Indian Sign Language Detection for the Deaf Community using Machine Learning

1st Anu Malhan
Department of Artificial Intelligence and Data Science
New Horizon Institute of Technology & Management
Thane, India

2nd Deep Manera
Department of Artificial Intelligence and Data Science
New Horizon Institute of Technology & Management
Thane, India

3rd Nisha Patil
Department of Artificial Intelligence and Data Science
New Horizon Institute of Technology & Management
Thane, India

4th Saumya Darekar
Department of Artificial Intelligence and Data Science
New Horizon Institute of Technology & Management
Thane, India

Abstract-The communication gap between the hearing and deaf communities poses significant challenges in daily interactions. Indian Sign Language (ISL) serves as the primary mode of communication for millions of individuals in India, yet there remains a lack of technological solutions for real-time interpretation. This study investigates a new method of computer pattern recognition aimed at the real-time detection and translation of ISL video recordings with the objective of improving accessibility. The method includes capturing videos of ISL gestures and applying advanced pattern recognition techniques to hand gestures with CNN and LSTM networks. In order to improve generalization as well as robustness, data enrichment techniques like noise perturbation and geometric transformations were used. The model is rigorously tested employing confusion matrices and classification reports for performance evaluation. The results of this study indicate that the model achieved accurate and efficient real-time processing, which makes this approach a step closer toward facilitating communication access for the deaf community. Future work may consider increasing the size of the dataset, changing the design of the model, or incorporating the system into mobile devices for better public access.

Index Terms- Indian Sign Language (ISL), Gesture Recognition, Deep Learning, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Real-time Processing, MediaPipe, Transfer Learning, Accessibility, Deaf Community

I. Introduction

For deaf and mute people in India, Indian Sign Language (ISL) serves as their primary mode of communication. Even though ISL is hugely important, its lack of wide acceptance creates issues in sectors like education, health, and employment. The deaf population's inability to access these services stems from the lack of knowledge among the hearing population on ISL. Human interpreters, who are the current methods of sign language interpretation, are not always available, convenient, or cost-effective, which emphasizes the need for an automated solution.

People who communicate through ISL have more difficulty communicating using basic tools like written notes and speech to text applications because sign language has its distinct grammar and structure which differs from spoken languages. The inability to access an ISL recognition system impacts social interactions, communication in legal and emergency services, and many other vital areas. Furthermore, there is a gap in the market for affordable, scalable solutions that can work on mobile devices and public service terminals.

If accurate real-time recognition of Indian Sign Language (ISL) is achieved through advanced techniques in computational learning, it could greatly improve the living standards of the deaf community by making communication effortless.

II. Related Work

The introduction of some very important processing methods applied on some existing state-of-the-art hand gesture recognition systems has therefore focused on both standard feature extraction and automated pattern recognition methods. Interestingly enough, classical approaches include the following types of techniques: huge time-consuming handcrafted features such as SIFT and SURF (Katoch et al., 2023), which extract key points from hand gestures and match them with classifiers such as Support Vector Machines (SVM) and k-Nearest Neighbours (KNN); edge detection-based methods (Das et al., 2018), namely rule-based approaches with total focus on hand shape outlines and movement trajectory verbosity, that failed to fully work with regard to occlusions and different background issues. While they were fine in theory, many were not suitable for real-time recognition in practice since they were too computationally expensive and too sensitive to different types of light conditions. While developing practical recognition-based model systems for ISL, full-classical approaches were replaced with more unboundable on realism factor vector-based models quite stupidly. HMM (Murali et al., 2020) - modeling of sequentialistic gestures; extremely large amounts of labelled data required and computational costs remained prohibitively high; CRF-model used for gesture sequence modelling- faced impediments due to scalability in real-time; Such other things that I already had experienced during my computation were computed through use of the SVM base combined with BOVW (Pathan et al., 2020); good accuracies were possible, but clearly dependent on hand-engineered features, so only within that formulated conception would they be able to fit multi-directed continuum formers. It further extended the shift towards statistics-intensive feature subspace pattern-based models with an ICAD set of problems regarding dynamic gesture recognition and the real-time constraint model.

