A framework to recognize the sign language system for deaf and dumb using mining techniques

Lokavarapu V. Srinivas, Chitri Raminaidu, Devareddi Ravibabu, Shiva Shankar Reddy
Department of Computer Science and Engineering, Sagi RamaKrishanam Raju Engineering College, Bhimavaram, India

Article Info

ABSTRACT

One way of communicating with the deaf is to speak sign language. The chief barrier to little Indian sign language (ISL) research was the language diversity and variations in place. It is essential to learn sign language to communicate with them. Most learning takes place in peer groups. There are very few materials available for teaching signs. Thus, signing is very challenging to learn. Fingerspelling is the first step in sign learning and is used whenever there is no appropriate sign or if the signatory is unfamiliar with it. Sign language learning tools currently available use expensive external sensors. Through this project, we will take this field further by collecting a dataset and extracting functionally helpful information used in several supervised learning methods. Our current work presents four validated fold cross results for multiple approaches. The difference from the previous work is that we used different figures for our validation set than the training set in four-fold cross-validations.

Corresponding Author:

Lokavarapu V. Srinivas
Department of Computer Science and Engineering, Sagi RamaKrishnam Raju Engineering College
West Godavari District, Chinnamiram, Bhimavaram, Andhra Pradesh 534204, India
Email: srinivas.srkrcse@gmail.com

1. INTRODUCTION

Nelson Mandela quoted that speak to a man in his language, which goes to his head. Speak to him in his speechnear his heart, and language is undoubtedly essential for human interaction [1]. It is a medium used by humans to communicate, express and understand real-world concepts. It is so ingrained in our daily routine that we often consider it self-evident and fail to recognize its importance. Sadly, people with impaired listening are usually forgotten and left out in the fast-changing society. Sign language (SL) still has no meaning when sent to non-signal language users, although it is a means of communication for deaf people. An ideal tool is used to communicate their thoughts to people with hearing impairments and a perfect interpretation of what the latter means for non-SL users. Many countries have their standards and sign gestures. An alphabet, for example, in Korean, does not mean the same as in Indian sign language (ISL) [2]. While this emphasizes diversity, the complexity of SLs is also highlighted. The actions of profound learning must be well versed to obtain decent accuracy. American sign language (ASL) is used in our proposed system to create our datasets. Identification of sign gestures is performed with either of the two methods. First, the signer uses a pair of data gloves during hand movements [3]. Secondly, a vision-based approachis further classified as dynamic and static recognition [4]. A static address is the 2D representation of gestures, while a dynamic is the live capture of gestures in realtime. And while it is more than 90% precise, it is uncomfortable to wear gloves and cannot be used in rainy weather. They are not easy to carry because they also need computers: i) moreover, half of the "deaf" population in the United States is above the age of 65, although the total number of "deaf" persons in the country is estimated at 600,000; ii) in India, at least 50

lakh youngsters are affected by deafness, according to a World Health Organization (WHO) assessment; iii) it is estimated that around 10,27,835 persons worldwide have some hearing or speech disability; this number includes about 5,45,179 men and 482,656 women. At least 2,66,586 (1,51,170 males and 1,15,416 women) have speech impediment; iv) over a quarter of those with hearing loss have it due to exposure to loud noise, making it the most prevalent cause of acquired hearing loss; and v) rehabilitation for 'disabling' hearing loss is necessary for 5% of the global population, or 430 million people (432 million adults and 34 million children). By 2050, about 700 million people, or one in ten, will have severely impaired hearing [5].

2.   RELATED WORK

Wang *et al.* [6] proposed a multidimensional hidden Markov models (HMM) based system for ASL recognition. They use cyberglove and flock of birds motion trackers to interpret ASL gestures. ASL has 26 alphabets and 36 basic handshapes. After being trained with 8 samples, the system is 95% accurate in its identifications on average. According to the evaluation results, the proposed method facilitates rapid training and online learning of new gestures. Atwood *et al.* [7] showed that machine learning (ML) could be used to develop interactive educational tools and assist deaf people with communication. They utilized a single hidden layer neural networks (NN) and a principal component analysis (PCA) model. An analysis of the principal components and NN is constructed using the data collected from three subjects. The best NN exhibited 95.8% accuracy compared to 96.1% for the PCA model. Chuan *et al.* [8] developed an ASL recognition system using an affordable and compact 3D motion sensor. The experiment showed that the k-nearest neighbour and support vector machine (SVM) were the best methods to classify ASL's 26 English alphabet letters.

Rahman *et al.* [9] developed a new model for existing ASL classification algorithms to perform better. A convolutional neural network (CNN) model was given pictures of the letters and numbers after preprocessing them. This approach's effectiveness was assessed using four ASL datasets readily accessible to the public. The computer model proposed by the researchers significantly increases all ASL signs' recognition accuracy (89%). The model predicts all the signs with 100% accuracy. Starner and Pentland [10] proposed using hidden Markov processes to recognize hand gestures in ASL without explicitly modelling the fingers. The first experiment showed that 99% of the words were correctly spelt. A word accuracy of 92% was achieved with the second experiment, which tracked hands without gloves. Bantupalli and Xie [11] devised a technique for identifying signs that uses an ensemble of two models to determine the motions used in SL. To train the algorithm to detect gestures, they employed a custom-recorded ASL dataset based on an existing dataset. They looked at two different classification strategies, namely the pooling layer and the SoftMax layer, for predictions of the outcomes. The SoftMax layer delivered superior results as a consequence of its unique characteristics. Using the media pipe hands API that Google developed, Chakraborty *et al.* [12] have categorized the English alphabet as it is represented by various hand gestures in the ISL. Utilizing this application programming interface (API) will allow you to determine the x, y, and z coordinates in three-dimensional space for each of the 21 landmarks on each hand. They discovered that by utilizing the media pipe API, they could accurately forecast the ASL and various international SLs. SVM, random forest (RF), k-nearest neighbor (KNN), and decision tree (DT) have been compared to each other in terms of accuracy, with SVM having the most significant accuracy percentage at 99%.

