Based and Dynamic Time Warping based Machine Learning Algorithms:

These algorithms extract information from the classification model in the form of rules that are easy to understand and very expressive. This algorithm is ideally suited for the study of data containing a combination of numerical and qualitative characteristics. DYNAMIC TIME WARPING BASED ALGORITHM: Dynamic Time Warping (DTW) is a time series alignment algorithm originally designed to recognize speech. It aims to coordinate two sequences of feature vectors by iteratively warping the time axis until an optimum fit is found between the two sequences (according to acceptable metrics). The algorithm for Dynamic Hand Gesture Recognition and Novel Sentence Analysis was proposed by Archana S. Ghotkar and Gajanan K. Kharate [7]. The Microsoft Kinect Sensor was used in this strategy. The main purpose of this paper was to design and create a new algorithm for the development of Indian sign language sentences, taking into account the constant recognition of sign language. For word recognition, this paper examined two algorithms. Rule based and methods based on Dynamic Time Warping were developed. For continuous word recognition, this approach was more precise than the rule-based method. Inverted indexing for sentence comprehension was used in the new method. The problems of traditional continuous comprehension of sentences in the presentation of sign language have been resolved and webcam and hence can be used with mobility by the hearing-impaired community. The system could also be used for learning Indian sign language

III DESIGN

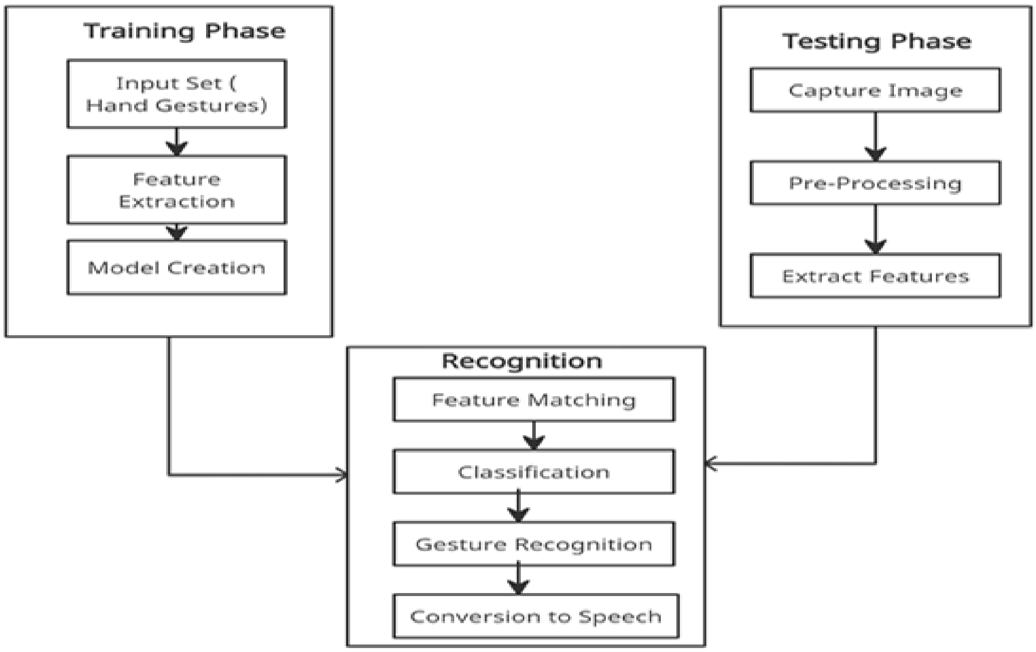

Fig 1. Steps in Design Process

Fig 1 shows the steps of the design process. Data Collection An algorithm for image recognition takes an image as an input, for example, "'A Letter' Hand Sign Gesture," B" Letter' Hand Sign Gesture, "C" Letter' Hand Sign Gesture," etc. The algorithm needs to be trained to understand the content of the image. We concentrate on two-class classifiers (binary). Under the hood, several popular object detectors have a binary classifier. Gestures are recorded with the help of a webcam which will be saved in a separate folder for each gesture created. OpenCV is a python package for creating real-time computer vision apps. It concentrates on image processing, filtering, video capture, and analysis, including features such as face and object detection. The process of dataset collection includes the following steps:

Capture: Using a webcam, 300 images are captured for each gesture. Images are captured within the defined rectangular boundary and are converted into a black and white image. Flipping: Images not captured in a specific format will be flipped and stored in a defined format.

