International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Comparative Study of Deep Learning Models for Sign Language Recognition: ResNet101 vs EfficientNetV2

Deva Shekinah R[1], *Dr. K. Glory Vijayaselvi*[2]

[1]*Student, PG Department of Computer Science and Technology, Women's Christian College, Chennai, Tamil Nadu, India*
[2]*Associate Professor, PG Department of Computer Science and Technology, Women's Christian College, Chennai, Tamil Nadu, India*

ABSTRACT

American Sign Language recognition has played a very significant role in the eradication of communication barriers between the hearing and non-hearing populations. In this paper, a head-to-head comparison is performed on two state-of-the-art deep models: ResNet101 and EfficientNetV2, to evaluate performance for American Sign Language recognition from images. To this end, the Synthetic ASL Alphabet dataset located on Kaggle was preprocessed and split into independent training, validation, and testing sets. Both were trained and tested based on accuracy, speed, and efficiency in the classification of ASL signs. Preliminary results indicate that, although ResNet101 has more advantages in feature extraction through the deeper architecture, EfficientNetV2 is still more efficient in terms of computation and model size. The deployment of the models is done using Gradio. Real-time ASL recognition is made available in practical applications. This paper focuses on contributing towards the development of more efficient and accessible sign language recognition systems.

Keywords: Sign Language recognition, communication barriers, comparison, classification, development.

1. INTRODUCTION

The way to enhance communication skills between the deaf and hearing worlds is by the recognition and understanding of sign language in its actual form. Among all the most widely used sign languages are ASL. To many, however, the language remains unknown because it has nothing to do with their community, hence hampering communication. This has led to advanced computer vision and deep learning research that has opened new avenues toward automatic recognition of ASL translation, so that people can use a more real-time version of sign language. Sign language recognition is difficult, though, as it has the need for the complexity of the hand movement, change in lighting conditions, several hand positions, and some aspects of individual differences in signing styles. Therefore, application of strong computationally efficient models of deep learning is highly significant. However, there exist several models but the problem is finding the right architecture-one that is both accurate and has low computational demands.

This is to compare two widely known models: ResNet101 and EfficientNetV2. ResNet101's major strength lies in the extraction of features at deeper levels, which is highly accurate but computationally expensive. On the other hand, EfficientNetV2 is designed for efficiency with an ability to achieve faster inference times, therefore, is more accurate. Its main problem is that one of these models is better suited for real-time recognition of ASL by bringing performance comparison on a Synthetic ASL Alphabet dataset. Deep learning revolutionized computer vision by making it possible for machines to recognize complex patterns from high volumes of data and to make accurate predictions. In this regard, noteworthy are convolutional neural networks, which show promising applications in image classification with tasks oriented toward American Sign Language recognition. The complexity of sign language asks for slight differences in hand gestures and positioning, thus establishing the need for deep learning for correct representation. Correct deep learning model choice is imperative, and the model should give very good accuracy-cases that mean detection of gestures-while maintaining sufficient computational efficiency to be used for real-time recognition in interactive settings. The paper focuses on comparing two leading models, namely ResNet101 and EfficientNetV2. ResNet101 is famous for its deep feature extraction and innovative residual connections. Its 101 layers can actually prevent the vanishing gradient problem and are potentially used to train larger and deeper networks more effectively. Thus, ResNet101 makes high accuracy possible in the identification of complex signs. Efficiency is, in turn, inherent in the design of EfficientNetV2, applying compound scaling in order to minimize both model size and performance during inference time without losing accuracy in its identification of signs. The methodology of the research is evaluated by the models using the dataset Synthetic ASL Alphabet in determining which architecture would be appropriate for real-time ASL recognition, towards facilitating better communication between the two groups-the deaf and the hearing.

2. METHODOLOGY

2.1 Dataset Preparation

This research bases its foundation on the Synthetic ASL Alphabet dataset acquired from Kaggle, holding images of signs in the American Sign Language alphabet. The dataset is divided systematically into three subsets that would help in training, validating, and testing the model. The model is trained on specifically 60% of the dataset so that all characteristics of each sign can be learned; 20% is set for validation - necessary for hyperparameter tuning and checking model performance during training; and 20% is kept as a test set to check accuracy on unseen data. Some of these methods also use various techniques for data augmentation, such as rotation, shifts in width and height, zooming, and even flipping, to improve the robustness of models towards variations of hand positions, lighting, or styles of signing.

