DEEP LEARNING BASED CLASSIFICATION OF DOUBLE-HAND SOUTH INDIAN SIGN LANGUAGE GESTURES FOR DEAF AND DUMB COMMUNITY

Ramesh M. Badiger [1], Rajesh Yakkundimath [2], Naveen Malvade[3]

Article History: **Received:** 30.05.2023 **Revised:** 02.07.2023 **Accepted:** 10.08.2023

Abstract

The field of gesture recognition, which is a part of computer science and language technology, focuses on using mathematical algorithms to analyze human gestures in non-verbal communication, particularly hand gestures or movements. This research paper presents the development of multi-stream deep transfer learning models, specifically Inception-V3, VGG-16, and ResNet-50, for recognizing signs of south Indian languages using double-hand gestures such as Kannada, Tamil, and Telugu. The classification performance of these models has been enhanced using a dataset of 10,000 double-hand gesture images. Among the models, Inception-V3 achieved the highest test accuracy of 89.5% and validation accuracy of 88.45% in classifying double-hand gesture images into ten categories. The results of this study could be used to create automated systems that help people with speech impairments or other functional limitations and enhance their talents.

Keywords: Sign language, gesture identification, transfer learning, Inception-V3, VGG-16, and ResNet-50.

[1]Department of Computer Science and Engineering, Tontadarya College of Engineering, Gadag 582 101, Karnataka, India
[2]Department of Computer Science and Engineering, K. L. E. Institute of Technology, Hubballi 580027, Karnataka, India
[3]Department of Information Science and Engineering, S.K.S.V.M. A. College of Engineering and Technology, Laxmeshwar 582116, Karnataka, India
[1, 2, 3] Visvesvaraya Technological University, Belagavi 590018, Karnataka, India

DOI: 10.31838/ecb/2023.12.6.286

1. Introduction

Designing solutions that can significantly improve the lives of individuals with disabilities is one of the biggest opportunities that Artificial Intelligence (AI) technologies have created. According to the World Health Organization, about 15% of the global population (1.2 billion people) has some type of disability. In the fight for inclusivity, companies, organizations, and individuals have developed AI tools capable of assisting people in performing daily tasks. Accessibility, disability, and artificial intelligence have become intertwined. Research in machine learning and AI is increasingly interested in either helping to medically heal disabilities or accommodating them better. One of the important areas in which AI has significantly advanced is the facilitation of non-verbal communication between people with disabilities through facial expressions, gestures, sign languages, and other body language. Sign language is a method of communication that uses hand gestures and movements rather than words. It is mostly used by deaf or hearing-impaired individuals.

Table 1: Various double-hand gesture images

Sl. No	Sign/Gesture	Letter in Kannada /Telugu/ Tamil/	Sl. No		Letter in Kannada /Telugu/ Tamil/ English Language
1		ಮೀನು / చేప / மீன் / Fish	6		ಸಮಯ / సమయం / நேரம் / Time
2		ಯಾವುದು / ఏది / எப்படி / Which	7		ಉದ್ದ / పొడవు / நீளம் / Length
3		ಅಡಿ / అడుగు / அடி / Foot	8		ಚಿಟ್ಟೆ / తిత్తిబ్బుటి / புறா / Butterfly
4		ಚರ್ಮ / చర్మం / தோல் / Skin	9		ಮನೆ / ఇల్లు / வீடு / House

5		ನವಿಲು / నెమలి / மயில் / Peacock	10		ಏನು / ఏమి / என்ன / What

A popular and developing area in computer vision and pattern recognition is the understanding of gestures and sign language. Fast and reliable hand gesture recognition is still a challenge, despite recent significant advancements. Table 1 shows some of the hand gesture illustrations. Table 2 lists all of the double-handed motions that were taken into consideration along with their corresponding characters in various south Indian languages.

Table 2: Label encoding/numerical code assignments for the considered Double-Hand gestures.

Letter in Kannada/Telugu/Tamil/ English Language	**Class Label (Alphanumerical Code)**
ಮೀನು /చేప /மீன் /Fish	**1**
ಯಾವುದು /ఏది /எப்படி /Which	**2**
ಅಡಿ /అడుగు /அடி /Foot	**3**
ಚರ್ಮ /చర్మం /தோல் /Skin	**4**
ನವಿಲು /నెమలి /மயில் /Peacock	**5**
ಸಮಯ /సమయం /நேரம் /Time	**6**
ಉದ್ದ /పొడవు /நீளம் /Length	**7**
ಚಿಟ್ಟೆ /తీత్తిబ్బుటి /புறா /Butterfly	**8**
ಮನೆ /ఇల్లు /வீடு /House	**9**
ಏನು / ఏమి /என்ன /What	**10**

