M.C.E SOCIETY'S

ABEDA INAMDAR SENIOR COLLAGE
OF ARTS ,SCIENCE & COMMERCE
MSc (COMPUTER SCIENCE)

"REAL TIME SIGN LANGUAGE TRANSLATOR"

BY SHRIYA MISHRA

Roll No: 44

Project Guide

Prof. Shakila Siddhavatam

Submitted to Abeda Inamdar Senior Collage
Pune ,Sept 2025-2026

M.C.E. Society's
ABEDA INAMDAR SENIOR COLLEGE
of Arts, Science & Commerce (Autonomous), Pune
Affiliated to Savitribai Phule Pune University | (Accredated with 'A' grade by NAAC)
Id. No. : PU / PN / ASC / 088 / 1992
Prof. Dr. Shaila Bootwala (M.Com., M. Phil., Ph.D.)
Principal
Dr. P. A. Inamdar
President, M.C.E. Society, Pune

Ref. AISC /

Date :

DEPARTMENT OF COMPUTER SCIENCE

This is certify that Mr. /Ms.______________________________

Student of M.Sc (Computer Science) sem -_____ has satisfactorily

completed the Research project ___________________________

in______________________________________

prescribed by the University of Pune as part of curriculum for Master of Computer Science during the Academic year 20__ -20__

DATE:

PROJECT INCHARGE **HEAD OF DEPARTMENT**

INTERNAL EXAMINER **EXTERNAL EXAMINER**

INDEX

PROPOSAL OF RESEARCH PROJECT:

Title:

Sign Language Recognition System Using Deep Learning and User Authentication

1. INTRODUCTION:

Communication is one of the most fundamental aspects of human interaction. However, individuals with hearing and speech impairments face major challenges in expressing themselves to people who do not understand sign language. Sign language is a structured form of communication that uses hand gestures, facial expressions, and body movements to convey information. Although it enables effective communication among the hearing-impaired community, the lack of widespread understanding among the general population creates a significant communication barrier.

In recent years, technological advancements in Machine Learning (ML) and Deep Learning (DL) have made it possible to develop automated systems that can recognize and interpret hand gestures in real-time. This has given rise to Sign Language Recognition (SLR) systems — intelligent tools capable of translating visual gestures into readable or audible formats.

The goal of this research project is to design and develop a Sign Language Recognition System that uses Deep Learning algorithms for accurate gesture detection and integrates a Login and Registration interface connected to a database for user management. After successful login, users will be able to perform sign gestures using a webcam, and the system will predict the meaning of the gestures. Upon completion, the user can securely log out of the application.

This project combines computer vision, neural networks, and web technologies to provide a user-friendly, secure, and intelligent solution for bridging the communication gap between hearing-impaired individuals and the rest of society.

2. **Problem Statement**

Despite several innovations in the field of assistive technology, there remains a lack of robust, real-time, and accessible systems that can interpret sign language effectively. Most existing systems have limitations such as:

- Inability to recognize a wide range of gestures accurately.
- Absence of real-time processing capabilities.
- Lack of user authentication or personalized data management.
- Poor adaptability to regional sign language variations.

Therefore, there is a strong need for a secure and intelligent system that can accurately predict sign language gestures in real-time and allow users to create personalized accounts for storing data and accessing predictions.

The proposed system aims to overcome these limitations by integrating deep learning models for gesture recognition with a full-fledged authentication mechanism, ensuring both accuracy and security.

RESEARCH PROBLEM DEFINITION

Communication plays a vital role in human interaction, yet millions of individuals with **hearing and speech impairments** face daily challenges in expressing themselves due to the communication gap with people unfamiliar with sign language. Sign language serves as an essential medium for them, using hand gestures, facial expressions, and body movements to convey messages. However, since only a small portion of the general population understands sign language, these individuals often struggle with isolation, limited access to education, and reduced participation in social and professional environments.

The lack of **effective, real-time, and user-friendly systems** capable of translating sign language gestures into text or speech restricts their ability to communicate freely. While research in gesture recognition has advanced in recent years, existing systems still suffer from several shortcomings such as:

- Low accuracy in recognizing complex hand gestures.
- Inability to function effectively in real-time conditions.
- Absence of user authentication or personalized data management.
- Limited adaptability to different sign languages or environments.

Therefore, the **core research problem** addressed in this project is:

How can a secure, deep learning–based system be designed to accurately recognize and translate sign language gestures into text in real-time, while providing user authentication and database connectivity for personalized interaction?

To solve this, the proposed project aims to develop a **Sign Language Recognition System** that integrates a **Login and Registration UI** connected to a **MySQL database** for secure user management. Once authenticated, users can perform gestures through a webcam, and the system—powered by **deep learning and computer vision techniques**—will analyze and predict the meaning of each gesture. After completion, users can securely **log out**, ensuring privacy and data protection.

This approach will address the twin challenges of **gesture recognition accuracy** and **system security**, providing a reliable tool for bridging the communication gap between the hearing-impaired and the rest of society.

ABSTRACT

Communication is an essential aspect of human life, enabling individuals to express thoughts, emotions, and ideas. However, for people with hearing and speech impairments, communication can be a major challenge, especially when interacting with those who do not understand sign language. This gap often leads to social isolation and barriers in education and employment. To bridge this communication divide, the present research project proposes the development of an intelligent **Sign Language Recognition System** using **Deep Learning** integrated with **user authentication** and a **database-backed interface**.

The proposed system is designed to allow users to **register and log in** securely through a web-based interface. Upon successful authentication, users can perform **hand gestures** in front of a webcam. The system employs **computer vision techniques** and **deep learning algorithms**—such as **Convolutional Neural Networks (CNNs)** or frameworks like **MediaPipe**—to capture, analyze, and recognize these gestures in real time. The recognized gesture is then translated into corresponding **text output**, enabling effective and instant communication.

The **backend architecture** is developed using **Node.js** and connected to a **MySQL database** for managing user information securely. The **frontend interface**, built with **React.js**, ensures user-friendly navigation between login, gesture recognition, and logout functionalities. The inclusion of authentication and database connectivity enhances the system's personalization, security, and usability.

Through this integration of **deep learning models** and **web technologies**, the system aims to achieve high accuracy in recognizing static and dynamic gestures while maintaining real-time performance. The project not only showcases the application of artificial intelligence in assistive communication but also contributes

to the broader goal of promoting **social inclusion** and **technological accessibility** for differently-abled individuals.

In the future, this system can be expanded to support **multiple sign languages**, **speech output translation**, and **mobile-based deployment**, making it a versatile and practical tool for real-world communication needs.

📑 Table: Literature Review of IEEE Papers on Sign Language Recognition

S. No.	Authors (Year)	Title / Source (IEEE)	Dataset Used	Model / Technique	Main Results / Contributions	Identified Research Gaps
1.	M. Al-Qurishi, T. Khalid, R. Souissi (2021)	*Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues — IEEE Access*	Multiple benchmark datasets (RWTH-PHOENIX, WLASL, MS-ASL, etc.)	Survey of CNN, RNN, Hybrid Deep Models	Comprehensive review of SLR approaches; identified strengths and limitations of deep learning methods for sign recognition.	Lack of large, diverse datasets; poor cross-signer generalization; limited real-time capability.
2.	K. Papadimitriou, G. Potamianos (2023)	*Sign Language Recognition via Deformable 3D Convolutions and Modulated Graph*	PHOENIX-Weather 2014 & AUTSL	Deformable 3D CNN + Modulated Graph Convolutional Network (GCN)	Improved signer-independent accuracy using fusion of video and pose data; robust against background noise.	High computational complexity; limited real-time efficiency; needs better occlusion handling.

