International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 12 | Dec 2025 www.irjet.net p-ISSN: 2395-0072

Sensor-Based Sign Language Recognition using a Smart Glove and Hybrid Deep Learning Models (Conv1D + BiLSTM).

S.P. Kullarkar [1], Anshu Gudhewar [2], Sidhesh Dhande [3], Ravi Shripad [4], Kautuk Butle [5]

[1]*Assistant Professor, Dept. of AI&DS, KDK. College Of Engineering, Maharashtra, India*
[2]*Student, Dept. of AI&DS, KDK. College Of Engineering, Maharashtra, India*
[3]*Student, Dept. of AI&DS, KDK. College Of Engineering, Maharashtra, India*
[4]*Student, Dept. of AI&DS, KDK. College Of Engineering, Maharashtra, India*
[5]*Student, Dept. of AI&DS, KDK. College Of Engineering, Maharashtra, India*

Abstract -*Communication is key to human life, but for those who are deaf or Problem of hearing, communication across species borders can be complicated, not everyone knows sign language. This creates a communication outlet which restricts their ability to interact socially, attend school or work in public. To fix this, our project proposes AI Based Sign Language Interpreter Glove, that will convert any hand gesture for sign language into a speech or text.*

The glove is designed to include flex sensors and motion sensors to measure fingers and hands bending and moving. That information goes to a microcontroller, which then processes the data and sends it to an AI model/Deep Learning Model trained to interpret various sign language gestures. After the detection of the signs, the signs are translated into text appearing on the monitor or the display attach to Glove then Converted into voice through a Text-to-Speech (TTS) system.

The presented solution is low-cost, portable, and easy-to-use, allowing daily use. It has the possibility of creating the conditions for individuals with disabilities to truly be themselves, participate in society, and experience the potential they have by allowing them to communicate freely and effectively with normal People.

The most important advantage of this project was its practicability. Compared to vision-based systems based on cameras and controlled environments, a glove-based solution is low-weight, portable and not reliant upon the lighting or background. It is also economical and affordable, making it an appropriate device for daily aid. The design is also scalable, too, so more gestures and more than one sign language can be included in the future.

Key Words: **Sign Language, Interpreter Glove, Artificial Intelligence, Deep Learning, Flex Sensors, Motion Sensor, Communication, Text-to-Speech, microcontroller, Embedded Systems**

1. INTRODUCTION

Communication is the foundation of human development, the individuals with speech and hearing Problem faces significant barriers in expressing themselves and not being understand in real time. Traditional sign language, are highly used within the deaf and mute People and is not widely understand by the Majority population, thus creating a communication gap between them. This gap must be filled by a new type of solution that would read sign language, interprets it and translate it to a universal format such as text or speech.

This work is aiming to explore AI-powered smart wearable glove for near real-time sign language detection and translation. Multiple sensors (e.g., flex sensors, accelerometers, and gyroscopes) will be mounted on the glove to measure precise finger bending motions, hand posture and gestures. These sensor signals will be analyzed with the help of Advance deep learning techniques to reliably capture sign language patterns. Upon detection of the signs, the system will convert them into text or speech output so that sign language users and non-signers could communicate smoothly.

The inclusion of Internet of Things (IoT) data connectivity also makes the system versatile in that via wireless or wired means, the identified output can be viewable on smart phones, computers, terms etc. This not only makes the technology portable and accessible but also ensures that it can be used in diverse real-world scenarios such as education, healthcare, workplace communication, and public services.

2. RELATED WORK

A. In 2024, Zhang has developed the wearable gesture recognition glove that recognize gesture in real time with 95% accuracy. The overall system is combination of emg sensor and computer vision that demonstrating the feasibility and portability in gesture recognizing.

B. In 2024, Filipowska and the team build a device which uses neural network to recognize gestures. This design focuses on fixing sensors drift and difference between finger movement. This are the combination of on device and cloud processing to make it reliable in real time processing,

C. In 2023, Some organizations work on full translation of gestures, where glove captures gestures and directly converted into a text and speech in real time. This system highlights the importance of fast system, energy efficient and make balanced between glove processing and phone or cloud support system.

D. In 2023, researchers like Achenbach develop the "Give me a sign" dataset, which includes 64 American gestures language handshapes in which recorded with a glove. Their project has proven the hand gestures can be recognize with trustworthy accuracy, also provide valuable dataset for future development.

