Improved CNN Based Sign Langiage Recognition For Specially Disabled People

Sonali Rangdale[1], Nagesh Raykar[2], Prashant Kumbharkar[3] , Santosh Borde[4]

[1]JSPMsRajashriShahuCollegeofEngineering,Pune-411033,India,
Pune-India,
Sonalirangdale1278@gmail.com

[2]ShriJagdishprasadJambharmalTimbrewalaUniversity,Jhunjhunun,India
Pune-India,
Nageshkraykar@gmail.com

[3]JSPMsRajashriShahuCollegeofEngineering,Pune-411033,India,
Pune-India,
Pbk.rscoe@gmail.com

[4]JSPMsRajashriShahuCollegeofEngineering,Pune-411033,India,
Pune-India,
spraoborde@gmail.com

***Abstract*:**
Since the majority of people do not understand sign language, there needs to be a bridge built so that the community can interact with the deaf. The use of image processing technology as a translator tool is one way that technology, which is constantly improving and working to better people, can be utilized to build a communication bridge between the community and deaf individuals. The goal of this study is to create a model for recognition of sign language with the help of Indian sign language. The model is implemented with modified CNN deep leaning methodology. The Convolution Neural Network (CNN) algorithm in the Deep Learning method can be a classification-tion tool, with the ability of the Convolution Neural Network (CNN) to learn several things. The model is implemented to capture images of palm signs using web camera, system conclude and display the name of capture images. With the use of convolution neural networks. The CNN is used in this model to convert the Indian Sign Language to text and voice. In this study, CNN method has been successfully carried out for sign language recognition with the results being able to increase the accuracy value to 99.4%. Sothat the results of the research increase the higher accuracy value.

***Keywords*: Sign Language, Hearing Disability, Convolutional Neural Network, OpenCV;**

(Article history: Received: Oct 31,2023 and accepted Jan, 31,2024)

I. INTRODUCTION

In our daily life, humans communicate with each other using words and gesture but for the deaf people or people who are hard of hearing, sign language is the way for communication .By considering this differently abled people and new innovative Technologies we have built the sign language recognition system. This will open different new opportunities for a sign language users to communicate friendly and efficiently and showcase their skills to global in easy way. This paper aims to cover most recent technologies to add advanced features in basic existed system. In this system we have used concept of image detection, feature extraction and text recognition and used CNN algorithm. This model will help the people to overcome the challenges they faced while communicating with hearing people and human computer interaction. This paper covers the existing sign language recognition system designed and methods and techniques. The data sets used in this model which is created by taking images through webcam and finally Conclusion and future directions in this area is discussed This model gives outcome in form of A to Z alphabets,0-9 numbers and also in form of voice.