The major conclusions of a survey of the literature conducted to ascertain the state-of-the-art in the development of computer vision and pattern recognition applications are presented. A novel machine learning-based hand gesture detection system is suggested in [1] using Human-Computer Interaction (HCI) on a cloud platform with support for the Internet of Things (IoT). A hand gesture dataset is developed and implemented, and the movements are then categorized using a convolutional neural network (CNN). [2] Introduced a deep learning-based system for recognizing hand shapes in American Sign Language (ASL). The system uses a CNN to learn hand form characteristics and a long short-term memory (LSTM) network to recognize signs. The study's recognition rate is 97.8%. [3] described an innovative technique that combines long short-term memory (LSTM) and convolutional neural networks (CNN) to translate Arabic Sign Language into text.

The method successfully achieved high accura cy rates of up to 96.8% for some signs when te sted on a dataset of independent users. In [5], a Twin Delayed Deep Reinforcement Memory Network (TD-DRMN) architecture-based artificial intelligence system for sign language prediction is proposed. The findings demonstrate that the TD-DRMN architecture outperformed the other models in predicting sign language gestures, achieving an accuracy of 99.8%. A deep learning-based sign word recognition system that is invariant to rotation, translation, and scale is proposed in [6]. The system employs long short-term memory (LSTM) networks to recognise sign words and convolutional neural networks (CNNs) to extract features from sign language images. The study achieved a 96% recognition rate when testing the system's precision against a dataset of Bengali Sign Language. CNN-based method for recognizing static hand gestures is presented in [7].

The results of the experiment show an improvement in accuracy of 98%, with training and validation loss over the optimal number of epochs coming out to be 0.021 and 0.064, respectively. A novel method for finger spelling localization and classification in Taiwan Sign Language (TSL) based on the You Only Look Once (YOLO) deep learning architecture is presented in [8]. The experiment using the 15,000 image training dataset and 15,000 image test dataset reveals that the model had a recognition rate of 84.99%. [9] Describes a sign language recognition system that makes use of computer vision techniques. The technology blends deep learning-based hand gesture recognition with machine learning algorithms to precisely recognize sign language motions. The study evaluates the system's accuracy on a dataset of 100 signs and gets a recognition rate of 96.3%. [10] Presents a comprehensive literature analysis of the major computer vision-based hand gesture detection methods. Methods based on deep learning and those based on features are discussed.

The challenges and limitations of hand gesture recognition methods are emphasized. [11] uses artificial neural networks (ANN) to identify datasets of sign language. The pre-training data for the ANN model, which attained an average recognition accuracy of 98.0%, was the image network sign language dataset. [12] proposes a real-time sign language recognition system that uses edge detection, pattern recognition, and skin color identification as computer vision techniques. 85% of the 100 Indian Sign Language dataset are correctly recognized by the system. [13] Presents a technique for hand sign detection from depth photos using multi-scale density characteristics, coupled with a comprehensive evaluation of the system's effectiveness. In this study, American Sign Language numbers' depth images are taken into account, and a recognition rate of 98.20% is achieved. With an average accuracy of 83.36%, a deep convolutional long short-term memory network hybrid model is utilized in [16] to recognize the recommended hand movements. To further validate the model's performance, benchmark hand gesture dataset and alternative Indian Sign Language dataset were utilized. The results indicated average recognition accuracy of 97% and 99.34 0.66%, respectively.

A substantial lot of research has been done on sign language and hand gesture identification, according to the literature review. The recognition of human poses and scenes from their photographs makes use of deep learning algorithms. The impetus for the existing work stems from the fact that, when considering double-hand gestures as types of gestures or signs, no work has been done on the identification of South Indian Sign Language (SISL) motions' images. There are four sections in the paper. In Section.2, the suggested task is described. The findings of the experiment are presented in Section 3. The conclusion is given in Section 4.

2. Proposed Methodology

The two primary parts of the suggested methodology are the preparation of the image dataset and the classification using deep learning. Figure 1 displays the block diagram outlining the steps taken in the suggested methodology.