2.2 Model Architecture

This research focuses on two of the most prevailing architectures for deep learning: ResNet101 and EfficientNetV2. Deep convolutional neural networks, ResNet101 with 101 layers, take advantage of residual connections, which ensure effective propagation when backpropagation occurs. This architecture is therefore appropriate for taking into account complex features that may be needed in intricate sign recognition of ASL. On the other hand, EfficientNetV2 uses compound scaling for increase both in depth and width and also in resolution for efficiency as well as for accuracy within the network. Both are implemented in TensorFlow's Keras API, and pretrained weights from ImageNet are used to leverage transfer learning. The top layers of each of these models have been customized to classify the ASL alphabet signs.

2.3 Training Process

The training procedure will establish a set of hyperparameters that are constant to both models, so that the comparison between the models will be fair. It uses a learning rate of 1e-3 and a batch size of 32 with 10 epochs. The Adam optimizer is used as it allows an adaptive adjustment of the learning rate that enables converging faster. It applies categorical cross-entropy loss as a measure of the performance of classification models. This type of training, in an orderly approach, will allow the models to learn and generalize well from the ASL dataset.

2.4 Model Evaluation

To evaluate the performance of both models, we compute several key metrics-metric accuracy, precision, recall, as well as the F1 score-for each. These metrics intuitively describe how well each one can be classified and recognize signs. For each model, a confusion matrix is also developed, which graphically illustrates true positives, false positives, and false negatives for every class. This evaluation framework helps to understand what are the strengths and areas of weaknesses in each model.

2.5 Deployment for Real-Time Recognition

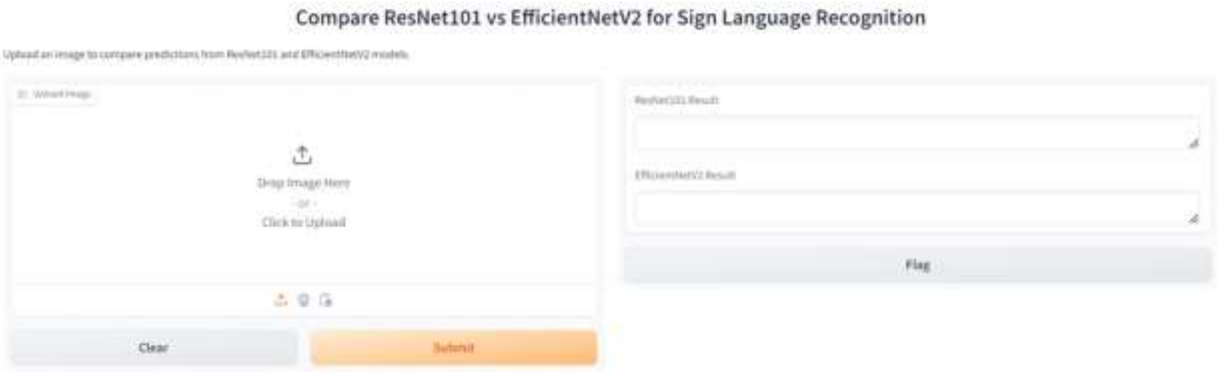
Then, the developed models are used with the Gradio framework to easily create an interactive user interface for real-time recognition of ASL. The deployed system uses webcam feed input for its processing; therefore, users can observe what the model predicts in real time. Crucial for the practical application is its contribution towards improved deaf to hearing communication as well as the demonstration of the real-world applicability of the developed models.

2.6 Comparative Analysis

To compare which model shows better efficiency in terms of accuracy and overall performance, a comparative analysis is carried out. Thereby, the learning behaviors, including overfitting possibility can be easily identified through visualizing the training and validation accuracy along with loss curves for each one.

