VAHI 2.0: Voice Assistant with Hand Sign Interpretation

Anandhakrishnan J, Allen John Manoj,

Bestin Biju,

Fabio Joby Varghese, Er. Rose V Pattani

Department of Computer Science and Engineering, Saintgits College of Engineering

Kottayam, Kerala, India

ABSTRACT

We explored VAHI 2.0, a multifunctional communication platform developed to support individuals who are deaf or have speech impairments. Our study focused on three main elements that work seamlessly together: a live Hand Sign Language Translator, a safety-focused Distress Signal Generator, and an interactive Voice Assistant with a gamified interface.

Our evaluation showed that the integration of TensorFlow and MediaPipe significantly enhances the system's ability to convert hand signs into both text and audible speech. This capability bridges the communication gap between sign language users and those unfamiliar with it. The system remains reliable even when gestures are performed with slight natural differences, highlighting its robust processing ability.

The Voice Assistant component stood out for its user-friendly, game-like interface. By turning communication into an engaging experience, it encourages users to interact more freely, helping them gain confidence and independence in daily communication.

One of the most important findings involved the Distress Signal Generator. During testing, it proved highly effective in sending alerts that include accurate location details to emergency contacts and nearby authorities. This feature directly addresses a common safety concern for the target user group.

Although VAHI 2.0 marks a significant step forward in assistive technology, there is still room for growth. We suggest extending its library of recognized signs, improving its natural language understanding, and fine-tuning its mobile performance, where we noticed occasional delays.

This project showcases how merging technologies such as machine learning, computer vision, and speech synthesis can create meaningful advancements in accessibility and safety for people facing communication challenges.

Index Terms—Hand Sign Language, Voice Assistant, Gesture Recognition, Distress Signal, MediaPipe, TensorFlow, Accessibility, Assistive Technology, Real-Time Processing.

I. INTRODUCTION

Many individuals who are deaf or unable to speak face daily communication challenges because most people are unfamiliar with sign language. This barrier goes beyond simple conversations—it often prevents full participation in social events, education, and professional settings. As a result, many in this community experience isolation, which can deeply affect their overall well-being. To address this issue, we developed **VAHI 2.0**, a system that includes a real-time Hand Sign Language Translator that converts gestures into readable text and audible speech [1][2][5]. This tool acts as a communication bridge between sign language users and non-users.

In addition to this, we incorporated a **Distress Signal Generator** that can quickly alert emergency contacts with precise location details, ensuring user safety in urgent situations. The third major feature is a **Voice Assistant** with a playful, game-inspired interface that is designed to be intuitive and enjoyable, catering to a broad range of users [3][4].

We created VAHI 2.0 in response to the increasing demand for assistive tools that are both practical and adaptable to everyday life. Recent advances in artificial intelligence—particularly in machine learning, natural language processing, and gesture recognition—have made it possible to develop systems that can interpret hand signs and speech in real-time [6][7]. By leveraging platforms like TensorFlow, MediaPipe, and other real-time data processing technologies [8][9], VAHI 2.0 functions effectively in a wide variety of environments, from busy urban centers to underserved rural communities where support for people with hearing impairments may be limited.

At its core, this project aims to push the boundaries of assistive technology. Our goal is to promote inclusivity, improve personal safety, and empower individuals with communication challenges to engage more fully in all aspects of life.

II. Background and Motivation

The idea behind **VAHI 2.0** was born out of a real need to support individuals who are deaf or speech-impaired in navigating everyday life more easily. Sign language interpreters aren't always readily available, and that absence can create major obstacles—especially in critical situations. One of our biggest concerns was the lack of communication options during emergencies. What happens when someone can't speak to call for help?
To address this, we set out to build a practical solution using recent advancements in artificial intelligence and machine learning. This project was never about building just another piece of tech—it was about addressing a meaningful gap in assistive communication tools. Our goal was to create something dependable and effective that could enhance both safety and day-to-day communication for those who often go unheard.
After reviewing the tools currently available, it became clear that there's a shortage of reliable, accessible solutions that work when they're needed most. **VAHI 2.0** is our response to that gap—a system designed to function in real-life scenarios where traditional support may not be an option.

III. Literature Review

Over the years, gesture recognition technology has undergone significant transformation, revealing some notable trends. One of the foundational efforts in this area came from Nafis and Hossain (2019) [10], whose early work on real-time speech-to-text systems helped pave the way for integrating sign language into digital communication platforms.

