Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 4, August 2026, pp. 3631~3645
ISSN: 2302-9285, DOI: 10.11591/eei.v15i4.11395

An assistive communication system for mute individuals using MATLAB-based image processing and sensor integration

Mohan P. Thakre[1], Krupali Kanekar[2], Badal Kumar[3], Alok Kumar[3], Supriya Nilesh Thakur[4], Pranali M. Thakre[5], Prashant K. Magadum[6]

[1]Department of Electrical Engineering, Walchand College of Engineering, Sangli, India
[2]Department of Computer Science Engineering, Ramrao Adik Institute of Technology, D. Y. Patil Deemed to be University, Navi Mumbai, India
[3]Department of Electrical Engineering, School of Electrical and Communication Sciences, JSPM University, Pune, India
[3]Department of Electrical Engineering, Sandip Institute of Technology and Research Centre, Nashik, India
[4]Department of Electrical Engineering, Marathawada Mitra Mandal's College of Engineering Karvenagar, Pune, India
[5]Department of Bachelor Computer Science, Willingdon College, Sangli, India
[6]Department of Electrical Engineering, Brahmdevdada Mane Institute of Technology, Solapur, India

Article Info

Article history:

Received Sep 14, 2025
Revised May 17, 2026
Accepted Jul 14, 2026

Keywords:

Arduino
Hand gestures
Image processing
MATLAB
Raspberry Pi

ABSTRACT

Interactions between non-verbal people and verbal communicators have always presented serious challenges. Although functional, current sign language systems often lack the flexibility and affordability needed for wider use. This document provides a novel speaking system created for those who are mute, supporting communication using two methodologies: i) recognition of hand motions with flex sensors and ii) detection of finger positions through image processing techniques combined with MATLAB. The setup in question uses a single camera and a Raspberry Pi to capture hand photos. Finger coordinates are then extracted from these photos and compared to an existing database for matching. An appropriate verbal output is then produced from the identified gesture. Flex sensor-based inputs are simultaneously connected to an Arduino, which uses a Bluetooth module to transmit text data to an external application for speech conversion. This dual-mode technology improves everyday communication for people who are silent by providing a customized and efficient framework for converting sign-based gestures into vocal communications. This approach makes a significant contribution to the development of inclusive technology by offering an affordable, adaptable, and practical alternative for assistive communication.

Corresponding Author:

Mohan Thakre
Department of Electrical Engineering, Walchand College of Engineering
Sangli, Maharashtra, India
Email: mohanthakre@gmail.com

1. INTRODUCTION

Assistive communication technologies are significant in enhancing interactions of speech and hearing-impaired persons. Among language and speech impaired (LIS) users, sign language is one of the most common means of communication. Sign languages take place in different regions of the world and are not always universal, however, there are some sign languages which are understood by a different group such as Indian sign language (ISL), American sign language (ASL), and British sign language (BSL). This causes difficulties during normal times of interaction between individuals who use signs and those who do not within the learning environment, health and public service provision, and social interaction. To decrease this

gap, a gesture to text and gesture to speech system has been introduced by the researchers which are used to recognize gestures through sensor, vision or hybrid sensor and vision-based systems [1].

The system for recognizing sign language has to fulfill a number of important criteria for a practical system. If possible, it should be inexpensive, simple, portable, accurate, and real-time. Meanwhile, it must be user-friendly and work across a variety of environmental circumstances. Many current systems are unable to meet these needs. These systems are simple, inexpensive and involve wearing devices. Vision-based systems are non-invasive but also have sensitivity to light conditions, background noise, camera quality, as well as computational constraints. These difficulties render the challenge of developing efficient and affordable assistive communication system as an important research problem.

In sensor-based approaches, the simplicity, portability and low cost of the hardware means that there have been a large number of studies carried out. Glove-based systems have been developed employing flex sensors and microcontrollers, which have the ability to recognize finger bending patterns and translate gestures into text and speech. For instance, a flex sensor-based glove and an Arduino Nano were used to accomplish gesture recognition, which was found to be accurate as 83.58% and to give audible messages [2]. Arduino Uno, Bluetooth modules, and a flex sensor have been used in similar solutions to send recognized gestures to a mobile phone as either text or speech with an assistive communication capability [3] which is of a low cost. In other works, the integration of flex sensors with accelerometers and application with Android system has been proposed for gesture classification with higher accuracy [4]. In the domain of gesture recognition, machine learning algorithms have also been utilized to identify human gestures from the sensor, and the best result was obtained by adversarial learning algorithm that achieved a recognition rate of 88.32% [5]. These systems offer a good recognition accuracy, however relying on remaining attached to the body, in the case of wearing a glove may decrease user acceptance, long-term usability and comfort.

Because of the lack of need for any such wearable devices, vision-based approaches have been the focus of attention. These systems are based on a camera that is able to capture hand movements and perform recognition by means of image processing or deep learning (DL) [6], [7]. A typical pipeline for the use of vision is the acquisition of an image, using some sort of pre-processing, feature extraction, classification, and comparing the extracted features with trained sets of data. Real time gesture recognition has also been achieved in small systems with embedded platforms like the Raspberry Pi [8]-[10]. To tackle these challenges, computer vision libraries (OpenCV) and DL frameworks (TensorFlow) have been adopted to facilitate real-time image processing, object detection and gesture classification [11]-[13]. There are also advanced detection models like you only look once (YOLO), single shot multi box detector (SSD), and Faster region-based convolutional neural network (R-CNN), which further improved real-time recognition performance, especially the YOLO based model is suitable for the process of high-speed detection [14]-[17]. In recent years, the architectures like YOLOv5, YOLOv8, and convolutional neural networks (CNN)-Transformer have enhanced the performance of detection accuracy and processing efficiency. But the variability of light, background noise, camera position, processing latency and the ability to perform the heavy processing required by the embedded low-cost devices remains a problem with vision-based systems [18], [19].

Combining several sensing modalities to address the drawback of each modality is called hybrid approach, which is a compromise between simply using any one of them. With such systems it is possible to achieve greater reliability when one recognition mode can cope with the deficiencies of the other. Cybernetwork enhances recognition by supplementing with sensor data, for instance, when the conditions in the camera are not suitable; in an appropriate environment rich in visual information, vision-based recognition has the potential to reduce the reliance on the input of a wearable sensor [20], [21]. Although it may be possible, most of the studies are based on sensors or cameras for recognition currently available [22], [23]. There is limited research on the development of low-cost combination of these two approaches with a final system that integrates both methods for real-time assisted communication [24], [25]. In this paper, the main research gap that is solved is the development of a hybrid assistive communication system with vision-based and sensor-based gesture recognition which is efficiently, low cost and adaptable. Thus, the main research gap solved in this paper is an efficiently, low cost, and adaptable hybrid assistive communication system, which is based on vision and sensor-based gesture recognition. With many systems, cost, comfort, accuracy, and speed of operation are commonly tested against one another. Combining a flex-sensor-based gesture detection system with camera-based image recognition facilitates enhanced usability and reliability for various user specifications and environmental conditions.

