International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Multiple Deep CNN models for Indian Sign Language translation for Person with Verbal Impairment

Kujani.T*[1], DhilipKumar.V[2]

Submitted: 22/07/2022 **Accepted**: 25/09/2022

Abstract: Understanding the sign language is very useful for people with verbal and hearing impairment. Sign language is a category of nonverbal communication for people weakened by speech and listening capability. Automatic translation of sign gestures into text have gained more attention in recent years. In this research work, a deep CNN based approach has been proposed for detecting existing 35 signs of Indian Sign Language (ISL) alphabets into text in an efficient manner using hand kinematics. A Convolution Neural Network (CNN) with custom developed number of convolution layers with a suitable optimizer is applied. CNN handle issues of lighting conditions and most importantly it is efficient in tackling the problems under computer vision. It is considered for detecting features with required training without any manual pre-processing. The proposed approach has achieved 92.85% of accuracy with ISL dataset and accuracy of this method is compared with the transfer learning models such as Densenet201 and Resnet50.

Keywords: *Hand Gesture Recognition, segmentation, Indian Sign Language, convolutional neural networks.*

1. Introduction

In everyday life, people who have problem in verbal communication, face issues of social separation and miscommunication. Technology is quickly varying and enlightening the way the world functions which can help to resolve such issues. Researchers are working on their own way for building hardware and software that can assist the deaf people to communicate and understand. People use Sign language for communication purpose. The hand movements and three-dimensional spaces are used to convey the meaning. Sign Languages are generally different from spoken languages. To produce the comprehensive messages, spoken language makes use of rules but complex grammar governs the sign language. The method developed for sign language recognition possesses an easy, competent and precise mechanism to transform sign language into text or speech. The deep learning techniques provides great support to classify the different alphabets and identify the sign languages. Generally, the sign languages can be conveyed using hand, head and different body movements. Any country's Sign Language consists of set of alphabets and it is similar to the written or spoken language of that country. ISL combines gestures of both the hand for representing the digits and alphabets. ISL has many features including number signs, and can represent family relations. The proposed approach is to efficiently recognize the alphabets more accurately and in short time compared with exiting techniques.

This approach for recognising the sign language contains simple steps and implementation done using deep CNN for training the dataset and predicting the results accurately. Presently, CNNs present the results with state-of the-art for both having image-oriented tasks including object detection, image segmentation and classification and also for video-based tasks which focuses on recognition activity and gestures. The proposed work concentrates to improve classification accuracy as well as efficiency. The contributions of this proposed work are summarized below:

1. An architecture has been developed using deep CNN for attaining high performance to translate ISL sign language into text.
2. The developed model is trained by exploiting the ISL dataset for sign gesture.
3. Validation of trained model is performed on real time video and translation to text is obtained for the given hand sign gesture and the performance is measured.

2. Related Works

In the past years, many research works had been done on sign language translation with good hand-crafted features. Some approaches were based on Local binary pattern and histogram of oriented gradients. In recent years, researches were using deep convolution neural networks for efficient training with the available dataset. Approaches based on learning frame wise representations and Markov models are considered as complex in case of detecting the real time dynamic hand gestures. Continuous Sign Language Recognition [18] with both visual feature extraction and text information as a cross model was developed on three famous datasets. ASLNN [14],for hand posture recognition technique using ANN proposed hand segmentation technique in their work for detecting the sign gestures and showed accuracy of nearly 96.78%. Stacked time-based fusion layers [5] for the purpose of feature extraction and bi-directional RNN for learning purpose. The dependencies of short and long term simultaneously were developed with the deep temporal convolution layers were used instead of RNN for sign language translation [7]. Two architectures for detecting gestures and to classify the sign gestures for evaluation, Levenshtein was considered in single time activation [15], based on Ego Gesture and NVIDIA Dynamic Hand

[1] *Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India*
ORCID ID : https://orcid.org/ 0000-0001-9388-6856
[2]*Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India*
ORCID ID : https://orcid.org/0000-0003-2545-2073
** Corresponding Author Email:kujani@veltech.edu.in*

Gesture Datasets.

Hand Assessment for real time recognition of Indian and American Sign Language [12] was developed and the system was able to translate the recognized alphabets into speech by using label matching algorithm. A method to overcome the issues related to labelling the words during sign language translation, which could eliminate the initial pre-processing which occurs during the temporal segmentation [11] was developed. Kinect sensor data was applied to classify the static gestures of ASL, the system was trained to classify the three characters using CNN. For increasing the performance, utilized various classifiers in the temporal convolution layer and combined the predictions with the CTC decoder for better performance [3]. Glove based gesture recognition was proposed for translating the ASL [2]. To convert the ASL for finger spelling into text using the Aforge.NET dataset was demonstrated and the images taken from the webcam are compared with those in the database [16]. Presentation of different methods containing Adaptive Statistical Database which evolves and adapts the changes according to the dynamic input given through the webcam [6]. A relevant work was developed by applying CNN for classifying 20 Italian gestures of complete body by using a Microsoft Kinect on images of people performing the gestures and achieved 91.7% cross-validation accuracy [19].