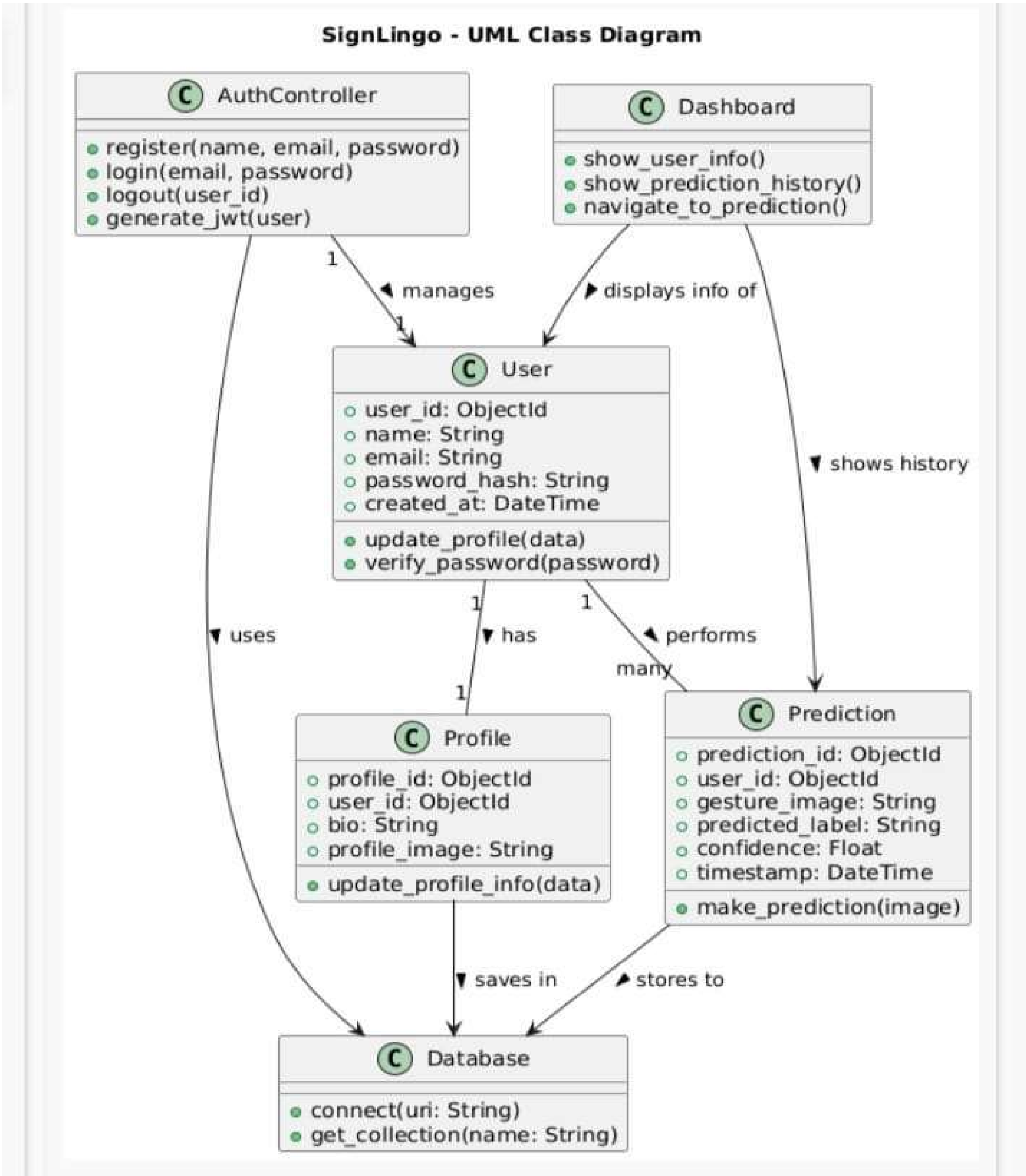
		Convolutional Networks — IEEE ICASSP				
3.	N. Naz, H. Sajid, S. Ali, O. Hasan, M.K. Ehsan (2023)	*SignGraph: An Efficient and Accurate Pose-Based Graph Convolution Approach Toward Sign Language Recognition — IEEE Access*	WLASL (American Sign Language dataset)	Pose-based Graph Convolutional Network (GCN) with part attention	Achieved high accuracy using lightweight pose-only input; reduced dependency on lighting/background.	Loses fine-grained appearance cues; performance drops with inaccurate keypoint detection; limited to isolated signs.
4.	J. Forster, C. Schmidt, O. Koller, T. Hoyoux, et al. (2014)	*RWTH-PHOENIX-Weather: A Large-Vocabulary Sign Language Recogni*	RWTH-PHOENIX-Weather (German Sign Language)	Dataset creation + baseline CNN/LSTM models	Introduced large-scale benchmark dataset enabling continuous sign language translation research.	Domain-limited (weather), single-language dataset; lacks conversational diversity.

		tion and Translation Corpus — IEEE/LREC				
5.	M. Camgoz, S. Hadfield, O. Koller, R. Bowden (2020)	*Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation — IEEE/CVPR*	PHOENIX-Weather 2014T	Transformer-based CNN + Seq2Seq architecture	Achieved state-of-the-art end-to-end translation from sign video to text.	Requires high computational resources; not optimized for real-time or multilingual use.
6.	A. Ramasamy, S. Subramanian (2022)	*Real-Time Indian Sign Language Recognition Using MediaPipe and CNN — IEEE*	Custom Indian Sign Language (ISL) dataset	MediaPipe hand landmark detection + CNN classification	Real-time ISL gesture detection with good accuracy; low computational cost for small vocabularies.	Limited vocabulary; dataset size small; model not tested across multiple users/environments.