Priya *et al.* [13] discuss some methods (SVM, KNN, logistic regression and CNN) that can be used to implement an approach to help communicate a non-signer with a signer much easier. They include the complete grammar of the ASL, consisting of 26 letters and 10 digits. Experimental results were promising, with an accuracy of 80.30% for SVM and 93.81% for deep neural network (DNN). The kernel-based approaches for identifying ASL were explored by Moghaddam *et al.* [14]. The experiment used 700 Persian alphabet signs signed by 35 people. The initial phase is picture scaling, thenconverting to grayscale and detecting a hand. The kernel discriminant analysis (KDA) and the kernel principal component analysis (KPCA) approaches were used for feature extraction. According to the experiments and test results, the KPCA-NN model had a maximum accuracy of 95.91%.

CNN are the basis of a model that Goswami and Javaji [15] propose for use in the identification and categorization of hand gestures. The collection comprises 26 hand motions, which may be mapped to the English alphabet from A to Z. A standard dataset known as hand gesture recognition, which can be found on the Kaggle website, has been taken into consideration in this study. An approach for deep learning (DL) based on CNN can automatically learn and extract features to categorize each motion. The suggested model presents the findings. 99% of the time, the results are correct. Shankar *et al.* [16] proposed object identification of computer vision that sees the highest use overall. It is a method for locating and shaping real-world things such as furnishings and artwork. Even though there are manydetection techniques, their accuracy and efficiency levels are inadequate. The YOLOv3 and YOLOv4 object recognition algorithms are used in this study, using DL techniques to identify the items.

Barbhuiya *et el.* [17] applied DL-based CNNs to model static signs in SL recognition robustly. Modified pre-trained AlexNet and modified pre-trained VGG16-based architectures are used to extract features. A multiclass SVM classifier follows this. The results are analyzed using the characteristics of the various layers to determine which ones provide the most outstanding recognition performance. The suggested system has 99.82% recognition accuracy, better than state-of-the-art approaches. Rao *et al.* [18] provide a CNN-based method for identifying ISL gestures. Videos taken using the front camera of a mobile device may be processed using this technology. A dataset including information on 200 ISL signs is compiled by hand. Three distinct datasets are used during the training process of CNN. As input for the first batch is provided with a dataset consisting of a single set alone. The second batch consists of two sets of training data, while the third one has three such groups. This CNN model has an overall recognition rate that is averaged out to be 92.88%.

Gupta *et al.* [19] worked on denoising a picture necessary to retrieve information from an image without making any mistakes. Denoising and improving picture quality utilizing wavelets and evolutionary computing. A novel SGO and APSO may coordinate an adaptive thresholding-based wavelet denoising technique.Using contour detection and a fuzzy c-means algorithm, Mariappan and Gomathi [20] built a real-time SL recognition system that could be carried about as a portable device. Face left and right hands, as well as body contours, may all be identified using outlines. The software they developed was tested using a dataset that included videos filmed by 10 different signers performing a variety of phrases and sentences. A level of accuracy equal to 75% was reached by using this approach.

Chong and Lee [21] developed anSL recognition system based on the leap motion controller (LMC). It was designed to recognize ASL, consisting of 26 letters and 10 digits. The recognition rates of 26 letters and 10 numbers for the SVM and DNN, respectively, were 80.30% and 93.81%. IP-supervised ML and deep understanding would be used in creating a fingerspelling alphabet recognition system for SL [22]. Grayscale images may be coloured by a technique known as "picture colourization", "colourizing" a picture involves taking a monochrome or "grayscale" image as input and producing a full-colour image in red, green, and blue (RGB) format. Image colourization is also often referred to as colouring a grayscale image [23]. A technique based on depth contrast features and per-pixel classification is first used to produce a segmented hand configuration. After that, a hierarchical mode-seeking algorithm is created and implemented to localize hand joint locations while considering kinematic restrictions [24]. Add 2D IP for feature extraction and a DT-based NN to texture-based face emotion identification [25].

Shotton *et al.* [26] use a representation of the body that is intermediate between the parts and the whole and is built in such a way that a correct categorization of the components per pixel will locate the joints of the body. On the other hand, the second strategy involves going backwards in time and immediately regressing the locations of the body's joints. Image sharpening produces a clear, high-resolution image using a fuzzy photo as input. The process of sharpening blurry images has significantly influencedvarious sectors, including astronomy, forensic science, and medical imaging [27]. Anand *et al.* [28] used ISL recognition system (ISLR) for the deaf and hard of hearing and used hand gesture photographs to ISL. The suggested ISLR system is a pattern recognition approach with two major modules: feature extraction and classification. A combination of discrete wavelet transform (DWT)-based feature extraction and closest neighbour classifier is used to identify SL. The categorization is accomplished using the time-honoured approach, which considers factors such as colour, shape, texture, petals, and sepals, among other characteristics. The methodologies of DL have led to significant advancements in picture analysis and categorization [29].