In this paper, a dual modality assistive communication system with wearable sensor input and vision-based recognition is introduced. The system is a proposed system based on flex-sensors attached to an Arduino micro-controller to detect the movement of the gestures in a plane with low cost sensors and an image-based recognition system based on a Raspberry Pi vision module. A hybrid design involves supplementing the first mode of operation with a second that will be as complementary as possible, to enhance the adaptability.

The major contributions in this work are the development of dual mode assistive communication system incorporating flex sensors, the Arduino microcontroller, the Raspberry pi and the camera module which is performed at low cost. The implemented system comprises a gesture recognition module that recognizes the hand gestures and translates the recognized gestures back to meaningful communication output. Moreover, a gesture or visual input recognition module based on vision is also implemented by means of image processing on embedded platform to assist the non-contact gesture and visual input recognition. The proposed hybrid architecture integrates sensor-based and vision-based techniques, providing enhanced usability, reliability, and adaptability across various operating conditions. The system has practical viability as a low cost, portable and easy-to-use assistive communication device for people with speech or hearing disabilities. The system suggested in this paper would try to facilitate better communication between people with hearing or speech impairments and the rest of the community. The multimodal recognition of the present invention uses low cost hardware and can be implemented to create an accessible, practical and flexible assistive communication technology.

The rest of this paper is organized as follows: in section 2, one gets a general overview of the system and overlook at basic concepts. The proposed system is described in terms of hardware development in section 3. The algorithm and flowchart are described in section 4. The results of the implementation are explained in section 5. Finally, section 6 presents the conclusion and possible directions for future research.

2. METHOD

2.1. Proposed system architecture

The idea behind the proposed system is that it will be a dual mode assistive communication system for hand gesture-to-text and hand gesture-to-speech conversion. This system comprises two working complementary recognition modules: sensor-based gesture recognition module and vision-based gesture recognition module. The sensor-based module is based on flex sensors placed on a wearable glove that capture finger movements, while the vision-based module is based on a camera connected to a Raspberry Pi, which analyses the images and detects the hands' gestures.

The overall design is presented in Figures 1 and 2. In this figure shows the block diagram of a sign language to speech system by the help of flex sensor, Arduino Uno, Bluetooth communication module (HC-05) Bluetooth module and mobile speaker. The proposed sign-language-to-text system consists of four components: image acquisition, sign-language dataset preparation, YOLO-based sign-language gesture detection, and text synthesized on an organic light emitting diode (OLED) screen. They are shown together in the Figure 2. Dual mode makes the system more flexible, as either direct sensor input or visual input (via the camera) can be used to trigger a gesture recognition regardless of the user's choice and the environment.

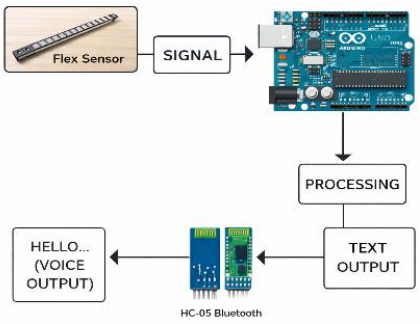

Figure 1. Block diagram of the sign language to speech and text output system

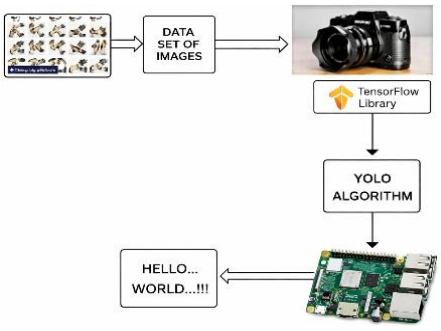

Figure 2. Block diagram of sign language to text output

The main components of the system to be developed are five flex sensors, an Arduino Uno microcontroller, the HC-05 Bluetooth module, the RPI Board, the camera module, the OLED display, and the mobile speaker is given in Table 1. The flex sensors are fixed to the fingers of the glove to detect bending of the flexer while the glove is in the gesture process. Each flex sensor will output a change of resistance depending on the bending angle of the hands' finger. This resistance variation is converted to an analog signal voltage which is fed to analog input pins on the Arduino Uno.

The operating circuit of flex-sensor based module is shown in Figure 3. A 10 kΩ resistor is used to create a voltage divider circuit with each flex sensor and then to the Arduino. The sensor is connected between one terminal with ground, and the other is connected with the supply voltage through the resistor.

The analog voltage values are passed to the Arduino who converts them into a digital value using its built-in analog-to-digital converter and applies a predetermined threshold to each detected gesture. The HC-05 Bluetooth module is connected to the serial communication pins of Arduino and sends the recognized gesture output to the mobile device or speaker.

The vision-based module is based on raspberry Pi and camera module, taking real-time pictures of hand gesture. Image processing and object detection techniques are used to process the captured images. The recognized gesture is shown as text on OLED display. Raspberry Pi enables the system to work as a miniaturized embedded system for real-time gesture recognition.

Table 1. Hardware specifications of the proposed system

Component	Purpose in the proposed system
Flex sensors	Measure finger bending and generate analog signals
Arduino Uno	Reads sensor values, performs threshold-based gesture recognition, and controls Bluetooth output
HC-05 Bluetooth module	Transmits recognized gesture data to a mobile device
Mobile speaker/text-to-speech (TTS) application	Converts recognized text into speech output
Raspberry Pi	Performs image acquisition and vision-based gesture recognition
Camera module	Captures real-time hand gesture images
OLED display	Displays recognized text output
10 kΩ resistors	Used in voltage divider circuits for flex sensor interfacing

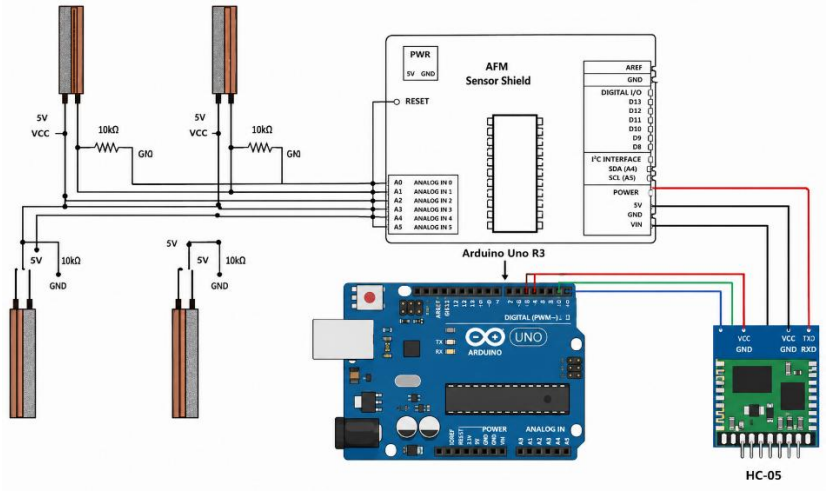

Figure 3. Speaking system circuit diagram using flex sensor

3. SENSOR AND VISION BASED GESTURE RECOGNITION

3.1. Sensor-based

The recognition process starts when the user executes a hand gesture present each time he puts on the glove. Each flex sensor's resistance changes with the bending of its respective finger. These resistance changes produce analog voltage signals that: i) are read by the Arduino Uno analog inputs and ii) create a basis for truth tables. The analog voltage signals which are outputted by these resistances are read by the analog input pins of the Arduino Uno, which also will form the base for the truth table.

