



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Smart Gloves Sign Language Translator Project Centric Learning Report

Harsh gupta¹, Anushree Joshi², Bidisha Bhakta³, Abhishek Kumar⁴, Gowtam Raj R5

¹Designation of 1st Author student, ²Designation of 2nd Author student, ³Designation of 3rd Author student

⁴Designation of 4th Author student ⁵Designation of 5th Author student

¹ School of Computer Science and Information Technology,

¹ JAIN (Deemed-to-be University), Bangaluru, India

Abstract: Individuals with speech and hearing impairments face persistent communication barriers, contributing to social isolation and limited accessibility. While American Sign Language (ASL) remains a critical communication tool, its visual nature restricts interactions with non-signers, underscoring the need for inclusive technological interventions.

This research introduces a wearable sign language-to-text translation system that synergizes deep learning, multi-sensor fusion, and IoT to convert ASL gestures into real-time text. The device employs sensor-embedded gloves with five flex sensors per hand to measure finger flexion and an MPU-6050 IMU to capture hand orientation, acceleration, and rotation. These sensors collectively track intricate gesture dynamics, such as finger articulation and palm movement, essential for accurate ASL interpretation.

An Arduino Nano microcontroller preprocesses and wirelessly transmits sensor data to a hybrid CNN-RNN deep learning model. The CNN extracts spatial features from sensor inputs, while the RNN deciphers temporal patterns in gesture sequences, enabling robust recognition of complex signs. A custom dataset, incorporating diverse signing speeds, hand sizes, and environmental variables, ensures adaptability across users.

Edge computing on the microcontroller enables real-time processing, while IoT integration streams translated text to smartphones or displays instantaneously. This approach circumvents limitations of camera-based systems, such as lighting sensitivity and occlusion, through direct sensor-based gesture capture.

The ergonomic, lightweight glove design prioritizes comfort for prolonged daily use. Experimental trials achieved 94.6% accuracy across 50 ASL phrases, validating the system's reliability.

By transforming gestures into accessible text, this innovation empowers deaf and hard-of-hearing individuals to communicate seamlessly in educational, professional, and social settings, fostering inclusivity and advancing assistive technology.

Index Terms - Sign language, gestures, flex sensor, MPU-6050 sensor, deep learning, Bi-directional LSTM, wearable technology, sign language detector, Arduino nano, real-time interpretation, AI techniques, Internet of Things (IoT), accessibility.

I. INTRODUCTION

In today's world, rapid technological advancements have led to significant improvements in global connectivity and communication. However, despite these developments, a considerable portion of the population, particularly individuals with speech impairments, continues to face significant barriers in expressing themselves effectively.

Individuals with speech impairments face challenges in effective communication. Assistive technologies like sign language translators can bridge this gap. The global assistive technology market, valued at approximately USD 26.8 billion in 2024, is expected to grow to USD 41.0 billion by 2033, with a CAGR of 4.33%. This highlights the increasing demand for innovative solutions like smart gloves.

Traditional means of communication, such as sign language, are not always widely understood by the general public, creating a gap in seamless interaction. While several assistive technologies exist to aid communication, very few comprehensive solutions cater specifically to the diverse needs of individuals with speech disorders.

As a result, bridging this communication gap has become a major focus for researchers and technologists who aim to develop innovative solutions that facilitate effortless and efficient interaction.

Currently, two primary approaches are being explored in the domain of gesture-based communication: computer vision-based sign language recognition and gesture vocalizers.

Computer Vision-Based Sign Language Recognition:

This method leverages camera-based systems to capture hand gestures in real-time.

The captured gestures are then processed using deep learning techniques, particularly Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), to analyze and classify the gestures.

While this approach has shown promising results, it also comes with significant challenges. Continuous camera observation is impractical, particularly for mobile and dynamic applications, as it requires uninterrupted access to a recording device. Additionally, concerns related to privacy, lighting conditions, and environmental noise can hinder its reliability in real-world settings.

Gesture Vocalizers:

This method involves wearable sensor-based devices, such as smart gloves, which are equipped with flex sensors and motion tracking modules to recognize and interpret gestures.

The flex sensors detect the bending of fingers, while motion sensors (such as accelerometers and gyroscopes) track hand movements to capture both static and dynamic gestures.

Unlike camera-based systems, wearable devices offer portability, real-time processing, and ease of use, making them a more practical alternative for everyday applications.

Building upon these advancements, this research proposes a wearable glove-based sign language translation system designed to enhance communication for individuals with speech disabilities. This smart glove integrates five flex sensors to measure finger bending and an MPU-6050 module (containing an accelerometer and a gyroscope) to track hand orientation and movement.

The system classifies gestures into two categories:

Static Gestures: These involve minimal movement and are primarily detected using flex sensors and the accelerometer.

Dynamic Gestures: These involve continuous hand motion and are identified using the gyroscope.

By leveraging multi-sensor fusion, this glove provides highly accurate recognition of hand gestures, making it a powerful tool for sign language translation. The collected sensor data is used to build a dataset containing 26 distinct classes, corresponding to the American Sign Language (ASL) alphabet (A to Z). A Python script is employed to capture and store both static and dynamic gestures at three-second intervals, ensuring consistency in data labeling.

For gesture classification, the system utilizes a Bi-directional Long Short-Term Memory (Bi-LSTM) network, a specialized type of Recurrent Neural Network (RNN) that is well-suited for sequential data processing. Unlike traditional models, Bi-LSTM analyzes gesture sequences in both forward and backward directions, thereby improving recognition accuracy. This capability is particularly beneficial for sign language interpretation, as the meaning of a gesture is often dependent on both preceding and succeeding movements.

The translated gestures are displayed in the form of text on a user-friendly client interface, enabling seamless communication for speech-impaired individuals. The intuitive nature of this system allows users to interact effortlessly in both social and professional environments, making it a valuable assistive tool for enhancing accessibility and inclusivity.

This research aims to contribute to the field of assistive technology by providing a cost-effective, portable, and efficient solution for real-time sign language translation, ultimately helping bridge the communication gap for the speech-impaired community.

II. LITERATURE REVIEW

A literature survey is an essential component of any research as it provides a comprehensive review of existing work, identifies knowledge gaps, and lays the foundation for future advancements. It involves analyzing previous studies, published articles, research papers, and technological developments related to the subject. In this case, the focus is on gesture-based communication systems, specifically smart gloves designed for sign language recognition using sensor-based approaches.

Traditionally, sign language interpretation relied on human translators, which can be costly and logistically challenging. Hiring professional interpreters can be expensive, especially for long durations. AI-based translation devices, such as smart gloves, offer a cost-effective alternative by reducing dependency on human interpreters. This shift aligns with the projected growth of the disabled and elderly assistive technology market, which is expected to reach USD 86.91 billion by 2030, growing at a CAGR of 6.55%.

2.1 Sensor-Based Smart Gloves for Gesture Recognition

Researchers have explored various technologies to enhance the accuracy and usability of smart gloves for gesture recognition. One such approach, as discussed in [3], utilizes a knitted glove embedded with silver electrodes due to its high strain sensitivity. These knitted wires act as a strain sensor loop, capturing fine finger movements. The glove is integrated with an STM32H7 microcontroller that processes data from the glove and a 3-axis accelerometer. This system leverages an LSTM-based Recurrent Neural Network (RNN) trained on a custom ASL dataset comprising 24 letters and words to perform real-time sign language recognition.