Devareddi and Srikrishna [30] worked on the vast majority of search engines still using the more conventional text-based algorithms and depending on captions and metadata to locate images. Over the last two decades, content-based image retrieval (CBIR), image classification, and image analysis have increased the number of retrieval operations. The field assessment is carried out with deaf persons using sophisticated communications technology to connect with hearing people in various settings [31]. Elmezain *et al.* [32] have done their work based on the hidden Markov model and offer a real-time system that automatically detects Arabic digits (0-9) in both isolated and continuous gestures (HMM). Many alternative HMM topologies with varying states may be used to handle solitary gestures, such as ergodic, left-right (LR), and left-right banded (LRB). Shankar *et al.* [33] was done their work to retrieve the images by using genetic algorithm (GA) model.

3. METHOD

One of the approaches to machine learning (ML) is supervised, which involves training a model using input and anticipated output data. It is determined using several machine learning techniques. Keras is now communicating with Tensorflow to build the model. Compiling a model requires the creation of a loss function and an optimal procedure. model. Compile (loss='name of loss function, optimizer=name of

optimizer'). The loss function measures the model's predictions. Figure 1 shows the data for training the model for the letters A to Z and 1 to 9.

The second set of data is loaded during this stage. The model never saw this data set, so thatit will be verified with reasonable accuracy. The model can be saved to 27 by the model name when fully trained. Save ("name of the file.h5") after the model has been finished training. The model is stored and then utilized in the actual world. This stage is known as the "model assessment" process. The model can therefore be used for the evaluation of new data. Figure 2 shows the data which is used for testing the model.
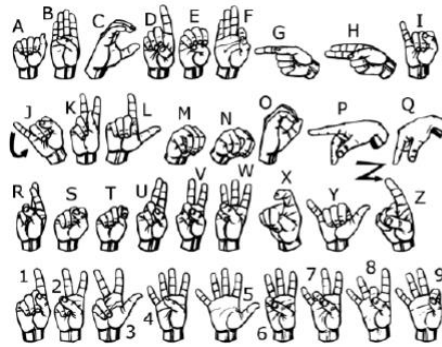
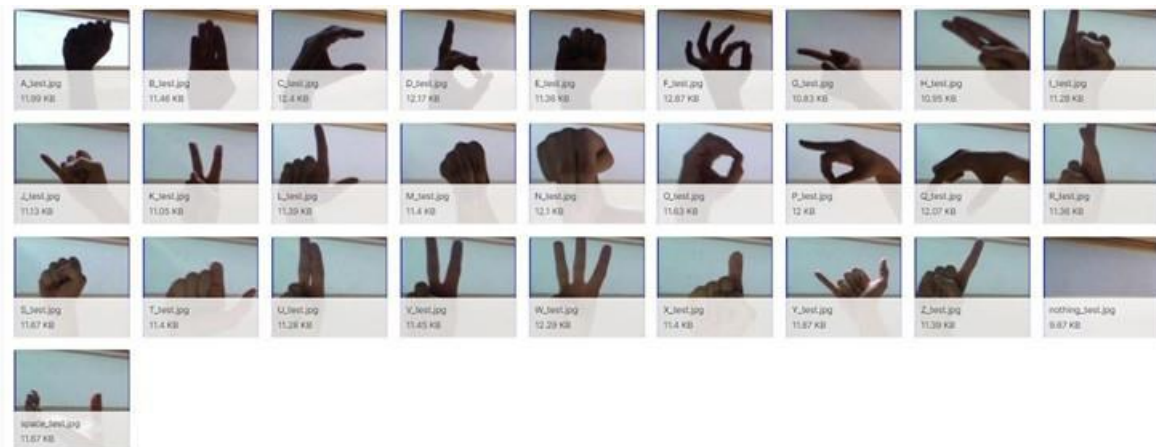Figure 1. Data used for training the model (for the letter–'A')

Figure 2. Data used for testing the model

3.1. Dataset

The SL recognition dataset consists of different images of the hands. The dataset consists of 768 instances collected from Kaggle data repository. The dataset is divided into 2 segments-training data and testing data. Each dataset is elaborately explained in the paragraphs below. The training dataset contains images, each of size 34 KB having 784 pixels in a grid structure with a 28*28 distribution. The photos include gestures which correspond to the English alphabet in multiple ways. Each alphabet has around 600 corresponding images to train the model. The test dataset is built by randomly selecting several images from the test dataset. It will check the accuracy of the model. Also, photos were added to the dataset to check if the model was appropriately trained. Table 1 lists every feature of the dataset.

Table 1. Summary of the dataset

Attributes	Description	Values
Hand shapes	The different shapes of hands for the other alphabet.	A-Z
Palm orientation	The palm is kept in the fixed position in front of the camera.	0,1
Location	The place where the system is placed	0,1
Movement	Camera-facing hand movement.	0,1
Classification	Target variables.	Letter from A-Z

