Gesture based Sign Language Recognition System

K Anitha
Department of Electronics and Communication Engineering
Kumaraguru College of Technology
Coimbatore, India
anithajaya9811@gmail.com

R Naveen Karthick
Department of Electronics and Communication Engineering
Kumaraguru College of Technology
Coimbatore, India
naveen2002karthick@gmail.com

K C Varsheni
Department of Electronics and Communication Engineering
Kumaraguru College of Technology
Coimbatore, India
varsheni.c@gmail.com

A Kalaiselvi
Department of Electronics and Communication Engineering
Kumaraguru College of Technology
Coimbatore, India
kalaiselvi.a.ece@kct.ac.in

Abstract - **The use of hand gestures in sign language has been a crucial method of non-verbal communication, particularly for individuals with hearing or speech impairments. Despite numerous sign language systems developed by various developers worldwide, most of them are not flexible or cost-effective for end-users. This project proposes a hand gesture recognition system that can automatically detect sign language, allowing for more effective communication between deaf and dumb people and normal individuals. Gesture and pattern recognition are rapidly evolving fields that play a crucial role in nonverbal communication. Hand gestures have become an integral part of human's daily lives, and the Hand Gesture recognition system offers a natural and approachable way of communicating with computers, which people are more familiar with. The proposed system combines several cutting-edge technologies such as MobileNet, Xception, ResNet 101, and DenseNet 121. Among the four algorithms, DenseNet 121 achieved the highest overall accuracy, with an accuracy of 98.9%. Real-time sign language detection with text-to-speech technology and the development of a Telegram bot for sign language detection are additional features of this project. The integration of these technologies has resulted in a novel and effective method of sign language communication that can enhance the lives of individuals with hearing or speech impairments.**

Keywords - Hand gesture recognition, MobileNet, Xception, ResNet 101, DenseNet 121, Text-to-speech, Telegram bot.

I. INTRODUCTION

Sign language is a crucial mode of communication for individuals who have hearing or speech impairments. Hand gestures play a vital role in this method of communication. The development of an effective system for recognizing hand gestures is essential for enhancing the lives of individuals with such disabilities[1]. While several systems for sign language have been developed worldwide, they are neither adaptable nor economical. Therefore, this paper proposes a system that can automatically detect sign language, providing a natural and user-friendly interface for communication.

This paper's objective is to create a robust and efficient hand gesture recognition system that can detect signs in real-time. The proposed system utilizes state-of-art algorithms such as MobileNet, Xception, ResNet 101, and DenseNet 121 transfer learning models to achieve high accuracy in hand gesture recognition. Additionally, the system will utilize text-to-speech technology to enable the system to communicate back to the user, and a Telegram bot has been developed for remote sign language detection. The goal is to provide a flexible alternative to traditional sign language systems and help people's with hearing or speech impairments, communicate more effectively with each other and with normal individuals.

The rest of this paper is structured as follows. A survey of the relevant literature for hand gesture recognition systems is described in Section II. Section III describes information about the dataset used. The methodology and implementation details of the proposed system, including the algorithms used, such as MobileNet, Xception, ResNet 101, and DenseNet 121, as well as text-to-speech technology and the telegram bot have been described in Section IV. Section V presents the results of the system, including its accuracy, and performance. Finally, Section VI provides conclusions and suggestions for future research directions in the field of system for recognizing hand gestures.

II. LITERATURE REVIEW

Several methods have been proposed in the literature for sign language detection. The most familiar one is the system that utilizes a combination of hand tracking, feature extraction, and machine learning algorithms to recognize Turkish sign language (TSL). This system has an accuracy of 97.1%[2]. Utilizing this concepts, further enhancement in detection techniques was proposed, which uses wearable sensors, including gloves or wristbands, to record hand and body movements and to detect sign language [3].

