IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 3, June 2025, pp. 2035~2043
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i3.pp2035-2043

Sign language recognition and classification using blended ensemble machine learning

Akash Rajan Rai, Sujata Rajesh Kadu
Department of Information Technology, Terna College of Engineering, University of Mumbai, Navi Mumbai, India

Article Info

Article history:

Received Mar 3, 2024
Revised Nov 25, 2024
Accepted Jan 27, 2025

Keywords:

Blended ensemble ML
Data augmentation
ISL recognition
Multi-thresholding
VGG-16

ABSTRACT

An efficient sign language recognition system (SLR) is the most significant for hearing-impaired people for communication. The body movements and hand gestures are utilized to characterize the vocabulary in dynamic sign language. The SLR is a challenging problem because the computational model requires simultaneous spatial-temporal modelling for a number of sources. To overcome this problem, this research proposes the blended ensemble machine learning (ML) approaches for SLR. Initially, the Indian sign language (ISL) dataset is collected for evaluating the effectiveness of the model. Then, the pre-processing is done by using data augmentation and normalization techniques. Then, the pre-processed data is provided to the segmentation process which is done by using multi-threshold entropy function. Then, VGG-16 is used for the feature extraction process to extract the features and finally, classification is carried out using ensemble ML. An effectiveness of the proposed method is validated based on accuracy, precision, recall, and F1-score, wherein it achieves better results of 99.57%, 0.92%, 0.95%, and 0.99% as compared to the existing works like support vector machine (SVM) and convolutional neural network (CNN).

Corresponding Author:

Akash Rajan Rai
Department of Information Technology, Terna Engineering College, University of Mumbai
Navi Mumbai, Maharashtra, India
Email: akashrai932@gmail.com

1. INTRODUCTION

Communication is significant for people to discover their necessities as well as interactions with other people. There are number of deaf and dumb people who majorly depend on sign language to communicate with others [1], [2]. The sign languages are measured as graphical and non-verbal form of communications utilized through differently-abled people to express themselves or interact with their surroundings. Globally, the sign language is the most emerging as well as challenging task. The sign language efficiently helps the hard-of-hearing as well as speed-impaired society in acquiring academic proficiency, professions as well as social rights [3], [4]. The sign language converts the words, sentences, numbers and letters of natural language to enable the vocally deactivated people to interact with the other people [5]. In sign language, the meaning and extraction of data is expressed through using hand gestures, movements of the body, facial expressions as well as emotions rather than sound, to send the messages. Moreover, sign language minimizes the communication gap among deaf and dumb people, facilitating smooth communication. The sign languages vary from region to region, and nation to nation [6], [7]. The number of researchers determine an exciting and exclusive form of communication in sign language over various nations. The machine learning (ML) and deep learning (DL) techniques have obtained better enhancement capabilities in sign language recognition (SLR) [8], [9]. The ML or DL techniques are implemented for the automatic recognition of sign language gestures to minimize communication gap with these people.

Various researchers design the new approaches for SLR from the advantages of existing approaches to enhance the model's performance [10], [11]. The SLR techniques are performed to enhance the efficiency of the model through minimizing the processing time, developing reliable databases, enabling quality enhancement. As an outcome, automatic SLR approaches are required to translate signs into equivalent text or sound without an assistance of translators [12], [13]. Mostly approaches are divided into two types; initially, the approach depends on hand shape as well as gesture movement. Then, the approaches depend on the sequence of image for every SLR. The researchers have introduced the SLR techniques for various languages such as Indian, Chinese, American, and German sign language. Though various improvements have been developed for SLR, still some models have failed to manage the real-time datasets and lack in some cases [14]–[16]. Nowadays, noteworthy progressions have been developed in the area of SLR, especially through ML approaches. However, attaining the better accuracy and robustness over various sign language dataset is challenging. Hence, the motivation of this research is to solve the afore-mentioned problem through proposing the blended ensemble ML approaches to improve the effectiveness of SLR. Compared to previous works like K-nearest neighbor (KNN), naïve Bayes (NB), random forest (RF), and support vector machine (SVM), the proposed blended ensemble ML approaches attain better accuracy and generalizability, representing the enhanced effectiveness.