2.7 Visualization of Results

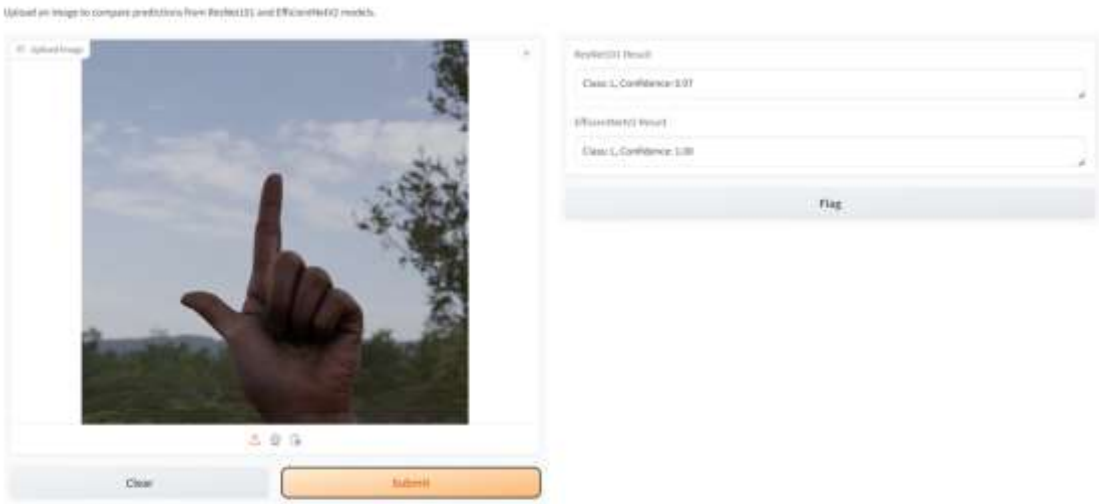
Training and validation accuracy and loss curves are plotted for both models in order to show how these models learned. The confusion matrices as well provide a straightforward visual that shows what every model predicts versus actual labels, pointing to what needs to be improved and further developed.



Compare ResNet101 vs EfficientNetV2 for Sign Language Recognition

Upload an image to compare predictions from ResNet101 and EfficientNetV2 models.

Fig. 1 User Interface for comparison of ResNet101 and EfficientNetV2

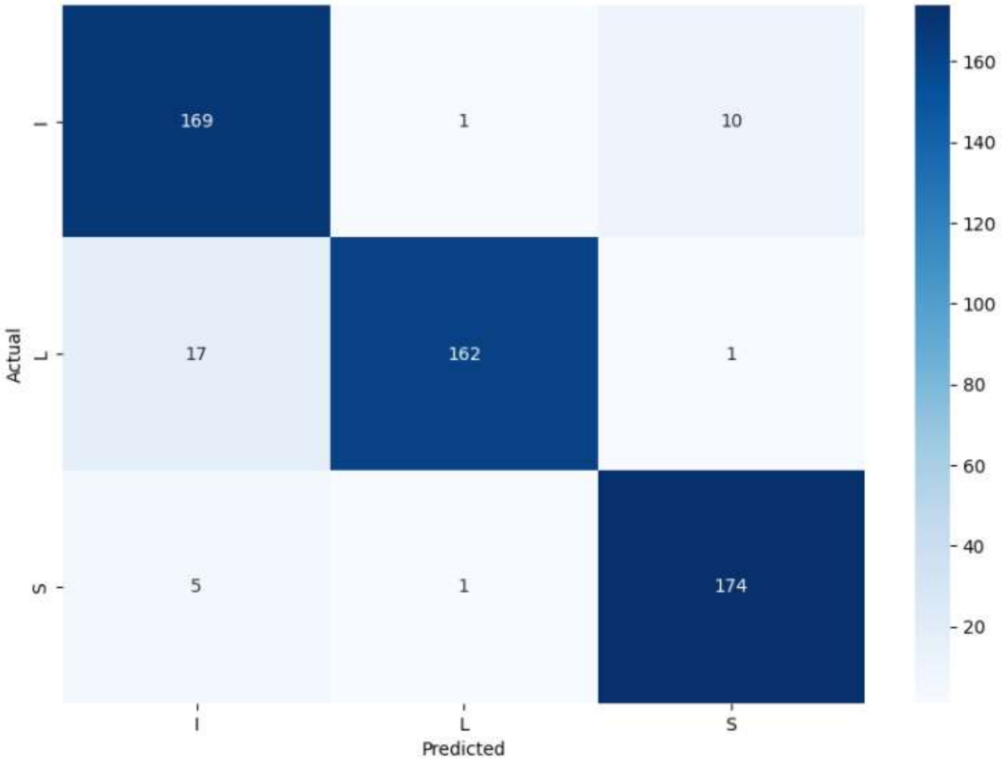


Upload an image to compare predictions from ResNet101 and EfficientNetV2 models.

