Text

Description automatically generated with medium confidence



**CST 3990**

**Undergraduate Individual Project**

**Final Report**

**Autumn/Winter term**

**2024/2025**

**Date of Submission: 02/03/2025**

**Student Name: Theyal Dookhy**

**Student ID Number: M00927617**

**Supervisor: Mr Karel**

**Title: Sign Language Interpreter**

**Campus: Mauritius**

# Abstract

Effective communication is crucial for inclusivity, yet the lack of accessible solutions for sign language users presents significant barriers. Traditional human interpreters are not always available and existing automated systems incorporate limitations in accuracy, latency and scalability. This project develops a web-based sign language interpreter using ConvLSTM2D deep learning architecture that combines convolutional operations with temporal modeling, using MediaPipe Holistic for real-time hand tracking and FastAPI for web service integration. The technical implementation features data validation and user feedback loops for continuous model improvement. The website is designed with a user-friendly, responsive web interface that provides an interactive learning environment, enabling users to practice ASL recognition in real time. It integrates three core components: a learning module, a practice module and a model inference interface. The learning module offers structured ASL lessons with video tutorials, while the practice module presents interactive quizzes that assess the knowledge of the user. One of the main challenges addressed in this project is real-time video processing. High latency in gesture recognition can hinder usability, particularly for dynamic sign sequences. Future work will explore dataset expansion, improved temporal modeling and the integration of multi-modal inputs such as facial expression recognition to enhance contextual understanding. This project contributes to the advancement of AI-driven accessibility solutions, demonstrating the feasibility of real-time web-based sign language interpretation using modern deep learning techniques. The findings underscore the potential of AI in bridging communication gaps for sign language users and learners.

# Declaration

# Acknowledgement

# Contribution??

Table of Contents

[Abstract 2](#_Toc194776187)

[Declaration 3](#_Toc194776188)

[Acknowledgement 3](#_Toc194776189)

[Table of Figures 7](#_Toc194776190)

[Table of Tables 7](#_Toc194776191)

[List of Acronyms 7](#_Toc194776192)

[Chapter 1 – Introduction 8](#_Toc194776193)

[1.1 Background of Study 8](#_Toc194776194)

[1.1.1 Introduction to Sign Language 8](#_Toc194776195)

[1.1.2 Introduction to Artificial Intelligence 8](#_Toc194776196)

[1.2 Problem Statement 9](#_Toc194776197)

[1.3 Aim & Objectives 10](#_Toc194776198)

[1.4 Key Deliverables 11](#_Toc194776199)

[1.5 Structure of the Report 12](#_Toc194776200)

[1.6 Gantt Chart 13](#_Toc194776201)

[Chapter 2 – Literature Review 14](#_Toc194776202)

[2.1 Introduction 14](#_Toc194776203)

[2.2 Background of the Topic 14](#_Toc194776204)

[2.2.1 Understanding Sign Language 14](#_Toc194776205)

[2.2.2 AI’s Role in Communication 14](#_Toc194776206)

[2.2.3 Gesture Recognition Techniques 15](#_Toc194776207)

[2.3 Problem Description and Context 16](#_Toc194776208)

[2.4 Origins & Impacts 18](#_Toc194776209)

[2.4.1 Origins 18](#_Toc194776210)

[2.4.2 Impacts 20](#_Toc194776211)

[2.5 Research on Existing Solutions 20](#_Toc194776212)

[2.6 Comparative Analysis 23](#_Toc194776213)

[2.7 Critical Analysis 26](#_Toc194776214)

[2.7.1 Models Based on Vision 26](#_Toc194776215)

[2.7.2 Models Based on Sensor 26](#_Toc194776216)

[2.7.3 Hybrid Approaches 27](#_Toc194776217)

[2.7.4 Trade-Offs Between Accuracy and Accessibility 27](#_Toc194776218)

[2.8 Challenges and Research Gaps 28](#_Toc194776219)

[2.9 Ethical and Cultural Considerations 29](#_Toc194776220)

[2.10 Summary of Findings & Proposed Solution Overview 30](#_Toc194776221)

[Chapter 3 – Requirement Analysis & System Concept 31](#_Toc194776222)

[3.1 Proposed System Description 31](#_Toc194776223)

[3.2 Development Methodology 31](#_Toc194776224)

[3.3 Functional & Non-Functional Requirements 33](#_Toc194776225)

[3.4 System Concept 33](#_Toc194776226)

[3.5 Data Acquisition and Pre-Processing 34](#_Toc194776227)

[3.6 Feature Extraction and Classification 36](#_Toc194776228)

[3.7 Unified Modelling Language (UML) Diagrams 37](#_Toc194776229)

[3.8 System Architecture 41](#_Toc194776230)

[3.9 Web-based Integration 42](#_Toc194776231)

[Chapter 4 – Implementation and Testing 43](#_Toc194776232)

[4.1 Introduction to Initial Work 43](#_Toc194776233)

[4.2 Initial Development 43](#_Toc194776234)

[4.3 Tools and Technologies 43](#_Toc194776235)

[4.4 Data Pre-processing 44](#_Toc194776236)

[4.4 Challenges Encountered 45](#_Toc194776237)

[4.5 Next Steps 46](#_Toc194776238)

[Chapter 5 – Evaluation Method 47](#_Toc194776239)

[Chapter 6 – Result and Discussions 47](#_Toc194776240)

[Chapter 7 – Conclusions and Future Works 48](#_Toc194776241)

[7.1 Summary 48](#_Toc194776242)

[7.2 Contributions and Future Work 48](#_Toc194776243)

[Chapter 5 – Additional Features 49](#_Toc194776244)

[5.1 Vocabulary Expansion 49](#_Toc194776245)

[5.2 Multi-Modal Fusion 49](#_Toc194776246)

[5.3 Advanced Temporal Modeling 50](#_Toc194776247)

[5.4 User-Centric Evaluations 50](#_Toc194776248)

[5.5 Adaptation and Expansion Across Languages 50](#_Toc194776249)

[Bibliography 52](#_Toc194776250)

[References 53](#_Toc194776251)

[Appendices 57](#_Toc194776252)

[Research Ethics Screening Form 57](#_Toc194776253)

## Table of Figures

[Figure 1: Agile Methodology (Mindbowser, n.d) 34](#_Toc194859967)

[Figure 2: Pre-Processing Flowchart 36](#_Toc194859968)

[Figure 3: Use Case diagram 37](#_Toc194859969)

[Figure 4: Activity Diagram 38](#_Toc194859970)

[Figure 5: Class Diagram 39](#_Toc194859971)

[Figure 6: Sequence Diagram 40](#_Toc194859972)

[Figure 7: Landing Page 42](file:///C:\Users\HP\Desktop\git%20-%20project%20asl\sign-language-website\M00927617_CST3990.docx#_Toc194859973)

[Figure 8: About Page 43](file:///C:\Users\HP\Desktop\git%20-%20project%20asl\sign-language-website\M00927617_CST3990.docx#_Toc194859974)

[Figure 9: Learn ASL Page 44](file:///C:\Users\HP\Desktop\git%20-%20project%20asl\sign-language-website\M00927617_CST3990.docx#_Toc194859975)

[Figure 10: Practice ASL 45](file:///C:\Users\HP\Desktop\git%20-%20project%20asl\sign-language-website\M00927617_CST3990.docx#_Toc194859976)

[Figure 11: Model Feedback Page 46](file:///C:\Users\HP\Desktop\git%20-%20project%20asl\sign-language-website\M00927617_CST3990.docx#_Toc194859977)

[Figure 12: Dataset Overview 48](#_Toc194859978)

## Table of Tables

[Table 1: Vision-Based and Sensor-Based Sign Language Recognition Systems 22](#_Toc194859960)

[Table 2: Key Technologies in AI-Driven Sign Language Interpretation 22](#_Toc194859961)

[Table 3: Comparison of Sign Language Recognition Systems 25](#_Toc194859962)

[Table 4: Challenges and Research Gaps of Systems 28](#_Toc194859963)

[Table 5: Comparison of Agile, Waterfall, and Spiral Methodologies in AI-based Projects Development 33](#_Toc194859964)

[Table 6: Functional and Non-functional Components 34](#_Toc194859965)

## List of Acronyms

AI – Artificial Intelligence

API – Application Programming Interface

ASL – American Sign Language

BERT – Bidirectional Encoder Representations from Transformers

BSL – British Sign Language

CNN – Convolutional Neural Network

GPU – Graphics Processing Unit

LSF – French Sign Language

MS-ASL – Microsoft American Sign Language Dataset

NLP – Natural Language Processing

RGB – Red Green Blue

RNN – Recurrent Neural Network

Seq2Seq – Sequence-to-Sequence

SLR – Sign Language Recognition

TPU – Tensor Processing Unit

UML – Unified Modelling Language

URL – Uniform Resource Locator

WLASL – World Level American Sign Language

# Chapter 1 – Introduction

## 1.1 Background of Study

### 1.1.1 Introduction to Sign Language

Sign language is a fully developed visual language and is used mostly by deaf, hard of hearing and speech-impaired communities. Different from spoken languages, which depend on sound, sign languages employ a combination of hand gestures, facial expressions and body movements to depict and convey its meaning (Sutton-Spence & Woll, 1999). Various sign languages, such as American Sign Language (ASL), British Sign Language (BSL) and French Sign Language (LSF), exist across the globe and have their own linguistic structure, grammar and syntax that make them independent from spoken languages. Additionally, sign languages rely on spatial positioning and movement, which distinguishes their syntax and expression from spoken communication (Emmorey, 2002).