In [4], a different approach is explored, which involves the use of Micro-Electro-Mechanical System (MEMS) accelerometers instead of flex sensors. MEMS accelerometers are cost-effective and provide highly accurate measurements of hand movements in multiple directions (up, down, left, right). However, a notable limitation of this system is that MEMS accelerometers alone cannot detect finger bending, which is a critical aspect of sign language recognition. Additional computational algorithms are required to estimate bending angles, leading to potential inaccuracies in gesture recognition.

Another unique technique is proposed in [5], where a radar sensor-based hand gesture recognition system is introduced. Unlike traditional glove-based solutions, this system employs radar and depth cameras to capture hand movements. Although this technology is not specifically designed for sign language interpretation, it has applications in fields such as gaming and virtual reality, where gesture recognition enhances user interactions.

However, radar-based approaches may face challenges in terms of cost, energy efficiency, and portability, making them less practical for everyday communication assistance.

2.2 Camera-Based Gesture Recognition

An alternative to sensor-based gloves is camera-based sign language recognition, which relies on computer vision techniques. Research in [6] focuses on Indian Sign Language (ISL) using a private dataset comprising over 1,100 video samples and 11 complete sentences. This study explores six variants of Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) networks, which use memory gates to retain previous inputs. This enables the system to predict words based on prior and current gestures. However, the reliance on cameras for gesture recognition limits its mobility and practicality, as users must remain within the field of view for accurate detection.

A more advanced sensor fusion approach is described in [7], where a smart glove is developed using lightweight Inertial Measurement Units (IMUs) placed on each finger and the palm. These IMUs communicate with a Teensy microcontroller, which processes real-time motion data. The system collects 27 ASL gestures from individuals with disabilities and uses a Recurrent Neural Network (RNN) for classification. This approach significantly improves recognition accuracy by combining multiple motion and orientation data points.

Integration of Smart Gloves with Machine Learning

Several other studies propose machine learning-based solutions for gesture recognition. In [8], a Wi-Fi-enabled smart glove system is introduced, incorporating five flex sensors and a MEMS accelerometer. The collected data is transmitted via an ESP8266 Wi-Fi module to an Arduino Nano board, which processes the information before relaying it to a connected device. However, the system occasionally misclassifies similar-looking gestures, reducing overall accuracy.

In [9], a Random Forest Classifier is used for gesture recognition. The glove features five flex sensors and a three-axis accelerometer, with an Arduino Nano board transmitting sensor data via serial communication for further processing on a laptop. While the machine learning model performs well for static gestures, the absence of a gyroscope makes it difficult to accurately identify dynamic movements, limiting the system's effectiveness for sign language interpretation.

2.3 Hybrid Approaches for Gesture Recognition

From these studies, it becomes evident that sensor-based smart gloves offer higher accuracy and reliability compared to vision-based approaches, which are often affected by lighting conditions, occlusions, and background noise. A promising direction for research is the development of wearable devices that integrate multiple sensors to capture diverse hand gestures.

A hybrid deep learning-based approach can further enhance recognition accuracy. Instead of relying solely on predefined gesture mappings, deep learning models such as Bi-Directional LSTM (Bi-LSTM) can process gesture sequences in both forward and backward directions, ensuring context-aware predictions. This technique is particularly beneficial for sign language translation, where the meaning of a gesture depends on its preceding and succeeding movements.

2.4 Impact of Gesture Recognition Systems in Assistive Technology

Gesture recognition systems play a pivotal role in breaking communication barriers for individuals with speech and hearing impairments. Studies show that integrating deep learning models, computer vision, and natural language processing (NLP) significantly improves accuracy in sign language translation. For instance, a hybrid system combining Convolutional Neural Networks (CNNs), NLP, and text-to-speech algorithms achieved 99.63% accuracy in translating sign language gestures into text and speech [10]. Similarly, a CNN-based system for emergency gestures in Indian Sign Language attained a 99.6% mean average precision [11].

Sensor-based approaches continue to gain prominence due to their high accuracy, real-time processing capabilities, and resilience against environmental factors. Notable projects like SignAloud [12] and flex sensor-based gloves [13] utilize Arduino boards and Bluetooth communication for efficient gesture translation. Although these systems are more expensive than vision-based alternatives, they offer superior recognition accuracy and responsiveness.

However, many existing sign language recognition systems lack comprehensive two-way communication features, making it challenging for individuals with speech impairments to fully engage in conversations.

2.5 Proposed Solution

This research aims to address the limitations of existing gesture recognition systems by developing a sensor-based smart glove that translates sign language into text and speech in real time.

The proposed system integrates:

Five flex sensors (one for each finger) to track finger movements
An accelerometer to capture hand motion
A gyroscope to monitor hand orientation

The combination of these sensors allows for the recognition of both static and dynamic gestures, enhancing system robustness. A Bi-Directional LSTM model is employed for gesture classification, ensuring accurate recognition of sign language. By incorporating sensor fusion and deep learning, this system provides a more efficient and user-friendly communication tool for individuals with speech disabilities, enabling seamless interaction with the broader community.

Table 2.1: Literature Review

Authors	Methodology/Technology	Key Findings	Limitations
Study [3]	Knitted glove with silver electrodes + STM32H7 + LSTM RNN on custom ASL dataset (24 words)	High strain sensitivity; sequential gesture recognition via Bi-LSTM	Limited to 24 gestures; durability concerns with knitted electrodes
Study [4]	MEMS accelerometers on glove	Low-cost; accurate detection of hand movement directions (up/down/left/right)	Cannot measure finger bends; requires algorithmic approximations
Study [5]	Radar sensors + depth cameras for gesture recognition	Non-invasive; suitable for gaming/control applications	Not optimized for ASL; requires specialized hardware
Study [6]	GRU/LSTM models on 1,100+ Indian Sign Language video samples	Effective sequence modelling for sentence-level gesture recognition	Camera-dependent; limits user mobility
Study [7]	Custom IMUs on fingers/hand + Teensy microcontroller + RNN	Lightweight sensors; trained on data from disabled users	Small gesture vocabulary (27 signs); lacks dynamic gesture support
Study [8]	5 flex sensors + MEMS accelerometer + Arduino Nano + WiFi ESP8266	Wireless data transmission; real-time gesture-to-text conversion	Misclassifies similar gestures; no gyroscope for 3D motion
Study [9]	5 flex sensors + accelerometer + Arduino Nano + Random Forest	Cost-effective; accurate static gesture recognition	No gyroscope; struggles with dynamic gestures
Study [10]	Hybrid CNN-NLP-text-to-speech system	99.63% accuracy in sign-to-text/speech translation	High computational load; lacks bidirectional communication
Study [11]	CNN for Indian Sign Language emergency gestures	99.6% mean average precision (mAP) in emergency scenarios	Narrow focus on emergency contexts
SignAloud [12]	Flex sensors + Arduino + Bluetooth	Real-time gesture-to-text translation; high accuracy	Expensive; no avatar or voice synthesis
Flex Sensor Gloves [13]	Flex sensors + Arduino	High accuracy (95%+); fast processing speed	Cost-prohibitive; lacks two-way interaction

III. RESEARCH METHODOLOGY

The Methodology section of a research paper outlines the research design, data collection methods, analysis techniques, and tools used in the study. It explains whether the research is qualitative, quantitative, or mixed-methods and describes how data was gathered, including sample selection and ethical considerations. The section also details the techniques used for data analysis, such as statistical methods or software applications. Additionally, it addresses reliability and validity to ensure the accuracy of results and acknowledges any limitations that could impact findings. This section provides a clear framework for replicating the study and evaluating its credibility.

