Vyanjak: Innovative Video Intercom and Notification System for the Deaf Community

Vandana Kate[a], Chanchal Bansal, Charu Pancholi, Ashvini Patidar, Gresey Patidar and Daksh Kitukale
Department of CSIT, Acropolis Institute of Technology and Research, MP, India

Keywords: Deaf Communication, Video Intercom, Offline Transcription, Vosk Library, Wearable Technology, IoT Integration, Random Forest, Sign Language Detection.

Abstract: Effective communication is essential for human connection, yet millions of deaf individuals face significant barriers that limit their social, educational, and professional opportunities. According to the World Health Organization (2024), over 430 million people globally experience disabling hearing loss, a number projected to reach 2.5 billion by 2050. This research introduces Vyanjak, an innovative video call intercom system designed to address communication challenges for the deaf community. The system includes a video call intercom, a wearable notification device, and an Android application, all working together to improve communication accessibility. Vyanjak operates on existing Wi-Fi networks and a dedicated ad hoc network, providing features such as video calls, group calls, and video messaging. The sign language-to-text conversion module utilizes Random Forest algorithms with 98% accuracy, enabling seamless communication between deaf individuals and others. For speech-to-text conversion, the system employs the Python library Vosk, offering offline functionality without internet dependency. The entire system is designed to work offline, ensuring reliable performance in any environment.The wearable device delivers customizable alerts via unique vibration patterns and color-coded lights, while the symbol-based interface is tailored for users with limited literacy. Offline AI-driven transcription converts both speech and hand signs into text across multiple languages. By integrating Internet of Things (IoT) devices such as smoke alarms, Vyanjak enhances safety of its users. This paper details the design, development, and potential impact of Vyanjak in improving communication, safety, and independence for deaf individuals, fostering inclusivity through bridging communication gaps.

1 INTRODUCTION

Communication is an essential component of human interaction, yet millions of individuals with hearing impairments encounter significant obstacles that hinder their full participation in society. According to the World Health Organization, over 430 million people worldwide experience disabling hearing loss, a figure projected to reach 2.5 billion by 2050. These challenges are particularly pronounced in social, educational, and professional contexts, where existing communication tools often depend on stable internet connectivity or fail to accommodate users with limited literacy, leaving many deaf individuals feeling isolated.

To address these challenges, Vyanjak presents an innovative communication solution that integrates a video call intercom, a mobile application, and a wearable notification device. Vyanjak is designed to function both online and offline, operating on existing Wi-Fi networks and a dedicated ad hoc network, ensuring consistent performance even in areas with limited connectivity. Central to the system is the video call intercom, which enables communication through sign language-to-text and speech-to-text conversion, utilizing advanced machine learning models to enhance accuracy. The intercom is equipped with a symbol-based interface that allows users to navigate easily by selecting familiar icons for frequently contacted entities, such as "+" for medical rooms.

Complementing the intercom is a wearable notification device that provides real-time alerts via customizable vibration patterns and color-coded lights. Each color corresponds to a specific contact, enabling users to identify incoming calls without needing to engage with their device actively. This feature en-

[a] https://orcid.org/0000-0002-2281-2187

Kate, V., Bansal, C., Pancholi, C., Patidar, A., Patidar, G. and Kitukale, D.
Vyanjak: Innovative Video Intercom and Notification System for the Deaf Community.
DOI: 10.5220/0013590200004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 2*, pages 253-259
ISBN: 978-989-758-763-4

sures users remain informed, even when not using the intercom or mobile app.

The Vyanjak mobile application, developed with Flutter for the frontend and Python for the backend, seamlessly integrates these features, allowing users to make video calls, send messages, and receive notifications while on the move. Additionally, the app supports customization of the contact list, enabling users to assign symbols or images to frequently contacted individuals for easier identification.

By providing a versatile and accessible communication platform, Vyanjak aims to enhance connectivity for the deaf community across various settings, including schools, hospitals, and workplaces, fostering greater independence and social interaction. This paper will outline the methodology for developing Vyanjak, detailing the hardware components, machine learning models utilized for sign language and speech recognition, and the design considerations for the mobile application and intercom system. The organization of the paper includes sections covering the problem statement, objectives, literature review, methodology, scope and limitations, expected outcomes, project timeline, resources, and conclusion.