Before classification, there is a stabilization process in which multiple readings take place which have been averaged together. The filtering eliminates noise, hand tremors and sudden changes in signal. Then the Arduino dispositive compares the processed values from the sensors with the ranges of values that are pre-programmed into the microcontroller. There is a different configuration of voltage ranges across the five flex sensors for each gesture determination. When the detected values are identical to a template or set of values, the associated speech command is produced.

Once recognized, the gesture information is sent using HC-05 Bluetooth module to a mobile device that is connected to the module. A mobile device or speaker system is used to read the received text to the user by means of a text-to-speech application. When the value from the sensor does not match any pre-programmed pattern, it keeps reading the value of the input until it finds one that matches a required gesture. The operation of this module is depicted in Figure 4.

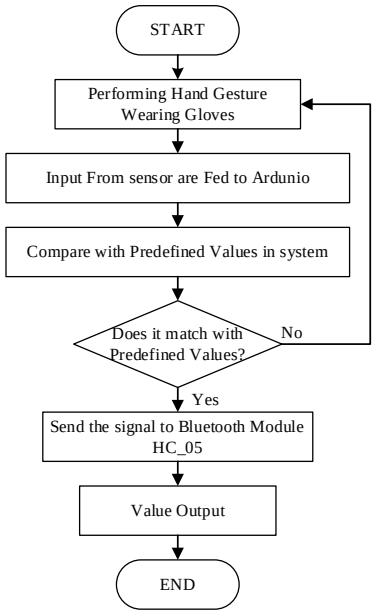

Figure 4. Steps for system using Arduino microcontroller

3.2. Vision-based

The recognition module of the vision has recorded the image of hand gestures using the camera of Raspberry Pi. Captured frames are used to process by object detection model based on YOLO. Figure 5 shows the complete processing workflow, which consists of image acquisition, preprocessing, annotation, generated dataset, model training, prediction, non-maximum suppression and output generation.

Resized and normalized captured images are fed into detection model. Data augmentation has been used to enhance model stability and less prone to over-fitting. The augmentation method is affine transforms and flipping the images horizontally. The gesture images are annotated in YOLO format, which is an annotation file that has the label of gesture class and the box coordinates of the gesture. The annotated dataset is then fed into the model training and testing. At detection, the trained model YOLO would return the class of gesture and the bounding box in one forward pass. As in Figure 6, the input image was segmented into several grid regions, each grid region directly nets out the probability of each class of object and the bounding box information.

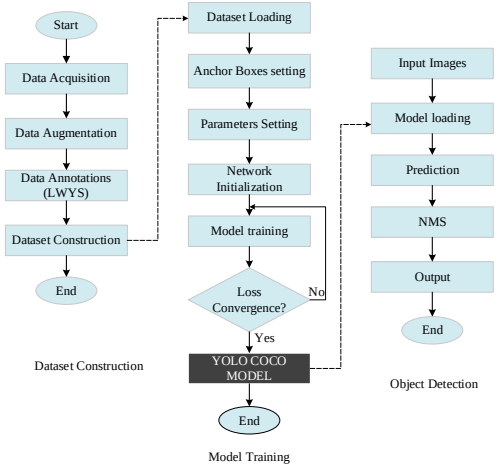

Figure 5. Flowchart of YOLO

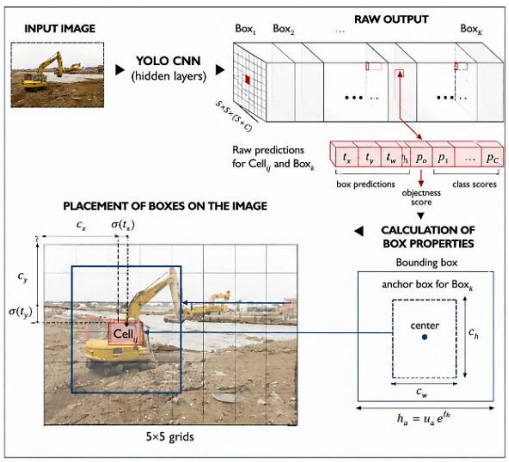

Figure 6. Working block diagram of YOLO

The output of each prediction grid was shown in Figure 7; where p C indicated that the probability score of class C. The output vector was used for the non-maximum suppression to delete the multiple detection image and leave the single best prediction. Then, the predicted gesture would be types of in text form on the OLED screen as in the vision-based workflow as in Figure 8.

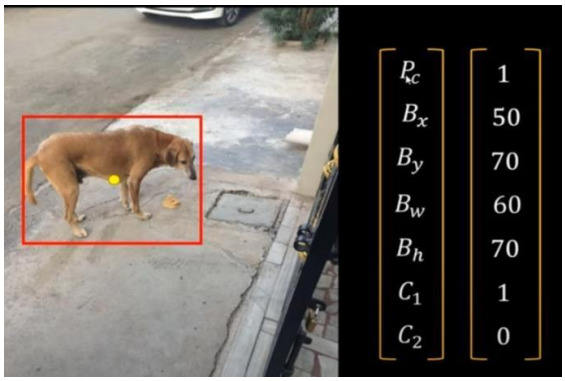

Figure 7. Output vector representation of YOLO

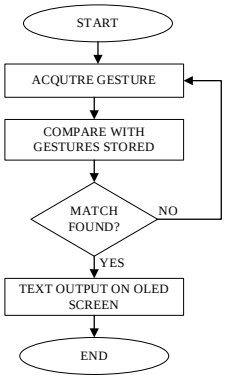

Figure 8. Steps for system using image processing

3.3. Dataset preparation and annotation

The data set of the vision-based module is made up of roughly 300-400 gesture images, which is split evenly according to gesture classes and each with about 20-25 images. Open-source annotation software (YOLO labeling format) is used to annotate the gesture images. At the beginning of each annotation file, the total number of gestures and the gesture classification id is entered with coordinates of the gesture's bounding box. Hands gesture classes are represented by class IDs from 1 to 5 which refer to the hand gesture classes selected.

