Vol. XVII & Issue No. 10 October - 2024
INDUSTRIAL ENGINEERING JOURNAL

MARATHI SIGN LANGUAGE RECOGNITION USING CONVOLUTION NEURAL NETWORK

Manan Shah, Dev Prajapati, Austin Isaac, Jainesh Chawan & Rohan Tapulli
Student, Computer Engineering, Fr. Conceicao Rodrigues College of Engineering, Mumbai, India
Email: mananshahamit@gmail.com
Email: devprajapati00@gmail.com
Email: rohan.tapulli@gmail.com
Email: austinjan13@gmail.com
Email: Jaineshrchawan@gmail.com

Ashwini Pansare
Assistant Professor, Computer Engineering, Fr. Conceicao Rodrigues College of Engineering
Mumbai, India
Email: ashwinip@fragnel.edu.in

Abstract

Hand Gesture to Text Converter is a growing topic of interest in the technology world. The major goal is to close the communication gap between mute people and normal people so that the speech-impaired person can converse with a non speech impaired person who cannot understand the signing simply and comfortably. Unpredictably identifying and separating the elements of sign images is a crucial problem in this subject. A text for the gesture would be generated and shown to the person accessing this model. The text is basically in the form of single letters of the Marathi alphabet. This paper's objective is to provide a practical solution that will help remove some barriers that could exist between a non-speech impaired person and a non speech impaired person. The goal of the model is to achieve language fluency and accuracy in identifying context and content of a gesture image solely from the target descriptions provided for images in the dataset and use that to caption other images fed to it. The training of the model is focused on achieving maximum similarity to a target recognition that is specified during the training phase. The development of a large vocabulary for signing recognition systems is currently receiving increased attention from researchers.

Keywords-Machine Learning, deep Learning, marathi sign recognition, sign language recognition, CNN.

1. INTRODUCTION

The deaf community uses sign language throughout India. However, deaf children are not taught in deaf schools using sign language. Teachers are not encouraged to use sign language as a teaching tool in teacher training programmes. There is no curriculum that includes sign language. The use of sign language can break down barriers to communication, but many parents of deaf children are unaware of this. Thus, the lack of knowledge and increase in demand has made the sign language awareness a necessity. The need for sign language interpreters is important in institutions and locations where hearing and deaf people interact. These are just one of the very few places where educating people about sign language has become essential. According to the 2011 Census, there are around 50 lakh deaf people living in India. Numerous organisations that support the deaf have noted the challenges caused by the long-standing disregard for the needs of the deaf community. It is necessary to update outdated teaching and training methods. The deaf community complained of a shortage of experienced interpreters for official situations such business meetings or training sessions. When they are needed, interpreters typically must be booked at least two weeks in ahead. However, last minute cancellations and change in their availability in an emergency situations leads to further challenges to speech impaired community. These issues make it extremely difficult to general and open conversation. Our program intends to make sign language interpreting simple and rapid for the general public to access. The field of study on hand gesture to text conversion is becoming more and more popular. The major goal is to close the communication gap between two of these people so that the deaf person can communicate with a non speech impaired person who cannot interpret the signing simply and comfortably. This study report illustrates the sign identification of the Marathi sign language via your hand which is four fingers and a thumb.

EXISITNG SYSTEMS AND RELATED WORK

[1]. Shinde et al. proposed to carry out sign detection in two parts, first being the use of camera and second is an offline system. The system proposes an offline way of learning the sign languages, their recognition to text and vice versa. The video based recognition requires the user to portray the sign with the help of his hand in front of the camera very accurately in order to have maximum accuracy. For the offline system the input image is

browsed from the predefined database and corresponding text is displayed, moreover for video system, it initially converts the photo to grayscale and then further matches it with the database and displays the output. Here, resizing of image, colour based extraction, noise reduction, calculating centre of gravity, translation into binary image and comparison with database is done during translation. The limitation was inefficiency of the model for some words and lack of accuracy due to change in hand size with every different human.