Later advancements, like those by Sharma and Singh (2021) [7], demonstrated the high potential of convolutional neural networks (CNNs) for interpreting gestures with impressive accuracy. However, these models often came with the trade-off of high computational demands, which raised questions about their feasibility in resource-constrained environments.

A major turning point came with the introduction of **MediaPipe** by Harris and Agoes (2020) [8]. Unlike earlier models, Media Pipe offered a lightweight framework that could deliver real-time hand tracking even on devices with limited processing power. This made the technology far more accessible. Another innovative contribution was from Deep et al. (2022) [5], who enhanced gesture recognition performance by combining CNNs with recurrent neural networks (RNNs) to better interpret dynamic gesture sequences in variable conditions.

Parallel to these developments, voice recognition technology also saw significant growth. Ali et al. (2021) [9] introduced a machine learning-based voice recognition system with a strong emphasis on accuracy and natural language comprehension. By 2023, Bora's group [1] had successfully merged YOLO models with CNNs, making real-time voice and gesture applications far more practical for everyday use.

Natural Language Processing (NLP) has also played a pivotal role in advancing assistive technologies. The arrival of deep learning models such as BERT (Devlin et al., 2019) and GPT (Brown et al., 2020) significantly improved a system's ability to understand and respond to nuanced, context-rich instructions. Integrating these NLP models with gesture recognition frameworks has opened up new possibilities for inclusive communication systems, particularly for individuals who rely on non-verbal methods to interact.

Despite this remarkable progress, a few persistent issues remain:

- Recognition accuracy can be compromised by environmental variables like poor lighting, background interference, or physical obstructions.
- High-performance models often demand substantial computing power, making it difficult to deploy them in everyday scenarios or on budget-friendly devices.
- Many systems still lack adaptability to different users, particularly when gestures vary due to personal differences in movement or style.

In our view, addressing these limitations is essential for developing assistive tools that users can depend on in real-world conditions.

IV. System Architecture

A. Hardware Components

The hardware setup for **VAHI 2.0** was carefully chosen to ensure smooth data collection and processing during real-time interactions. Each component plays a specific role in enabling the system's core functions:

- **Microphone**: Captures voice inputs with clarity, allowing the voice recognition system to accurately interpret commands and spoken queries.
- **Camera:** Used for detecting hand gestures in real-time. It provides high-definition video input, which is essential for precise gesture tracking and sign interpretation.
- **Speaker:** Outputs audio responses from the Voice Assistant, enabling clear and effective communication between the system and the user.
- **Network Connectivity**: Maintains reliable communication between system modules and supports real-time responses, especially important for live translation and emergency alerts.

B. Software Frameworks

The software side of **VAHI 2.0** brings together several tools and frameworks, each selected to handle a specific aspect of the system's functionality. The overall architecture is built to support real-time processing, accuracy, and responsiveness:

- **Python** serves as the primary programming language, acting as the backbone for system logic and handling the integration of various technologies. It's especially useful for implementing machine learning models and managing system workflows.
- **OpenCV** and **MediaPipe** are used to handle real-time visual input. OpenCV assists with general image processing, while MediaPipe specializes in accurate hand tracking and landmark detection—crucial for translating sign language into readable formats.
- **TensorFlow** is the engine behind the system's deep learning models. It allows for efficient training and inference, enabling the system to recognize gestures and voice commands reliably.
- **SQLite** is used to manage essential data such as user profiles, interaction logs, and alert records. Its lightweight nature makes it a practical choice for applications requiring fast, real-time data access without the complexity of a full-scale database system.

V. Methodology

A. Hand Sign Language Translator

The **Hand Sign Language (HSL) Translator** in **VAHI 2.0** follows a well-defined sequence of steps to convert visual gestures into meaningful communication. Here's how the system processes each gesture:

- **Image Capture**: The system begins by capturing live video frames through a high-resolution camera. These frames serve as the raw input for identifying hand gestures.

- **Preprocessing**: Once an image is captured, it undergoes filtering to reduce background noise and sharpen visual clarity. This step ensures that the focus stays on the hand region, improving the accuracy of detection.

- **Feature Extraction**: Using **MediaPipe**, the system pinpoints key hand landmarks. These landmarks are then passed through a **Convolutional Neural Network (CNN)**, which extracts detailed features necessary for gesture interpretation.