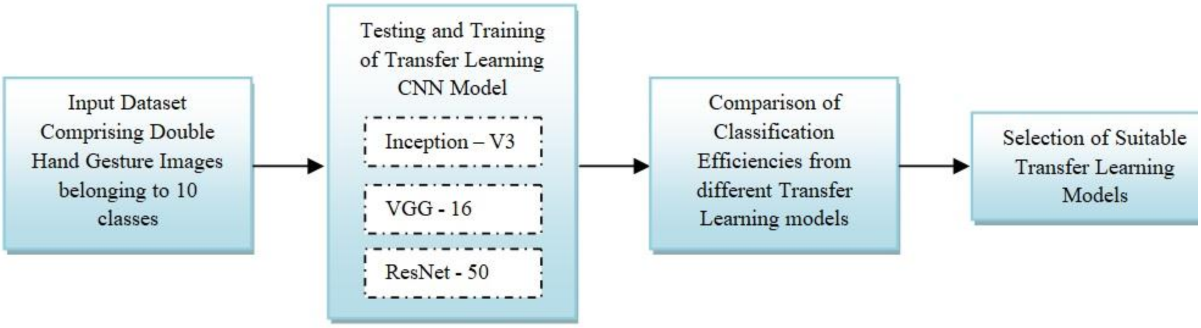

Figure 1. Block diagram of the proposed methodology.

2.1 Dataset Preparation

Ten different double-hand gesture types are taken into consideration when creating the dataset. The photos were taken using a 24.2 megapixel Nikon D3300 digital SLR camera in natural light against a black and green background. The dimensions of each image are 1080 x 2400 pixels. 500 photos are included in each category of double-handed gestures in the original image dataset, which has 5000 total images. Later, 10,000 additional photos are added to the dataset of images using a variety of traditional image augmentation techniques, including translation, arbitrary rotations, shearing, scaling, and flipping. Images from the dataset were gathered in a diversity of sizes. To cut down on the computational time required for further processing and their storage on the medium, the image size is decreased to 300 X 300 pixels. The dataset is divided into three sub-datasets: 7000 images for training, 1500 images for validation, and 1500 images for testing. Table 1 shows the specifics of the final image dataset that was employed in the project.

2.2 CNN Classifiers

In the present work, three leading transfer learning CNN models namely, Inception-V3 [14], VGG-16 and ResNet-50 [4] are deployed for the recognition and classification of Double-Hand gesture images. The descriptions of individual CNN models are available in literature [15]. The steps involved in training the considered CNN models on the prepared dataset are given in Algorithm 1.

Algorithm 1. Training and testing pre-trained CNN models on double-hand gesture image dataset.

Input: Double-hand gesture image dataset.

Output: Identified images.

Start

Step 1. Import the image dataset consisting Double-Hand gesture images.

Step 2. Generate training and testing image datasets in batches.

Step 3. Load the base CNN model and customize only the final layer.

Step 4.Compile and fit the base CNN model with different values of dropout, and different optimizers and activation functions.

Stop.

3. Experimental Results and Discussion

The MATLAB R2022b programming environment's Deep Learning Toolbox is used in the studies for classifying double-handed gestures. The Deep Network Designer tool is used to import the pre-trained CNN models under consideration and prepare them for transfer learning by altering the appropriate layer properties. The output or classification layer and final learnable layer in each model are changed to correspond to the classes in the newly created double-hand gesture image dataset. All the CNN model training and testing operations are performed on a single workstation running on Windows 11 operating system, configured with an Intel Core i9-13,900 processor, 16 GB of RAM, and NVIDIA GPU with 12 GB of memory. To get control over model training, the significant training options such as initial learn rate, validation frequency, number of epochs, and mini-batch size are initialized to 0.0001, 10, 30, and 35 respectively. The 'ReLu' function is used to activate all hidden layers, whereas the'softmax' function is used to activate the output layer. A stochastic gradient descent (SGD) technique with a categorical cross-entropy logarithmic loss function is used to fine-tune the network. Batch normalization is used to increase accuracy overall and speed up learning.

The customized CNN models are trained and validated using the augmented image dataset. The individual CNN model training progress, validation accuracy, loss, and other details of training are graphically shown in Figs. 2–4. Figs. 6–7 depict the confusion matrices for the considered CNN models when evaluated with the test image dataset. The performance comparison of all the CNN models is accomplished based on the evaluation metrics such as precision, recall, and F1 scores derived from the confusion matrices. The average evaluation metric scores obtained by all the considered CNN models for the 35 classes of double-hand gesture images are tabulated in Table 1. From Table 3, it is clear that the Inception-V3 model has the best performance, followed by the ResNet-50 and VGG-16 models in that order. The Inception-V3 model has yielded the highest evaluation metrics as well as the maximum validation and testing accuracies of 93.45% and 94.10%, respectively. The performance comparison results of all the considered pre-trained CNN models are graphically shown in Fig. 8.