The implementation of the system proposed by [2]. Anupama et al. is mainly divided into the following steps 1. Image capture through webcam continuously 2. Passing the image to the computer and guessing the hand gesture. 3. Converts hand gestures to command to the computer. 4. Operate the command. Image capture and extraction involves finding images from the depth camera, rescaling the raw image, background elimination and extraction. This work was focused on recognizing the hand gesture but the major limitation is the use of expensive cameras and that the model works on American Sign Language on which much work has been done in recent history.

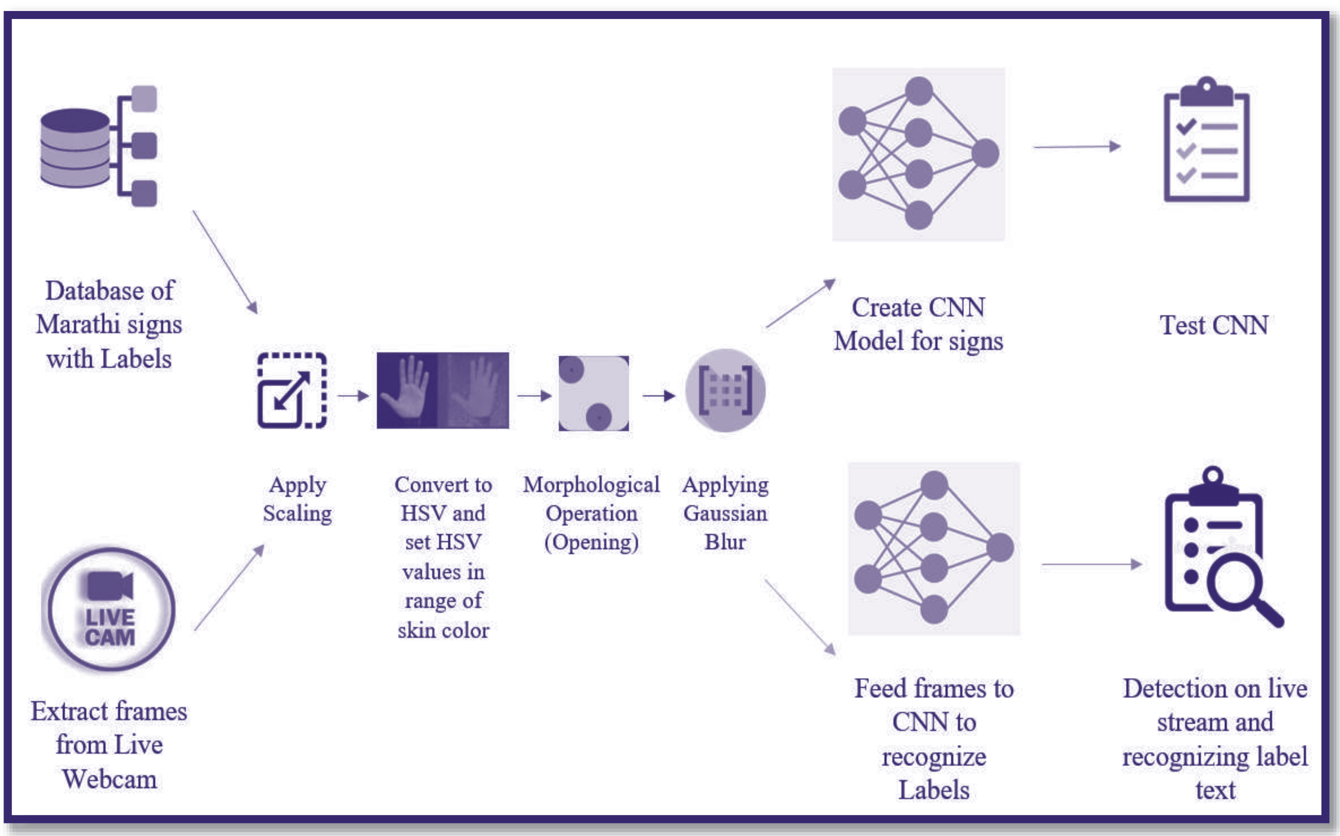
The system developed by [3]. Xingyan et al. involves recognizing gestures by making use of the Fuzzy C-Means Clustering Algorithm which can be used in a mobile application. The system gives a 85.83% accuracy for recognizing gestures. The limitations include incorrect extraction of objects if it is larger than the hand, overlapping objects due to poor lighting environment or changes and reduction in performance of the model if the hand is more than 1.5 metres away from the camera.

