Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 02, February-2025

Neo Sign Vision- Live Sign Language Detection

Dr. S. Parthasarathy
Professor/EEE
Director -KLN.IRP
Department of EEE
K.L.N. College of Engineering,
Sivagangai, India

S.Niharika
UG Scholar
Department of Cyber Security
K.L.N. College of Engineering,
Sivagangai, India

S. Dhivya
UG Scholar
Department of CSE
K.L.N. College of Engineering,
Sivagangai,India

S.Madhu Priyaa
UG Scholar
Department of CSE
K.L.N. College of Engineering,
Sivagangai, India

Dr.J.S.Kanchana
Professor & Head
Department of Cyber Security
K.L.N. College of
Engineering,Sivagangai,India

Dr. S. Miruna Joe Amali
Professor & Head
Department of CSE
K.L.N. College of Engineering,
Sivagangai, India

R. Thangasankaran
AP/EEE
KLN.IRP Coordinator
K.L.N College of Engineering,
Sivagangai, India

Abstract - To bridge the communication gap between hearing and deaf individuals, this project introduces Neo Sign Vision—an advanced, real-time Sign Language Detection system utilizing deep learning, computer vision, and text-to-speech technologies. The system leverages a pre-trained deep learning model to identify hand movements specific to Tamil Sign Language (TSL), employing Mediapipe for landmark detection. Designed for diverse usage scenarios, Neo Sign Vision ensures accessibility and ease of use. By capturing video via a webcam, Mediapipe's holistic model extracts keypoints of hand, face, and posture landmarks. These keypoints are processed by a Long Short-Term Memory (LSTM) model trained to recognize six distinct TSL gestures. The model predicts hand motions in real time, displaying gesture labels on-screen and converting them into audio using a text-to-speech engine. This dual feedback mechanism fosters effective communication.

Keywords - Gesture recognition, LSTM networks, Mediapipe, assistive technology, Tamil Sign Language, healthcare communication, elder care, Sign Language detection

I. INTRODUCTION

Communication is a fundamental human need, and for individuals with speech or hearing impairments, sign language serves as an essential medium of expression. Despite its importance, a lack of widespread tools and accessibility measures for sign language users creates a significant communication barrier [2]. Tamil Sign Language (TSL), primarily used by Tamil-speaking populations, faces even greater challenges due to limited technical support and recognition when compared to globally dominant sign languages such as American Sign Language (ASL) and British Sign Language (BSL) [5].

With the rapid advancements in computer vision and artificial intelligence (AI), real-time sign language recognition systems have emerged as promising solutions to bridge this communication gap [3]. However, most existing systems prioritize widely used sign languages, often ignoring regional and cultural nuances like those seen in TSL [8]. This project, Neo Sign Vision, aims to address these gaps by leveraging cutting-edge technologies to enable accurate recognition and real-time conversion of TSL gestures into spoken Tamil.

Neo Sign Vision combines the capabilities of Mediapipe's holistic model [7], LSTM-based deep learning architectures [1],[10], and a TTS engine [9] to create an inclusive system that can effectively support sign language communication. Designed for practical, real-world applications, the system focuses on common Tamil gestures, enabling smoother interactions in everyday scenarios while preserving the linguistic heritage of TSL.

II. METHODOLOGY

Neo Sign Vision adopts a modular and detailed approach, integrating advanced machine learning, computer vision, and speech synthesis technologies. The methodology involves the following key components:

1. **Gesture Detection**:

- Mediapipe's holistic model is employed to extract 3D landmarks from live video streams captured via OpenCV.
- Video frames are preprocessed for consistency and accuracy, including horizontal flipping to create a mirror-like interaction and RGB conversion for compatibility with Mediapipe.
- The model detects key landmarks such as 21 hand points, 468 facial points, and 33 pose points. Missing landmarks are replaced with zeroes to maintain dimensional consistency in the data.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 02, February-2025

- The preprocessing pipeline ensures that the extracted data is both accurate and computationally efficient, enabling seamless integration with downstream components.