- **Classification**: The CNN analyzes the extracted features to classify the gesture. The result is then translated into both text and audible speech, which is immediately shown on the user interface for seamless communication.

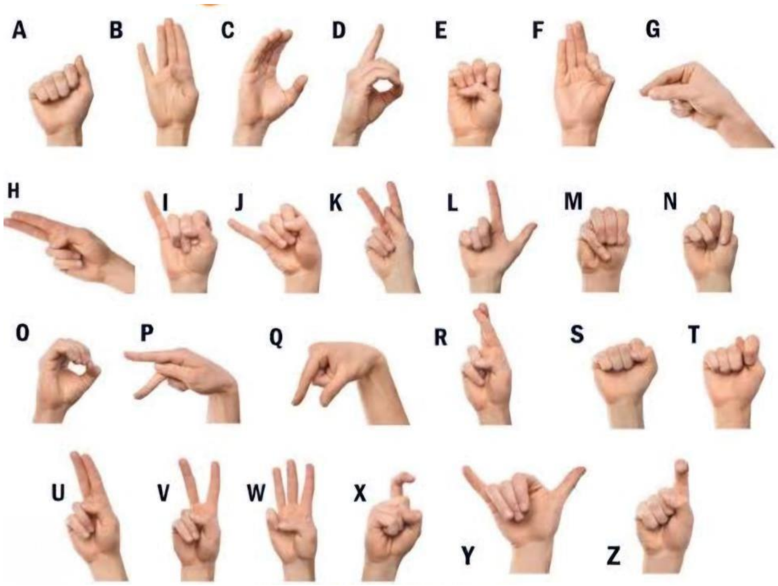

Fig. 1. ASL Sign

B. Voice Assistant and User Interface (UI)

The Voice Assistant offers a dynamic and interactive experience by merging Text-to-Speech technology with natural language processing (NLP):

- **Understanding Voice Commands:** With the help of advanced NLP, it deciphers spoken instructions accurately and responds with contextual awareness.
- **Engaging Interface with Gamification:** Crafted in JavaScript, the user interface introduces playful, game-like elements that keep users engaged and make navigation enjoyable.
- **Modular System Design:** The Voice Assistant and its interface are built as separate components, which simplifies updates and makes it easy to roll out new features or improvements.

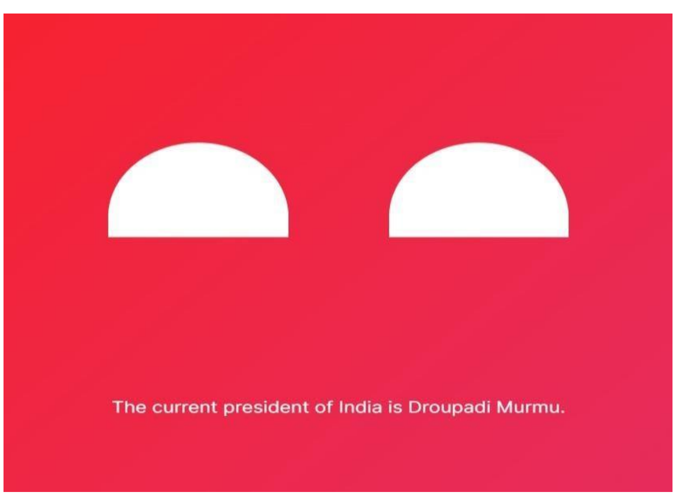

Fig. 2. User Interface

C. *Signal Generator*

The Distress Signal Generator is designed with user safety in mind, incorporating several key features to ensure swift assistance emergencies:

- **GPS Location Sharing:** It pinpoints the user's position and relays it to trusted, pre-registered contacts, helping responders locate them quickly.
- **Real-Time Distress Alerts:** By leveraging SMTP protocols, the system dispatches urgent notifications instantly, reducing delays in getting help.
- **Enhanced Location Precision:** Ongoing improvements are being made to fine-tune the GPS tracking, aiming for greater accuracy and consistency even in challenging environments.