[4]. E. Stergiopoulou et al. proposed a system based on Self-Growing and Self-Organized Neural Gas (SGONG) which is a fitting procedure where the SGONG method is applied to get the shape of the hand region based on which the morphological details of the palm is retrieved. The disadvantage of this system is that the gestures are only taken from the right hand and must be vertically aligned, the palm of the hand should face the camera and the system doesn't work if the background is not uniform and plain.

PROPOSED SYSTEM

The proposed system comprises of generating the dataset using 42 Marathi signs and various operations are performed on the like augmentation and skin masking. The dataset is divided into 80% training data, 10% test data and 10% validation data. The training data was fed into a Linear as well as Sequential model with one input layer, 4 hidden layer and one output layer and a test accuracy of the Linear model was 87.5% while for the Sequential model was 86.5%. The detailed workflow is demonstrated below in Fig 1.

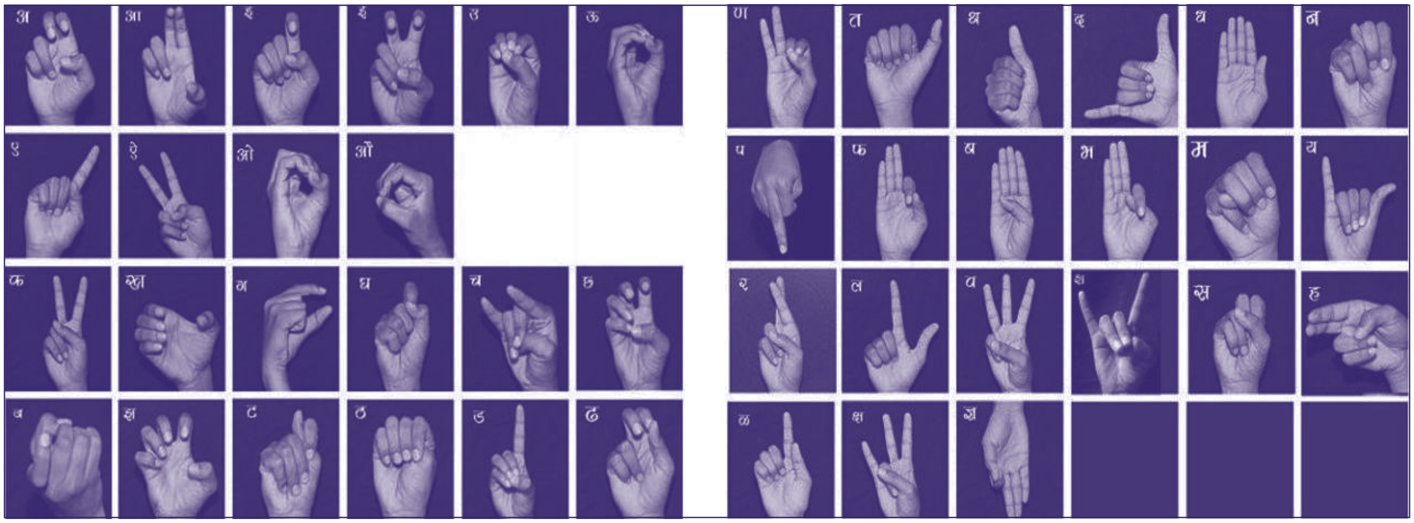
Fig 1. Proposed System



METHODLOGY

A. Data acquisition: The Marathi Sign Dataset was created by taking pictures of 42 Marathi signs as shown in Fig 2. Each sign in the Marathi sign Dataset represented the Marathi barakhadi and the images were taken with a high resolution camera in front of an even background. The dataset contained hand pictures of various individuals i.e., men, women and children to ensure the variability in the dataset to prevent overfitting. The individuals whose hand pictures were obtained were of various ages and of different colour to provide variance in the dataset. The total number of pictures obtained in the dataset are 42*20 which results in 840 images.