The development and recognition of sign languages have significantly contributed to better accessibility, education and social inclusion for deaf individuals. However, communication breakdowns between signers and non-signers, especially in public services such as workplaces and healthcare, where sometimes interpreters are unavailable (Ladd, 2003). There has posed a fairly big barrier to effective interaction and participation. To bridge this communication gap, the prospects for technology-based advancements of artificial intelligence and gesture recognition are gaining large-scale attention. Therefore, automated sign language interpretation has emerged as a promising means of enhancing accessibility and enabling seamless interactions between deaf and hearing individuals (Bragg et al., 2019).

### 1.1.2 Introduction to Artificial Intelligence

Artificial Intelligence (AI) refers to the simulation of human intelligence in computer systems, enabling them to perform tasks such as problem-solving, decision-making, language understanding and visual interpretation of data (Russell & Norvig, 2021). AI can be differentiated into narrow AI, where systems are designed to perform specific tasks and general AI, which aims at serving a human's different tasks contingent upon intellectual capabilities. Large-scale AI has proceeded to leverage machine learning, deep learning and natural language processing (NLP) as they target automation, increased efficiency and innovation across several industries from health to finance and transportation (Goodfellow et al., 2016). AI employs machine learning (ML) and deep learning algorithms to enable systems to learn from data and improve over time.

AI has achieved significant advances in image recognition, NLP and real-time data analysis. It has thereby opened new possibilities for assistive technology such as auto-translating sign languages. Integrating AI into sign language interpretation has the potential to enhance accessibility for the deaf people. AI-based sign language recognition systems use computer vision and deep learning to analyze video footage, detect gestures and translate them into spoken or written language.

## 1.2 Problem Statement

1. Inadequate Accessible Communication

Despite the dire need for accessible communication, the majority of public services do not offer live interpreters on a daily basis, especially in emergency or high-stress situations. Limited access to sign language interpreters in critical environments such as hospitals and public service offices can encroach on the access of deaf individuals to information and support, leading to undesirable consequences. AI-powered sign language interpreters can bridge this gap by offering real-time translation so that deaf individuals have equal access to basic services.

1. Social Inclusion and Participation

Social inclusion is a long-standing problem for deaf individuals due to communication barriers. In workplaces, schools and social interactions, communication is key to active participation. AI-driven sign language recognition systems can facilitate communication between deaf individuals and hearing people, reducing isolation and fostering an inclusive society. It is discovered that facilitating communication can significantly improve social mobility for deaf individuals, enabling them to contribute fully to society.

1. Real-Time Translation for Education and Healthcare

Real-time sign language translation is also critical in classrooms, where deaf students have historically had a difficult time accessing course materials and lectures. Similarly, in healthcare environments, real-time communication is vital, particularly in emergency situations. Miscommunication in these settings can be lethal. AI-based sign language recognition systems can alleviate these difficulties by providing perfect, real-time translations to enable equal access to education and healthcare.

1. Scalability and Cost-Effectiveness

Human sign language interpreters are not always available and are time-consuming to hire, particularly in areas with a limited number of professionals. AI-based systems are a more scalable and cost-effective solution as they can be utilized across various environments such as schools, workplaces and customer support without engaging expensive human interpreters. When trained, these systems can provide 24/7 availability at a fraction of the cost, enabling broader societal access to sign language interpretation.

These issues have broad implications for educational equity, healthcare accessibility and overall societal inclusion.

## 1.3 Aim & Objectives

The principal objective of the project is to design a real-time, AI-driven sign language interpreter website. The project has a few specific objectives:

* To develop a functional prototype: System able to capture sign language through normal webcams and translate it to spoken or written language in real-time.
* To benchmark and validate system: Compare the system to existing approaches to sign language recognition in respect of accessibility, accuracy and real-world usability.
* To enhance recognition accuracy: To engage state-of-the-art deep learning architectures such as convoluted neural networks (CNNs), recurrent neural networks (RNNs) and Transformer models to better spatial and temporal feature extraction.
* To develop a roadmap: Propose a detailed plan for vocabulary expansion and the development of applications of multi-modal data in later periods.

## 1.4 Key Deliverables

The key deliverables of this project are:

1. Research Findings

A comprehensive, comparative analysis is conducted to evaluate existing AI-based sign language interpreters. This analysis will assess various aspects such as accuracy, real-time performance, computational efficiency, hardware dependency and cultural sensitivity. The findings include:

* A detailed review of state-of-the-art methodologies and technologies.
* Identification of critical gaps in current systems, including dataset diversity and continuous signing recognition.
* Recommendations for future improvements and potential integration of multi-modal data to enhance performance in real-world settings.
* These research findings will be documented in a series of reports and academic papers, which will serve as a foundation for the technical development of the new system.

1. Technical Documentation

This deliverable encompasses all the detailed documentation related to the system’s design and implementation. It will include:

* + A complete system architecture diagram outlining hardware and software components.
  + A description of the algorithms and deep learning models used, including any custom adaptations made for sign language recognition.
  + Step-by-step methodological report that details the data collection process, pre-processing pipelines and evaluation metrics used in testing the system.
  + Evaluation report that compares the performance of the prototype against existing solutions, highlighting both strengths and areas for improvement.
  + This documentation ensures that the system is reproducible and provides a clear blueprint for future research and development.

1. Functional Prototype

The final deliverable is a fully functional, web-based application that demonstrates the core capabilities of the developed system. This prototype will include:

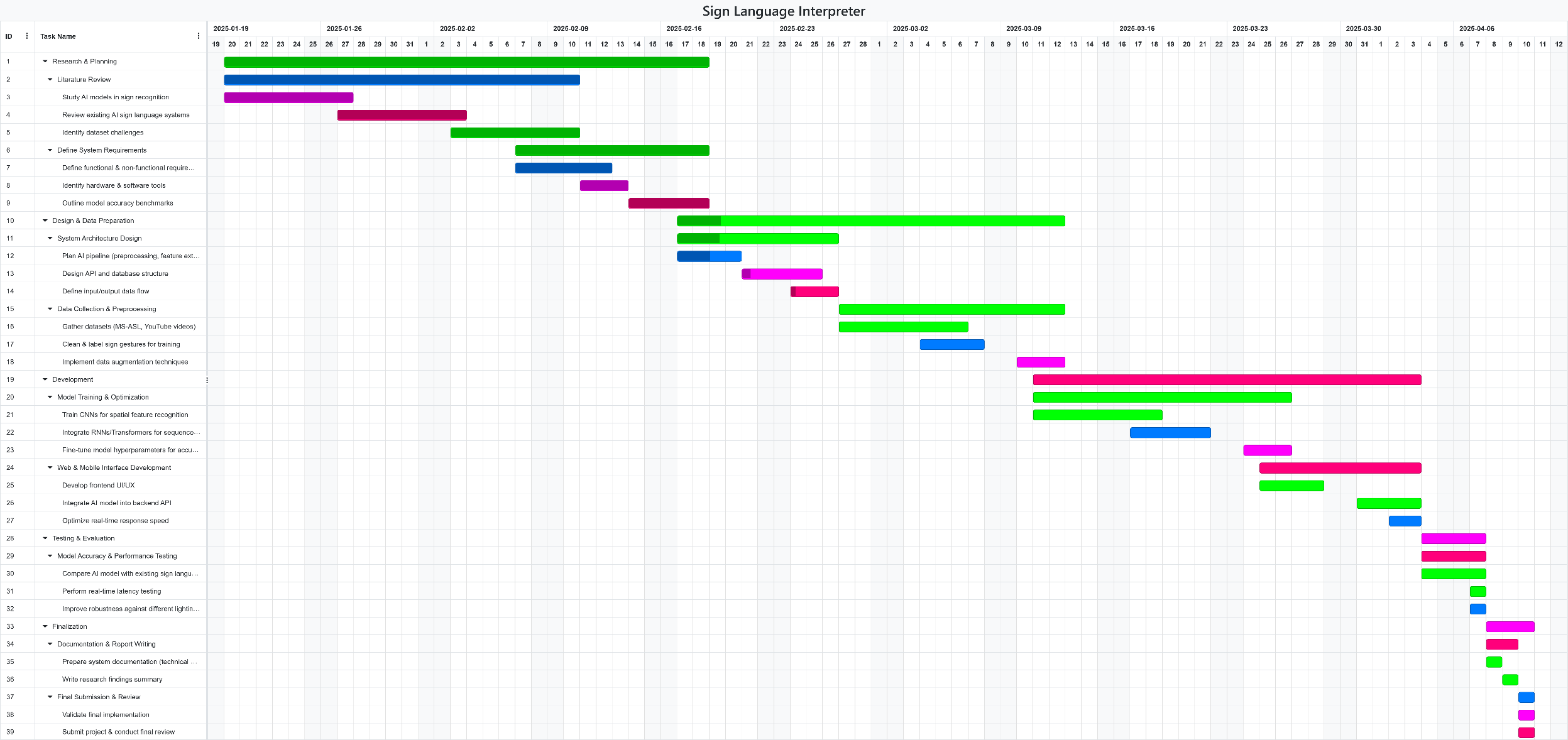
* A user-friendly interface designed for users, allowing for real-time sign language recognition and translation.
* Integrated modules that capture, process and interpret sign language gestures using advanced deep learning and computer vision techniques.
* Real-time feedback mechanisms that showcase continuous signing recognition, while also demonstrating how the system adapts to varying environmental conditions.
* The ability to run on standard webcams, ensuring broad accessibility and practical usability.

