TITLE:

EduSign – Indian Sign Language (ISL) Learning Platform

Submitted by

Swapna K (24PG00192)

M. Reagan (24PG00075)

(Submitted as part of the Project Course – Summative Assessment 2025–26)

Under the guidance of

Mr. Ashith Sagar Naidu

Course: Master of Computer Applications (M.C.A)

Project Report

III Semester 2025 – 2026

School of Engineering

CERTIFICATE

This is to certify that the major project work carried out entitled **"EduSign – Indian Sign Language (ISL) Learning Platform."** submitted to the School of Engineering, Chanakya University in partial fulfilment of the requirements of the degree of Master of Computer Applications in the academic year 2025-2026 is a record of the original work done by **M. Reagan (24PG00075) AND Swapna K (24PG00192)** under my supervision and guidance and that this major project work has not formed the basis for the award of any Degree / Diploma / Associateship / Fellowship or similar title to any candidate of any University.

Place: Devanahalli

Date:17-01-2026

Signature of Student: Signature of the Guide

1.M Reagan
2.Swapna K

AKNOWLEDGEMENT

We express our sincere gratitude to all those who contributed to the successful completion of our project, **“EduSign – Indian Sign Language (ISL) Learning Platform.”**

We would like to extend our deepest thanks to our project guide **Mr. Ashith Sagar Naidu**, for his constant guidance, technical expertise, and valuable feedback throughout the development of this project. His encouragement and constructive suggestions helped us overcome challenges and refine both the technical and conceptual aspects of our work.

We are also immensely thankful to **Dr. Bharath Setturu**, whose academic vision and leadership have played a significant role in shaping the learning and research culture at **Chanakya University**. His emphasis on innovation, inclusivity, and socially impactful technology inspired us to work on a project that addresses real-world educational needs, particularly for users and the hearing-impaired community.

We sincerely thank the School of Engineering, **Chanakya University**, and the faculty members of the **Master of Computer Applications (MCA)** program for providing the necessary infrastructure, academic support, and encouragement required to successfully carry out this project.

TABLE OF CONTENT

Sr. No.	Title	Page No.
01	Introduction	2
02	Literature Review	3-4
03	Data Sources [Code]	5-7
04	Methodology	8-14
05	Results and Discussion	15-22
06	Challenges Faced	23-24
07	Conclusion	25
08	Future Scope	26
09	References	27

ABSTRACT

EduSign is a real-time, interactive, and user-friendly Indian Sign Language (ISL) learning platform created to help users to learn and practice sign language in a natural and engaging way. By using live gesture recognition through a webcam, the platform allows users to actively participate in learning rather than just watching demonstrations.

With the support of computer vision, deep learning, and modern web technologies, EduSign provides instant visual feedback, making the learning process both effective and enjoyable.Traditional ISL learning methods usually depend on static images or prerecorded videos, which offer limited interaction and no real-time correction. EduSign overcomes this limitation by letting users perform signs live while the system continuously analyzes their hand movements and gesture patterns. MediaPipe is used to accurately extract hand and body landmarks, while deep learning models such as Dense Neural Networks (DNN) and Long Short-Term Memory (LSTM) networks are used to recognize both static gestures and dynamic movements.

The platform follows a clear and gradual learning path. Children start with simple static gestures like alphabets (A–Z) and numbers (0–9), and slowly progress to motion-based words, colours, days of the week, and full sentence-level gestures. Throughout this process, EduSign provides immediate and easy-to-understand feedback, including predicted labels, confidence scores, stability indicators, and visual skeleton overlays. To keep learners motivated, the platform also includes fun gamification features such as badges, rewards, progress tracking, and animated mascots.Technically, EduSign is developed using a modular client–server architecture. The frontend is built with React, while the backend uses Flask, with real-time communication enabled through Socket.IO. User progress and achievements are securely stored using Firebase.

Overall, the system demonstrates accurate real-time gesture recognition, smooth interaction, and a supportive learning environment tailored for children. EduSign not only works as an effective educational tool today but also provides a strong foundation for future enhancements, such as more advanced sentence recognition and mobile-based learning support.

I. INTRODUCTION

EduSign is a real-time and user-friendly Indian Sign Language(ISL) learning platform created to help users to learn sign language in an interactive and engaging way. Instead of only watching to help peoples to learn sign language in a interactive and engaging way. Instead of watching images or videos, peoples can learn using a realtime sign recognition system like using a webcam to check whether the gesture is correct or not, by combining computer vision, deep learning and modern web technologies, EduSign provides sign language learning into an active, hands-on- experience for the children.

In India, Indian Sign Language(ISL) is a way through which users's communication with hearing and speech impairments. However, most existing ISL learning resources depend heavily on static pictures or recorded videos. While these is a way of learning but they do not allow learners to practice signs on their own or receive instant correction. As a result, children may feel confused or unsure about their gestures. EduSign overcomes this problem by giving real-time feedback, helping learners understand and improve their signs as they practice.The platform is designed with a step-by-step learning approach suitable for beginners. users first learn basic static gestures such as alphabets (A–Z) and numbers (0–9), which are the building blocks of Indian Sign Language. Once they gain confidence with the basics, they move on to motion-based lessons that include words, colours, days of the week and general words which includes both static and motion words. This allow the users's to grow their skills in learning the Indian Sign Language.

Along with learning individual words, EduSign also helps children practice full sentences in Indian Sign Language. This allows them to see how separate signs naturally connect to express complete ideas. By recognizing longer gesture sequences and patterns over time, the sentence-level module makes learning more practical, engaging, and closer to how sign language is used in real-life communication.

Overall, EduSign aims to create a supportive,engaging, fun, and inclusive learning environment for children. By combining real-time gesture recognition with a playful and gamified interface, the platform encourages regular practice, builds confidence, and makes learning Indian Sign Language both effective and enjoyable. EduSign demonstrates how technology can be used meaningfully to support inclusive education and empower young learners.

II. LITERATURE REVIEW

Sign Language Recognition (SLR) has been an important research which is been going on currently for the need to reduce communication barriers for the deaf and the hearing population. Early works on sign language recognition systems the focus was more on features such as skin colour detection, contour extraction, and geometric analysis of hand shapes [9]. Although these approaches demonstrated the feasibility of gesture recognition, they were highly sensitive to lighting conditions, background clutter, and camera variations. This were the limitations in real-time and practical recognizing environments.

With the rapid advancement of deep learning, researchers began adopting data-driven approaches for the sign language recognition such as Convolutional Neural Networks (CNNs) emerged as an effective solution for recognizing static hand gestures. CNN-based models are able to automatically learn important visual features from images or hand landmark data, which reduces the need for manual feature extraction. Many studies have shown that CNNs perform much better than traditional machine learning methods like SVM and k-NN, especially for recognizing static gestures. Research specifically focused on Indian Sign Language (ISL) has also reported higher accuracy and better reliability with CNN-based approaches, even when gestures are performed under different lighting and background conditions.

However, static gesture models were not only the solution for recognizing all the Indian Sign Language signs which also include motion signs. Many ISL gestures consist of dynamic hand movements that cannot be captured using single-frame analysis. So to overcome this limitation, researchers introduced sequence-based models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. LSTM models are well-suited for capturing temporal dependencies across consecutive frames and making them well suited for dynamic sign recognition. Many studies have shown that LSTM-based approaches can successfully recognize motion-based gestures, words, and short phrases by tracking hand and body landmark movements across consecutive frames.

A significant advancement in continuous sign language recognition and translation was introduced by Camgoz_Sign_Language_Transformers_Joint_End-to-End_Sign_Language_Recognition_and_Translation. through transformer-based architectures [10].This work proposed an end-to-end framework that jointly performs Continuous Sign Language Recognition (CSLR) and Sign Language Translation (SLT) using transformer encoders and decoders. By combining transformer encoders and decoders with Connectionist Temporal Classification (CTC) loss, the model learns how signs align over time without needing detailed frame-by-frame labels. Their results achieved state-of-the-art performance on large datasets, clearly showing how attention mechanisms are powerful for understanding long and complex sign sequences.

Despite these advancements, most existing sign language recognition systems were mainly focused on translation accuracy rather than supportive learning. Many systems are designed for adult users and are designed to convert sign language videos into spoken or written text. As a result, they often do not provide real-time feedback, learning guidance, or interactive features that are important in educational features. In addition, transformer-based models require high computational power and large datasets, which makes them not suitable for lightweight, real-time learning platforms intended for users. [4].

Recent studies focused on Indian Sign Language (ISL) have explored landmark-based recognition methods using frameworks like MediaPipe. These approaches reduce computational complexity compared to raw video processing while still delivering good accuracy. Many systems combine MediaPipe-based landmark extraction with CNN or LSTM models to achieve better real-time performance. However, most of this work is limited to simple tasks such as recognizing alphabets or numbers and does not offer a well-structured learning path that progresses across different difficulty levels.From an educational point of view, very limited research has concentrated on child-friendly ISL learning platforms. Most existing solutions depend on pre-recorded videos or static demonstrations, which provide limited interaction. Without real-time correction and confidence-based guidance, children may lose interest, and their learning process becomes slower and less effective.