II. LITERATURE SURVEY

Recognizing hand gestures is a crucial component in the last two decades of re-search. Scientists have done much study and experimented with a range of methods for recognizing gestures. The paper by nishi Intwala, arkav Banerje, meenakshi, and nikhil Gala propose a novel approach for Indian sign language converter using Convolution- al neural network. The main motive behind this translator system is smooth communication between impaired and normal person.To address this issue the author developed an Indian sign language converter model that can be accurately identify all al- phabet from real time images or hand gesture. They created data set of Indian sign language i.e 52,000 images of 26 alphabet in different background. For creating dataset 5 steps are: A. Dataset creation B. Image Cropping C. Image Resizing D. Image Flipping E. Training the Classifier F. Segmentation. They tested this model on a dataset of 52,000 hand gesture images and demonstrated that it achieves an accuracy of 96% in identifying the alphabet [1].The author suggest that their sign converter sys- tem has significant potential to detect hand gesture for smooth communication. The authors note that this approach has several advantages over current methods, including its low cost, portability, and ease of use, huge dataset, accuracy. Mobile Net is a very effective approach to classify the large amount of dataset with high accuracy. Overall, the paper highlights the potential of deep learning and mobile Net is improve to the classification and accuracy [2].Previous studies have explored the use of various machine learning, deep learning and python. Traditional methods include thresh- old-based segmentation, feature extraction, and classification using machine learning algorithms. For ex. A Singha et al. Use the dataset of 240 images for 24 sign. In their methodology they extracted the Eigen values from the images and clustering is done based on the Euclidean distance. Their model has static kind of gesture and they achieve 97% accuracy [3].The ok an et al. proposed a hierarchical CNN based design. This model consists two convolutional neural network. Detecting hand gesture by using light weight CNN and to predict the gesture used bulky 3CDCNN.The bhagat et al [4].Is used the image processing and deep learning techniques. For ISL alphabet and number used CNN to train 36 static gesture. To achieve the 1-1 mapping between depth and RGB pixel they used computer vision techniques. They got 98.81 accuracy at the time of training images. Computer vision used for static and LSTM is used for 10 ISL dynamic word gesture with accuracy 99.08% at the time of training [5].Kaustubh mani tripaathi, shruti patil created a translator for Indian sign language. Convolutional is one of the main methods to do images recognition and image classification. Firstly, the output in form of text and after character is recognized, the output will converted text into speech format. For this they used the pyttsx library in python [6].Sachin bhat, ameuthesh and sujith are created system Indian sign language to text and voice message. For tracing correct movement they used the flex sensors. According to the movement of the fingers there is an associated voltage drop because variation of resistance. It's in analogy in nature [7].

III. EXISITING APPROACH

In existing system authors worked on single and double handed gestures. Recognition system of Indian sign language is developed by using machine learning algorithms such as SVM, K Nearest Neighbors, Logistic Regression and Back propagation Algorithm. These algorithms was used for train their system [8.9, 10]. They have used different techniques for data preprocessing i.e., to remove image noise and to convert RGB image to grey scale image. They have used tensor flow and keras libraries to format datasets. Different algorithms and approaches used in existing system have different accuracy. Highest accuracy we have found during survey is in KNN and CNN algorithm [11, 12].

A. Praposed Technique:

The Hand segmentation is an important step in the model. First capture hand image using webcam. To get an accurate model large data set is needed. CNN is used for training and testing. Sign gestures captured in camera is processed using image pro- cessing [13]. Then accurate classifiers are applied to differentiate different signs. This signs are get matched with dataset and translated output is displayed on the screen in the format of text and voice. Different python libraries of different versions are used to do the process and to convert text to speech pyttsc library of python is used. In this paper, a CNN based recognition system based on Indian Sign Language has been proposed [14].

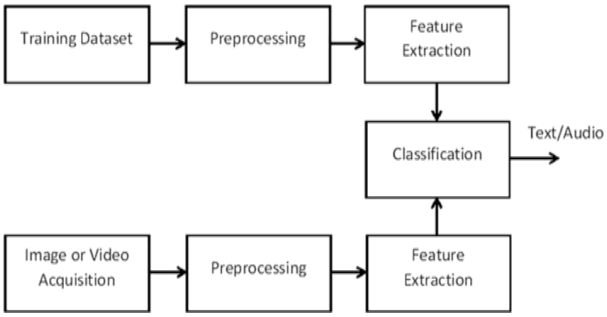

Fig.1:Architecture for sign language recognition System

IV. METHODOLOGY

Today's evolved technology has strength to help the especially abled people. A system which is built in this study that can recognize Indian sign language using CNN. The sign language user can use web camera and the identified alphabets and numbers can be displayed on the screen and voice of letters and numbers is produces. This system uses tensor flow and keras library. We have created the dataset using webcam and fetched images using OpenCV library [15]. The proposed model uses Indian Sign Language dataset to recognize the signs made by hand gestures. The Architecture of system is shown in below diagram. The methodology of the proposed system consists of following steps. Each step after implementation provides satisfactory results [16].

