RINA - RESPONSIVE INTERPRETER FOR NON-VERBAL ASSISTANCE (BASED ON INDIAN SIGN LANGUAGE)

Project Reference No.: *49S_BE_6546*

College : H.K.E. Society's Sir. M. Visvesvaraya College of Engineering, Raichur
Branch : Department Of Computer Science and Engineering
Guide : Dr. R K Ashwin
Students : Ms. Shreelatha G
Ms. Snehalatha
Ms. Spoorthi
Ms. V Kavya

Keywords:

Indian Sign Language (ISL), MediaPipe Holistic, LSTM (Long Short-Term Memory), Real-time Gesture Recognition, Assistive Technology, AWS, Human-Computer Interaction

Introduction:

Communication is a fundamental human right, yet the deaf and hard-of-hearing community in India faces a persistent digital divide. While technology has advanced, accessible solutions for Indian Sign Language (ISL) remain scarce and often limited to static content. Early research in this field focused primarily on static alphabets or digits using traditional image processing, which failed to capture the dynamic, temporal nature of real-world signing. Modern deep learning approaches like CNNs and LSTMs have shown promise, but their effectiveness is often hindered by a lack of standardized, large-scale ISL datasets. This project introduces RINA, a system designed to bridge this gap. Originally, the vision for this project was to implement a real-time Generative AI model. However, due to the extreme scarcity of comprehensive, high-quality ISL datasets required for training generative architectures, the project was strategically pivoted. RINA now functions as a robust, landmark-driven recognition system that uses a custom-built dataset to provide immediate, context-aware assistance through high-quality pre-recorded ISL video responses, serving as a scalable foundation for future generative development.

Objectives:

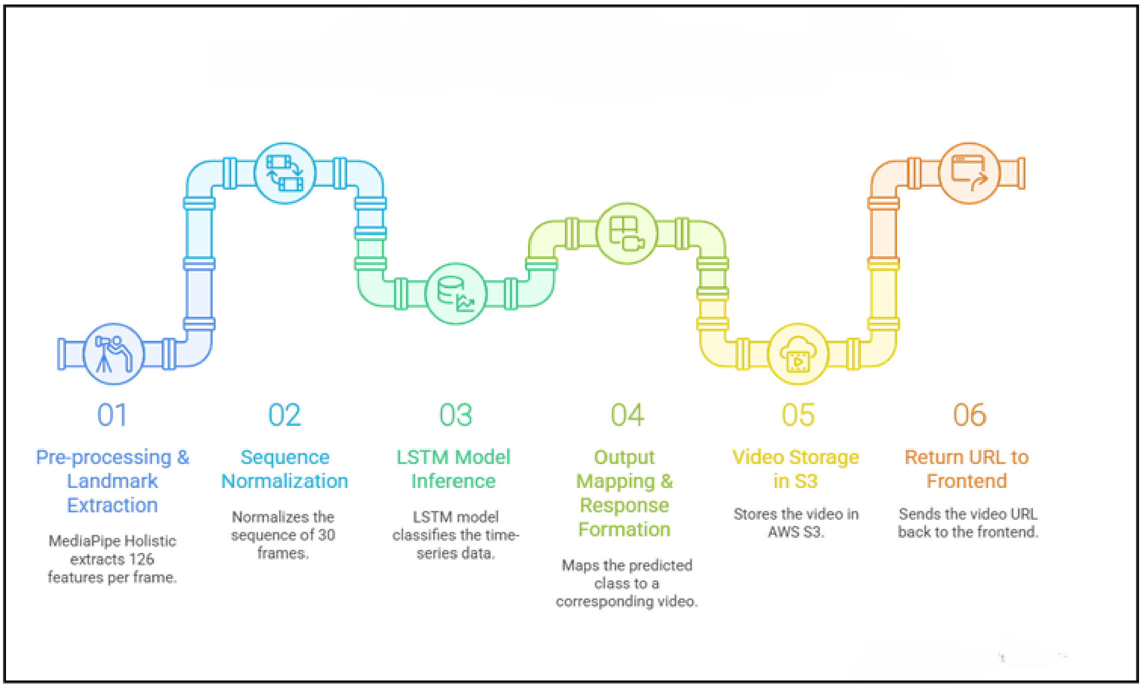
1. To design and develop a comprehensive Bharath Sign Language Recognition and Learning System for accessibility and communication.
2. To enable real-time Indian Sign Language (ISL) recognition using webcam-based input and deep learning models.
3. To facilitate two-way communication between signers and non-signers in educational environments.

4. To provide sign-based visual responses by mapping recognized gestures to ISL videos (RINA system).
5. To convert audio and text inputs into ISL visual outputs for better inclusivity.
6. To develop interactive ISL learning modules with alphabets, numbers, sentences, and quizzes.
7. To create a custom ISL dataset due to the lack of standardized datasets.
8. To preprocess and organize high-quality gesture data for accurate model training.
9. To deploy the system on scalable platforms like AWS for accessibility and reliability.
10. To ensure future scalability by allowing easy addition of new gestures and learning features.

