Digit Recognition of Hand Gesture Images in Sign Language Using Convolution Neural Network Classification Algorithm

M. Navyasri and G. Jaya Suma

1 Introduction

A static sign language is a hand gesture images used for communicating instead of spoken words. Every person in the world wants to convey the information or share their emotion to other person. Communication plays an important role in day-to-day life; not only a person who knows the language will communicate with words but also through his body language or through his facial expression he will communicate to others. The person who does not know the same language, but if he can able to speak will communicate to others. The person who could not able to speak or hear, also communicates with others through hand gestures which is called as sign language. The person who knows the nonverbal language using hand gestures can communicate to the person who understand the sign language. One deaf and impaired person will face many daily life challenges. A translation process is required which can be used to interpret static nonverbal language which is a hand gesture images to text and then to voice conversion can fill the bridge gap of communication among deaf impaired person and normal person who does not know sign language. There are many sign languages based on country urban, rural and tribal areas. Machine translation which is used to translate sign language to text and voice, and voice to sign language is used in the field of education to teach and train special abled people. The remaining paper is structured as follows. Section 2 discusses the literature reviews and related works of sign language recognition. In Sect. 3, the proposed methodology using convolution

M. Navyasri (✉)
JNTUK, Kakinada, Andhra Pradesh, India
e-mail: navyasrimullapudi@gmail.com

KCCITM, Greater Noida, Uttar Pradesh, India

G. J. Suma
IIIT, Department of IT, JTUGV, Vizianagaram, Andhra Pradesh, India

B. P. Swain and U. S. Dixit (eds.), *Recent Advances in Electrical and Electronic Engineering*, Lecture Notes in Electrical Engineering 1071,
https://doi.org/10.1007/978-981-99-4713-3_32

neural network classification algorithm is illustrated followed by evaluation metrics in Sect. 4. Lastly, the conclusion and summary work are presented in Sect. 5.

2 Literature Survey

Kohlon and Singh, in the text to sign language machine translation review [1], deal with the state of the art in the advanced deep learning technologies to build the translation optimal. Farooq et al., in the sign language translation, challenging and limitations [2], provide a systematic review in all the aspects of sign language in multidisciplinary subjects. Halvardsson et al. and Rastgoo et al. [3, 4], the hand gesture recognition applying deep learning approaches for spatio temporal information using the deep learning, LSTM, SSD. And the dynamic sign language recognition using 2D convolution neural networks, 3D hand key points, SVD. Halvardsson et al. [3] use the transfer learning and convolution neural network system to provide the image perception and the mini-batch gradient algorithm during the pre-training and the accuracy model. In Sharma and Singh [5], Indian sign language recognition has created 65 different uncontrolled environments and shows the encouraging performance. In Adeyanju et al. [6], a review of hand gesture image recognition with the expert system learning methods demonstrates the remarkable success to achieve good results. In Sharma and Anand [7], Indian sign language recognition deep models perform a systematic evaluation and statistical deep models to pre-train using gradient-based optimization hyperparameters only a few ISL recognition. In Imran et al. [8], the communication gap has less understood based on the country sign and visual gesture language to communication among deaf people.

3 Proposed Methodology

The dataset covers 5000 raw image files from sign 0 to sign 9 (500 files of each sign) and 5000 corresponding output image files (applying Media Pipe) downloaded from Kaggle. This is an American Sign Language Digits Dataset, from sign 0 to sign 9. This dataset uses depth information for generating hand key points (using Media Pipe), which enriches the dataset and enhances the accuracy during classification. This is an American Sign Language Digits Dataset, using Media Pipe framework, which accurately detects the hand and 21 hand key points from a raw RGB image, and stores the co-ordinate values of these key points. The dataset contains 5000 such raw image files from sign 0 to sign 9 (500 files of each sign) and 5000 corresponding output image files (applying Media Pipe) (Fig. 1).

The graphical representation of digit images is displayed in Fig. 2, where Horizontal axis represents the label of respective digit and Vertical axis represents the no of image files.

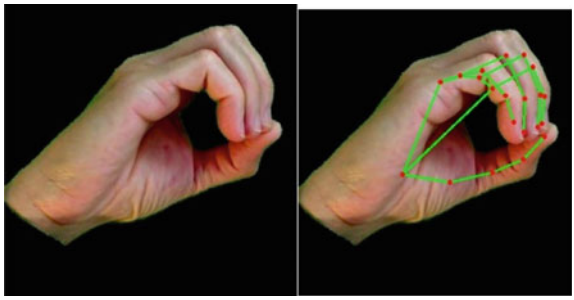

Fig. 1 Sample digit image 0 (raw image and image with media pipe)