ANN for classification of signs using error back propagation methods for recognizing the ISL was developed [13]. The model based on Inception v3 network type of CNN for recognizing the ASL sign language is demonstrated [1]. A model to use the ISL words commonly used by farmers was demonstrated using GoogleBet and BiLSTM classifier [23]. Classification of the ISL sign language using k-nearest and naive bayes classifier was developed [22]. The model for hand gestures by merging the layers of CNN with Recurrent Neural Network Long short-term memory (RNN-LSTM) for temporal feature extraction and to classify the sequence of frames was developed [17].

3. Materials and Methods

3.1 Dataset Collection

The dataset considered for implementation is taken from Kaggle, Indian Sign Language Dataset [20], which contains 35 directories for 1 to 9 and a to z alphabets. Fig 1. depicts the data available for each digit and alphabets in Kaggle Dataset for Indian Sign Language. The difference between the ASL dataset and ISL, is in ISL both hands are used for gesture and hence it is complex when compared to ASL.

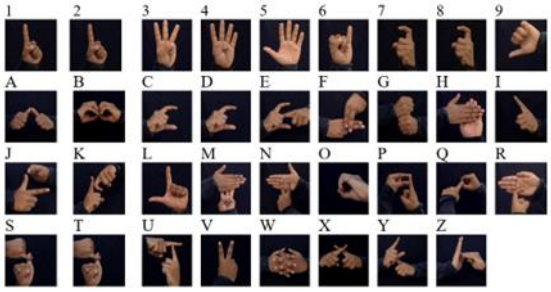


Fig 1. ISL Dataset Sample Images

3.2. Convolution Neural Network (CNN)

A Convolutional Neural Network is a distinct kind of multi-layer neural networks, especially developed for working on pixel images for identifying the visual patterns with less pre-processing steps. CNN is good to classify the images. CNN has the following layers: Convolution Layer followed by activation layer, ReLU Layer, Max Pooling Layer, Fully Connected Layer and dense layer. The convolution operation is major fundamental building block of a convolutional neural network which is performed by applying the sliding over the filter which is also called as kernel.

The convolutional layer contains a kernel/filter with a chosen size that can slide over the pixels multiplying and summing values, finally resulting with a simplified matrix. The motivation to use the CNN is the ability to learn the features and also provides the mechanism to optimize the loss function. The applied Classification function in our approach at the final layer is SoftMax which is a vector function where it takes many inputs and results many outputs as probability values. The general architecture of CNN is shown in Fig 2.

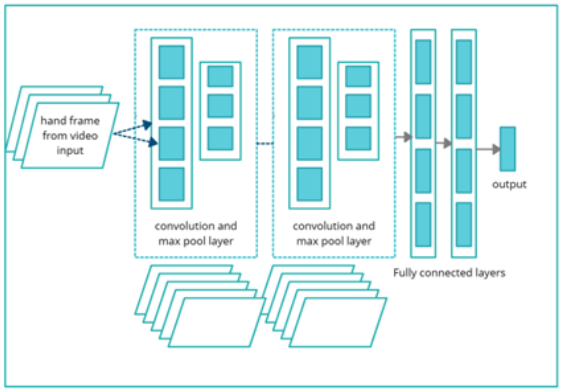


Fig 2. General Structure of CNN Architecture

The neural network architecture has been developed with the different models such as LeNet5, AlexNet, ResNet, Inception, DenseNet, MobileNet and VGGNet. Transfer Learning is a technique of machine learning in which the developed models are fed for training on available larger data sets and enable them to fit for custom dataset. The knowledge which was learned from the pretrained network are considered for classification of similar kind of task, and hence therefore reducing the burden of retraining the network. The short demanding time and data requirements are the advantages of these methods. VGG19, DenseNet are few examples of transfer learning networks which can help to achieve phenomenal accuracy on ImageNet dataset. This approach is compared with the DenseNet201 and Resnet50 model.

3.2.1 DenseNet201 Model

DenseNet201[10] model have 201 deep layers in it. This model can be applied when the accuracy is declining due to vanishing gradient which occurs due to longer path in between the input and final layer. In DenseNet, if the number of layers is L, there will be $\frac{L(L+1)}{2}$ direct connections. In DensetNet, the primary important things to be considered are the growth rate, Dense Blocks, Transition blocks and composite layers. The architecture of DenseNet201 is shown in Fig 3. Each layer output the number of feature maps which is the growth factor. Dense blocks possess the convolution layers, the transition layer aggregates the feature maps from the dense block and reduce its dimensions. Composite function includes Batch Normalization, RELU and convolution in sequence. The architecture of DenseNet201 structure is shown in Fig 3.