Further to make the detection more accurate, a method using Support Vector Machine(SVM) for Indonesian (ISL) gestures for sign language detection was proposed in. This paper implements visual detection in four steps: Basic Encoder-Decoder, Visual Attention, Semantic Attention, and Deep Reinforcement Learning. Further to improve the classification a method was proposed in the literature, based on models such as the Vision detecting CNN, language generation RNN, HRNE, LSTM with attention, and Deep Reinforcement Learning. The paper provides a detailed explanation and explores the effectiveness of the hLSTM framework for enhancing model accuracy[4].

Apart from detection, classification techniques was also incorporated using CNN model to classify sign language gestures captured by a camera. The system was trained and tested using the American Sign Language (ASL) alphabet dataset, and the dataset is preprocessed by converting the images to grayscale and resizing them to a resolution of 64x64 pixels. The system achieved an accuracy of 97.5% [5]. Later, an E-CNN (Ensemble Convolutional Neural Network) based model utilizes a combination of hand tracking, feature

2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) | 979-8-3503-0681-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAECA56562.2023.10199750

extraction, and E-CNN model to recognize Indonesian Sign Language (ISL) gestures, achieving an accuracy of 94.24% [6]. To further improve the efficiency an approach using a Microsoft Kinect sensor utilizes depth information and a dynamic time-warping algorithm was proposed to recognize Chinese Sign Language (CSL) gestures, achieving an accuracy of 91.11% [7].

Further Large datasets of sign language are utilized in data-driven approaches to train machine learning models to recognize signs and gestures. Rule-based approaches identify signals and gestures by using established rules and criteria. [8],[9].

The use of accelerometry and surface electromyography (sEMG) signals for automatic sign language recognition (ASLR) have been used in Colombian sign language. Combining these signals improved recognition accuracy, highlighting the potential of multimodal approaches in ASLR systems[10]. By combining MediaPipe's multimedia processing capabilities with machine learning models, the researchers demonstrate the feasibility of achieving accurate and real-time sign language gesture recognition[11].

Computer vision-based systems use cameras to record and analyze sign language users' hand and body motions, and machine learning algorithms to identify certain gestures and signs [12]. Various machine learning models such as decision trees, SVMs, ANNs, and deep learning models such as CNN and RNN have been reviewed and the accuracies were ranging from 70% to 99%, depending on the dataset and gesture complexity [13]. An approach using LSTM, achieving an accuracy of 92.25% was achieved. The dataset consisted of the 26 ASL alphabet signs, captured from a single camera. Training, validation, and test sets were made from the dataset, with 70%, 15%, and 15% of the images in each set, respectively. [14]-[17].

Recently, a hand gesture recognition system was proposed using VR space and triboelectric communication. This method used the smart glove to detect sign language and has good efficiency [18]. Recently a CNN based Indian Sign language detection was proposed to detect the Indian sign language. In this work the author has used a CNN model with varying randomness for the filters in CNN. These filter characteristics improve the feature extraction of the images [19]. Apart from this, a unique method for recognizing the Indian sign language was proposed. In this work the input is a continuous video sequence of the signs. [20]. The detection was carried out a using a neural network. Based on the above studies this paper aims in developing a gesture based sign language recognition using deep learning methods.

III. DATASET

The ASL Alphabet dataset from Kaggle is a comprehensive collection of American Sign Language alphabet graphics with 87,000 photos divided into 29 classes. The dataset includes three essential classes for real-time applications and categorization: space, delete, and nothing. The dataset has been split into 80:20 ratios for training and validation. With its diverse range of classes and a large number of photos, this dataset could help in the creation of more precise and trustworthy recognition of sign language systems.

IV. METHODOLOGY AND ALGORITHMS

The proposed methodology employsseveral image pre-processing techniques to enhance the quality of the ASL Alphabet dataset. This included resizing all images to 224x224 pixels, normalizing pixel values, and applying data augmentation techniques such as rotation, shear, zoom, and horizontal flip. Data augmentation techniques specific to sign language recognition, such as adding random noise and blurring have also been applied. In addition, transfer learning have been used to fine-tune pre-trained models (MobileNet, Xception, ResNet101, and DenseNet121) for ASL recognition. Finally, a real-time implementation of the model using text-to-speech technology and a Telegram bot have been developed that predicts the sign in an image uploaded by the user. These processes collectively contributed to the development of a more accurate and reliable sign language recognition system. The workflow model shown in Fig.1 depicts a step-by-step process that involves data collection, image pre-processing, model training, evaluation, and deployment to develop a more accurate and reliable sign language recognition system.