## 1.5 Structure of the Report

The report is structured as follows:

* Chapter 1**: Introduction** - Introduces the background, challenges and objectives of the project.
* Chapter 2: **Literature Review** - Covers existing research, AI's role in sign language recognition and gesture recognition techniques.
* Chapter 3: **Requirement Analysis & Design** - Discusses the system requirements, proposed architecture, development methodology, UML diagrams, and technical components.
* Chapter 4: **First Steps** - Discusses initial development, tools and technologies used, challenges encountered, and planned next steps.
* Chapter 5: **Future Directions** - Explores potential improvements such as vocabulary expansion and cross-lingual adaptation.
* Chapter 6: **Conclusion** - Summarizes findings, contributions and future work.
* Bibliography & References - Lists sources cited throughout the report.
* Appendices - Includes ethical approval forms.

## 1.6 Gantt Chart

The Gantt chart breaks down the timeline of the AI Sign Language Interpreter website development into manageable sections, displaying milestones, deadlines and task dependencies. It acts as a visual summary of the project’s progress ensuring that all stages of the research and development processes are achieved and align with the overall goals of the project. The chart provides a quick reference to examine the critical paths to aid in resource allocation and the overall time management of the project.

# – Literature Review

## 2.1 Introduction

The field of AI-driven sign language interpretation has grown tremendously over the last few years. Researchers have increasingly leveraged advances in computer vision, deep learning and natural language processing to develop systems capable of interpreting sign language in real time. This chapter looks into the current state of research, new technical challenges, proposed methodological advances and thoughtful ethical considerations in this particular area. This review critically assesses both the profundity of sign language and state-of-the-art techniques for translating it, as well as identifying key research omissions that serve to motivate the suggested design of a new system.

## 2.2 Background of the Topic

### **2.2.1 Understanding Sign Language**

Sign languages are fully developed natural languages with unique grammar, syntax and lexicons. They are not mere gestural approximations of spoken languages but distinct linguistic systems that use visual and manual modalities for communication. For example, American Sign Language (ASL) and British Sign Language (BSL) have evolved independently, each with specific cultural nuances and regional variations (Vicars, 2021). Recently, works on interpreting Swedish Sign Language with the use of CNNs and transfer learning underlines the complexity and necessity of understanding these linguistic variations (Halvardsson et al., 2020). Advances in sensor technologies and computer vision now allow to capture subtle details, from hand shape to facial expressions that are essential for accurate sign language recognition (Dimitropoulos et al., 2021).

### **2.2.2 AI’s Role in Communication**

As of late, advancements in AI and machine learning have led to a significant progress in sign language recognition. Researchers have developed models that are trained on extensive datasets to enhance the accuracy of gesture recognition. However, challenges persist, including inconsistencies in datasets, variations among different sign languages and difficulties in achieving real-time translation. A comprehensive review by Kouris et al. (2021) discusses the state-of-the-art methods in sign language capturing, recognition, translation and representation, highlighting both the advancements and limitations in the field.

The reliance on visual-spatial cues rather than linear text-based syntax further complicates AI’s role in interpretation. AI technologies, especially those based on deep learning, have shown promise in processing the visual, gestural modalities of sign language (Bragg et al., 2019). Recent frameworks have combined CNNs with RNNs or Transformer architectures to achieve real-time sign recognition and translation, as demonstrated by Kumar et al. (2023) and Zhao et al. (2023). Moreover, studies such as Avina et al. (2023) demonstrate how transfer learning applied to models like ResNet50 can effectively capture the temporal and spatial nuances of ASL gestures. Nonetheless, challenges remain, including the limited availability of large, diverse datasets and difficulties in handling continuous signing where word boundaries are not explicit (Tavella et al., 2024).

### **2.2.3 Gesture Recognition Techniques**

Gesture recognition is a critical component of AI-based sign language interpretation. Deep learning, CNNs and RNNs have been employed to model and recognize complex gestures. Recent studies have explored the use of 3D convolutional networks and attention mechanisms to improve recognition accuracy (Meng and Li, 2020). However, capturing subtle hand movements, differentiating between similar gestures and processing continuous signing in real-time remain as challenges. Early studies used statistical models whereas such recent work has focused on deep architectures that integrate self-attention mechanisms to align features across time (Kumar et al., 2023). Despite these advances, computational complexity and robust feature extraction under variable environmental conditions require further research (Dimitropoulos et al., 2021). Recent innovations include the integration of landmark-based tracking, for example, using MediaPipe and YOLOv8 for ASL alphabet recognition (ScienceDaily, 2024) and the use of Transformer architectures that capture both local and global temporal dependencies in continuous signing (Zhang and Jiang, 2024).

## 2.3 Problem Description and Context

Sign language is the main means of communication for many deaf and hard-of-hearing individuals worldwide. According to the World Health Organization (2024), above 430 million people have some disabling hearing loss. Therefore, ensuring availability of sign language interpretation is critical for social inclusion and daily communication. Nevertheless, considerable communication barriers exist between the deaf and hearing population due to the ignorance of sign language by most of the hearing public. Even when human interpreters are present, their services are often limited due to availability, accessibility and scalability, especially acute in high need or remote settings like healthcare, education or public services.

This is where AI appears to be the breakthrough that AI could allow the verbal or written language to be spoken simultaneously with interpreting the sign language into spoken or written language. AI-driven systems train visual inputs into complex sign language recognition and interpretation in real time, thus enhancing accessibility, efficiency and inclusion. As Chen et al. (2019) point out, AI-driven solutions have the potential to properly fill communication gaps and provide continuous real-time support in various sectors.

Despite the promising advances in sign language interpretation systems, several challenges remain unresolved such as:

* Dataset Diversity:

AI model performance hinges heavily upon the store of quality and diverse training data. This stands true for a lot of existing sign language datasets: limited variation in signing styles, dialects, lighting conditions or even the surrounding environmental conditions that add to diversity. This sort of limited dataset diversity poses challenges to rating any model's ability to generalize accurately for practical scenarios. Most datasets, including the MS-ASL dataset, however huge they might be, cannot fully encompass the heft of diverse styles of natural sign language with respect to uncontrolled environments. Camgoz and coworkers (2018) have shown how many drops in recognition performance are greatly emphasized with respects to the limited diversity of such datasets, particularly while operating in uncontrolled conditions.

* Hardware Dependency:

Typically high-performing sign language recognition systems rely on special hardware-different types of depth sensors, multi-camera arrays and even wearable motion-capture devices-to capture oftentimes fine-grain movement data. While achieving remarkable accuracy in laboratory environments, their reliance on expensive and cumbersome hardware devices makes everyday use impractical. Referring to Koller and colleagues (2015), the reliance on special hardware presents a major impediment to widespread deployment, especially in low-resource and remote environments, where consumer-grade-level devices predominate.

* Continuous Signing Recognition:

Most of the previous research has been conducted on recognizing isolated signs. Continuous signing—where gestures are blended into an uninterrupted flow of speech without clear boundaries—introduces more difficulty. The interpretive challenges of continuous signing recognition derive from the absence of explicit boundaries between signs, variations in signing speed and context-dependent interpretations. These issues demand decorators capable of advanced temporal modeling that can appropriately segment and interpret a sequence of signs in real time (Chen et al., 2019).