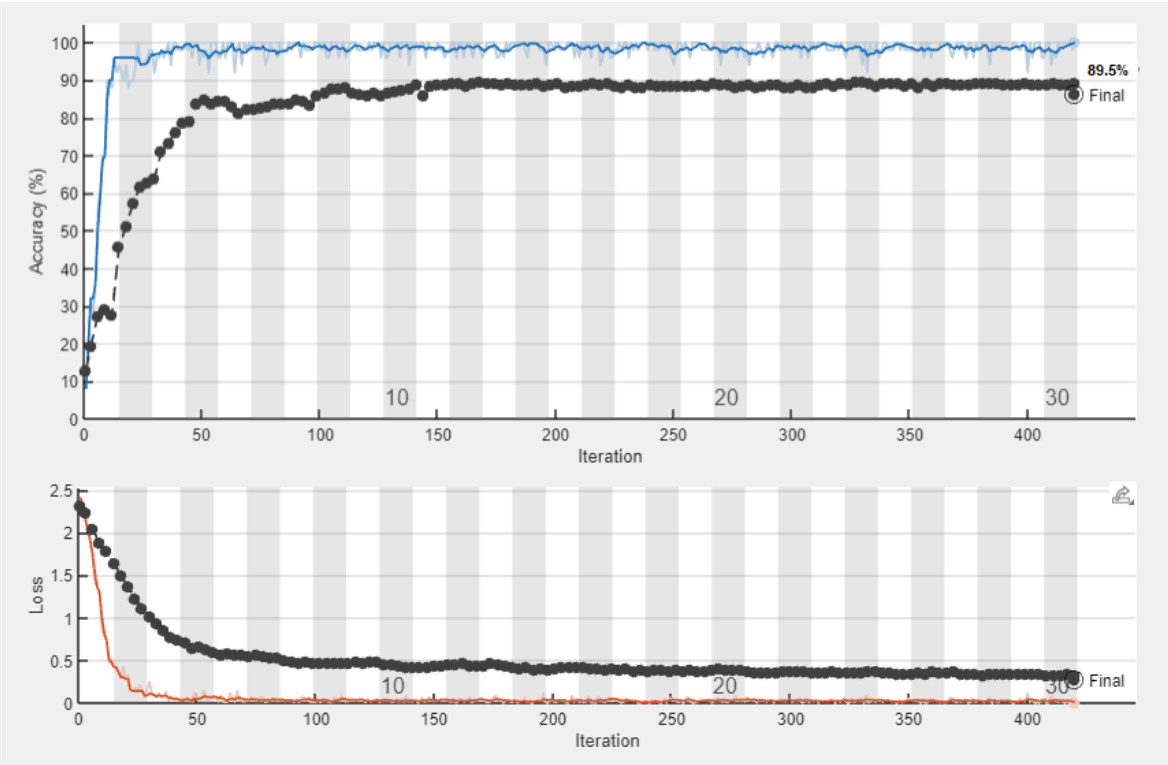

Figure 2: Plot of validation accuracy and loss graph while training Inception CNN model.