As a result, learning becomes slower and less effective. EduSign builds based on existing research while directly addressing the gaps identified in earlier studies. Instead of focusing mainly on translation, EduSign is designed around learning and practice. It provides real-time feedback as users perform ISL gestures, helping them understand and correct their signs immediately The platform integrates CNN-based models for static gestures such as alphabets, numbers, and static words, along with LSTM-based sequence models for dynamic gestures including motion words, colours, days, and sentences. By using MediaPipe-based landmark extraction [1] and lightweight deep learning architectures. EduSign is able to deliver smoother real-time performance that is well suitable for interactive learning.

III. DATA SOURCES

The data used in the EduSign project was entirely self-collected to ensure accuracy, consistency, and suitability for a child-centric Indian Sign Language (ISL) learning platform. Publicly available ISL datasets are limited and not optimized for real-time educational applications, which motivated the creation of a custom dataset tailored to EduSign's learning modules.All gesture data was recorded using a standard webcam in a controlled indoor environment. The gestures were performed at a natural pace suitable for users, and each sign was recorded multiple times to capture variations in hand position, speed, and orientation.

The dataset includes:

- ISL Alphabets (A–Z)
- Numbers (0–9)
- A–Z vocabulary words
- Days of the week
- Colours
- Motion-based general words
- Static general words
- Basic sentence-level gestures

Instead of storing raw video frames, landmark-based features were extracted using Google MediaPipe:

- MediaPipe Hands for static gestures
- MediaPipe Holistic for motion-based words and sentences

The extracted **(x, y, z)** coordinates of hand, face, and pose landmarks form the input features for both model training and real-time inference.All extracted features are stored in NumPy **(.npy)** format, allowing fast loading, efficient memory usage, and seamless integration with TensorFlow/Keras models. The data is organized class-wise, enabling structured training and easy future expansion.

A custom dataset was chosen due to the lack of standardized ISL datasets for users, the need for real-time compatibility, and the requirement for consistent landmark formats. This approach ensured better control over data quality and improved alignment between training conditions and real-world usage in the EduSign platform.

CODE:

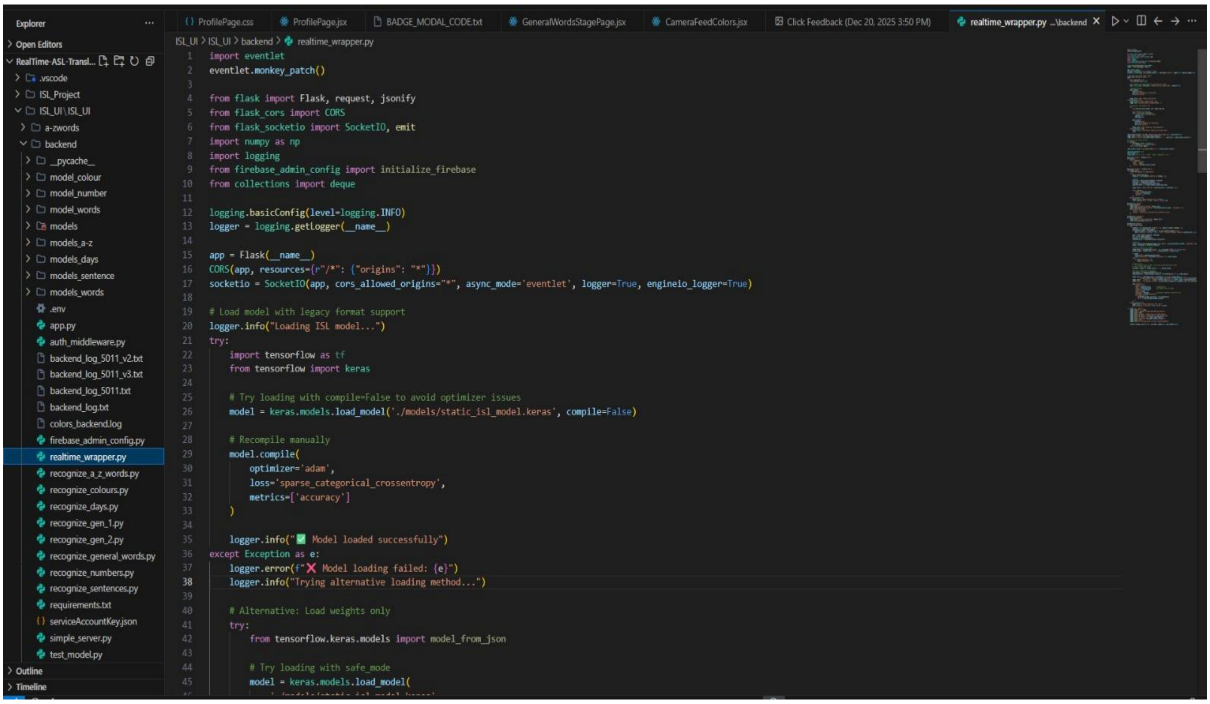


Figure 1: realtime_wrapper.py

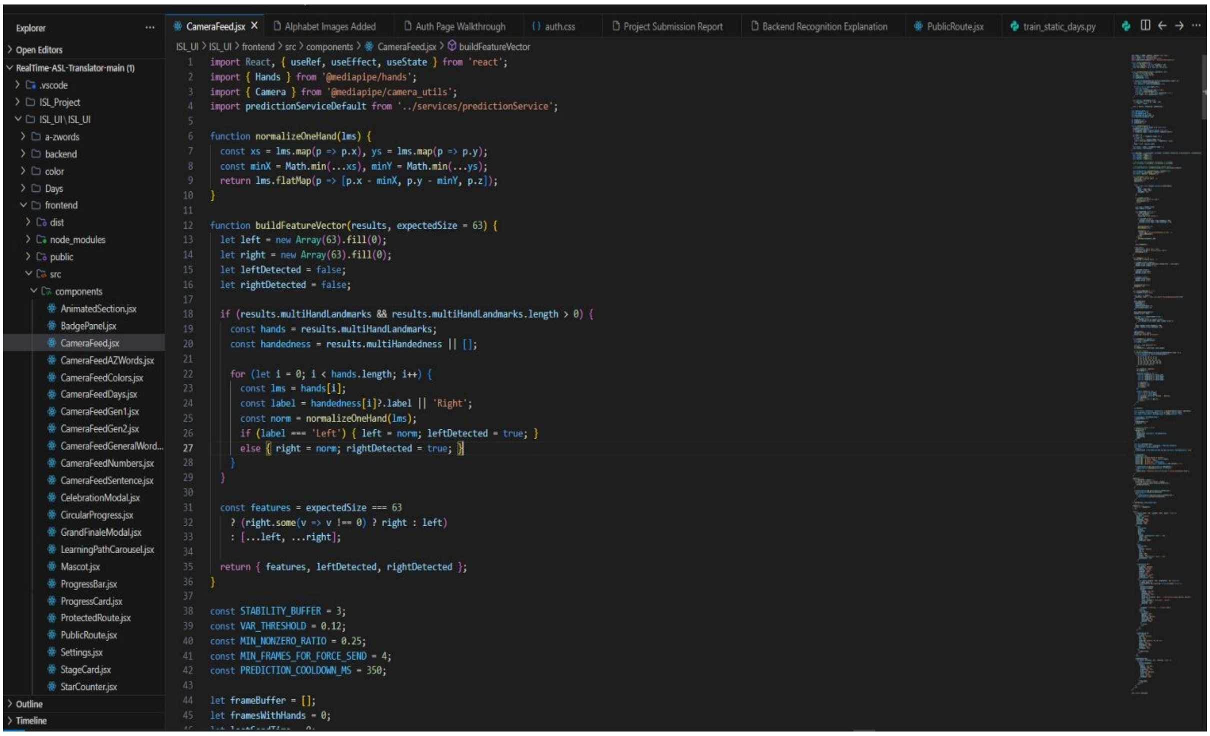


Figure 2: cameraFeed.jsx

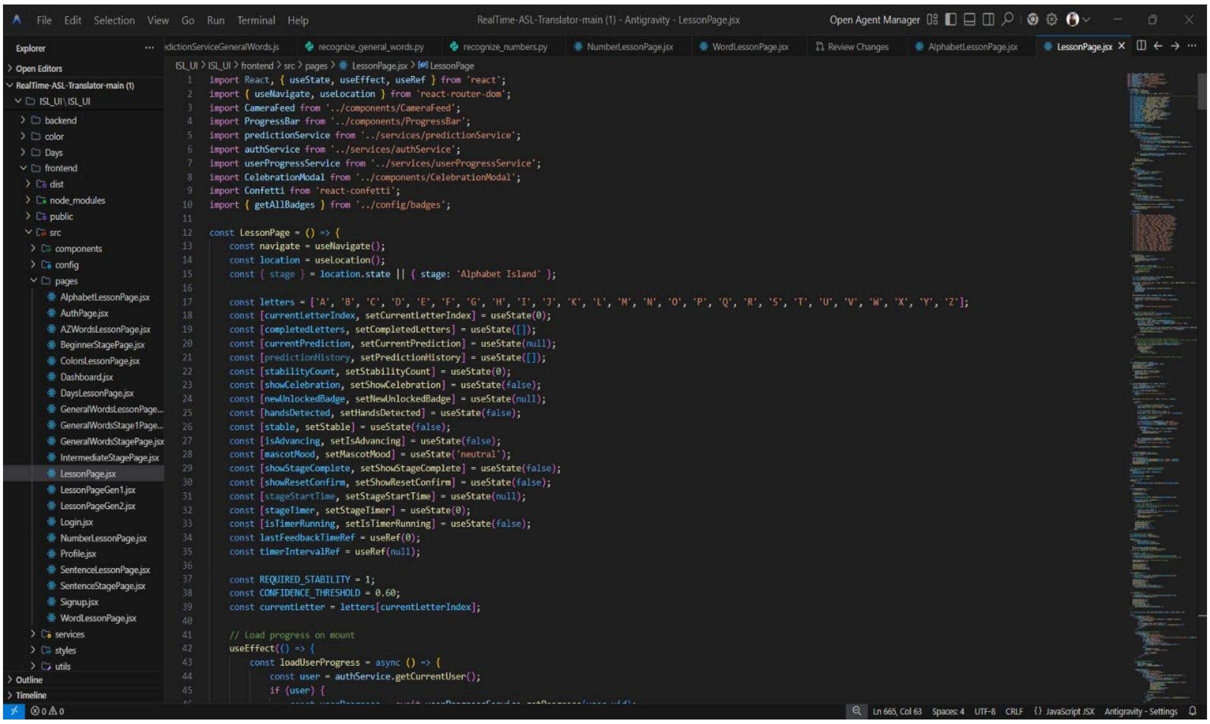


Figure 3: lessonPage.jsx

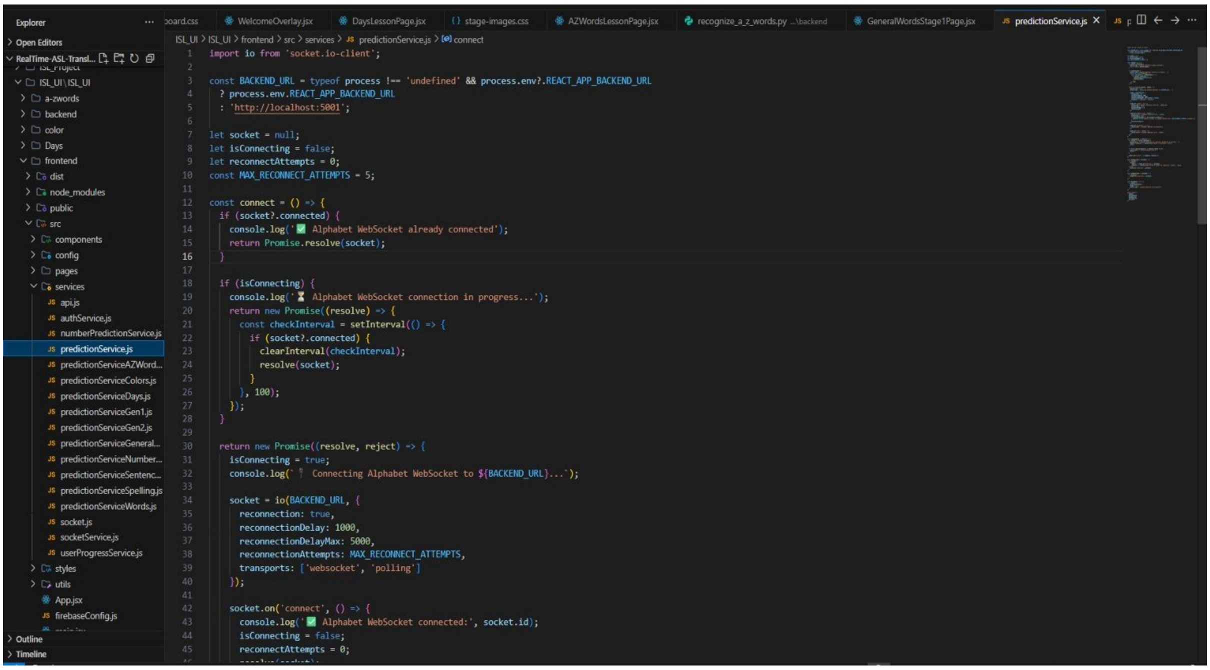


Figure 4: predictionService.js

IV. METHODOLOGY

The methodology of the EduSign project is designed as a clear, structured, and end-to-end learning pipeline that supports real-time Indian Sign Language (ISL) learning for children`s and everyone. The system brings together live video capture, landmark extraction, deep learning–based gesture recognition, and interactive feedback in a seamless manner. Special attention is given to real-time responsiveness, stable predictions, and creating a learning experience that feels friendly and encouraging for children.

4.1 System Overview

EduSign follows a real-time client–server architecture. The learning process begins when a user performs ISL gestures in front of a webcam. The live video stream is continuously processed to extract hand and body landmarks using MediaPipe. These landmarks are converted into numerical feature vectors, normalized, and then sent to train using deep learning models for recognition.

Based on the level, the system uses different models:

- **Static models** for alphabets, numbers, days, and static words.
- **Motion based** model for motion-based words, colours, and sentence-level gestures.

Before showing result output to the learner, predictions are checked using confidence thresholds and stability measures. Only reliable predictions are displayed on the user interface. The learner's progress, scores, and achievements are then saved in the backend for tracking and motivation.

4.2 Flowchart

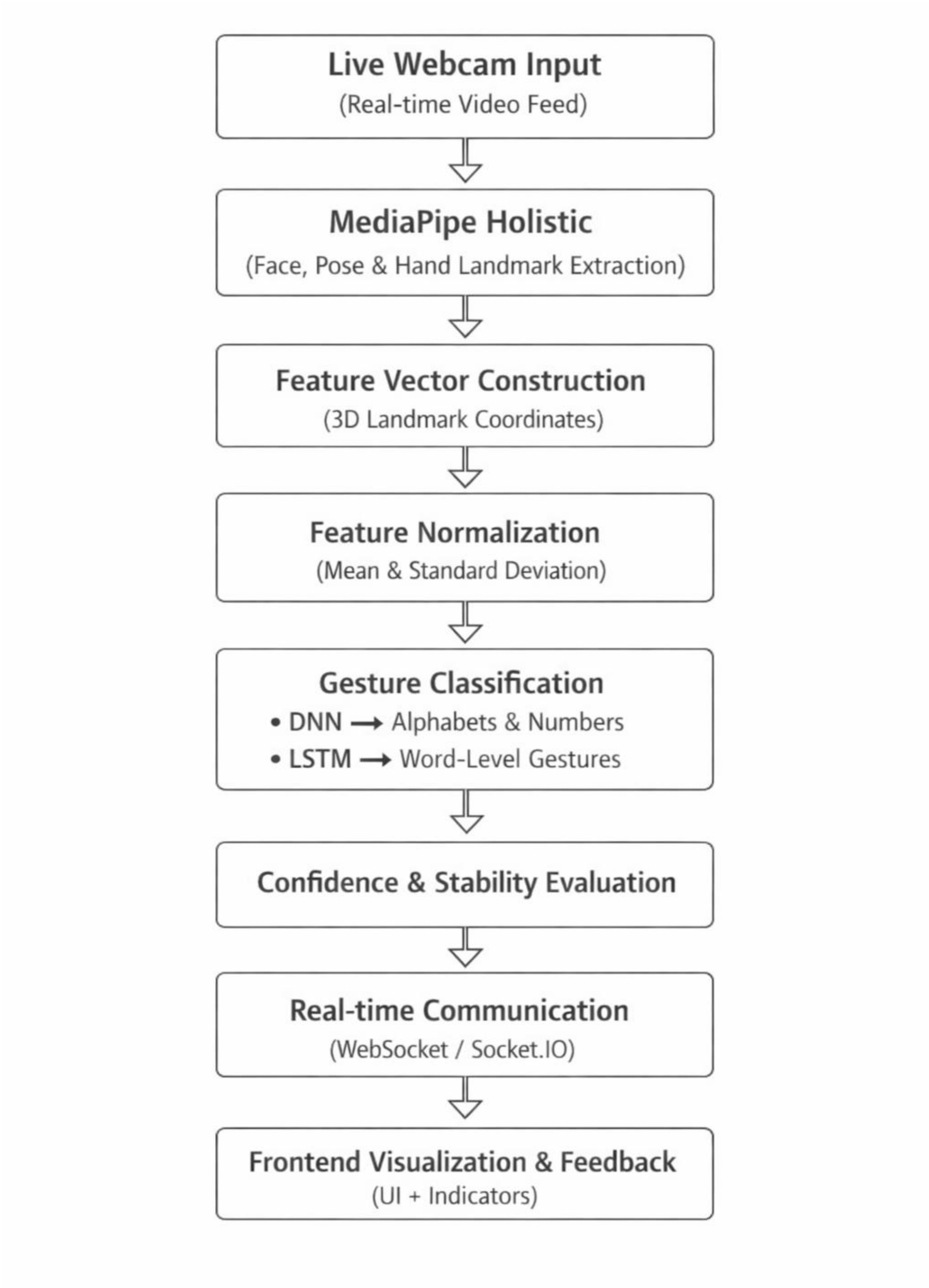


Figure 4.1: Flowchart of the EduSign real-time gesture recognition system

4.3 Model Architecture Logic

EduSign uses specialized deep learning models that are carefully chosen based on the type of Indian Sign Language (ISL) gesture being learned. Rather than using a single, generic model for all gestures, the system applies different architectures for static and dynamic gestures. This design choice helps improve recognition accuracy and makes the learning process more effective and reliable for learners.

i. **Static Gesture Recognition (DNN-Based)**

Static gestures such as alphabets, numbers, days, and static vocabulary words are recognized using a Dense Neural Network (DNN). These gestures are defined primarily by hand shape rather than motion. The model processes single-frame landmark feature vectors consisting of either 63 or 126 coordinates, depending on whether one or both hands are involved.