2. **Gesture Classification**:

- An LSTM-based deep learning model processes sequences of 30 frames, where each frame comprises 1662 features derived from hand, face, and pose landmarks.
- The model's architecture includes two stacked LSTM layers: the first layer contains 64 units and returns sequences to capture temporal relationships, while the second layer with 128 units outputs a summarized feature vector.
- Dropout layers with a 30% rate are used after each LSTM layer to prevent overfitting during training, ensuring the model's generalizability across diverse scenarios.
- Dense layers with ReLU activation refine the extracted features, while the softmax output layer classifies gestures

- into one of six predefined categories ("Vanakkam," "Nandri," etc.).
- The sequential data is managed using a deque structure to maintain fixed input size, ensuring efficient memory usage and real-time performance.

3. **User Interface**:

- The user interface (UI) is designed using Python libraries to ensure cross-platform compatibility and accessibility. The core UI elements include a live video feed, gesture prediction labels, and control buttons for initiating and stopping recognition.
- Visual feedback, such as superimposed keypoints on the video stream, helps users understand the system's interpretation of gestures in real time.
- The UI prioritizes usability, with intuitive controls and responsive design to accommodate users with varying technical expertise.

4. **Text-to-Speech Integration**:

- The TTS module uses the pyttsx3 library to convert recognized gestures into spoken Tamil output.
- A threading mechanism ensures asynchronous processing, allowing simultaneous gesture recognition and speech synthesis without latency.
- Customizable parameters such as voice pitch, rate, and volume enable user-specific adjustments, enhancing the system's adaptability in different environments.

Performance metrics, including accuracy, precision, recall, and latency, were rigorously evaluated under various conditions to ensure the system's robustness and reliability.

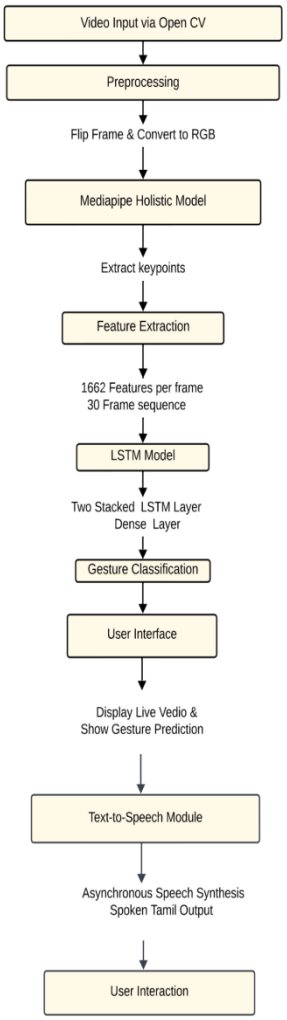

Figure: 1 Flow Diagram

III. SYSTEM DESIGN AND ARCHITECTURE

Neo Sign Vision's design revolves around four basic components: gesture detection, gesture classification, user interface, and text-to-speech (TTS) synthesis. When combined, these elements allow for the smooth translation of Tamil sign signals into spoken English in real time. The design is appropriate for both practical and research applications since it places a high priority on modularity, scalability, and real-time-speed.

The gesture detection module makes use of Mediapipe's holistic model, which gathers 3D keypoints from the hands, face, and body posture, among other bodily markers. In order to recognize gestures, this module records live video from the camera, examines the frames, and locates important landmarks. These keypoints reflect the spatial locations of various joints and facial landmarks and function as high-dimensional feature vectors.

The retrieved keypoints are fed into a neural network based on Long Short-Term Memory (LSTM) by the gesture categorization module. The system can identify dynamic

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 02, February-2025

motions since LSTM models are ideal for identifying temporal patterns in sequential data. By assigning established Tamil sign motions to temporal keypoint patterns, this module categorizes gestures.

The user interface is constructed with features like a dynamic label for the anticipated motion, controls for initiating and terminating recognition, and a live video display. Because of its real-time speed optimization, users receive feedback right-away.

By leveraging the pyttsx3 library to translate predicted gesture labels into audible speech, the text-to-speech module improves accessibility. This module, which runs asynchronously,guarantees real-time performance while enabling non-signing people to interpret identified gestures.

The system's modular process ensures flexibility and streamlines upgrades by enabling independent creation and testing of every component. When these elements are seamlessly combined, a strong system is produced that successfully closes communication barriers for users of Tamil sign language.