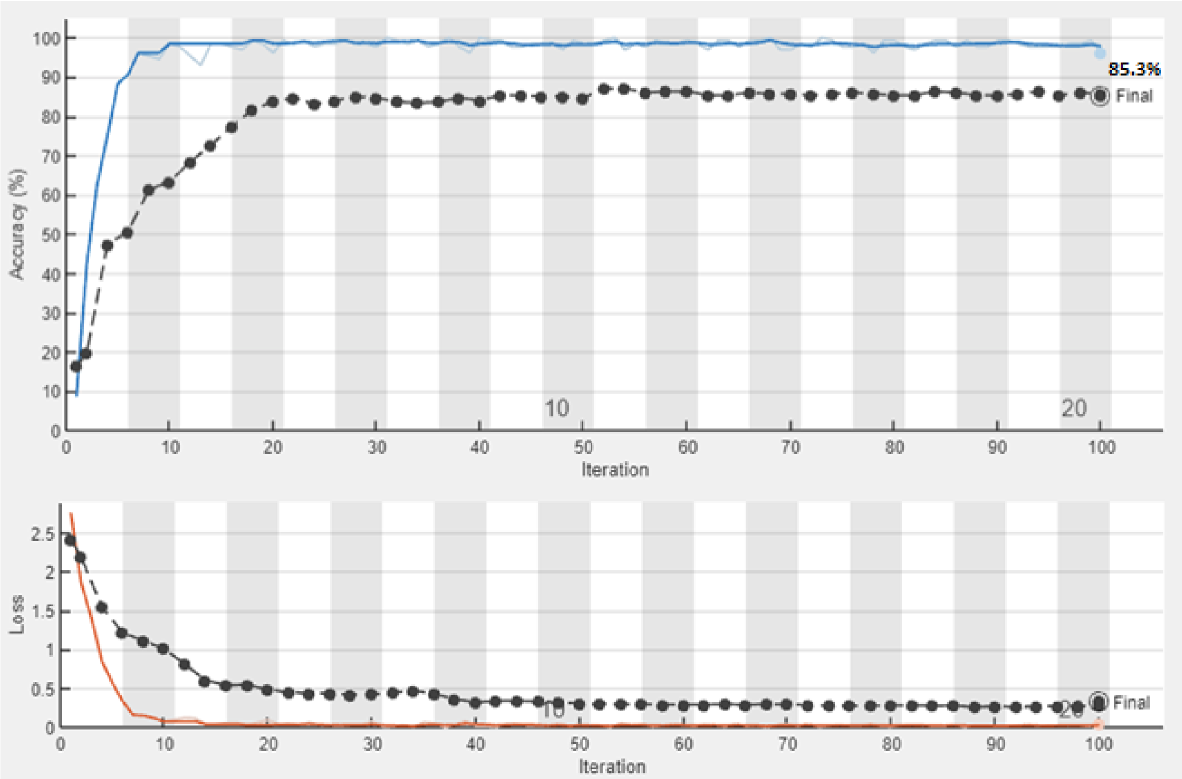

Figure 3: Plot of validation accuracy and loss graph while training ResNet 50 CNN model.

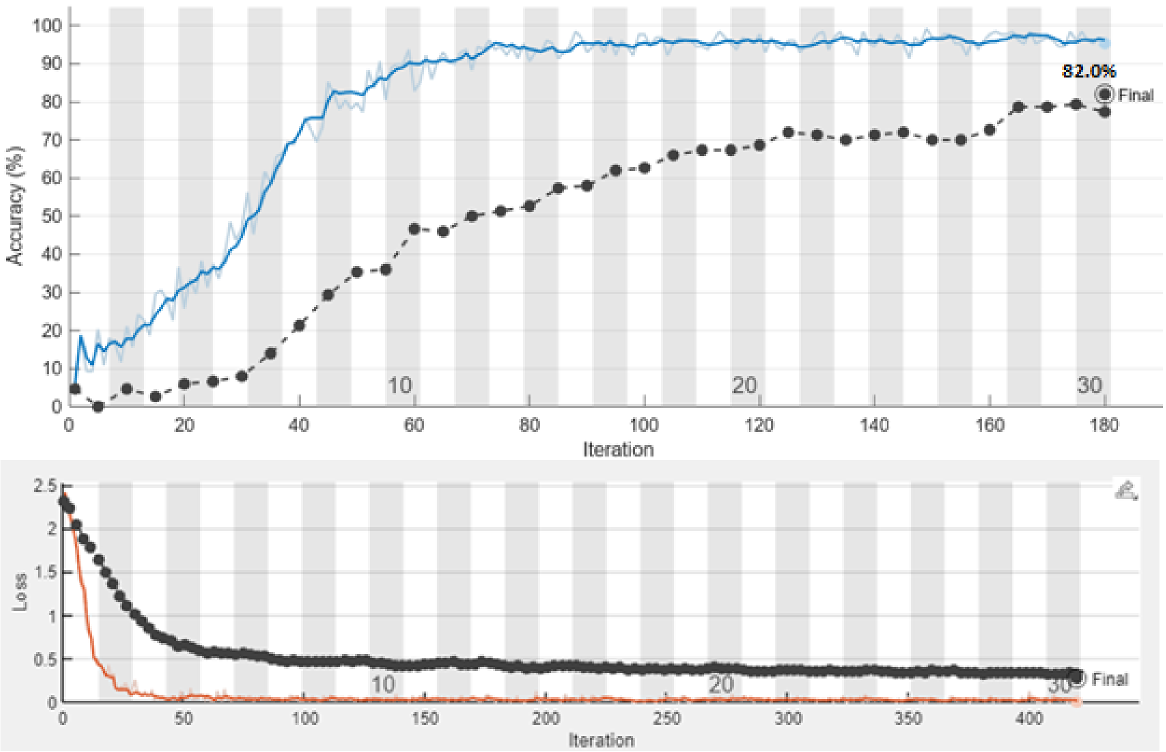

Figure **4:** Plot of validation accuracy and loss graph while training VGG 16 CNN model.