Based on the nature of sign language, the researchers utilize various approaches for SLR. Various types of SLR are analyzed in this section to eliminate the communication gap. Because of the varying nature of signs in every sign language, the sign recognition is challenging. Athira *et al*. [17] introduced the SVM for self-determining vision-based SLR approach. The suggested approach had the capability to recognize the single and double handed of static as well as dynamic gestures using real-time video of Indian sign language (ISL). In that pre-processing stage, skin color segmentation approach was utilized for the ISL extraction. The suggested approach efficiently minimized a computational speed to an excessive amount by the utilization of Zernike moments frame extraction approach. However, the introduced approach was performed only with five static symbols for SLR. Katoch *et al*. [18] presented the SVM and convolutional neural network (CNN) for the classification of sign language. The suggested approach utilized the bag of visual words (BOVW) to recognize the ISL alphabets (A-Z) as well as digits (0-9) in real-time video stream. Speeded up robust features (SURF) were extracted from the histograms, and images were produced to map the sign with consistent labels. A collaborative graphical user interface (GUI) was generated for easy access of SLR. However, the suggested approach utilized the large cluster data for enhance the model performance.

Sharma and Singh [19] developed the robust computer-vision based CNN by depth wise separable convolution (DSN) for the recognition of sign language. Initially, the ISL dataset was created from 65 users in an uncontrolled environment. Then, intra-class variance in the database was performed by augmentation approach to enhance the generalization capability. The CNN was utilized for the process of feature extraction as well as classification of ISL sign language. The suggested approach easily solved the determination of two-hand ISL gesture problem. But the suggested approach had poor performance in case of determine the similar gestures. Natarajan *et al*. [20] presented the complete DL approach to handle the SLR, production tasks, and translation in real-time cases. The suggested approach utilized a MediaPipe library and hybrid CNN with Bi-directional long short-term memory (Bi-LSTM) for image and text extraction. Alternatively, the sign gesture videos of speech sentences were computed by the utilization of hybrid neural machine translation (NMT) + MediaPipe + dynamic generative adversarial network (GAN) approach. Nonetheless, the suggested approach in such protocols significantly affected the performance. Nandi *et al*. [21] introduced the CNN integrated with augmentation, batch normalization, dropout, stochastic pooling as well as difGrad optimizer for fingerspelling static sign recognition approach for ISL alphabet. The training, testing and losses of the suggested approach were attained for multiple separate optimizers, as well as three types of pooling approaches. Nonetheless, the suggested approach was futile in detecting the dynamic and real-time signs. From this analysis, some limitations have been identified; executed only with the minimum static symbols, poor performance, and futile to recognize the dynamic and real-time signs. Based upon this inference, this research proposed employs ISL recognition, and the new approach is discussed in detail in the following section. The major contributions of this paper are listed as follows:

- The preprocessing is performed by using data augmentation and data normalization techniques. The multi threshold image entropy technique is used for segmentation.
- In this research, the novel and robust of blended ensemble ML approach is proposed for SLR which is utilized in ISL translation system.
- The proposed method effortlessly challenges a complex problem for determining two-hand gestures of ISL with better recognition outcomes over other the existing works.

The structure of the paper is arranged as follows: section 2 provides the proposed method. Section 3 presents the classification using blended ensemble ML techniques. Section 4 provides the results and discussion, and section 5 covers the conclusion of this research.

2. PROPOSED METHOD

In this section, the proposed methodology is presented to describe various steps which are carried out in this research. The data flows in this proposed method involves collection of data, preprocessing, segmentation, feature extraction and classification. Figure 1 shows the process intricate in SLR.