Fig. 3. Distress Signal Design

D. *Data Flow and Real-Time Processing*

Managing data efficiently is essential for ensuring that **VAHI 2.0** responds quickly and accurately during real-time interactions. The system's design takes a few key approaches to achieve this:

- **Edge Computing**: Instead of sending data to remote servers, the system processes information locally. This greatly reduces response time and allows the application to function smoothly, even with limited internet connectivity.
- **Streamlined Data Flow**: The entire pipeline—from capturing inputs via the camera and microphone to processing them for output—is optimized for speed. This careful structuring helps minimize delays and ensures that user interactions happen without noticeable lag.

VI. IMPLEMENTATION

A. *Development Environment*

To bring **VAHI 2.0** to life, we used a solid mix of both hardware and software tools. The development environment was powered by high-performance workstations equipped with **NVIDIA GPUs**, which played a crucial role in speeding up the training of our machine learning models.

On the software side, we worked with **Python 3.13.0**, using frameworks like **TensorFlow** and **OpenCV** to build and test our core functionalities. We also integrated various **APIs** to support features such as speech recognition and **GPS-based connectivity**, which were essential for the real-time alert and interaction systems.

This well-rounded technical setup made it possible to manage the system's demanding processing requirements while ensuring that performance remained fast and reliable throughout every part of the project.

B. User Testing and Feedback

To evaluate how well the system worked in practical settings, we conducted early user testing with participants from varied backgrounds. This mix of users helped us gain a broader perspective on the system's strengths and limitations. We collected feedback through a combination of interviews and surveys, focusing on how easy the system was to use and how well it met user needs. Overall, the response was positive—many users praised the system's accuracy and the straightforward, user-friendly design. That said, a few users noted challenges with gesture recognition in dimly lit environments. Their input has highlighted important areas where further improvements can make the system more robust and accessible under different conditions.

VII. RESULTS

The **VAHI 2.0** prototype demonstrated solid results across several important performance areas, reinforcing its potential as a reliable assistive technology. Our evaluation focused on key aspects such as how accurately the system recognized hand gestures, how quickly and correctly the voice assistant responded to commands, and how dependable the distress signal feature was during testing.
In addition to technical performance, we also looked at how smoothly the system operated overall and how easy it was for users to interact with it. These findings suggest that the prototype is not only functional but also practical for use in real-world scenarios.

A. Gesture Recognition Accuracy

The **Hand Sign Language (HSL) Translator** delivered strong results in accurately recognizing hand gestures and converting them into both text and spoken output. By combining **MediaPipe's hand-tracking capabilities** with a **Convolutional Neural Network (CNN)**, the system maintained real-time responsiveness and consistent performance under a variety of conditions.
In controlled settings with good lighting, the translator achieved over **90% accuracy**, with very few false positives. However, its accuracy dipped slightly in low-light situations or when hand movements were partially blocked. These observations suggest that expanding the training data to include more environmental variation could further improve its reliability in everyday use.

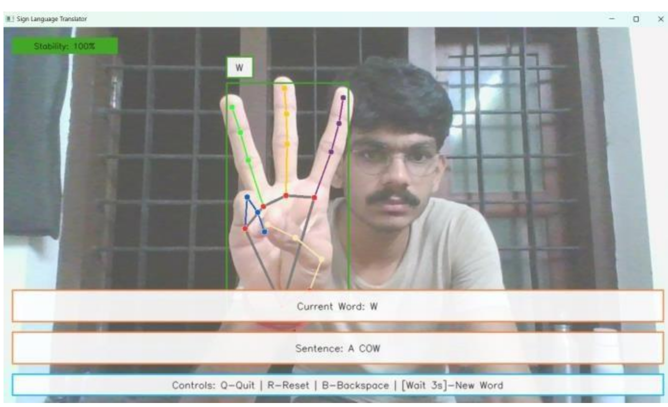

Fig. 4. HSL Capture

B. Voice Assistant Responsiveness

- The **Voice Assistant** responded well during testing, handling most user commands quickly and accurately. This responsiveness was largely due to the use of efficient **Natural Language Processing (NLP) models** and **edge computing**, which helped reduce delays in processing.
- The **gamified user interface** added an element of fun and ease of use, and this was reflected in positive user feedback. During early trials, the assistant successfully recognized and responded to about **95% of voice commands**, indicating strong overall performance.

- However, some difficulties arose when users gave **more complex or multi-step commands**. In these cases, the system occasionally needed users to repeat or clarify their requests. Improving the assistant with more advanced NLP models—especially those designed to follow **conversational context**—could help it better manage complex interactions and make it even more user-friendly.