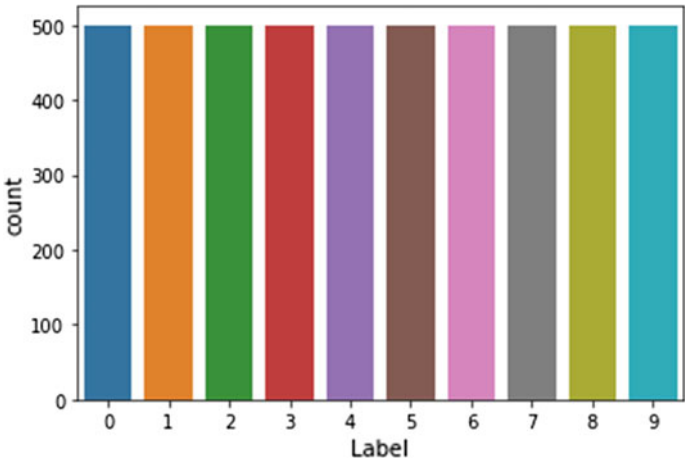

Fig. 2 Graphical representation of digit dataset

Convolution neural networks is a deep learning algorithm and an important classification algorithm that can take an input image, assign biases and weights to numerus objects in the input image [9, 10]. Compared to remaining classification algorithms, requirement of pre-processing is extensively lesser in Convolution Net. Like the pattern connectivity of neurons in the brain of human, so the architecture is analogues and inspired by the arrangement of Visual cortex in Convolution Net. The respective field is the visual field, a region where individual neurons respond (Table 1).

In the dataset, 75% of data is used for training and 25% of data used to test by the proposed model. The diagrammatic representation of convolution neural network algorithm which applied on digit dataset is displayed in Fig. 2. A convolution neural network algorithm will go through several steps for training the input data or input hand gesture images [9]. The algorithm has four layers which are rectified linear, convolution layer, pooling layer, unit layer, and fully connected layer. The first step is convolution layer. Any image is measured in the form of pixel values which is a matrix. In convolution layer, important features are extracted from an image which is in the form of matrix of pixels. They are operation of convolution phase which is

Table 1 Model summary for digit recognition using CNN algorithm

Layer	Shape of the Output	Param
CONV 2D (1^{ST})	(None,150,150,32)	2432
MAX POOLING 2D	(None, 75, 75, 32)	0
CONV 2D (2^{ND})	(None, 75, 75, 64)	18,496
MAX POOLING 2D	(None, 37, 37, 64)	0
CONV 2D (3^{RD})	(None, 37, 37, 96)	55,392
MAX POOLING 2D	(None, 18, 18, 96)	0
CONV 2D (3^{RD})	(None, 18, 18, 96)	83,040
MAX POOLING 2D	(None, 9, 9, 96)	0
FLATTEN	(None, 7776)	0
DENSE	(None, 512)	3,981,824
ACTIVATION	(None, 512)	0
DENSE	(None, 10)	5130

Model Sequential

focused on feature detectors, basically works as neural network's filters. The second phase is ReLU layer or rectified linear unit, which will reconnoitre the functions of linearity in the framework of convolution neural network. The third step in the algorithm is pooling [10] that can provide a method for sampling feature maps. The next step is Soft Max, a method of smoothen the combined feature map into a consecutive long vector and the last phase is Fully Connected layer, which is simply a feed forward network and it is an output of ending pooling.

4 Evaluation Metrics

The assessment metrics that are considered in this analysis are Accuracy, Precision, Recall and F1-score. The proposed measures play a significant role in the comparative analysis of different classification algorithms.

$$\text{Accuracy} = \frac{\text{TrPos} + \text{TrNeg}}{\text{TrPos} + \text{TrNeg} + \text{FalPos} + \text{FalNeg}}$$

$$\text{Precision} = \frac{\text{TrPos}}{\text{TrPos} + \text{FalPos}}$$

$$\text{Recall} = \frac{\text{TrPos}}{\text{TrPos} + \text{FalNeg}}$$

$$\text{F1} - \text{Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Table 2 Classification report for digit recognition

Digit	Pr	Re	f1-sc	Sup
0	1	1	1	146
1	1	1	1	123
2	0.98	0.99	0.99	126
3	0.99	0.98	0.99	116
4	1	0.97	0.99	118
5	1	0.98	0.99	118
6	1	0.92	0.96	123
7	0.89	0.95	0.92	141
8	0.92	0.92	0.92	106
9	0.94	0.99	0.96	136
Accuracy			0.97	1250

Where Pr = Precision, Re = Recall, f1-sc = F1-score, sup = Support

where

TrPos True (Correctly labelled) Positive
TrNeg True (Correctly labelled) Negative
FalPos False (In Correctly labelled) Positive
FalNeg False (In Correctly labelled) Negative

The experiments are done with the accuracy of 97.12 using the modules TensorFlow with keras in Python. The results are analysed using F1-score, Precision and Recall which are listed in the classification report of digit recognition in the Table 2. The output graph of confusion matrix is also analysed, and digit 0,1,4,5 and 6 hand gesture images are recognized correctly compared to other digits (Figs. 3, 4).