In order to enhance the performance of the training, there are images that are produced based on data augmentation. Further, horizontal flipping and affine transformations to enhance the diversity of the dataset that is used to train the model, in order to make the model more generalizable. This final dataset is split into 3 subsets: training, validation and testing is given in Table 2.

Table 2. Dataset description

Parameter	Description
Total images	350 images after augmentation
Number of classes	5 gesture classes
Images per class	20–25 images per class
Final images per class	70 images per class after augmentation
Annotation format	YOLO format
Annotation tool	makesense.ai open-source annotation tool
Augmentation methods	Affine transformation and horizontal flipping
Training split	245 images, 70%
Validation split	52 images, 15%
Testing split	53 images, 15%
Image size	416×416

4. YOLO MODEL CONFIGURATION AND TRAINING

4.1. Model training

A YOLO object detection model is used to train the vision-based gesture recognition model. YOLO is chosen model since its object localization and classification requires only one forward pass, making it suitable for real-time embedded applications. The model's spatial abstraction of the image is done using convolutional layers as illustrated in Figure 9. The YOLO architecture in turn predicts the object class, the coordinates of the bounding box, and the confidence score.

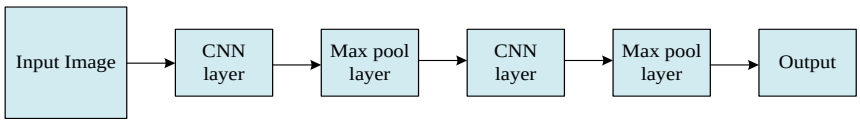

Figure 9. Block diagram of CNN networking

This training process starts with loading the data sets, setting anchor boxes, initialization of parameters and training the network. The anchor boxes are generated to enhance the ability of using the accuracy of object localization. Model is trained iteratively until convergence of the loss function. The loaded

trained model is communicated to the embedded platform for real-time prediction is given in Table 3. Inference happens when the image to be classified is fed into the YOLO network and the final prediction is attained with the application of non-maximum suppression.

Table 3. Model training configuration

Parameter	Value
Detection model	YOLO object detection model
YOLO version	YOLOv3-Tiny with common objects in context (COCO) pre-trained weights
Framework	TensorFlow 2.8.0
Input image size	416×416 pixels
Batch size	16
Number of epochs	100
Learning rate	0.001
Optimizer	Adam optimizer
Loss function	YOLO detection loss using bounding box loss, objectless loss, and classification loss
Confidence threshold	0.50
Intersection over union (IoU) threshold for NMS	0.45
Hardware used for training	Intel Core i5 system with 8 GB RAM and NVIDIA GTX 1650
Hardware used for inference	Raspberry Pi 4 Model B with 4 GB RAM

4.2. Software implementation

The sensor-based module is programmed using Arduino IDE. The five flex sensors provide analog sensor input and the analog values they provide are then preprocessed using a process of averaging followed by comparing the analog values against the known range of analog values for each flex sensor to recognize gestures (pattern recognition), and then send the recognized gesture over the HC-05 Bluetooth module. A smartphone application uses the Bluetooth receiver to interpret text and switch it to speech.

The image processing part of the module is done with Python modules. Model training and inference is using TensorFlow, and image acquisition, resizing, frame processing, and display operations can be completed by OpenCV. When speech feedback is needed, the google text-to-speech (gTTS) library is used for the text-to-speech conversion. There are two simulations of the flex-sensor circuit in TinkerCAD - one in which the flex is actuated, in "Live Construction", as shown in Figure 10, and another in which the flex is not actuated, in "Pulse Construction", as illustrated in Figure 11. The software information included for greater reproducibility is given in Table 4.

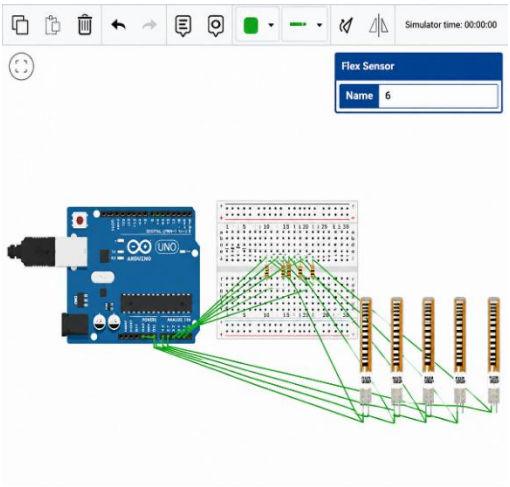

Figure 10. Simulation circuit connections

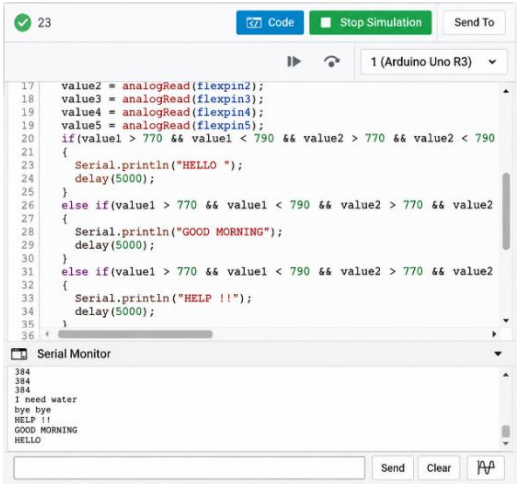

Figure 11. Simulation results

4.3. Evaluation and working procedure

The performance of the proposed system is being evaluated on each module independently for sensor-based as well as vision based. For the module with sensor, several times of each nominated gesture, the person wearings the glove should perform that gesture. The flex sensor data is used to compare and store the reading in the Arduino microcontroller in predefined ranges of flexibility. A gesture is correctly recognized when the generated text or speech output is the same as the gesture performed.

Table 4. Software specifications

Software/Library	Purpose
Arduino IDE 2.2.1	Programming and uploading code to the Arduino Uno microcontroller
Python 3.8.10	Vision-based processing, YOLO execution, and speech conversion
TensorFlow 2.8.0	YOLO-based gesture model training and inference
OpenCV 4.5.5	Image capture, frame resizing, preprocessing, and real-time video handling
gTTS 2.2.4	Conversion of recognized text output into speech
TinkerCAD	Simulation of the Arduino Uno and flex sensor circuit
Operating system	Windows 10 64-bit for training, Arduino programming, and simulation
Raspberry Pi OS version	Raspberry Pi OS Bullseye 64-bit
Python version	Python 3.8.10
TensorFlow version	TensorFlow 2.8.0
OpenCV version	OpenCV 4.5.5

The test set consists of unseen gesture images or camera frames in the realtime camera frames for vision-based module. The predicted class is compared with the ground-truth label (True Class). The evaluation takes both accuracy of classification and real-time response. Furthermore, detection performance is measured with our confidence score, mean average precision (mAP) and inference speed. The following evaluation metrics should be formally reported:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision \times Recall}$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives. For object detection and mAP, confidence score, IoU and also frames per second (FPS) should be reported. These indices give a formal criterion to compare the recognition accuracy, detection performance and real-time speed.