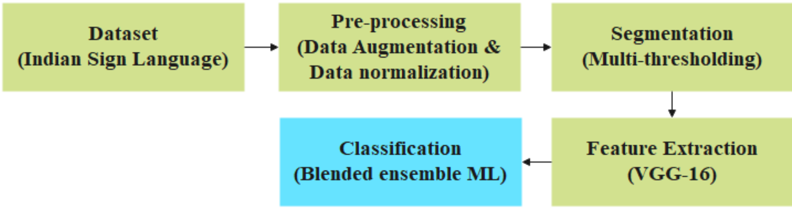

Figure 1. Workflow of the proposed method

2.1. Dataset

Data acquisition is significant for research works and it is essential for ML development. The one of the most significant challenges faced during the research is that there are no benchmark datasets available for ISL [22]. Hence, this research tried to physically develop the dataset that comprises of alphabets (A-Z) and numerical values (0-9) by utilizing the external web camera. Sign is accomplished by 65 various people which consequences in a total of 1690 gestures. From this collected dataset, every image is scaled down from a maximum definition of size 256×256. Figure 2 shows the samples collected data of both alphabets and numerical values.

Figure 2. Sample dataset of hand gesture

2.2. Pre-processing

In this section, the collected data is utilized as input for the pre-processing. These techniques involve data augmentation, and normalization are applied in the collected ISL gestures data. In data augmentation, the collected input ISL dataset has only a limited number of samples, hence data augmentation is performed to enhance the size of the image samples [23], [24]. The data is enhanced by crating the new similar samples through transforming the actual data. The model gets trained utilizing 49,920 images (80%) and tested utilizing 12,480 (20%) of the total of 62,400 images. The trained images are augmented by using various affine operations such as zooming, rotation, skewing, shearing, height and width shift. As a consequence, every image produces 90 new images, hence newly produced images are multiplied into 49,920 images, which resulted in 4,542,720 images in augmented training data. Then, the augmented data is fed to the normalization technique to standardize the input to a larger.

The data normalization is utilized to estimate if the data distribution in every input pixel is unchanging or regular, alongside faster training convergence. The normalization takes maximum and minimum values to set the data in between the range of 0 and 1, respectively. The normalization is calculated by using (1):

$$X_{new} = \frac{\mathrm{X}-\mathrm{X_{min}}}{\mathrm{X_{max}}-\mathrm{X_{min}}} \tag{1}$$

Where, X is the normalized value, $\mathrm{X_{min}}$ and $\mathrm{X_{max}}$ are the minimum and maximum of every feature. Then, the normalized data is forwarded for the segmentation process.

2.3. Segmentation

In this section, the pre-processes output is utilized as input to segment the image data. Image segmentation is an approach for extracting a target object from the background. The pre-processed data are classified according to the threshold value and depicted as single and multi-threshold segmentation. Hence, this research uses the image segmentation which is based on maximum entropy values. This model is utilized to validate the optimal threshold value, as well as is used to optimize the data presented in the background. Hereafter, an identification of image greyscale is accomplished to extract the targeted areas and the greyscale image range is explained as $[0, N - M]$. The pixels which embrace a greyscale value are considered as targeted area (T) and background area (B). These two areas involve randomness, which is then combined with possibility concentrations represented in (2) to (4):

$$E(T) = -\sum_{i=0}^{t} \frac{P_k}{P_l} In \frac{P_k}{P_l} \tag{2}$$

$$E(B) = -\sum_{i=l+1}^{M-1} \frac{P_k}{1-P_l} In \frac{P_k}{1-P_l} \tag{3}$$

$$\Psi(t) = E(T) + E(B) \tag{4}$$

Where $\Psi(t)$ is utilized to attain the greater value, and t is the optimal threshold value. Hence, a larger entropy thresholding approach is utilized for multi-threshold segmentation and the problems in this approach are examined as N dimensions. Furthermore, the greyscale image values considered to segment the input data and then exposed to the $l + 1$ region. An objective function of this approach is provided in (5).

$$O([t_1, t_2, \ldots t_l]) = E_0 + E_1 + \cdots + E_l = -\sum_{k=0}^{t_1-1} \frac{P_i}{\omega_0(t_1)} In \frac{P_i}{\omega_0(t_1)} + \sum_{k=0}^{t_2-1} \frac{P_i}{\omega_0(t_1,t_2)} In \frac{P_i}{\omega_0(t_1,t_2)} + \cdots + \sum_{k=t_l}^{M-1} \frac{P_i}{\omega_0(t_l,M-1)} In \frac{P_i}{\omega_0(t_l,M-1)} \tag{5}$$

Where P_i and ω is the possibility of greyscale value in ith image and class. An optimal threshold value is determined at the value of $O([t_1, t_2, \ldots t_l])$, influencing a similar value of $t_1^*, t_2^*, \ldots t_l^*$. Still, this segmentation owns the efficient segmentation accuracy and the segmented value is fed to the feature extraction process.