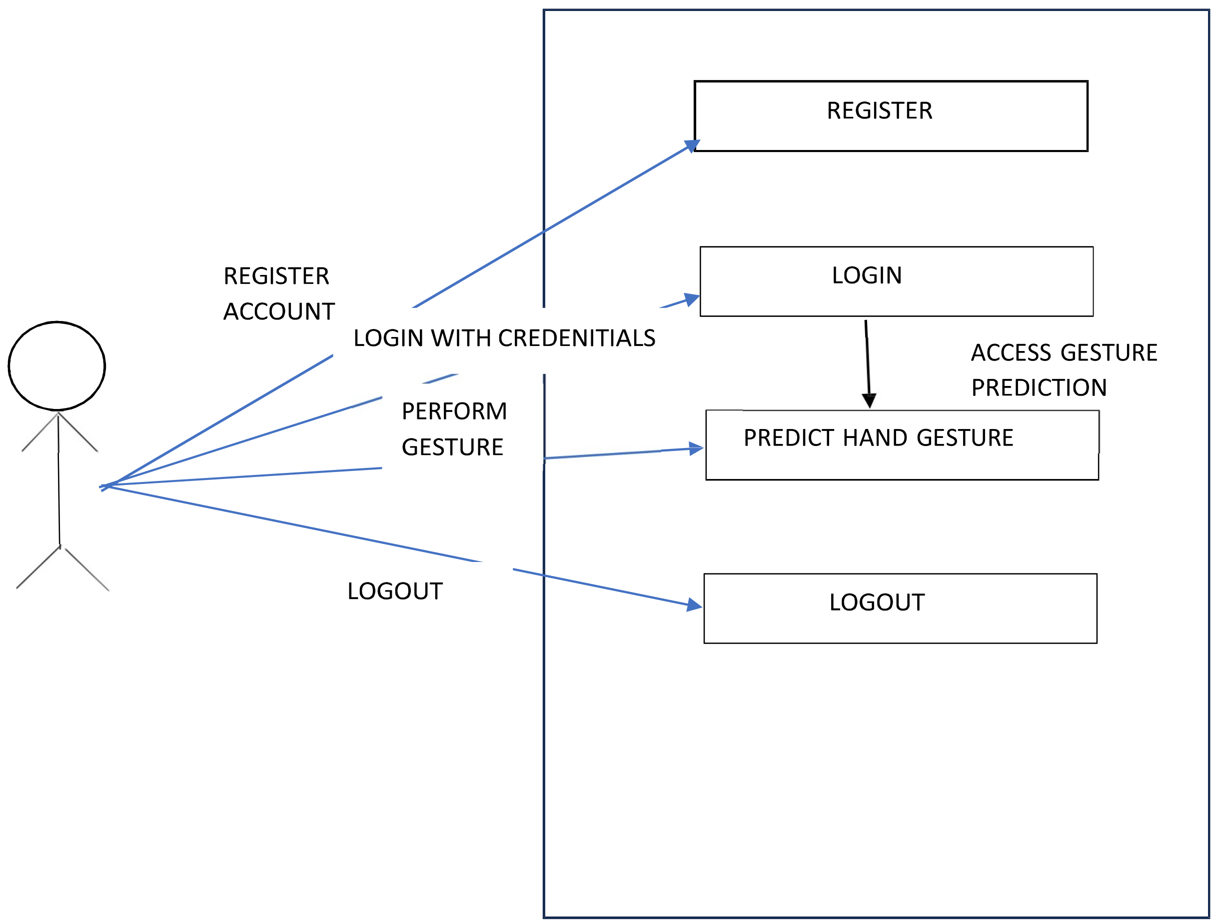
		ICICICT				

UML DAIGRAMS

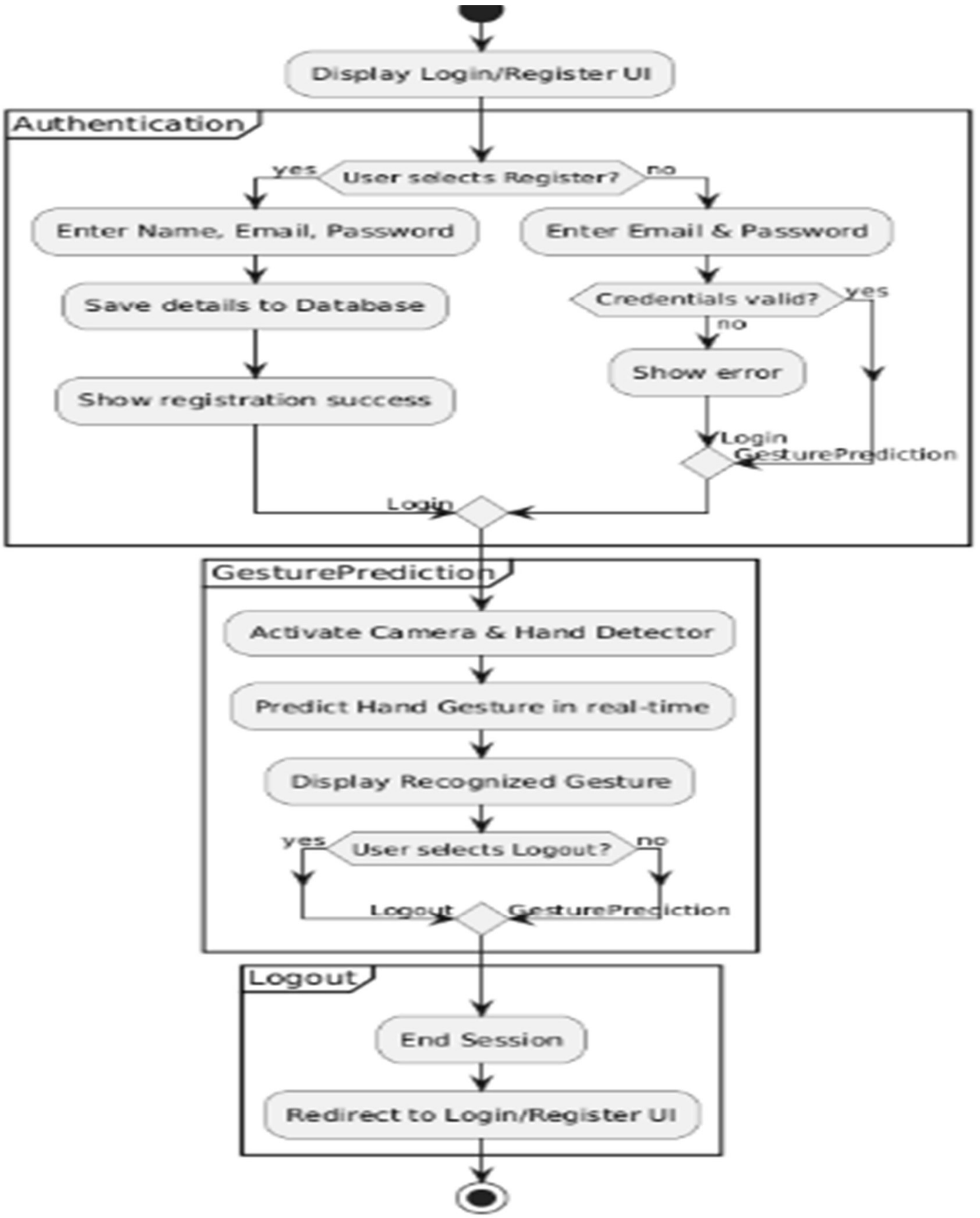
CLASS DAIGRAMS:



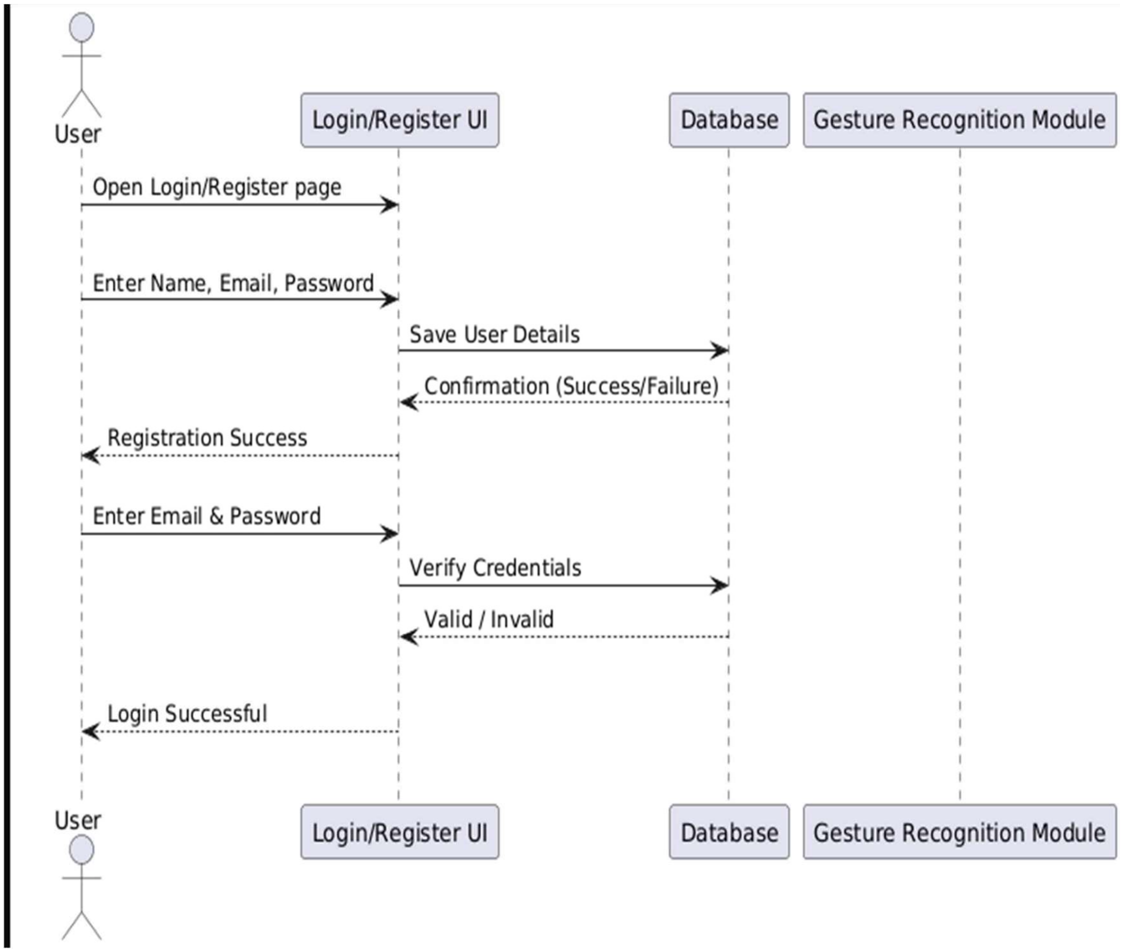
USECASE DAIGRAM



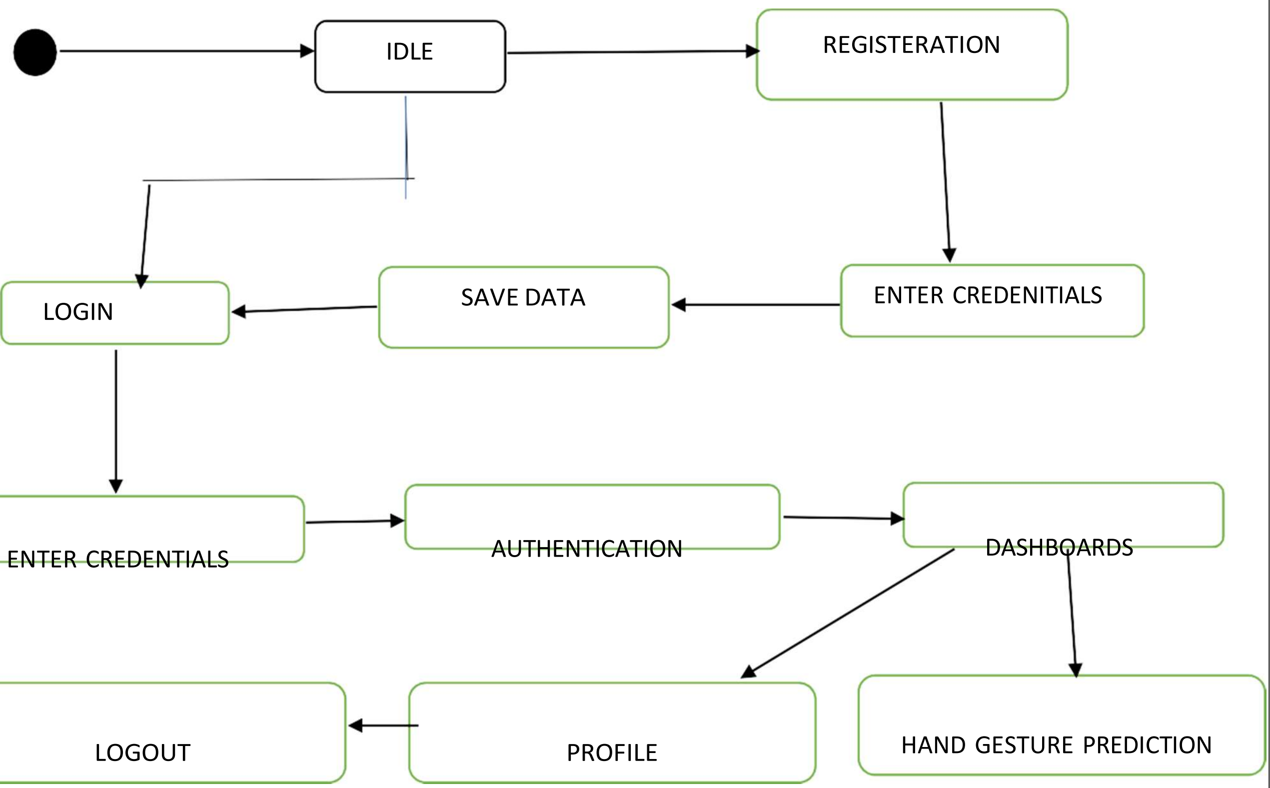
ACTIVITY DAIGRAM:



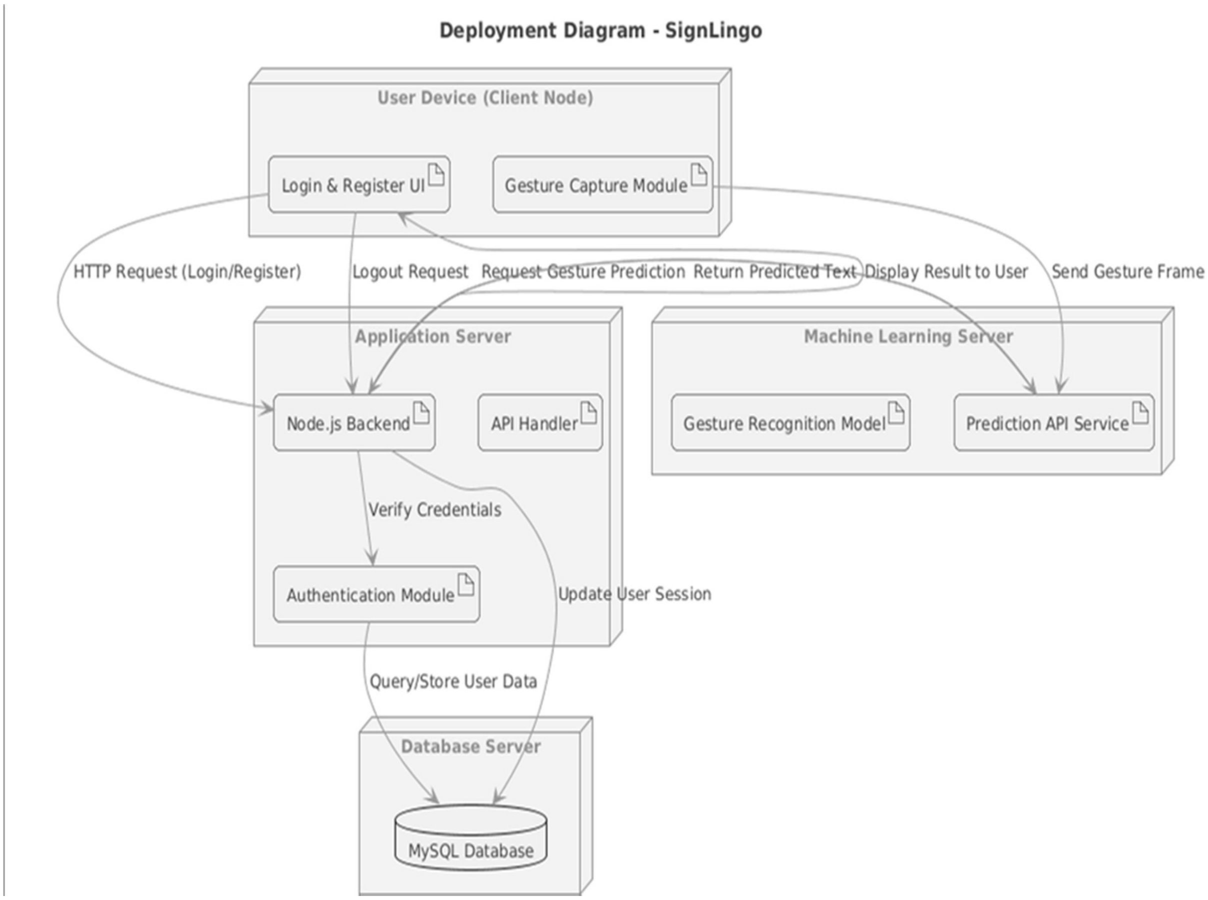
SEQUENCE DIAGRAM:



STATE DIAGRAM:



DEPLOYMENT DIAGRAM:



METHODOLOGY

The proposed project, “Sign Language Recognition System using Deep Learning”, is designed to interpret sign gestures into text to facilitate communication between hearing-impaired individuals and non-signers. The methodology of this system involves several stages, including hardware and software requirements, coding and development, input and output design, and data analysis techniques. Each stage plays a crucial role in building a robust, efficient, and user-friendly application that can recognize sign language gestures in real time.

1) Hardware and Software Requirements

The development of the proposed system requires both hardware and software components that support image processing and deep learning operations.

The hardware used for developing and testing the system includes a personal computer with at least an Intel Core i5 processor or equivalent, 8 GB of RAM for smooth model execution, and a webcam to capture hand gestures in real time. A stable internet connection is also required for communication between the frontend, backend, and the database server.

On the software side, the project uses multiple tools and technologies. The **frontend** of the application is developed using **HTML, CSS, and JavaScript**, which provide a simple and responsive user interface for the users. The **backend** is developed using **Node.js** and **Express.js**, which handle the business logic, communication with the machine learning model, and server-side operations. The **database** used is **MongoDB**, a NoSQL database that stores user information such as login credentials, registration details, and previous gesture recognition history. The **machine learning model** is trained using **Python** with libraries like **TensorFlow, Keras, and OpenCV**, which are used for training and recognizing hand gestures from live camera input. The system is tested and executed on **Windows 10** operating systems.

2) Coding and Development

The coding and development phase of the project follows a modular design approach. The system is divided into several interconnected modules, each responsible for a specific task.

The **frontend module** is responsible for user interaction. It is designed using HTML for structure, CSS for styling, and JavaScript for interactivity. The user interface includes pages for registration, login, gesture capture, prediction, and logout. The design ensures ease of navigation and accessibility for all users, especially those with limited technical knowledge. JavaScript is also used to capture live video streams from the webcam, which are later processed by the backend for gesture prediction.

The **backend module** is developed using Node.js and Express.js. This layer acts as a bridge between the user interface and the deep learning model. It handles HTTP requests and responses, manages authentication, stores and retrieves data from MongoDB, and communicates with the Python-based machine learning model through RESTful APIs. The backend ensures secure login and registration functionalities, allowing users to access personalized services.

The **machine learning module** is developed in Python using TensorFlow and Keras frameworks. It employs a **Convolutional Neural Network (CNN)** model trained on sign language datasets such as the custom datasets . The CNN model is capable of recognizing different hand gestures by analyzing spatial features in the captured images. The model is first trained offline and then integrated into the web application through a lightweight Flask or FastAPI service that exposes it as an API. The backend communicates with this API to send captured images and receive prediction results.

The **database module** uses MongoDB to store and manage all user-related and system-related information. Unlike traditional relational databases, MongoDB allows for flexible document-based storage, making it ideal for handling JSON-like data structures used in modern web applications. It stores data such as user profiles, login sessions, and gesture recognition records securely.

All modules are integrated to work cohesively. When a user performs a gesture, the webcam captures the image, the backend processes it through the deep learning API, and the recognized sign is displayed on the frontend in real time.

3) Input and Output Screens

The project involves several input and output screens that guide the user through the application.