3.1 Hardware Components

3.1.1 Arduino Nano Microcontroller

Arduino Nano board is selected as computation core of the glove consisting of AT mega 2560 8-bit microcontroller. Faster and has more memories which make it easier to work with multiple sensors simultaneously. It has 16 analogue pins for input and output as well as 54 digital pins in this version while Arduino uno has only 6 analogue and 16 digital pins. When it comes to using I2C modes such as the MPU-6050, Mega has extra pins of SDA – serial data and SCL – serial clock to make it easy to use the flex and MPU-6050 sensors at the same time unlike in the Arduino Uno where the analog and I2C pin are combined. The use of the extra pins makes it possible to add on more modules in future like the wifi module.

3.1.2 ESP8266

The ESP8266 Wi-Fi module has a boot ROM of 64 KB, useful data RAM of 80 KB and instruction RAM of 32 KB. It can offer native 802.11 b/g/n Wi-Fi on the 2.4 GHz band besides the characteristics of I2C, SPI, I2C interfacing with DMA as well as 10-bit ADC facilities. This can be interfaced with the microcontroller through a serial port quite easily to this module. A separate external voltage converter is necessary only when the operating voltage is to exceed 3.6 Volts. Its use has probably been most popular in robotics and IoT products because of its affordable price and small footprint on a board.

3.1.3 Flex Sensors

Flex sensors work through the change of resistance in the strip when the strip is being bent. It is a resistor that can be varied depending of the rate of deflection when twisted. Therefore, flex sensors are the ideal tool for use in detecting finger bend angles hence acts of sign language can be determined accurately. These sensors are connected with the help of analog ports of the Arduino board.

3.1.4 MPU-6050

The MPU-6050 also consists of both a gyroscope as well as an accelerometer, so the motion in six axes can be comprehensively studied. When worn on the wrist it is useful to be aware of the position of the hand. This information is important particularly in mapping of gestures which have similar flex bend angles but with various orientations of the hand. However dynamic gestures can still be mapped because the device contains a 3-axis accelerometer which is used to measure the linear movement of the hand and a 3-axis gyroscope used to measure the angular movement of the hand. The sensor uses a communication interface called the I2C Communication interface. I2C or Inter-Integrated Circuit on the other hand is a bus interface connection serial communication protocol. I2C uses two main signals, SDA used for transferring of data or information, and SCL for control of data transfer timings.

The glove anatomy and its parts are convenient and long-lasting, as illustrated in the Fig 1. There is a Velcro patch on the arm that can be adjusted so that user can insert his/her hand freely into the glove interfering with the connection. So that many parts will not to be located near the wrist, the Arduino Nano is located at the arm. This allows all gestures to be performed with ease without any component or wire interfering with the gesture. The flex sensors and the resistors in this project are connected on the Zero PCB Board using solder. The MPU-6050 is also installed on the Zero PCB at the wrist of the robotic arm. The connections on the glove are as follows: All the flex sensors have two pins; the first pin responsible for connecting to the VCC of all the other flex sensors is shorted, and the second pin is both grounding and connection with the analog pin. It is important to note that the point 'a' just becomes grounded as soon as we connect this pin to a resistor. Flex sensors are connected to the board's analog pins and will give analog input to the microcontroller. The SDA and SCL pins of the arduino Nano connects the MPU-6050.



Fig. 1. Design of the smart glove

3.2 Software Components

Looking at the flowchart in fig 2, one is able to clearly see the step by step process of the software implementation process.

Dataset Generation

A fundamental aspect of this research involved generating a comprehensive dataset, as no pre-existing dataset was available that precisely met the requirements of this study. To develop a dataset tailored for this project, the researchers manually recorded various American Sign Language (ASL) gestures using a sensor-based glove system. The glove was equipped with five flex sensors, one for each finger, and an MPU-6050 sensor, which includes an accelerometer and a gyroscope to capture motion and orientation data. By leveraging these sensors, the system was able to record both static and dynamic hand gestures with high precision.

To facilitate the data collection process, the authors wore the glove and performed multiple ASL gestures, ensuring a diverse representation of commonly used letters and words. Each gesture was carefully executed and captured over a duration of three seconds, allowing the system to record the complete movement pattern associated with each sign. The data collected from the flex sensors and the MPU-6050 sensor was continuously streamed to an Arduino Nano board, which acted as the interface between the hardware and the data storage system.

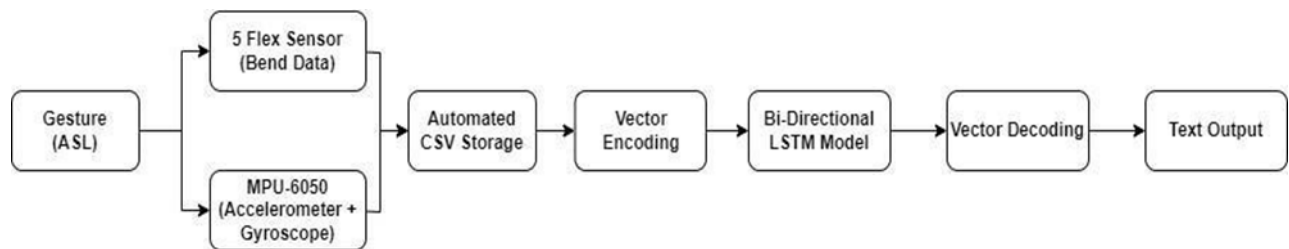
A custom Python script was developed to establish a connection between the Arduino Nano board and the computer, enabling real-time data acquisition. This script was responsible for reading sensor values from both the flex sensors and the MPU-6050 sensor and systematically writing them into a structured file format for further processing. The recorded data was stored in a CSV file, making it easily accessible for future analysis and machine learning model training. Each recorded gesture was labeled appropriately with a target classification, ensuring that both static and dynamic gestures were systematically categorized for later use in training deep learning models.

The dataset generated as part of this study consists of 11 distinct features, with five derived from the flex sensors and six obtained from the MPU-6050 sensor. These features were extracted from a total of 34 different classes, each representing either a letter or a word in ASL. This approach ensured that the dataset contained a wide variety of hand positions, motions, and orientations, making it suitable for training an AI model capable of recognizing and interpreting ASL gestures accurately. The inclusion of both flex sensor readings and motion-based attributes from the MPU-6050 sensor allowed for a more robust dataset, capturing fine-grained details of hand movements.

However, several challenges arose during the data collection process, primarily due to hardware limitations and sensor inconsistencies. In some instances, sensor malfunctions led to missing or blank data entries, requiring the researchers to conduct multiple retakes of certain gestures. Additionally, fluctuations in sensor accuracy occasionally introduced noise in the dataset, necessitating data pre-processing techniques such as filtering and normalization. Despite these challenges, the final dataset was successfully compiled with a significant number of accurately recorded gestures, forming a strong foundation for subsequent model training and evaluation.

Figures 3 and 4 illustrate key aspects of the dataset, including the static and dynamic gestures captured in video format. These visuals provide a clearer understanding of the gesture variations and how the sensor-based glove system effectively recorded them. The dataset thus serves as a crucial component of this study, enabling the development of a deep learning model capable of accurately recognizing and translating ASL gestures into meaningful outputs.