2.4. Feature extraction

The segmented data is given to the feature extraction to extract features in the dataset. The general aim of feature extraction is to reduce the dimensionality and data compaction. Hence, VGG-16 pre-trained architecture is used for the feature extraction process. The VGG-16 [25], [26] is the most popular image feature extraction employed to extract a large amount of data. The VGG-16 architecture consists of various layers such as convolution, fully connected (FC) as well as pooling, also applicable in AlexNet architecture. The input size of this architecture is fixed to $224 * 224$ pixels, accompanied by a filter size of $3 * 3$. At the end of this architecture, it involves the activation function which is accomplished for distributing the probabilities classes to output layers. In this feature extraction for the proposed method, the initial layer of VGG-16 architecture is forwarded to the convolutional layer with $224 * 224$ pixel image size. In the rectified linear unit (ReLU) activation function, $224 * 224$ image size is forwarded by primary stack of two convolutions with the appropriate region of $3 * 3$. Every layer of this architecture consists of 64 filters and a padding is often reserved with a pixel size of 1, whereas stride value is set to 1. A spatial resolution is taken in this model, wherein the activation function is dealt with a pooling layer with the stride value of 2×2 pixels. Moreover, in the following step, the outcome is once more forwarded to the convolutional layers and later fed to a pooling layer, which is resulted in an outcome of $56 * 56 * 128$ pixels. This layer endures the final process of convolutional as well as pooling layer. The flatten layer is the located among FC layers, which are then utilized as the convolutional stacks. Finally, the FC layer is performed as an output layer which involves 1000 neurons, whereas, the initial two FC layers represent an outcome of 4096 neurons. Then, the extracted features are forwarded to the classification process.

3. CLASSIFICATION USING BLENDED ENSEMBLE MACHINE LEARNING TECHNIQUES

In this section, an outcome of extracted features is utilized as input for classifying the various sign languages. The ML algorithms-based recognition and classification of sign languages provide a better performance. In this classification process, the blended ensemble ML approaches are used for recognition and classification of sign language. Ensemble learning is the procedure of integrating various learnings to enhance the performance, where every learning approach trains on various subsets of features. The ensemble learning performance belonging to the working style and capability of the selected base classifiers. Ensemble learning models endeavor to enhance predictability through combining the various approaches into highly dependable individual model. The detailed information about this blended ensemble approach is described.

3.1. K-nearest neighbor

The KNN [27] algorithm is unpretentious and easy-to-implement when compared to the logistic regression (LR). The aim of this algorithm is to measure the distance among selected point and other points. To select the k points with minimal distance, this model creates statistics on classification types of those selected k points. The KNN algorithm believes that the objects are similar to each other. The distance is mainly calculated by using Euclidean and Manhattan distance, the mathematical formula of these distance is formulated in (6) and (7):

$$Euclidean\ distance: d(x,y) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2} \tag{6}$$

$$Manhattan\ distance: d(x,y) = \sqrt{\sum_{k=1}^{n}|x_k - y_k|} \tag{7}$$

where x_k, y_k- kth dimensional values of the two points which are utilized in dimensional space.

3.2. Naïve Bayes

The NB [28] is one of the ML approaches, which utilizes the Bayes theorem to solve the problem probability function. To determine the posterior distribution, background knowledge is explained as prior distribution, as well as associated with the perceptional data in the procedure of probability function. The Bayesian statistics is an approach for determining the data and evaluating the parameters according to the Bayes' theorem, which is expressed in (8) and the $p(features)$ are estimated by using (9) based on the total probability rule.

$$p(classes|features) = \frac{p(features|classes)p(classes)}{p(features)} \tag{8}$$

$$P(B) = \sum_{i=1}^{n} p(A_i)P(B|A_i) \tag{9}$$

Eventually, the probability depends on a particular group and the greater value is taken as classification outcome. There are various types of NB such as Bernoulli, multinominal and Gaussian NB, which differentiate the feature vector as discrete or continuous. The Bernoulli NB classifier is suitable in a situation where the feature vectors imitate to the Bernoulli distribution such as binary distribution. The multinomial NB classifier is applicable for the discrete feature vector and imitates the multivariate distributions. Finally, the Gaussian NB is employed only once the feature vectors have continuous variables and imitate or estimate to the actual distribution.