The correctly classified digits and incorrectly classified digits are displayed in Figs. 5 and 6. It is observed that the hand gesture image of digit 9 is classified as 5 and 7. And the hand gesture image of 7 is identified as 8 because of the hand positions.

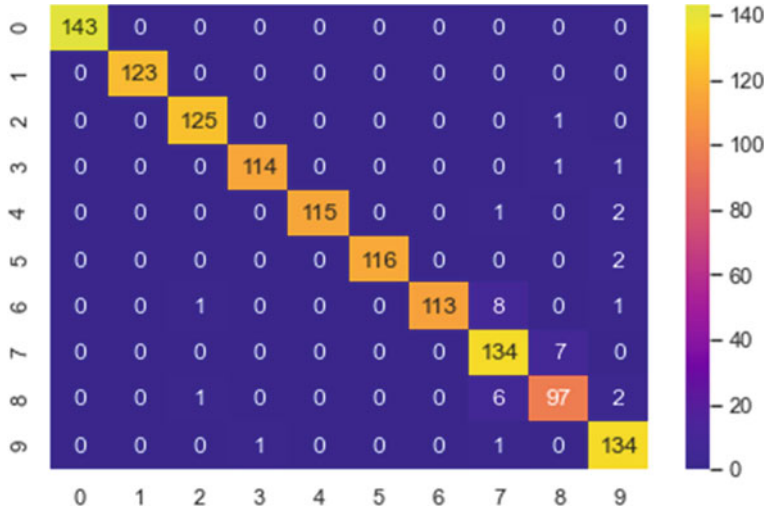

Fig. 3 Confusion matrix output graph for digit recognition

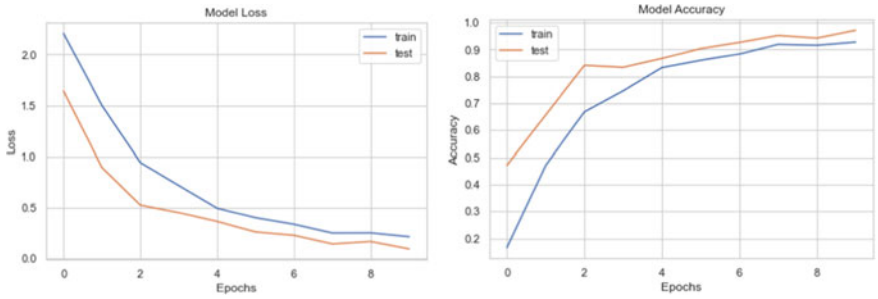

Fig. 4 Model loss and model accuracy output graph for digit recognition

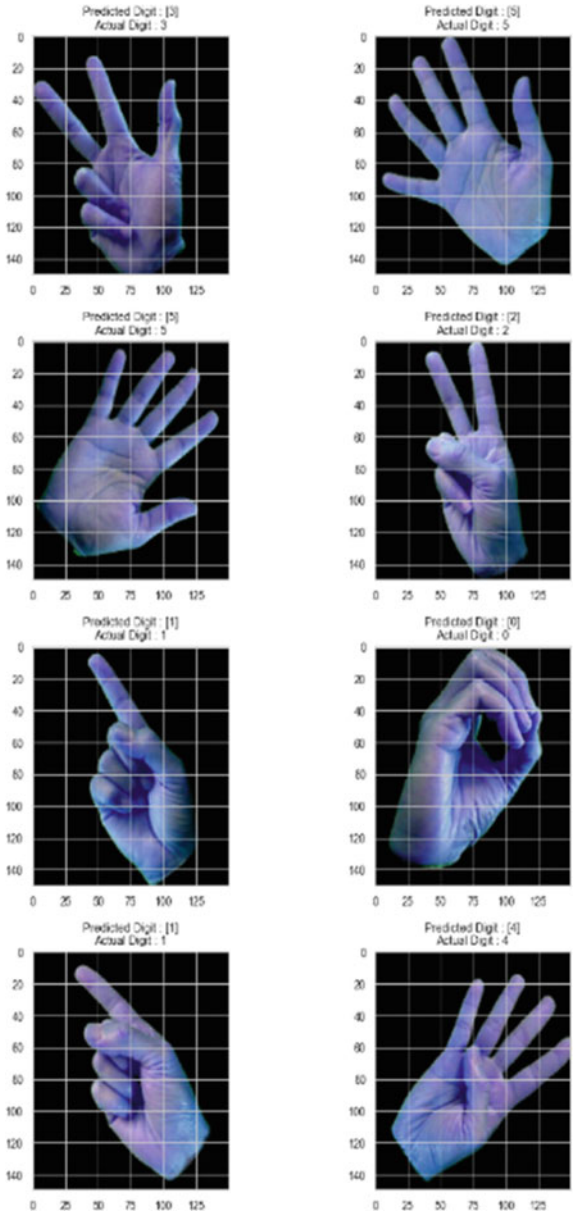

Fig. 5 Correctly classified digits

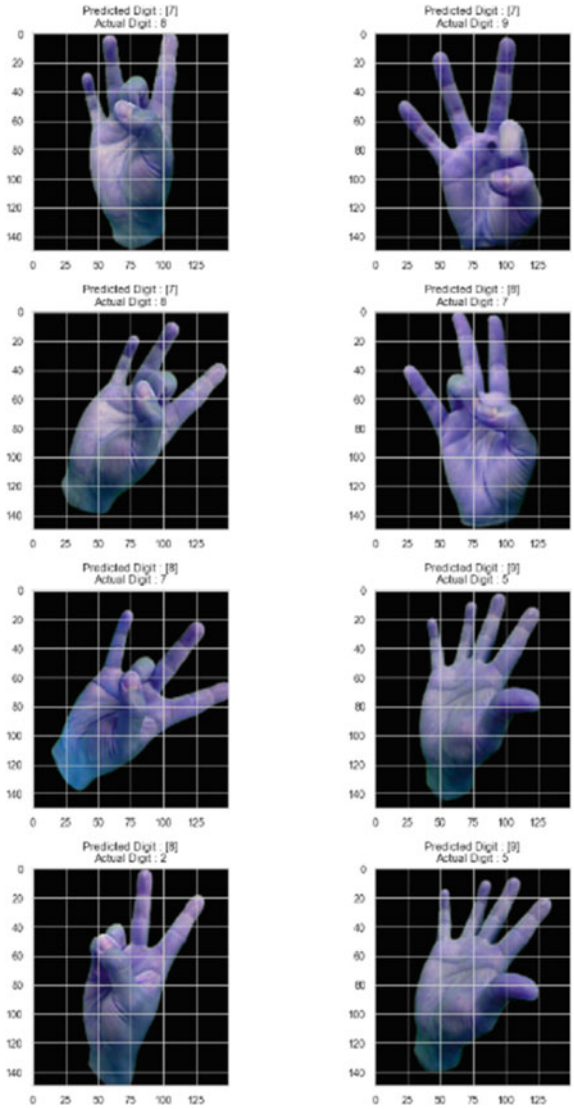

Fig. 6 Incorrectly classified digit

5 Conclusion

Digit recognition using hand gestures in sign language for deaf and muted people is a helpful way to fill the gap among the deaf muted person and common person. The manuscript presents the performance of convolutional neural network on digit hand gesture static images with Media Pipe. 97% of accuracy is obtained in the experiment analysis. In future, the partial voice of muted people will be added with dynamic video of sign language to recognize the sentence and emotion of that person

and the recurrent neural network with long short-term memory will be applied on hand gesture images to get more accuracy compared to convolution neural network.

References

1. Kahlon NK, Singh W (2023) Machine translation from text to sign language: a systematic review. Univ Access Inf Soc 22:1–35