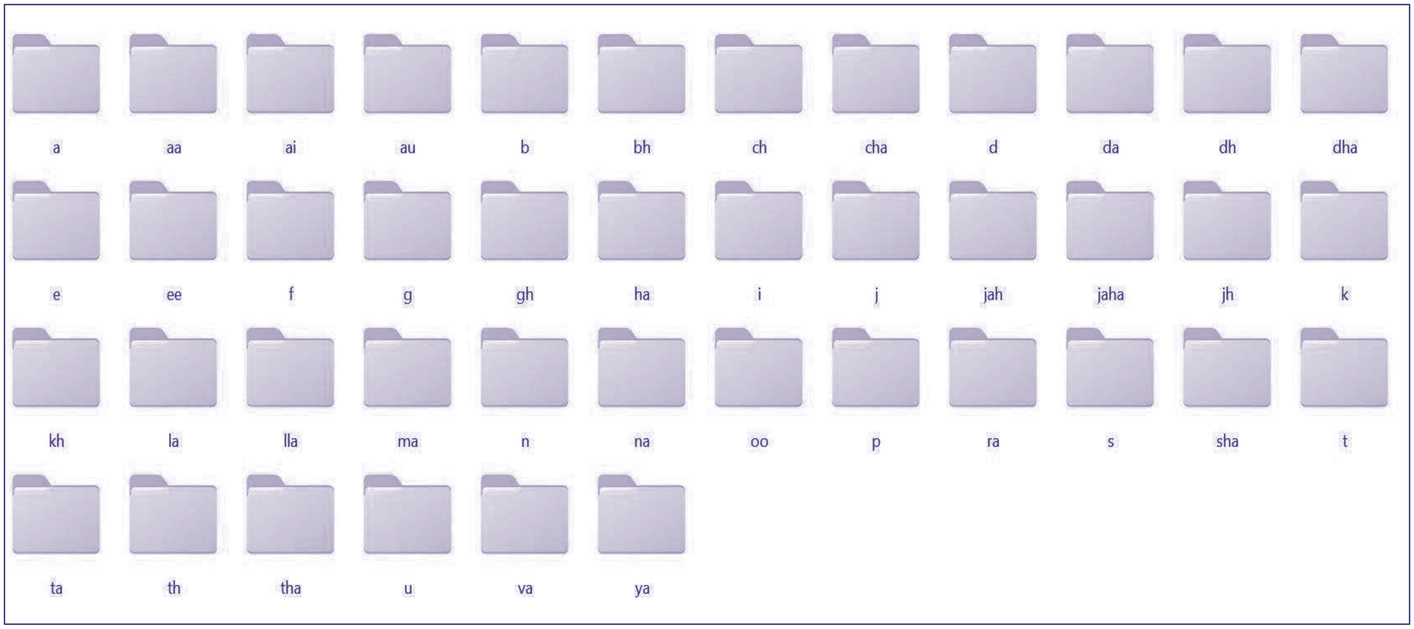
Fig 2. Marathi Barakhadi Signs



B. Data augmentation: Data augmentation can be done in two ways, by adding new observations or by generating new observations with the help of machine learning algorithms. In our proposed method we augmented dataset images using a machine learning algorithm. The Marathi Sign Dataset containing 840 images were augmented to 43,050 images. For each Marathi sign the number of images that were generated were 1025. The augmentation process included various steps. The original images were rotated clockwise, anticlockwise, flipped, left and right shift, zoomed in and out and the brightness of the images was varied.