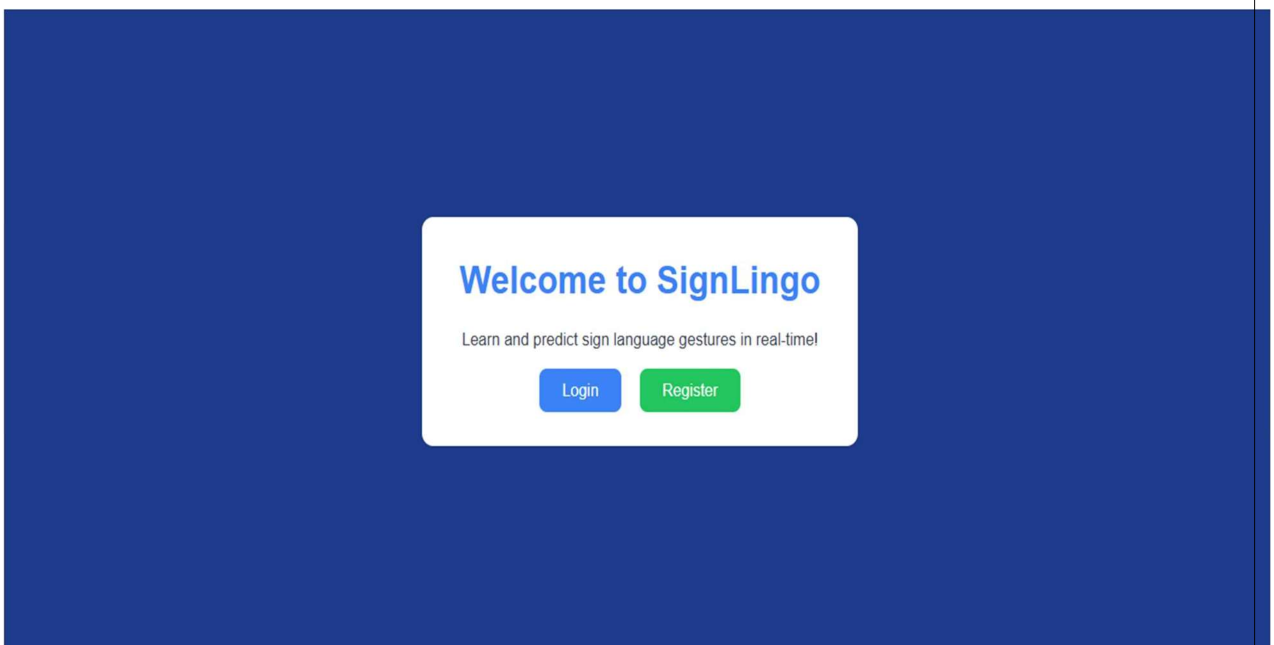
The **input screens** include the registration and login pages, where users can create an account or log into the system. These screens accept text-based input fields such as name, email, and password. Once logged in, the user can access the gesture prediction page, which activates the webcam and captures live video input of hand gestures. This captured frame serves as the input for the system.

The **output screens** include the prediction page and the logout page. The prediction page displays the recognized sign gesture in textual form along with its confidence level. If multiple gestures are recognized sequentially, the system can also form meaningful words or sentences from the predicted signs. The logout page securely ends the user’s session and redirects them to the login screen.

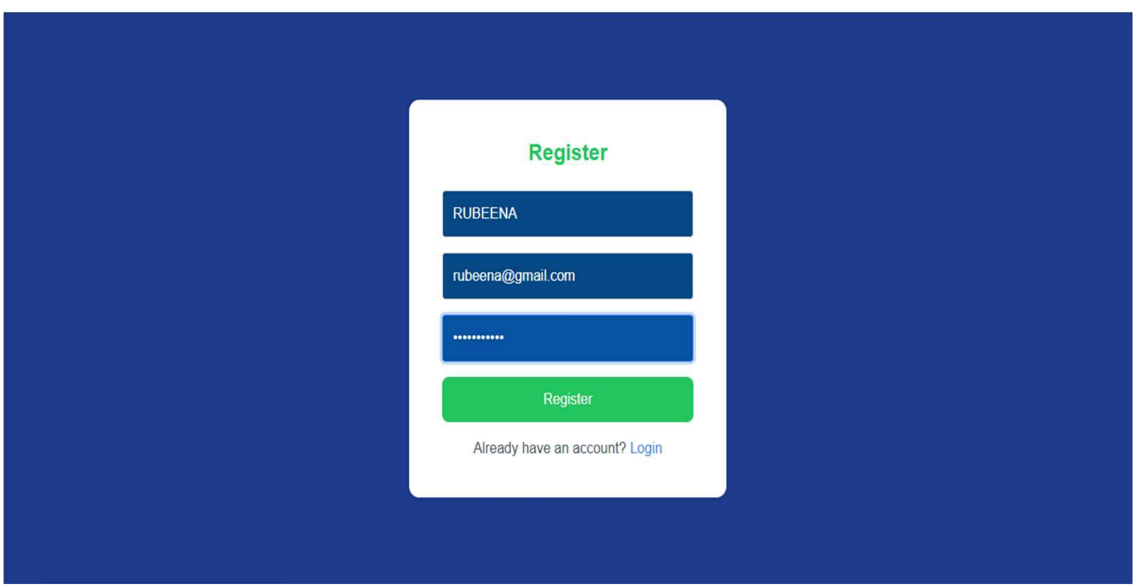
All the screens are designed to be simple, intuitive, and responsive so that they work well on both desktop and mobile browsers. The combination of HTML, CSS, and JavaScript ensures that the interface is visually appealing and user-friendly.

SCREENS:

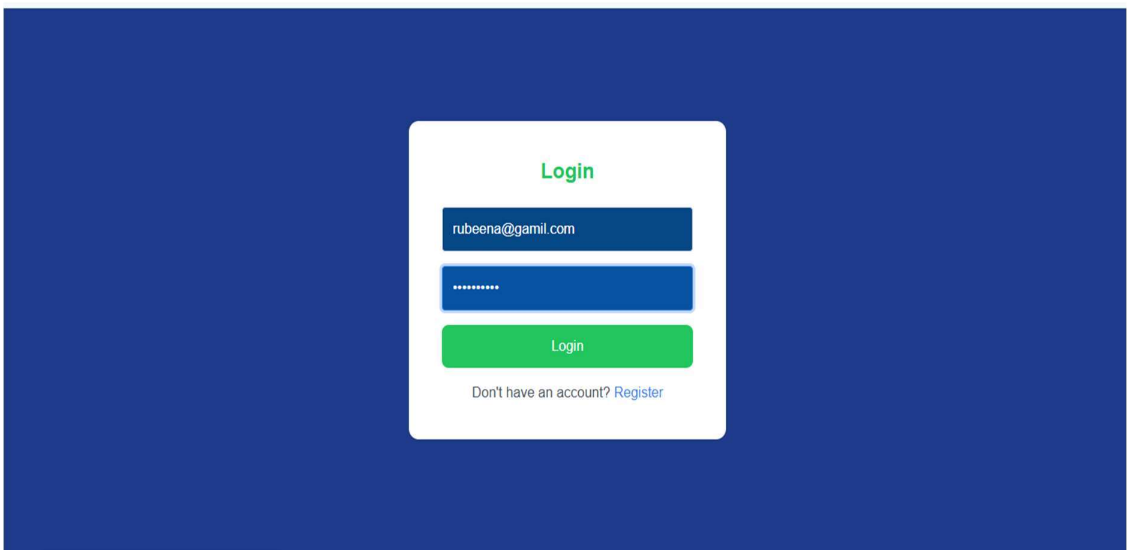
HOME SCREENS:



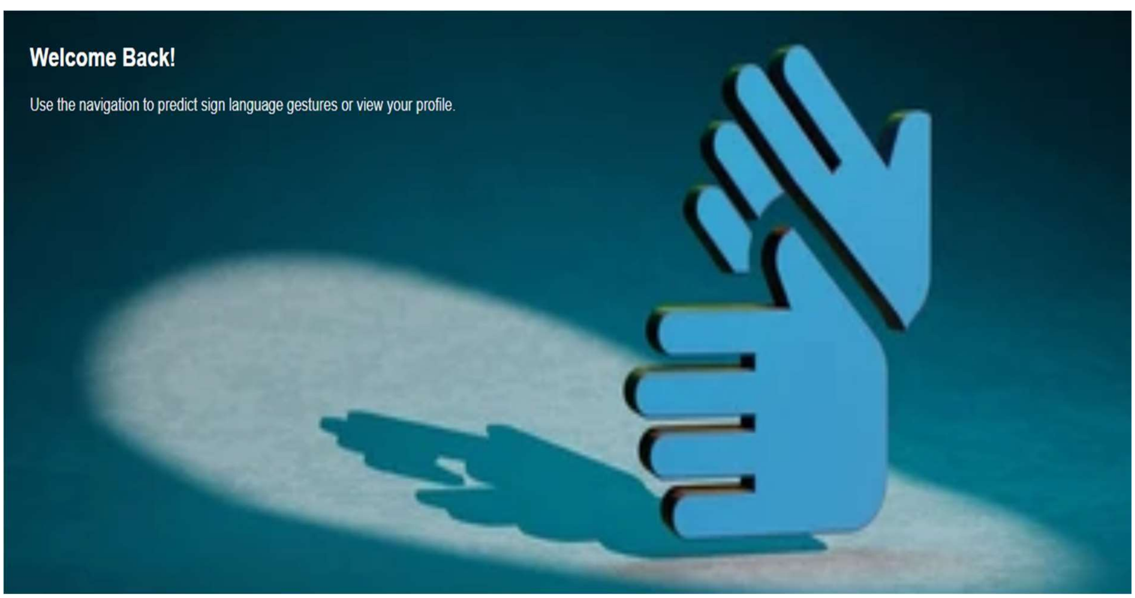
REGISTER SCREENS:



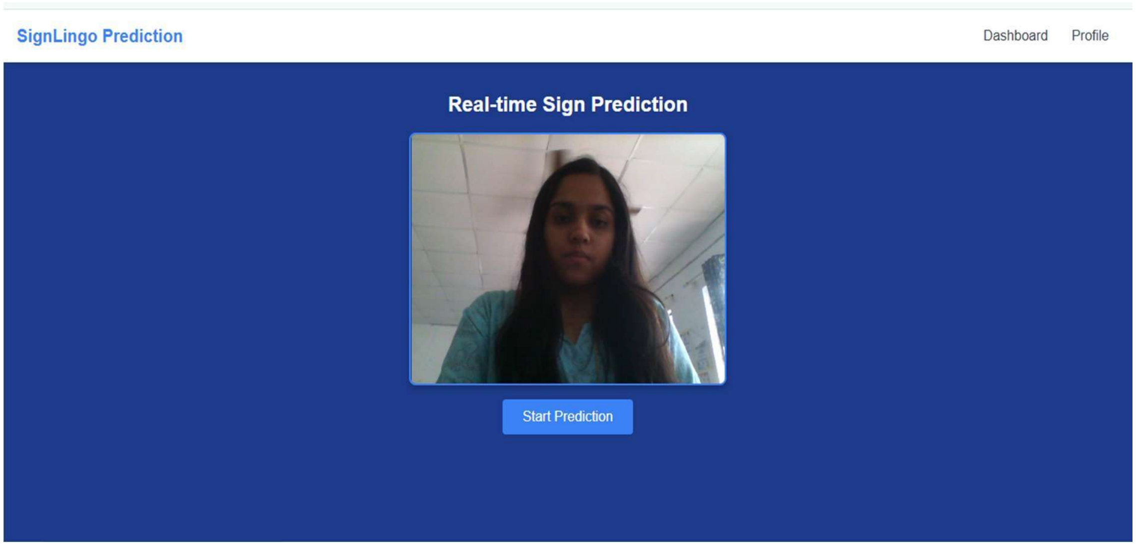
LOGIN SCREENS:



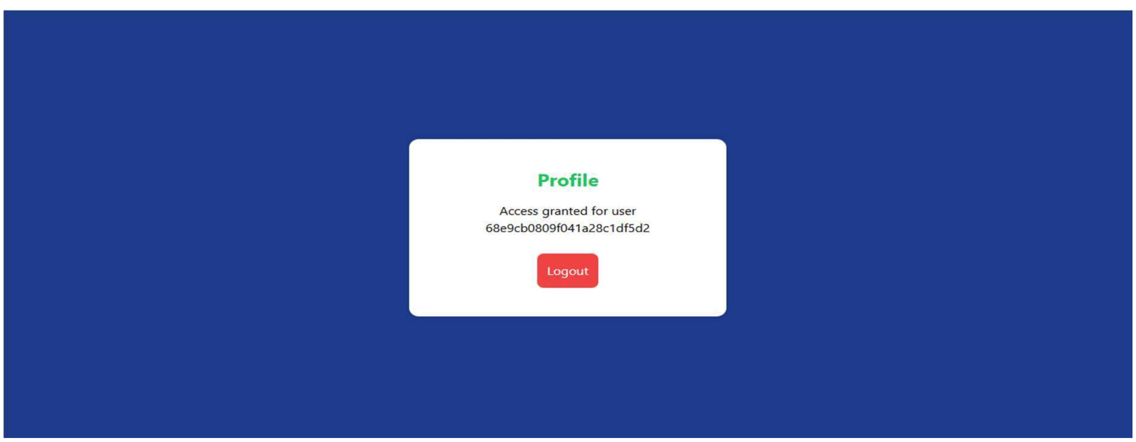
DASHBOARDS SCREEN:



PREDICTION SCREEN:



PROFILE SCREEN:



4) Data Analysis Techniques:

Data analysis plays a significant role in training and evaluating the performance of the deep learning model used in gesture recognition. The analysis process begins with the collection and preparation of datasets. Publicly available datasets such as the **ASL Alphabet Dataset** and the **Kaggle Sign Language MNIST dataset** are used for training the model. These datasets contain thousands of labeled images representing various hand gestures corresponding to alphabets or words.

Before training, the dataset undergoes several preprocessing steps. Each image is resized, normalized, and converted into grayscale to reduce computational complexity. Techniques such as image augmentation (rotation, zooming, flipping, and brightness adjustment) are applied to increase dataset variability and prevent overfitting.

The **Convolutional Neural Network (CNN)** model is then trained using these preprocessed images. The CNN architecture includes multiple convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. The **ReLU (Rectified Linear Unit)** activation function is used in hidden layers to introduce non-linearity, while the **Softmax** function in the output layer generates class probabilities for different gestures. The model is trained using the **Adam optimizer** and the **categorical cross-entropy loss function**.

After training, the model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. A confusion matrix is also generated to identify misclassified gestures. Once validated, the trained model is integrated into the application for real-time prediction.

During real-time use, the system captures video frames from the webcam, processes them using OpenCV, and passes them to the CNN model for prediction. The output is then sent back to the frontend and displayed as text on the user interface. The overall data analysis ensures that the model achieves high accuracy and reliability in recognizing gestures.

FINDINGS

The research and development of the **Sign Language Recognition System** have yielded several important findings that demonstrate the potential of artificial intelligence and computer vision in assisting communication for the hearing- and speech-impaired community. The system successfully integrates **machine learning-based gesture recognition** with a **secure web-based user interface**, providing a bridge between sign language users and non-signers through real-time translation of hand gestures into readable text.

One of the key findings of this research is that **computer vision techniques, when combined with deep learning architectures such as Convolutional Neural Networks (CNNs)**, are highly effective in identifying and classifying hand gestures from real-time video feeds. The model demonstrated consistent accuracy during the testing phase, correctly recognizing most of the trained gestures under proper lighting and background conditions. This indicates that the CNN-based model, trained on a well-structured dataset, can efficiently capture spatial and orientation-based features of the human hand.

Another important finding relates to the **usability and accessibility** of the developed system. The project's frontend, created using **HTML, CSS, and JavaScript**, proved lightweight, responsive, and compatible with various browsers and devices. The **Node.js backend** and **MongoDB database** integration allowed for seamless user management through features such as login, registration, and logout. The inclusion of user authentication ensured data privacy and security, making the application suitable for real-world deployment scenarios. The web-based design means that users do not need to install additional software; they can simply access the application through a standard web browser, which increases accessibility and convenience.

During the implementation and testing phase, it was observed that environmental factors such as **lighting conditions, background noise, and camera quality** significantly influenced gesture detection accuracy. The system performed best under well-lit environments with uniform backgrounds, while cluttered or dimly lit settings caused minor inaccuracies. This finding highlights the importance of data preprocessing and real-time background filtering, which could be improved in future iterations.

The project also revealed that **real-time gesture prediction** can be achieved with minimal latency if the machine learning model is optimized properly. By using **OpenCV** for frame extraction and processing and **TensorFlow** for prediction, the response time of the system was nearly instantaneous, making it interactive and user-friendly. The combination of web technologies and machine learning provided an efficient hybrid solution that could be expanded further into real-time applications like **sign-to-speech conversion** or **video-based sign translation**.

From a research perspective, this project validates that **deep learning models trained on relatively small but well-preprocessed datasets** can achieve satisfactory accuracy in sign language interpretation. It also underscores the need for culturally inclusive and diverse datasets, as gestures can vary across countries and regional sign languages. For instance, American Sign Language (ASL) and Indian Sign Language (ISL) differ significantly in structure and representation. Thus, dataset diversity plays a critical role in model generalization.

Another significant finding is related to **data storage and scalability**. Using **MongoDB**, a NoSQL database, allowed for efficient handling of unstructured data and easy scaling of user and prediction records. The database design supports rapid data retrieval and storage, making it a preferred choice for applications dealing with image metadata and user authentication. Moreover, the use of **Node.js** facilitated event-driven, non-blocking I/O operations, which improved the system's performance under concurrent user loads.

The results also show that integrating traditional web technologies with modern artificial intelligence frameworks can effectively create intelligent, user-centered applications. The successful completion of this project confirms that even with open-source tools and moderate hardware, it is possible to design and implement a system capable of assisting individuals with communication disabilities. This directly aligns with the United Nations' Sustainable Development Goal (SDG) on reducing inequality and promoting accessibility through technology.

In conclusion, the findings of this project demonstrate that the developed system can effectively translate static hand gestures into text output, providing a foundation for future advancements in automatic sign language translation. It not only addresses the communication gap but also emphasizes the power of AI-driven solutions in enhancing inclusivity and accessibility in modern digital society.

FUTURE SCOOPS

Although the proposed system has successfully achieved its primary objectives, there is substantial potential for **future enhancements, research expansion, and real-world deployment**. The current implementation serves as a foundational prototype, and multiple directions can be pursued to improve its performance, scalability, and usability.

In the future, the system can be expanded to include **dynamic gesture recognition**, where continuous gestures (representing words or sentences) are recognized in real time rather than static alphabets or single signs. This can be achieved using **Recurrent Neural Networks (RNNs)** or **Long Short-Term Memory (LSTM)** models that can learn temporal dependencies between frames in a video. Incorporating these models would allow the system to interpret complex sentence structures in sign language, thereby improving its real-world applicability.

