IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 2, April 2025, pp. 1355~1362
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i2.pp1355-1362

Enhancing accessibility with long short-term memory-based sign language detection systems

Jyoti Wadmare[1], Reena Lokare[2], Ganesh Wadmare[3], Dakshita Kolte[1], Kapil Bhatia[1], Jyotika Singh[1], Sakshi Agrawal[1]
[1]Department of Computer Engineering, K. J. Somaiya Institute of Technology, Mumbai, India
[2]Department of Information Technology, K. J. Somaiya Institute of Technology, Mumbai, India
[3]Department of Artificial Intelligence and Data Science, K. J. Somaiya Institute of Technology, Mumbai, India

Article Info

Article history:

Received Feb 27, 2024
Revised Oct 25, 2024
Accepted Nov 14, 2024

Keywords:

Architecture
Long short-term memory
MediaPipe holistic API
OpenCV
Recurrent neural network model
Sign language detection

ABSTRACT

Individuals who are deaf or experience difficulties with hearing and speech predominantly rely on sign language as their medium to communicate, which is not universally comprehended leading to obstacles in achieving effective communication. Advances in deep learning technologies in recent years have enabled the development of systems intended to autonomously interpret gestures in sign language and translate them into spoken language. This paper introduces a system built on deep learning methodologies for recognizing sign language. It uses long short-term memory (LSTM) architecture to distinguish and classify hand gestures which are static and dynamic. The system is divided into three primary components, including dataset collection, neural network assessment, and sign detection component that encompasses hand gesture extraction and sign language classification. The module to extract hand gestures makes use of recurrent neural networks (RNNs) for the detection and tracking of hand movements in video sequences. Another RNN that is incorporated by classification module categorizes these gestures into established sign language classes. Upon evaluation on a custom dataset, the proposed system attains an accuracy rate of 99.42%, thus making it visualize its promise as an assistive technology for handicapped hearing individuals. This system can further be enhanced by including further classes on sign language and real-time gesture interpretation.

Corresponding Author:

Jyoti Wadmare
Department of Computer Engineering, K. J. Somaiya Institute of Technology
Somaiya Ayurvihar Complex, Eastern Express Highway, Near Everard Nagar, Sion (East)
Mumbai-400 022, India
Email: jyoti@somaiya.edu

1. INTRODUCTION

Techniques such as computer vision and machine learning are employed to recognize hand gestures, facial expressions, and body movements that are essential components of sign languages used in various geographical areas. Therefore, this technology is very important to help the deaf and non deaf communicate with people who do not understand sign language. This idea can find application in educational institutions, healthcare centers, welfare departments, and social interaction contexts by converting gesture into text or speech or other modalities of expression. Involvement of social settings with the use of sign language will break all barriers of communication and increase the level of integration.

The recent work on the detection of sign language has achieved tremendous significance for many languages, considering different methods and datasets. Models with complex features that include flow-guided and 3D and 2D deformable convolutions together with sequence attention mechanisms are proposed in order to classify the different classes of signs in the dataset from sign language correctness discrimination (SLCD) [1]. In American sign language, a vision-based system has achieved an 83% success rate in detecting continuous signs and movement epenthesis (ME) with datasets such as the Boston University database [2]. Similarly, in Pakistan sign language, the adoption of multiple kernel learning within support vector machine architectures has demonstrated enhanced accuracy and performance metrics compared to traditional methods [3].

Moreover, a hybrid deep neural network integrating 3D deep neural networks and attention-based bidirectional long short-term memory (Bi-LSTM) models has been proposed for identifying Indian and Russian sign motions, validated against an Indo-Russian sign language dataset [4]. In another context, a deep learning solution for Bangla sign language (BSL) recognition has achieved impressive accuracy rates by combining spatial, skeletal, and edge recognition techniques across multiple convolutional neural network (CNN) models and ensemble algorithms [5]. Additionally, boundary-adaptive encoder-decoder models for Chinese sign language [6], ResNet-based British sign language detection [7], supervised learning for Indian sign language [8], CNN-based Bengali sign recognition [9], and specialized CNNs for Arabic sign language (ArSL) [10] all demonstrate the field's expanding capabilities. These advancements underscore the diversity of approaches and significant strides made in sign language detection, aiming to reduce communication gaps for the deaf people across different linguistic and cultural contexts.