Fig 3. Augmented Dataset

Fig 4. Final Augmented Dataset



C. Data pre-processing: In our proposed system data pre-processing is the process of converting an augmented dataset to skin masked images before applying any morphological operations. The first step in this method is to convert the augmented dataset into HSV images. This is done by using a color space transform. The range of HSV values that were used were in the range of (5,50,50) to (15,255,55) which lies in the skin colour range. In the next step morphological operations were applied which include erosion followed by dilation. Floating pixels and thin lines are eliminated through morphological erosion, leaving only substantial objects. The remaining lines and shapes look thinner and smaller. Morphological dilatation fills up small gaps in objects and increases object visibility. Shapes with filling appear larger and lines appear thicker. After the morphological operations on the dataset Gaussian blur was applied on it to reduce noise (high frequency components) and lastly the pre-processed image was masked with the original image to extract the exact sign in each image. The final pre-processed image will contain only the Marathi sign with background being eliminated in order to remove the unnecessary noise and data.

Fig 5. Data Pre-processing

Fig 6. Final Pre-processed Dataset

D. Suitable Model Selection: The data was trained on various deep learning models in order to choose the best effective model for recognition, as shown in Table 1 below. The linear model had the highest accuracy with the least amount of overfitting, according to the analysis.

TABLE 1

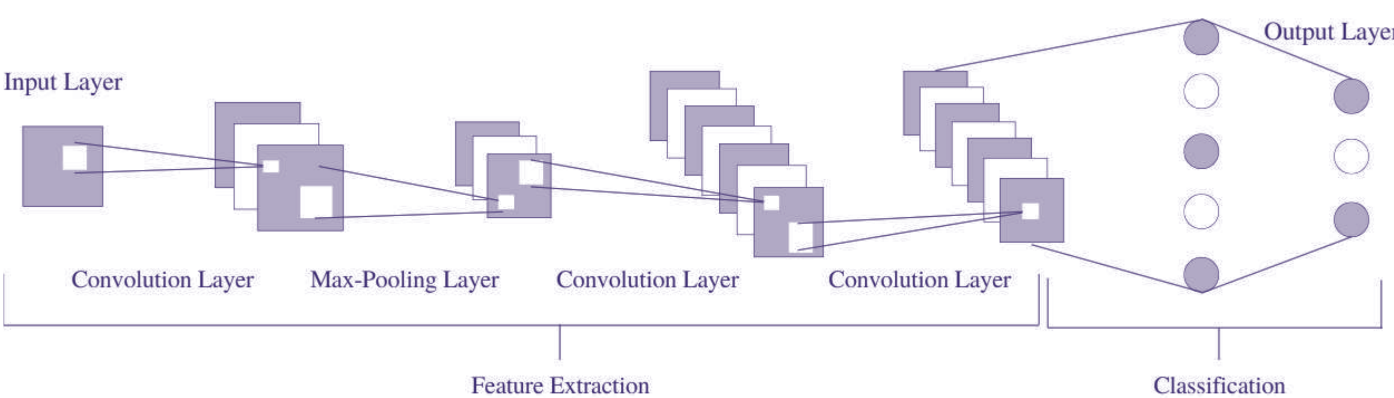
Sr no	Model Name	Training Accuracy
1	Linear	99.10
2	Sequential	96.50

E. Network Architecture and Implementation of CNN: The suggested application first collects frames from the webcam's live video before choosing the appropriate frames using a key frames-by-key frames extraction method. The main technology for video abstraction is key frame extraction, which can significantly reduce redundant information in the video. Our critical frame extraction processes generally consist of three steps:

1) computing image entropy
2) locating nearby extreme points
3) carrying out density cluster

After the key frames are extracted, we go on to the hand segmentation techniques. We used Skin mask mode processing to preprocess the images obtained and now we will use these as the input for our CNN model and we will get the gestures as output that we will be displaying along with the live video as text. Following the proper model selection methods, the architecture diagram for the linear model shown in Fig. 7 is selected. The linear model contains 4 hidden layers in the CNN model. The first hidden layer contains 128 nodes and we have used Relu as the activation function for this layer. In the first layer we have used Batch normalization to normalize the contributions to a layer for every mini batch and to drastically reduce the number of epochs required to train the deep neural network. The next layer is the Max-pooling layer which has 64 nodes and is used to select the maximum element from the region so the output will be of the most prominent features. The next 2 layers are the Convolution layers having nodes 64 each. The output layer is the final layer, and it employs SoftMax activation. The optimizer and loss function are specified using the compile method. Adam optimizer and categorical entropy are utilised, respectively. The model is fitted on the training set, tested on the test set with a batch size of 32, and validated on validation data for 15 epochs in a batch size of 64.