3.2. Algorithms used

3.2.1. Glove-based method

The user is the starting point and the final destination of the system currently in use for identifying hand movements. There are five stages: input, acquisition, preprocessing, classification, and results. Display an entire architecture is shown in Figure 3. Figure 1 depicts the user with right-hand input devices. A 5DT data glove 5 Ultra on the back of the user's hand identifies hand orientation, and a 3-axis accelerometer on the front gathers finger flexion data. analog-to-digital converter (ADC) turns accelerometer data into 14-bit integers and transmits them to the computer for preprocessing. The data glove and accelerometer data are calibrated and standardized at that stage. After that, it is stored so that it may be utilized as a parameter in the subsequent categorization procedure. Here, artificial intelligence (AI) methods like artificial neural network (ANN) are employed to identify motion using the information learned from the preceding steps' data [12]. The classification results are given to users through a trained network interface.
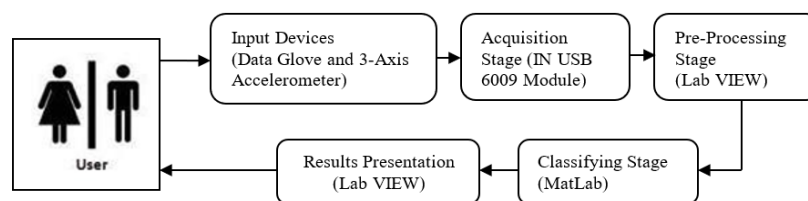

Figure 3. System architecture

3.2.2. Classification

For this endeavour, we use a multilayer perceptron Feed forward neural network (FFNN) trained with backpropagation. It was decided to employ several ANN to identify the input devices' data reliably. Each fingertip's flexion sensor contributes five pieces of information, while an accelerometer's three axes provide twelve pieces of information. The input layer of the system's network will employ 8. There are as many neurons in the output layer as letters in the ASL alphabet, plus 752 more for motions that aren't letters, such as transition gestures; the notes to be included are those based on static activity. Therefore, Moghaddam *et al.* [14] are removed, leaving the output layer at 25 neurons. After being triggered by a set of input neurons, the signals from those neurons will be sent along into the hidden layer and finally into the output layer. The output layer's values may take any value between 0 and 1, inclusive. Prototype matching, sometimes called statistical template matching, uses statistics to find the closest match between obtained information values and "templates" [3]. This method is unique because it can be implemented quickly without requiring substantial training or calibration. The ANN is the most widely used ML approach for pattern identification. Since the data from the data glove can be used to train this approach, it is possible to classify postures and discriminate between static and dynamic movements. Recognizing SL is one of several areas where long-term fuzzy logic has been used and requires human judgement [4]. Linear discriminant analysis (LDA) is a practical approach used in ML to deliver accurate and less difficult classification via dimensionality reduction with enhanced grouping [4]. HMMs are a well-liked method that has shown their usefulness in several fields, including computer vision, voice recognition, molecular biology, and SLR [5]. In addition to the HMM, the KNN is used in classifying hand gestures [6], and the KNN classifier, in conjunction with SVM, has been utilized in categorizing postures. The KNN contributes to research to improve the recognition of ASL signals [3].

3.2.3. Linear discriminant analysis

For reducing the dimensionality of a dataset, a common approach is LDA. Dimensionality reduction is the reduction of a dataset from n variables to k variables, where the k variables are some combination of the n variables that preserves or maximizes some valuable property of the dataset. In the case of LDA, the new variables are chosen (and the data reprojected) in a way that maximizes the linear separability of a particular set of classes in the underlying data. LDA makes a few data-related simplifications:

− The data have a Gaussian distribution, and when they are displayed, each feature has the form of a bell curve.
− The variance of each character is the same; that is, the values of each variable deviate from the mean by an amount roughly equivalent tothe average. The LDA model extrapolates each feature's mean and variance from the chronic kidney disease (CKD) data under these presumptions.

The conventional method for estimating the mean value (mu) of each input (x) for each class (k) is dividing the total importance by the total number of values included in the CKD dataset. Here yields an estimate of the mean value (mu).

$$mu(k) = 1/n(k) * sum(x) \tag{1}$$

Where, $mu(k)$ is the average value of x for class k, and n(k) is the number of occurrences of class k for each feature in the CKD dataset, the mean value of x for class k is denoted by. Calculating the variance across all classes requires taking the average squared deviation of each result from the mean and applying (1).

$$sigma\text{\textasciicircum}2 = 1/(n-K) * sum((x-mu)\text{\textasciicircum}2) \tag{2}$$

Where, $sigma\text{\textasciicircum}2$ represents the variance over all inputs (x), n represents the number of instances, K represents the number of classes, and mu is the mean for the input x. LDA generates predictions by estimating the likelihood that a fresh set of inputs belongs to the target class value. This estimation is done in the context of a target class value. The output value is the value that has the greatest likelihood of being correct and thus concludes the classification process. Using Bayes' theorem, the model may provide a probability estimate. Bayes' theorem may estimate the probability of an output class (k) given the input class (x) by considering the possibilities of each type and the chances of data belonging to each category in the CKD dataset.

$$P(Y = x|X = x) = (PIk * fk(x)) / sum(PIl * fl(x)) \tag{3}$$

The base probability of each class (k), as shown in the training data, is denoted by the notation PIk (e.g. 0.5 for a 50-50 split in a two-value target class). It is known as the prior probability in Bayes' theorem.

$$PIk = nk/n \tag{4}$$

The likelihood that x is a class member is represented by the formula f(x) in the previous sentence. For f, a distribution function based on the Gaussian model is utilized (x). We arrive at the equation presented in (5) by applying the Gaussian to the equation and then simplifying the result. A discriminant function is what it is called, and the class that is determined to have the most significant value will be the output classification (y):

$$Dk(x) = x * (muk/sigma\text{\textasciicircum}2) - (muk\text{\textasciicircum}2/(2 * sigma\text{\textasciicircum}2)) + ln(PIk) \tag{5}$$

the discriminant function for class k given input x is denoted by the Dk(x) symbol. The parameters $mu(k)$, $sigma\text{\textasciicircum}2$, and PIk are all determined from the data using (2)-(5), respectively.