The proposed sign language detection system utilizes long short-term memory (LSTM) networks for accurate gesture recognition. A custom dataset comprising 5,400 image samples was curated, featuring gestures such as "hello" and "thank you". Preprocessing steps involved resizing, normalization, and noise reduction of the video data. Key features were extracted using MediaPipe Holistic, and these features were fed into the LSTM network for static and dynamic gesture prediction. The model underwent training for 600 epochs, achieving high accuracy, which was evaluated using standard metrics including accuracy and precision. Subsequently, the system was deployed for real-time detection, effectively predicting gestures captured by a camera. The document has the following structure: section 1 introduces the concept of sign language detection. Section 2 reviews various existing sign language detection systems developed for different languages. The system proposed under development is described in section 3. The outcomes and findings of the study is highlighted in section 4. Conclusions are presented in section 5.

2. RELATED WORK

The strong foundation to develop the sign language detection system came through a thorough survey of research papers which helped in understanding the need for the creation of a robust system for recognizing a broad range of sign language gestures and real-time detection. Lipi *et al.* [11] highlight BSL used in Bangladesh and parts of India, and it is critical for persons who have hearing impairment. Authors propose deep learning approach using "BSLword" dataset for BSL word recognition, which contains 30 words and 1,200 images. Using CNN and the Adam optimizer, it detects BSL static-gesture words from photos with accuracy of 92.50%. By addressing a void in BSL word recognition, this study improves assistive devices for the hard of hearing. Ismail *et al.* [12] emphasizes on the recognizing human gesture in videos as a challenging problem in computer vision due to environmental variables. Authors address using a Microsoft Kinect v2 camera to capture a collection of 20 ArSL words, together with 7,350 RGB and 7,350 depth footage. Their recognition accuracy of dynamic ArSL motions is 100%. This is accomplished using LSTM and gated recurrent unit (GRU) which merges 2D and 3D CNN with recurrent neural networks (RNNs).

Agrawal *et al.* [13] offer a real-time American sign language identification model that employs webcam images and transfer learning to achieve 98% accuracy with minimum training data. Different methods used in this model are CNN and single shot detector. It can recognize ASL letters and motions with accuracy. It is implemented in Jupyter Notebook with CUDA and CUDA deep neural network (cuDNN) for GPU acceleration which prioritizes local environment for real-time detection. Hafit *et al.* [14] focuses on Malaysian sign language (MSL) to be the primary source of communication in Malaysia with limited public knowledge and resources for learning. Authors intend to improve MSL identification by utilizing phone cameras, classified learning modules, evaluation tests, and a feedback module. It also analyzes images captured from phone camera for sign language detection and converts it into sign meaning. This strategy facilitates communication between the deaf population and the wider public. It encourages a broader understanding and acceptance of MSL.

Abdulhamied *et al.* [15] use camera action detection for real-time prediction of sign language actions such as "thank you" without external devices. By utilizing LSTM and dropout layers, it attains 99.35% accuracy. Dropout layers were employed to enhance accuracy by mitigating overfitting in deep learning models. This approach helps in improving accessibility without the need for external devices and helping the community's deaf and mute members. Ong *et al.* [16] focuses on baby sign language, a nonverbal form based on hand gestures, allows children with hearing problems to communicate and can also be taught to normal newborns before they can speak. Authors aim to work on a real-time baby sign language detecting system. It detects faces using skin colour, hands using palm center and convex hull, and analyses movements. When tested in Malaysia, it achieved over 85% accuracy for signs such as "mom," "eat," and "milk" across multiple racial categories.