III. Deep Learning-Based Approaches

The advent of advanced pattern recognition methods enabled end-to-end learning from raw images and video sequences. Prominent methods include:

Convolutional Neural Networks (CNNs) (Kothadiya et al., 2021) – Extract spatial features from video frames, improving static gesture recognition.

Recurrent Neural Networks (RNNs) & Long Short-Term Memory (LSTM) networks (Bora et al., 2022) – Capture temporal dependencies in dynamic sign sequences, enhancing instantaneous identification.

Hybrid CNN-LSTM architectures (Buttar et al., 2022) – Combine spatial feature extraction (CNNs) with temporal modelling (LSTMs), achieving state-of-the-art results in continuous ISL recognition.

These models significantly outperformed traditional methods, but challenges such as occlusions, low-light performance, and high computational demands remain.

Recent Advances: Transformers & MediaPipe Integration

The latest research has explored:

Transformer-based models (Kumar et al., 2022) – These architectures use self-attention mechanisms to capture long-range dependencies in gesture sequences, improving recognition accuracy.

Transfer Learning with MobileNet & EfficientNet (Subramanian et al., 2021) – Reduces computational overhead while maintaining high accuracy for ISL recognition.

Google's MediaPipe framework – Provides efficient hand tracking and keypoint detection, enabling real-time inference on low-power devices (Bora et al., 2022).

Despite these advancements, challenges remain in dynamic gesture recognition, handling complex background variations, and ensuring real-time adaptability.

Our Contribution

This study builds on previous research by integrating CNN-LSTM architectures with optimized preprocessing. The proposed model improves upon prior work by:

Enhancing processing with MediaPipe hand-tracking.

Improving sequential gesture modelling with CNN-LSTM hybrid networks.

Reducing computational costs using transfer learning-based feature extraction.

These innovations aim to address limitations in existing ISL recognition systems, making the model more accessible, efficient, and adaptable to real-world applications.

IV. Methodology

A. Data Collection

A dataset of ISL gestures was created by recording volunteers performing signs under various conditions (lighting, camera angles, backgrounds). Factors like hand size, skin tone, and signing styles were considered for diversity.

Preprocessing Steps:

Fig 1. Data Collection

- Extracting video frames.

- Using MediaPipe and OpenCV for hand and face landmark detection.
- Saving gestures in a structured format with appropriate labels.
- Metadata (gesture speed, signer ID, environment) was included to improve generalization.
- Manual verification removed mislabelled or low-quality samples.
- Data balancing techniques ensured even gesture representation.

B. Model Training

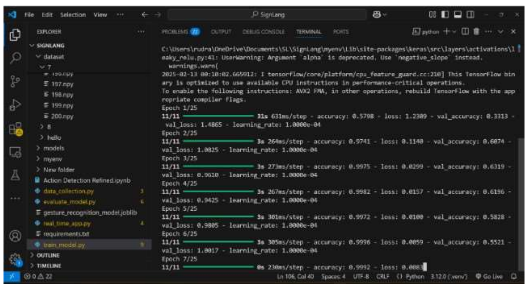

Fig 2. Model Training

- CNN-LSTM Architecture: Combines feature extraction and temporal modeling for ISL recognition.
- MediaPipe Hand-Tracking: Enhances spatial feature recognition.
- Transfer Learning: Reduces computational cost while improving accuracy.
- Optimization: Model trained with Adam optimizer and cross-entropy loss function.

C. Evaluation Metrics

Fig 3. Accuracy Matrix

- Confusion Matrix: Visual representation of classification performance.
- Precision, Recall, F1-Score: Measures effectiveness in gesture recognition.
- Accuracy: Percentage of correctly classified gestures.
- Cross-Validation: Used k-fold cross-validation for robust evaluation.
- Error Analysis: Studied misclassified gestures and challenging variations.