3.3. Proposed model
3.3.1. Image segmentation

The segmentation of images is one of the most critical and complex problems for the analysis of images. The segmentation of ideas is an essential part of picture processing. Computer vision divides an image into meaningful regions or objects by partitioning it. Otsu is one of the most effective approaches for determining picture thresholds among all segmentation methods becauseof its straightforward computation. The Otsu method is known for being both easy to do and very successful, making it one of the most widely used techniques. Its foundation is thresholding technology. It relies on the threshold's ideal value that optimizes the generated object's variation and background classes across classes. This value may be found by experimenting with different values. The work flow of proposed work was shown in Figure 4. This was done by giving an input as image through web application and after that it will preprocess the data and it will feed an image to the deep learning model and after that by observing the object of the user and it will recognize the entire object by letter.

3.3.2. Image segmentation algorithm

Let's say that the intensity of a greyscale picture may be described as L different degrees of grey [1,2,..., L]. The total number of points is stated as = +   + ⋯ +, and the symbol gives the number of points with the grey level at I. It is possible to think of the histogram of this grayscale picture as an occurrence distribution of probability.

$$p(i) =, x \geq 0, \sum x = 1$$

Using the threshold value of t, the image pixels are segmented into the foreground and background components known as C0 and C1, respectively. Where C0 stands for pixels that are included within levels [1,2, …t], and C 1 indicates pixels that are contained throughout levels [t+1,., L]. It is possible to describe the average and this class's probability of occurrence as [16]. Where w and z denote probabilities of foreground

part and background part. Besides, μ, α refer to the mean in grey levels of the foreground, the experience of the grey image and the entire gray level image. The between-class variance σ of the C0 and C1 is given by:

$$\sigma = w\,(\mu - \alpha) + z\,(\mu - \alpha)$$

in the discrimination analysis, the separable degree η of the class is $\eta = max\ \sigma$. Finally, maximizing σ to choose the optimal threshold.

$$tt = arg\ max\ \sigma$$

A large number of classes is one of the main reasons for comparatively less precision (26). So we tried to divide the classification issue into multiple levels or hierarchies. We first train an SVM model with a linear kernel to classify alphabets into a handed hand or two.
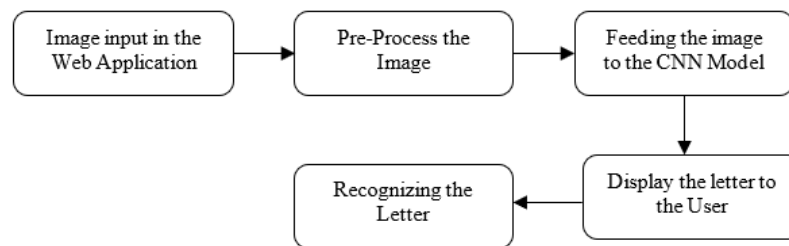
Figure 4. Proposed framework

4.   RESULTS

The input image is normalized. The convolution base is first created for the model, and pooling is done. Then the flattening layer is added, and the dense layers are added (hidden and output layer). The model is compiled and compiled. The optimizer is adam, and the loss function is categorical cross-entropy. Data is trained and tested on this model for 10 epochs and then saved the model. The accuracy graph between the training and validation data is shown in Figure 5. Then the model was integrated with the web application. The input is taken from the user, the information is preprocessed, and then the letter is predicted using the model; then, the model matches the sign with the classes and displays the class's alphabet.

This model has been accurate at 95%. Afterwards, we trained the multiclass SVM models of a linear kernel to classify one-handed alphabet (accuracy 56%) and two-handed alphabets (accuracy 60%). A single-handed or two-handed alphabet is first organized, then inserted into the appropriate model and labelled, depending on classification. In contrast, the individual model results on histograms of oriented gradient (HOG) features were almost the same overall performance and four-fold cross-validation (CV) accuracy of 53.23% as direct multiclass SVM, as shown in Figures 5-7, shows some of the output screens of alphabets from 'A' to 'Z'.
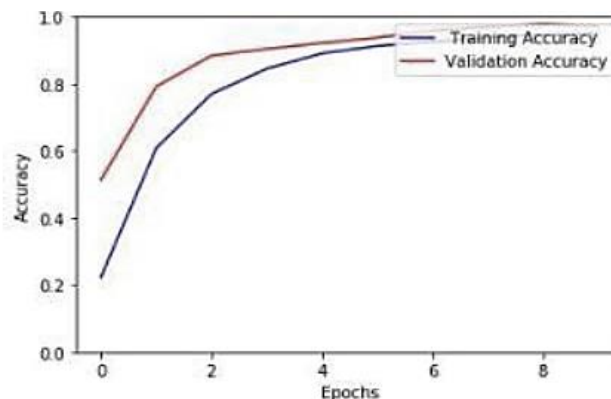
Figure 5. Model accuracy

Screenshot of our gesture 'A'

Screenshot of our own gesture 'D'

Screenshot of our own gesture 'F'

Screenshot of our gesture 'G'

Screenshot of our own gesture 'H'

Screenshot of our own gesture 'I'

Screenshot of our gesture 'K'

Screenshot of our gesture 'L'