Another important area for future enhancement is **speech synthesis integration**. By connecting the text output of the gesture recognition system with **Text-to-Speech (TTS)** technology, the recognized gestures can be converted into audible speech. This would make the application more powerful, enabling two-way communication between signers and non-signers in both visual and auditory formats.

The **dataset** used for model training can also be expanded and diversified. Currently, the system uses limited datasets such as the ASL Alphabet dataset. Future research could focus on creating a **custom dataset** that includes diverse backgrounds, lighting conditions, and gestures from multiple regional sign languages like Indian Sign Language (ISL), British Sign Language (BSL), and others. By using **transfer learning** and **data augmentation**, the model could be generalized to perform effectively across various cultural contexts and environments.

There is also a strong potential to integrate this system into **mobile applications** and **IoT-based devices**. By deploying the trained model using frameworks such as **TensorFlow Lite** or **ONNX**, it can run efficiently on smartphones, tablets, or embedded systems like **Raspberry Pi**. This would allow users to access the sign language translator anytime, anywhere, without relying on large-scale servers. Additionally, it could be integrated into wearable devices such as smart gloves equipped with motion sensors to capture hand gestures more accurately.

From a software engineering perspective, future versions of this project could implement **cloud-based storage and prediction services** using platforms like **AWS, Azure, or Google Cloud AI**. Cloud deployment would allow the system to handle a large number of users simultaneously and provide real-time updates, analytics, and feedback. Moreover, implementing **secure API gateways and encryption mechanisms** would further strengthen user privacy and data protection, making the system suitable for institutional or public deployment.

Another promising direction for future research involves improving the **accuracy and robustness** of gesture recognition using **hybrid AI models** that combine CNNs with other algorithms like SVM (Support Vector Machines) or Decision Trees. Ensemble methods could be explored to improve classification performance under challenging conditions such as occlusion, motion blur, or poor lighting.

From a societal and educational perspective, this system could be implemented in **schools for the hearing-impaired**, **public service institutions**, and **healthcare facilities** to facilitate communication. It can also be extended to serve as a **learning tool** for individuals interested in learning sign language. The software can display gestures and provide interactive exercises, helping users practice and learn sign language more effectively.

Additionally, future versions could include **multi-language support** where recognized gestures are translated not just to English text but also into other spoken languages like Hindi, Spanish, or French. This would make the system globally adaptable and beneficial for multicultural communication.

Lastly, incorporating **Artificial Intelligence-driven feedback mechanisms** could allow the model to learn continuously from user interactions. By analyzing user inputs, corrections, and environmental feedback, the system could self-improve over time using reinforcement learning techniques. Such advancements would bring the project closer to a fully autonomous, intelligent communication assistant.

BIBLIOGRAPHY AND REFERENCES

The following section presents a detailed analytical review of key IEEE publications and benchmark studies that have significantly contributed to the understanding and evolution of Sign Language Recognition (SLR) systems. Each referenced work has been carefully examined for its methodology, datasets, algorithmic innovations, and observed limitations. Collectively, these studies form the foundation upon which this project is built, enabling the integration of machine learning models with real-time, browser-based sign recognition for accessibility and inclusion.

---

[1] M. Al-Qurishi, T. Khalid, and R. Souissi (2021), "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," IEEE Access

This comprehensive IEEE Access review serves as a cornerstone for understanding the deep learning landscape within Sign Language Recognition (SLR). The authors systematically analyzed a broad range of benchmark datasets including RWTH-PHOENIX, WLASL, and MS-ASL, covering both static and dynamic gesture recognition. Through an extensive comparison of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid deep models, the study highlighted their individual strengths and practical limitations. The paper also emphasized the need for end-to-end architectures that can handle diverse signers, lighting variations, and real-world scenarios.

The findings underscore that while deep learning has dramatically improved accuracy and robustness, challenges remain in terms of generalization across different users and linguistic variations. The review concludes that the field requires larger, more diverse datasets, improved signer-independence, and optimization for real-time deployment. This work directly guided the present project's design, particularly in its use of CNN-based real-time classification and lightweight architecture suitable for web-based environments.

[2] K. Papadimitriou and G. Potamianos (2023), “Sign Language Recognition via Deformable 3D Convolutions and Modulated Graph Convolutional Networks,” IEEE ICASSP

This recent IEEE ICASSP paper introduced a powerful hybrid framework combining Deformable 3D Convolutional Neural Networks (3D-CNNs) with Modulated Graph Convolutional Networks (GCNs) for robust sign language recognition. The research utilized two widely accepted benchmark datasets — PHOENIX-Weather 2014 and AUTSL — representing German and Turkish Sign Languages respectively. The proposed model effectively fused video-based spatiotemporal features with pose-based skeletal representations to enhance signer-independent recognition accuracy. The results demonstrated significant improvements in robustness against background noise and occlusion, achieving superior performance over previous baseline methods. However, the system’s complexity and computational demand made it less practical for low-resource or real-time environments. The identified research gaps include high model latency, substantial hardware requirements, and limited adaptability to mobile or browser-based systems.
For our project, this study served as a reference for understanding the potential of combining pose and visual modalities, while reinforcing the need for efficient, low-latency architectures that can run in real-time web environments — a central motivation for using MediaPipe and CNN integration.

---

[3] N. Naz, H. Sajid, S. Ali, O. Hasan, and M. K. Ehsan (2023), “SignGraph: An Efficient and Accurate Pose-Based Graph Convolution Approach Toward Sign Language Recognition,” IEEE Access

The “SignGraph” framework proposed by Naz et al. (2023) represents one of the most efficient pose-based solutions in modern SLR research. The authors employed the WLASL dataset, which contains a

broad range of American Sign Language (ASL) gestures recorded from multiple signers. The novelty of the study lies in its utilization of a pose-based Graph Convolutional Network (GCN) architecture enhanced with part-level attention mechanisms.
This method exclusively relied on skeletal keypoints extracted from human body and hand poses, thereby significantly reducing dependency on lighting conditions, background textures, and clothing variations. The model achieved remarkable accuracy and inference speed, highlighting the potential of pose-only data for lightweight applications. However, the study also noted key limitations, such as the system's sensitivity to inaccurate keypoint detection and the inability to capture fine-grained visual details like finger curvature or subtle hand textures.
This research proved influential to the conceptualization of our own system, particularly regarding the importance of integrating pose detection frameworks like MediaPipe for feature extraction before CNN-based classification. Moreover, the identified gap—loss of visual cues in pose-only recognition—helped shape the decision to include a combined visual and skeletal approach in the future scope of the project.

---

[4] J. Forster, C. Schmidt, O. Koller, T. Hoyoux, et al. (2014), "RWTH-PHOENIX-Weather: A Large-Vocabulary Sign Language Recognition and Translation Corpus," IEEE/LREC

The RWTH-PHOENIX-Weather corpus remains one of the most influential datasets in continuous sign language research. Developed by Forster et al. (2014), this dataset comprises large-scale recordings of German Sign Language (DGS) broadcasts, aligned with corresponding gloss and text translations. The work primarily focused on dataset creation and the establishment of baseline recognition models based on CNN-LSTM hybrid architectures.
Its contribution was instrumental in enabling future studies on continuous sign language translation — moving beyond isolated gesture recognition to full-sentence comprehension. However, the

dataset's thematic constraint to weather forecasts and its focus on a single language limit its generalization to conversational or multilingual contexts.

In the context of the present project, this research highlights the critical role of large annotated datasets in model training and benchmarking. While our project employs smaller datasets suited for real-time ISL gesture recognition, the long-term vision aligns with the creation or adoption of larger, open-source datasets covering broader gesture vocabularies.

---

[5] M. Camgoz, S. Hadfield, O. Koller, and R. Bowden (2020), "Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation," IEEE/CVPR

The IEEE/CVPR paper by Camgoz et al. (2020) represents a major milestone in SLR research, introducing transformer-based architectures for end-to-end sign-to-text translation. Using the PHOENIX-Weather 2014T dataset, the authors developed a model that integrates convolutional neural networks for spatial encoding with transformer-based sequence-to-sequence learning for language modeling.

This approach marked a transition from traditional modular pipelines to unified models capable of simultaneously recognizing and translating sign sequences into natural language sentences. The system achieved state-of-the-art performance, demonstrating the potential of attention mechanisms in capturing long-term dependencies across gesture sequences.

Nevertheless, the architecture required significant computational resources and large annotated datasets, making real-time or multilingual deployment challenging. For this project, the paper provided inspiration for future work involving transformer-based integration once sufficient ISL data and computational infrastructure become available. It also emphasized the importance of balancing model complexity with accessibility and responsiveness in user-facing systems.