Bahia and Rani [17] emphasize that sign language recognition is increasing with the use of image and video databases. The paper highlights different data gathering techniques, including kinect and goal based, leap motion, and vision-based systems, and classifies them into basic (sign characters), moderate (sign words), and advanced levels (sentence interpretation). It emphasizes how important publicly available datasets are to supporting continuous advancements in communication technologies. Monteiro *et al.* [18] state that video platforms are critical for sign language content, however finding specific videos is difficult owing to metadata constraints. Community members frequently share links manually, which may not always match the community's information demands. The enhancements focus on reducing face detection and segment processing time by 96% while minimizing accuracy loss. By utilizing a phased classifier and keyframe-based methodology, efficiency is increased, increasing the accessibility of sign language material on these sites.

Wei *et al.* [19] uses reinforcement learning to find semantic boundaries in sign language films which focuses on weakly supervised continuous sign language recognition (SLR). The method aims learning discriminative video clip representations using multi-scale perception scheme. The alignment to sign words is performed by moving the frame through gloss labels with no explicit time-bound. It uses multiscale perception for discriminative representation learning and performs well on the datasets proposed. Hu *et al*. [20] focuses on overfit of deep learning methods used recently for sign language detection due to limited data. Authors introduce SignBERT+, which combines hand posture tokens from an off-the-shelf detector with self-supervised learning to enhance sign language interpretation. Gesture state and spatial-temporal position encoding are incorporated in every token. This approach simulates detection errors and improves context awareness. SignBERT+ sets a new standard in sign language translation and recognition.

Bejuri *et al.* [21] focus on sign language for teaching mathematics in lower school more accessible, especially for those with disabilities, vocal cord infections, or spasmodic dysphonia. The purpose is to create a webcam-based sign language detecting system to recognize and decipher the hand gestures of educators to efficiently communicate ideas. High accuracy is promised by this technology. This algorithmic technique is intended to enrich knowledge and serve society by improving inclusive educational procedures. Nugraha *et al.* [22] at special school (SLB) Bina Insani Depok use a motion capture sign language translator to solve communication difficulties for students with disabilities, particularly the deaf. It integrates with the learning management system (LMS) and improves access and translation efficiency into sign language. The analysis, design, development, implementation, and evaluation (ADDIE) architecture is used in its implementation, and user acceptability testing (UAT) is used for validation. The results show improved efficiency in navigating users to the homepage of the LMS then to the sign language translator page, and altering words into gesture-based linguistic.

Purkayastha *et al*. [23] developed a sign language recognition software to help people with speech and hearing impairments in communicating more successfully. Using a CNN, the program identifies and classifies hand motions in photos before translating them into English alphabets. Constructed using CNNs, matplotlib, NumPy, OpenCV, TensorFlow, Anaconda, and Python, it guarantees reliable and easy-to-use functionality for the community of individuals with disabilities. Rajalakshmi *et al.* [24] created a sign language detection system employing an architecture of hybrid neural network to assist peoply having difficulty in hearing and speech. It uses a 3D convolutional net for static movements and a hybrid of semantic spatial and temporal data extraction for dynamic gestures. It was tested on a new Indian and Russian sign language dataset where it outperformed baseline models with accuracy of 99.76% for static and 99.85% for dynamic gestures. Alam *et al.* [25] investigates how smartphones can aid in the accessibility of individuals with speech difficulties and hearing impairments. It focuses on the use of smart devices for tasks such as voice recognition and sign language detection using machine learning and deep learning. It attempts to fill the research void and provide insightful analysis for accessible technology professionals by examining works published between July of 2012 and July of 2023.

Based on the related work in sign language detection, it is found that major research gaps include the availability of limited diverse sign language datasets and lack of real-time detection systems. These gaps hinder the development of accurate, robust sign language recognition capable of handling dynamic environments. Our project addresses these issues using deep learning and image processing techniques, offering advantages over glove-based systems.

3. METHOD

The methodology addressed shortcomings in sign language identification by curating a dataset of 5,400 gesture images captured from a video. It used an LSTM network with convolutional layers for feature extraction and GRU-based LSTM layers to capture temporal relationships. A total of 600 epochs of training were conducted, using the Adam optimizer and dropout layers to achieve regularization. The proposed system is implemented using Python and the Keras deep learning library. Figure 1 shows the workflow of the system proposed.