Screenshot of our own gesture 'O'

Screenshot of our own gesture 'P'

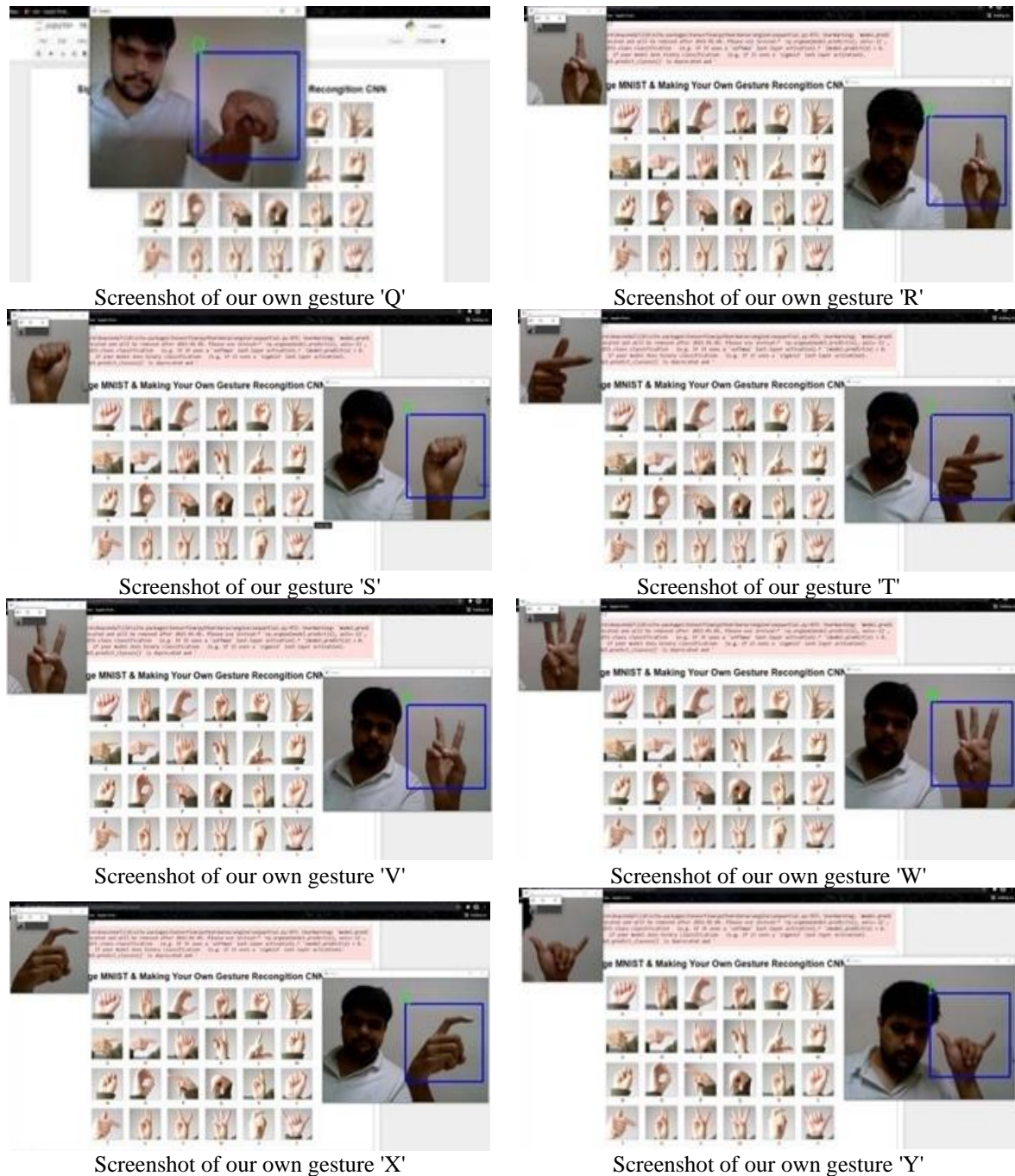Figure 6. Screenshots for alphabets from 'A' to 'P'

Screenshot of our own gesture 'Q'

Screenshot of our own gesture 'R'

Screenshot of our gesture 'S'

Screenshot of our gesture 'T'

Screenshot of our own gesture 'V'

Screenshot of our own gesture 'W'

Screenshot of our own gesture 'X'

Screenshot of our own gesture 'Y'

Figure 7. Screenshots for alphabets from 'Q' to 'Y'

5. CONCLUSION

The proposed system for recognizing SL characters can be extended further to identify gestures. It will be better to display sentences as the most suitable language translation rather than displaying letter labels. The readability is also enhanced. The scope of several SLs may be expanded. Some more training material may be needed to identify the letter appropriately. Multiple pictures are necessary to clarify and analyze information in contemporary applications. Some functionalities will need to be disabled to enable the various programmes' operations. An image is converted from one form to another during digitalization, scanning, communication, and storage. Therefore, an image enhancement process has to be carried out by the output image. This process includes several approaches to developing the visual presence of a picture. Thus, this process must be carried out. Image enhancement improves the interpretability of pictures for individuals, their information awareness, and the information available to other autonomous imaging systems. The

drawing will then be extracted using various techniques to make it more computer-readable. The creation of expert knowledge, the detection of edges, and the synthesis of false information from diverse sources may all be done effectively with the help of systems that recognize SLs. Convergence in a NN works toward the goal of producing an accurate rating for the output.