The complete working procedure of the proposed system is summarized:

- The user performs a hand gesture using either the sensor glove or the camera-based interface.
- In the sensor-based mode, the flex sensors sense the bend of the finger and uses analog signal send to the Arduino Uno.
- The Arduino translates the analog data to digital data, averages the received data and checks it against the gesture limits.
- If a proper gesture has been initialized, the matching text will be sent to speech through HC-05 Bluetooth.
- In the vision-based mode, the Raspberry Pi camera captures the hand gesture image.
- The captured image is pre-processed and passed to the trained YOLO model.
- On detection, the gesture is then registered by the YOLO model with identical outcomes being eliminated with non-maximum suppression.
- The recognised gesture is either displayed as text on the OLED screen or it is fed to the text-to-speech to have an audio output. A specific method is followed to standardize the process.

This makes each one much clearer and makes easier another implementation of the system by the hardware and software, the dataset and the model.

5. RESULTS AND DISCUSSION

5.1. Implementation results

The proposed assistive communication system was developed and evaluated with sensor-based and vision-based recognition modules. In the sensor-based recognition module, flex sensors from a glove, an Arduino Uno micro-controller, and an HC-05 Bluetooth module were used to convert hand gestures into text and speech output. Camera and DL-based image processing was used for vision-based recognition module to use hand gestures for converting hand motions into text output.

5.1.1. Results of sensor-based gesture recognition

The beginnings of the sensor-based prototype were made with a net fabric glove as the sensor input device. This caused the flex sensors to be fed inconsistent signals as the net glove was a poor fit for the hand.

Because of this the gesture recognition did not work as desired. The gloves were eventually replaced with a cotton glove. The system using flex sensor is implemented as shown in Figures 12 and 13 while the output results corresponding to these are shown in Figures 14 and 15.

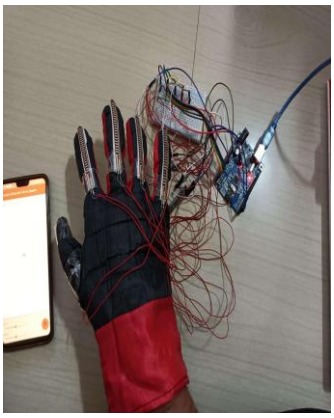

Figure 12. Implementation example 1

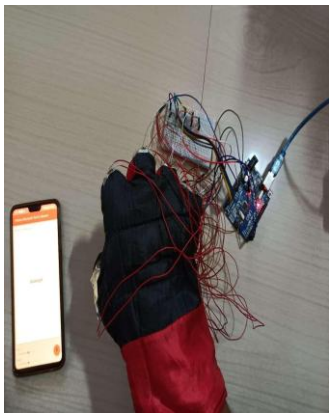

Figure 13. Implementation example 2

Figure 14. Implementation outcome 1 (zoomed)

Figure 15. Implementation outcome 2 (zoomed)

The Arduino Uno reads the analog voltage values from the flex sensor continuously, checks if the values fall into the gesture specific threshold range; if yes, it sends the necessary message-'Hello', 'Bye', 'Emergency', 'Call the Doctor' through Bluetooth using HC-05 Bluetooth module to the mobile via an application which in turn translates the message into speech output. The main observations of the sensor-based implementation are summarized in Table 5.

Table 5. Sensor-based implementation results

Parameter	Observation
Input device	Flex sensor glove
Microcontroller	Arduino Uno
Communication module	HC-05 Bluetooth module
Output format	Text and speech
Recognized messages	Hello, bye, emergency, and call the doctor
Initial limitation	Loose net glove caused unstable sensor readings
Improvement made	Cotton glove used for better fitting and stable bending
Main error source	Variation in finger bending angle and user gesture style
Overall performance	Reliable under controlled conditions

As shown in Table 5, the results proved the feasibility of the sensor-based module of low cost and real-time gesture recognition. But its accuracy is highly affected by the fit of the glove, the placement of the sensor and the finger movements.

5.1.2. Results of vision-based gesture recognition

The Vision-based system was evaluated by real time gesture images. The implementation results of 4 sample gestures are shown in Figures 16 to 19. The system takes a gesture images through the camera preprocess gesture features extraction and gesture classification through a trained DL model. The tests were conducted on an x 86 platform-based Windows operating system and on an ARM 64 Raspberry Pi platform.

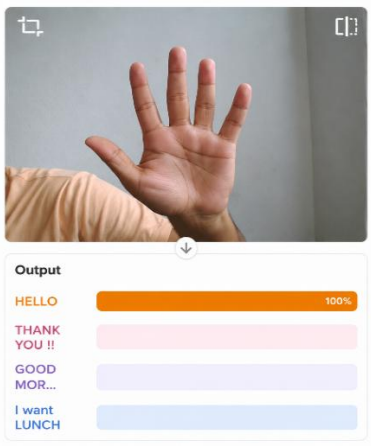

Figure 16. Implementation of hand gesture 1

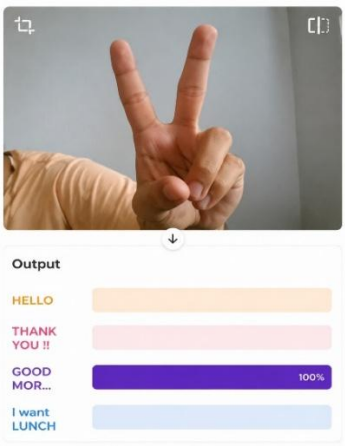

Figure 17. Implementation of hand gesture 2

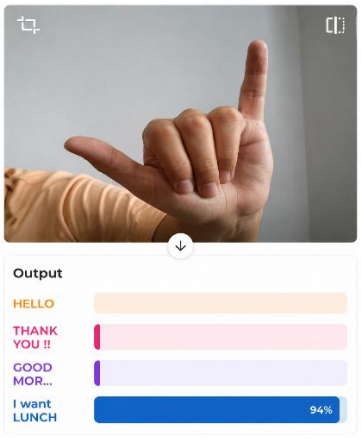

Figure 18. Implementation of hand gesture 3

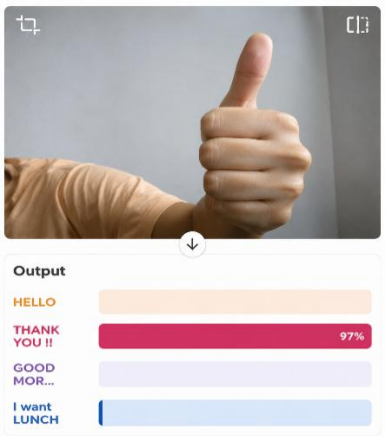

Figure 19. Implementation of hand gesture 4

The recognition results on the Windows platform were more conclusive due to the greater computing power and better DL library support. On the Raspberry Pi platform, the performance dropped due to its hardware constraints and a mismatch of the version of TensorFlow with the environment of the Raspberry Pi OS, so resulted in slow recognition, misclassification, and even gesture misrecognition while testing in real-time. The main findings of the vision-based implementation are presented in Table 6.