Different steps involved in the proposed system are as follows:

Step 1: Dataset collection - The initial stage is to acquire a dataset of sign language gestures recorded with a video camera. The custom dataset contains 5,400 images divided into six categories: hello, I love you, thank you, close, come, and eat. Each image was individually labelled and separated into training and validation sets for system evaluation.

Step 2: Data preprocessing - The dataset contains pre-processed images. Noise is reduced, and the quality is improved. This guarantees coherence and precision for further analysis.

Step 3: Feature extraction - The edges and contours of hand movements are identified as crucial features using feature extraction which were inputs to the LSTM network which improved gesture recognition in detail with precision.

Step 4: LSTM network architecture - The LSTM network, a category of RNN, is used for classification of sign languages which makes gesture predictions using many LSTM cells with feature vectors of previously processed images and stores information.

Step 5: MediaPipe holistic - The key points are extracted from the user's hand using MediaPipe holistic application programming interface (API) which is an initial step in sign language detection. It identifies different holistic key areas from the body such as hands, arms, fingers, and wrists which are utilized in feature extraction from hand gesture and are assigned to predefined sign language classes as shown in Figure 2.

Step 6: Modeling training - The sign language gesture is trained with the use of the LSTM network on the training set where feature vectors are fed into the LSTM network and the required parameters are adjusted so that the error of predicting the output is minimum. This is done for 600 epochs till proper precision is attained.

Step 7: Model evaluation - The trained LSTM network is validated with the sign language gesture validation set in which feature vectors are fed to the network and comparison is done between the gestures predicted by the network with the real gestures and assessed using different performance indicators such as accuracy, recall, precision, and F1 score.

The proposed methodology provides a strong foundation for developing sign language detection systems in real-time with the combination of accurate hand motion analysis and deep learning. It helps in improving accessibility in different applications and helps in enhancing communication in the deaf community.

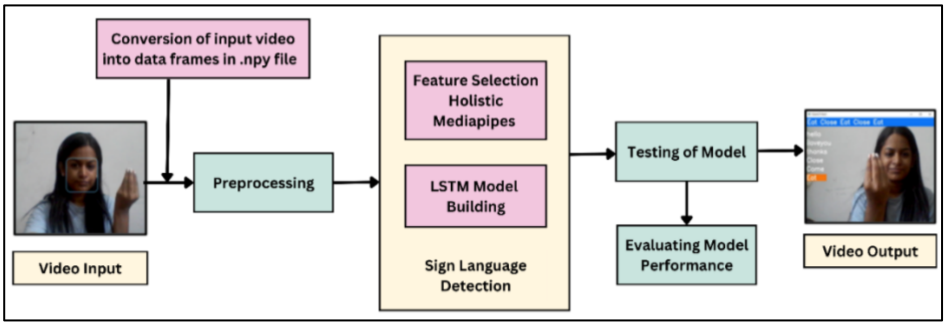


Figure 1. Flow diagram of the system proposed

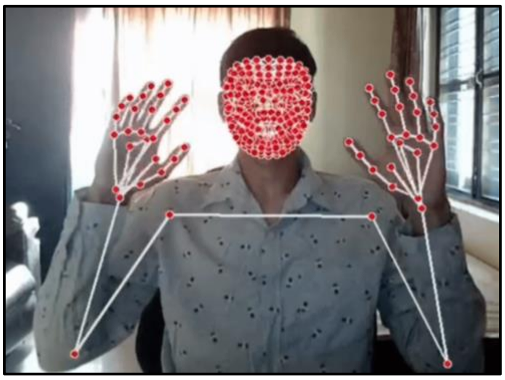
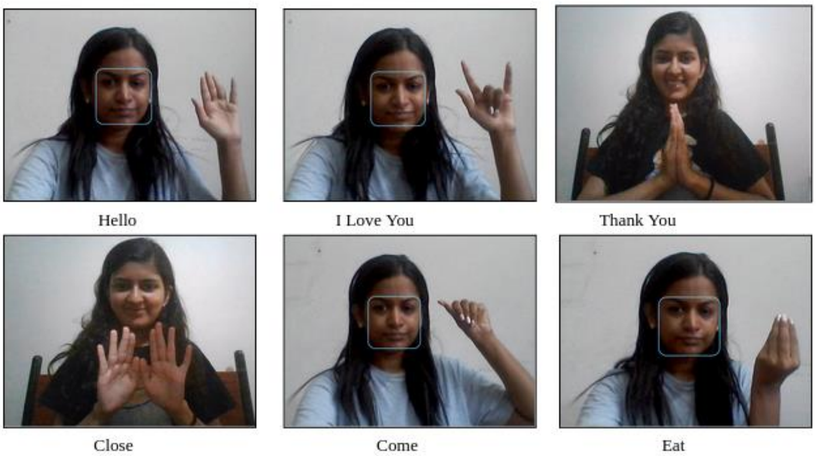