REFERENCES

[1] D. Metaxas, M. Dilsizian, and C. Neidle, "Scalable ASL sign recognition using model-based machine learning and linguistically annotated corpora," *8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community, Language Resources and Evaluation Conference 2018 Miyazaki, Japan*, Feb. 2018.

[2] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8925, Springer International Publishing, 2015, pp. 572–578. doi: 10.1007/978-3-319-16178-5_40.

[3] D. Bhavana, K. K. Kumar, M. B. Chandra, P. V. S. K. Bhargav, D. J. Sanjana, and G. M. Gopi, "Hand sign recognition using CNN," *International Journal of Performability Engineering*, vol. 17, no. 3, pp. 314–321, 2021, doi: 10.23940/ijpe.21.03.p7.314321.

[4] A. A. Hosain, P. S. Santhalingam, P. Pathak, J. Kosecka, and H. Rangwala, "Sign language recognition analysis using multimodal data," in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Oct. 2019, pp. 203–210. doi: 10.1109/DSAA.2019.00035.

[5] M. Assan and K. Grobel, "Video-based sign language recognition using hidden Markov models," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1998, vol. 1371, pp. 97–109. doi: 10.1007/BFb0052992.

[6] H. Wang, M. C. Leu, and C. Oz, "American sign language recognition using multi-dimensional hidden Markov models," *Journal of Information Science and Engineering*, vol. 22, no. 5, pp. 1109–1123, 2006.

[7] J. Atwood, M. Eicholtz, and J. Farrell, "American sign language recognition system," *24-787 Artificial Intelligence and Machine Learning for Engineering Design*. Pittsburgh, pp. 1–4, 2012.

[8] C.-H. Chuan, E. Regina, and C. Guardino, "American sign language recognition using leap motion sensor," in *2014 13th International Conference on Machine Learning and Applications*, Dec. 2014, pp. 541–544. doi: 10.1109/ICMLA.2014.110.

[9] M. M. Rahman, M. S. Islam, M. H. Rahman, R. Sassi, M. W. Rivolta, and M. Aktaruzzaman, "A new benchmark on American sign language recognition using convolutional neural network," in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dec. 2019, pp. 1–6. doi: 10.1109/STI47673.2019.9067974.

[10] T. Starner and A. Pentland, "Real-time American sign language recognition from video using hidden Markov models," in *Proceedings of International Symposium on Computer Vision - ISCV*, 1995, pp. 265–270. doi: 10.1109/ISCV.1995.477012.

[11] K. Bantupalli and Y. Xie, "American sign language recognition using deep learning and computer vision," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec. 2018, pp. 4896–4899. doi: 10.1109/BigData.2018.8622141.

[12] Chakraborty *et al.*, "Indian sign language classification (ISL) using machine learning," *American Journal of Electronics and Communication*, vol. 1, no. 3, 2021, doi: 10.15864/ajec.1304.

[13] U. H. Priya, S. K. Prasad, M. M. Jacob, R. R. Krishna, and P. R. Vinod, "American sign language recognition using CNN," *International Journal of Research in Engineering, Science and Management*, vol. 3, no. 7, pp. 333–336, 2020.

[14] M. Moghaddam, M. Nahvi, and R. H. Pak, "Static Persian sign language recognition using kernel-based feature extraction," in *2011 7th Iranian Conference on Machine Vision and Image Processing, MVIP 2011 - Proceedings*, Nov. 2011, pp. 1–5. doi: 10.1109/IranianMVIP.2011.6121539.

[15] T. Goswami and S. R. Javaji, "CNN model for American sign language recognition," in *Lecture Notes in Electrical Engineering*, Oct. 2021, vol. 698, pp. 55–61. doi: 10.1007/978-981-15-7961-5_6.

[16] R. S. Shankar, L. V. Srinivas, P. Neelima, and G. Mahesh, "A framework to enhance object detection performance by using YOLO algorithm," in *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, Apr. 2022, pp. 1591–1600. doi: 10.1109/ICSCDS53736.2022.9760859.

[17] A. A. Barbhuiya, R. K. Karsh, and R. Jain, "CNN based feature extraction and classification for sign language," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 3051–3069, Jan. 2021, doi: 10.1007/s11042-020-09829-y.

[18] G. A. Rao, K. Syamala, T. V. S. Divakar, and N. S. kumar, "Selfie sign language recognition with shape energy features and mahalanobis distance classifier," in *Lecture Notes in Electrical Engineering*, vol. 690, Springer Nature Singapore, 2021, pp. 161–168. doi: 10.1007/978-981-15-7504-4_17.

[19] V. M. Gupta, K. Murthy, and R. S. Shankar, "A novel approach for image denoising and performance analysis using SGO and APSO," *Journal of Physics: Conference Series*, vol. 2070, no. 1, Nov. 2021, doi: 10.1088/1742-6596/2070/1/012139.

[20] H. M. Mariappan and V. Gomathi, "Real-time recognition of Indian sign language," in *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, Feb. 2019, pp. 1–6. doi: 10.1109/ICCIDS.2019.8862125.

[21] T.-W. Chong and B.-G. Lee, "American sign language recognition using leap motion controller with machine learning approach," *Sensors*, vol. 18, no. 10, p. 3554, Oct. 2018, doi: 10.3390/s18103554.

[22] H. B. D. Nguyen and H. N. Do, "Deep learning for American sign language fingerspelling recognition system," in *2019 26th International Conference on Telecommunications (ICT)*, Apr. 2019, pp. 314–318. doi: 10.1109/ICT.2019.8798856.