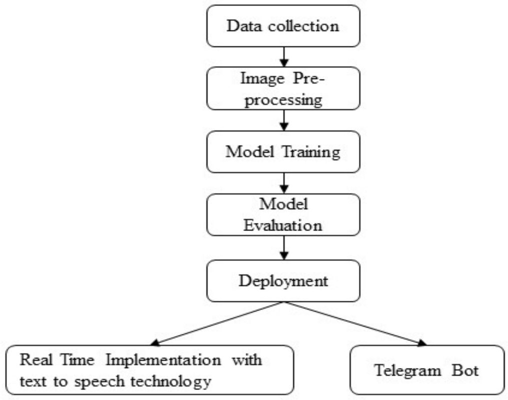

Fig.1 Workflow Model

A. MobileNet

MobileNet V2 is a deep learning architecture that has been known for its speed and efficiency, making it a suitable option for real-time applications such as sign language recognition. MobileNet V2 is based on depthwise separable convolutions, which lowers the number of parameters required for training while still maintaining high accuracy. Through the experiments, we found that the MobileNet V2 architecture achieved high accuracy while maintaining fast inference times, making it a good choice for real-time sign language recognition applications. The architecture diagram of MobileNet V2 has been depicted in Fig.2.

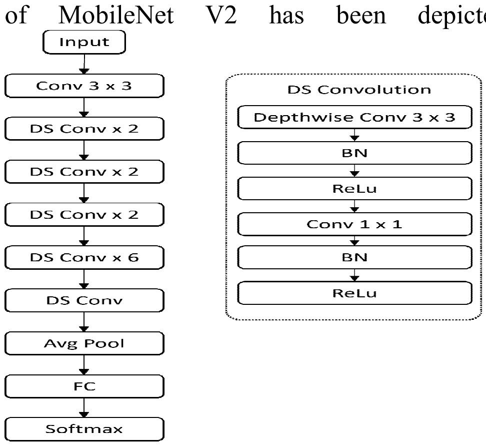

Fig.2 Architecture of MobileNet V2

B. *Xception*

Xception is a deep CNN architecture and is a variant of the Inception architecture.

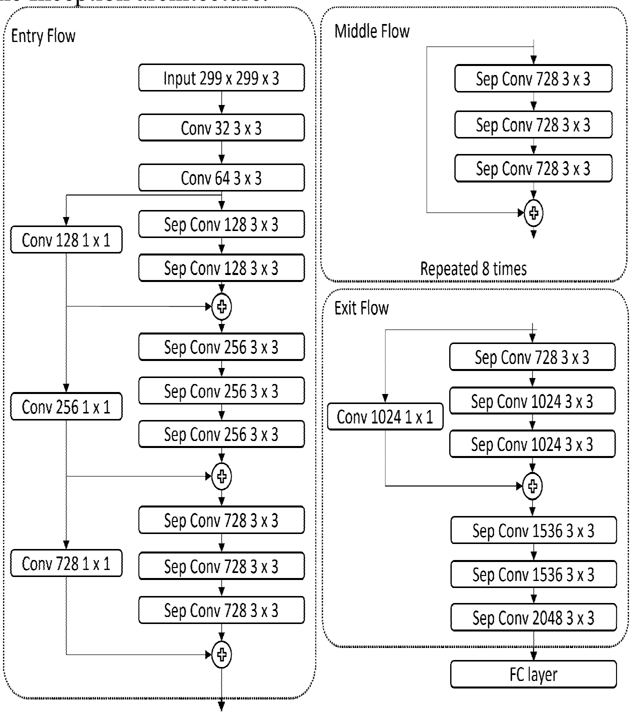

Fig.3. Architecture of Xception

Xception employs depth-wise separable convolutions which is achieved by performing spatial convolutions and channel-wise convolutions separately. Xception has 3 parts namely Entry flow, Middle flow, and Exit flow, and flow has different layers. The architecture of the Xception has been depicted in Fig.3.