The network consists of multiple fully connected layers combined with Batch Normalization to stabilize training and Dropout to reduce overfitting. This architecture ensures reliable recognition even when minor variations occur in hand position, orientation, or finger alignment.

ii. **Dynamic Gesture Recognition (LSTM with Attention)**

Dynamic gestures such as motion-based words, colours, and sentences require understanding movement over time. For this purpose, EduSign employs a multi-layer Bidirectional LSTM architecture that processes temporal sequences of landmark data.

Word-level gestures are represented using 30-frame sequences, while sentence-level gestures use 60 frames to capture longer temporal dependencies. An Attention Mechanism is integrated into the network to automatically emphasize the most informative frames within a sequence, such as the peak or transition phase of a gesture. This allows the model to distinguish visually similar gestures based on subtle motion and timing differences rather than static posture alone.

4.4 Dataset and Preprocessing

To maintain accuracy and consistency, all datasets used in EduSign are self-collected in controlled lighting conditions using our local laptop webcam. The data includes both static and dynamic ISL gestures and is carefully aligned with the different learning stages of the platform.For feature extraction, MediaPipe Hands is used for static gestures, while MediaPipe Holistic is used for dynamic gestures such as words and sentences. The extracted landmark coordinates **(x, y, z)** are normalized by centering them relative to the wrist position and scaling them using the standard deviation. This step helps the system handle variations in camera distance, hand size, and user positioning.

All processed landmark features are stored in NumPy **(.npy)** format, which allows fast data loading during model training and supports efficient, real-time inference during live learning sessions.

4.5 Architecture of the System

The EduSign system is built using a layered architecture to ensure scalability, modularity, and smooth interaction between components.

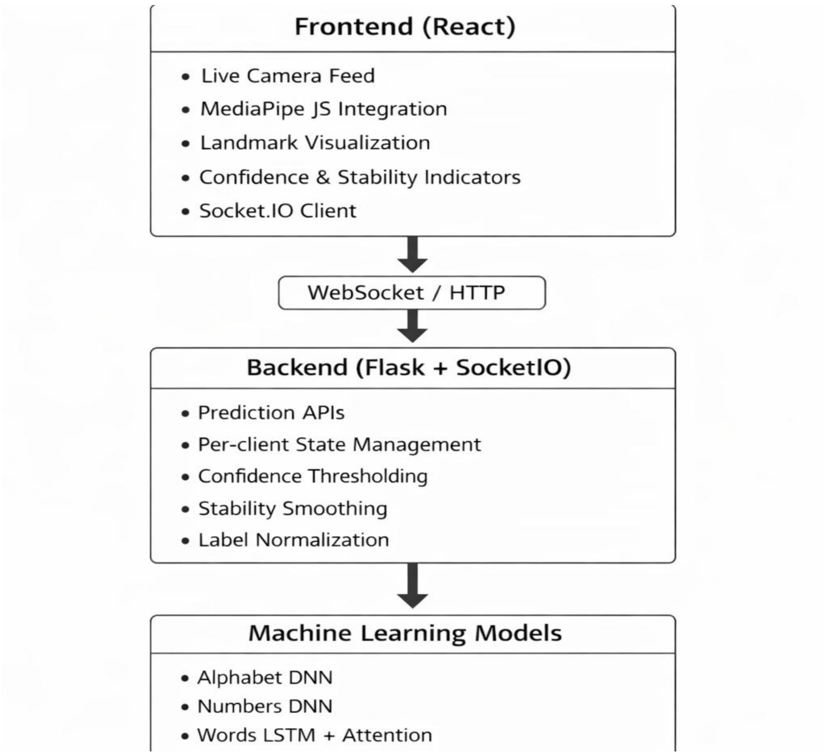


Figure 4.2: Architecture of the EduSign system.

4.6 Tools and Technologies Used

- **MediaPipe Hands & Holistic:** Used for real-time extraction of hand coordinates for basic signs, and full-body holistic landmarks for sentence recognition.
- **TensorFlow with Keras:** Used to develop and train DNN (Dense Neural Network) and LSTM (Long Short-Term Memory) deep learning models.
- **Python:** Used for backend development and model integration.
- **Flask and Flask-SocketIO:** Used to build backend microservices and handle real-time communication.
- **Socket.IO:** Enables low-latency, bidirectional communication between frontend and backend.
- **ReactJS:** Used to create an interactive, child-friendly frontend interface.
- **Firebase:** Used for user authentication, progress tracking, and secure data storage.
- **HTML, CSS, and JavaScript:** Used for UI design and visual feedback.

4.7 Dataset Description

The datasets used in the EduSign project are self-collected to ensure accuracy, consistency, and suitability for user-oriented Indian Sign Language (ISL) learning. All gesture data was recorded using a standard webcam under controlled lighting conditions. The data collection strategy is structured to support the diverse learning modules in EduSign: **Sign Basics, Word Wonderland, and Sentence Safari.**

1. **Static Gesture Dataset (Sign Basics & Static Words)**

The static gesture dataset consists of single-frame landmark feature vectors and is used to recognize ISL signs that depend mainly on hand shape and finger positioning, rather than movement. This dataset forms the foundation of beginner-level learning in EduSign and is primarily used to train Dense Neural Network (DNN) models.

A. **Basic Signs**

This section introduces learners to the fundamental building blocks of Indian Sign Language:

- **Alphabets (A–Z):**
  The dataset includes 26 classes, one for each alphabet letter. These signs helps user`s to learn basic finger formations and hand orientations that are essential for ISL.

- **Numbers (0–9):**
  This section contains 10 classes representing numeric hand signs. Since many number signs differ only slightly in finger positions, they are well suited for static classification using DNN models.

B. Word Wonderland – Static Section

This section focuses on vocabulary that can be recognized using a fixed hand posture:

- **Static Words (General Stage 2):**
  Words such as "Bad", "Food"," and "Home" where the meaning is conveyed mainly through a specific hand shape rather than movement.
- **Days of the Week:**
  Seven classes representing Monday to Sunday. While these signs may involve small movements in real-life use, the system is trained using key static poses. During live recognition, stability checks and temporal smoothing help ensure accurate and consistent predictions.

2. Dynamic Gesture Dataset (Word Wonderland & Sentence Safari)

The dynamic gesture dataset is designed to recognize motion-based Indian Sign Language gestures, where meaning is conveyed through a sequence of movements rather than a single hand shape. Since these gestures rely on the flow and timing of motion, EduSign uses LSTM models to capture temporal patterns across consecutive video frames. Each gesture is recorded as a sequence of frames, allowing the model to learn how movements evolve over time and accurately recognize dynamic ISL signs.

A. Word Wonderland – Motion Sections

This section focuses on commonly used ISL words that involve clear and meaningful hand movement:

- **A–Z Vocabulary Words:**
  A collection of 26 dynamic words linked to alphabet letters, helping user's strengthen vocabulary while reinforcing their understanding of alphabets.
- **Motion Words (General Stage 1):**
  Includes action-based signs such as "Hello," "Thank You," and "Come," where direction, speed, and rhythm of movement are important for recognition.

- **Colours:**
  A set of 11 colour signs (for example, Red, Blue, and Green), which often involve repeated or directional hand movements.

Each word gesture is represented using 30 consecutive frames, providing sufficient temporal information for reliable recognition.

For dynamic gesture recognition, the input to the system is a temporal sequence of frames represented as (T × F), where T denotes the number of frames in a gesture sequence and F represents the feature size per frame. To extract meaningful features, MediaPipe Holistic is used to capture full-body context from each video frame. This includes 468 facial landmarks, 33 pose landmarks, 21 landmarks from the left hand, and 21 landmarks from the right hand.

In total, this results in up to 543 landmarks per frame, corresponding to approximately 1629 feature values when considering the three-dimensional (x, y, z) coordinates. For improved efficiency and real-time performance, only the most relevant subsets of landmarks are selected based on the task being performed. By learning temporal motion patterns across these landmark sequences, the dynamic gesture dataset enables EduSign to accurately recognize words and full sentences. This approach supports more natural and expressive ISL learning, allowing children to progress smoothly from basic signs to fluent sentence-level communication.

Data Processing & Storage

For each video frame captured, MediaPipe extracts accurate (x, y, z)landmark coordinates. Static gesture recognition uses MediaPipe Hands for quick and efficient single-frame processing, while dynamic and sentence-level recognition relies on MediaPipe Holistic to capture hand movements along with body posture and facial cues.To ensure consistent performance across different users and camera setups, all landmark data is normalized by taking the wrist as a reference point and scaling the values using standard deviation. This step makes the system more robust to changes in distance, orientation, and user movement.