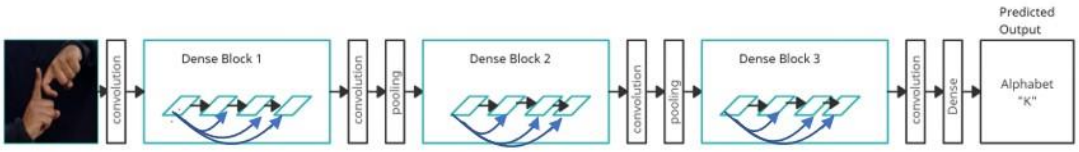


Fig 3. DenseNet201 Architecture

3.2.2 Resnet50 Model

Resnet50 [8] model is a subclass of CNN used to solve training and testing error that occurs when the number of layers increase. The main intuition in Resent is skip connection which allows the network to skip through layers when it feels that they may be less relevant in training. In Resnet there will be both convolution layer and also a direct connection which is called Identify connection. When Y=F(X)+X, the idea is to make the F(X)=0. Fig 4. shows the residual block. Therefore, the value of Y=X, which is the main logic of residual network, that is to add the actual input into the loss function and try to make the F(X)=0, so that we can have Y=X, which can increase the overall accuracy.

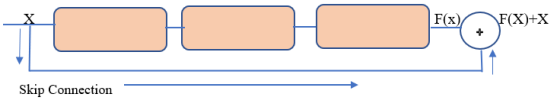


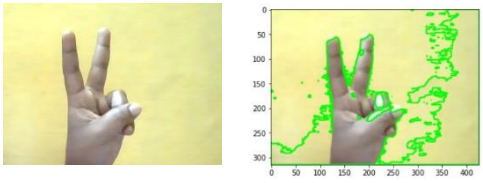
Fig 4. Residual Block

3.3 Hand Segmentation

This work adopts the hand segmentation based on threshold grey scale images, since it is fast as well as reliable and applied background subtraction in the image. Thresholding is the significant method in segmentation process done on grey scale images. Different threshold methods such as binary, Inverted Binary, Truncated, To Zero and To Zero Inverted are available. In our approach, we have employed binary threshold in which if the pixel intensity is greater than the given threshold, the value 255 will be set, else, it will be replaced with zero. Background subtraction technique is generally used for detection of moving objects and play an important role in object detection. The idea is to detect the active objects from the referred frame and current frame. The current frame will be fed continuously and the average of all the frames will be found. Finally, the absolute different between the frame will be found. In OpenCV, accumulatedWeighted() function is used to find the average of all the frames.

3.3.1 Contour Detection and Identifying the letters of ISL

Contour detection plays a main role used for identifying the letters or digits gestures. Contours are the curve connecting all continuous points. For getting better accuracy, threshold is obtained before contours. In this work, once the hand segment is obtained from the input video, contours are drawn. Since contours can be detected for any objects in the input, the largest hand contour alone is extracted. Using the properties of Contours, it is possible to detect the symbols of any sign language. Fig. 5 shows the results of finding the contours of the image. The steps to perform the contour detection is given in algorithm 1.



(a)input image (b) detected contours

Fig 5. Contour Detection

Algorithm 1: Hand Segmentation

Input: frame from video, background=none, accumulated_weight, threshold=25

Step 1: calculate the running average based on background value
Step 2: if background = none, then
update the new background value from the input frame value
Step 3: compute the weighted_average and update the background value
Step 4: measure the absdiff between background and current frame, diff.
Step 5: perform the binary threshold on the diff image
Step 6: fetch the contours from the frame
Step 7: if the contour length=0 then no contour identified, else the return the maximum contour length

4. Proposed Approach

The objective of this research is to develop a technique using CNN for real time sign language translation to text for verbal impaired people. In this work, ISL has been considered for translation in text using Deep CNN model for detecting hand gestures. The proposed system consists of three stages: 1) preparation of standard dataset of Sign Language; 2) Hand tracking and training the images using CNN, 3) Model Creation for prediction and evaluation of the predicted model. Fig 6. depicts the complete work flow of the proposed model.

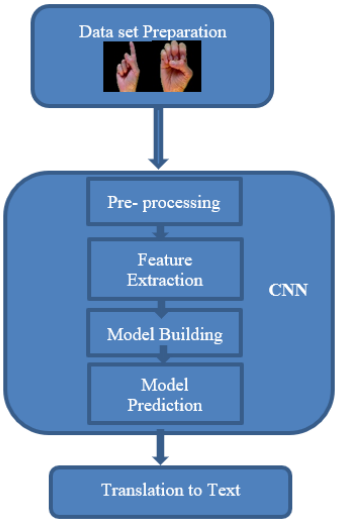


Fig 6. Flow of Proposed System

4.1 Proposed Model

The sequential model of CNN with input size 64,64 is created with four convolution layers, maxpool layers, one flatten layer followed by four dense layers. The last dense layer has 35 class labels for identifying the 35 symbols from ISL dataset. In each layer, the activation function, RELU is used and in the last layer, Softmax activation function is applied to categorize the classes. The summary of model developed is shown in Fig 7.