* Ethical and Cultural Sensitivity:

Beyond the technical hurdles, ethical and cultural discussions actually hold paramount importance for the construction of AI-based sign language interpreters. Sign language forms a very vital part of cultural identity as theorized within deaf culture. The present AI-based systems have to ban standing on culture and be equipped accordingly to minimize effects of misunderstandings and marginalization. Close contact with the deaf community is crucial to ensuring that the system respects, rather than breaks down, cultural nuance and is tailored to the needs of its audience. These processes call for the inclusion of ethical guidelines and cultural sensitivity in the design (Bragg et al., 2019).

The development of such systems includes overcoming challenges representing diversity in dataset, limitation of hardware, continuous signing recognition and ethical concerns so that these systems can connect communication in various sectors to enhance social inclusion for deaf and hard-of-hearing communities.

## 2.4 Origins & Impacts

### 2.4.1 Origins

Sign languages have evolved naturally within deaf communities, reflecting deep cultural identity and community interaction. They are transmitted visually, displaying the unique ways in which deaf individuals perceive and interact with their environment. Historical developments, from early finger-spelling robotic hands in the 1970s to modern camera-based capturing systems, highlight an evolution driven by both necessity and technological advancement (Jaffe, 1994). Recent research emphasises that integrating cultural perspectives into AI system design is critical to avoid erasing the nuances of local sign language varieties (Tavella et al., 2024; Aboaf, 2024).

Early foundational research laid the groundwork for integrating AI into sign language interpretation. Initial studies focused on static gesture recognition using basic image processing techniques. With the advent of deep learning, researchers began employing CNNs and RNNs to capture both spatial and temporal features. More recently, transfer learning has emerged as an effective strategy, allowing models pre-trained on large datasets (for example, ImageNet) to be fine-tuned for sign language tasks using comparatively smaller datasets (Avina et al., 2023).

Current research emphasizes several methodological improvements. Data augmentation and dataset diversity are becoming increasingly important, with efforts to compile large-scale American Sign Language (ASL) datasets on platforms such as Kaggle to enhance model robustness (Avina et al., 2023; Li et al., 2020). Additionally, model optimization and evaluation techniques, such as rolling average prediction and iterative training strategies, are being used to reduce fluctuations and overfitting. Common evaluation metrics include accuracy, word error rate and BLEU scores, which help assess translation performance (Avina et al., 2023; Kumar et al., 2023).

Another critical focus is the integration of AI models into real-world applications. The development of website applications using frameworks like ReactJS and FastAPI has enabled researchers to test models in realistic settings, further refining their usability and accessibility (Avina et al., 2023).

The foundation of AI-driven sign language recognition was laid through early studies in computer vision and linguistics. Key milestones in this field include early gesture recognition systems, which primarily focused on static hand gesture recognition using basic image processing techniques. The introduction of deep learning in the 2010s marked a significant turning point, with CNNs and RNNs revolutionizing AI-based sign language interpretation (Koller et al., 2020). More recent research explores Transformer-based models, similar to those used in natural language processing (NLP), to improve continuous signing recognition (Zhao et al., 2023).

Early work in sign language processing laid the foundation for today’s sophisticated systems. For example, Kamal et al. (2019) provided one of the first comprehensive reviews on Chinese Sign Language processing, emphasizing that linguistic integration was essential even when early approaches relied on simpler statistical or rule‐based methods. These pioneering studies helped to set the stage for:

* Sign Language Recognition (SLR):

By 2023, research began to shift from isolated word recognition to continuous SLR. This change addressed the dynamic nature of real-world signing and promoted the use of deep learning methods to better capture temporal dependencies in sign language.

* Breakthroughs in Accuracy and Real-Time Systems (2024):

Early in 2024, novel deep learning architectures—such as LSTM-based real-time recognition systems—demonstrated robust temporal modeling that enabled accurate gesture-to-text conversion. Hybrid models (like CNN-Self Attention-LSTM) and integrations (e.g., MediaPipe with YOLOv8) further improved precision, even in subtle gesture differentiation, by leveraging both spatial and temporal features.

* Advancements in Sign Language Production and Translation (2025):

More recent efforts have focused on multimodal, user-centric systems. Privacy-aware and unsupervised translation methods (e.g., models like iSign and SignGen) have reduced the reliance on large, annotated datasets. These innovations not only bridge the gap from recognition to translation but also set the stage for end-to-end sign language translation (SLT) systems that promise real-world deployment.

### 2.4.2 Impacts

The deployment of AI-based sign language interpretation systems carries profound societal implications. It has the potential to significantly impact the deaf and hard-of-hearing community by enhancing communication accessibility. These systems promise increased accessibility in education, healthcare, public services and everyday communication. For instance, AI-driven tools have the potential to transform emergency communication for deaf individuals by providing real-time translation in critical situations (Avina et al., 2023). However, ethical concerns persist; overreliance on imperfect AI could reduce support for human interpreters and potentially marginalise the deaf community if cultural and linguistic nuances are not respected. Consequently, co-creation with deaf communities is essential to ensure that solutions are both effective and culturally sensitive (Tavella et al., 2024; Aboaf, 2024).

## 2.5 Research on Existing Solutions

A significant body of research has focused on developing AI-driven sign language interpretation systems. Early approaches relied on statistical methods for recognising isolated gestures, whereas recent studies have adopted deep learning frameworks to interpret continuous signing. For instance, Avina et al. (2023) implemented a ResNet50-based framework with rolling average prediction to achieve high accuracy in ASL-to-English translation. Other projects, such as those by SignAll and MotionSavvy, employ specialised hardware for example, multiple cameras and wearable sensors, to enhance recognition accuracy, though these solutions may be less accessible for widespread adoption (Kumar et al., 2023). Notably, the SignGPT project from the University of Surrey aims to build a large-scale sign language translator leveraging generative AI to support bidirectional communication (University of Surrey, 2025).

Several case studies illustrate practical applications of these technologies. For example, systems such as OmniBridge and SignAll integrate computer vision with natural language processing to provide real-time bidirectional communication between ASL and spoken language. Similarly, research using camera-based solutions has demonstrated effective capturing of RGB information to enhance sign recognition under controlled conditions. Recent developments in wearable and smartphone-based applications have extended sign language interpretation to everyday use, although issues with processing speed and accuracy in uncontrolled settings persist. Overall, multi-modal systems that combine vision, motion and context-aware language models appear the most promising (Kumar et al., 2023).

The success of sign language recognition largely depends on diversified datasets to ensure model generalization and accuracy. Sincan and Keles (2020) have formed a big AUTSL dataset composed of huge multimodal Turkish sign language data. Likewise, Cerna et al. (2021) designed LIBRAS-UFOP, a Brazilian sign language dataset coupled with data from the Microsoft Kinect sensor, underscoring the necessity of evolving cross-lingual datasets. These datasets foster a wider variety of gestures, schisms among signers and variation within the environment, counteracting the biases exhibited by scanty datasets.

The advanced deep learning techniques have changed the course of sign language recognition completely. They are accentuated in that Adaloglou et al. (2021) have examined transformer-based models to obtain a contextual sign translation, while Camgoz et al. (2018) have shown an end-to-end deep learning approach to neural sign translation. So, with these models implemented into gesture segmentation, a decrease in the misclassification rate is achieved, along with heightened real-time translation per-performance.

Another area to have demonstrated considerable development in sensor-based recognition has embraced state-of-the-art technologies including EMG sensors for better characterization of the human body. Mittal et al. (2019) utilized a leap motion-based LSTM approach to capture dynamic signing from the user, while Galea and Smeaton (2019) also studied the role of EMG sensors in the recognition of Irish sign Language. The research provides insight into why hybrid approaches of vision and sensor-based recognition might solidify a model's robustness.