E. In 2022, Saeed and their team published a review of glove-based Techniques. In this they highlighted the newly trends like combining flex sensors and motion sensors, using neural networks for better accuracy and connecting glove with mobile and web applications. Their review paper given a clear direction for designing and implementation of more practical system.

F. In 2021, Deshmukh et al. Apply the supervised machine learning algorithms to improve gestures classification using large datasets. While accuracy improved high computational requirement which cause restriction in deployment on low power microcontrollers.

G. In 2020 Saggio and their team find out that CNN works best when was enough data by comparing traditional recognition method with deep learning models and simpler model like k-nn were still effective when resources are limited.

H. In 2019, Gupta and Mehta enhanced the techniques by combining flex sensor with an accelerometer to detect both finger and hand orientation. Their model translated gestures into recorded voice messages, the accuracy declined under hand position, reveals the need for better sensors

I. In 2018, Sharma presents the low-cost hand gesture recognition system by integrating flex sensor with an Arduino UNO this prototype translates specific gestures into a text. This technology term as assistive technology but fall behind advanced motion tracking and speech output features.

3. METHODOLOGY

This methodology is adopted in the research of designing, developing, and implementation of Sensor-Based Sign Language Recognition using a Smart Glove capable of translating gestures into audio or text translation. This system integrates sensor-based techniques with a Deep learning bidirectional LSTM model (hybrid model).

The entire workflow is divided into Several phases, as described as follows: 1) System design and hardware 2) Data Acquisition 3) Data Preprocessing 4) Model Development 5) Model Deployment 6) Real-Time Testing and Validation.

1. System Design and Hardware

This system architecture is designed around a glove-based system, which consists of a combination of a flex sensor, an Arduino nano, an MPU6050 (accelerometer gyroscope module), jumping wires, a speaker and a Glove. The Flex Sensor is mounted on all five figures which are attach to the MPU6050 and Arduino nano to record bending and hand orientation, as well as fingers movement. The flex sensors are corresponded to each finger. Their resistance changes when they are bended more with the movement of fingers and hand, which is read by Arduino Nano as varying analogy voltage levels. The MPU6050 provides real time hand movement and finger movement data, add an additional Layer of orientation tracking.

The Arduino nano is the central processing system for data acquisition. It interacts with the sensors and records their values through analogy pins and I2C pins, and transmits the readings to the computer for further processing. The entire setup is powered by USB or Battery pack. Ensures portability and ease of use.

The entire system is assembled on palm of glove and the Arduino nano is connected to mpu6050 and mpu6050 is connected to the 5-flex sensor which are mounted on figures to record readings of movement the entire system is power by USB or a battery pack, ensuring portability and ease of use. Ensures portability and ease of use.

2. Architecture

2.1 Convolution neural network (CNN)

In 1998, the first classical CNN architecture was designed by LeCun et al., which achieved impressive performance in the handwriting digit recognition task. Different from the traditional neural network structures, there is local connectivity and weight sharing in CNNs, which improves effective learning and reduces the overfitting problem of parameters.

There are three main layers of a CNN: convolution, pooling, and fully connected. The convolution layer, pooling layer and fully connected layer. The convolution layer takes small areas of the input data and preprocesses them one at a time using filters (known as kernels). These filters can find pattens or features, such as edges, colors. Then, a ReLU function removes all the negative values and retains only the useful positive values to make the result clean and easy to work with. The pooling layer reduces the di mentality of feature map by max or average pooling

operations, making simple computations and providing spatial invariance. The fully connected layer usually consists of a few hidden layers and a soft max function, which extracts features from the data and outputs the probability distribution of each class.

2.2 Bidirectional long short-term memory (BiL STM)

The Recurrent Neural Network (RNN) are employed for data processing tasks since they possess short-term memory capacity and can handle temporal dependencies. However, classical RNNs have difficulties in capturing long term dependencies, as noted by Greff et al. in 2015, the LSTM model is proposed to address this by maintain long term dependencies better than RNNs, taking advantage of memory cells with gates.

An LSTM contain three gates: forget gate, input gate and output gate. In which the forget gate decides which information throw away from the previous cell state, the input gates decide which information added to current cell state and output gate decides which information passed to the result. These gates contain a nonlinear activation function 'sigmoid' or 'tanh' let LSTM network to regulate information flow in a dynamic way thus being able to capture long range dependencies in sequential data.

However, it has a drawback that the standard LSTM handle information in one direction only, store everything they had until that time and cannot use future spatial context. To solve this problem the Gravel et al. introduce the Bidirectional LSTM (BiLSTM) in 2013. It consists of LSTM layer in opposite direction, one read the data forward (past to future) and one reads the data backward (future to past). It combines the result from both directions to model dependencies for both past and future context.