Processing:

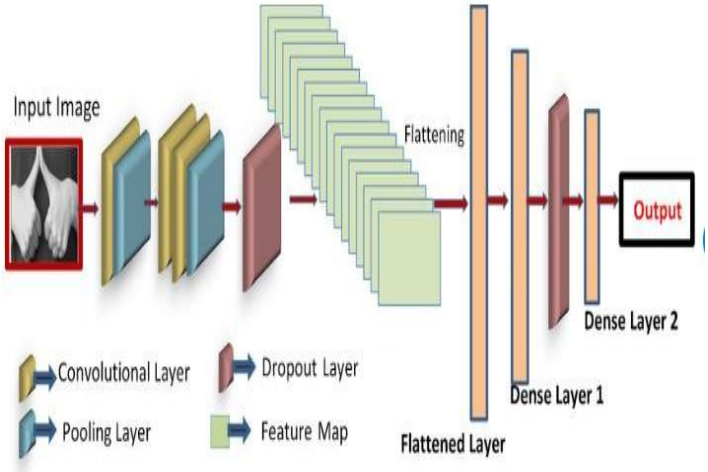

Fig.2:Proposed Model for recognition of sign language using CNN

The captured image is in RGB color format. In this step image processing takes place that is transformation of RGB image to grayscale i.e, 3D to 1D. The 1D format is the grayscale image. We have used average method for this transformation.

$$GREY=\frac{R+G+B}{3} \quad \text{----------(1)}$$

Using this strategy, you need 33% blue, 33% red, and 33% green. This system meets the requirements and functions well. Image noise is eliminated during image processing, preparing the image for algorithm application. The format of the dataset is designed to closely resemble the traditional ISL DATASET. Because of gesture motions, there are no cases for 9=J or 25=Z. Each training and test case encodes a label (0–25) as a one-to-one map for each alphabetic letter A–Z [17,18,19]. The normal MNIST handwritten digit dataset is approximately half the size of the training data (27,455 cases) and test data (7172 cases), but they are otherwise similar. They have a header row labelled pix-el1,pixel2...,.pixel784, which represents a single 28x28 pixel image with grayscale values between 0-255.

Training:

The TensorFlow library with CNN is used for training and classification purpose. The model is trained using supervised learning techniques, where the model is trained on Labeled data. Different metrics are used to test its accuracy and performance such as accuracy, precision, recall and F1 score. By training a CNN on a large dataset gives more accuracy and high performance. In this model it is more than 90%.

Classification:

The next step is extract meaning features from the preprocessed image. This has been done by using CNN algorithm. CNN adds filters to detect certain features of the image. The working of CNN depends on the types of filters applied. Particularly CNN is used for image classification. CNN provides better solutions. CNN is little bit same as neural network. Main building blocks of CNN are convolutional layer, pooling layer and fully connected layer. Where Convolutional layer extracts feature from input images by convolving filters. Filter is just small kernel that learned. Pooling layer reduces the input dimensionality. Pooling layer operates each feature map independently. Fully connected layer are similar to multilayer which will have perceptron. They have whole connection with all neurons in the subsequent layer. CNN algorithm is applied to dataset to classify different signs by passing the input images across different layers and achieved accuracy more than 90%.

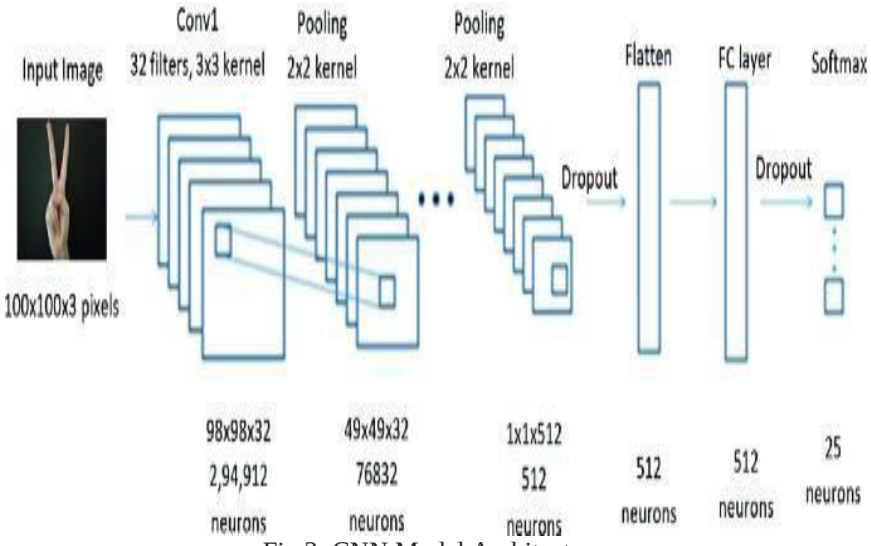

Fig.3: CNN Model Architecture

Prediction:

Finally, the signs are predicted which are made by the user in real time. And the out- put is generated in the format of text and voice. As the dataset used in this system is of high quality i.e., 500 images for each letter and CNN algorithm is used, the system has good accuracy and high performance. The data flow of Sign language recognition system involves different steps i.e., image capturing, dataset creation, image preprocessing, segmentation, feature extraction, classification and prediction of signs which is being made by user in real time. Input image passes through this flow of processes and predict the final output in form of voice.

Testing Phase:

After training phase it is observed we can see that our models performed very well on the validation sets. We can now use the model to predict labels of the whole testing dataset. Then we will Compute the accuracy score between the predicted labels and the true labels to see if it performs well. Scikit Learn has a built-in function for that.

Dataset Description:

Large dataset collection is the first and most important step in this model. It is created using OpenCV. Through web camera images of hand gestures are captured and stored in respective folder. Images of A to Z alphabets and 0-9 numbers are captured and stored. 500 images of each letter is stored. After that, labelling is done to each folder of 500 images each. Dataset is divided into 8:2 ratio i.e., 80% of data is used for training and 20% of data is used for testing. Test.csv and train.csv files are created after successfully done with dataset division.

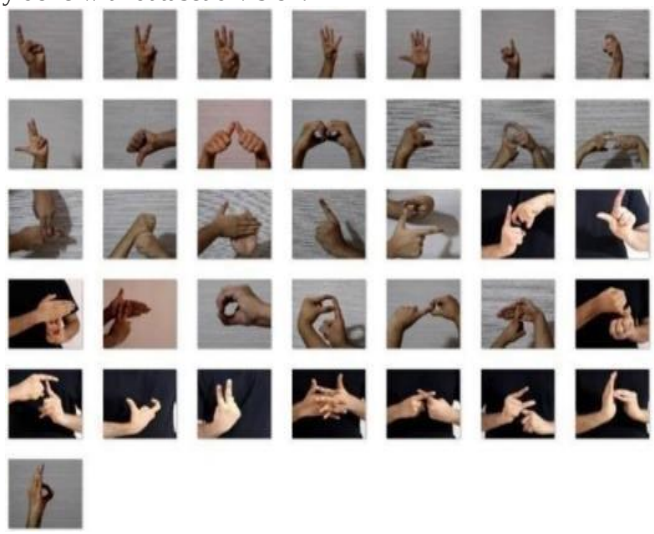

Fig 4: Sign Languae Dataset

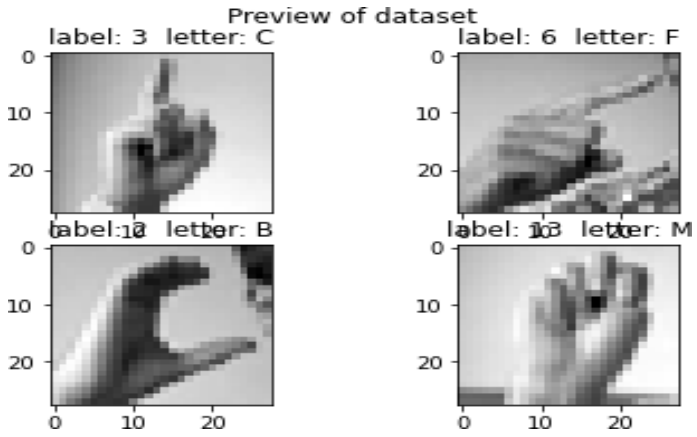