2. Farooq U, Rahim MSM, Sabir N et al (2021) Advances in machine translation for sign language: approaches, limitations, and challenges. Neural Comput Appl 33:14357–14399
3. Halvardsson G, Peterson J, Soto-Valero C et al (2021) Interpretation of Swedish sign language using convolutional neural networks and transfer learning. SN Comput Sci 2(207):1–15
4. Rastgoo R, Kiani K, Escalera S (2022) Real-time isolated hand sign language recognition using deep networks and SVD. J Ambient Intell Human Comput 13:591–611
5. Sharma S, Singh S (2022) Recognition of Indian sign language (ISL) using deep learning model. Wireless Pers Commun 123:671–692
6. Adeyanju IA, Bello OO, Adegboye MA (2021) Machine learning methods for sign language recognition: a critical review and analysis. Intell Syst Appl 12:200056
7. Sharma P, Anand RS (2021) A comprehensive evaluation of deep models and optimizers for Indian sign language recognition. Graph Vis Comput 5:200032
8. Imran A, Razzaq A et al (2021) Dataset of Pakistan sign language and automatic recognition of hand configuration of Urdu alphabet through machine learning. Data Brief 36:107021
9. Yim J, Ju J, Jung H, Kim J (2015) Image classification using convolutional neural networks with multi-stage feature. In: Kim JH, Yang W, Jo J, Sincak P, Myung H (eds) Robot intelligence technology and applications 3. Advances in intelligent systems and computing, vol 345. Springer, Cham, pp 587–594
10. Suma J, Oguri S (2020) A multi-biometric iris recognition system using convolution neural network. i-manager's J Pattern Recognit 7(1):1–7