[6] A. Ramasamy and S. Subramanian (2022), “Real-Time Indian Sign Language Recognition Using MediaPipe and CNN,” IEEE ICICICT

Ramasamy and Subramanian (2022) proposed one of the most practical and region-specific approaches for real-time sign recognition. Unlike most existing works focused on ASL or European sign languages, this research targeted Indian Sign Language (ISL).
The authors developed a lightweight system combining Google’s MediaPipe hand landmark detection with a Convolutional Neural Network (CNN) classifier. The dataset used was a custom collection of ISL gestures captured under controlled lighting conditions.
The system demonstrated high recognition accuracy for a limited vocabulary, achieving low latency suitable for real-time webcam-based applications. The use of MediaPipe allowed for efficient extraction of hand joint coordinates, significantly reducing preprocessing time. However, the study acknowledged limitations such as small dataset size, lack of multi-user testing, and limited environmental diversity.
This work closely aligns with the practical goals of our project. The decision to integrate MediaPipe for hand tracking and CNN for classification was directly inspired by this study’s success in achieving real-time responsiveness. The identified gaps — especially dataset expansion and cross-user validation — have been incorporated into the “Future Enhancements” section of our project report.

Synthesis and Comparative Reflection

A comparative analysis of the above six references reveals clear trends and recurring challenges in the field of sign language recognition. While early studies (e.g., Forster et al., 2014) focused primarily on dataset creation and baseline architectures, more recent works (Naz et al., 2023; Papadimitriou & Potamianos, 2023) explore advanced graph-based and transformer-based models that

significantly enhance accuracy but introduce high computational demands.
The consistent research gaps identified across studies include:

1. **Insufficient Dataset Diversity** — Most datasets are domain-specific or language-limited (e.g., German or American Sign Language), lacking regional adaptation such as Indian Sign Language (ISL).
2. **Real-Time Processing Limitations** — High-performing models often require substantial GPU resources, reducing usability in low-end devices or browser-based environments.
3. **Environmental Sensitivity** — Varying lighting conditions, occlusions, and camera angles still challenge model robustness.
4. **Signer Independence** — Recognition accuracy tends to drop significantly for users not represented in the training dataset.
5. **Integration with Web Systems** — Few research works bridge the gap between deep learning models and web-based user authentication, accessibility, or UI integration.

By addressing these gaps, the present project makes a distinct contribution: it combines **deep learning-driven gesture recognition** with **web technologies (Node.js, MongoDB, HTML, CSS, and JavaScript)** to deliver a **real-time, user-friendly, and accessible sign language recognition system**. The research insights drawn from these IEEE references directly informed the architectural choices, preprocessing strategies, and scope for future expansion.

REFRENCES

This project and the accompanying research have been supported by a variety of academic, technical, and online resources. A significant portion of the theoretical foundation and technical understanding has been drawn from peer-reviewed IEEE research publications and digital learning platforms.

The study referred to works such as *“Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues”* by M. Al-Qurishi, T. Khalid, and R. Souissi (IEEE Access, 2021), which provided a comprehensive overview of existing models and open challenges in sign language recognition. Similarly, the paper by K. Papadimitriou and G. Potamianos (IEEE ICASSP, 2023), *“Sign Language Recognition via Deformable 3D Convolutions and Modulated Graph Convolutional Networks,”* contributed valuable insights into advanced deep learning models combining 3D CNNs and GCNs for enhanced accuracy and signer independence.

Further, research such as *“SignGraph: An Efficient and Accurate Pose-Based Graph Convolution Approach Toward Sign Language Recognition”* by N. Naz, H. Sajid, S. Ali, O. Hasan, and M.K. Ehsan (IEEE Access, 2023), offered significant understanding into the role of graph convolutional networks in pose-based gesture recognition. The foundational dataset paper *“RWTH-PHOENIX-Weather: A Large-Vocabulary Sign Language Recognition and Translation Corpus”* by J. Forster et al. (IEEE/LREC, 2014), was also referenced to understand dataset creation and evaluation in large-scale SLR systems. Additionally, *“Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation”* by M. Camgoz, S. Hadfield, O. Koller, and R. Bowden (IEEE/CVPR, 2020), inspired the exploration of transformer-based architectures for end-to-end translation models.

For Indian Sign Language recognition specifically, *“Real-Time Indian Sign Language Recognition Using MediaPipe and CNN”* by A. Ramasamy and S. Subramanian (IEEE ICICICT, 2022), was particularly relevant, highlighting real-time gesture recognition approaches and the use of MediaPipe for lightweight processing.

In addition to these IEEE resources, **Google Scholar** and **IEEE Xplore** digital libraries were extensively used to locate and access the latest research papers and datasets. **GitHub** played a crucial role in providing open-source implementations, datasets, and pre-trained models which served as reference frameworks during the development and testing of the proposed system. **YouTube tutorials** and developer channels were instrumental in understanding complex concepts related to machine learning model training, CNN architectures, dataset preprocessing, and frontend-backend integration.

Stack Overflow, **Google Colab documentation**, and **Medium technical blogs** further assisted in resolving implementation challenges and understanding the nuances of TensorFlow, OpenCV, and MediaPipe frameworks.

Therefore, this research and project development were made possible through the collective contributions of IEEE research papers, open educational resources on **YouTube**, coding examples and repositories from **GitHub**, as well as extensive information and datasets accessed through **Google** and **IEEE Xplore** platforms. These sources together provided the theoretical knowledge, technical frameworks, and practical guidance necessary for the successful design, development, and analysis of the sign language recognition system.

CONCLUSION

The development of the Sign Language Recognition System marks a significant step toward bridging the communication gap between the hearing-impaired community and the rest of society. This project successfully integrates the power of **deep learning**, **computer vision**, and **web technologies** to create a robust, interactive, and accessible platform capable of translating sign gestures into textual or verbal output. The use of **HTML, CSS, and JavaScript** for the frontend provided an intuitive and user-friendly interface, while **MongoDB** ensured secure and efficient data storage and management at the backend.

Throughout the implementation of this project, a variety of existing models, techniques, and datasets were studied through IEEE research papers and other scholarly sources. It was observed that while significant progress has been made in sign language recognition globally, there remain substantial gaps, especially concerning **real-time processing**, **multilingual recognition**, and **cross-signer adaptability**. By reviewing and synthesizing findings from papers such as *Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues (Al-Qurishi et al., 2021)* and *Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation (Camgoz et al., 2020)*, it became evident that combining pose-based and image-based models could greatly enhance recognition accuracy and system robustness.

The practical implementation of the system demonstrated how technologies such as **MediaPipe** for hand landmark detection and **Convolutional Neural Networks (CNNs)** for image classification can be effectively utilized to identify and interpret sign gestures in real time. Furthermore, integrating the recognition model with a dynamic web-based interface not only enhances accessibility but also ensures that the system can be deployed on various platforms and used by people with minimal technical knowledge. The data flow from user interaction to recognition output and the storage of relevant information in the MongoDB database ensured a seamless and efficient process across all modules.

This project also underscored the importance of **dataset diversity** and **model training** for improving recognition accuracy. Most available datasets, such as RWTH-PHOENIX, WLASL, and AUTSL, cater primarily to American or European sign languages, which limits their direct application to Indian Sign Language (ISL). Hence, the creation of custom datasets tailored to local linguistic and cultural contexts is crucial for future improvements. The implementation of preprocessing techniques, proper annotation, and the use of data augmentation can further optimize model performance, especially in cases where large-scale datasets are unavailable.

Another important finding from this research is the necessity of **real-time optimization** and **cross-device compatibility**. Although many deep learning models achieve high accuracy in controlled environments, their efficiency tends to drop in real-world scenarios where lighting, background, and gesture speed vary significantly. The proposed system was therefore designed with a focus on lightweight models that could run smoothly even on standard computing devices or mobile platforms without requiring high-end GPUs.

In addition to technical achievements, the project highlights the broader **social and humanitarian significance** of sign language recognition technologies. By enabling smooth interaction between hearing-impaired and non-signing individuals, this innovation fosters inclusivity, accessibility, and empowerment. It aligns with the vision of using artificial intelligence and machine learning not just for technological advancement but for meaningful social transformation.

However, despite the progress made, this study recognizes that several challenges still persist. The model's performance can further be improved by incorporating multimodal data (such as facial expressions and body posture), integrating **Natural Language Processing (NLP)** for contextual understanding, and adopting **transformer-based hybrid architectures** to enhance semantic translation from gestures to meaningful text. Additionally, further efforts must be directed toward building larger, open-source, annotated datasets for Indian and regional sign languages to promote broader research collaboration.

In conclusion, the Sign Language Recognition System developed in this project represents a practical, research-driven, and socially relevant contribution to the field of assistive technologies. It demonstrates the feasibility of combining deep learning models with efficient web-based interfaces for real-time gesture recognition. The project not only validates the power of AI in solving accessibility challenges but also opens up new pathways for future work, including **real-time mobile deployment**, **speech-to-sign translation**, and **integration with educational or healthcare platforms**.

Ultimately, this work aspires to serve as a foundation for future innovations that make communication more inclusive and technology more humane, ensuring that every individual — regardless of their hearing ability — can connect, express, and be understood in the digital age.