Extracting: The image features are extracted within the defined boundary and the remaining part of the image will be cropped. Display: The captured frames will be shown on the screen before we manually save it into a specified folder.

Storing: Images for gestures are captured and stored in a folder for the respective gesture.

Pre Processing: The input image is pre-processed to reduce the size of the image. Cropping and resizing are done as a part of the pre-processing. This is necessary for the next step, feature extraction. Filtering is the first preprocessing phase. From the acquired image, unnecessary noise is removed. Background subtraction forms the next major stage. This processing results in a binary image in which white is colored with the pixels that form the hand and all the others are black. This processing includes the classification as part of human skin or not of each pixel of the image. CNN Algorithm: CNN algorithm is used for classification purposes.

CNNs are composed of four types of layers i.e., Input layer, Convolutional layer, Pooling layer, and Fully-connected layers. When these layers are piled, a CNN architecture will be created. The input layer will contain the image's pixel values. The convolutional layer determines the scalar product of the weights of each pixel and determines the output neurons. Introduction to Convolutional Neural Networks aims to add the rectified linear unit (usually abbreviated to ReLu) to the activation output provided by the previous layer with an 'elementary' activation function such as sigmoid. The pooling layer applies a function so that all the negative values are replaced with zero. The fully connected layer is the layer where actual classification takes place. Class scores are generated and the compressed image with class scores is converted into a list. How CNN Works? An input image may be deformed. These deformed images should also be classified by classifier because they are also the images to be predicted. In normal technique , when both the images are compared, the image classifier will not be able to predict the deformed image. A computer understands an image using numbers at each pixel. For example, in a binary image, black pixel is considered with a value 1 and while pixel will have-1 value. CNN compares these images piece by piece. By finding rough features in roughly the same positions in two images, CNN gets a lot better in seeing similarity than whole image matching schemes Training For training, a list of training images has to be built from the file system. The sub folders need to be analysed in the image directory and split into stable training, testing, and validation sets. The next step is to return a data structure describing the lists of images for each label and their paths. The training step then creates a graph from the saved file and returns a Graph object holding the trained Inception network, and various tensors that will be manipulated. Then the model tar file is downloaded and extracted. If the pretrained model to be used doesn't already exist, it is downloaded from the TensorFlow.org website and unpacked it into a directory. The given list of floats is written to a binary file. Running the image through simple distortions like cropping, scaling, and flips during training, can improve the results. These reflect the kinds of variances expected in the actual world, and can help the model cope better with natural data. A network of operations has to be built to apply the specified parameters to an image in this step. Cropping: Cropping is done by randomly placing a bounding box in the full image. The cropping parameter determines how big that box is in comparison to the input image. If it's 0, the box will be the same size as the input, with no cropping. The crop box will be half the width and height of the input if the value is 50%. Scaling: Scaling is similar to cropping, only difference is that the bounding box is always centred and the size of the bounding box fluctuates randomly within the given range. If the scale percentage is zero, for example, the bounding box will be the same size as the input and no scaling will be done. If you set it to 50%, the bounding box will be somewhere between half the width and height and full size. The tensorflow library downloads the inception model. This model has a set of rules for machine learning. These include read input files , index them and convert them to a format such that machine learning can be applied. The captured image may vary in width and height. These are converted o a fixed height and width of small dimensions so that it can be stored and trained more easily. Gaussian filters are applied to reduce the noise in the dataset. Then it is passed to tensorflow, which is the image processing library. Seventy percent of the files present in the dataset are used for training

IV IMPLEMENTATION

Overview of Technologies Used This project aims to transform the language of signs into a recognized language. Here, as a hand gesture, feedback is given. In the translate API tool that utilizes the algorithm CNN (convolutional neural network) and machine learning, the entire translation is completed. The input image in the form of a hand gesture is fed via the OpenCV method to the translator API tool. To create models, Tensorflow library is used. It is capable of classification and prediction. The GUI is created with the PyQt5 tool. All the signs are kept as an array of objects using numpy library, which is used to implement text to sign conversion. tesseract is a library that extracts text from images. To process input images, PIL is used. Music is played using the mixer library from pygame.