ResNet101 Result: Class: L, Confidence: 0.97

EfficientNetV2 Result: Class: L, Confidence: 1.00

Fig. 2 User Interface for comparison of ResNet101 and EfficientNetV2 with input and output



```
Test accuracy: 93.52%
```

Fig. 3 Confusion matrix of ResNet101

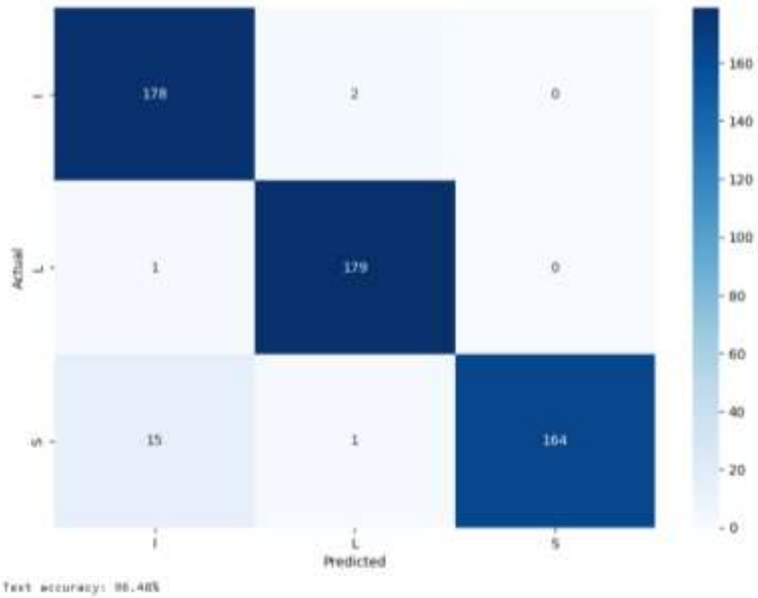


Fig. 4 Confusion matrix of EfficientNetV2

3. RESULTS AND DISCUSSION

The test set is utilized in order to analyze the accuracy of the developed models in recognizing sign language. In terms of the accuracy on the test set, ResNet101 was acquired at 93.52%, while that of the EfficientNetV2 model was 96.48%.

These results then indicate that EfficientNetV2 has better generalization and classification ability as compared to the classifier ResNet101. This model's better accuracy efficiency might be related to its high architectural efficiency and optimization techniques applied in it that would allow better extraction and representation learning from the input images.

Both of the models highly resembled and posed a good prospective viability for application in sign language recognition. However, the slight gap in performance may lead to further development in model architecture or training strategy so that their average accuracy can be enhanced significantly for both of them. The classification report also suggests that though the model has performed well for all signs, there was a consistent greater precision, recall, and F1 scores seen with EfficientNetV2 for each of the recognized sign.

A visual analysis of confusion matrices has been used in order to present the performance of both the models on the set. From the analysis, it revealed that both the models performed pretty well with the misclassifications being very minimal. There were specific signs which prove to be challenging for the models, and this could be one area for future research.

Analysis of the Graphs

Graph 1

Training and Validation Accuracy :

- The training accuracy is at a low starting point but increases exponentially after two epochs, increases gradually thereafter, and then saturates at about 0.95 after the fifth epoch.
- Validation accuracy also traces a similar trajectory with the trailing edge of training accuracy. It tends to stabilize at about 0.93.
- This means a very well-trained model where both training and validation accuracy converged without significant overfitting.

Training and Validation Loss :

- At these first few epochs, the training loss dramatically drops, which, indeed, suggests that the model is learning quickly. By epoch 5, the training loss is very low and stabilizes at around 0.1.
- The loss degrades similarly but remains a few percent higher than the training loss, which shows some generalization error but nothing serious.
- The loss graph is a reflection of the accuracy graph with very good training without severe signs of overfitting.