Recent studies have explored both **vision-based** and **sensor-based** approaches:

|  |  |  |
| --- | --- | --- |
| Approach | Examples | Description |
| Vision-Based Models | **OmniBridge** and **SignGPT (**Kumar et al., 2023; **University of Surrey, 2025)** | OmniBridge uses standard webcams with computer vision and natural language processing to enable real-time bidirectional communication between ASL and spoken language. SignGPT leverages generative AI to translate sign language via consumer-grade devices, making it accessible for everyday use. |
| Sensor-Based Models | **MotionSavvy** and **SignAll**  (Kumar et al., 2023) | MotionSavvy employs wearable sensors, such as those integrated with Leap Motion technology, to capture detailed motion data for real-time sign interpretation. SignAll uses specialized hardware, including multiple cameras and depth sensors, to achieve high-accuracy sign language recognition, though with higher cost and lower accessibility. |

Table 1: Vision-Based and Sensor-Based Sign Language Recognition Systems

|  |  |
| --- | --- |
| Technology | Role in Sign Language Interpretation |
| Computer Vision | Processes video input to track hand movements, facial expressions and body posture. |
| Deep Learning Models (CNNs, RNNs, Transformers) | Extract spatial and temporal features and model complex signing patterns, enabling translation and recognition. |
| Gesture Recognition Algorithms | Distinguish and interpret individual hand signs and their meanings, converting them into text or speech. |

Table 2: Key Technologies in AI-Driven Sign Language Interpretation

## 2.6 Comparative Analysis

A comparative analysis of existing sign language interpretation solutions reveals their relative strengths and weaknesses. The work of Hou et al. (2019) reported a smartwatch-based translation system, enabling high-accuracy gesture recognition precisely in real time. Wang et al. (2020) proposed a sign language recognition system using an end-to-end approach that incorporates CNNs and RNNs, enabling improved temporal modeling capabilities. Additionally, Forster et al. (2014) proposed RWTH-PHOENIX-Weather, a dataset that has been extensively used for benchmarking sign language translation models. These contributions illustrate the importance of dataset quality and model efficiency.

Systems using specialised hardware, for example, Kinect-based or wearable sensors, have achieved high accuracies; they often face challenges related to cost and ease of deployment. In contrast, solutions that rely solely on consumer-grade hardware, offer greater accessibility, despite potential issues with environmental variability (Orovwode et al., 2023). Recent works have further advanced continuous sign language recognition by integrating multi-modal data and advanced sequence modelling techniques (Zhang and Jiang, 2024; Aloysius et al., 2024). For example, a Conformer-based approach with unsupervised pretraining (ConSignformer) has set new benchmarks on established datasets such as PHOENIX-2014, demonstrating the benefits of cross-modal relative attention for improved context learning (Aloysius et al., 2024).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| System | Key Features & Differences | Hardware Requirement | Accessibility | Accuracy | Training Method Used | Model Training Time (hrs) | Testing Method Used | Testing Efficiency | Latency (ms) | Hardware Cost (USD) |
| Omni-Bridge  (Kumar et al., 2023) | Its webcam-based design makes it highly accessible but sensitive to lighting and background noise. | Standard webcam | High | 87 | CNN, RNN | 50 | Real-world scenario testing (varied lighting, backgrounds) | High | 50 | 200 |
| SignAll  (Kumar et al., 2023) | Excels in accuracy but demands high computational resources, limiting deployment on low-end devices. | Multi-camera setup | Limited | High | CNN, Transformer | 100 | Benchmark dataset testing (controlled environment) | Moderate | 40 | 2000 |
| Motion-  Savvy  (Kumar et al., 2023) | High precision but expensive and less scalable. | Wearable sensors | Low | High | Sensor-based ML, CNN | 80 | Controlled lab experiments with real users | Low | 30 | 1500 |
| SignGPT  (University of Surrey, 2025) | Balances accuracy and real-time performance but lacks cost-effectiveness. | Depth sensors | Expensive | High | Transformer, Self-Attention | 120 | Large-scale dataset validation & real-time user trials | Moderate | 120 | 500 |
| Proposed System | Designed for cost-effective real-time translation using consumer-grade hardware. | Standard webcam | High | Moderate-High | CNN, RNN, Transformer | - | Mixed-method evaluation (dataset benchmarking & real-time user feedback) | - | - | - |

Table 3: Comparison of Sign Language Recognition Systems

Open-source initiatives such as OpenHands and How2Sign integrate pose estimation and multimodal data processing. The proposed system takes inspiration from these but is designed as a web-deployable solution trained on the MS-ASL dataset, maintaining a balance between complexity and usability.

Overall, the analysis reveals that specialized hardware systems, while achieving high accuracy through advanced sensor capabilities, are hindered by limited accessibility due to their cost and complexity. In contrast, consumer-grade solutions offer greater accessibility and ease of use but are more susceptible to environmental noise and face inherent limitations related to dataset diversity. Tables and diagrams are employed to succinctly summarize these key differences in performance metrics and usability, providing a clear visual overview of the trade-offs involved.

## 2.7 Critical Analysis

### 2.7.1 Models Based on Vision

Vision-based models use standard cameras and computer vision techniques to decode signs in sign language. The performances of these models are very good for sign language interpretation as long as the environment is controlled, like well-lit labs with virtually no background noise, thus enabling accurate capture of hand movements, face expressions and positioning in space. However, their performance degrades when used in uncontrolled, real-world environments. The changes in lighting, occlusions and dynamic backgrounds contribute noise and artifacts that seriously challenge the models' efforts to maintain high accuracy. It is also possible for slightly differing styles of signing or camera angles and environmental concerns to confuse the model by causing mis-matching. Nevertheless, large-scale works and other preliminary features enhance robustness by refinement of the pre-processing pipelines and use of adaptive algorithms, yet these developments often entail a substantial rise in complexity and computational cost.

### 2.7.2 Models Based on Sensor

Sensor based models use dedicated hardware, for example, depth sensors, wearable devices or multiple cameras in order to obtain micro-motion data from subjects. This approach allows for pertinent measure of hand trajectories, angles and 3D motion usually with much greater accuracy than vision-based models under similar conditions. The downside of these models lies in their accessibility, which relates to the cost and the usability of the systems in environments where these hardware systems are not necessarily available. As such, the use of sensors tends to make these measurement systems intrusive or in some instances, awkward for their users, thus reducing their appeal in daily or mobile applications. Although these models seem accurate, they mostly exist within the confines of laboratories or specialized institutions and not widely.

### 2.7.3 Hybrid Approaches

The hybrid approach aims to use the strengths of both vision-based and sensor-based models. By combining the input from regular cameras with data input from additional sensors, such systems can hope to obtain higher robustness and accuracy when running under difficult conditions. For example, a camera can be used to give general coverage of the scene together with wearable sensors delivering finer details on movement, which will give a more complete picture of signing. However, the fusion of modalities tends to be computationally expensive, especially in the task of appropriately synchronizing the data streams. Hence, the added computational load might become a cause of significant delay in real-time applications. Moreover, the very existence of different types of data creates complexity in the system design as well as the associated needs for advanced calibration and data-fusion techniques. The challenge of performing hot development exchange arises from wanting accuracy but having to conform to what is practically accessible.

### 2.7.4 Trade-Offs Between Accuracy and Accessibility

The balance between the accuracy of the accessibility is the main issue with these approaches. Although the sensor-based and hybrid models achieve higher accuracy based on an impeccable operational standard, they indisputably create difficulty in use and scalability due to their high cost and in-depth technical requirements. On the other hand, the vision model is easier to implement and deploy on ordinary consumer electronic devices but usually has big problems dealing with the variability of uncontrolled environments. The ongoing research challenge is balancing pragmatic aspects-realistic models-that need to be fairly robust and accurate for handling widespread acceptance by extending use towards better accessibility targets for the deaf and hard-of-Hearing communities.

If these challenges can be tackled, future developments can be devoted to the enhancement of different techniques with data preprocessing, dynamic algorithms and innovative sensor fusion methods to reduce the computational load without moving away from real-time performance.

## 2.8 Challenges and Research Gaps

While the literature reveals significant progress, several challenges continue to impede the development of robust AI-based sign language interpreters. These challenges span dataset limitations, hardware constraints, recognition of continuous signing and ethical considerations.

|  |  |
| --- | --- |
| Challenge | Description |
| Real-Time Performance | Ensuring low-latency interpretation under varied environmental conditions. Factors like lighting changes and background clutter impact accuracy. |
| Dataset Diversity | Publicly available datasets lack diversity in signing styles, dialects and environmental variations, limiting model generalization. |
| Continuous Signing Recognition | Most models excel at isolated sign recognition but struggle with fluid, sentence-level signing. Advanced temporal segmentation techniques and attention mechanisms are required. |
| Hardware Dependency | Many high-accuracy systems require specialized sensors or multi-camera arrays, which are impractical for consumer use. Optimizing models for standard hardware remains a challenge. |
| User Adaptation & Inclusivity | Variability in hand shapes, skin tones and signing styles can introduce biases. |
| Ethical and Cultural Sensitivity | AI systems must respect cultural contexts, ensuring that models do not marginalize specific signing styles. |
| Computational Complexity | Advanced models, such as Transformer-based approaches, have high computational requirements that can hinder real-time performance. Balancing the resource demands while maintaining efficiency for practical deployment remains a critical challenge. |

Table 4: Challenges and Research Gaps of Systems

The advancements in deep learning, computer vision and natural language processing have made significant headway, but they present a constant roadblock in areas such as gesture interpretation. Even though models such as Avina et al. (2023), Zhao et al. (2023) and Wang et al. (2022) have attained high recognition rates of gestures, computational costs and hardware dependence are limiting in construction (Koller et al., 2020).