Implementation details of the Modules

Sign to text and text to sign: Tensor Flow is the main library used in conversion from sign to text. Similarly the conversion from text to sign happens.

Voice to text and text to voice: speech recognition library by google was used to recognize the voice. Similarly, GTTS(Google Text To Speech) was used to convert the given text into voice.

Emergency Text message: This feature enables the user to send a text message by showing a sign to the camera. This is achieved by passing the sender id, name and number through a dictionary. An API call is then made to send a message.

Opening Applications: This feature enables the user to open applications using particular signs to the camera. This is achieved by passing the path of the application as a parameter

Play music using gestures: This feature enables the user to start and stop music. The mixer library from pygame supports this feature.

Difficulties encountered and Strategies used to tackle. The key challenges while implementing the project has been the lack of common datasets, occluded features and variance in the language with locality. Our project uses a custom dataset with real time collection to overcome this. Some characters share similar gestures which leads to misinterpretation of the gesture. This was solved by collecting large datasets for different gestures. The accuracy of the predicted gesture was low at the start of the project but it was improved by increasing the resolution of the images so that more critical details from images were extracted. Training took a lot of time. To cut down on training time, transfer learning was implemented. Transfer Learning is the process of taking a model that has been trained on a big dataset and applying its knowledge to a smaller dataset

V RESULTS

The proposed system for Indian Sign Language (ISL) character classification was evaluated on a dataset containing hand gesture images representing alphabets and numerals. The model was trained using a Convolutional Neural Network (CNN), and its performance was assessed based on training loss and accuracy. Visualisation of the Training performance of the CNN Model

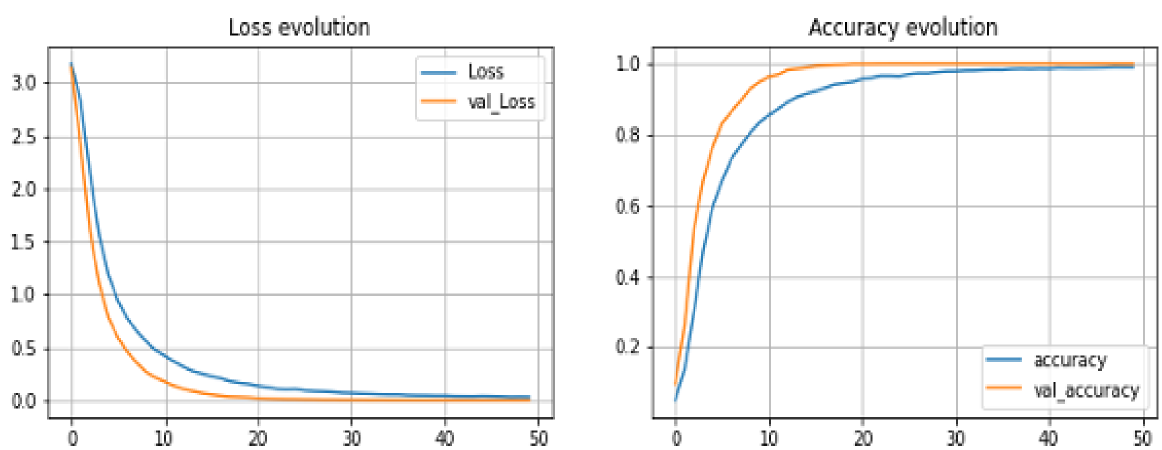

Fig 2. Training Loss and Accuracy

Fig 2. shows the training proceeds, the accuracy of the model improves and the loss gradually decrease. Out of 9,000 images collected, 70% of the images was used for training and 30% of the images were used for testing. Overall accuracy was calculated using the following formula:

$Accuracy = Number\ of\ correct\ predictions\ Total\ /number\ of\ predictions\ made$

The overall accuracy was found to be 98.9% for sign to text conversion and 99% for text to sign conversion. Overall, the results indicate that the proposed CNN-based ISL recognition system is highly effective, offering a reliable solution for translating sign language gestures into readable text. Future enhancements could focus on expanding the dataset, incorporating more complex gestures, and improving real-time performance for broader usability.