C. *ResNet 101*

ResNet-101 is a deep CNN architecture that was introduced in 2016 as part of the ResNet family of models.

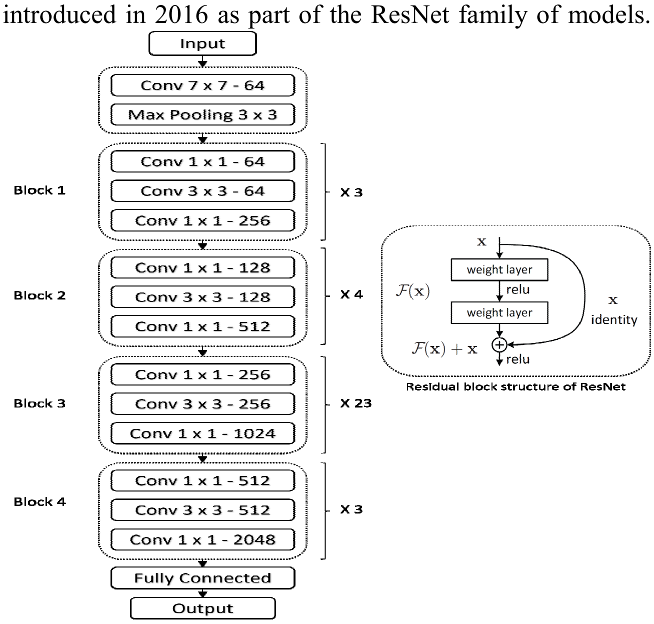

Fig.4 Architecture of ResNet 101

It is a variant of the original ResNet architecture that consists of 101 layers and employs residual connections to allow the model to discover more effective ways to represent the input data. Residual connections help to mitigate the vanishing gradient problem, which can occur in very deep networks and make it difficult to train them effectively. The architectural diagram has been shown in Fig.4.

D. *DenseNet 121*

DenseNet-121 is a deep CNN architecture that was introduced in 2016. It is part of the family of DenseNets, which are designed to address the vanishing gradient problem in deep neural networks by encouraging feature reuse and reducing the number of parameters. A convolutional layer, four dense blocks, and a final classification layer are among the 121 layers in DenseNet-121. Each dense block consists of several convolutional layers that are densely connected. The outputs of all previous layers are concatenated and used as a input for the block's following layer, allowing each layer to access the feature maps of all preceding layers. This dense connectivity helps to propagate gradients and encourages feature reuse, which leads to better training efficiency and improved accuracy. The architectural diagram of the DenseNet 121 is depicted in Fig.5.

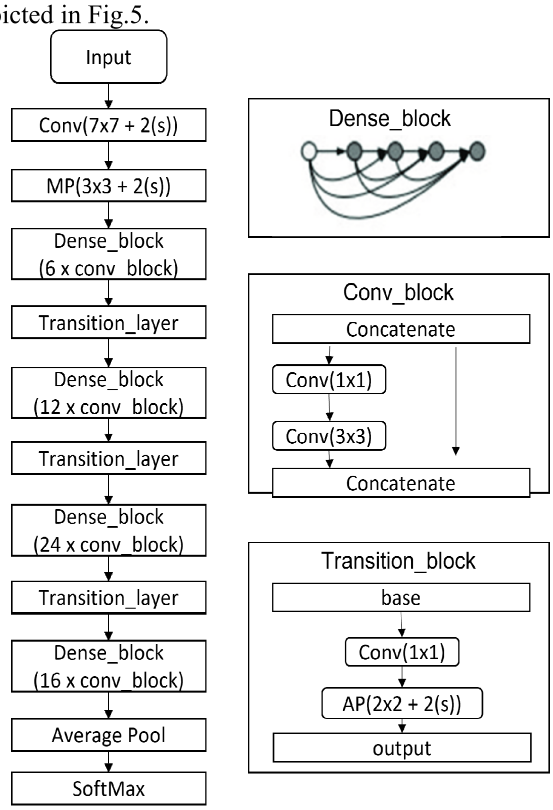

Fig.5 Architecture of DenseNet 121

E. *Deployment*

The proposed methodology uses pysttx3, which provides a useful tool for generating high-quality and natural-sounding speech output in the implementation of sign language detection. By combining deep learning gesture recognition with TTS technology, the proposed method contains a powerful system that has the potential to improve communication and accessibility for people with hearing disabilities.