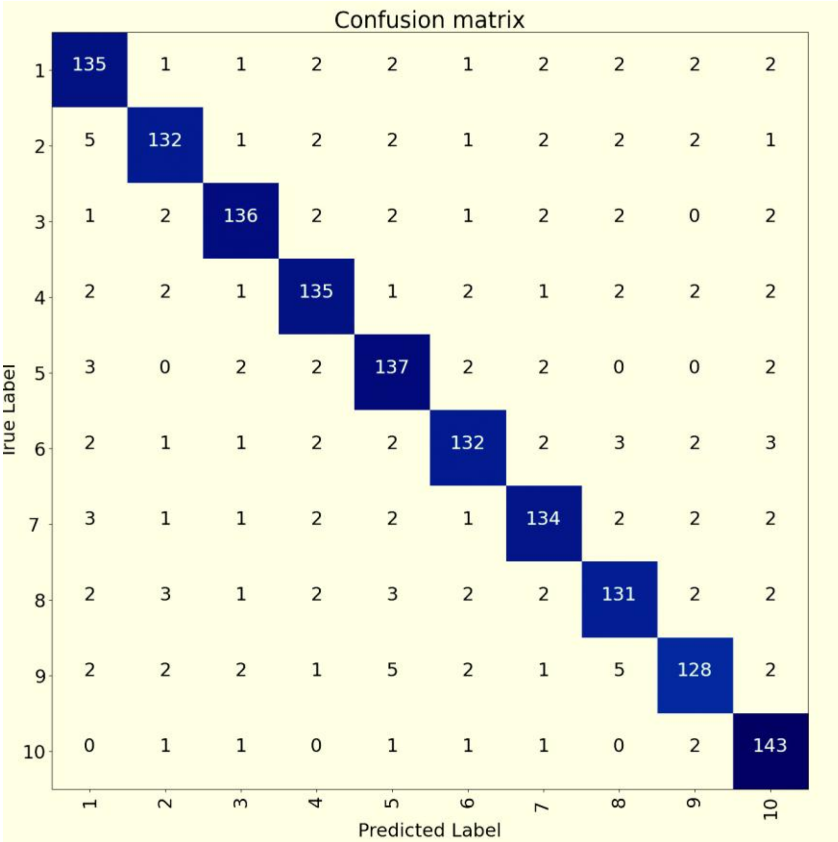

Figure 5. Confusion matrices plotted on the test dataset for the trained Inception-V3 CNN model.

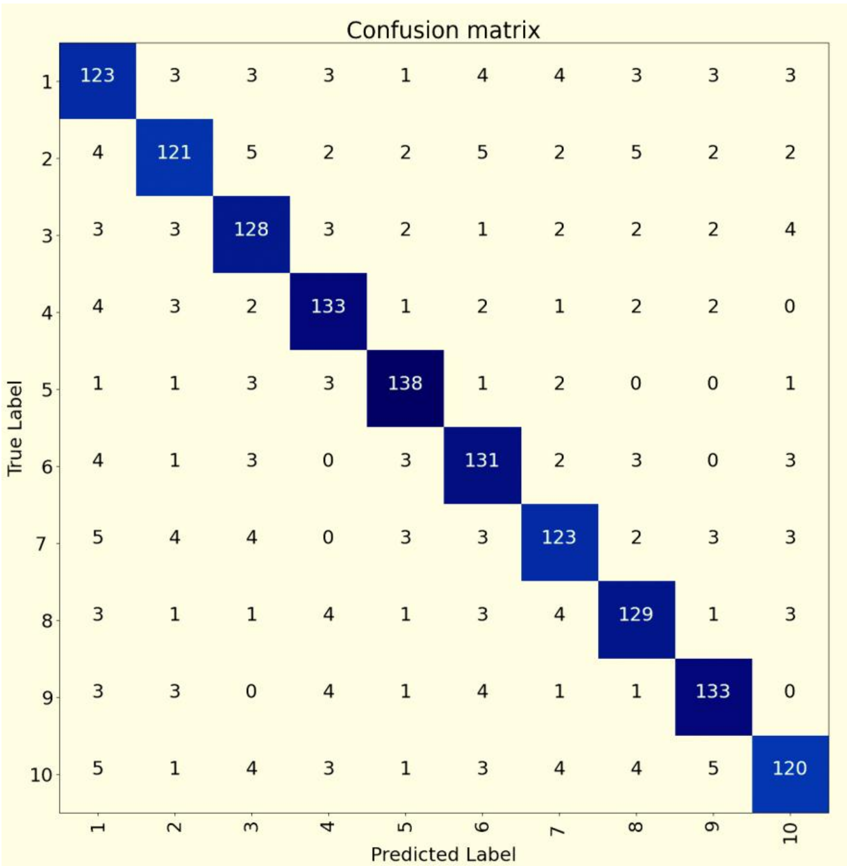

Figure 6. Confusion matrices plotted on the test dataset for the trained ResNet-50 CNN model.