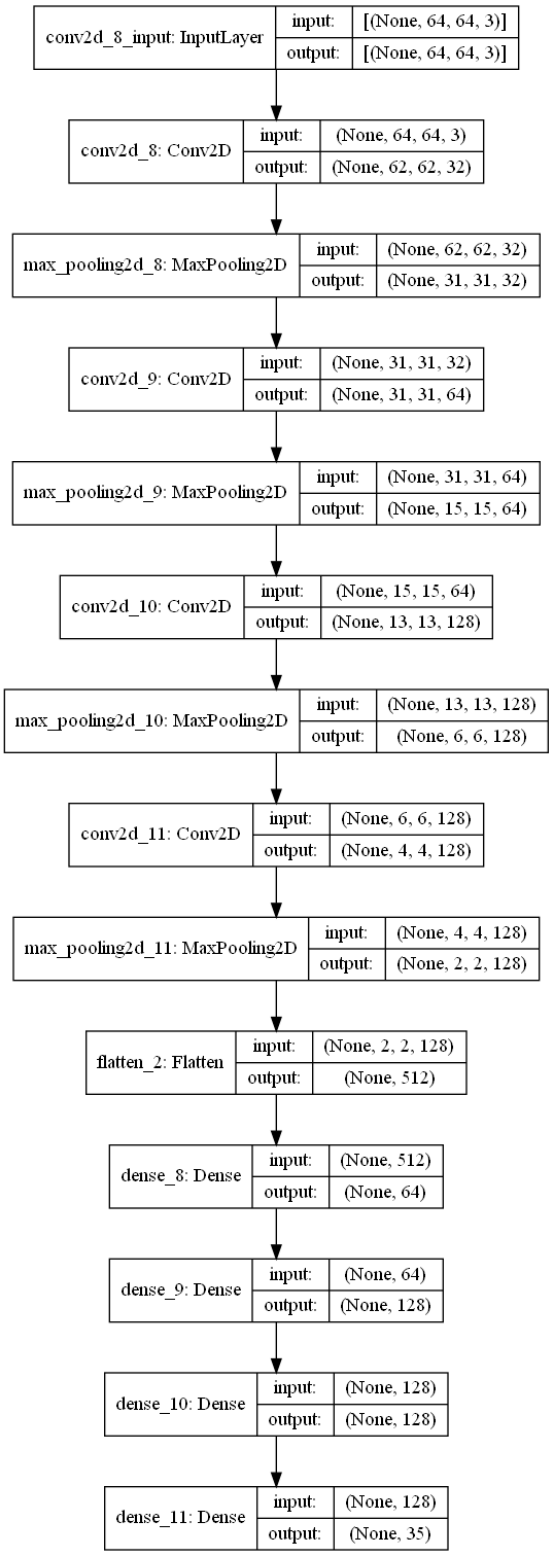


Fig 7. Proposed Model summary using CNN for ISL

4.1 Softmax Function Computation

In Softmax activation function, the vector of input values is considered for computation and converts them to probability values while preserving the rank order of the input values.

$$s_i = \frac{e^{X_i}}{\sum_n e^{X_i}}$$

where, Xi is the given vector values and si is the resulting index of soft max value.

where, Xi is the given vector values and si is the resulting index of soft max value.

For instance, $X = \begin{bmatrix} 1.2 \\ 2.6 \\ 0.5 \\ -1.9 \end{bmatrix}$ then the value of s_i is $s_i = \begin{bmatrix} 0.18 \\ 0.72 \\ 0.09 \\ 0.01 \end{bmatrix}$

cThe last layer our CNN model contains Softmax as its activation function, since we are trying multi class classification. The Softmax converts the last layer results in the form of probability distribution.

Further, the model was compiled using Adam and Stochastic Gradient Descent optimizers for minimizing the loss.

Compilation of the model is done using the following parameters:

1. learning_rate=0.001
2. loss function: categorical_crossentropy, since it is a multiclass classification model
3. optimizer: adam and stochastic gradient descent
4. metrics: accuracy→ the number of images classified correctly
5. Callbacks and checkpoints for early stopping and ReduceLROnPlateau are used for monitoring the loss and adjusting the learning rate.

where, Xi is the given vector values and si is the resulting index of soft max value.