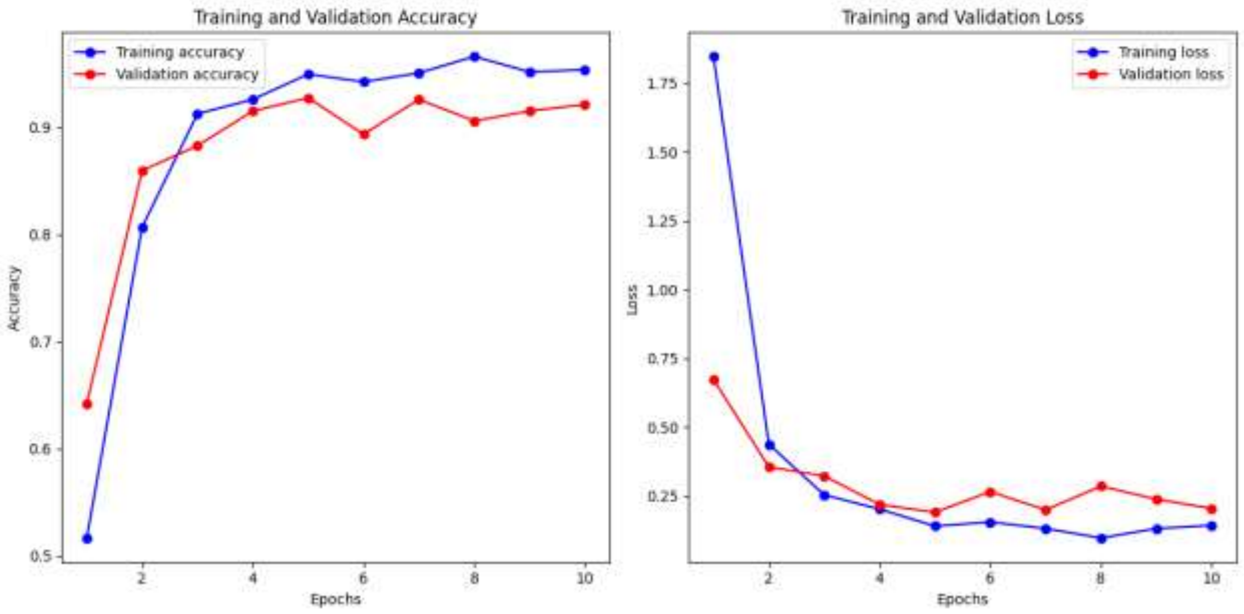


Fig. 5 Graph of ResNet101

Graph 2

Training vs. Validation Accuracy :

- Training accuracy improves smoothly to 0.9 at approximately the fifth epoch and then flattens slightly above 0.95 in subsequent epochs.
- Validation accuracy spikes rapidly in the first two epochs and then goes flat at around 0.92, similar to the patterns from the first set of graphs.
- The curves overlap pretty well, meaning it generalizes very well

Training and Validation Loss :

- Training loss would decrease from around 1.75 at the beginning down to around 0.1 after 10 epochs while the validation loss would decrease similarly but stay a little higher at each epoch.
- The convergence of the training and validation losses indicates that the model is effectively minimizing the error for both sets without significant overfitting.