Figure 1: The Intercom System

2 LITERATURE REVIEW

This section explores various works related to communication devices and systems for the deaf and hearing-impaired communities, focusing on advancements in sign language recognition, speech-to-text conversion, and mobile applications. (Nathan et al., 2016) studied mobile applications designed for the deaf community. Their work emphasizes the significance of mobile technologies in bridging communication gaps among hearing-impaired individuals, focusing on accessibility and user-friendliness in mobile applications. (Wei et al., 2010) proposed a solution to enhance voice quality in IP video intercom systems by implementing the G.711 encoding technique and the UDP transmission protocol. This innovation addresses delays and real-time voice communication issues, improving the user experience for video-based intercom systems. (Dai and Chen, 2014) developed an optimization scheme for visual intercom systems to improve visitor identification and video monitoring capabilities triggered by entrance calls. Their approach addresses the challenge of unsmooth video transmission, demonstrating strong market potential for such systems. In a subsequent study, (Wei et al., 2011) reiterated the importance of voice communication quality in IP video intercom systems. Their work further refines G.711 encoding and UDP protocol integration to mitigate delays and voice conflicts during video communication. (Mande and Lakhe, 2018) presented a system for automatic video processing utilizing Raspberry Pi and the Internet of Things (IoT). Their solution enhances video communication by leveraging IoT technologies to improve efficiency in sectors like healthcare and surveillance. (Sharma et al., 2013) proposed a sign language recognition system using the SIFT algorithm to recognize Indian Sign Language signs with 95% accuracy. The system focuses on real-time image capture and feature extraction, presenting an innovative method for converting sign language into text. (Naidu et al.,) introduced an application that facilitates communication for deaf, dumb, and paralyzed individuals. The app translates speech and text into sign language, employing video-based interaction on smartphones to bridge communication barriers. (Jhunjhunwala et al., 2017) developed a system using a glove equipped with flex sensors to recognize American Sign Language (ASL) signs. This system converts sign language into text and speech, providing real-time communication via an Arduino Nano. (Abraham and Rohini, 2018) created a real-time device using Arduino Uno, flex sensors, and an Android app to convert sign language gestures into text and speech. Their work also integrates a neural network-based prediction system for gesture recognition, improving the accuracy of sign language conversion. (Yadav et al., 2020) introduced a system that converts ASL gestures into text and speech using Convolutional Neural Networks (CNN). Their method focuses on recognizing fingerspelling gestures in real-time, providing a robust solution for sign language communication. (Vinnarasu and Jose, 2019) designed a speech-to-text conversion system that also provides concise text summaries, facilitating effective documentation of lectures or oral presentations. This solution enhances comprehension and serves as a tool for deaf individuals in academic settings.(Wagner, 2005) discussed various techniques for real-time intralingual speech-to-text conversion. Her research focuses on providing fast and accurate transcription for individuals with hearing impairments in settings such as conferences or counseling sessions. (Sharma and Sardana, 2016) presented a speech-to-text system using a bidirectional nonstationary Kalman filter. This method achieves 90% accuracy in noisy environments and offers a more noise-robust solution compared to traditional HMM-based systems.(Choi et al., 2018) introduced CCVoice, a mobile recording service that uses Google Cloud Speech API to convert audio to text. This service is designed to assist individuals in gathering voice evidence, particularly for victims of harassment or assault.

3 METHODOLOGY

The Vyanjak communication system consists of three primary components: the intercom, the wearable notification device, and the mobile application. These elements work together to provide a complete communication system for the deaf and hearing-impaired.

3.1 System Components

Intercom Device

The intercom supports video calls, group calls, and video messaging while operating on Wi-Fi or its own ad hoc network. It includes a symbol-based UI for users with limited literacy and can function offline using AI-powered transcription.

- **Features**:
  1. Works on existing Wi-Fi or its own network.
  2. Supports video calls, group calls, video messaging.
  3. Customizable alarms using video messages.
  4. Symbol-based UI for high illiteracy.
  5. Offline AI-powered transcription (speech-to-text, hand signs-to-text).

Wearable Notification Device

The wearable device delivers notifications using unique vibration patterns and color-coded lights for easy recognition of incoming communications.