For instance, $X = \begin{bmatrix} 1.2 \\ 2.6 \\ 0.5 \\ -1.9 \end{bmatrix}$ then the value of s_i is $s_i = \begin{bmatrix} 0.18 \\ 0.72 \\ 0.09 \\ 0.01 \end{bmatrix}$

The last layer our CNN model contains Softmax as its activation function, since we are trying multi class classification. The Softmax converts the last layer results in the form of probability distribution.
Further, the model was compiled using Adam and Stochastic Gradient Descent optimizers for minimizing the loss.
Compilation of the model is done using the following parameters:

- learning_rate=0.001
- loss function: categorical_crossentropy, since it is a multiclass classification model
- optimizer: adam and stochastic gradient descent
- metrics: accuracy is the number of images classified correctly
- Callbacks and checkpoints for early stopping and ReduceLROnPlateau are used for monitoring the loss and adjusting the learning rate.

4.2 Optimization of weights

Adaptive Moment Estimation optimizer is used to minimize the loss during the training and validation. Adam Optimizers is used since the greater number of parameters are used and it provides better accuracy as it combines both momentum and Root Mean Square Propagation (RMSP) and provides optimized results. Momentum helps to smoothening and RMSProp for changing the learning rate in efficient manner. The following is the steps in ADAM optimization algorithm implementation.

Algorithm 2: To compute weight and bias using ADAM optimizer:

Input: $Vd_w = 0, Vd_b = 0, Sd_w = 0 \; and \; Sd_b = 0$

where, $Vd_w \; and \; Vd_b \; are \; with \; respect \; to \; momentum \; and$

$Vd_w \; and \; Vd_b \; are \; with \; respect \; to \; RMSProp$

1. For iteration t, Compute
$\frac{\partial L}{\partial w} \; and \; \frac{\partial L}{\partial b} \; by \; using$ mini_$batch_size$

2. The following two steps are with respect to momentum

$$\text{(i) } Vd_w = \beta_1 \, Vd_w + (1-\beta)\frac{\partial L}{\partial w}$$

$$\text{(ii) } Vd_b = \beta_1 \, Vd_b + (1-\beta)\frac{\partial L}{\partial b}$$

3. The following two steps are with respect to RMSProp

$$\text{(i) } Sd_w = \beta_2 \, Sd_w + (1-\beta)\left(\frac{\partial L}{\partial w}\right)^2$$

$$\text{(ii) } Sd_b = \beta_2 \, Sd_b + (1-\beta)\left(\frac{\partial L}{\partial b}\right)^2$$

4. The new updated weight is computed as

$$W_t = W_{t-1} - \frac{\eta * Vd_w}{\sqrt{Sd_w + \varepsilon}}$$

where, η is the learning rate and ε is small positive integer.

5. The new updated bias is computed as

$$b_t = b_{t-1} - \frac{\eta * Vd_b}{\sqrt{Sd_b + \varepsilon}}$$

5. Experimental Setup and Results

A public dataset for Indian Sign Language [20] , is used for evaluating and comparing our method with existing methods. For training and testing, nearly 28035 and 13965 samples were considered. The metrics to evaluate the model are classification accuracy and logarithmic loss. Experiments were conducted on ISL dataset using deep CNN Model proposed for 10 epochs with 877 steps per epoch and achieved the accuracy of 92.8%. For comparison purpose, the model performance is shown with transfer learning techniques DenseNet201 and Resnet50.

The Table 1. depicts the parameters and recognition rate. Table 2. shows the performance comparison of the proposed work with other works.

Table 1. Parameters And Recognition Rate

learning_rate= 0.001
function: categorical_crossentropy
optimizer: adam and stochastic gradient descent
metrics: classification accuracy

Model	Total Samples	Train Size	Test Size	Epochs	Steps per Epoch	Recognition Rate%
Proposed Model	42000	28035	13965	10	877	92.85
Densenet201	42000	28035	13965	10	1753	99.49
Resnet50	42000	28035	13965	10	1753	98.61

Table 2. Comparison of Performance with other methods

Methods	Training Accuracy
Dynamic temporal warping [9]	.540
Histogram of 3D joints [21]	.789
Hidden Markov model [4]	.900
DGSLR using ANN [14]	.967
DenseNet201 Model	.994
Resent50 Model	.986
Proposed method using CNN on ISL dataset	.928

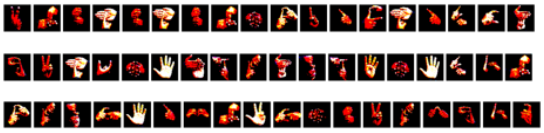


Fig 8. Results from Training the ISL data for hand gestures

Fig 8. shows the plots obtained from different training samples and the sample predicted images taken from the ISL dataset. The Fig 9. shows the samples of correctly predicted output of alphabets after training the model from ISL Dataset in video input.