Figure 2. MediaPipe holistic keypoints

4. RESULTS AND DISCUSSION

After training, the LSTM network is utilized in real-time to detect and recognize sign language gestures by processing a sequence of feature vectors extracted from pre-processed images as input and generating the predicted gesture as output. The camera is capturing the hand gestures images, which are pre processed to extract the features. The features are fed into the LSTM network, which predicts the sign language gesture.

Figure 3 indicates the sample dataset for each class along with its output respectively. Figure 3(a) displays the sample dataset consisting of six sign language classes that is given input to the system. These include hello, I love you, thank you, close, come, and eat. Each class includes video signs of a person performing the corresponding sign language gesture. Figure 3(b) depicts the output for class 'eat' which is a closed hand with fingers touching the thumb and moving to the mouth to represent the act of eating.



(a)

(b)

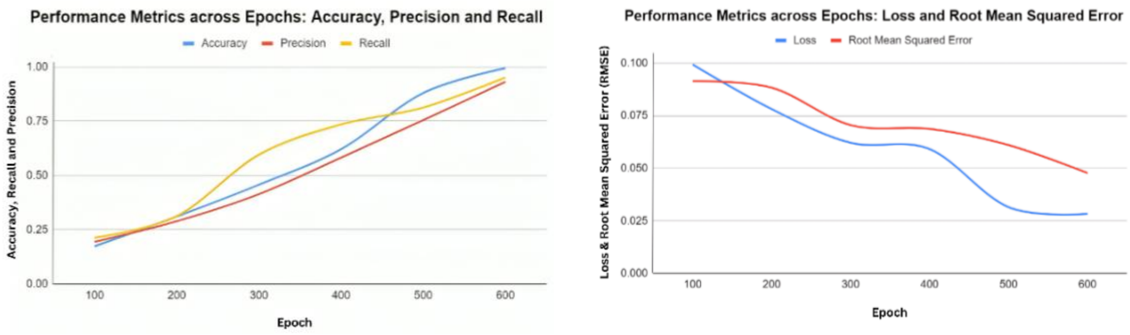
Figure 3. Sample dataset for each class with its output (a) sample dataset and (b) output classes

The proposed system was assessed using the performance indicators such as accuracy, precision, recall, loss, and root mean squared error, as shown in Table 1, with parameter values changing every 100 epochs to indicate the training progress of the proposed sign language detection system using deep learning techniques. With an accuracy of 99.42%, the system demonstrated a high level of accuracy in identifying motions in sign language. The precision and recall ratings for all sign language classes were also high, demonstrating that the system can reliably classify hand motions across all sign languages.

Figure 4 represents the model performance metrics over 600 epochs. Figures 4(a) and 4(b) illustrate the progression of accuracy, precision, recall, and loss/root mean squared error across epochs for the sign language detection system proposed. The graphs depict a consistent upward trend in accuracy, precision, and recall, starting from 17.42% and reaching 99.42%. Conversely, the loss and root mean squared error exhibit a steady decline, indicating improved model performance with training epochs.