Fig 7. CNN Linear Model Architecture



RESULTS

The accuracy given by the Linear model for the training dataset was 97.40 % with a final loss score of 0.1345 for the final 100 epoch and the accuracy given by the sequential model was 95% with a final loss score of 0.1550. The test accuracy for the Linear model was 87.5% while for the Sequential model was 86.5%. The end system was able to detect and recognize Marathi sign using Linear CNN model.

Fig 7. Result 1

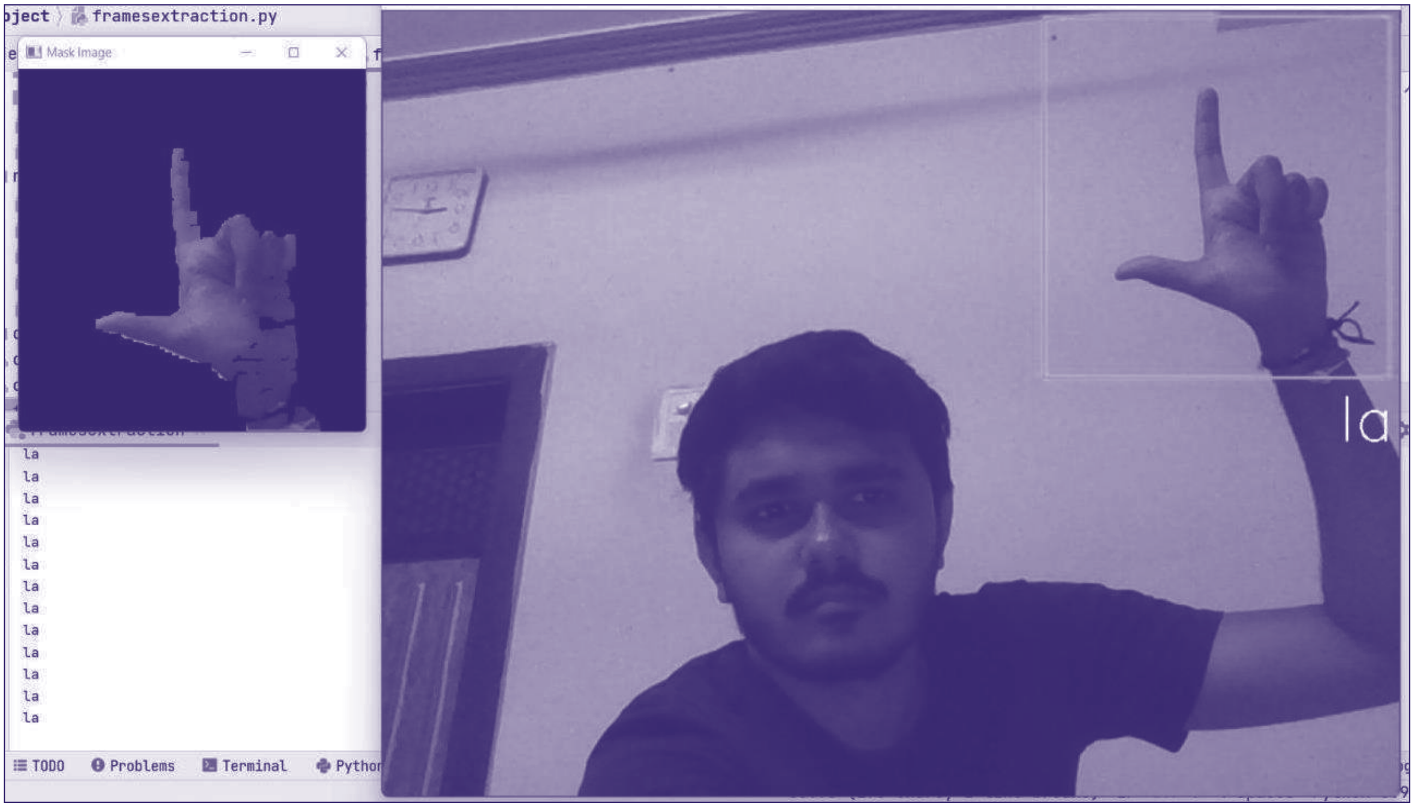
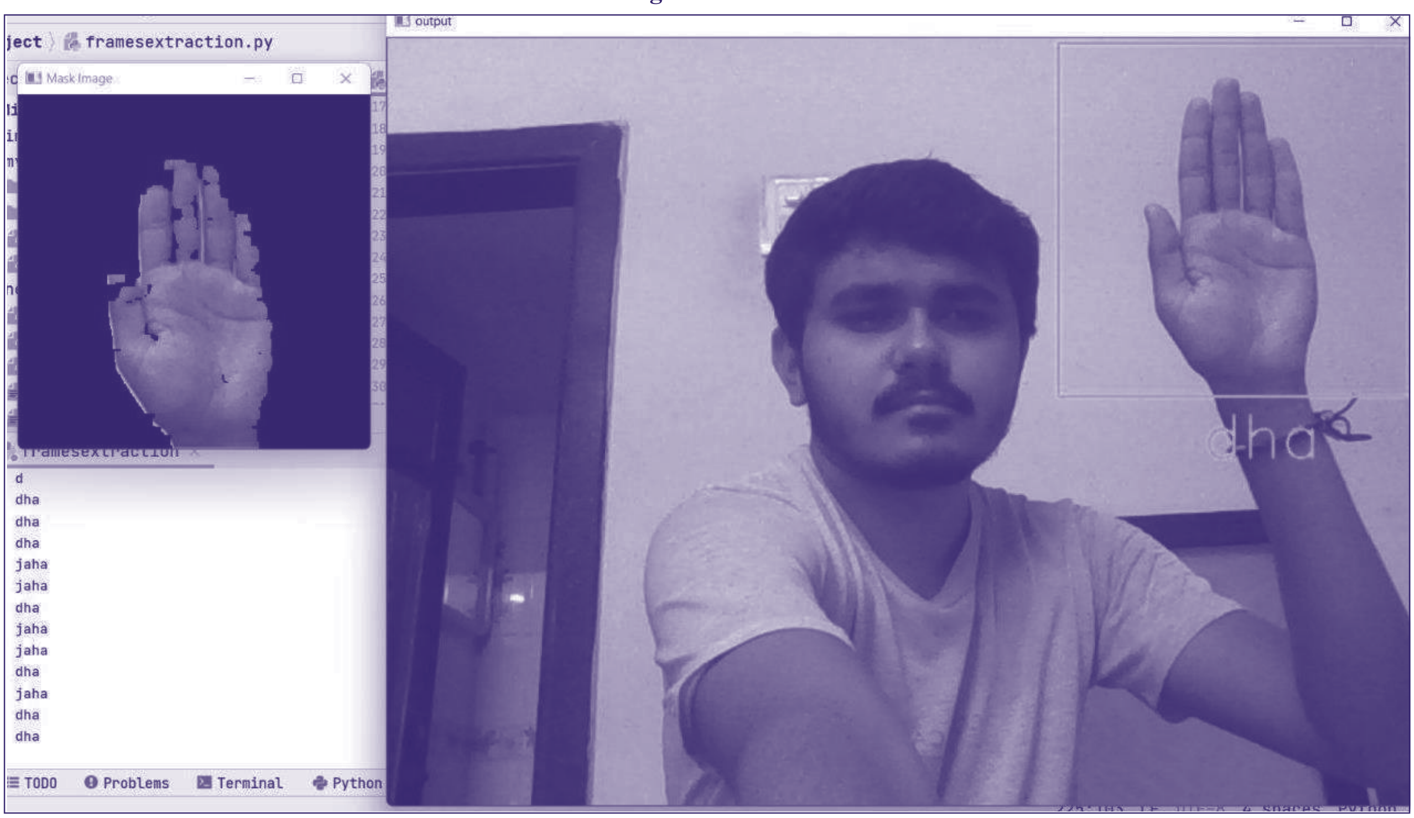