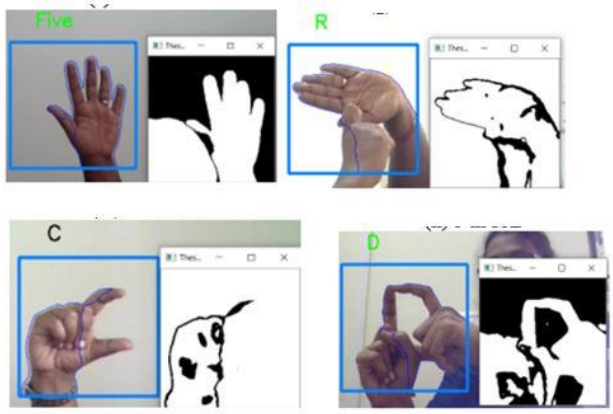


Fig 9. Sample of Correctly Predicted Output from Video frames

5.1 Performance Metrics

The classification metrics considered for evaluating the performance of the model are Accuracy, Recall, Precision and F1-Score. Accuracy is the common metric used in classification problems which describes the total number of correct predictions with respect to total number of predictions made in dataset. Confusion matrix helps to find the classes correctly and incorrectly predicted. The possible outcomes of classification are True Positive (TP), False Negative (FN), True Negative (TN) and False Positive (FP). True Positive (TP) is the one in which the actual and predicted value is positive. False Negative (FN) is one in which actual is positive and predicted is negative. True Negative (TN) is one in which the actual is negative and predicted also negative. False Positive (FP) is one in which the actual is negative and

predicted is positive. The equation for the metrics is represented below equation (1) to (4).

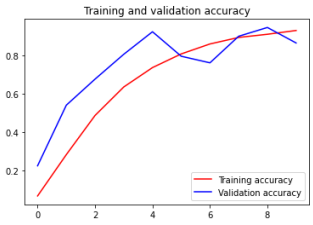
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \ldots\ldots\ldots\ldots (1)$$

$$Precision = \frac{TP}{TP + FP} \ldots\ldots\ldots\ldots (2)$$

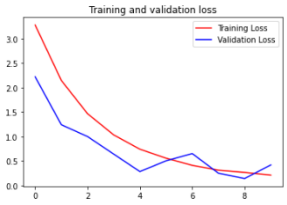
$$Recall = \frac{TP}{TP + FN} \ldots\ldots\ldots\ldots (3)$$

$$Fscore = \frac{2}{\frac{1}{Recall} + \frac{2}{Precision}} \ldots\ldots\ldots\ldots (4)$$

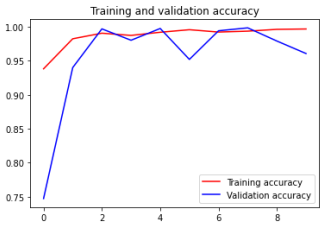
Fig 10. demonstrates the accuracy and loss results obtained by executing the model using proposed CNN and DenseNet201 model for the specified number of epochs. The accuracy improves gradually on increasing the number of epochs. In this work, training was conducted with 10 epochs for ISL dataset.



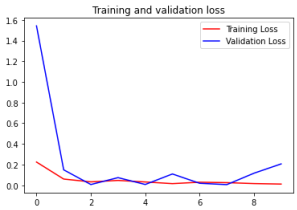
(a) ISL Training Accuracy using CNN



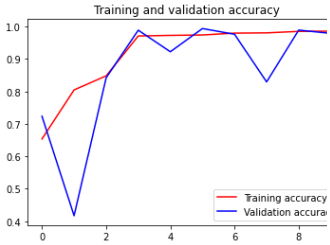
(b) ISL Training Loss using CNN



(c) ISL Training Loss using DenseNet201



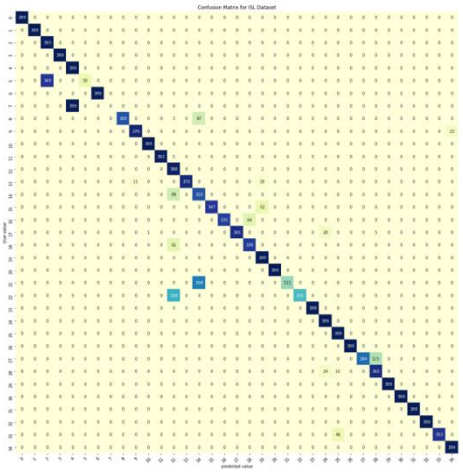
(d) ISL Training Accuracy using DenseNet201