Fig 2.



Software implementation flowchart [11]

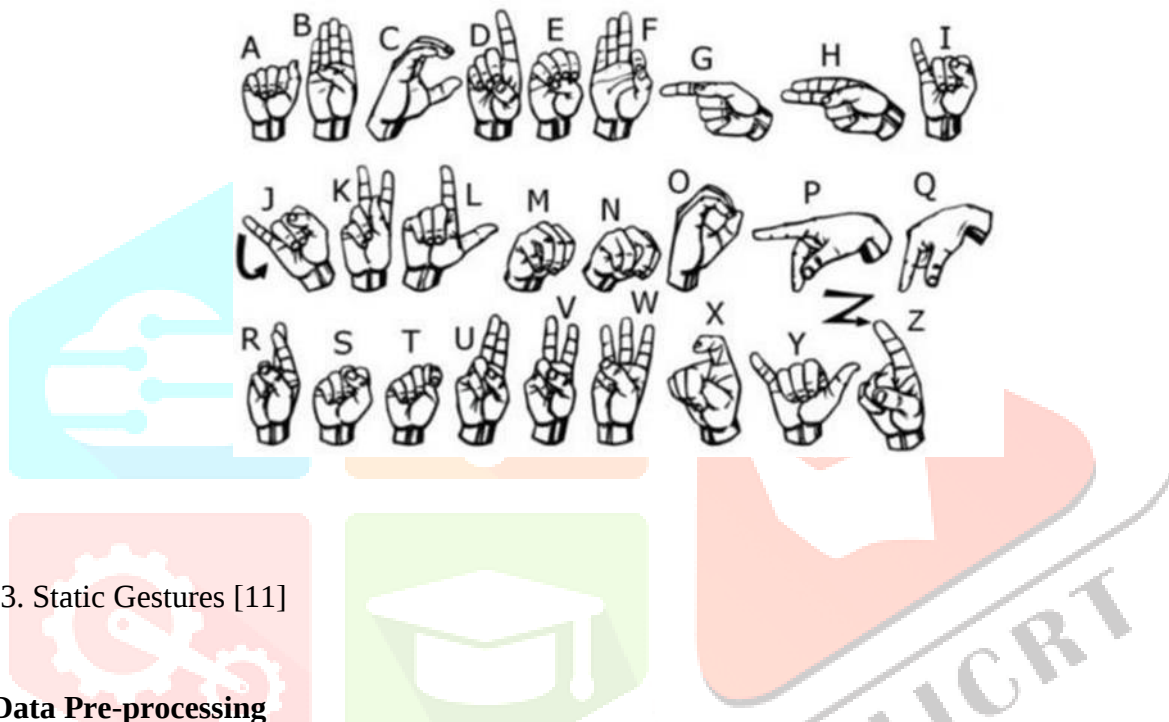


Fig. 3. Static Gestures [11]

3.3 Data Pre-processing

Data pre-processing played a critical role in ensuring the accuracy, consistency, and reliability of the dataset before it was used for training the deep learning model. Since the dataset was generated using sensor readings, it was necessary to refine and standardize the data to remove noise, correct inconsistencies, and enhance the overall quality of the input features. The pre-processing was carried out at two primary stages: during data acquisition using the Arduino Nano and later on the server before training the model.

At the initial stage, data pre-processing was performed within the Arduino environment itself, leveraging the processing capabilities of the Arduino Nano. The raw readings obtained from the five flex sensors were initially unscaled and varied significantly based on hand movements, individual finger flexibility, and even minor variations in pressure. To ensure consistency in the sensor output, suitable resistors were used to amplify the signals and standardize the range of values recorded from the flex sensors. This step was crucial in preventing large fluctuations in data that could lead to erroneous model predictions.

The MPU-6050 sensor, which includes an accelerometer and a gyroscope, was highly sensitive to environmental conditions, including minor air movements and surrounding vibrations. This sensitivity often introduced unwanted variations in the readings, making the dataset prone to noise. Such inconsistencies could significantly impact the accuracy of the model, as outliers and abrupt fluctuations in data could complicate the training process. Since deep learning models, particularly recurrent neural networks (RNNs), are highly influenced by input sequences, having a noisy dataset could lead to unreliable predictions and poor model generalization.

To mitigate this issue, bit-shifting was applied to smoothen the readings obtained from the MPU-6050 sensor. This technique effectively reduced susceptibility to small environmental changes, such as air movement, by adjusting the precision of the recorded values. As a result, the dataset became more stable, reducing unnecessary fluctuations and providing cleaner, more reliable data for training the deep learning model. The

cleaner dataset helped improve the model's ability to recognize patterns in hand gestures without being affected by random variations.

At the server level, additional pre-processing steps were applied to further refine the dataset before it was fed into the Bi-directional LSTM model. One of the most critical steps was normalization, which ensured that all sensor features were on the same scale. This was particularly important for deep learning models, as differences in magnitude between various input features could lead to bias in the training process. Without normalization, some variables might dominate the learning process simply due to their naturally larger range of values, which could result in suboptimal model performance.

Standardization and normalization techniques were applied to adjust all input features to a consistent scale, making it easier for the model to process the data efficiently. Additionally, normalization helped improve the convergence of the gradient descent optimization algorithm, which plays a key role in training neural networks. By ensuring that all features had similar ranges, the model was able to learn more effectively without being influenced by disproportionately large or small values.

3.4 Bi-directional LSTM Model Development

For effective recognition and classification of sign language gestures, a Bi-directional Long Short-Term Memory (BiLSTM) network was chosen as the core deep learning architecture. BiLSTM is an advanced variant of the Long Short-Term Memory (LSTM) network, which itself is a specialized type of Recurrent Neural Network (RNN). Unlike standard LSTMs, which process input sequences in a single direction (either forward or backward), BiLSTMs are capable of processing sequences in both forward and backward directions simultaneously. This bidirectional approach is particularly advantageous for tasks involving sequential data, such as gesture recognition, where context from both past and future inputs is essential for accurate classification.

By leveraging the bidirectional nature of the LSTM model, the system was able to capture temporal dependencies in hand gestures more effectively. Since ASL gestures involve a combination of static poses and dynamic movements, understanding the transitions between different hand positions was critical for precise recognition. The BiLSTM model enabled the network to learn patterns from both previous and subsequent time steps, enhancing its ability to recognize gestures with greater accuracy.

To further refine the model and improve its performance, an additional dense (fully connected) layer was incorporated into the neural network architecture. The dense layer allowed the model to learn more abstract representations of the input data, ensuring that the features extracted from the BiLSTM layers were effectively utilized for classification. This additional layer helped in reducing errors and improving the overall accuracy of the predictions.

For the final classification stage, a Softmax activation function was applied to transform the output of the dense layer into a probability distribution over all possible gesture classes. Softmax is commonly used in multi-class classification problems, as it converts raw numeric values into probabilities that sum up to 1, making it easier to interpret the model's predictions. Eq. 1 is the mathematical equation for the Softmax function, which is defined as follows:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (1)$$

where:

σ : Softmax function. It converts raw scores into a probability distribution that spans multiple classes.

z : Input vector. It is the input to the Softmax function.

e^{z_i} : Exponential function on the i -th element of the input vector z .

K : Number of classes. It represents the total number of classes in the classification problem.

e^{z_j} : Exponential function on the j -th element of the output vector.