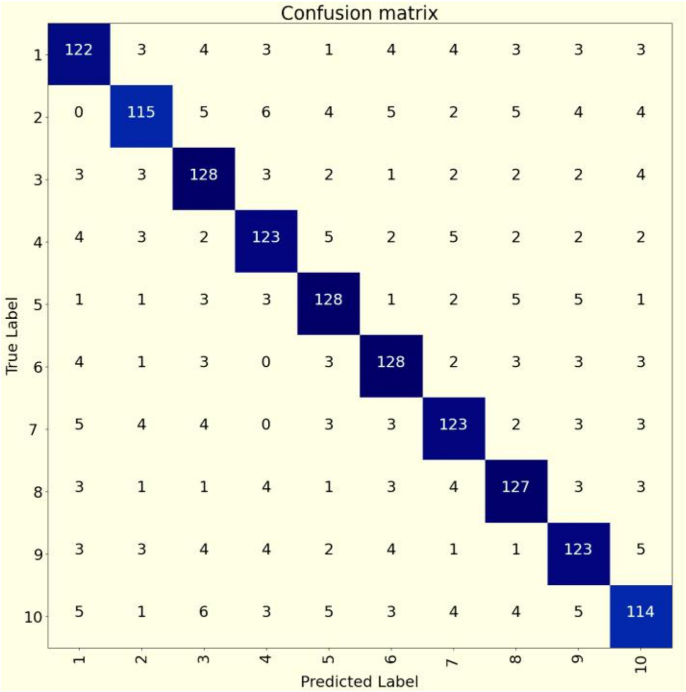

Figure7. Confusion matrices plotted on the test dataset for the trained Inception-V3 CNN model.

Table 3. Evaluation metrics derived from the confusion matrices plotted for the considered CNN models.

Sl. No	CNN model	Average Performance metrics across all the Double-Hand gesture classes			Average Validation Accuracy (%)	Average Test Accuracy (%)
		Precision	Recall	F1 Score		
1	Inception-V3	0.8940	0.8960	0.8920	88.45%	89.5%
2	ResNet-50	0.8530	0.8520	0.8490	84.23%	85.3%
3	VGG-16	0.8200	0.8200	0.8180	81.03%	82.0%

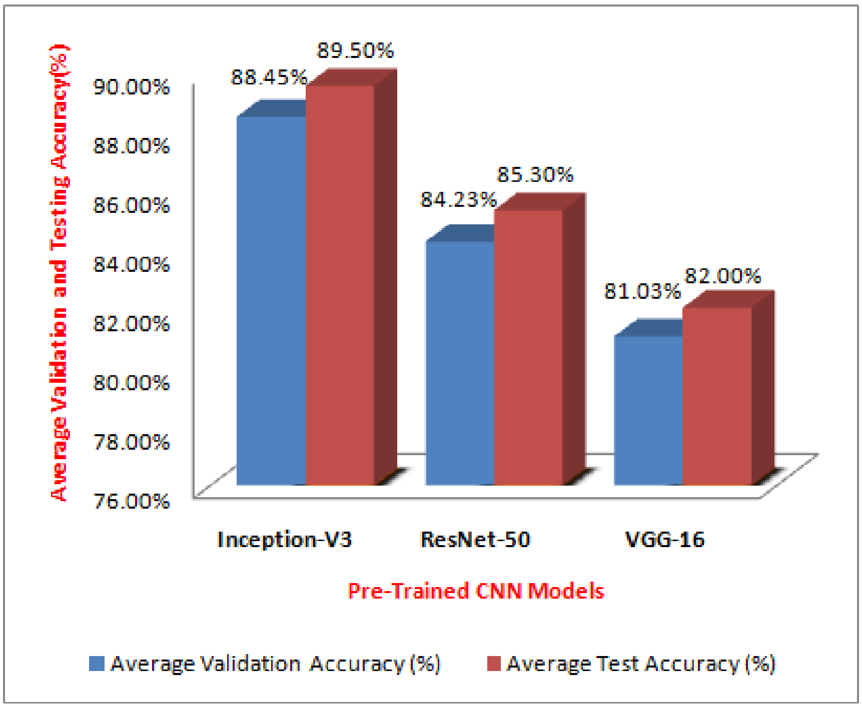

Figure 8. Performance evaluation results of all the pre-trained CNN models

4. Conclusion

The current study uses popular pre-trained CNN models, namely Inception-V3, ResNet-50, and VGG-16, to classify double-hand gesture images across 10 different categories. These models were customized to accommodate the image classes and fine-tuned to improve classification performance. While all three models produced impressive results, Inception-V3 exhibited superior performance with an average classification accuracy of 89.5%. The results of this study have potential applications in developing automated systems for identifying South Indian language gestures from still images and streaming videos. Future research could explore using the latest CNN models and incorporating publicly available image datasets to further enhance the image data set used in this study.