Because LSTM looks at data from both directions, it can understand the full context of information in a more effective way. which helps it to perform better in gesture prediction, emotional prediction, and language modelling. Therefore, BiLSTM is an enhanced version of LSTM, which captures more complex temporal features from input data sequence space and provides better prediction accuracy.

The architecture of the CNN + BiLSTM model combines the ability of CNN and sequential learning from BiLSTM networks. The CNN layer serves as a primary feature extractor. Local spatial patterns are captured from input images or time series data throughout convolution and pooling operations to reduce noise and dimensions, while keeping relevant features.

The resulting feature map created by the CNN are flattened and fed into the BiLSTM layer, which are capable of capturing temporal poisoning sequence and context relationship between features. The forward and backward layers learn how features change over a time and how they related to each other. It can use this ability to process dependencies in both directions improving its contextual understanding as well as prediction performance.

The fully connected dense layer collects all the features and supplies them to a softmax output layer for classification. This combination of spatial and time-series based learning is very useful for solving challenging pattern recognition problems including detection of defects, emotion analysis, and gesture predictions. This CNN + BiLSTM model can be seen as a new hybrid deep learning architecture that links between features extracted by static images and dynamic sequences, enabling the robust and efficient prediction.

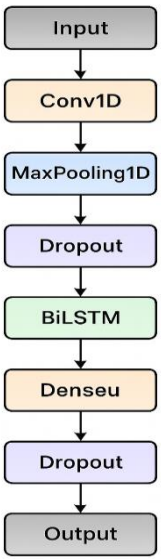

Fig. 2.1 CNN + BILSTM Model Flow

3. Dataset Description

In this research, we have developed the gesture dataset using sensor-based data acquisition to recognize and classify 50 different hand gestures. The data are gathered through a glove prototype based on Arduino Nano integrated with five flex sensors and an MPU6050 sensor (3-axis accelerometer and 3-axis gyroscope). A custom dataset was created to match unique movements of the hand.

The data collected from five participants where each performed all 50 gestures five times, providing total of 1,250 samples for the data collection. For every gesture, 5 seconds of data were collected at a sampling rate of 100 Hz resulting in generating both angular velocity and acceleration data for the x, y and z axes. These configurations ensure that each gesture includes unique space-temporal information to obtain high recognition accuracy.

In order to maintain data integrity, we applied a noise reduction process by using a Butterworth low-pass filter

(cutoff frequency: 20 Hz) and normalized the signals via Z-score standardization. The cleaned data were segmented for 2s duration with a 50% overlap with each window being labelled with the occurrence of its corresponding gesture. At last, we split the dataset into 70:15:15 ratio, 70% for training, 15% for validations and 15% for testing sections. This 50-gesture dataset is quite ideal for deep learning model (cnn+bilstm) to predict gestures efficiently and real-time from sensor data.

4. Implementation Flow

4.1 System initialization

When we start the system upon the power up of wearable glove, it first checks all the hardware components such as flex sensors and motion sensors to ensure they are working correctly. At the same time, the pre trained CNN+BiL STM model is loaded into memory.

4.2 Sensor Data Acquisition

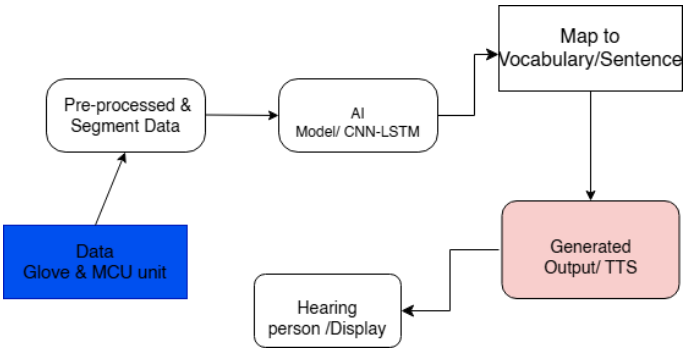

Fig.4.1 Flow of System

In this part, when glove is in use, its continuously collects the real time sensor data of figures movements and hand movements from the embedded sensors. The spatial orientation is tracked by the accelerometer and gyroscope. These are then record the real time and the user's gestures can be captured. Which in return assist the Model to recognize the Gestures. This stream of data is then passed on for processing. This data stream is then forwarded to the next process.