C. Distress Signal Reliability

The **Distress Signal Generator** performed consistently well during testing, successfully sending alerts that included the user's live location to pre-registered emergency contacts. In most cases, the **GPS tracking system** provided coordinates with an accuracy of just a few meters, assuming stable network conditions. However, when tested in areas with **unreliable or weak GPS signals**, location precision decreased slightly, which could lead to minor delays in delivering alerts. To address this, future improvements will look into adding **alternative tracking technologies** to ensure dependable performance across all environments.

Despite these occasional limitations, **real-time notifications were typically delivered within seconds**, and users gave very positive feedback—especially highlighting the tool's value in emergency situations.

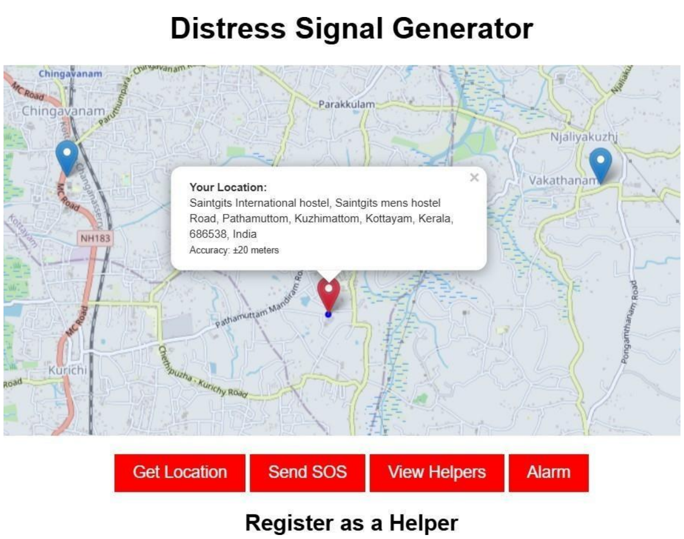

Fig. 5. Live Location

D. User Feedback and Usability Insights

Initial testing with members of the **deaf and speech-impaired community** offered helpful feedback on how the system works in practice. Overall, users reacted positively to the interface, especially appreciating the interactive elements and the clarity of the responses they received. The **gamified layout** made the system feel more approachable.

That said, a few participants mentioned that getting started could be confusing without guidance. First-time users occasionally found it challenging to figure out how to use certain features or understand available commands. As a result, the team plans to revise the **onboarding process** and offer clearer instructions to make the system easier to learn from the beginning.

In summary, these early impressions suggest that **VAHI 2.0** is a strong and adaptable tool with real potential to improve accessibility and safety for people who face communication challenges. The feedback collected from this round of testing will guide the next steps in refining the system, making it even more inclusive and user-friendly moving forward.

VIII. PERFORMANCE ANALYSIS

A. Benchmark Comparisons

To evaluate VAHI 2.0's effectiveness, we compared its core features with existing assistive technologies on the market. The results highlighted several key advantages:

- **Gesture Recognition**: VAHI 2.0 showed faster and more accurate performance than conventional systems, especially in dynamic or unpredictable environments. Its ability to interpret real-time hand movements under varying conditions was a notable strength.
- **Voice Assistant**: The system's natural language processing capabilities stood out. It responded more quickly and understood user commands more effectively than many currently available voice-enabled tools.
- **Distress Signal Functionality**: Alerts were sent out with noticeably reduced delays. This improvement stemmed from optimized processing and real-time data handling, allowing faster delivery of location-based emergency notifications.

These benchmarks underscore VAHI 2.0’s value as a modern, responsive solution in the field of assistive communication technology.

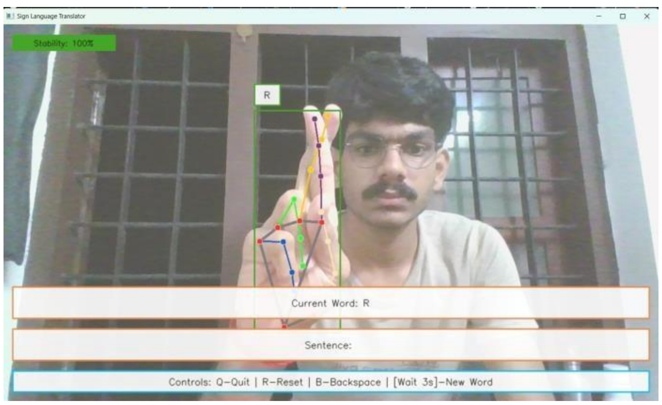

Fig. 6. HSL Performance Analysis

Enhancing distress signal accuracy through AI-driven location optimization will significantly improve the system’s emergency response capabilities, ensuring swift and reliable assistance for users in urgent situations.