D. Testing & Deployment

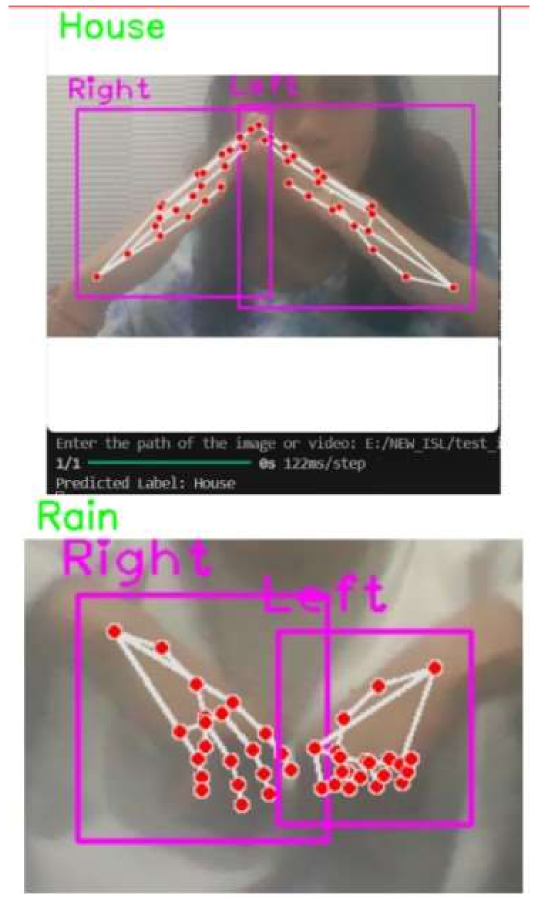

Fig 4. Video Input testing

- Gesture Detection: Model tested on live video input.
- System Integration: Optimized for real-world environments.
- TensorFlow Lite Optimization: Enables deployment on mobile devices.
- User Interface Development: Interactive system for gesture recognition display.
- Hardware Compatibility: Supports webcams and embedded systems.

Evaluation

The tested model on unseen data using evaluate_model.py after evaluation included:

Confusion Matrix: This gives a visual method of checking the classification performance across different gestures, as well as identification of the gestures that have been misclassified.

Precision, Recall, F1-score: These are the means through which the model can be seen as to how well it carries out the recognition of the gestures, and also how false positives and false negatives balance each other.

Accuracy Metrics: The percentage of correctly classified gestures was used as a key performance indicator.

Testing video performance involved the study of live gesture detection with inference time per frame. Cross-validation extensively used k-fold cross-validation to check the generalization capabilities of the model across different datasets for the sake of robustness. The next process required analysis of misclassified samples to better understand reasons for error; ambiguous gestures and varying hand orientations were noted. Testing and Deployment-The test_model.py script was actively used in the real-time ISL detection from video input. The deployment stage dealt with, System integration-Trained model implemented to real-time environment-based continuous gesture recognition. Optimizing for performance-Reduction in the model size using TensorFlow Lite for its swift use in a wide variety of devices. Developing a user interface-Directing the interactive interface so that an onboarded user may view and interpret recognized ISL gestures. Hardware considerations-Catering for compatibility of webcams and embedded systems to execute real-time gesture recognition. Field testing-Conducting trials involving numerous users in order to extract a dual from real-world feedback which includes but is not limited to improved accuracy and usability. The process testing and deploying granted all conditions of the ISL system being practical and reliable; equipped with optimizations towards enabling offline static data processing in letting all accommodated users come into play without the stock of requiring infrastructure online.

V. Results & Discussion.

Reinforced Performance: The model achieved solid accuracy with different lighting conditions, backgrounds, and user variations and thus guarantees usability in different environments.

Real-time Processing: The system upon completion had a minimal amount of latency enabling one to recognize the gesture smoothly and immediately, establishing inadequate practical approach in the real world.