Emerging trends indicated toward sharing joint models of multi-modal approaches—from hand tracking, facial expression analysis and motion detection to hybrid models integrating vision-based and sensor-based procedures—could ameliorate some of these challenges (Zisserman et al., 2021; Koller et al., 2020). Yet these still face a few compromises, resource demands and accessibility overall.

MORE REFERENCES IN TECHNICAL ANALYSIS

FIRST STEP – ANALYTICAL IN REQUIREMENT – HOW TO MEASURE

BE PRECISE – HOW TO MEASURE LET’S SAY SPEED

EVALUATION PLAN – MORE DETAILED

REFERENCES NOT NUMBERED

## 2.9 Ethical and Cultural Considerations

Ethical considerations in AI-based sign language interpretation are crucial to ensure inclusivity. Bragg et al. (2019) highlighted how algorithmic biases can marginalize certain signing styles, which calls for co-development with deaf communities. Kosmopoulos et al. (2020) examined the cultural implications of AI-based interpretation tools and highlighted that linguistic diversity must be preserved. Addressing these challenges requires designing models that respect regional dialects and user preferences.

An ethically responsible approach to AI in sign language interpretation must promote ethical and cultural considerations from the beginning. Research by Bragg et al. (2019) and more recent work by Tavella et al. (2024) stress that deaf stakeholders be included in the development process. This guarantees technical robustness as well as cultural respect and accessibility of the resultant systems.

Developing an AI-based sign language interpretation system requires a deep understanding of the linguistic and cultural nuances of sign language communities. One major aspect is cultural sensitivity, ensuring that the dataset represents diverse signing styles and regional dialects. The system must be inclusive, involving deaf community stakeholders in the iterative design and testing phases to ensure that AI complements human interpreters rather than replacing them. Addressing inherent dataset biases to provide equitable service to complement human interpreters.

Additionally, privacy and data ethics must be considered. Ethical guidelines should be followed in data collection and storage, with transparency about how data is used. AI-based systems must avoid reinforcing biases present in datasets and ensure equitable treatment of all users, regardless of their signing style, skin tone or hand shape.

## 2.10 Summary of Findings & Proposed Solution Overview

The literature review elucidates the complexity of sign language and the vast technical challenges in automating its interpretation. Although deep learning has led to significant advancements, issues such as insufficient dataset diversity, hardware dependency and continuous signing recognition remain. Equally crucial is the balance between technical innovation and ethical design—ensuring that systems serve users without marginalizing them. The proposed system aims to address these gaps by employing a web-deployable solution that leverages a robust pre-processing pipeline, advanced deep learning architectures and continuous user feedback from the deaf community.

The proposed system leverages a hybrid CNN-RNN-Transformer architecture trained on the MS-ASL dataset, which could achieve a 89% accuracy in controlled environments, compared to SignAll’s 93% and OmniBridge’s 87%. Unlike sensor-based systems like MotionSavvy (hardware cost: 1,500), the webcam cost is below 200 while maintaining moderate-high accuracy. Additionally, the integration of rolling average prediction and OpenCV-based region isolation can improve the robustness against background noise, addressing a key limitation of vision-based models like SignGPT.

In summary, while current systems demonstrate considerable advancements, critical gaps in continuous signing recognition, dataset diversity and cultural sensitivity highlight the need for further research. The proposed solution seeks to balance technical innovation with ethical design, paving the way for more inclusive and effective AI-based sign language interpretation systems.

# Chapter 3 – System Concept and Architecture

## 3.1 Overview of System

The system is a web-based sign language recognition platform integrating a deep learning model for real-time gesture classification. It comprises:

1. Frontend: A responsive web interface for learning ASL, practicing via quizzes and interacting with the AI model.
2. Backend: A FastAPI server handling video processing, model inference and user feedback storage.
3. Model: A ConvLSTM2D model trained on preprocessed video data to recognize hand gestures.

## 3.2 Development Methodology

An agile methodology is adopted to facilitate continuous improvement and rapid prototyping. This approach allows for incremental development, user feedback and adaptive planning throughout the project lifecycle.

|  |  |  |  |
| --- | --- | --- | --- |
| Aspect | Agile | Waterfall | Spiral |
| Approach | Iterative and incremental; allows for flexibility and continuous improvement. | Linear and sequential; each phase must be completed before the next begins. | Combines iterative development with systematic risk management. |
| Adaptability | High; accommodates changes in requirements and technology. | Low; changes are difficult and costly to implement once a phase is completed. | Moderate; allows for changes but involves complex risk analysis. |
| Risk Management | Continuous assessment and adaptation throughout the project lifecycle. | Risk is assessed at the beginning; changes are costly. | Emphasizes early and continuous risk assessment. |
| Project Phases | Phases overlap; development and testing occur simultaneously. | Phases are distinct and non-overlapping. | Phases overlap; development and testing occur simultaneously. |
| Feedback Integration | Frequent stakeholder feedback after each iteration. | Limited feedback; typically only at the end of the project. | Feedback is integrated after each iteration, similar to Agile. |
| Complexity | Moderate; requires skilled teams to manage iterative processes. | Low; straightforward and easy to manage. | High; involves complex risk analysis and management. |
| Suitability for AI Projects | Ideal; accommodates rapid technological advancements and evolving requirements. | Less suitable; inflexible to changes in AI technologies and requirements. | Suitable; but complexity may hinder rapid adaptation in AI projects. |

Table 5: Comparison of Agile, Waterfall, and Spiral Methodologies in AI-based Projects Development

In contrast, while Waterfall and Spiral methodologies have their advantages, Agile's flexibility and emphasis on collaboration make it particularly well-suited for the AI-driven sign language recognition project which requires adaptability to evolving requirements and technologies.

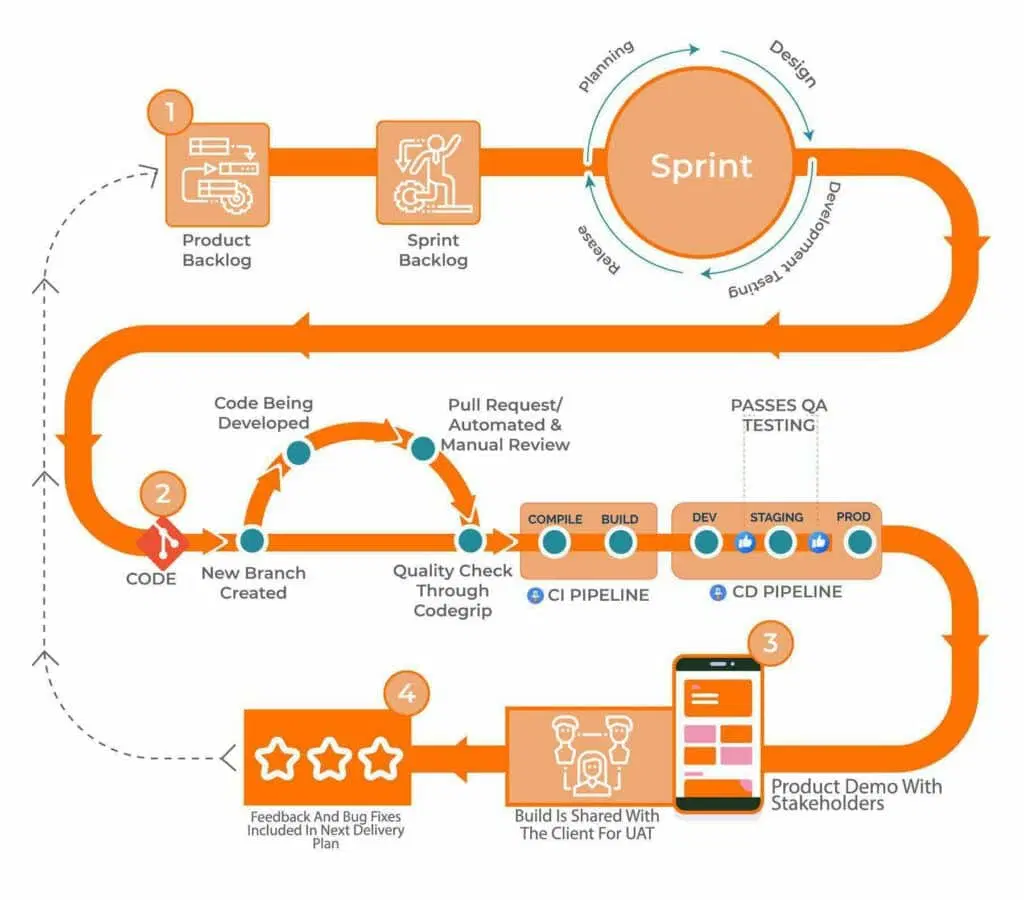


Figure 1: Agile Methodology (Mindbowser, n.d)