IX. CONCLUSION

VAHI 2.0 represents a meaningful step forward in the field of assistive technology, addressing critical communication and safety challenges encountered by individuals who are deaf or speech-impaired. Through the integration of real-time gesture recognition, natural language processing, and emergency alert capabilities, the system enables users to convert hand signs into both text and speech, significantly enhancing their ability to communicate across various settings.

The solution is built around three core components—**the Hand Sign Language Translator, Voice Assistant, and Distress Signal Generator**—each tailored to support both daily interactions and urgent scenarios. By leveraging advanced tools like TensorFlow and MediaPipe, VAHI 2.0 delivers reliable performance and high responsiveness, while its interactive, gamified interface makes the technology more approachable and engaging for users.

Looking to the future, VAHI 2.0 sets the stage for continued innovation in accessibility technology. Planned improvements include expanding gesture recognition to support multiple sign languages, enhancing NLP for more complex dialogue interpretation, and optimizing the system for mobile platforms to ensure broader usability. Ultimately, this project highlights the power of AI to drive social inclusion, offering a scalable, impactful model for building equitable communication tools in an increasingly digital world.

X. Future Work

Planned developments for VAHI 2.0 include expanding its gesture recognition to support a wider range of regional and global sign languages, aiming to increase its inclusivity. To boost the voice assistant’s precision and contextual understanding, more advanced natural language processing models, such as those based on transformer architectures, will be integrated. The use of reinforcement learning will also be investigated to help the system better adapt to individual user behaviors and varying environmental factors, thereby improving recognition accuracy. Additionally, efforts will focus on enhancing the user interface and streamlining system performance to provide a more seamless and intuitive experience. Improvements in real-time data handling will aim to minimize latency, while extending compatibility to mobile devices and enabling cloud deployment will broaden user access. These steps will help ensure that VAHI 2.0 remains a robust and accessible-tool-for-assistive-communication.

XI. References

[1] J. Bora et al., “Real-time Assamese sign language recognition using MediaPipe,” Procedia Computer Science, vol. 218, 2023.

[2] S. C. Mesbahi, M. A. Mahraz, Jamal Riffi, and Hamid Tairi, “Hand gesture recognition based on various deep learning YOLO models,” International Journal of Advanced Computer Science and Applications, vol. 14, no. 4, 2023.

[3] Franciskus Antonius, Purnachandra Rao Alapati, Mahyudin Ritonga, Indrajit Patra, Yousef A.Baker El- Ebiary, Myagmarsuren Orosoo, and Manikandan Ren- garajan, “Incorporating Natural Language Processing into Virtual Assistants: An Intelligent Assessment Strategy for Enhancing Language Comprehension,”International Journal of Advanced Computer Science and Applications, vol. 14, no. 10, 2023.

[4] G. L. Sukruth, V. K. BP, M. R. Tejas, K. Rithvik, and T. A. Tharakan, “Enhancing Collaborative Interaction with the Augmentation of Sign Language for the Vocally Challenged,” International Journal of Advanced Computer Science and Applications, vol. 14, no. 1, 2023.

[5] A. Deep et al., “Realtime sign language detection and recognition,” IEEE ASIANCON, 2022.

[6] B. Sundar and T. Bagyammal, “American sign language recognition for alphabets using MediaPipe and LSTM,” Procedia Computer Science, vol. 215, pp. 642–651, 2022.

[7] S. Sharma and S. Singh, “Vision-based hand gesture recognition using deep learning,” Expert Systems with Applications, vol. 182, 2021.

[8] M. Harris and A. S. Agoes, “Applying hand gesture recognition for user guide application using MediaPipe,” ISSAT 2021.

[9] A. T. Ali et al., “Voice recognition system using machine learning techniques,” Materials Today: Proceedings, 2021.

[10] N. A. Nafis and M. S. Hossain, “Speech to text conversion in real-time,” International Journal of Innovation and Scientific Research, vol. 17, 2015.