VI CONCLUSION

Machine translation is a very hot research subject in the field of natural language processing at present. Machine learning helps to train a human brain-like translation system. Open CV, CNN and translate API are capable of recognizing and translating sign language into text and speech. In text processing, the CNN Translator API offers better performance. Sign language recognition is a crucial communication aid for the speaking and hearing impaired. This instrument can help bridge the gap between individuals who are natural and deaf/dumb. This work aims to address these challenges by focusing on ISL character classification, compiling a dataset, and analyzing various features. The research seeks to develop a machine learning model capable of recognizing hand gestures used in sign language for finger spelling. The Convolution Neural Network is used to classify sign language characters, including alphabets and numerals, with exceptional precision. A user-independent model is trained using classification-based machine learning algorithms on a dataset of images, and performance is evaluated through testing. Convolutional Neural Network (CNN) algorithms are applied to improve accuracy in recognizing ISL gestures.

REFERENCES

[1] Jain, S., Raja, K.S. and Mukerjee, M.P.A., 2016. Indian sign language character recognition. *Indian Institute of Technology, Kanpur Course Project-CS365A*.

[2] Goyal, S., Sharma, I. and Sharma, S., 2013. Sign language recognition system for deaf and dumb people. *International Journal of Engineering Research Technology, 2*(4), pp.382-387.

[3] Kim, Y. and Toomajian, B., 2016. Hand gesture recognition using micro-Doppler signatures with convolutional neural network. *IEEE Access*, *4*, pp.7125-7130.

[4] Tripathi, K. and Nandi, N.B.G., 2015. Continuous Indian sign language gesture recognition and sentence formation. *Procedia Computer Science*, *54*, pp.523-531.

[5] Nanivadekar, P.A. and Kulkarni, V., 2014, April. Indian sign language recognition: database creation, hand tracking and segmentation. In *2014 International conference on circuits, systems, communication and information technology applications (CSCITA)* (pp. 358-363). IEEE.

[6] Ghotkar, A.S. and Kharate, G.K., 2015. Dynamic hand gesture recognition and novel sentence interpretation algorithm for indian sign language using microsoft kinect sensor. *Journal of pattern recognition research*, *1*, pp.24-38.

[7] Chen, Z.H., Kim, J.T., Liang, J., Zhang, J. and Yuan, Y.B., 2014. Real-time hand gesture recognition using finger segmentation. *The Scientific World Journal*, *2014*(1), p.267872.

[8] Badhe, P.C. and Kulkarni, V., 2015, November. Indian sign language translator using gesture recognition algorithm. In *2015 IEEE international conference on computer graphics, vision and information security (CGVIS)* (pp. 195-200). IEEE.

[9] Singha, J. and Das, K., 2013. Recognition of Indian sign language in live video. *arXiv preprint arXiv:1306.1301*.

[10] Geetha, M. and Manjusha, U.C., 2012. A vision based recognition of indian sign language alphabets and numerals using b-spline approximation. *International Journal on Computer Science and Engineering*, *4*(3), p.406.

[11] Oudah, M., Al-Naji, A. and Chahl, J., 2020. Hand gesture recognition based on computer vision: a review of techniques. *journal of Imaging*, *6*(8), p.73.

[12] Rokade, Y.I. and Jadav, P.M., 2017. Indian sign language recognition system. *International Journal of engineering and Technology*, *9*(3), pp.189-196.

[13] Nagendraswamy, H.S. and Kumara, B.C., 2017. LBPV for recognition of sign language at sentence level: An approach based on symbolic representation. *Journal of Intelligent Systems*, *26*(2), pp.371-385.

[14] Dudhal, A., Mathkar, H., Jain, A., Kadam, O. and Shirole, M., 2019. Hybrid SIFT feature extraction approach for Indian sign language recognition system based on CNN. In *Proceedings of the international conference on ISMAC in computational vision and bio-engineering 2018 (ISMAC-CVB)* (pp. 727-738). Springer International Publishing.