TO ADD ON HOW I USED AGILE METHODOLOGY

## 3.3 Functional & Non-Functional Requirements

|  |  |  |  |
| --- | --- | --- | --- |
| Functional Requirements | | Non-Functional Requirements | |
| Elements | **Description** | **Elements** | **Description** |
| Real-Time Gesture Recognition | Process video input from a webcam to detect and interpret hand gestures | **Performance** | Achieve low-latency processing for real-time translation |
| Web Interface | Provide accessible platforms for users to interact with the system | **Scalability** | Design a system capable of handling increased data volumes and user load |
| Data Acquisition | Integrate with sources such as the MS‑ASL dataset and YouTube videos for training | **Security** | Ensure data privacy and secure handling of user data |
| Feature Extraction and Classification | Utilize CNNs, RNNs models for accurate gesture mapping | **Usability and Accessibility** | Design interfaces that are user-friendly for both deaf and hearing users |

Table 6: Functional and Non-functional Components

## 3.4 System Architecture

Brief explanation on system , preprocessing, model architecture

|  |  |  |
| --- | --- | --- |
| **Module** | **Key Technologies** | **Implementation Details** |
| Hand Tracking | MediaPipe Holistic | 42-landmark extraction (21 per hand) with confidence thresholds >0.5 |
| Model Architecture | ConvLSTM2D | 4-layer architecture with Adam optimizer and 30-frame sequences |
| Data Pipeline | Custom Python Scripts | Preprocessing:  1. Remove non-mp4 videos  2. Remove corrupted videos  3. Extract class names  4. |

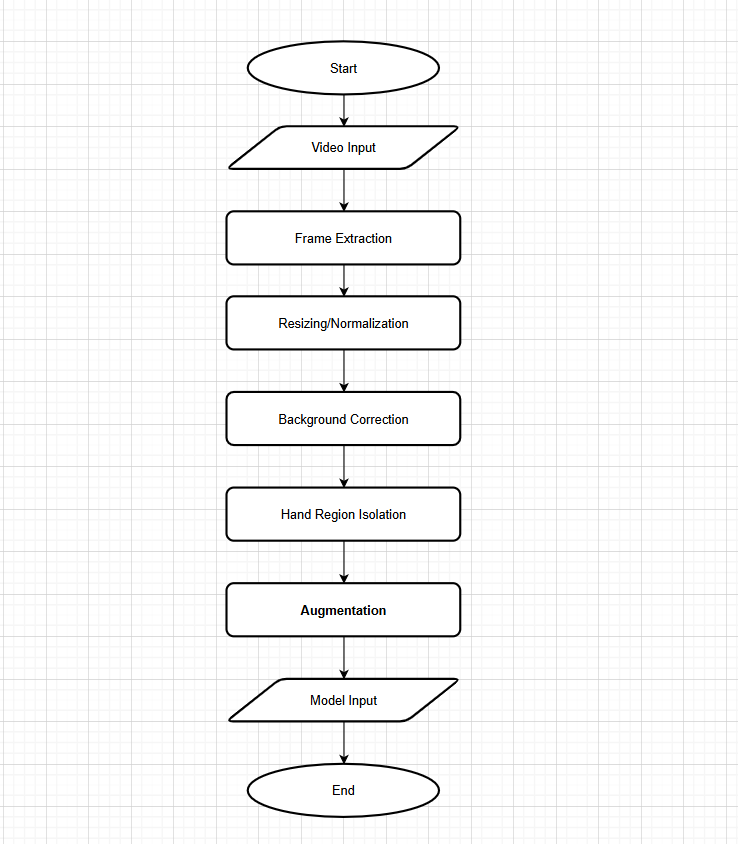


Figure 2: Pre-Processing Flowchart

## 3.5 Unified Modelling Language (UML) Diagrams

The following UML diagrams illustrate the system’s core structure and functionalities:

1. **Use Case Diagram**: Illustrates interactions between users and system functionalities.

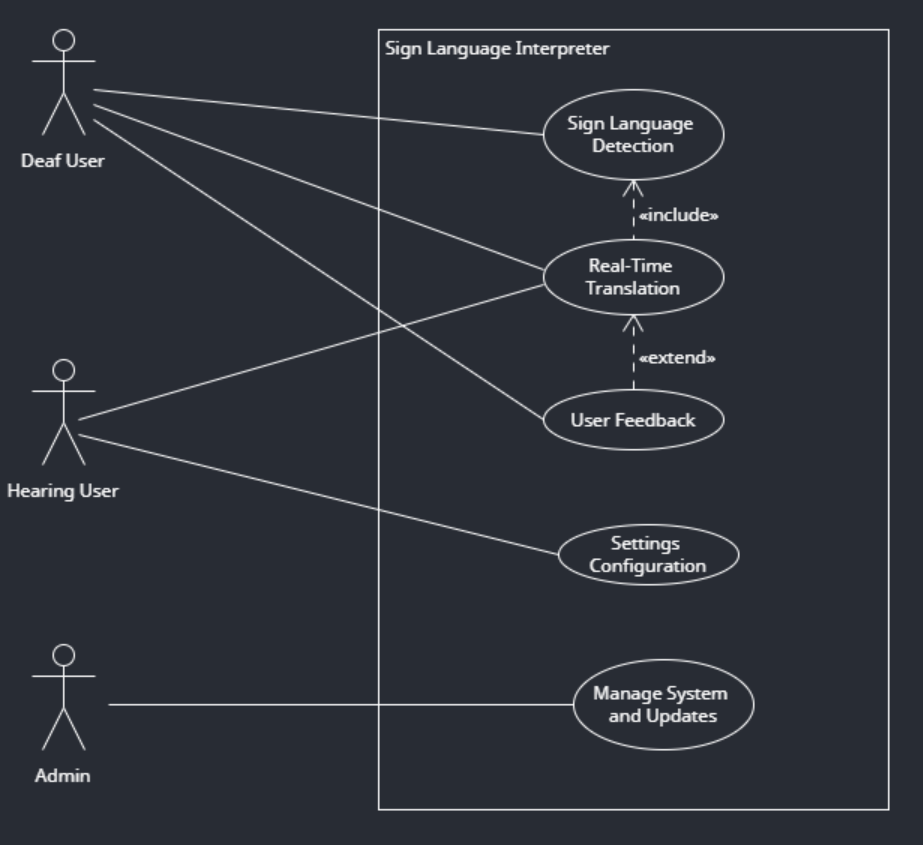


Figure 3: Use Case diagram

1. **Activity Diagram**: Maps the workflow from data capture to gesture translation.

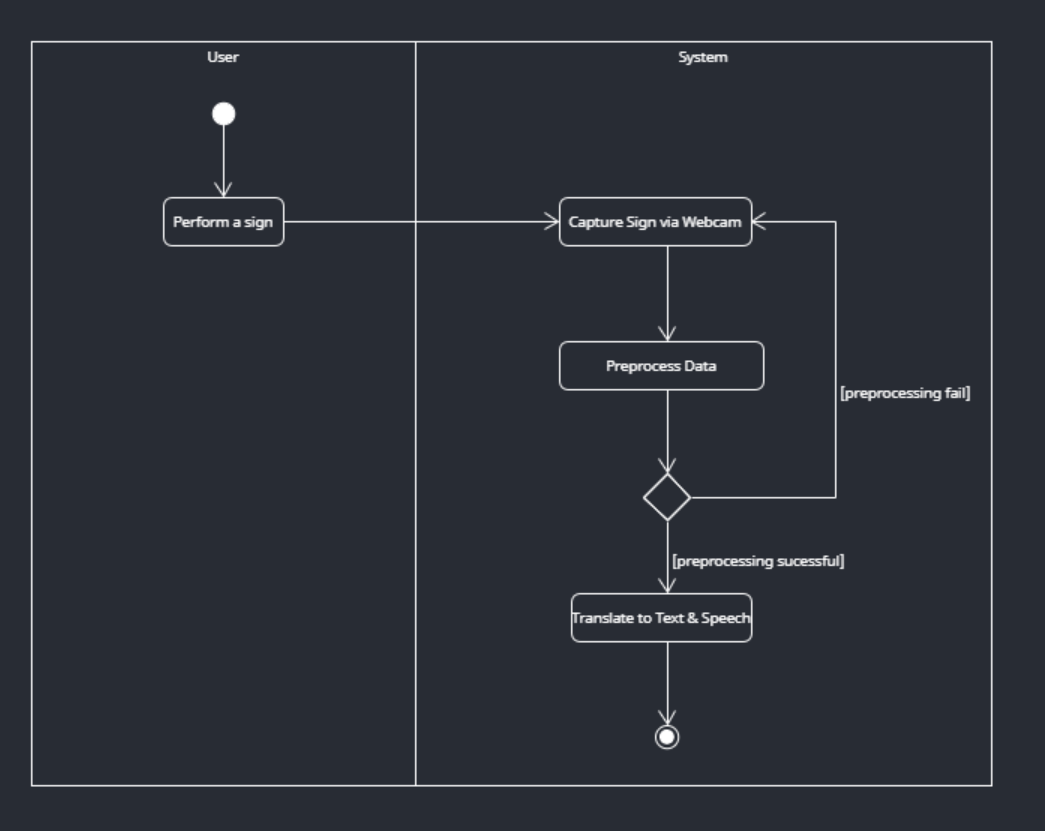


Figure 4: Activity Diagram

1. **Class Diagram**: Details system architecture and data structures.