Fig 5: Indian Sign Language Dataset

V. EXPERIMENTAL AND RESULT SECTION

Indian sign language words can be identified with greater than 90% accuracy thanks to the sign language recognition technology. 26 alphabets and 10 ISL numbers can be efficiently and accurately recognized by it in palm gesture labeling. Both text and audio output are predicted. Training accuracy is 99.27, and test accuracy is 99.81 once batch normalization is included. Even with batch normalization, this only takes 40 epochs, or nearly half the time. The graph shows that there is no discernible drop in loss after 15 epochs, if we look closely. In order to terminate training after 15 or 20 epochs, we can employ early stopping. Without batch normalization, that amounts to around one-fifth of the time.

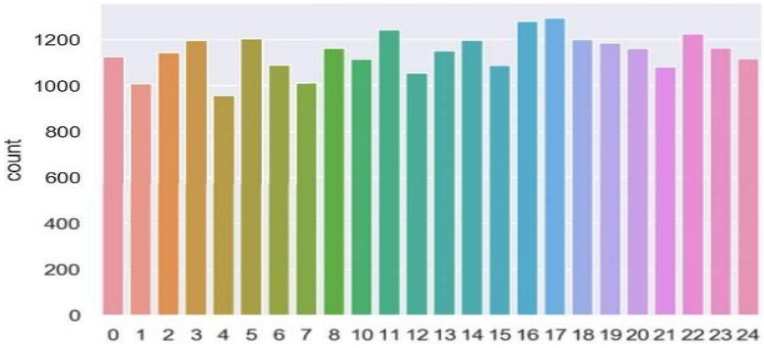

Fig6:Dataset Frequency Distribution

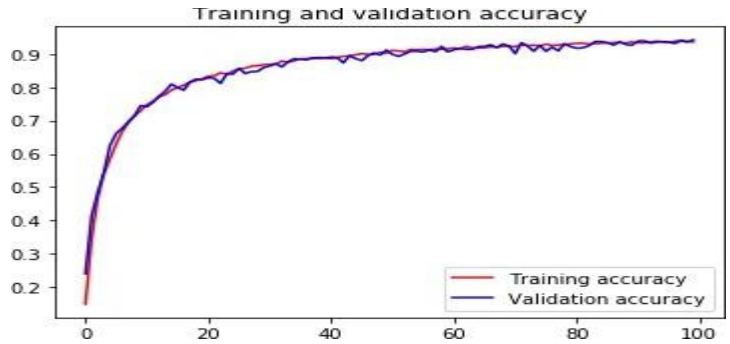

Fig.7: Training and validation accuracy plot

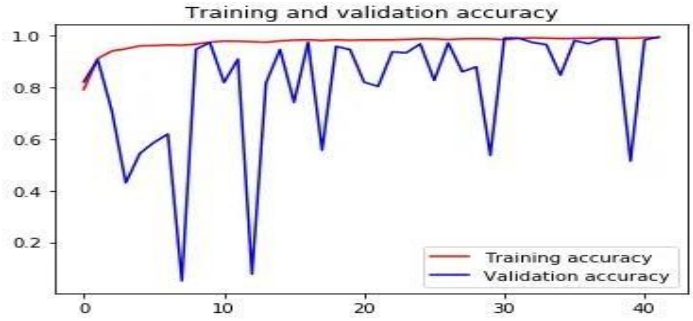

Fig.8: Training and validation accuracy plot 1

Table1: Result of CNN Model ISL

Model	No. of Conv layers	Augmentation	Batch Normalisation	Dropout	Training Acc	Test Acc
model1	1	yes	yes	no	97.52	97.71
model2	2	no	no	no	100	91.06
model3	2	no	yes	no	100	95.60
model4	2	yes	yes	no	94.58	98.41
model5	2	yes	yes	no	99.32	99.71
model6	2	yes	yes	no	98.70	99.51
model7	2	yes	yes	0.4	94.09	98.42
model8	2	yes	yes	(0.4 & 0.4)	84.27	91.98
model9	3	yes	yes	no	99.23	99.33
model10	3	yes	yes	no	91.90	98.1
model10	3	yes	yes	0.4	91.90	98.1
model11	3	yes	yes	0.2	92.32	97.5
model12	4	no	yes	no	100	96.10
model13	4	yes	yes	no	99.28	99.83
model14						
with decaying learning rate	4	yes	yes	no	99.71	100