A Telegram bot have been developed for sign language detection which is designed to receive an image from the user, then processed using the deep learning model to classify the sign language gesture. The corresponding sign language output is then generated, which is sent back to the user through the Telegram bot.

V. RESULTS

A. MobileNet

As stated earlier the training and testing images are given to the various transfer learning models. The specified models were implemented using python. The table 1 shows the accuracy and loss for all the four models used in the proposed work.

Algorithms	Accuracy		Loss		Precision	Recall
	Training	Validation	Training	Validation		
MobileNet	0.907	0.873	0.435	0.521	0.755	0.726
Xception	0.957	0.941	0.211	0.272	0.865	0.812
ResNet 101	0.976	0.864	0.208	0.579	0.958	0.857
DenseNet 121	0.989	0.981	0.131	0.236	0.967	0.892

Table 1 Comparison of the 4 algorithms

The accuracy and loss curves for the four models are shown in figures 6-9.

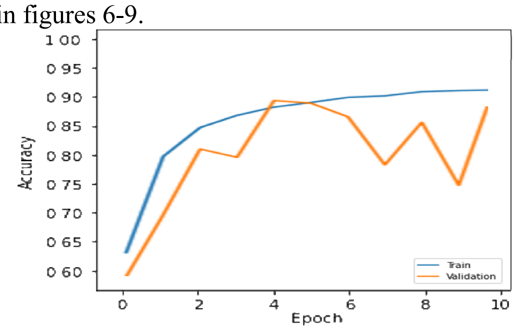

Fig.6(a) Accuracy graph for MobileNet

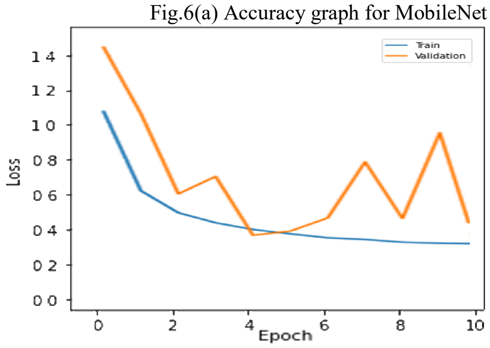

Fig.6(b) Loss graph for MobileNet

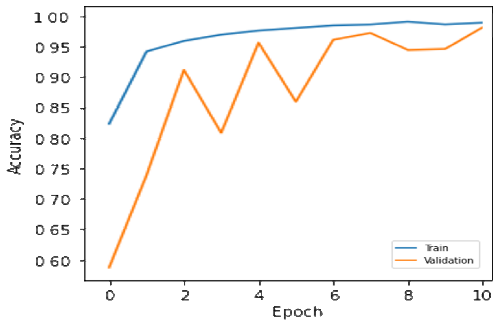

Fig.7(a) Accuracy graph for Xception

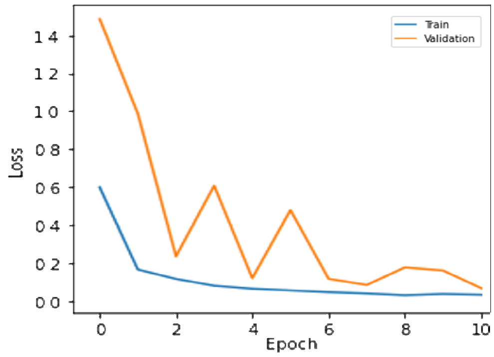