[23] R. S. Shankar, G. Mahesh, K. V. S. S. Murthy, and D. Ravibabu, "A novel approach for gray scale image colorization using convolutional neural networks," in *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, Jul. 2020, pp. 1–8. doi: 10.1109/ICSCAN49426.2020.9262377.

[24] C. Dong, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using Microsoft Kinect," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2015, vol. 2015-Octob, pp. 44–52. doi: 10.1109/CVPRW.2015.7301347.

[25] D. R. Babu, R. S. Shankar, G. Mahesh, and K. V. S. S. Murthy, "Facial expression recognition using bezier curves with hausdorff distance," in *2017 International Conference on IoT and Application (ICIOT)*, May 2017, pp. 1–8. doi: 10.1109/ICIOTA.2017.8073622.

[26] J. Shotton *et al.*, "Efficient human pose estimation from single depth images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2821–2840, Dec. 2013, doi: 10.1109/TPAMI.2012.241.

[27] R. S. Shankar, G. Mahesh, K. V. S. S. Murthy, and J. Rajanikanth, "A novel approach for sharpening blur image using convolutional neural networks," *Journal of critical reviews*, vol. 7, no. 07, pp. 139–148, Apr. 2020, doi: 10.31838/jcr.07.07.22.

[28] M. S. Anand, N. M. Kumar, and A. Kumaresan, "An efficient framework for Indian sign language recognition using wavelet transform," *Circuits and Systems*, vol. 07, no. 08, pp. 1874–1883, 2016, doi: 10.4236/cs.2016.78162.

[29] R. S. Shankar, L. V. Srinivas, V. S. Raju, and K. Murthy, "A comprehensive analysis of deep learning techniques for recognition of flower species," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Feb. 2021, pp. 1172–1179. doi: 10.1109/ICICV50876.2021.9388503.

[30] R. B. Devareddi and A. Srikrishna, "Review on content-based image retrieval models for efficient feature extraction for data analysis," in *2022 International Conference on Electronics and Renewable Systems (ICEARS)*, Mar. 2022, pp. 969–980. doi: 10.1109/ICEARS53579.2022.9752281.

[31] V. López-Ludeña, C. González-Morcillo, J. C. López, E. Ferreiro, J. Ferreiros, and R. San-Segundo, "Methodology for developing an advanced communications system for the deaf in a new domain," *Knowledge-Based Systems*, vol. 56, pp. 240–252, Jan. 2014, doi: 10.1016/j.knosys.2013.11.017.

[32] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A hidden markov model-based continuous gesture recognition system for hand motion trajectory," in *2008 19th International Conference on Pattern Recognition*, Dec. 2008, pp. 1–4. doi: 10.1109/ICPR.2008.4761080.

[33] R. S. Shankar, K. Sravani , L. V. Srinivas and D. Ravibabu, "An approach for retrieving an image using genetic algorithm," *International Journal of Latest Trends in Engineering and Technology*, vol. 9, no. 2, 2017, doi: 10.21172/1.92.10.

BIOGRAPHIES OF AUTHORS

Lokavarapu V. Srinivas ⓘ 🔢 SC ◖ is an Assistant Professor in the Department of Computer Science and Engineering in Sagi Rama Krishnam Raju Engineering College, Bhimavaram, Andhra Pradesh, India. He Received B.Tech and M.Tech Degrees from Andhra University, Visakhapatnam. Currently He is pursuing Ph. D in Computer Science and Engineering at Andhra University. His research areas are machine learning, edge computing, natural language processing and image processing. He can be contacted at email: srinivas.srkrcse@gmail.com.

Chitri Raminaidu ⓘ 🔢 SC ◖ is an Assistant Professor at the department of Computer Science and Engineering in Sagi Ramakrishnam Raju Engineering College, Bhimavaram, Andhra Pradesh, India. He received B. Tech Degree in Sri Vasavi Institute of Engineering and Technology, Department of Information Technology in 2012. He holds an M.Tech Degree in Sri Vasavi Institute of Engineering and Technology, Department of Computer Science and Engineering. His research areas are image processing, machine learning and deep learning. He published 8 papers in International Journals and Conferences. He can be contacted at email: raminaidu.chitri@gmail.com.

Mr. Devareddi Ravibabu ⓘ 🔢 SC ◖ received his B. Tech (Computer Science and Engineering) Degree in 2005 from Andhra University, obtained M. Tech (Computer Science and Technology) Degree from Andhra University in 2009 and doing Research in Computer Science and Engineering in Acharya Nagarjuna University, Guntur, India. My current research interests are in internet of things, image processing, computer vision and artificial intelligence. Email: ravibabu.devareddi@gmail.com.

Shiva Shankar Reddy ⓘ 🔢 SC ◖ is an Assistant Professor at the Department of Computer Science and Engineering in Sagi RamaKrishnam Raju Engineering College, Bhimavaram, Andhrapradesh, INDIA. He is pursuing PhD degree in Computer Science and Engineering with a specialization in Medical Mining and Machine Learning. His research areas are Image Processing, Medical Mining, Machine Learning, Deep Learning and Pattern Recognition. He published 30+ papers in International Journals and Conferences. He has filed 05 patents. His research interests include image processing, medical mining, machine learning, deep learning and pattern recognition. He can be contacted at email: shiva.csesrkr@gmail.com.