3.3. Random forest

RF [29] is an illustrative ensemble learning approach that depends on strategies of the bagging family. The RF utilizes decision tree (DT) as base learners and develops various DT by utilizing the sampling approach without placement. The RF is an approach for aggregating or even bagging data which is deployed to minimize a significant parameter of the variance in the output. In classification process, each tree considers the output as the classes and the classes with greater number of outcomes are selected as final output. The test samples acquire outcomes on every developed DT as well as estimate the test sample category by the voting strategy. At the working of RF, various samples are arbitrarily chosen to develop the number of DT, while in the procedure of DT development, the selected features and its split nodes are arbitrary. For the classification task, the output of RF is the class selected by the majority of the class.

3.4. Support vector machine

SVM [30] is widely used for the problems of classification, detection as well as regression. SVM is the most energy efficient as it utilizes the subset of training points in decision function. The SVM employs efficiently in high-dimensional feature vectors, hence hyperplane dimensions are often less than 1 in feature vector. The process of identifying the desired hyperplane is determining a maximum margin. The purpose is to enhance the margin for improving the robustness, as well as for reducing the classification's error rate. The sample points which determine a greater margin are known as support vectors. The SVM is used for making predictions and is essential for fit on sparse data. The goal of SVM is to classify the sample into one or more classes through the extension that specifies to the multi-class problem. There are a number of hyperplanes that discriminate the two classes; however, the aim is attained by identifying an optimal separating hyperplane which places outermost from both classes.

4. RESULTS AND DISCUSSION

This section illustrates the results and discussion of the proposed method for SLR. The proposed method for SLR is executed by utilizing the platform of python 3.8 with Windows 10 OS, 16 GB RAM with intel-i7 processor. The proposed blended ensemble ML approach is analyzed based on various assessment metrics named accuracy, precision, recall, and F1-score. The mathematical expressions of these metrices are expressed in (10) and (11).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} (10)\ Precision = \frac{TP}{TP+FP} \tag{10}$$

$$Recall = \frac{TP}{TP+FN} (12)\ F1 - score = \frac{2TP}{2TP+FP+FN} \tag{11}$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

4.1. Performance analysis

In this section, the effectiveness of the proposed method for SLR is analyzed by using the standard dataset. Tables 1 to 3 show the outcomes during segmentation, feature extraction and classification for SLR on ISL dataset. The performance metrices named accuracy, precision, recall, and F1-score are utilized to estimate an effectiveness of the proposed method.

Table 1 shows the multi-thresholding outcomes with various segmentation techniques for SLR on ISL dataset. The multi-thresholding attains better performance as compared to the existing methods with various assessment metrices. The existing works such as like Otsu thresholding are compared with the proposed multi-thresholding technique. The suggested multi-thresholding technique attains the accuracy, precision, recall, and F1-score of 99.57%, 0.92, 0.95, and 0.99, correspondingly these results show effective outcomes.

Table 1. Performance analysis for segmentation

Methods	Accuracy (%)	Precision	Recall	F1-score
Edge-based	91.23	0.82	0.84	0.88
Region-based	93.06	0.83	0.88	0.90
Adaptive thresholding	95.28	0.86	0.91	0.93
Otsu thresholding	97.38	0.90	0.93	0.96
Multi-thresholding	99.57	0.92	0.95	0.99

Table 2 presents the VGG-16 architecture with various pre-trained models for SLR on ISL dataset. The VGG-16 attains superior performance as compared to the existing methods with various assessment metrices. The existing works such as like ResNet50, ImageNet, InceptionNet and ALexNet are compared with the proposed technique. The proposed VGG-16 technique simultaneously attains the accuracy, precision, recall, and F1-score of 99.57%, 0.92, 0.95, and 0.99 and these results shows effective outcomes.