Fig.7(b) Loss graph for Xception

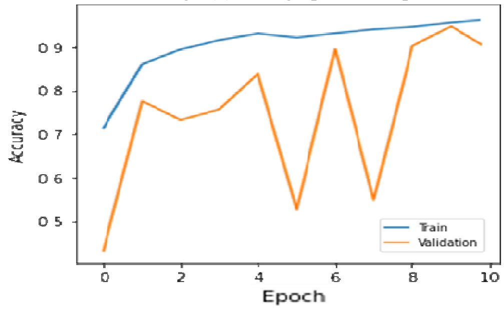

Fig.8(a) Accuracy graph for ResNet 101

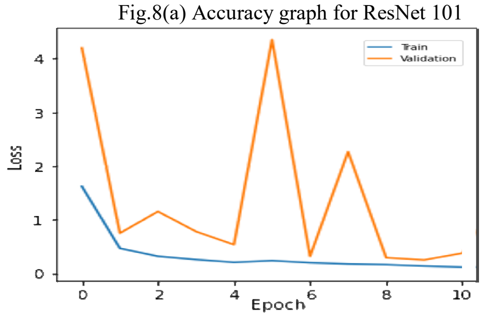

Fig.8(b) Loss graph for ResNet 101

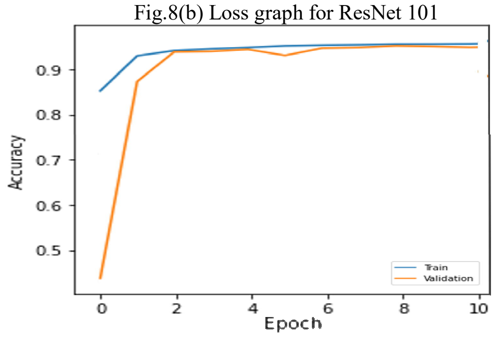

Fig.9(a) Accuracy graph for DenseNet 121

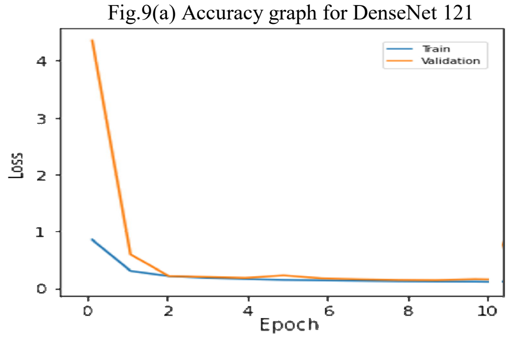

Fig.9(b) Loss graph for DenseNet 121

The performance of the proposed gesture-based sign language detection model for the 4 algorithms has been evaluated by a confusion matrix shown in Fig.10. The horizontal rows describe the ground truth classes, while the vertical columns describe the predicted classes. The offdiagonal elements describe the misclassified classes, whereas the diagonal elements represent the number of classes that were correctly classified for each class in the dataset.

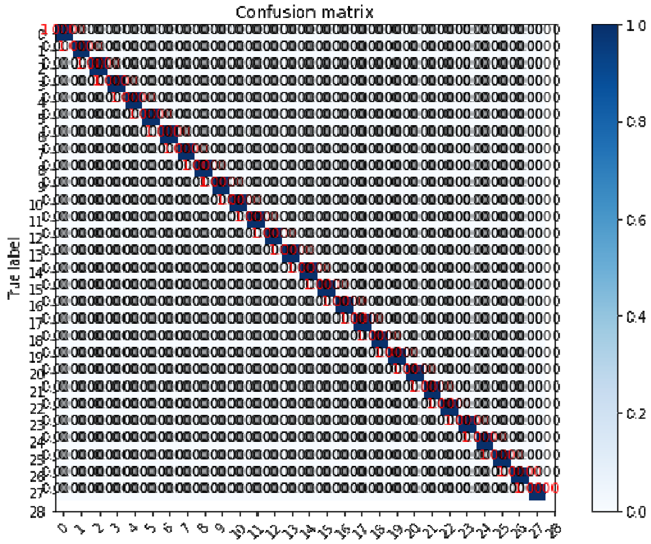

Fig.10 Confusion Matrix

From Table 1, it is evident that the model using the Densenet 121 gives a good detection accuracy, with an accuracy of 98.9%. Also the loss of the Densenet model is less compared to the other three model. The major reason for this high accuracy and less loss in densenet 121 is due to the more number of layers in its architecture. This gives a good training and validation parameters.