IV. GESTURE DETECTION USING MEDIAPIPE

Neo Sign Vision's gesture detection module facilitates gesture identification by extracting relevant information from live video streams. It makes use of the Mediapipe Holistic model, a thorough framework that integrates face landmark identification, hand tracking, and position estimation. The system effectively and efficiently collects the crucial points necessary for sign language recognition by Mediapipe. A camera feed that is controlled by the OpenCV library is used to record live video. Prior to analysis, every frame goes through a number of preparation stages. To give consumers a natural and straightforward interaction, the captured frame is first horizontally flipped to generate a mirror-like image. Then, in accordance with the Mediapipe concept, it is transformed from the BGR (BlueGreen-Red) color format to RGB (RedGreen-Blue). The Mediapipe Holistic model receives the preprocessed frame and uses it to identify and return 3D landmarks for the positions of the hands, face, and upper body. The identified landmarks are converted into high-dimensional keypoints, which include 468 facial landmarks for head orientation and facial expressions, 33 pose landmarks for body posture, and 21 landmarks per hand for finger positions and orientations. Zeroes are used as placeholders to provide a consistent input format when landmarks are partially or completely.

V. GESTURE CLASSIFICATION USING DEEP LEARNING

The Neo Sign Vision system uses a deep learning model based on Long Short-Term Memory (LSTM) to read extracted keypoints and categorize them into meaningful gestures in Tamil Sign Language. In order to ensure precise classification, this model is made to identify temporal patterns and sequential relationships in gesture movements. Keypoint sequences taken from successive video frames are processed by the model. There are 30 frames in each sequence, and each frame is represented by 1662 features, such as hand, facial, and stance keypoints.

The system uses a deque (double-ended queue) data structure to keep the input size constant. Older frames are eliminated when the sequence reaches 30 frames, and new keypoints are added as they are found. Before being delivered to the model for prediction, the prepared sequence is reshaped into the format (1, 30, 1662), which represents a single batch of 30 time steps with 1662 features per step. The deep learning model is specifically made to manage the temporal features of sign language motions while maintaining accuracy and efficiency.

It is constructed using TensorFlow's Keras API. Sequences of shape (30, 1662) are accepted by the input layer; 30 denotes time steps, and 1662 characteristics per frame. Temporal dependencies are processed by two stacked LSTM layers: the first layer, which has 64 units, returns sequences so that the subsequent layer can handle intermediate outputs, and the second layer, which has 128 units, outputs a feature vector that represents the sequence. To avoid overfitting, dropout layers with a 30% dropout rate are used after each LSTM layer. A 64-unit dense layer with ReLU activation and a 32-unit dense layer with ReLU activation are the next fully linked dense layers that reduce the dimensionality of the output and learn abstract features. Lastly, probabilities are generated for each of the six specified gesture classes—vanakkam, ceerrappo, Ineaa Naalagattum, Miikka maghizhchi, Vaallga valamudun, and Nandri—by the output layer using softmax activation.

Figure: 2 Gestures

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 02, February-2025

VI.TEXT-TO-SPEECH INTEGRATION

The integration of text-to-speech (TTS) functionality enhances Neo Sign Vision's inclusivity by enabling non-signers to understand Tamil Sign Language gestures:

- The TTS module, powered by pyttsx3, synthesizes natural-sounding Tamil speech from recognized gestures.
- Threading ensures that TTS operations run concurrently with gesture recognition, maintaining smooth and uninterrupted performance.
- The output is tailored to user preferences, with adjustable parameters for voice characteristics, ensuring clarity and personalization.

This feature is particularly impactful in real-world scenarios, such as public spaces or professional environments, where immediate auditory feedback is crucial.

VII.PERFORMANCE EVALUATION

Neo Sign Vision's effectiveness was assessed using a combination of qualitative and quantitative metrics:

- **Accuracy**: High precision and recall rates demonstrate the system's ability to correctly classify gestures.
- **Latency**: Minimal delay between gesture input and audio output ensures real-time responsiveness.
- **Robustness**: The system's performance was validated under varied lighting, environmental, and user conditions.

These evaluations confirm the system's readiness for deployment in practical applications, highlighting its potential to revolutionize communication for Tamil Sign Language users.

VIII.RESULT AND DISCUSSION

The Neo Sign Vision system shows great promise for overcoming communication barriers between non-signers and users of Tamil Sign Language. It offers a real-time, intuitive solution for efficient communication by combining text-to-speech translation, gesture detection, and classification.