Acknowledgement
The authors wish to thank the anonymous reviewers for their useful comments which have enhanced the value of the paper.

Declaration of Competing Interest
The corresponding author certifies that there is no conflict of interest.

Funding
There was no specific grant from a funding agency in the public, commercial, or nonprofit sectors for this research.

5. References

1. Adhikari, S., Gangopadhayay, T.K., Pal, S., Akila, D., Humayun, M., Alfayad, M. and Jhanjhi, N.Z. A Novel Machine Learning–Based Hand Gesture Recognition Using HCI on IoT Assisted Cloud Platform. *Computer Systems Science and Engineering*, 2023, 46(2), pp. 2123-2140.
2. Al Amin Hosain, P.S.S., Pathak, P., Rangwala, H. and Košecká, J., FineHand:Learning Hand Shapes for American Sign Language Recognition. 2020 arXiv:2003.08753**.**
3. Balaha, M.M., El-Kady, S., Balaha, H.M., Salama, M., Emad, E., Hassan, M. and Saafan, M.M. A Vision-Based Deep Learning Approach for Independent-Users Arabic Sign Language Interpretation. *Multimedia Tools and Applications*, 2023, 82(5), pp.6807-6826.
4. He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, 7(3), pp. 171–180. . https://doi.org/10.48550/arXiv.1512.03385
5. Karthick, G.M., Kirubanantham, P., Saranya, A. and Sayeekumar, M. Artificial Intelligence Based Sign Language Prediction by Using the Twin Delayed Deep Reinforcement Memory Network architecture. *International Journal of Intelligent Systems and Applications in Engineering*, 2023, 11(2), pp.200-211.
6. Miah, A.S.M., Shin, J., Hasan, M.A.M., Rahim, M.A. and Okuyama, Y. Rotation, Translation And Scale Invariant Sign Word Recognition Using Deep Learning. *Comput. Syst. Sci. Eng*, 2023, 44(3), pp.2521-2536.
7. Pinto, R.F., Borges, C.D., Almeida, A.M. and Paula, I.C. Static hand gesture recognition based on convolutional neural networks. *Journal of Electrical and Computer Engineering*, 2019, pp.1-12.
8. Nakjai, P., Maneerat, P. and Katanyukul, T. Thai finger spelling localization and classification under complex background using a YOLO-based deep learning. *In Proceedings of the 11th International Conference on Computer Modeling and Simulation*, 2019, pp. 230-233.
9. Obi, Y., Claudio, K.S., Budiman, V.M., Achmad, S. and Kurniawan, A. Sign language recognition system for communicating to people with disabilities. *Procedia Computer Science*, 2023, 216, pp.13-20.
10. Oudah, M., Al-Naji, A. and Chahl, J. Hand gesture recognition based on computer vision: a review of techniques. *Journal of Imaging*, 2020, 6(8), pp.1-29.
11. Praveen, N., Hand Gesture Communication Using Deep Learning And Ann Algorithm. International *Research Journal of Modernization in Engineering Technology and Science*, 2023, 5(3), pp.2999-3005.
12. Rao, K.V., Attimarad, S.G., Sanjana, A.N., Shwetha, K. and Shraddha, M.K., Real Time Sign Language Recognition Using Computer Vision. *International*

Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, 10.1109/ICSPC51351.2021.9451709.

13. Sahana, T., Paul, S., Basu, S. and Mollah, A.F. Hand sign recognition from depth images with multi-scale density features for deaf mute persons. *Procedia Computer Science*, 2020, 167, pp.2043-2050.
14. Szegedy, C., Liu, W., Jia, Y., Scott Reed, P,S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. Going Deeper with Convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, https://doi.org/10.48550/arXiv.1409.4842, pp.1-12.
15. Too E C, Yujiana Li, Njukia S, Yingchun L. A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, 2019, 161, pp.272-279.
16. Venugopalan, A. and Reghunadhan, R. Applying Hybrid Deep Neural Network for the Recognition of Sign Language Words Used by the Deaf COVID-19 Patients. *Arabian Journal for Science and Engineering*, 2023, 48(2), pp.1349-1362.