Deployment

The proposed method has a real-time detection using a web camera and the MobileNet model has been used to predict the class. The predicted class and the corresponding confidence score are then displayed on the video stream in real time. The system handles three levels of confidence, indicating high, low, or no confidence in the prediction. Additionally, a text-to-speech feature have been incorporated that announces the predicted class using a synthesized voice. This feature makes the system more accessible and user-friendly, particularly for individuals with hearing impairments or those who may not be proficient in sign language. The output images have been shown in Fig.11(a) and Fig.11(b).

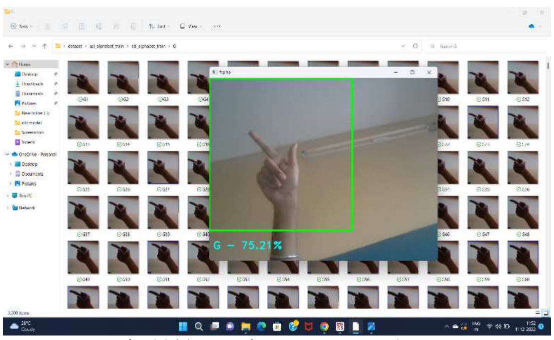

Fig.11(a) Sample Output Image 1

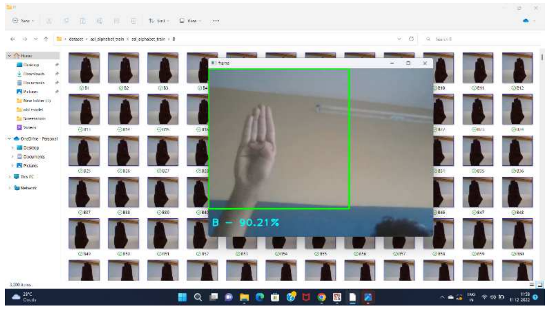

Fig.11(b) Sample Output Image 2

A Telegram bot have been developed that uses the proposed system to predict the class of hand gestures in user-submitted images. The bot accepts an image as input from the user and predicts the class. The predicted class and corresponding confidence score are then returned to the user as a message. The bot provides an easy-to-use interface for users to quickly obtain predictions without requiring any technical expertise. The predicted image from the telegram bot has been depicted in Fig.12.

Fig.12 Sample Predicted Image using Telegram Bot

II. CONCLUSION

In this paper, we proposed and evaluated four different deep learning algorithms, including MobileNet, Xception, ResNet 101, and DenseNet 121, for real-time sign language detection. The results showed that all four algorithms achieved high classification accuracy but among the four algorithms, DenseNet 121 achieved the highest accuracy, with an accuracy of 98.9%. To further improve the usability and accessibility, text-to-speech technology and a Telegram bot have been integrated. The Telegram bot acts as a bridge between sign language users and non-signers, allowing the sign language gestures to be translated into text messages.

ACKNOWLEDGMENT

We wish to show sincere gratitude to “Ms. A. Kalaiselvi” who helped a lot with this project and thank our college for their cooperation in the completion of this project.

REFERENCES

[1] A. Kumar, K. Thankachan and M. M. Dominic, "Sign language recognition," 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 2016, pp. 422-428, doi: 10.1109/RAIT.2016.7507939.

[2] T. Karayılan and Ö. Kılıç, "Sign language recognition," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 2017, pp. 1122-1126, doi: 10.1109/UBMK.2017.8093509.

[3] A. S. Antony, K. B. V. Santhosh, N. Salimath, S. H. Tanmaya, Y. Ramyapriya and M. Suchith, "Sign Language Recognition using Sensor and Vision Based Approach," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-8, doi: 10.1109/ACCAI53970.2022.9752580.