Table 3 exhibits the result of the proposed Blended Ensemble ML method with various classification models for SLR on ISL dataset. The proposed approach attains significant performance as compared to the existing methods with various assessment metrices. The existing works such as like KNN, NB, RF, and SVM are compared with the proposed blended ensemble ML technique. The proposed method attains the accuracy, precision, recall, and F1-score of 99.57%, 0.92, 0.95, and 0.99, respectively. The blended ensemble ML method accomplishes effective outcomes when compared to individual approaches.

Table 2. Performance analysis for feature extraction

Methods	Accuracy (%)	Precision	Recall	F1-score
ResNet50	94.29	0.85	0.89	0.93
ImageNet	95.03	0.87	0.91	0.94
InceptionNet	96.29	0.88	0.93	0.96
AlexNet	98.47	0.91	0.94	0.97
VGG-16	99.57	0.92	0.95	0.99

Table 3. Performance analysis for classification results on ISL dataset

Methods	Accuracy (%)	Precision	Recall	F1-score
KNN	95.74	0.86	0.90	0.94
NB	96.38	0.87	0.92	0.95
RF	97.34	0.88	0.93	0.97
SVM	98.98	0.91	0.94	0.98
Blended Ensemble ML	99.57	0.92	0.95	0.99

4.2. Comparative analysis

Table 4 presents the comparative results of the proposed work with existing works. The proposed blended ensemble ML is validated with method, datasets, accuracy, precision, and recall for ISL dataset. The existing works such as [17], [19]–[21] are used to analyze in contrast to the proposed method.

Table 4. Comparative results of proposed work with existing work on ISL dataset

Methods	Accuracy (%)	Precision	Recall	F1-score
SVM [17]	90.1	N/A	N/A	N/A
CNN-DSC [19]	92.43	0.89	0.94	0.98
VGG-16+Bi-LSTM [20]	98.56	N/A	0.9843	0.9758
CNN [21]	99.36	N/A	N/A	N/A
Proposed blended ensemble ML	99.57	0.92	0.95	0.99

4.3. Discussion

This section illustrates the limitations of existing works and explains how the proposed method overcomes such limitations. The limitations of the existing works such as SVM [17] only performed with the 5 static symbols for SLR. SVM-CNN [18] utilized the large cluster data for enhancing the model's performance. The CNN-depth separable convolution (DSC) [19] had poor performance in case of recognition of the similar gestures. CNN [21] approach recognized only the static signs. However, the proposed blended ensemble ML method tackles these limitations. The proposed attains better results and accomplishes an accuracy of 99.57%, precision of 0.92, recall of 0.95, and F1-score of 0.99. On the other hand, the existing work CNN-DSC [19] attains the accuracy of 92.43%. Thereby, the existing works exhibit poor performance in contrast to the proposed blended ensemble ML method.

5. CONCLUSION

The aim of this research is to accomplish the real-time recognition of alphabets and numerical values in ISL by providing the image processing-based recognition method. Hence in this research, the blended ensemble ML approach is introduced for ISL recognition and classification. In this introduced method, the blended ensemble learning involves four main ML approaches such as KNN, NB, RF, and SVM. As a result, this research enhances the SLR accuracy, as well as minimizes the model's computational complexity by utilizing the effective blended ensemble ML approach. The efficiency of the model which is trained and tested on the ISL collected dataset is studied. The experimental results reveal that the proposed method accomplishes the accuracy of 99.57%, precision of 0.92, recall of 0.95, and F1-score of 0.99 respectively. In future work, the performance of the proposed method will be enhanced for a number of real-time applications.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Akash Rajan Rai	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	
Sujata Rajesh Kadu	✓	✓	✓							✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Indian Sign Language at https://www.kaggle.com/datasets/prathumarikeri/indian-sign-language-isl, reference number [22].