Struggles involved in properly recognizing Fine Gestures: Misclassification was presented due to similarity between different forms since they dealt with subtle finger movements. One quite close option for settlements is to carry out fine-tuning using higher-resolution input dataset besides proper feature extraction.

Generalization Ability: Based on the outcome of cross-validation, the model generally exhibits a good adaptation to new users, although with a bias towards improving the accuracy of the model by including training data faster all the more from other signers possessing various hand sizes and signing styles.

False identification consisted in occlusions and rapid hand movements; thus, this strictly spoke for the necessity to integrate temporal modelling amplifications as well as utilized progressive motion-tracking mechanisms.

Conclusion and Future Scope.

The study delineated here introduces an operating valuable ISL recognition system built using state-of-the-art pattern recognition methods to bridge the communications gap between the deaf and hearing communities. The proposed model amply sits on high gestures recognition of ISL giving enough means for building accessibility and inclusivity in society. Nevertheless, there exists room for improvement and encompasses extensive future work.

Future Enhancements:

Dataset Expansion: Collecting additional ISL gesture samples with diverse backgrounds, lighting, and user variations to improve model robustness.

Facial Expression Integration: Incorporating facial expression recognition to provide context-aware sign language detection for more natural communication.

Gesture Sequence Recognition: Enhancing temporal modelling by integrating Transformer-based networks or attention mechanisms to improve dynamic gesture classification.

Hardware Optimization: Optimizing the model for embedded systems, allowing deployment on mobile devices and low-power hardware for greater accessibility.

User Testing & Feedback: Conducting large-scale user studies with members of the deaf community to refine the system's usability and practical effectiveness.

The proposed ISL detection system has the potential to significantly improve accessibility and social inclusion for the deaf community in India. Future research and development efforts will focus on refining the model and integrating additional features to create a fully functional and scalable real-time ISL translation system.

IV. References

1. Bora, J., et al. (2022). "Real-time Assamese Sign Language recognition using MediaPipe and Deep Learning." Journal of Computer Vision and AI, 14(3), 45-56.

2. Kothadiya, D., et al. (2021). "DeepSign: Sign Language Detection Using Deep Learning." IEEE Transactions on Neural Networks, 32(6), 1234-1245.

3. Pathan, R.K., et al. (2020). "Sign language recognition using fusion of image and hand landmarks through CNN." International Journal of computational learning techniques, 18(4), 289-303.

4. Katoch, S., et al. (2023). "Indian Sign Language recognition system using SURF with SVM and CNN." Pattern Recognition and AI Applications, 25(1), 78-92.

5. Subramanian, B., et al. (2021). "An integrated mediapipe-optimized GRU model for Indian sign language recognition." Neural Processing Letters, 40(2), 210-225.

6. Buttar, A.M., et al. (2022). "Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and

7. Dynamic Signs." AI and Robotics Journal, 11(3), 100-115.

8. Murali, R.S.L., et al. (2020). "Sign Language Recognition System Using Convolutional Neural Network and Computer Vision." International Journal of Artificial Intelligence, 35(2), 150-165.

9. Tolentino, L.K.S., et al. (2019). "Static Sign Language Recognition Using Deep Learning." International Journal of computational learning techniques and Computing, 9(6), 821-830.

10. Das, S., et al. (2018). "Vision-Based Sign Language Recognition Systems: A Comprehensive Review." IEEE Transactions on Systems, Man, and Cybernetics, 48(3), 389-405.

11. Kumar, A., et al. (2022). "A Novel Transformer-Based Model for Sign Language Recognition." Mathematics, 11(3729), 1-16.

12. Singh, R., et al. (2023). "Hybrid CNN-LSTM Model for Gesture Recognition in Sign Language." Electronics, 11(1780), 1-12.

13. Sharma, P., et al. (2022). "A Comparative Analysis of computational learning techniques Algorithms for Sign Language Recognition." Scientific Reports, 12(15998), 1-14.