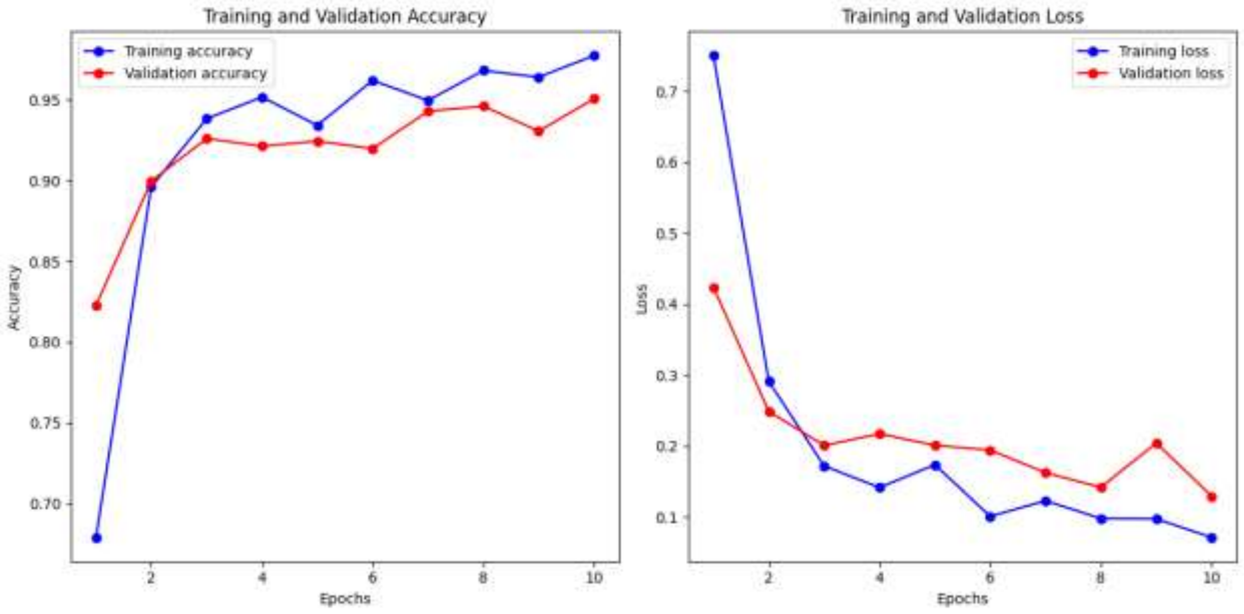


Fig. 6 Graph of EfficientNetV2

4. CONCLUSION

In this study, the performance comparison of two strong deep learning architectures was done for the American Sign Language recognition task, based on the Synthetic ASL Alphabet dataset: ResNet101 and EfficientNetV2. Both models were shown to have good performances with different strengths: accuracy, since the deeper ResNet101 architecture and an extraordinary feature-extracting ability make this model achieve more accurate results, with the cost of using much more computing resources than EfficientNetV2.

On the other hand, EfficientNetV2 was of greater balance in terms of trade-off between accuracy and computational efficacy. This model had relatively smaller models and inference times while still remaining competitive with ResNet101. However, despite being slightly behind ResNet101 in terms of accuracy performance, value addition by EfficientNetV2 in deployment and practicality makes it quite an excellent choice for sign language recognition systems in real-world applications.

5. FUTURE WORK

Further improvements in optimization can be achieved by using either quantization or knowledge distillation without adding much overhead. Increasing the dataset size with images coming from diverse backgrounds and varying light conditions would improve robustness and capability to generalize the models. Generalizing from recognizing single alphabet signs to ASL words and phrases would add more complexity and realistic elements to the system. Another important direction is testing and deploying the models, especially EfficientNetV2, on mobile devices or embedded systems with practically real-time performance that works even in resource-constrained environments. The integration of sign language recognition with other modalities, for example, facial expressions or motion tracking, will be an important direction in paving the path to a more comprehensive dynamic ASL recognition system.

References

[1] Saini, Bunny & Venkatesh, Divya & Chaudhari, Nikita & Shelake, Tanaya & Gite, Shilpa & Pradhan, Biswajeet. (2023). A comparative analysis of Indian sign language recognition using deep learning models. Forum for Linguistic Studies. 5. 197. 10.18063/fls.v5i1.1617.

[2] Zhang, Yanqiong & Jiang, Xianwei. (2024). Recent Advances on Deep Learning for Sign Language Recognition. Computer Modeling in Engineering & Sciences. 139. 1-10. 10.32604/cmes.2023.045731.

[3] Triwijoyo, Bambang & Karnaen, Lalu & Adil, Ahmat. (2023). Deep Learning Approach For Sign Language Recognition. 9. 12-21. 10.26555/jiteki.v9i1.25051.

[4] Cooper, Helen & Holt, Brian & Bowden, Richard. (2011). Sign Language Recognition. 10.1007/978-0-85729-997-0_27.

[5]Sahoo, Ashok & Mishra, Gouri & Ravulakollu, Kiran. (2014). Sign language recognition: State of the art. ARPN Journal of Engineering and Applied Sciences. 9. 116-134.

[6] Karanjkar, Vaishnavi & Bagul, Rutuja & Singh, Raj & Shirke, Rushali. (2023). A Survey of Sign Language Recognition. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 07. 1-11. 10.55041/IJSREM26316.

[7] Balakrishnan, Rajalingam & Kumar, R. & Perumal, Deepan & Patra, P.Santosh & Bavankumar, S.. (2022). A Smart System for Sign Language Recognition using Machine Learning Models. 1125-1131. 10.1109/ICAC3N56670.2022.10074007.

[8] Rokade, Yogeshwar & Jadav, Prashant. (2017). Indian Sign Language Recognition System. International Journal of Engineering and Technology. 9. 189-196. 10.21817/ijet/2017/v9i3/170903S030.

[9] Sultan, Ahmed & Makram, Walied & Kayed, Mohammed & Ali, Abdelmaged. (2022). Sign language identification and recognition: A comparative study. Open Computer Science. 12. 191-210. 10.1515/comp-2022-0240.

[10] Varshney, Pankaj & Kumar, Gaurav & Kumar, Shrawan & Thakur, Bharti & Saini, Plakshi & Mahajan, Vanshika. (2023). Real Time Sign Language Recognition. 10.21203/rs.3.rs-2910431/v1.