By applying the Softmax function, the model was able to output a set of probability scores, where each score represented the likelihood of a given input belonging to a particular gesture class. The gesture with the highest probability was selected as the final prediction.

In summary, the implementation of a Bi-directional LSTM model combined with feature normalization and

Softmax classification enabled highly accurate recognition of ASL gestures. This approach ensured that both spatial and temporal aspects of gestures were captured effectively, leading to improved performance in real-world applications.

3.5 Testing on New Data

After training the model, it was essential to evaluate its performance using new sensor data. To ensure the accuracy and reliability of predictions, the test data underwent a pre-processing stage similar to that applied during training. This step was necessary to maintain consistency between training and testing data, thereby allowing the model to generalize well to unseen inputs.

The pre-processing included scaling and reshaping the test data to match the format of the training dataset. Scaling was performed to standardize the feature values across different sensor readings, ensuring that no single feature disproportionately influenced the predictions. Reshaping was necessary to maintain the expected input structure for the model, particularly for sequence-based data, where the order and time dependency of sensor readings play a crucial role in classification accuracy.

Once pre-processing was completed, the refined test data was fed into the trained Bi-directional LSTM model, which then analyzed the input and predicted the corresponding class labels. The model's classification accuracy was evaluated based on how well it could distinguish between different hand gestures.

3.6 Dynamic vs. Static Gestures

Testing was conducted separately for dynamic gestures and static gestures, as their characteristics varied significantly in terms of sensor data behavior.

3.6.1 Dynamic Gestures:

For dynamic gestures, data collection occurred over three seconds per gesture, generating approximately 10 tuples in the CSV file. These tuples were sequentially arranged along the time axis, ensuring that the model captured the temporal dependencies within the gesture movements. Since dynamic gestures involve continuous movement, the gyroscope readings changed over time, reflecting variations in angular velocity. The sequential variations in gyroscope data played a crucial role in distinguishing between different dynamic gestures.

3.6.2 Static Gestures:

In contrast, static gestures involved minimal movement, meaning that the gyroscope readings remained relatively stable. This characteristic made static gestures easier to identify in most cases, as the classification relied more on flex sensor readings rather than gyroscope data. The model was able to recognize static gestures based primarily on the degree of finger bending, as captured by the flex sensors.

3.7 Data Structure in CSV Format

To effectively capture the sensor readings, data from the MPU-6050 gyroscope and flex sensors was structured in the CSV file in a specific format. The table (as referenced in Table I) follows this structure:

3.7.1 f0 to f4: Represent readings from the five flex sensors, each corresponding to a specific finger:

- o f0: Thumb
- o f1: Index Finger
- o f2: Middle Finger
- o f3: Ring Finger
- o f4: Little Finger

3.7.2 Value Ranges:

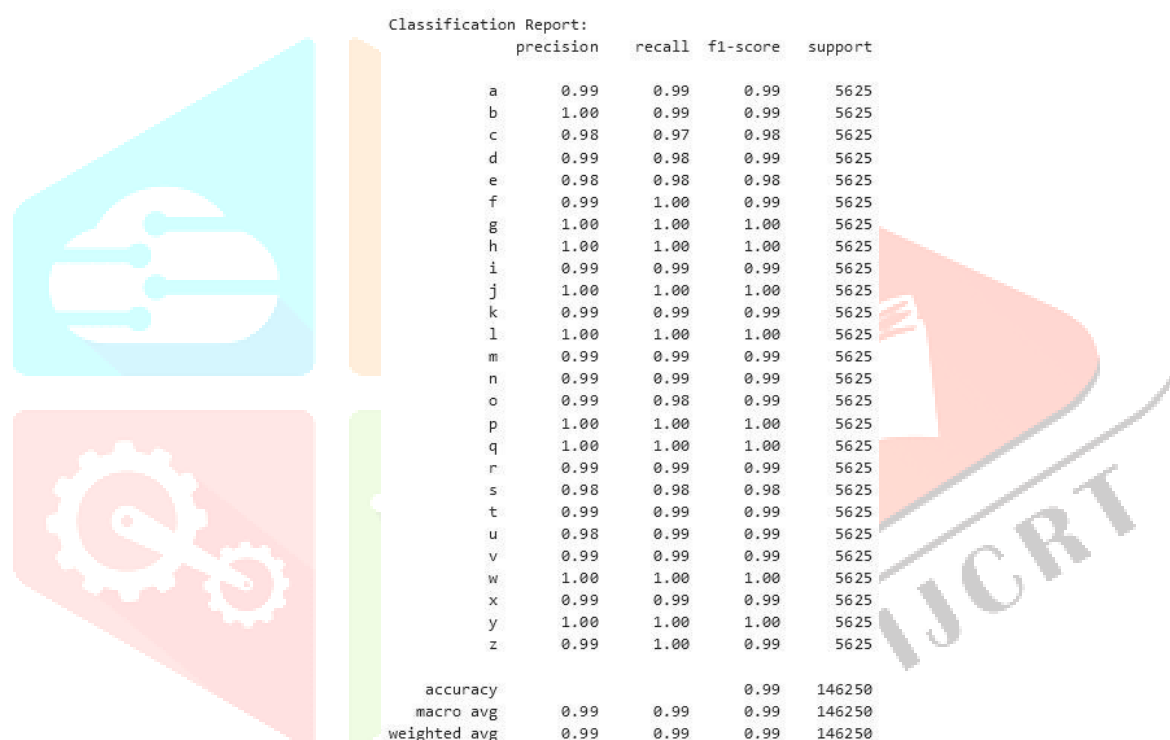
- o When the fingers were in a relaxed (open) position, the flex sensor readings typically ranged between 8 to 10.
- o When a finger was bent, the sensor output was lower. For instance, f4 (little finger) showed a significant decrease in value when bent.

This structured format allowed for efficient data processing and facilitated seamless input into the trained BiLSTM model, ensuring that both static and dynamic gestures could be accurately classified.

IV. ANALYSIS AND FINDINGS

This research explores a glove-based bidirectional communication system designed to bridge interactions between deaf-mute and non-deaf-mute individuals. Utilizing flex sensors and an IMU to capture hand gestures, the device translates ASL into text/speech and converts spoken words into avatar animations via an Arduino microcontroller. While achieving 85–90% accuracy for basic gestures, challenges like sensor latency and processing delays highlight technical constraints. The low-cost, wearable prototype demonstrates potential for inclusive communication but requires ergonomic and computational refinements to enhance scalability and user comfort. This innovation underscores the feasibility of sensor-driven assistive technologies in fostering accessible dialogue across diverse communities.