- **Key Features**:
  1. Customizable vibration patterns.
  2. Color-coded lights for contact identification.
  3. Optionally displays text for those who can read.

Mobile Application

The mobile app integrates with the intercom and wearable device to allow users to make video calls, send messages, and manage alerts as shown in Figure 2 illustrating various user interface screen . The app ensures that users can receive calls, notifications, and reminders through a symbol-based interface, catering to individuals with low literacy levels.

- **Home Screen**: Users can select contacts with symbols and initiate calls or messages.
- **Notification Screen**: Displays video messages, reminders, or alerts.
- **Video Call Screen**: Supports real-time video calls with sign language-to-text and speech-to-text conversion.
- **Settings Screen**: Manages system preferences and displays device status, including battery and connectivity.

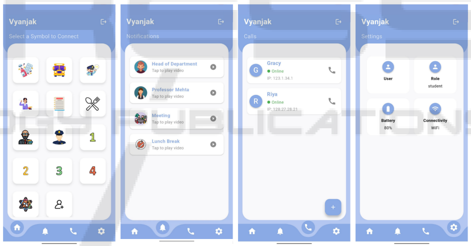


Figure 2: Mobile App Screens

3.2 System Workflow

The following outlines the workflow of the Vyanjak system:

1. User logs into the mobile app and sets up the device.
2. User selects a contact and initiates a call or message via the intercom or mobile app.
3. The notification device vibrates and flashes a color to indicate incoming communication.
4. Real-time communication transcribes sign language and speech into text.
5. The system provides customizable vibration patterns and light signals, offering multi-modal communication support.

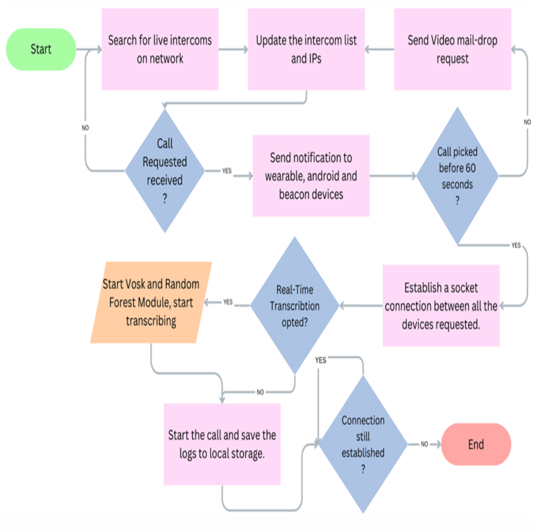


Figure 3: The Adhoc Network Calling Flow Chart

The flowchart in Figure 3 illustrates the process flow for handling incoming and outgoing calls in the Vyanjak system. The steps can be described as follows:

1. The system starts by searching for live intercom devices connected on the network.
2. After detecting live intercoms, the system updates the intercom list and their corresponding IP addresses.
3. If a call request is received, the system sends a notification to the wearable device, mobile app (Android), and beacon devices.
4. The system checks if the call has been picked up within 60 seconds:
   - If the call is picked up, it establishes a socket connection between all the devices involved in the call.
   - If the call is not picked up within 60 seconds, a video mail-drop request is triggered, allowing the user to leave a video message for the recipient.
5. The system checks if real-time transcription is opted for by the user. If so, the Vosk speech recognition module and the Random Forest hand-sign recognition module are activated to transcribe the conversation.
6. The system continuously monitors the call connection. If the connection is lost, the call is terminated, and the session ends.
7. During the call, the system logs the call details to local storage for future analysis.

3.3 Internal Hardware Components

The proposed system consists of various hardware and software components as listed in Table 1 and 2. The hardware components include:

- **Raspberry Pi 4B**: Central processing unit for video calls and sign recognition.
- **Camera Module**: Captures real-time hand signs.
- **ESP-32 Microcontroller**: Manages notification alerts.
- **mDNS and Wi-Fi**: Ensures connectivity even without the internet.