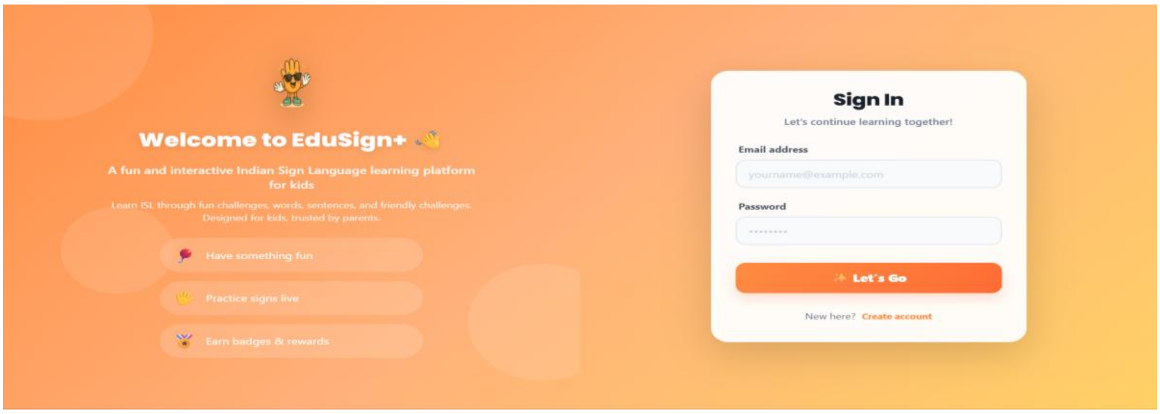
The processed landmark features are stored in NumPy (**.npy**) format, enabling fast data loading, efficient model training, and smooth real-time inference during live system use.

V. RESULTS AND DISCUSSION

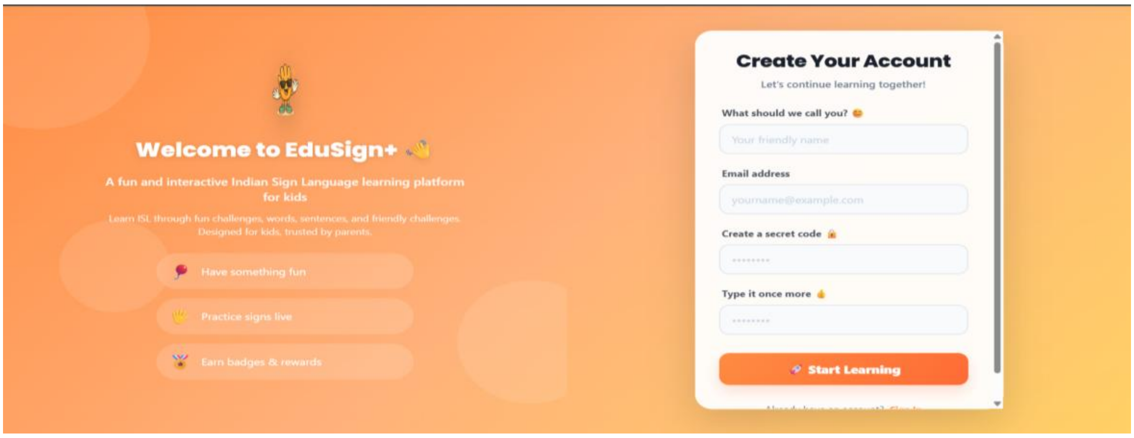
This section presents the final results and outcomes of the EduSign project after successful implementation of all planned modules. The system was tested extensively using real-time webcam input, and screenshots from the deployed web application are included to demonstrate functionality, usability, and learning flow.

5.1 Authentication and User Access

EduSign implements a secure and reliable authentication system using Firebase Authentication to ensure safe user access. Learners can register and log in using either an email and password, making the process simple and flexible. User sessions are securely maintained across logins, allowing user to continue learning without repeated sign-ins. The login and signup interfaces are designed with a clean layout, bright visuals, and minimal complexity, ensuring that user’s can easily understand and navigate the process.



Figures 5.1.1: Login page of the EduSign platform.



Figures 5.1.2: Signup page of the EduSign platform.

The authentication system performs efficiently, providing strong security while maintaining a user experience.

5.2 Dashboard and Navigation

The Dashboard serves as the central control panel of the EduSign platform and plays a key role in guiding learners through their learning journey.

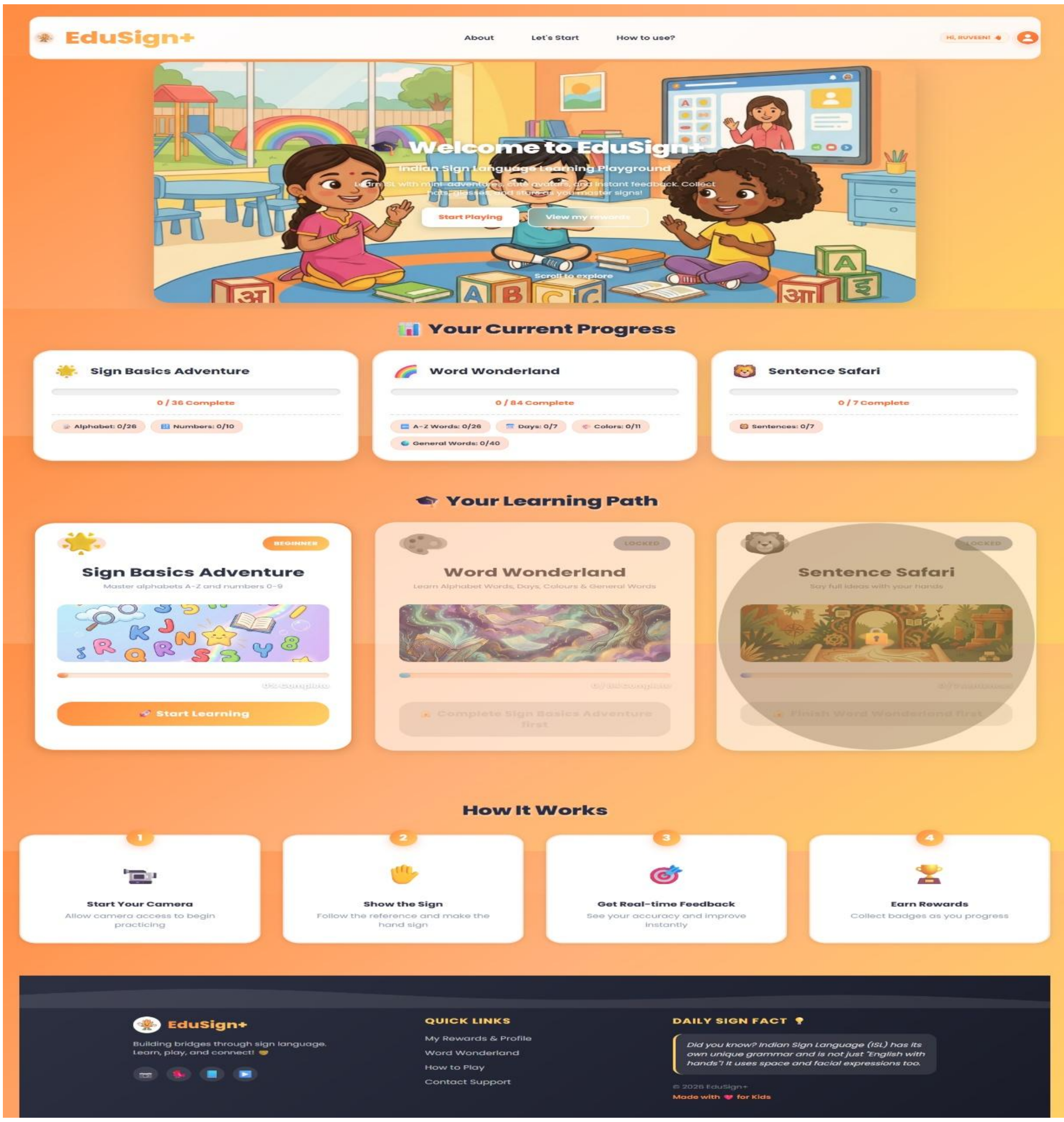


Figure 5.2 shows the Dashboard interface with stage-wise navigation.

It clearly displays all learning stages, including the Beginner Stage, the Intermediate Stage (Word Wonderland), and the Sentence Safari stage. Locked and unlocked stages are visually distinguished, helping users to understand their progress and upcoming lessons. The dashboard allows smooth and intuitive navigation between different learning modules, ensuring that learners can move easily from one activity to another. The dashboard provides a structured learning path and improves user engagement by visually guiding learners through different stages.

5.3 Profile Page and Achievement System

EduSign provides each learner with a personalized Profile Page that reflects their individual learning progress and achievements. The profile displays earned and locked badges, allowing users to clearly see what they have accomplished and what they can unlock next. In addition, learners can unlock fun mascot accessories such as hats, glasses, and shoes as they progress through different stages, making the experience more engaging and rewarding. All progress-related data is securely stored and retrieved using Firebase Firestore, ensuring consistency across sessions.