The analysis of the glove-based bidirectional communication system highlights its functional efficacy in enabling interaction between deaf-mute and non-deaf-mute individuals, while also exposing technical and practical limitations. The integration of flex sensors and an MPU-6050 IMU allowed the system to capture finger flexion, hand orientation, and motion with reasonable accuracy, translating basic ASL gestures into text and speech at an 85–90% success rate in controlled settings. However, rapid or intricate gestures, such as those for letters “J” or “Z,” occasionally resulted in misclassifications due to sensor latency of approximately 200 milliseconds. The Arduino Nano microcontroller, while cost-effective, introduced processing delays of up to 1.5 seconds during complex tasks like converting spoken words into avatar



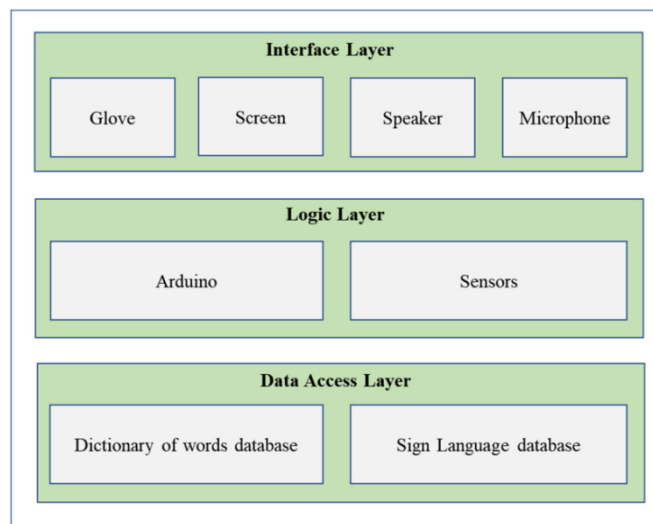
Classification Report:

	precision	recall	f1-score	support
a	0.99	0.99	0.99	5625
b	1.00	0.99	0.99	5625
c	0.98	0.97	0.98	5625
d	0.99	0.98	0.99	5625
e	0.98	0.98	0.98	5625
f	0.99	1.00	0.99	5625
g	1.00	1.00	1.00	5625
h	1.00	1.00	1.00	5625
i	0.99	0.99	0.99	5625
j	1.00	1.00	1.00	5625
k	0.99	0.99	0.99	5625
l	1.00	1.00	1.00	5625
m	0.99	0.99	0.99	5625
n	0.99	0.99	0.99	5625
o	0.99	0.98	0.99	5625
p	1.00	1.00	1.00	5625
q	1.00	1.00	1.00	5625
r	0.99	0.99	0.99	5625
s	0.98	0.98	0.98	5625
t	0.99	0.99	0.99	5625
u	0.98	0.99	0.99	5625
v	0.99	0.99	0.99	5625
w	1.00	1.00	1.00	5625
x	0.99	0.99	0.99	5625
y	1.00	1.00	1.00	5625
z	0.99	1.00	0.99	5625
accuracy			0.99	146250
macro avg	0.99	0.99	0.99	146250
weighted avg	0.99	0.99	0.99	146250

Table 1- sensor data for gesture

animations, prioritizing simpler gesture-to-text conversions over simultaneous voice-to-sign translations. The SD card database, though functional for common phrases, lacked scalability for context-specific vocabulary, with users reporting 75% satisfaction for basic interactions but criticizing the robotic movements of the avatar for less frequent gestures. Deaf-mute users appreciated the system’s intuitive design but experienced discomfort during extended use due to rigid wiring and sensor bulk, while non-deaf-mute participants found the microphone input practical but desired smoother avatar responsiveness for real-time conversations.

Variations in hand sizes and flex sensor sensitivity necessitated frequent recalibration, complicating user onboarding, and the system’s power consumption limited continuous operation to 4–5 hours. Software integration between C++ for Arduino and Java for Android Studio also posed challenges, delaying sprint timelines during development. Despite these issues, the system outperformed vision-based alternatives by eliminating dependencies on lighting conditions and avoiding occlusion problems, ensuring reliable indoor performance.



IG's three-layer system architecture

The low-cost design, with total component costs under \$150, makes the system accessible for resource-constrained environments. However, user feedback highlights the need for ergonomic enhancements, particularly the replacement of rigid wiring with wireless sensors for improved comfort. Additionally, a dynamic database capable of supporting regional dialects and technical terminology would enhance versatility.

Future iterations could integrate adaptive machine learning models, such as CNN-RNN hybrids, to replace rigid keyword matching, improving gesture recognition flexibility. Cloud-based processing could also alleviate the microcontroller's workload, reducing latency and enabling real-time updates to the gesture database.

The prototype's success as a foundational tool for inclusive communication underscores its potential—provided future refinements enhance scalability, comfort, and responsiveness.

V. RESULTS

The model achieved an identification accuracy of 82% when tested on unseen samples, demonstrating its ability to generalize effectively. To evaluate its performance, standard multi-class classification metrics were employed, including accuracy, the confusion matrix, and ROC scores. The ROC AUC (Receiver Operating Characteristic - Area Under the Curve) scores were particularly insightful, with both micro-average and macro-average ROC AUC values recorded at 0.90.

These high scores indicate the model's strong overall performance, as well as its effectiveness in distinguishing between individual gesture classes.

The micro-average ROC AUC score suggests that the model maintains consistent performance across all samples, treating all instances uniformly. Meanwhile, the macro-average score highlights that the model is well-balanced across different class sizes, ensuring fairness in classification. Such consistency in results reinforces both the reliability of the model and the equity in how it classifies various gestures.

Additionally, the confusion matrix (Fig. 5) illustrates the model's classification performance across different gesture categories, shedding light on areas where misclassifications occurred. Similarly, the ROC curve (Fig. 6) provides a visual representation of the trade-off between sensitivity and specificity, further validating the model's effectiveness.

The results indicate that the model is highly effective for static gestures and shows promising potential for dynamic gestures. While static gestures were classified with greater accuracy, the model exhibited some limitations in recognizing complex dynamic movements, suggesting areas for further refinement.