Table 1: Hardware Components and Descriptions

Component	**Description**
Raspberry Pi	Main processor for the Intercom Device handling video calls and communication.
ESP32	Microcontroller used in the Notification Device for vibration and light feedback.
Camera Module	Captures video for calls on the Intercom Device.
Vibration Motor	Provides tactile feedback in the Notification Device.
LED Lights	Provide visual alerts for incoming calls or messages on the Notification Device.
Microphone and Speaker	Enable audio communication for the video call system.
Power Source	Power supply for Raspberry Pi and ESP32 devices.
Display	Shows caller information and text messages on the Intercom Device.

Table 2: Software Components and Descriptions

Software Component	**Description**
Python	Used for developing the system's backend and communication logic on the Raspberry Pi.
NodeMCU Framework (ESP32)	Manages the Notification Device's firmware and communication features.
Google Firebase (optional)	Could be used for managing communication between devices and storing app configurations.
Wi-Fi Protocols	Ensures communication between the Intercom Device, Notification Device, and Mobile App.
Open-Source Libraries	For handling video processing, communication, and real-time notifications.
Mobile Application (Android/iOS)	Symbol-based app for controlling intercom functions, making and receiving calls, and customizing symbols, lights, and vibrations.

3.4 Random Forest Classifier for Hand Sign Recognition

The Random Forest classifier is employed for interpreting hand signs. Below are the steps involved in sign recognition:

1. **Image Capture**: Real-time hand gesture images are captured.
2. **Feature Extraction**: OpenCV is used to extract features from images.
3. **Sign Recognition**: Random Forest classifier converts signs into text.
4. **Speech Output**: Text is further converted into speech for enhanced communication.

3.5 Random Forest Classifier for Hand Sign Recognition

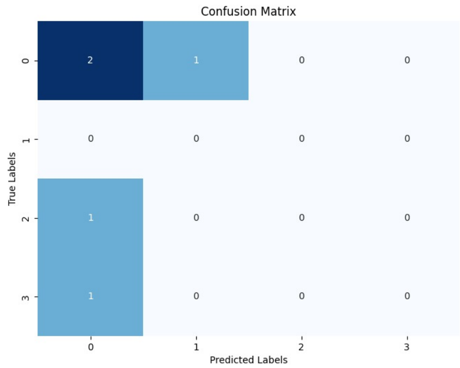
The Random Forest classifier is employed for interpreting hand signs and has proven highly effective in handling diverse datasets, including ASL (American Sign Language), ISL (Indian Sign Language), BSL (British Sign Language), and custom sign language images. Below are the steps involved in the sign recognition process:

1. **Image Capture**: Real-time hand gesture images are captured using a camera module.
2. **Feature Extraction**: OpenCV is used for preprocessing and extracting significant features from the captured images.
3. **Sign Recognition**: A Random Forest classifier is used to recognize hand signs and convert them into text. The classifier was trained using a dataset comprising 7 distinct classes, each containing 500 images resized to 48x48 pixels. The dataset was split into 75% for training and 25% for testing, and the model's performance was evaluated over 60 iterations.
4. **Speech Output**: The recognized text is then converted into speech, providing an additional communication modality for users.

3.6 Random Forest Model Performance

The Random Forest classifier demonstrated superior performance in terms of both accuracy and computational efficiency when classifying hand signs. Below is a detailed analysis of the model's metrics and performance:

- **Confusion Matrix**: Figure 4 shows the confusion matrix, providing insights into the true positive and false negative rates for the 7 distinct hand sign classes.



ConfusionMatrix.jpg

Figure 4: Confusion Matrix for the Random Forest classifier

- **Classification Report**: The classification report shown in Figure 5 includes metrics such as precision, recall, and F1-score for each class. This report highlights the robustness of the model in classifying ASL, ISL, BSL, and custom sign languages.

Classification Report:

	precision	recall	f1-score	support
1	0.50	0.67	0.57	3
2	0.00	0.00	0.00	0
4	0.00	0.00	0.00	1
5	0.00	0.00	0.00	1
accuracy			0.40	5
macro avg	0.12	0.17	0.14	5
weighted avg	0.30	0.40	0.34	5

Figure 5: Classification Report of the Random Forest Model

4 USABILITY AND CUSTOMIZATION

The system is designed to provide a highly intuitive experience for individuals who may be deaf, hearing-impaired, or illiterate. The symbol-based UI across both the intercom and mobile app makes it accessible for users with limited literacy. The notification device supports users by offering non-verbal cues, customizable vibration patterns, and color-coded lights. This multi-modal approach ensures that users can recognize incoming calls or messages based on sensory feedback alone.