Figure 5.3: shows the Profile Page with badges, avatar, and progress indicators.

This achievement system effectively motivates learners, encourages regular practice, and gives children a sense of accomplishment.

5.4 Beginner Stage – Sign Basics Adventure

The Beginner Stage focuses on helping children build a strong foundation in Indian Sign Language by learning basic signs. This stage includes alphabet and number recognition using static gesture models. For alphabet learning, the system recognizes ISL alphabets from A to Z using a Dense Neural Network (DNN) model. Learners receive real-time visual feedback through hand skeleton overlays, along with confidence and stability-based confirmation to ensure accurate recognition.

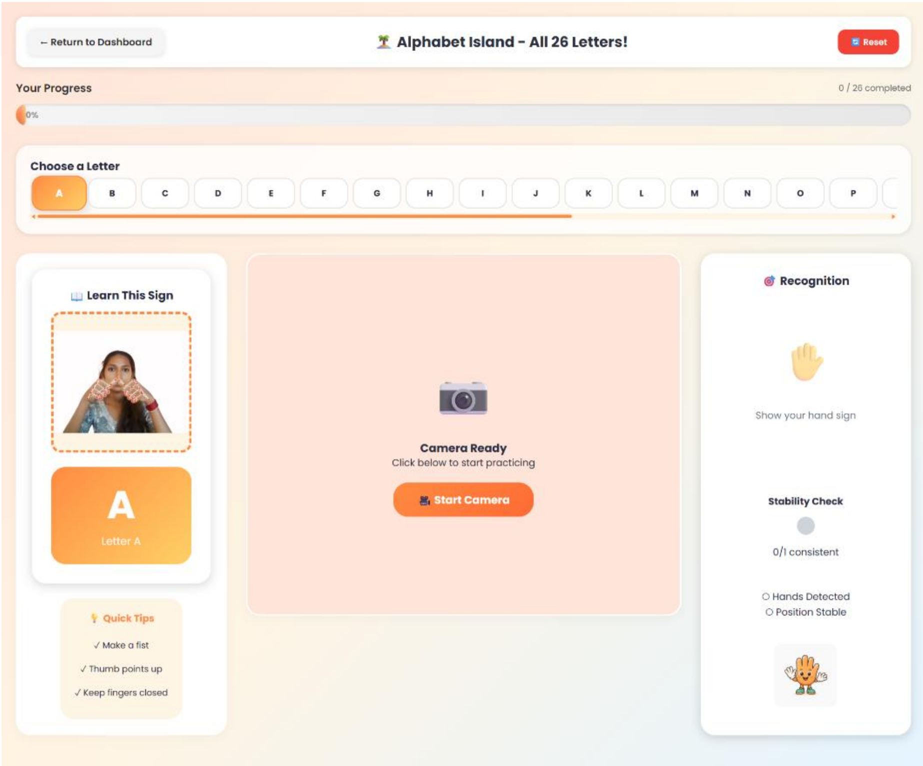


Figure 5.4.1: shows the Alphabet Lesson page.

The system achieved an accuracy of 95–97% for alphabet recognition, demonstrating stable and reliable real-time performance.

Number recognition is also introduced at this stage, covering digits from 0 to 9. The system applies dominant-hand detection and normalization techniques to improve recognition under different conditions.

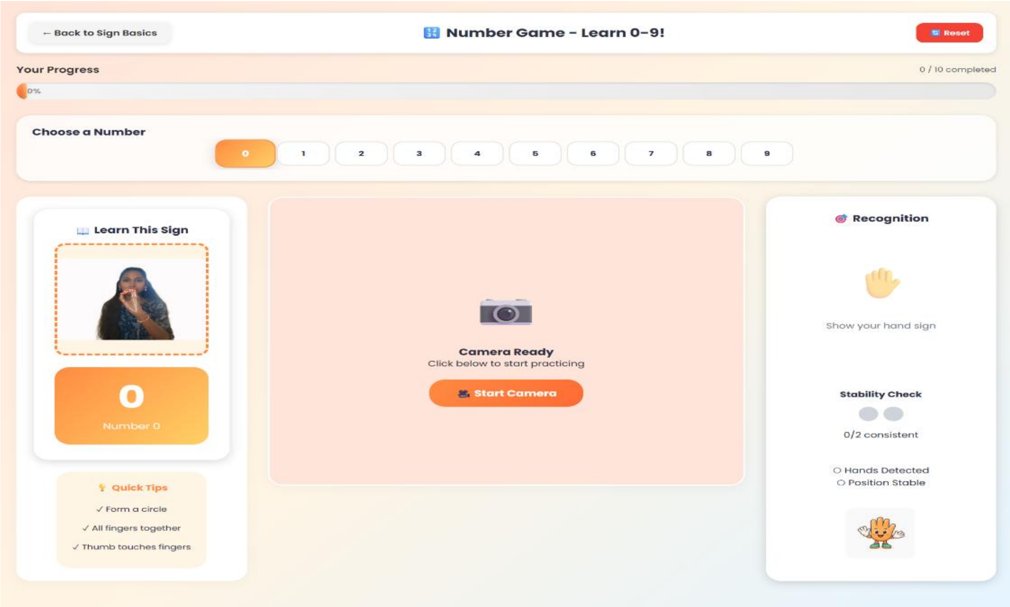


Figure 5.4.2: shows the Number Lesson page.

After optimization, number recognition achieved an accuracy of 95–99%, confirming the effectiveness of the static gesture models for beginner-level learning.

5.5 Intermediate Stage – Word Wonderland

The Intermediate Stage, known as Word Wonderland, introduces learners to a broader range of ISL vocabulary. This stage includes multiple sections such as A–Z vocabulary words, days of the week, colours, and general words. Motion-based A–Z vocabulary words are recognized using LSTM-based sequence models, where each gesture is analyzed over 30 frames to capture movement patterns.