Methodology:

The RINA system follows a modular, six-stage pipeline:

1. **Data Collection & Customization:** Due to the lack of existing datasets, we recorded approximately 500-720 videos for 10 distinct ISL question classes (e.g., "What is AI?", "What is Blockchain?").
2. **Landmark Extraction:** We utilized **MediaPipe Holistic** to track 126 keypoints (hands, pose, and face) per frame, converting visual movement into numerical feature vectors.
3. **Sequence Normalization:** Each gesture is standardized into a fixed sequence of 30 frames to provide consistent temporal input for the model.
4. **LSTM Model Training:** A **Long Short-Term Memory (LSTM)** network was chosen to learn the spatiotemporal patterns of the signs. Unlike standard models, LSTMs effectively solve the vanishing gradient problem, allowing the system to "remember" the hand's trajectory over time.
5. **Real-time Inference Pipeline:** The system captures live webcam frames, extracts landmarks on the fly (averaging 0.12s per frame), and predicts the gesture with a high confidence threshold.
6. **Responsive Output Generation:** Once a sign is recognized, the system is designed to automatically fetch and play a corresponding high-quality ISL answer video. While the architecture is structured for cloud-based retrieval, current implementation successfully demonstrates this playback locally using a synchronized multimedia directory.
7. **System Deployment Strategy:** The project is developed with a cloud-ready architecture intended for **AWS (EC2 and S3)** to ensure a public, scalable infrastructure. Initial deployment phases were initiated to host the LSTM model on EC2 and store assets in S3; however, the system currently operates in a robust local environment as we refine the network configurations and API integrations required for full cloud transition.



Result and Conclusion:

1. The RINA system achieved an overall accuracy of 91.11% on the test dataset.
2. It recorded high precision (91.85%) and F1-score (91.27%), ensuring reliable classification.
3. Certain gestures like “What is AI?” and “What is Solar System” achieved perfect prediction scores.
4. The confusion matrix showed strong ability to distinguish similar dynamic gestures.
5. MediaPipe Holistic provided accurate landmark extraction for gesture representation.
6. The LSTM model effectively captured temporal patterns in ISL gestures.
7. The system delivered real-time predictions with low latency and stable performance.

In Conclusion,

1. The project proves that computer vision and deep learning are effective for ISL recognition.
2. The RINA visual-response approach improves understanding through sign-based video outputs.
3. The system works efficiently even without large-scale ISL datasets.
4. It reduces dependency on human interpreters and related limitations.
5. The solution is cost-effective, scalable, and accessible via AWS deployment.
6. It enhances communication between signers and non-signers, promoting inclusivity.

7. The system provides a strong foundation for future expansion and improvements.

Project Outcome & Industry Relevance

1. Achieves real-time ISL recognition with ~91% accuracy and low latency (0.12s/frame).
2. Introduces a **visual-to-visual communication system** using ISL video responses instead of text/speech.
3. Enhances **Human-Computer Interaction** by enabling natural and inclusive communication.
4. Reduces dependency on interpreters, improving accessibility for the deaf community.
5. Applicable in **education and healthcare** for inclusive communication and learning.
6. Uses **MediaPipe Holistic + LSTM**, making it cost-effective and deployable on standard devices.
7. Supports **cloud scalability via AWS**, enabling wider real-world deployment.
8. Contributes a custom ISL dataset and provides a foundation for future AI-based sign language systems.

Working Model vs. Simulation/Study

1. **Software-Based Implementation**

   The project is implemented as a real-time software system that processes Indian Sign Language gestures using live webcam input. It has been developed, executed, and tested through practical experimentation. The system demonstrates consistent performance under real-time conditions. This confirms its applicability in real-world environments.

2. **Real-Time Input Processing**

   The system continuously captures video input through a webcam and processes gestures as they are performed. It analyzes each frame dynamically and generates outputs without delay. This ensures smooth interaction between the user and the system. Such real-time responsiveness is essential for practical communication applications.

3. **Integration of Computer Vision and Deep Learning**

   The project integrates MediaPipe Holistic for extracting keypoint landmarks and an LSTM model for recognizing gesture patterns. These components operate together in a continuous pipeline to ensure accurate predictions. The model effectively captures both spatial and temporal features of gestures. This integration reflects a complete and applied system design.

4. **End-to-End Functional Pipeline**

   The system includes all stages such as input capture, preprocessing, feature extraction, model prediction, and output generation. Each stage is executed seamlessly in real time. Recognized gestures are mapped directly to corresponding ISL video responses. This demonstrates the completeness of the implementation.

5. **Automated Output Response**

   Once a gesture is recognized, the system automatically plays the relevant ISL video response. This enables clear and meaningful communication for users. The response generation is immediate and does not require manual intervention. It highlights the automation capability of the system.