[4] K. Amrutha and P. Prabu, "ML Based Sign Language Recognition System," 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2021, pp. 1-6, doi: 10.1109/ICITIIT51526.2021.9399594.

[5] Yuqian Chen and Wenhui Zhang, "Research and implementation of sign language recognition method based on Kinect," 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, 2016, pp. 1947-1951, doi: 10.1109/CompComm.2016.7925041.

[6] Tejas Sukumar, Shashank K S and Mohammed Safeel. "Sign Language Recognition Techniques-A Review. "International Conference for Innovationin Technology (INOCON).pp.1-9,2020.

[7] E. Pereira-Montiel, E. Pérez-Giraldo, J. Mazo, D. Orrego-Metaute, E. Delgado-Trejos, D. Cuesta-Frau, J. Murillo-Escobar, Automatic sign language recognition based on accelerometry and surface electromyography signals: A study for Colombian sign language, Biomedical Signal Processing and Control, Volume 71, Part B, 2022, 103201, ISSN 1746-8094.

[8] Arpita Halder, Akshit Tayade, Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning, International Journal of Research Publication and Reviews, Volume 2, Issue 5, 2021, pp.9-17, ISSN 2582-7421.

[9] Raval, Jinalee Jayeshkumar, and Ruchi Gajjar."Real-time Sign Language Recognitionusing Computer Vision. "International Conference on Signal Processing and Communication (ICPSC), pp.542-546,2021.

[10] J. Singh and D. Singh, "A Comprehensive Review on Sign Language Recognition Using Machine Learning," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-6, doi: 10.1109/ICRITO56286.2022.9965118.

[11] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.

[12] A. Das, S. Gawde, K. Suratwala and D. Kalbande, "Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images," *2018 International Conference on Smart City and Emerging Technology (ICSCET)*, Mumbai, India, 2018, pp. 1-6, doi: 10.1109/ICSCET.2018.8537248.

[13] V. Deepika, A. Kalaiselvi and G. Dhivyaarthi, "Monitoring of Hydroponics Plant and Prediction of Leaf Disease using IOT," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2021, pp. 1-5, doi: 10.1109/ICAECA52838.2021.9675651.

[14] Soundarya. C, Kalaiselvi. A and Surya. J, "Brain Tumor Detection Using Image Processing," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2022, pp. 582-587, doi: 10.1109/ICACCS54159.2022.9785298.

[15] K Shantani, G Bhavisha, C Sanjitha, S Nagarathinam, A Kalaiselvi, "Detection of Lymphoma from the Bone Marrow Microscopic Images using Convolutional Neural Networks," 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2022, pp. 1151-1158, doi: 10.1109/ICIRCA54612.2022.9985564.

[16] Kalaiselvi, A., Nagarathinam, S., Paul, T. D., & Alagumeenaakshi, M. Detection of Autism Spectrum Disorder Using Transfer Learning. *Turkish Journal of Physiotherapy and Rehabilitation*, *32*, 2, 2021, pp. 926 – 933.

[17] Wen, F., Zhang, Z., He, T. *et al.* AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove. *Nat Commun* **12**, 5378 (2021). https://doi.org/10.1038/s41467-021-25637.

[18] Najla Musthafa, C.G. Raji,Real time Indian sign language recognition system,Materials Today: Proceedings,Volume 58, Part 1,2022,Pages 504-508,ISSN 2214-853, https://doi.org/10.1016/j.matpr.2022.03.011.

[19] Sharma, S., Singh, S. Recognition of Indian Sign Language (ISL) Using Deep Learning Model. *Wireless Pers Commun* **123**, 671–692 (2022). https://doi.org/10.1007/s11277-021-09152-1

[20] Prasad, R.K., Boruah, A., Majumdar, S. (2023). Real-Time Hand Gesture Recognition Using Indian Sign Language. In: Saraswat, M., Chowdhury, C., Kumar Mandal, C., Gandomi, A.H. (eds) Proceedings of International Conference on Data Science and Applications. Lecture Notes in Networks and Systems, vol 551. Springer, Singapore. https://doi.org/10.1007/978-981-19-6631-6_64