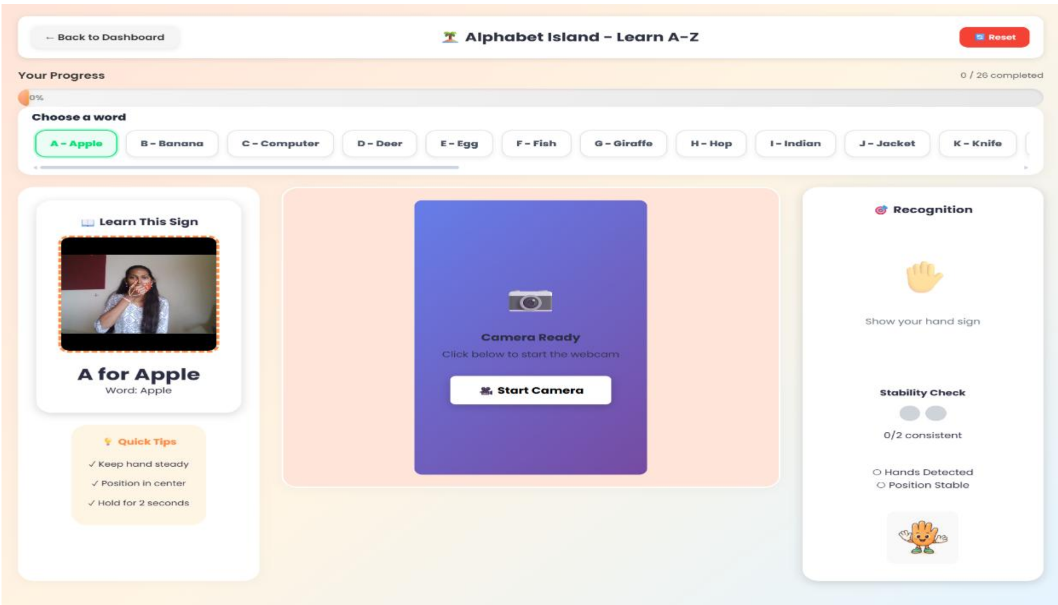
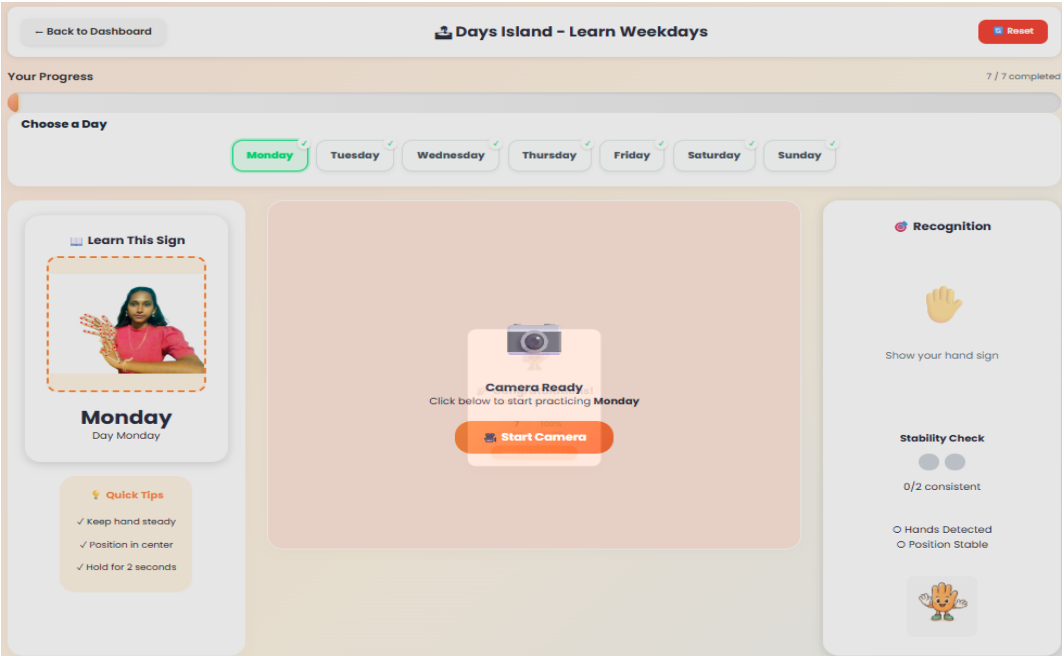
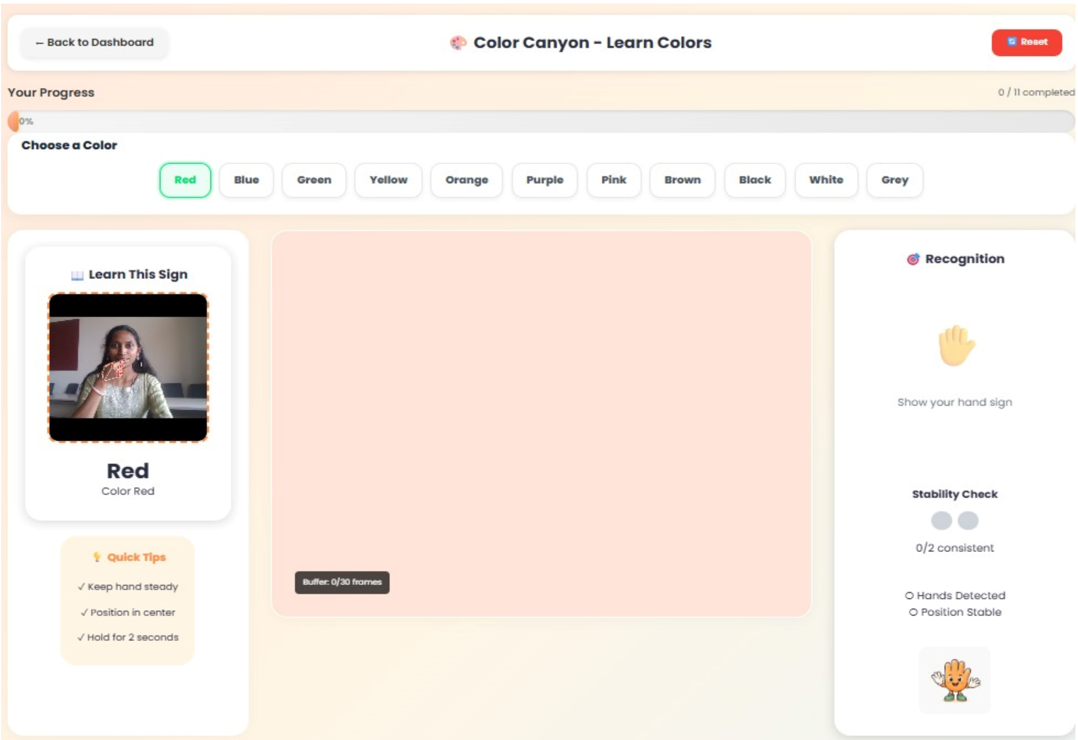


Figure 5.5.1: shows the A–Z Words lesson page.

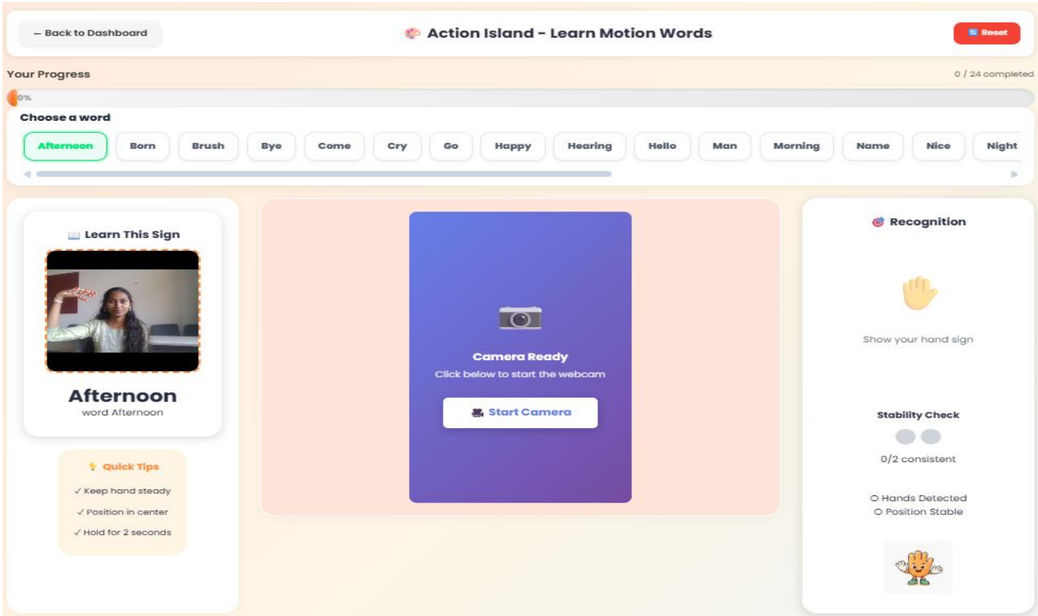
The Days and Colours sections further expand vocabulary learning, with seven gestures for days and eleven gestures for colours. These gestures are also recognized using motion-based LSTM models.

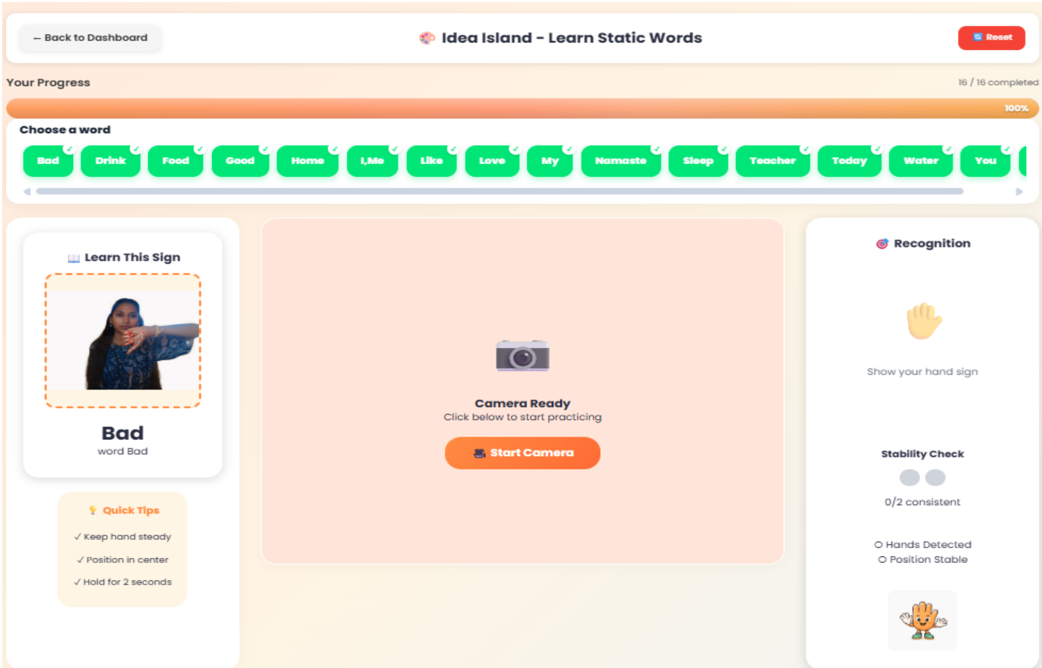




Figures 5.5.2 and 5.5.3: Days and Colours lessons.

In addition, the General Words section includes both motion-based words and static words, with separate models used for each type to ensure accurate recognition.





Figures 5.5.4 and 5.5.5: Motion Words and Static Words lesson pages.

In this word-level recognition achieved an accuracy of 95-98%, and stability mechanisms significantly reduced prediction flickering during live use.

5.6 Sentence Stage – Sentence Safari

The Sentence Stage, referred to as Sentence Safari, represents the advanced level of learning in EduSign. This stage focuses on recognizing short, predefined ISL sentences that involve coordinated hand movements, facial expressions, and body posture. The system uses an LSTM-based sequence model with a 60-frame input to capture longer and more complex gesture patterns. Holistic landmark features are incorporated to provide better contextual understanding.