Fig 8. Result 2



CONCLUSION

The research carried out has proposed a deep neural network Linear and Sequential model with the idea of helping out the mute people and to make their lives a bit easier. The proposed Linear model uses Torch CNN model while the Sequential model uses TensorFlow Keras models. After experimenting with both we decided to apply the Linear model because it shows more accuracy in terms of train and test dataset. The dataset was created by taking high resolution images of Marathi signs with the help of 20 individuals and the dataset was pre-processed to clear noise and background so the gestures could be recognized much easily. Our model recognizes 42 signs of Marathi. The end system is a Graphical user interface as shown in Fig 9. using tkinter python library that uses the model at the back end to predict signs so the user will be able to see recognized sign on the interface.

Fig 9. Graphical User Interface

FUTURE WORK

Recognition of thorough sentences of Marathi sign languages not just letters and numbers can be challenging. Different hand gestures and words can be identified and used as computer input in the future. The hand signs used to represent numbers can also be translated into orders that will carry out relevant actions immediately. Future research can focus on improving the ability to recognise in different lightning circumstances, which is a hurdle in this project.

REFERENCES

[1] Shinde, Amitkumar & Kagalkar, Ramesh. (2015). Advanced Marathi Sign Language Recognition using Computer Vision. International Journal of Computer Applications. 118. 1-7. 10.5120/20802-3485.

[2] Hand Gesture Recognition for Mute People Anupama H S1 , Dr. Cauvery N K2 , Dr. Lingaraju G M3 Asst Professor, CSE, RV College of Engineering, Bangalore, India1 Professor, ISE, RV College of Engineering, Bangalore, India 2 Professor, ISE, MSRIT, Bangalore, India 3

[3] Xingyan Li. (2003). "Gesture Recognition Based on Fuzzy C Means Clustering Algorithm", Department of Computer Science. The University of Tennessee Knoxville.

[4] E. Stergiopoulou, N. Papamarkos. (2009). "Hand gesture recognition using a neural network shape fitting technique," Elsevier Engineering Applications of Artificial Intelligence, vol. 22(8), pp. 1141– 1158, doi: 10.1016/j.engappai.2009.03.008

[5] J. P. Singh, A. Gupta and Ankita, "Scientific Exploration of Hand Gesture Recognition to Text," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 363-367, doi:

10.1109/ICESC48915.2020.9155652.

[6] A. Correa, K. Shetty, R. Binny and A. Pansare, "An Image Captioner for the Visually Challenged," 2021 2nd Global Conference for Advancement in Technology (GCAT), 2021, pp. 1-6, doi: 10.1109/GCAT52182.2021.9587875.

[7] L. Menezes, E. Gnanaraj, S. Bindra and A. Pansare, "Early Stage Stroke Prediction Using Artificial Neural Network," 2021 2nd Global Conference for Advancement in Technology (GCAT), 2021, pp. 1-6, doi: 10. 1109/ GCAT52182.2021.9587590.

[8] Salian, Shashank & Dokare, Indu & Serai, Dhiren & Suresh, Aditya & Ganorkar, Pranav. (2017). "Proposed system for sign language recognition". 058-062. 10.1109/ICCPEIC. 2017.8290339.

[9] Tolentino, Lean Karlo & Serfa Juan, Ronnie & Thio-ac, August & Pamahoy, Maria & Forteza, Joni & Garcia, Xavier. (2019). Static Sign Language Recognition Using Deep Learning. International Journal of Machine Learning and Computing. 9. 821-827. 10.18178/ijmlc.2019.9.6.879.

[10] Shirbhate, Radha S., et al. "Sign language Recognition Using Machine Learning Algorithm." International Research Journal of Engineering and Technology (IRJET) 7.03 (2020): 2122-2125.