Table 6. Vision-based implementation results

Parameter	Observation
Input device	Camera module
Processing platform	Windows system and Raspberry Pi
Recognition method	DL-based image processing
Output format	Text output
Best performance condition	Uniform lighting and simple background
Major limitations	Lighting variation, motion blur, background clutter, and camera angle
Raspberry Pi issue	TensorFlow compatibility and lower processing capability
Improvement required	Model optimization and hardware acceleration

As presented in Table 6, the results indicate that vision-based recognition is effective under controlled lighting and background conditions. However, its real-time performance on embedded platforms depends on software compatibility, processor capability, and model complexity.

5.2. Quantitative performance evaluation

Recognition accuracy, precision, recall, F1-score and object detection performance of proposed system were measured to evaluate proposed system performance. Recognition accuracy of the whole system was measured in 80-85%. Recognition precision and recall in the vision-based recognition module were 0.83 and 0.81 respectively and F1-score was 0.82. It shows that the balance between false positive and correct detection was well maintained by the proposed system. The quantitative performance results are shown in Table 7.

Table 7. Performance evaluation of the proposed system

Metric	Obtained value	Interpretation
Recognition accuracy	80-85%	Shows acceptable overall gesture recognition performance
Precision	0.83	Indicates that most detected gestures were correctly classified
Recall	0.81	Indicates that most actual gestures were successfully detected
F1-score	0.82	Shows balanced precision and recall performance
mAP	To be reported from final YOLO testing	Required for class-wise object detection evaluation
Best operating condition	Controlled lighting and clear background	Provides more stable vision-based recognition
Main failure condition	Poor lighting, motion blur, and similar gestures	Causes false detection or misclassification

As shown in Table 7, with a precision value of 0.83 we can see the system had relatively few false positive detections. As a recall score of 0.81 indicates most of the gestures intended to be detected could be successfully identified although some gestures were missed in challenging conditions. The F1-score of 0.82 confirms the system maintained a reasonable balance between detection reliability and completeness. For the vision-based YOLO model, mAP should be used when performing the final testing. mAP gives a class-wise evaluation of detection accuracy in more detail than accuracy, which simply gives an overall accuracy of the detection. As such, using mAP in the final manuscript would strengthen the technical strength and reproducibility of the results.

5.3. Comparison with existing methods

The system was used as a benchmark to compare with the mainstream sensor-based and vision-based assistive communication systems as reported in the literature. The previous systems based on flex-sensing are reported to have a recognition accuracy of 83.58%, while earlier sensor system based on machine-learning has a recognition accuracy of 88.32%. YOLO and CNN based vision system can provide near real-time recognition, but it takes a much higher computing power. The comparison with existing methods is provided in Table 8.

Table 8. Comparison with existing gesture recognition methods

Method	Main technology	Reported performance	Major limitation
Flex-sensor glove-based system [2]	Flex sensors and Arduino Nano	83.58% accuracy	Requires wearable glove
Machine-learning sensor-based system [5]	Sensor data with adversarial learning	88.32% recognition rate	Depends on sensor quality and training data
Vision-based recognition systems [17]-[19]	YOLO/CNN-based detection	High real-time detection capability	Sensitive to lighting and hardware capability
Proposed system	Flex sensors+vision-based recognition	80-85% accuracy, F1-score 0.82	Affected by lighting, sensor bending variation, and embedded hardware limits

As shown in Table 8, while the proposed system has a marginal decrease in accuracy compared with some individual sensor-based machine learning approaches, but the value of the proposed system is its hybrid system design. The system doesn't rely solely on one recognition method, the flex-sensor module brings robust recognition while controlled hand movement, and the vision-based module enables contactless gesture recognition, the system achieves much more practical application flexibility.

5.4. Discussion

Experimental results indicate that such a dual-mode system can facilitate real-time assistive communication with low-cost hardware. The sensor-based module was relatively stable provided the gloves had a good fit and the gestures were executed smoothly. Replacing the net glove with a cotton one improved the stability of sensor readings and proved the importance of mechanical design on recognition rate. The vision module achieved an acceptable result in the Windows platform, but failed to meet good performance in the Raspberry Pi. The reason is Really the Windows system had a much more powerful computer and preferable environment with DL setup. However, the Raspberry Pi implementation was limited by available computational power and the compatible problem of TensorFlow. Most of recognition errors were due to lighting variation, cluttered background, motion blur, camera orientation and finger gesture similarity. In the sensor-based module, the errors were Mostly due to unclear finger bending by user or bend angle were not similar with stored thresholds. These failure cases tell us that, both modules have to be calibrated and under test environment controlled. The hybrid design could mitigate the drawbacks of each approach respectively. The sensor-based approach is robust to lighting and background variations, while vision-based approach need less wearable input. The combined approaches enhance the system adaptability for real world assistive communication.

While the propose system shows an excellent performance, there are a couple limitations. The sensor-based module takes good fitting of glove and exact position of flex sensors. Any displacement of sensors may affect the resistance and produce wrong labels. While the vision-based module is sensitive to bad illumination noise camera position and hardware constraint during embedded implementation. In the future, the performance of the gesture model could be improved by increasing the dataset size, number of gesture classes and training of the model with broader lighting and background conditions. The vision module can be optimized by leveraging lightweight architectures, model compression or hardware acceleration like graphics processing unit (GPU), Edge tensor processing unit (TPU), and TensorFlow Lite. For the sensor-based module, instead of using the fixed value threshold the adaptive thresholding or machine learning classification can be used. In sum, the success of this system demonstrates that this hybrid assistive communication system is able to fulfill the fundamental functional requirements of accurate real-time gesture-to-text and gesture-to-speech conversions. So, this system is a low-cost portable solution with real-world applicability.

6. CONCLUSION

This paper describes a two-mode aided human-computer communication system, which allowed communication in the text and speech mode through hand gestures. This consisted of the sensor-based module, which used flex-sensors, Arduino Uno, and HC-05 Bluetooth and the vision-based module, which used the camera, Raspberry Pi, and image processing. The sensor-based module was capable of recognizing 17 of the predefined gestures. The vision-based module provided for contactless gesture recognition.