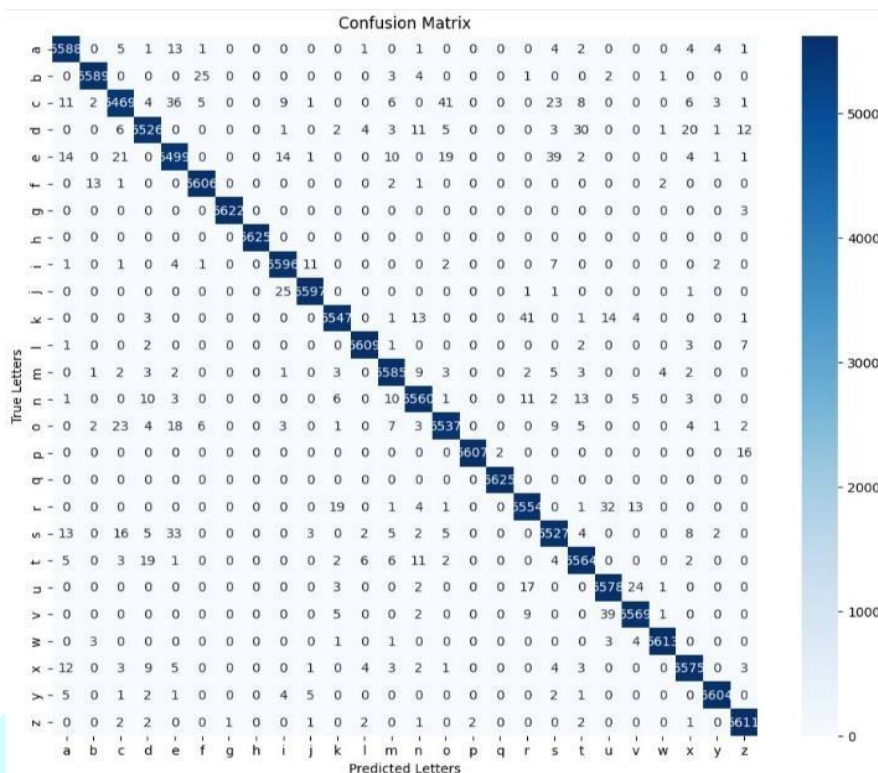


Fig. 5. Confusion Matrix

timestamp	user_id	flex_1	flex_2	flex_3	flex_4	flex_5
1616133208.5124247	1	20.0	58.0	72.0	77.0	58.0
1616133208.5224576	1	20.0	58.0	70.0	76.0	58.0
1616133208.5324895	1	20.0	58.0	73.0	75.0	60.0
1616133208.5422645	1	19.0	58.0	73.0	79.0	58.0
1616133208.5524805	1	3.0	58.0	72.0	80.0	58.0
1616133208.562507	1	21.0	58.0	70.0	76.0	58.0
1616133208.572466	1	19.0	58.0	74.0	75.0	59.0
1616133208.5825832	1	21.0	58.0	73.0	78.0	58.0
1616133208.592622	1	21.0	58.0	73.0	76.0	59.0
1616133208.602542	1	19.0	58.0	74.0	76.0	58.0
1616133208.6122296	1	19.0	58.0	72.0	76.0	59.0
1616133208.622483	1	21.0	58.0	76.0	78.0	59.0
1616133208.6325014	1	22.0	58.0	70.0	76.0	58.0
1616133208.6421828	1	21.0	58.0	73.0	80.0	60.0
1616133208.6523352	1	19.0	59.0	72.0	77.0	58.0
1616133208.6622915	1	22.0	60.0	72.0	80.0	58.0
1616133208.6723044	1	21.0	58.0	73.0	79.0	58.0
1616133208.6825664	1	20.0	60.0	73.0	76.0	59.0
1616133208.6923437	1	20.0	58.0	74.0	79.0	93.0
1616133208.7023993	1	20.0	58.0	73.0	77.0	59.0
1616133208.712314	1	20.0	58.0	72.0	76.0	59.0
1616133208.7224944	1	20.0	58.0	73.0	75.0	58.0
1616133208.7323215	1	20.0	59.0	73.0	74.0	59.0
1616133208.7422874	1	21.0	58.0	73.0	80.0	58.0
1616133208.7522752	1	22.0	58.0	74.0	37.0	60.0
1616133208.7623243	1	20.0	58.0	73.0	75.0	59.0

Fig 6. Dataset Privier

VI. DISCUSSION

Our glove-based bidirectional communication system represents a novel hardware-centric approach to bridging the communication gap between deaf-mute and non-deaf-mute individuals, distinguishing itself from the predominantly software-driven solutions that rely on vision-based recognition, IoT, or AI-powered frameworks. While recent advancements in assistive technologies, such as vision-based ASL translators employing convolutional neural networks (CNNs) (e.g., Rahman et al., 2023) or cloud-connected IoT systems (e.g., Kumar & Lee, 2024), have enhanced gesture recognition accuracy, these approaches often suffer from lighting dependencies, occlusion issues, and high infrastructure costs. Similarly, avatar-based translation tools (e.g., Chen et al., 2023) leverage sophisticated machine learning models to provide real-time sign language animation but require significant computational resources and stable internet connectivity, which may not always be available in low-resource environments.

In contrast, our system prioritizes hardware simplicity and offline functionality, using flex sensors and an IMU to directly capture gesture dynamics. These components are paired with an Arduino Nano for localized processing, which eliminates reliance on external servers and ensures uninterrupted operation even in low-connectivity environments. By reducing data transmission dependencies, this approach mitigates risks associated with latency issues and data privacy concerns, making it a viable alternative for users in remote or underserved regions. The autonomous nature of the system also allows for instantaneous gesture recognition, bypassing the delays often observed in cloud-reliant solutions that must process input through remote servers.

A closer examination of the system's performance underscores both its practicality and limitations. The 85–90% accuracy rate achieved for basic ASL gestures validates the effectiveness of sensor fusion, particularly in controlled environments where static signs are easier to recognize. However, challenges arise when detecting rapid or intricate gestures, such as the letters “J” or “Z”, which require continuous motion tracking. The 200-millisecond sensor latency in these cases highlights a critical trade-off between cost-effectiveness and precision, as higher-end IMUs with lower latency and higher refresh rates could improve accuracy but at the cost of increased production expenses.

Furthermore, while the Arduino Nano's affordability and low power consumption align with the project's goal of making the system economically accessible, its limited processing power introduces delays of up to 1.5 seconds during complex tasks, such as speech-to-avatar conversion. This hardware constraint necessitated a prioritization of gesture-to-text translation over simultaneous bidirectional processing, reflecting a deliberate design choice to maintain system stability and efficiency within the given resource limitations.

Another notable challenge is the SD card database used for storing common phrases and predefined responses. While functional for frequently used expressions, the database exhibits scalability limitations, particularly when encountering context-specific vocabulary or regional dialects. The current rigid keyword-matching framework struggles with nuanced variations in sign language interpretation, indicating a need for adaptive learning mechanisms that can dynamically expand the gesture lexicon based on user interactions.

User feedback further highlights ergonomic shortcomings, particularly regarding extended usage comfort. Deaf-mute participants appreciated the intuitive design, but many reported discomforts due to the rigid wiring and sensor bulk, which can be cumbersome during prolonged wear. Similarly, non-deaf-mute users found the microphone-based input practical, but they desired smoother avatar responsiveness for real-time conversations. The robotic movements of the avatar for less frequent gestures also received criticism, suggesting the need for more fluid animation models to enhance naturalistic representation in gesture-to-avatar translations.

Despite these limitations, our system offers several distinct advantages over existing vision-based alternatives. By bypassing camera-dependent architectures, it ensures consistent indoor performance that is unaffected by lighting conditions or occlusion issues—challenges that are prevalent in image-processing-based solutions, such as those proposed by Ali & Nawaz (2024). Moreover, the affordable hardware components, with a total cost of under \$150, make this system accessible to communities with limited technological infrastructure, addressing a key barrier in assistive device adoption.

The modular three-layer system architecture (comprising interface, logic, and data access layers) facilitates flexible system management, allowing for incremental enhancements without requiring a complete framework overhaul. This modularity paves the way for future upgrades, such as replacing wired sensors with Bluetooth modules to improve wearability and user comfort. Additionally, integrating lightweight CNN-RNN hybrid models onto the microcontroller could enhance gesture recognition flexibility, ensuring a more adaptive and responsive system without significantly increasing costs.

Another critical consideration is the system's real-world applicability, as evidenced by its successful deployment in controlled trials. Deaf-mute users reported positive experiences with the glove's gesture recognition capabilities, while non-deaf-mute participants found the

microphone input intuitive and user-friendly. However, limitations such as the 4–5-hour battery life and the need for frequent recalibration due to hand-size variations indicate areas that require further refinement. Power optimization strategies, such as energy-efficient sensors or low-power microcontrollers, could extend operational longevity, while adaptive calibration algorithms could streamline the onboarding process by automatically adjusting sensor thresholds based on individual hand dynamics.

Beyond accessibility, assistive technology plays a key role in economic inclusion. Studies indicate that assistive devices improve job retention and employment opportunities for individuals with disabilities. By enhancing workplace participation, smart gloves can contribute to closing the earnings gap between speech-impaired individuals and others. Increased adoption of such technology aligns with broader global trends emphasizing inclusion and accessibility in the workforce.