4.1 Customization Options

- **Symbols**: Unique icons can be assigned to each contact for easy recognition on both the intercom and mobile app.
- **Vibration Patterns**: Users can set specific vibration patterns for different contacts or message types. For example, three short vibrations could indicate a family member calling, while one long vibration signals an urgent message from a healthcare provider.
- **Light Signals**: LED colors and flashing patterns can also be personalized to help the user recognize who is calling or the importance of the notification.

5 IMPACT AND BENEFITS

5.1 Impact

- **Inclusivity:** Bridges the 45% employment gap for deaf individuals by enabling effective communication, thus enhancing workforce inclusivity.
- **Safety:** Integration with IoT devices, such as smoke alarms, enhances safety and enables quick decision-making.

5.2 Benefits

- **Education Sector:** Creates an inclusive environment that aids student success.
- **Employment:** Boosts job opportunities for disabled individuals, enhancing the economy.
- **Healthcare:** Improves patient communication and the quality of care.
- **Security:** Real-time notifications assist in making quick decisions.
- **Independence:** Supports autonomy for those with speech or hearing impairments.
- **Elderly Care:** Provides accessible alerts and communication tools for elderly individuals.

6 CONCLUSION

The Vyanjak system offers a versatile communication solution for the deaf community by combining machine learning, IoT integration, and a user-friendly interface. Its offline functionality and real-time transcription make it a reliable aid in various settings such as schools, hospitals, and workplaces, enhancing independence and interaction for users.

REFERENCES

Abraham, A. and Rohini, V. (2018). Real-time conversion of sign language to speech and prediction of gestures using artificial neural network. In *Procedia Computer Science*, volume 143, pages 587–594.

Choi, J., Gill, H., Ou, S., Song, Y., and Lee, J. (2018). Design of voice to text conversion and management program based on google cloud speech api. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1452–1453, Las Vegas, NV, USA.

Dai, H. and Chen, S. (2014). An optimization scheme of visual intercom call monitoring. In *The 7th IEEE/International Conference on Advanced Infocomm Technology*, pages 85–90, Fuzhou, China.

Jhunjhunwala, Y., Shah, P., Patil, P., and Waykule, J. (2017). Sign language to speech conversion using arduino. In *Proceedings of the International Conference on Advances in Computing*.

Mande, V. and Lakhe, M. (2018). Automatic video processing based on iot using raspberry pi. In *2018 3rd International Conference for Convergence in Technology (I2CT)*, pages 1–6, Pune, India.

Naidu, V. P. V., Hitesh, M. S., and Dhikhi, T. Department of computer science and engineering, saveetha school of engineering, saveetha university, chennai.

Nathan, S., Hussain, A., and Hashim, N. L. (2016). Studies on deaf mobile application. In *Proceedings of the AIP Conference*, volume 1761, page 020099.

Sharma, N. and Sardana, S. (2016). A real-time speech-to-text conversion system using bidirectional kalman filter in matlab. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2353–2357, Jaipur, India.

Sharma, S., Goyal, S., and Sharma, I. (2013). Sign language recognition system for deaf and dumb people. In *International Journal of Engineering Research and Technology*, volume 2, pages 382–387.

Vinnarasu, A. and Jose, D. (2019). Speech to text conversion and summarization for effective understanding and documentation. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(5):3642–3648.

Wagner, S. (2005). Intralingual speech-to-text conversion in real-time: Challenges and opportunities. In *Proceedings of the International Conference on Speech Processing*.

Wei, F., Jun, Z., Ping, W., and Yachao, Z. (2010). Study on g.711 voice communication of ip video intercom system. In *2010 International Conference on Digital Manufacturing & Automation*, pages 490–492, Changcha, China.

Wei, F., Zhang, J., Wang, P., and Zhang, Y. (2011). Study on g.711 voice communication of ip video intercom system. In *Proceedings of the Digital Manufacturing & Automation Conference*, volume 2, pages 490–492.

Yadav, B. K., Jadhav, D., Bohra, H., and Jain, R. (2020). Sign language to text and speech conversion. In *International Journal of Advance Research, Ideas and Innovations in Technology*.