System overall recognition accuracy was between 80% and 85%. The vision-based module had precision of 0.83, recall of 0.81, and F1-score of 0.82 showing there was relatively balanced detection performance. The flex-sensor module worked well under uncontrolled situation after the replacement of loose net glove with cotton glove. In the tested environment, vision-based module out-performed in Windows platform, hardware limit on Raspberry Pi caused slower FPS and incompatible with tensorflow. The major contribution of this paper is that propose a low-cost mobile gesture recognition hybrid setup that integrates wearable and camera-based gesture recognition systems.

This dual-mode system greatly enhances the flexibility, portability, and usability, compared to one single recognition method. Future work includes to increase the sensor calibration accuracy, enlarge the number of gesture classes for a more complete language, find the best lightweight DL models for the Raspberry Pi to improve recognition accuracy, and add other language support to the text to speech. To wrap up, this proposed system has good potential for an affordable assistive communication device for speech/hearing disability people.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the valuable guidance and technical support provided by Walchand College of Engineering, Sangli, Maharashtra, India whose insights greatly contributed to the successful completion of this work.

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
Mohan P. Thakre	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		
Krupali Kanekar	✓			✓		✓	✓			✓		✓		
Badal Kumar		✓			✓	✓		✓	✓					
Alok Kumar		✓		✓		✓				✓				
Supriya Nilesh Thakur	✓				✓			✓	✓		✓			
Pranali M. Thakre	✓		✓			✓				✓				
Prashant K. Magadum		✓	✓		✓				✓		✓	✓		

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.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] P. Telluri, S. Manam, S. Somarouthu, J. M. Oli, and C. Ramesh, "Low cost flex powered gesture detection system and its applications," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Jul. 2020, pp. 1128–1131, doi: 10.1109/ICIRCA48905.2020.9182833.

[2] L. A. Szolga, "On flight real time image processing by drone equipped with raspberry Pi4," in *2021 IEEE 27th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, Oct. 2021, pp. 334–337, doi: 10.1109/SIITME53254.2021.9663650.

[3] J. Marot and S. Bourennane, "Raspberry Pi for image processing education," in *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug. 2017, pp. 2364–2366, doi: 10.23919/EUSIPCO.2017.8081633.

[4] M. J. C. Samonte, R. A. Gazmin, J. D. S. Soriano, and M. N. O. Valencia, "BridgeApp: An assistive mobile communication application for the deaf and mute," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2019, pp. 1310–1315, doi: 10.1109/ICTC46691.2019.8939866.

[5] M. Sobhan, M. Z. Chowdhury, I. Ahsan, H. Mahmud, and M. K. Hasan, "A communication Aid system for deaf and mute using vibrotactile and visual feedback," in *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Sep. 2019, pp. 184–190, doi: 10.1109/ISEMANTIC.2019.8884323.

[6] V. N. Bharathi, S. Rizwan, R. R. Gangavarapu, and A. Thangapandi, "Real-time bidirectional assistive communication system for speech and hearing-impaired persons using deep learning models," in *2025 International Conference on Emerging Technologies in Engineering Applications (ICETEA)*, Jun. 2025, pp. 1–5, doi: 10.1109/ICETEA64585.2025.11099784.

[7] G. J. Ruslim, N. M. Salim, I. S. Edbert, and D. Suhartono, "Sign language detection to enhance online communication for deaf and mute individuals through deep learning models," in *2024 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, Nov. 2024, pp. 1–6, doi: 10.1109/CENIM64038.2024.10882726.

[8] S. Sasi, B. V. Srividya, and A. R. Aswatha, "Assistive device based on machine learning approach for communication of visually challenged and muted community," in *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, Sep. 2023, pp. 1048–1054, doi: 10.1109/ICOSEC58147.2023.10275924.

[9] E. S. Haq, D. Suwardiyanto, and M. Huda, "Indonesian sign language recognition application for two-way communication deaf-mute people," in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, Nov. 2018, pp. 313–318, doi: 10.1109/ICITISEE.2018.8720982.

[10] M. S. Verdadero and J. C. D. Cruz, "An assistive hand glove for hearing and apeech impaired persons," in *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Nov. 2019, pp. 1–6, doi: 10.1109/HNICEM48295.2019.9072695.

[11] S. S. Shinde, R. M. Autee, and V. K. Bhosale, "Real time two way communication approach for hearing impaired and dumb person based on image processing," in *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Dec. 2016, pp. 1–5, doi: 10.1109/ICCIC.2016.7919572.

[12] M. Kundurthi, T. Gurunathan, M. A. Khan, M. S. Raza, and A. Gangopadhyay, "Real-time object tracking on the Edge," in *2025 IEEE 9th International Conference on Fog and Edge Computing (ICFEC)*, May 2025, pp. 58–59, doi: 10.1109/ICFEC65699.2025.00011.
[13] S. Vigneshwaran, M. S. Fathima, V. V. Sagar, and R. S. Arshika, "Hand gesture recognition and voice conversion system for dump people," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, Mar. 2019, pp. 762–765, doi: 10.1109/ICACCS.2019.8728538.
[14] S. N. Kulkarni and S. K. Singh, "Object sorting automated system using Raspberry Pi," in *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*, Oct. 2018, pp. 217–220, doi: 10.1109/CESYS.2018.8724056.
[15] N. Supmool, P. Kositanon, S. Chaichalotornkul, and U. Manupibul, "Hand sign language translator using flex sensors and gyro sensors in pattern recognition method," in *2024 16th Biomedical Engineering International Conference (BMEiCON)*, Nov. 2024, pp. 1–5, doi: 10.1109/BMEiCON64021.2024.10896294.
[16] A. Suri, S. K. Singh, R. Sharma, P. Sharma, N. Garg, and R. Upadhyaya, "Development of sign language using flex sensors," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, Sep. 2020, pp. 102–106, doi: 10.1109/ICOSEC49089.2020.9215392.
[17] A. K. Panda, R. Chakravarty, and S. Moulik, "Hand gesture recognition using flex sensor and machine learning algorithms," in *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, Mar. 2021, pp. 449–453, doi: 10.1109/IECBES48179.2021.9398789.
[18] P. Modi, D. Menon, A. Verma, and A. S. Areeckal, "Real-time object tracking in videos using deep learning and optical flow," in *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, Jan. 2024, pp. 1114–1119, doi: 10.1109/IDCIoT59759.2024.10467997.
[19] N. Kumari, A. K. Bhatt, R. K. Dwivedi, and R. Belwal, "Accuracy testing of data classification using tenserflow a python freamwork in ANN Designing," in *2018 International Conference on System Modeling & Advancement in Research Trends (SMART)*, Nov. 2018, pp. 44–48, doi: 10.1109/SYSMART.2018.8746945.
[20] M. Sukkar, D. Kumar, and J. Sindha, "Real-time pedestrians detection by YOLOv5," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2021, pp. 01–06, doi: 10.1109/ICCCNT51525.2021.9579808.
[21] S. Mostafi, W. Zhao, S. Sukreep, K. Elgazzar, and A. Azim, "Real-time jaywalking detection and notification system using deep learning and multi-object tracking," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Dec. 2022, pp. 1164–1168, doi: 10.1109/GLOBECOM48099.2022.10000957.
[22] O. Vaidya, S. Gandhe, A. Sharma, A. Bhate, V. Bhosale, and R. Mahale, "Design and development of hand gesture based communication device for deaf and mute people," in *2020 IEEE Bombay Section Signature Conference (IBSSC)*, Dec. 2020, pp. 102–106, doi: 10.1109/IBSSC51096.2020.9332208.
[23] V. Jamdar, Y. Garje, T. Khedekar, S. Waghmare, and M. L. Dhore, "Inner voice - an effortless way of communication for the physically challenged deaf & mute people," in *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*, Sep. 2021, pp. 1–5, doi: 10.1109/AIMV53313.2021.9670911.
[24] A. K. Tripathy, D. Jadhav, S. A. Barreto, D. Rasquinha, and S. S. Mathew, "Voice for the mute," in *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, Feb. 2015, pp. 1–6, doi: 10.1109/ICTSD.2015.7095846.
[25] M. P. Thakre, P. S. Borse, N. P. Matale, and P. Sharma, "IOT based smart vehicle parking system using RFID," in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2021, pp. 1–5, doi: 10.1109/ICCCI50826.2021.9402699.