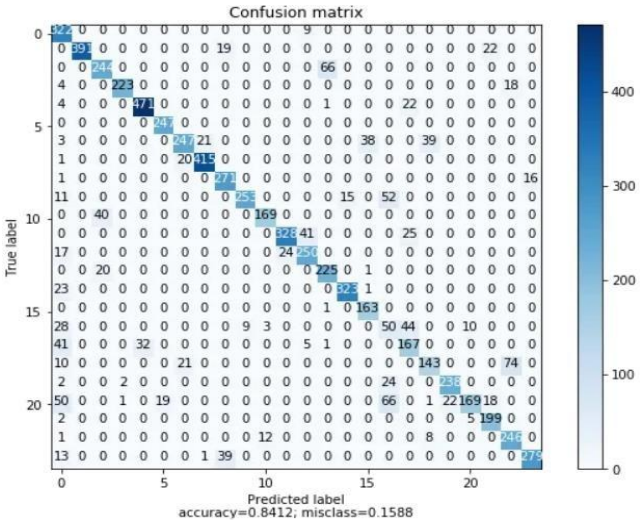

Fig.9: Confusion matrix on Testing Dataset

Above diagram shows the result window of the system which predicts the alphabet C. It also pronounce the letter C. In this way, the system predicts the ISL signs of all 26 alphabets and 10 digits in the format of text and voice in Fig.9 efficient way and with higher accuracy. The Indian sign language dataset contains signs for A to Z alphabets and numbers from 0 to 9. Dataset used for training and testing phase to build the model. We have used OpenCV library to develop dataset and CNN algorithm is applied on trained dataset.

Comparative Analysis: The existing system used SVM and logistic regression algorithms for this translation system. The model we have created using the CNN algorithm has more accuracy as compared to the existing system. This system has a feature to create a dataset while in existed system ISL dataset is uploaded from existing sources. It can generate out-put in the text as well as voice format. The system is more efficient to reduce the gap between specially abled and normal people. This System gives more than 90%accuracy for ISL dataset.

VI. Conclusion and Future work

This model helps many dumb and deaf people. The proposed model takes image of a palm gesture as input and recognize sign language in the format of both text and voice. It is helpful for normal people to understand what especially disabled people want to say. Our project aims to make communication simpler between special case people and normal people which is achieved with higher accuracy. This demonstrates, for instance, how our model confuses the U with the R, which is understandable given how similar these two signs are. More variables for these classes can be added to our model to help it identify higher level characteristics that will distinguish the two. With the help of this classifier, users may type letters in sign language on a computer. Because typing is faster, it is not particularly useful. The training data, however, largely limits the technology underlying it. Classifier implemented in this article could translate sign language to text if the dataset weren't the alphabet but rather any word that can be expressed in sign language. Learning from videos is useful because the majority of the words need motion, and there are a lot more classes to guess. The computational power required to train our dataset will also be considerably more consequential due to its increased size. The solution to this issue will be to use CNN architecture for multi-resolution processing.

REFERENCES

[1] Soma Shrenika myneni madhu bala “SIGN LANGUAGE RECOGNITION USING TEMPLATE MACHINE LEARNING TECHNIQUE ” Institute of Electrical and Electronics Engineers (2020).

[2] Saurabh Kumbhar, Abhishek Landge, Akash Kulkarni, Devesh Solanki, Vidya Kurtadikar “Indian Sign Language Recognition System” Volume 6, Issue 6,(2021).

[3] Varshitha P1, Penujuru Pranitha2, RNithin3, Mylarapu Vineetha4, Dr. Jalaja G5 “Indian sign language recognition system” International Research Journal of Engineering and Technology. (IRJET)Volume: 09 Issue: 06 (2022).