Table 1. Performance metrics for every 100 epochs

Epoch	Accuracy	Loss	Precision	Recall	Root mean squared error
100	0.1742	0.0993	0.1944	0.2133	0.0913
200	0.3109	0.0781	0.2891	0.3124	0.0883
300	0.4574	0.0621	0.4141	0.5954	0.0705
400	0.6211	0.0591	0.5822	0.7350	0.0687
500	0.8798	0.0315	0.7549	0.8125	0.0610
600	0.9942	0.0283	0.9312	0.9514	0.0477



(a) (b)

Figure 4. Model performance metrics over 600 epochs of (a) accuracy and (b) loss

4.1. Comparison with existing system

Table 2 provides a comparative overview of the existing sign language detecting technologies. It outlines the methods used, the datasets used, and the corresponding accuracy. This analysis is used to compare performance across various approaches in the field. Finally, the varied range of methodologies presented in Table 1 reveals considerable advances in sign language detection. From basic models like hidden Markov models (HMM) and support vector machines to advanced deep learning architectures like CNN and LSTM, these systems use specific datasets to achieve accuracies ranging from 62.63 to 99.42%. These achievements highlight the movement toward more accurate and accessible technologies for improving communication and inclusivity in difficult of hearing communities.

Table 2. Summary of different sign language detection techniques for different languages and their accuracies

Technique/model used	Language detected	Dataset	Accuracy (%)
Faster R-CNN with feature pyramid networks (FPN) and flow-guided feature [1]	Various sign languages	SLCD dataset: 20,792 sign language videos and 1,054,598 pictures	96.36
Multi-layered CNN [11]	British sign language	BSLword dataset: 1,200 images	92.50
LSTM [15]	American sign language	ASL dataset: 87,000 images	99.35
CNN [16]	Baby sign language	900 images (36 characters, 25 samples each)	98.05
Glove-based, kinect-based, leap motion, vision-based [25]	Indian sign language	ISLRTC datasets (Indian sign language): 36,000 images	62.63
LSTM network with convolutional layers for feature extraction and GRU-based LSTM layers – Our method	English sign language	5,400 images	99.42

5. CONCLUSION

In conclusion, deep learning approaches, notably LSTM networks, have proven to be extremely effective for detecting sign language motions, with an accuracy of 99.42%. Compared to conventional glove-based systems, this method has several advantages, including increased flexibility in various backdrops and the removal of the requirement for certain color bands. The system's effectiveness highlights how much better communication between hearing and hearing-impaired people may be achieved. Future scope should focus on enhancing system accuracy by broadening the dataset to include more signs and gestures. Furthermore, investigating CNN integration with hybrid models may improve performance even further. The system's capacity extends to recognize multiple sign languages and connecting it with speech recognition systems would allow for more complete communication solutions. Real-time applications in classrooms, public spaces, and workplaces have the potential to greatly improve accessibility for the hearing impaired, highlighting continuous prospects to advance and deploy deep learning techniques in sign language identification.

REFERENCES

[1] M. Zhang, S. Yang, and M. Zhao, "Deep learning-based standard sign language discrimination," *IEEE Access*, vol. 11, pp. 125822–125834, 2023, doi: 10.1109/ACCESS.2023.3330863.

[2] A. K. Talukdar and M. K. Bhuyan, "Vision-based continuous sign language spotting using gaussian hidden markov model," *IEEE Sensors Letters*, vol. 6, no. 7, pp. 1–4, Jul. 2022, doi: 10.1109/LSENS.2022.3185181.

[3] F. Shah, M. S. Shah, W. Akram, A. Manzoor, R. O. Mahmoud, and D. S. Abdelminaam, "Sign language recognition using multiple kernel learning: a case study of pakistan sign language," *IEEE Access*, vol. 9, pp. 67548–67558, 2021, doi: 10.1109/ACCESS.2021.3077386.

[4] E. Rajalakshmi *et al.*, "Multi-semantic discriminative feature learning for sign gesture recognition using hybrid deep neural architecture," *IEEE Access*, vol. 11, pp. 2226–2238, 2023, doi: 10.1109/ACCESS.2022.3233671.

[5] K. A. Shams *et al.*, "Multimodal ensemble approach leveraging spatial, skeletal, and edge features for enhanced bangla sign language recognition," *IEEE Access*, vol. 12, pp. 83638–83657, 2024, doi: 10.1109/ACCESS.2024.3410837.