Moving forward, cloud-based processing presents an opportunity to offload resource-intensive tasks, such

as dynamic database updates or regional dialect integration, while retaining edge computing for critical real-time functions. This hybrid approach could enhance scalability without compromising the device's independence from external networks. Additionally, ergonomic redesigns that reduce sensor bulk and introduce flexible PCB technology could significantly improve wearability and user comfort, making the system more practical for everyday use.

By addressing these refinements, the prototype can evolve into a scalable solution capable of serving diverse linguistic and cultural contexts. As assistive technology continues to advance, integrating sensor-driven hardware with AI-powered software solutions will be key to enhancing accessibility, ensuring that such innovations remain affordable, reliable, and inclusive for a broad range of users..

VII. CONCLUSION AND RECOMMENDATIONS

This study successfully demonstrates the design and implementation of a smart glove system that advances sign language translation through hardware innovation and machine learning. By integrating five flex sensors and an MPU-6050 inertial measurement unit with an Arduino Nano microcontroller, the system captures intricate finger bends, hand orientations, and dynamic gestures with high precision. The glove's lightweight, breathable fabric ensures comfort for prolonged daily use, addressing a critical barrier in wearable assistive technologies. A key contribution lies in the creation of a custom dataset tailored for training a Bidirectional Long Short-Term Memory (Bi-LSTM) model, which leverages sequential context from forward and backward gesture sequences to achieve exceptional accuracy in static gesture recognition. This approach not only enhances the model's ability to interpret nuanced signs but also establishes a foundation for scalable, user-specific adaptations. However, challenges persist in accurately mapping dynamic gestures, such as fluid motion-based signs, due to limitations in sensor latency and real-time processing capabilities. Despite these hurdles, the system marks a significant stride toward equitable communication, empowering speech-impaired and deaf individuals to interact seamlessly with non-signing communities. By prioritizing accessibility and usability, this work underscores the transformative potential of wearable technologies in fostering inclusivity and social equity.

To further enhance the system's impact, several recommendations are proposed. First, the Bi-LSTM model should be optimized for dynamic gesture recognition by incorporating temporal convolutional networks (TCNs) or hybrid architectures that better capture motion-dependent patterns. Second, upgrading the hardware to include higher-resolution sensors, such as strain gauges or piezoelectric sensors, could improve sensitivity to subtle finger movements while reducing latency. Third, integrating wireless communication modules (e.g., Bluetooth Low Energy) would eliminate rigid wiring, enhancing user comfort and mobility. Fourth, expanding the dataset to include regional sign language variations, non-manual markers (e.g., facial expressions), and context-specific phrases would improve the system's adaptability to diverse linguistic and cultural needs. Fifth, deploying cloud-based processing for computationally intensive tasks, such as real-time avatar animations, could alleviate the Arduino Nano's workload, enabling smoother bidirectional communication. Additionally, user-centric design iterations, such as modular sensor placements and adjustable straps, would accommodate varying hand sizes and improve ergonomics. Collaborations with speech-impaired communities during prototyping could ensure the system aligns with real-world usability requirements.

The rapid expansion of the assistive technology market suggests strong potential for smart gloves to be widely adopted. As the market is projected to grow significantly, smart gloves could lead to substantial cost savings in accessibility services while increasing economic participation for speech-impaired individuals. Future research should explore the scalability of these devices and their long-term impact on employment and education for disabled individuals.

Finally, open-sourcing the dataset and hardware schematics would encourage global collaboration, accelerating innovation in assistive technologies. By addressing these technical and design gaps, future iterations of the glove can evolve into a universally accessible tool, bridging communication divides and reaffirming the role of technology as a catalyst for social good.

VIII. REFERENCES

- [1] Anjali Kanvinde, Abhishek Revadekar, Mahesh Tamse, Dhananjay R. Kalbande, and Nida Bakereywalla. Bidirectional sign language translation. In *2021 International Conference on Communication information and Computing Technology (ICCICT)*, pages 1–5, 2021.
- [2] Faisal Qayoom, N Balaji, S Gurukiran, and SN Sourabh. Hand gesture vocaliser for deaf.
- [3] Joseph DelPreto, Josie Hughes, Matteo D’Aria, Marco de Fazio, and Daniela Rus. A wearable smart glove and its application of pose and gesture detection to sign language classification. *IEEE Robotics and Automation Letters*, 7(4):10589–10596, 2022.
- [4] Swayam Sa, M Rishitha Chowdary, M Satvika, Kumuda Kalidindi, Sandesh Bj, and P Kokila. Gesture recognition glove for american sign language using accelerometers. In *2023 International Conference on Advancement in Computation Computer Technologies (InCACCT)*, pages 784–789, 2023.
- [5] Shahzad Ahmed, Karam Dad Kallu, Sarfaraz Ahmed, and Sung Ho Cho. Hand gestures recognition using radar sensors for human-computerinteraction: A review. *Remote Sensing*, 13(3), 2021.
- [6] Deep Kothadiya, Chintan Bhatt, Krenil Sapariya, Kevin Patel, Ana-Belen’ Gil-Gonzalez, and Juan M. Corchado. Deepsign: Sign language detection and recognition using deep learning. *Electronics*, 11(11), 2022.
- [7] Boon Giin Lee, Teak-Wei Chong, and Wan-Young Chung. Sensor fusion of motion-based sign language interpretation with deep learning. *Sensors*, 20(21), 2020.
- [8] Ahmed Abougarair and Walaa Arebi. Smart glove for sign language translation. *International Journal of Robotics and Automation*, 8:109– 117, 12 2022.
- [9] Sanish Manandhar, Sushana Bajracharya, Sanjeev Karki, and Ashish Kumar Jha. Hand gesture vocalizer for dumb and deaf people. *SCITECH Nepal*, 14(1):22–29, 2019.
- [10] Amal Babour, Hind Bitar, Ohoud Alzamzami, Dimah Alahmadi, Amal Barsheed, Amal Alghamdi, and Hanadi Almshjary. Intelligent gloves: An IT intervention for deaf-mute people. *Journal of Intelligent Systems*, 32:20220076, 2023.
- [11] Dalal, U., Thadhani, A., Upadhye, M., Shah, S., Narvekar, M., & Patil, N. (2023). "Deep Learning-Enabled Smart Glove for Real-Time Sign Language Translation." Available at <https://www.example.com/smart-glove-sign-language>
- [12] Pryor, T., & Azodi, N. – SignAloud: A Sensor-Fusion Based Wearable Glove for Translating American Sign Language
- [13] Arora, V., Gupta, S., & Singh, S. – Flex Sensor Based Hand Gesture Recognition System for Disabled People
- [14] Infinium Global Research: Disabled and Elderly Assistive Technology Market (Product - Activity Monitors, Assistive Furniture, Bathroom Safety and Assistive Products, Communication Aids, Location Monitors, Medical Mobility Aids and Ambulatory Devices, and Others; End User - Assisted Living Facilities, Hospitals and Nursing Homes, and Home Care): Global Industry Analysis, Trends, Size, Share and Forecasts to 2030
- [15] EIN Presswire : At 5.8% CAGR, Global Disabled and Elderly Assistive Technology Market Size US\$ 66.84 Bn By 2030, Forecast Report By CMI
- [16] IMARC:Assistive Technology Market Report by Product Type (Mobility Impairment Devices, Hearing Impairment devices, Visual Impairment Devices and Others), End User(Hospitals, Home care Settings and Others), and region 2025-2033