[4] Muthu Mariappan H,Dr Gomathi V “Real-Time Recognition of Indian Sign Language ”Second International Conference on Computational Intelligence in Data Science (2019)

[5] Diksha Hatibaruah , Anjan Kumar Talukdar† and Kandarpa Kumar Sarma‡ “A Static Hand Gesture Based Sign Language Recognition System using Convolutional Neural Networks” IEEE 17th India Council International Conference (INDICON) (2020).

[6] Raykar, N., Kumbharkar, P., Jayantilal, D.H. (2023). Assembled LSTM Technique Used for Phonetic-Based Algorithm for Demographical Data. In: Singh, S.N., Mahanta, S., Singh, Y.J. (eds) Proceedings of the NIELIT's International Conference on Communica- tion, Electronics and Digital Technology. NICE-DT 2023. Lecture Notes in Networks and Systems, vol 676. Springer, Singapore. https://doi.org/10.1007/978-981-99-1699-3_36 (2023)

[7] Nishi Intwala, Arkav Banerjee, Meenakshi and Nikhil Gala “Indian Sign Language con- verter using Convolutional Neural Networks”2019 5th International Conference for Con- vergence in Technology (I2CT) Pune, India. Mar 29-31,(2019)

[8] Nagesh Raykar, Prashant Kumbharkar, Dand Hiren Jayatilal. Hybrid LSTM technique for phonetic. Int J Adv Electr Eng 2023;4(1):18-22. DOI: 10.22271/27084574.2023.v4.i1a.33 (2023)

[9] Pratik Likhar, Neel Kamal Bhagat, Dr. Rathna G N “Deep Learning Methods for Indian Sign Language Recognition” IEEE 10th International Conference on Consumer Electron- ics (2020).

[10] Raykar,N.,Khedkar,G.,Kaur,M. et al. A novel traffic load balancing approach for scheduling of optical transparent antennas (OTAs) on mobile terminals. Opt Quant Electron 55, 962 (2023). https://doi.org/10.1007/s11082-023-05201-0

[11] Chakravartula Raghavachari and Shanmugha Sundaram G A “Deep Learning Framework for Fingerspelling System using CNN”International Conference on Communication and Signal Processing.

[12] Nagesh Raykar, Dr. Prashant Kumbharkar, Dr. Dand Hiren Jayatilal. De-duplication avoidance in regional names using an approach based on pronunciation. Int J Adv Electr Eng 2023;4(1):10-17. DOI: 10.22271/27084574.2023.v4.i1a.32 (2023)

[13] Y. Liu, X. Yao, Ensemble learning via negative correlation, Neural Networks, Volume 12, Issue 10, Pages 1399-1404, ISSN 0893-6080, https://doi.org/10.1016/S0893-6080(99)00073-8 (1999).

[14] MacKay D.J.C. (1995) Developments in Probabilistic Modelling with Neural Networks — Ensemble Learning. In: Kappen B., Gielen S. (eds) Neural Networks: Artificial Intelli- gence an [11] dIndustrial Applications. Springer, London. https://doi.org/10.1007/978-1- 4471-3087-1_37 (1995)

[15] Polikar R. Ensemble Learning. In: Zhang C., Ma Y. (eds) Ensemble Machine Learning (2012)

[16] Kang, Byeongkeun, Subarna Tripathi, and Truong Q. Nguyen. ”Real- time sign language fingerspelling recognition using convolutional neural networks from depth map.” arXiv preprint arXiv: 1509.03001 (2015).

[17] Suganya, R., and T. Meeradevi. ”Design of a communication aid for physically chal- lenged.” In Electronics and Communication Systems (ICECS), 2015 2nd International Conference on, pp. 818-822. IEEE, (2015).

[18] Sruthi Upendran, Thamizharasi. A,” American Sign Language Interpreter System for Deaf and Dumb Individuals”, International Conference on Control, Instrumentation, Communi- cation and Computa (2014)

[19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Re- siduals and Linear Bottlenecks," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.(2018)