4.3 Preprocessing and feature extraction

In this part, the raw data is refined and structured to enhance processing. The noise is eliminated and signals are smoothed and normalized. The resulting data is divided into small time windows representing a single gesture each. From these chunks, the relevant features are extracted, such as finger bending, hand movements, and orientation changes. This ensures that the input data is given to recognize the model is well organized and consists of essential signals needed for accurate gesture prediction.

4.4 Gesture recognizing (CNN + BiLSTM model)

In this stage, the pre-processed features are input to a hybrid deep leaning model with CNN and BiLSTM. The convolution layer captures the spatial dependencies and local patterns from input sequences, whereas the BILSTM layers can learn the temporal relationships in forward and backwards. This combination is beneficials to gesture recognition improvement of accuracy and robustness of the model.

4.5 Textual Conversation

Once the gesture is recognized, it is translated into text, converting the hand movement into words that can be easily understood. For example, if the glove identifies the gesture, it immediately maps the output to the data flow in a smooth way to the speech synthesis.

4.6 Speech Synthesis

It then uses, text-to-speech (TTS) engine to read the response. It was this process that enables the system that "speak" the gesture for those who have difficulty communicating with people who do not understand sign language.

4.7 Output delivery

The resulting output speech is then played through a speaker, audio device. A corresponding text can also be shown on screen or mobile or web application directly by the user for convenient interaction and accessibility.

5. Limitation

Action and interaction are certain and more precise than context in themselves. However, the dataset was created are biased. Meaning that, the result cannot be applied to all sign language. Since language does not have any fixed path or linear structure. the researchers force them into one which does not match how spoken language works. Since culture and ethics are deeply involved in the sign language. it is important that the deaf people actively involved and creating developing systems.

6. Challenges and Future Scope

The gesture prediction system is effective, but some issues can affect its performance and scalability of the system. The variability of data requires additional research. The data related to the size of the user's hand, the speed of hand gestures and the tightness of the glove leads to instability in sensor readings, which influence the reliability and predictability of model. The dataset is limited sized i.e. 50 gestures and 1250 samples does not allow the model to make generalizing the proposed model to make reasonable

predictions for gestures not seen in the dataset. Additionally, the absence of a GUI should be noted; no web-based user control tools can be used because the device is unable to handle such dataset.

There is lot of more noise in the sensor data due factors such as sensors movements in different positions, during room setup or signal drift. The noise could not be removed for idealistic analysis, but it is up to the future work to design advanced data preprocessing methods. Despite the ability to recognize gestures in real-time, delays are introduced from continuous data retrieval and computation, and the time to move the sensor to a different label after the code has already been trained.

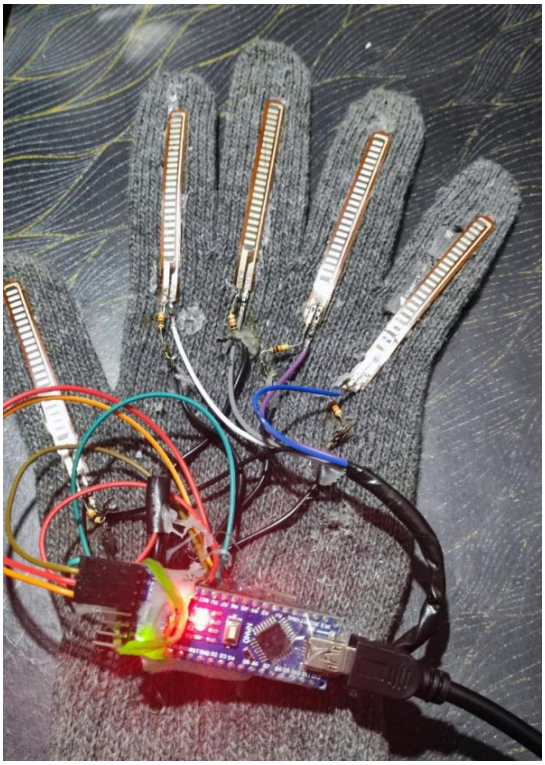

Fig. 6.1. AI gesture Interpreter Glove

REFERENCES

[1] B. Sarada, Ch. Sri Divya, S. Begum, P. Kousalya, and E. Simhadri, "Smart Glove for Gesture Recognition using IoT," International Journal of Scientific Research in Engineering and Management (IJSREM), Apr. 2025.

[2] F. Berjin Singh et al., "Gesture Interpretation Glove for Assistive Communication," Asian Journal of Applied Science and Technology (AJAST), Apr.–Jun. 2025.