[6] S. Huang and Z. Ye, "Boundary-adaptive encoder with attention method for Chinese sign language recognition," *IEEE Access*, vol. 9, pp. 70948–70960, 2021, doi: 10.1109/ACCESS.2021.3078638.

[7] U. Saeed *et al.*, "British sign language detection using ultra-wideband radar sensing and residual neural network," *IEEE Sensors Journal*, vol. 24, no. 7, pp. 11144–11151, Apr. 2024, doi: 10.1109/JSEN.2024.3364389.

[8] L. V. Srininvas, C. Raminaidu, D. Ravibabu, and S. S. Reddy, "A framework to recognize the sign language system for deaf and dumb using mining techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 1006–1016, Feb. 2023, doi: 10.11591/ijeecs.v29.i2.pp1006-1016.

[9] A. A. Alani and G. Cosma, "ArSL-CNN a convolutional neural network for Arabic sign language gesture recognition," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, pp. 1096–1107, May 2021, doi: 10.11591/ijeecs.v22.i2.pp1096-1107.

[10] F. M. J. M. Shamrat, S. Chakraborty, M. M. Billah, M. Kabir, N. S. Shadin, and S. Sanjana, "Bangla numerical sign language recognition using convolutional neural networks (CNNs)," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, pp. 405–413, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp405-413.

[11] K. A. Lipi, S. F. K. Adrita, Z. F. Tunny, A. H. Munna, and A. Kabir, "Static-gesture word recognition in Bangla sign language using convolutional neural network," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20, no. 5, pp. 1109–1116, Oct. 2022, doi: 10.12928/telkomnika.v20i5.24096.

[12] M. H. Ismail, S. A. Dawwd, and F. H. Ali, "Dynamic hand gesture recognition of Arabic sign language by using deep convolutional neural networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 2, pp. 952–962, Feb. 2022, doi: 10.11591/ijeecs.v25.i2.pp952-962.

[13] A. K. Agrawal, J. Kumar, A. Kumar, and P. Arvind, "Analogous sign language communication using gesture detection," *Australian Journal of Electrical and Electronics Engineering*, vol. 21, no. 4, pp. 486–498, Oct. 2024, doi: 10.1080/1448837X.2024.2337495.

[14] H. Hafit, C. W. Xiang, M. M. Yusof, N. Wahid, and S. Kassim, "Malaysian sign language mobile learning application : A recommendation app to communicate with hearing-impaired communities," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 6, pp. 5512–5518, Dec. 2019, doi: 10.11591/ijece.v9i6.pp5512-5518.

[15] R. M. Abdulhamied, M. M. Nasr, and S. N. A. Kader, "Real-time recognition of American sign language using long-short term memory neural network and hand detection," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 1, pp. 545–556, Apr. 2023, doi: 10.11591/ijeecs.v30.i1.pp545-556.

[16] N. T. J. Ong, A. S. H. Ali, and Y. A. H. M., "Employment of baby sign language in interactive learning for hearing-impaired people in Malaysia," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 361–367, Apr. 2020, doi: 10.11591/ijeecs.v18.i1.pp361-367.

[17] N. K. Bahia and R. Rani, "Multi-level taxonomy review for sign language recognition: emphasis on Indian sign language," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 1, Jan. 2023, doi: 10.1145/3530259.

[18] C. D. D. Monteiro, F. M. Shipman, S. Duggina, and R. Gutierrez-Osuna, "Tradeoffs in the efficient detection of sign language content in video sharing sites," *ACM Transactions on Accessible Computing*, vol. 12, no. 2, Jun. 2019, doi: 10.1145/3325863.

[19] C. Wei, J. Zhao, W. Zhou, and H. Li, "Semantic boundary detection with reinforcement learning for continuous sign language recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 3, pp. 1138–1149, Mar. 2021, doi: 10.1109/TCSVT.2020.2999384.

[20] H. Hu, W. Zhao, W. Zhou, and H. Li, "SignBERT+: Hand-model-aware self-supervised pre-training for sign language understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 11221–11239, Sep. 2023, doi: 10.1109/TPAMI.2023.3269220.