BIOGRAPHIES OF AUTHORS

Dr. Mohan P. Thakre received the B.Tech. and M.Tech. degrees in Electrical Power Engineering from Dr. Babasaheb Ambedkar Technological University (Dr. BATU), Maharashtra, India, in 2009 and 2011 respectively, and the Ph.D. degree in Electrical Engineering from Visvesvaraya National Institute of Technology (VNIT), Nagpur, Maharashtra, India in 2017. Currently, he is an associate professor at the Department of Electrical Engineering, Walchand College of Engineering, Sangli, Maharashtra, India. His research interests include FACTS and power system protection and renewable energy. He can be contacted at email: mohanthakre@gmail.com.

Dr. Krupali Kanekar received Ph.D. Degree in Electrical Engineering from Sandip University, Nashik in 2024, M.E degree in Electronics and Telecommunication Engineering from Ramrao Adik Institute of Technology, D. Y. Patil deemed to be university, Navi Mumbai in 2014. She did her B.E. in Electronics and Telecommunication Engineering from Don Bosco Institute of Technology, Mumbai in 2008. She worked as Trainee engineer in Mumbai Doordarshan, Worli, Mumbai. Since 2010 she is working as an Assistant professor in Ramrao Adik Institute of Technology, D. Y. Patil deemed to be university, Navi Mumbai. She can be contacted at email: krupali.kanekar@rait.ac.in.

Dr. Badal Kumar received his B.Tech. in Electrical and Electronics Engineering from Biju Patnaik University of Technology (2011) and M.E. in Digital Techniques and Instrumentation from S. G. S. Institute of Technology and Science (2015). He completed his Ph.D. from National Institute of Technology Manipur in 2022. He began his academic career in 2015 and served at Oriental College of Technology and Medi-Caps University. After his Ph.D., he worked at K. K. Wagh Institute of Engineering Education and Research, Shreeyash College of Engineering and Technology, and SVERI's College of Engineering in various academic and leadership roles. Currently, he is an Associate Professor in the Department of Electrical Engineering at JSPM University. His research interests include microgrids, smart grids, power system control, metaheuristic algorithms, artificial intelligence, and renewable energy systems. He can be contacted at email: kumarbadal89@gmail.com.

Dr. Alok Kumar received his B.Tech. degree in Electrical and Electronics Engineering from Biju Patnaik University of Technology, Rourkela, Odisha, in 2012. He completed his M.Tech. from the National Institute of Technology (NIT), Hamirpur, India, in 2015, and earned his Ph.D. from the Indian Institute of Technology (Indian School of Mines), Dhanbad, India. He has more than seven years of teaching experience in various engineering colleges across India. Currently, he is working as an Associate Professor in the Department of Electrical Engineering at SVERI's College of Engineering, Pandharpur, District Solapur, India. He current research interests include condition monitoring of electrical power apparatus and power system analysis. He can be contacted at email: kumaralok340@gmail.com.

Dr. Supriya Nilesh Thakur is currently serving as an Assistant Professor in the Department of Electrical Engineering at Marathwada Mitra Mandal's College of Engineering, Karvenagar, Pune, Maharashtra. She is actively involved in teaching, academic development, and research activities in the field of Electrical and Electronics Engineering. She has completed her Bachelor's degree in Electronics Engineering from Rajarambapu Institute of Technology, Sakharale, Islampur, Maharashtra, India. She pursued her Master's degree in Electronics and Telecommunication Engineering from MGM College of Engineering, Kalamboli, Panvel, Navi Mumbai, Maharashtra. Further strengthening her academic and research credentials, she completed her Ph.D. in Electrical and Electronics Engineering from the Sandip University, Nashik, Maharashtra, under the School of Engineering and Technology (SOET). She has a strong passion for teaching and mentoring students in core electrical and electronics subjects. Her research interests mainly focus on photovoltaic modules and their electrical parameters. She can be contacted at: supriyathakur78@gmail.com.

Mrs. Pranali M. Thakre received the MCA degree from Rashtrasant Tukadoji Maharaj Nagpur University (RTMNU), Maharashtra, India, in 2013 respectively, currently, she is an academic advisor at the Department of B.Sc. Computer Science (Entire) (BCS), Willingdon College, India. She can be contacted at email: pranalimohanthakre@gmail.com.

Dr. Prashant K. Magadum received his B.E. Electrical Electronics and Power Engineering from Swami Ramanand Teerth Marathawada University, Nanded, Maharashtra, India and M.E. Electrical (Power System) from Shivaji University, Kolhapur, Maharashtra, India respectively. He had completed Ph.D. degree in Electrical and Electronic Engineering from Visvesvaraya Technological University, Belgum, Karnataka, India. He is currently working as Associate Professor and Head, Department of Electrical Engineering, Brahmadevdada Mane Institute of Technology, Solapur, Maharashtra, India. He has having more than 20 years of teaching experience and has published more than 4 papers in various national and international journals. His research interest includes the design and control of multilevel converters for grid applications, FACTS based on modular multilevel converters and its controls. He can be contacted at email: magadumpk@gmail.com.