REFERENCES

[1] M. I. Saleem, A. Siddiqui, S. Noor, M. A. Luque-Nieto, and E. Nava-Baro, "A machine learning based full duplex system supporting multiple sign languages for the deaf and mute," *Applied Sciences*, vol. 13, no. 5, 2023, doi: 10.3390/app13053114.

[2] G. H. Samaan *et al.*, "MediaPipe's landmarks with RNN for dynamic sign language recognition," *Electronics*, vol. 11, no. 19, 2022, doi: 10.3390/electronics11193228.

[3] M. S. Amin and S. T. H. Rizvi, "Sign gesture classification and recognition using machine learning," *Cybernetics and Systems*, vol. 54, no. 5, pp. 604–618, 2023, doi: 10.1080/01969722.2022.2067634.

[4] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A. B. Gil-González, and J. M. Corchado, "Deepsign: sign language detection and recognition using deep learning," *Electronics*, vol. 11, no. 11, 2022, doi: 10.3390/electronics11111780.

[5] C. Lu, M. Kozakai, and L. Jing, "Sign language recognition with multimodal sensors and deep learning methods," *Electronics*, vol. 12, no. 23, 2023, doi: 10.3390/electronics12234827.

[6] Q. M. Areeb, Maryam, M. Nadeem, R. Alroobaea, and F. Anwer, "Helping hearing-impaired in emergency situations: a deep learning-based approach," *IEEE Access*, vol. 10, pp. 8502–8517, 2022, doi: 10.1109/ACCESS.2022.3142918.

[7] O. M. Sincan and H. Y. Keles, "Using motion history images with 3D convolutional networks in isolated sign language recognition," *IEEE Access*, vol. 10, pp. 18608–18618, 2022, doi: 10.1109/ACCESS.2022.3151362.

[8] S. Das, M. S. Imtiaz, N. H. Neom, N. Siddique, and H. Wang, "A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier," *Expert Systems with Applications*, vol. 213, 2023, doi: 10.1016/j.eswa.2022.118914.

[9] J. Bora, S. Dehingia, A. Boruah, A. A. Chetia, and D. Gogoi, "Real-time assamese sign language recognition using mediapipe and deep learning," *Procedia Computer Science*, vol. 218, pp. 1384–1393, 2022, doi: 10.1016/j.procs.2023.01.117.

[10] M. Hrúz, I. Gruber, J. Kanis, M. Boháček, M. Hlaváč, and Z. Krňoul, "Ensemble is what we need: isolated sign recognition edition," *Sensors*, vol. 22, no. 13, 2022, doi: 10.3390/s22135043.

[11] R. Rastgoo, K. Kiani, and S. Escalera, "Real-time isolated hand sign language recognition using deep networks and SVD," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 1, pp. 591–611, 2022, doi: 10.1007/s12652-021-02920-8.

[12] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, "Deepsign: Sign language detection and recognition using deep learning," *Electronics*, vol. 11, no. 11, 2022, doi: 10.11591/ijece.v12i3.pp2996-3004.

[13] S. Güney and M. Erkuş, "A real-time approach to recognition of Turkish sign language by using convolutional neural networks," *Neural Computing and Applications*, vol. 34, no. 5, pp. 4069–4079, 2022, doi: 10.3390/electronics11111780.

[14] A. Singh, A. Wadhawan, M. Rakhra, U. Mittal, A. Al Ahdal, and S. K. Jha, "Indian sign language recognition system for dynamic signs," *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO),* Noida, India, 2022, pp. 1-6, doi: 10.1109/ICRITO56286.2022.9964940.

[15] S. K. Singh and A. Chaturvedi, "A reliable and efficient machine learning pipeline for american sign language gesture recognition using EMG sensors," *Multimedia Tools and Applications*, vol. 82, no. 15, pp. 23833–23871, 2023, doi: 10.1007/s11042-022-14117-y.

[16] A. S. Nandhini, D. S. Roopan, S. Shiyaam, and S. Yogesh, "Sign language recognition using convolutional neural network," *Journal of Physics: Conference Series*, vol. 1916, no. 1, 2021, doi: 10.1088/1742-6596/1916/1/012091.