6. **Local Deployment and Testing**

   The system is deployed in a local environment and tested across multiple gesture inputs. Performance metrics such as accuracy and response time have been evaluated. The results indicate stable and reliable system behavior. This confirms that the implementation is practically validated.

7. **Cloud Deployment Capability**

   The architecture is designed to support deployment on cloud platforms such as AWS using EC2 and S3 services. Initial steps toward deployment have been explored for scalability. This allows the system to be extended for broader accessibility. It reflects readiness for real-world deployment scenarios.

8. **Practical System Execution**

   The project operates as a real-time interactive system with continuous user input and output generation. It demonstrates actual execution rather than conceptual analysis. The system handles live gesture recognition and response delivery effectively. This establishes its relevance for real-world applications.

Project Outcomes and Learnings

1. **Development of Real-Time ISL Recognition System**

   The project resulted in a system capable of recognizing Indian Sign Language gestures in real time. It achieved high accuracy (~91%) with low latency. The system performs consistently across multiple gesture inputs. This validates its effectiveness for practical communication use.

2. **Application of Deep Learning (LSTM) for Temporal Data**

   The use of LSTM helped in understanding how sequential gesture data is processed. It enabled capturing temporal patterns in hand movements effectively. The project provided hands-on experience with sequence modeling. This strengthened knowledge in deep learning techniques.

3. **Efficient Landmark-Based Feature Extraction**

   MediaPipe Holistic was used to extract keypoint landmarks for hands, face, and pose. This approach simplified gesture representation into numerical data. It improved processing speed and accuracy. The learning emphasized the importance of feature engineering.

4. **End-to-End System Design and Integration**

   A complete pipeline was built from data collection to real-time prediction and output generation. All modules were integrated to function seamlessly together. This improved understanding of system architecture and modular design. It also enhanced implementation skills.

5. **Dataset Creation and Preprocessing Experience**

   A custom dataset was created due to the lack of standard ISL datasets. This involved recording, labeling, and normalizing gesture sequences. The process highlighted the importance of data quality. It also improved data handling and preprocessing skills.

6. **Exposure to Deployment and Scalability Concepts**

   The system was designed with cloud deployment capability using AWS services. Initial deployment steps were explored for scalability. This provided insight into real-world application deployment. It enhanced understanding of making systems accessible and scalable.

References:

1. K. A. et al., “Indian Sign Language Recognition Using Deep Learning,” 2019.https://www.semanticscholar.org/paper/deebe9ccb4aa24a2b5fabc631568b9e5138a4ccb
2. M. Kothadiya and H. Bhatt, “Deepsign: Sign Language Detection and Recognition,” 2020.https://www.semanticscholar.org/paper/701661911d8d8573512da67e969854a04778c714

3. S. Jain, R. Khanna, and P. Arora, “Deep Learning Based Indian Sign Language Recognition System,” 2019.https://ieeexplore.ieee.org/document/8951649
4. P. Kannan, A. Sharma, and S. Reddy, “Real-Time Indian Sign Language Recognition Using Depth Sensor,” 2016.https://ieeexplore.ieee.org/document/7725348

Future Scope:

The future scope of this project includes:

1. **Expansion to a Full Vocabulary ISL System**: The system can be extended from 10 gestures to hundreds of ISL signs by collecting a larger and more diverse dataset. This would enable broader communication and support day-to-day conversational needs.
2. **Continuous Sign Language Recognition**: Instead of recognizing isolated signs, future models can interpret continuous sentence-level signing. Advanced deep learning architectures like Transformers or Bi-LSTMs can help decode entire sign sequences in real time.
3. **RINA as a Generic AI Assistant**: The current system uses fixed answers for each gesture, but future versions can integrate an LLM based backend so that RINA becomes a generic question-answering AI, capable of responding dynamically to any ISL query—similar to a sign-language powered ChatGPT.
4. **Live ISL-to-Speech and Speech-to-ISL Translation**: Real-time translation can be implemented where the system: o Converts ISL gestures into spoken speech, and o Converts spoken language into animated sign outputs, enabling two-way live communication.
5. **Multilingual Output Support**: Output responses can be expanded to multiple Indian languages such as Hindi, Telugu, Tamil, and Kannada, enabling broader use across different regions.
6. **Full Interactive Learning Mode**: The system can be enhanced to serve as a learning platform, teaching ISL through: o Multiple example sentences o Real-time feedback on user hand positions o Practice modules o Progress tracking and adaptive difficulty levels
7. **User-Adaptive Personalized Models**: The system can learn a user’s unique signing style over time and adapt to variations in speed, orientation, or hand shape automatically.
8. **Mobile Application Deployment**: A TensorFlow Lite mobile app can make the system portable, allowing real-time sign recognition on Android and iOS devices without requiring a high-end system.
9. **Scalable Cloud Architecture**: Enhancing AWS deployment with API endpoints, load balancing, and auto-scaling would allow thousands of users to access the system simultaneously in a production environment.