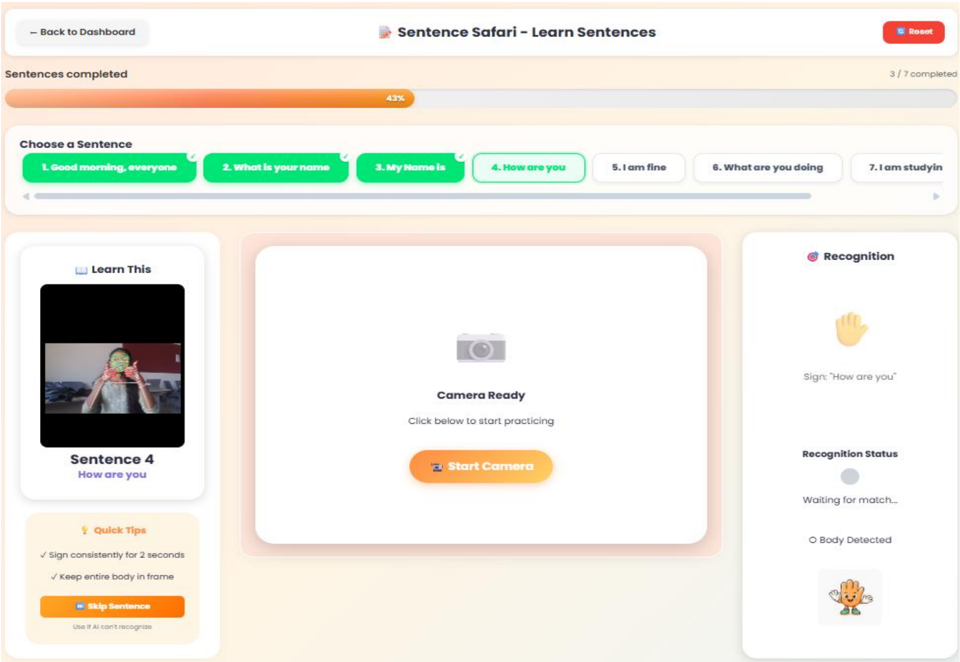


Figure 5.12: shows the Sentence Lesson page.

The results demonstrate that sentence recognition works reliably for predefined sentences, highlighting the feasibility of sentence-level ISL learning in a real-time environment.

VI. CHALLENGES FACED

The development of the EduSign platform presented several technical and design-related challenges. Although the core idea of creating a user-friendly ISL learning system was clear, implementing it in a real-time, accurate, and engaging manner required addressing multiple practical difficulties. The major challenges encountered and the corresponding solutions are discussed below.

1. **Prediction Flickering and Model Stability**

One of the earliest challenges was unstable predictions during real-time gesture recognition. Small hand movements or finger jitters caused the system to rapidly switch between predictions, which was confusing for users. To solve this, a stability buffer was introduced. Instead of showing every frame-level prediction, the system confirms a sign only when the same result remains consistent across multiple frames. This greatly improved stability while maintaining responsiveness.

2. **Confusion Between Visually Similar Signs**

Visually similar signs, such as the number “2” and the alphabet “V”, often caused misclassification due to identical finger positions. To reduce this confusion, separate models were trained for alphabets and numbers, and the datasets were refined. These changes significantly improved recognition accuracy and reduced user frustration.

3. **Lighting and Environment Dependence:**

Since EduSign uses a standard webcam, performance was affected by lighting conditions. In low light, MediaPipe occasionally struggled to detect hand landmarks. While lighting could not be controlled directly, visual indicators were added to alert users when detection quality was poor, helping them adjust lighting or camera position.

4. **Real-Time Performance vs Accuracy:**

Dynamic and sentence-level gesture recognition required higher computational power, which initially caused lag on the client side. To address this, a client–server architecture was adopted. Model inference was handled by a Flask backend, while real-time communication was optimized using Socket.IO. This ensured smooth interaction without sacrificing accuracy.

5. **Balancing Learning and Engagement:**

   Early versions of the interface were technically sound but lacked appeal for children. To improve engagement, gamification elements such as animated mascots, badges, rewards, and colorful stages were added. These features made learning more enjoyable and increased motivation.

Overall, addressing these challenges helped shape EduSign into a stable, responsive, and engaging ISL learning platform. Each improvement enhanced both system performance and the learning experience for users.

VII. CONCLUSION

The EduSign project successfully demonstrates the design and development of a real-time, interactive Indian Sign Language (ISL) learning platform tailored specifically for users. By integrating computer vision, deep learning, and modern web technologies, EduSign provides an engaging and practical environment where learners can actively practice sign language and receive instant feedback on their performance.

The system effectively recognizes static ISL gestures such as alphabets and numbers using deep learning based models and extends this capability to motion-based word recognition through sequence learning techniques. The use of MediaPipe for landmark extraction ensures accurate and lightweight feature representation, making the platform suitable for real-time usage on standard devices without specialized hardware.

A key strength of EduSign lies in its modular architecture. Separate recognition services for alphabets, numbers, words, and sentences allow the system to scale easily and maintain stable performance. Real-time communication through Socket.IO enables seamless interaction between the frontend and backend, ensuring smooth feedback during live gesture practice.

In addition to technical accuracy, the project emphasizes usability and engagement. The inclusion of a gamified learning structure with stages, badges, progress tracking, and visual celebrations transforms ISL learning into an enjoyable experience for users. This balance between technical robustness and user-friendly design makes EduSign more effective than traditional static learning methods.

Overall, EduSign achieves its objective of providing an accessible, interactive, and educational ISL learning solution. The project lays a strong foundation for future enhancements such as advanced sentence recognition, personalized learning analytics, and broader vocabulary expansion. As a semester-end project, EduSign reflects a successful application of machine learning and software engineering concepts to solve a meaningful real-world problem.

VIII. FUTURE SCOPE

EduSign already provides a strong and practical foundation for helping user learn Indian Sign Language (ISL) in real time. However, there are many opportunities to further enhance the platform and make it even more effective, engaging, and scalable.

From a technical point of view, future versions of EduSign could explore transformer-based deep learning models, especially for sentence-level recognition. Transformers are well suited for understanding long and complex gesture sequences, as they can capture relationships across an entire sentence more effectively. Combining different models such as DNNs, LSTMs, attention mechanisms, and transformers could further improve accuracy and make the system more robust to different signing styles.

Sentence-level learning can also be improved by adding context-aware and grammar-sensitive models, allowing the system to better understand natural variations in signing. Another useful enhancement would be step-by-step sentence construction, where each word is verified before forming a complete sentence. This approach can help users to learn more confidently and reduce mistakes.

To make learning more enjoyable, EduSign can introduce additional gamified features such as sign-based quizzes, gesture challenges, interactive stories, and reward-based progression. More expressive mascot interactions, including voice feedback and personalized encouragement, could further motivate young learners.

The platform could also benefit from adaptive learning techniques that automatically adjust lesson difficulty based on a user’s performance. Expanding the dataset to include more users, different lighting conditions, and varied signing speeds would improve real-world performance and reliability.

In the future, EduSign could be extended to support mobile devices, offline learning using lightweight models, and advanced technologies such as Augmented Reality (AR) and Virtual Reality (VR) to create immersive learning experiences. With these enhancements, EduSign has the potential to grow into a comprehensive, intelligent, and highly engaging ISL learning ecosystem for everyone.

IX. REFERENCES

[1] Google. MediaPipe: A Framework for Building Perception Pipelines. Available: https://developers.google.com/mediapipe

[2] Benjamin Batte, University of the Cumberlands (formerly Cumberland College), Date Written: April 25, 2025. “AI-Powered Sign Language Translation System: A Deep Learning Approach to Enhancing Inclusive Communication and Accessibility in Low-Resource Contexts.”

[3] S. Sudeshna, B. Sai Nikhitha, B. Vinaya, D. Dhanasree, D. Chandrika. Speak-to-Sign: “A Real-Time NLP-Driven Framework for Indian Sign Language.”

[4] Conference paper First Online: 20 December 2024. “AI-Enhanced Sign Language Interpreter.”

[3] Goodfellow, I., Bengio, Y., & Courville, A. (2016).” Deep Learning. MIT Press, Cambridge, MA.”

[4] Mittal, A., Kumar, P., & Roy, P. (2020). “Hand Gesture Recognition Using Convolutional Neural Networks,” International Journal of Computer Vision and Image Processing, vol. 10, no. 2.

[5] Koller, O. (2020). “Quantitative Survey of the State of the Art in Sign Language Recognition,” Computer Vision and Image Understanding, Elsevier.

[6] Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2015). “Sign Language Recognition Using Convolutional Neural Networks,” Proceedings of the European Conference on Computer Vision (ECCV) Workshops.

[7] Hochreiter, S., & Schmidhuber, J. (1997). “Long Short-Term Memory,” Neural Computation, MIT Press, vol. 9, no. 8, pp. 1735–1780.

[8] Zhang, J., Zhou, W., Xie, C., Pu, J., & Li, H. (2019). “Chinese Sign Language Recognition with LSTM Networks,” IEEE Transactions on Multimedia, vol. 21, no. 2, pp. 341–353.

[9] Warden, P. (2018). TinyML: Machine Learning with TensorFlow Lite. O’Reilly Media.

[10] Starner, T., Weaver, J., & Pentland, A. (1998). “Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371–1375.

[11] Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020). “Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).