[17] P. K. Athira, C. J. Sruthi, and A. Lijiya, "A signer independent sign language recognition with co-articulation elimination from live videos: an Indian scenario," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 3, pp. 771–781, 2022, doi: 10.1016/j.jksuci.2019.05.002.

[18] S. Katoch, V. Singh, and U. S. Tiwary, "Indian sign language recognition system using SURF with SVM and CNN," *Array*, vol. 14, 2022, doi: 10.1016/j.array.2022.100141.
[19] S. Sharma and S. Singh, "Recognition of indian sign language (ISL) using deep learning model," *Wireless Personal Communications*, vol. 123, no. 1, pp. 671–692, 2022, doi: 10.1007/s11277-021-09152-1.
[20] B. Natarajan *et al.*, "Development of an end-to-end deep learning framework for sign language recognition, translation, and video generation," *IEEE Access*, vol. 10, pp. 104358–104374, 2022, doi: 10.1109/ACCESS.2022.3210543.
[21] U. Nandi, A. Ghorai, M. M. Singh, C. Changdar, S. Bhakta, and R. K. Pal, "Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling," *Multimedia Tools and Applications*, vol. 82, no. 7, pp. 9627–9648, 2023, doi: 10.1007/s11042-021-11595-4.
[22] P. Arikeri, "Indian sign language dataset," *Kaggle*. 2020. [Online]. Available: https://www.kaggle.com/datasets/prathumarikeri/indian-sign-language-isl/data
[23] C. Chen *et al.*, "Enhancing MR image segmentation with realistic adversarial data augmentation," *Medical Image Analysis*, vol. 82, 2022, doi: 10.1016/j.media.2022.102597.
[24] W. M. Salama and M. H. Aly, "Deep learning in mammography images segmentation and classification: automated CNN approach," *Alexandria Engineering Journal*, vol. 60, no. 5, pp. 4701–4709, 2021, doi: 10.1016/j.aej.2021.03.048.
[25] S. Sharma, K. Guleria, S. Tiwari, and S. Kumar, "A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer disease using MRI scans," *Measurement: Sensors*, vol. 24, 2022, doi: 10.1016/j.measen.2022.100506.
[26] G. S. Nijaguna, J. A. Babu, B. D. Parameshachari, R. P. de Prado, and J. Frnda, "Quantum fruit fly algorithm and ResNet50-VGG16 for medical diagnosis," *Applied Soft Computing*, vol. 136, 2023, doi: 10.1016/j.asoc.2023.110055.
[27] F. Utaminingrum, I K. Somawirata, and G. D. Naviri, "Alphabet sign language recognition using k-nearest neighbor optimization," *Journal of Computers*, vol. 14, no. 1, pp. 63–70, 2019, doi: 10.17706/jcp.14.1 63-70.
[28] J. K. Alwan, D. S. Jaafar, and I. R. Ali, "Diabetes diagnosis system using modified Naive Bayes classifier," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 3, pp. 1766–1774, 2022, doi: 10.11591/ijeecs.v28.i3.pp1766-1774.
[29] S. Rhouas, A. El Attaoui, and N. El Hami, "Enhancing currency prediction in international e-commerce: Bayesian-optimized random forest approach using the Klarna dataset," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 3, pp. 3177–3186, 2024, doi: 10.11591/ijece.v14i3.pp3177-3186.
[30] M. Bansal and S. Gupta, "Detection and recognition of hand gestures for indian sign language recognition system," *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC),* Solan, India, 2021, pp. 136-140, doi: 10.1109/ISPCC53510.2021.9609448.

BIOGRAPHIES OF AUTHORS

Akash Rajan Rai received a bachelor degree in information & technology from the University of Mumbai in 2020. He is currently pursuing his master's degree in information technology from Terna Engineering College. His areas of research include machine learning, data analysis, and web development frameworks. He can be contacted at email: akashrai932@gmail.com.

Sujata Rajesh Kadu received a Ph.D. degree in electronics and telecommunication from the University of Mumbai in 2022. She is currently an Assistant Professor at the University of Mumbai, Terna Engineering College. Her areas of research include signal processing, image segmentation and classification, object-based image analysis, and multiresolution segmentation. She can be contacted at email: sujatakadu@gmail.com.