[21] W. M. Y. W. Bejuri, N. N. Zakaria, M. M. Mohamad, W. M. Yassin, S. S. S. Ahmad, and N. H. Choon, "Sign language detection using convolutional neural network for teaching and learning application," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 1, pp. 358–364, Oct. 2022, doi: 10.11591/ijeecs.v28.i1.pp358-364.

[22] D. Nugraha, S. Faizah, and M. Zaenudin, "Enhancing interaction and learning experience for deaf students through sign language translator," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 3, pp. 1730–1738, Jun. 2024, doi: 10.11591/ijeecs.v34.i3.pp1730-1738.

[23] S. Purkayastha, M. J. Roy, and N. Pradhan, "A demonstration for the detection of sign language," *Journal of Information and Optimization Sciences*, vol. 40, no. 8, pp. 1611–1621, Nov. 2019, doi: 10.1080/02522667.2019.1704519.

[24] E. Rajalakshmi *et al.*, "Static and dynamic isolated Indian and Russian sign language recognition with spatial and temporal feature detection using hybrid neural network," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 1, pp. 1–23, Jan. 2023, doi: 10.1145/3530989.

[25] I. Alam, A. Hameed, and R. A. Ziar, "Exploring sign language detection on smartphones: a systematic review of machine and deep learning approaches," *Advances in Human-Computer Interaction*, vol. 2024, pp. 1–36, Mar. 2024, doi: 10.1155/2024/1487500.

BIOGRAPHIES OF AUTHORS

Dr. Jyoti Wadmare is Assistant Professor in Department of Computer Engineering, at KJSIT. She has teaching experience of 17 years with an AI background. Her major domain of interest is the conjunction of AI and computer vision. Testimonials of work includes many conferences' presentations and articles published that quite clearly states advancement in this area by her and has filed a patent and acquired four copyrights. She can be contacted at email: jyoti@somaiya.edu.

Dr. Reena Lokare is Assistant Professor in Department of Information Technology at KJSIT as well as Ph.D. scholar in University of Mumbai having 21 years of teaching experience. She has done several publications in national and international conferences. Her areas of interest are artificial intelligence, bioinformatics, automata theory, data structures and algorithms, and soft computing. She can be contacted at email: reena.l@somaiya.edu.

Ganesh Wadmare is an Assistant Professor in the Department of Artificial Intelligence and Data Science at KJSIT and a Ph.D. scholar in Savitribai Phule Pune University, with academic experience for over 19 years. He has extensive exposure and experience in the field of artificial intelligence and renewable source of energy. He has published his research papers in both national and international conferences. He can be contacted at email: gwadmare@somaiya.edu.

Dakshita Kolte is studying B.Tech. in computer engineering at KJSIT. She has developed artificial intelligence machine learning, solutions that combine artificial intelligence (AI), and web-based technologies. She has a track record for participating in some of the most renowned competitions such as mastek project deep blue, aavishkar, and creative ideas and innovations in action. She has four copyrights for her work. She can be contacted at email: d.kolte@somaiya.edu.

Kapil Bhatia working on artificial intelligence and machine learning, has opted for B.Tech. in computer engineering at KJSIT. He is well adept at developing solutions that are the interplay of web-based technologies, the internet of things, and artificial intelligence. His participation in the well-known contests such as aavishkar, mastek project deep blue, and creative ideas and innovations in action speak volumes about his exceptional expertise. He has also bagged four copyrights on his work. He can be contacted at email: kapil.bhatia@somaiya.edu.

Jyotika Singh is Security Researcher-I at Forcepoint. She has done B.Tech. in computer engineering from K. J. Somaiya Institute of Technology, Mumbai, Maharashtra, India. She has interests in machine learning and data science with a good network in the field. She can be contacted at email: jyotika.s@somaiya.edu.

Sakshi Agrawal is a data science enthusiast who has pursued B.Tech. in computer engineering from K. J. Somaiya Institute of Technology, Mumbai, Maharashtra, India. Her areas of interest include Python, databases, and machine learning with strong experience in those areas. She can be contacted at email: sakshi.ga@somaiya.edu.