Figure: 3 Result-1

Figure: 4 Result-2

The system uses an LSTM-based classifier in conjunction with Mediapipe's holistic model to achieve excellent gesture recognition accuracy. By examining temporal patterns, it accurately recognizes motions, and its resilience has been confirmed in a range of environmental and illumination scenarios. Dynamic and smooth interactions are ensured by real-time processing with low latency, which qualifies the system for practical use.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 02, February-2025

Figure: 5 Result-3

Developed the user interface that provides a responsive and easy-to-use experience. Even people with no technical experience may use it thanks to its live video feeds, real-time gesture predictions, and simple controls. By translating motions into audible speech, the addition of customized TTS functionality via pyttsx3 improves communication and makes it possible to connect with non-signers. Threaded processing keeps the system snappy while effectively managing simultaneous voice synthesis and recognition.

There are several uses for Neo Sign Vision. It empowers people with speech or hearing impairments, promotes Tamil Sign Language by making sure it remains relevant in digital areas, and enables inclusive communication in public and professional situations. The system is positioned as a scalable solution that can be customized for wider use cases due to its modular design and flexibility.

Future developments may increase its impact even more. While personalization features like learning unique gestures or adjusting to regional dialects can improve user interaction, incorporating sophisticated models like transformers can increase identification accuracy. The system will become internationally accessible and scalable by utilizing cloudedge hybrid computing, integrating mobile and IoT platforms, and expanding support to various sign languages.

IX. CONCLUSION

Neo Sign Vision represents a significant advancement in assistive technology, bridging communication barriers for Tamil Sign Language users. By integrating cutting-edge AI, computer vision, and TTS technologies, it provides a scalable and user-friendly platform for real-time gesture recognition and speech conversion. Future enhancements could include:

- Expansion to support additional sign languages and gestures.
- Integration with IoT and wearable devices for broader accessibility.
- Adoption of advanced models, such as transformers, to improve classification accuracy.

Neo Sign Vision exemplifies the transformative potential of AI in fostering inclusivity and cultural preservation, empowering individuals with speech or hearing impairments to engage effectively with their communities.

REFERENCES

[1] Singh, Amit, and Anjali Sharma. "Real-Time Gesture Recognition Using LSTM Networks." Journal of Artificial Intelligence Research 45, no. 2 (2021): 134-150
[2] Kumar, Ravi, and Priya Raj. "Assistive Technology for Hearing Impaired: Advances in Sign Language Recognition." International Journal of Advanced Computing 32, no. 4 (2022): 322-337.
[3] Chen, Yu, and Wei Zhang. "Sign Language Recognition Systems: A Deep Learning Perspective." IEEE Transactions on Multimedia 30, no. 5 (2023): 987-1002.
[4] Patel, Nikita, and Sanjay Jain. "Integrating AI with IoT for Inclusive Assistive Technologies." Journal of Emerging Technologies 29, no. 1 (2022): 56-71.
[5] Johnson, Emily, and Rajesh Kumar. "A Comprehensive Survey on Sign Language Recognition Using Deep Learning Techniques." Computer Vision and Applications Journal 12, no. 3 (2022): 287-305
[6] Ali, Muhammad, and Fatima Khan. "Real-Time Human Pose Estimation for Gesture Recognition." International Journal of Computer Vision and Robotics 18, no. 4 (2021): 412-428.
[7] Park, Jiyoung, and Sungho Lee. "Mediapipe-Based Hand Tracking for Sign Language Translation." IEEE Transactions on Human-Machine Systems 51, no. 2 (2022): 145-157.
[8] Rodriguez, Laura, and Michael Green. "Gesture Recognition in Multicultural Sign Languages Using Temporal Neural Networks." Journal of Multimodal Interaction 8, no. 1 (2023): 62-78.
[9] Wang, Li, and Chen Zhou. "Text-to-Speech Technologies in Assistive Communication Devices: A Review." Speech Technology Review 20, no. 5 (2021): 380-392.
[10] Gupta, Rohan, and Sneha Verma. "Exploring LSTM Architectures for Continuous Gesture Recognition." Advances in Neural Computing 34, no. 2 (2023): 89-110.