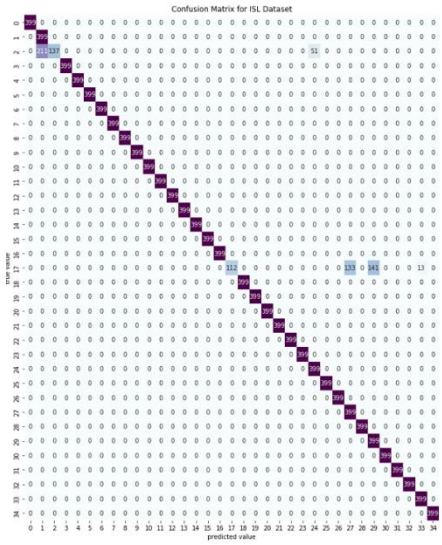
(e) ISL Training Loss using Resnet50

(f) ISL Training Loss using Resnet50

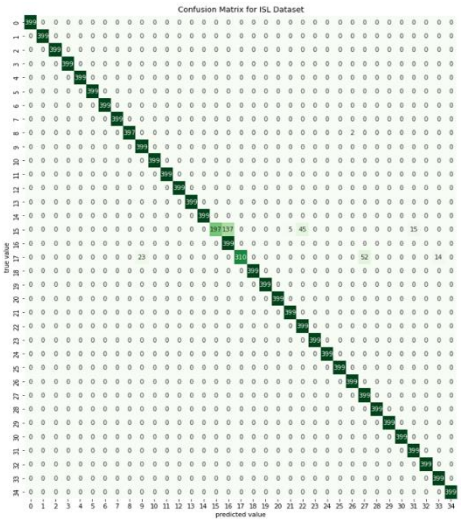
Fig 10. Training Loss and Accuracy for ISL Dataset using proposed CNN, DenseNet201 and Resnet50



(a) Confusion Matrix using proposed CNN



(b) Confusion Matrix using DenseNet201



(c) Confusion Matrix using Resnet50

Fig 11. Confusion Matrix for ISL using proposed CNN, Densenet201, and Resnet50

Fig 11. shows the confusion matrix for the 35 characters from ISL dataset using proposed a) CNN, b) DenseNet201 and c) Resnet50 model. From the confusion matrix of validation data, it is observed for the proposed CNN obtains the weighted average of .89 for precision, .86 for recall and .85 for f1-score. In DenseNet201, weighted average of Precision is .97, recall is .96 and f1-score is .95. In Resnet50, weighted average of Precision is

.98, recall is .98 and f1-score is .98. Fig 12. depicts the calculated values of performance metrics including the precision and recall for all symbols present in the ISL dataset after training the model.

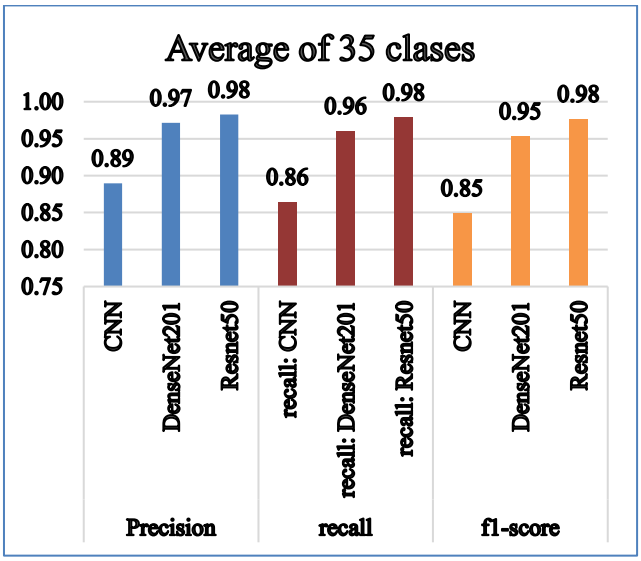


Fig 12. Performance metrics for proposed CNN, DenseNet201, Resnet50

5. Conclusion and future enhancement

The proposed methodology demonstrates the work done for translating the Indian Sign Language into text which can help the people with verbal impairments. The first part of the work was to collect the data from the standard dataset and make it suitable for our implementation. As the second part of work, training was done using deep CNN and prepared the model after the preliminary pre-processing procedures. Next, the developed model is loaded and the results are predicted for the given input. The experiment was conducted in a simpler manner and the accuracy was about 92.8% for Indian Sign Language using the proposed CNN model. In ISL dataset, approximately 42,000 images were used for implementation. The research work can be enhanced for other countries gestures and forming words as well as translation to speech along with embedding in mobile and IOT technologies.

Acknowledgement

We are thankful to the open-source resources available on Internet for giving more insight to carry out the work.

Authors' contributions

All authors have contributed to the work described in this paper and approved the final manuscript.

References

[1] 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), 2018, pp. 1-6, doi: 10.1109/ICSCET.2018.8537248.

[2] Abhishek, K. S., Qubeley, L. C. F., & Ho, D. (2016, August). Glove-based hand gesture recognition sign language translator using capacitive touch sensor. In 2016 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC) (pp. 334-337). IEEE.

[3] Beena, M. V., Namboodiri, M. A., & Dean, P. G. (2017). Automatic sign language finger spelling using convolution neural network: analysis. Int J Pure Appl Math, 117(20), 9-15.