[3] R. Damdoo et al., "An Integrative Survey on Indian Sign Language Recognition and Translation (ISLRT)," Springer Open AI Applications Journal, 2025.

[4] Sign Speak Research, "Comparing Sign Language AI and Human Interpreters: Evaluating Efficiency for Greater Social Sustainability," Sustainability in Technology Journal, 2025.

[5] B. Batte et al., "AI-Powered Sign Language Translation System: A Deep Learning Approach to Enhancing Inclusive Communication and Accessibility in Low-Resource Contexts," IEEE Access, 2025.

[6] M. Kumar and N. Gupta, "Real-Time Hand Gesture Recognition using Deep Learning," International Journal of Computer Science Trends and Technology (IJCST), vol. 13, no. 1, 2024.

[7] Chakraborty et al., "Deep Neural Network-based Indian Sign Language Recognition," IEEE Access, vol. 12, pp. 11834–11845, 2024.

[8] P. R. Bhatia et al., "Hybrid CNN-BiLSTM Approach for Gesture Recognition," IJITEE, vol. 13, no. 6, 2024.

[9] Y. Zhang et al., "Artificial Intelligence in Sign Language Recognition," MDPI Sensors, vol. 24, 2024.

[10] Tech Science, "Recent Advances on Deep Learning for Sign Language Recognition," TechScience Press Journal, vol. 12, no. 8, 2024

[11] N. Jain and R. Sharma, "Sensor-based Sign Recognition for Indian Languages," IJRTE, vol. 12, no. 2, 2023.

[12] Wu J, Ren P, Song B, Zhang R, Zhao C, Zhang X (2023) Data glove-based gesture recognition using CNN-BiLSTM mode with attention mechanism.

[13] P. K. Rao et al., "Low-Cost Smart Glove for Sign Language Interpretation," Procedia Computer Science, vol. 218, 2023.

[14] Patel et al., "Wearable Smart Glove for Communication Assistance," Journal of Intelligent Systems and Computing, vol. 35, pp. 255–268, 2023.

[15] H. Alzarooni et al., "AI-Based Sign Language Interpreter (GestPret)," International Journal of Artificial Intelligence and Applications, 2023.

[16] R. Banerjee and A. Ghosh, "Sensor-Fusion-based Indian Sign Language Detection using Machine Learning," JETIR, vol. 10, 2022.

[17] Thakur and D. Sharma, "A CNN-LSTM Hybrid Model for Gesture Recognition," IEEE International Conference on IoT and AI Systems, pp. 213–220, 2022.

[18] M. Joshi et al., "Hand Motion Recognition using Accelerometer and Gyroscope Sensors," International Conference on Computing and Communication Systems (ICCCS), 2021.

[19] T. G. Kumar and S. K. Pillai, "Real-time Gesture Recognition using Smart Glove and ML Models," IJRTE, vol. 9, no. 5, 2021.

[20] R. Rastgoo et al., "Sign Language Recognition: A Deep Survey," Expert Systems with Applications, 2021.

[21] P. Agarwal and S. Kumar, "Smart Wearable for Indian Sign Language Detection using IoT and AI," International Conference on Smart Computing and Communication, 2020.

[22] N. K. Bansal et al., "Feature Extraction and Classification of Sign Gestures using CNN," IEEE INDICON Conference, 2020.

[23] Baumgärtner et al., "Automated Sign Language Translation: The Role of Artificial Intelligence in Sign Language Translation," Frontiers in Artificial Intelligence, 2020.

[24] S. Chatterjee, "Motion Sensor-based Hand Gesture Recognition," IJERT, vol. 8, no. 7, 2019.

[25] Raj and R. K. Singh, "Wearable Glove for Sign-to-Speech Conversion," IJCSIT, 2019.

[26] R. S. Patil and P. Deshmukh, "Accelerometer-based Indian Sign Language Recognition," IRJET, 2019.

[27] R. S. Patil and P. Deshmukh, "Accelerometer-based Indian Sign Language Recognition," IRJET, 2019.

[28] Sharma and N. Saini, "Arduino-based Smart Glove for Speech-Impaired Communication," IJIRSET, 2018

[29] P. Gupta et al., "Sign Language Recognition using Flex Sensors and Microcontrollers," ICCES Conference Proceedings, 2018.

[30] S. N. Verma, "A Low-Cost Prototype for Sign Gesture Identification," IJCSMC, 2018.