[4] Caon, M., Yue, Y., Tscherrig, J., Mugellini, E., & Khaled, O. A. (2011, October). Context-aware 3d gesture interaction based on multiple kinects. In The First International Conference on Ambient Computing, Applications, Services and Technologies (pp. 7-12).

[5] Philip, A. M., and D. S. . Hemalatha. "Identifying Arrhythmias Based on ECG Classification Using Enhanced-PCA and Enhanced-SVM Methods". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 5, May 2022, pp. 01-12, doi:10.17762/ijritcc.v10i5.5542.

[6] Cui, R., Liu, H., & Zhang, C. (2019). A deep neural framework for continuous sign language recognition by iterative training. IEEE Transactions on Multimedia, 21(7), 1880-1891.

[7] Glenn, C. M., Mandloi, D., Sarella, K., & Lonon, M. (2005, June). An image processing technique for the translation of ASL finger-spelling to digital audio or text. In Instructional Technology and Education of the deaf Symposium, Rochester, NY (pp. 1-7).

[8] Guo, D., Wang, S., Tian, Q., & Wang, M. (2019, August). Dense Temporal Convolution Network for Sign Language Translation. In IJCAI (pp. 744-750).

[9] Gill, D. R. . (2022). A Study of Framework of Behavioural Driven Development: Methodologies, Advantages, and Challenges. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(2), 09–12. https://doi.org/10.17762/ijfrcsce.v8i2.2068

[10] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[11] Hossny, M., Filippidis, D., Abdelrahman, W., Zhou, H., Fielding, M., Mullins, J., ... & Nahavandi, S. (2012, January). Low cost multimodal facial recognition via kinect sensors. In LWC 2012: Potent land force for a joint maritime strategy: Proceedings of the 2012 Land Warfare Conference (pp. 77-86). Commonwealth of Australia.

[12] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[13] Modiya, P., & Vahora, S. (2022). Brain Tumor Detection Using Transfer Learning with Dimensionality Reduction Method. International Journal of Intelligent Systems and Applications in Engineering, 10(2), 201–206. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/1310

[14] Huang, J., Zhou, W., Zhang, Q., Li, H., & Li, W. (2018, April). Video-based sign language recognition without temporal segmentation. In Thirty-Second AAAI Conference on Artificial Intelligence.

[15] Kakoty, N. M., & Sharma, M. D. (2018). Recognition of sign language alphabets and numbers based on hand kinematics using a data glove. Procedia Computer Science, 133, 55-62.

[16] Kishore, P. V. V., & Kumar, P. R. (2012). Segment, track, extract, recognize and convert sign language videos to voice/text. International Journal of Advanced Computer Science and Applications, 3(6).

[17] Kolivand, H., Joudaki, S., Sunar, M. S., & Tully, D. (2021). A new framework for sign language alphabet hand posture recognition using geometrical features through artificial neural network (part 1). Neural Computing and Applications, 33(10), 4945-4963.

[18] Kopuklu, O., Gunduz, A., Kose, N., & Rigoll, G. (2019, May). Real-time hand gesture detection and classification using convolutional neural networks. In 2019 14th IEEE International

Conference on Automatic Face & Gesture Recognition (FG 2019) (pp. 1-8). IEEE.

[19] Modi, K., & More, A. (2013). Translation of Sign Language Finger-Spelling to Text using Image Processing. International Journal of Computer Applications, 77(11).

[20] Obaid, F., Babadi, A., & Yoosofan, A. (2020). Hand gesture recognition in video sequences using deep convolutional and recurrent neural networks. Applied Computer Systems, 25(1), 57-61.

[21] Papastratis, Ilias, et al. "Continuous sign language recognition through cross-modal alignment of video and text embeddings in a joint-latent space." IEEE Access 8 (2020): 91170-91180.

[22] Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2014, September). Sign language recognition using convolutional neural networks. In European Conference on Computer Vision (pp. 572-578). Springer, Cham.

[23] Prathum Arikeri. 2021 https://www.kaggle.com/prathumarikeri/indian-sign-language-isl

[24] Linda R. Musser. (2020). Older Engineering Books are Open Educational Resources. Journal of Online Engineering Education, 11(2), 08–10. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/41

[25] Rafibakhsh, N., Gong, J., Siddiqui, M. K., Gordon, C., & Lee, H. F. (2012). Analysis of xbox kinect sensor data for use on construction sites: depth accuracy and sensor interference assessment. In Construction Research Congress 2012: Construction Challenges in a Flat World (pp. 848-857).

[26] Sahoo, A. K. (2021, June). Indian Sign Language Recognition Using Machine Learning Techniques. In Macromolecular Symposia (Vol. 397, No. 1, p. 2000241).

[27] Venugopalan, A., & Reghunadhan, R. (2021). Applying deep neural networks for the automatic recognition of sign language words: A communication aid to deaf agriculturists. Expert Systems with Applications, 185, 115601