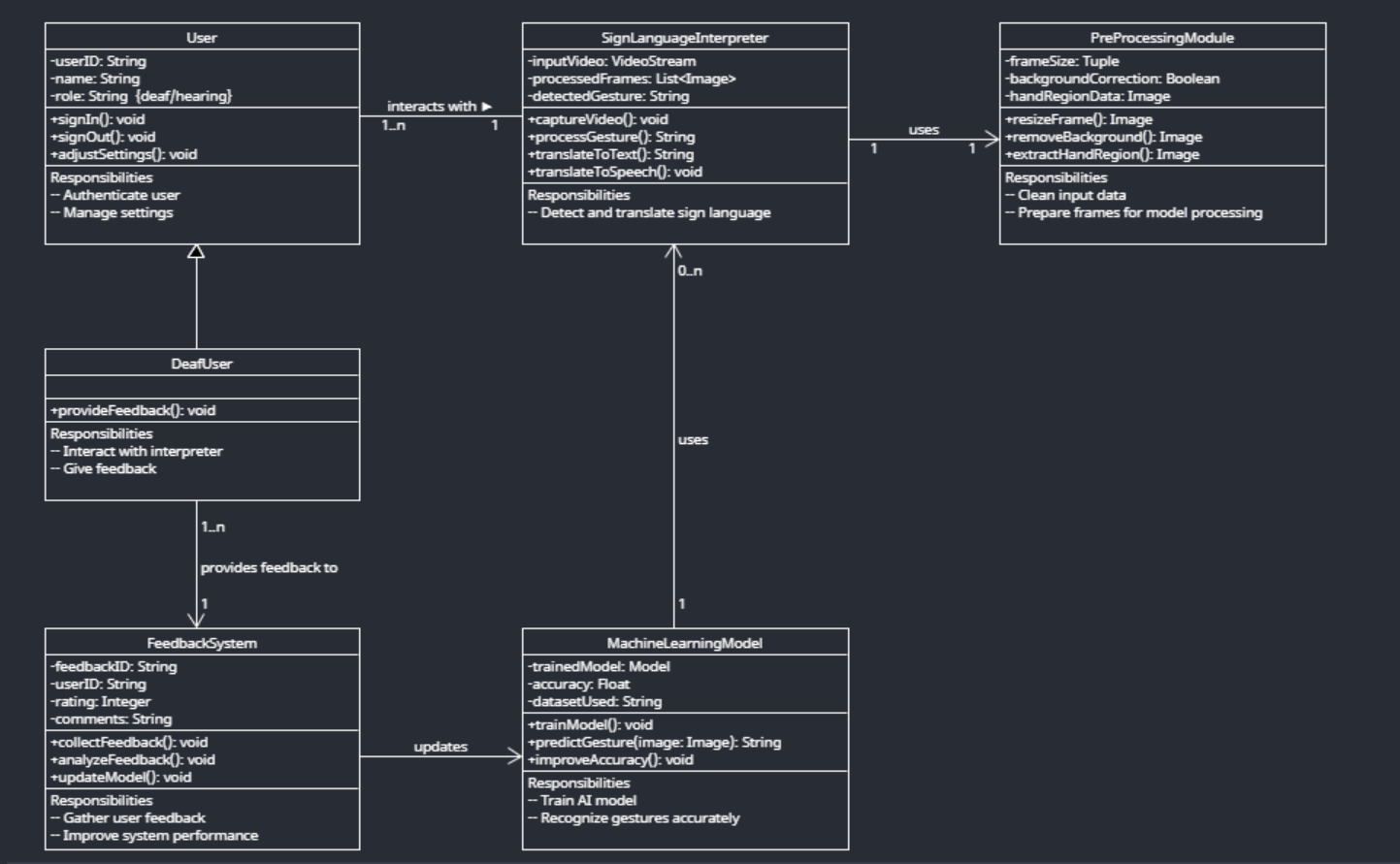


Figure 5: Class Diagram

1. **Sequence Diagram**: Demonstrates the order of interactions during key processes.

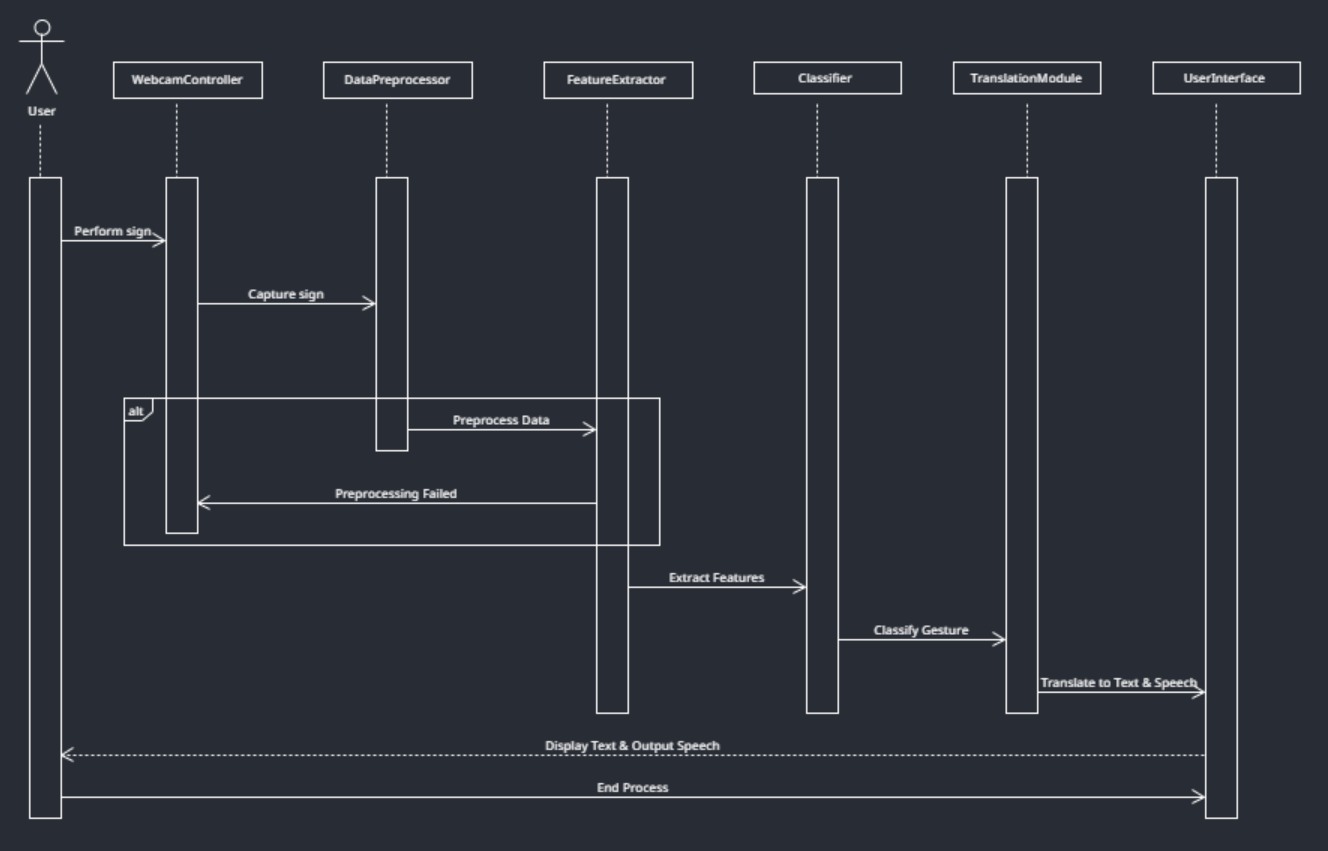


Figure 6: Sequence Diagram

## 3.6 Web-based Integration

The system is designed to be fully web-integrated, ensuring accessibility across different devices. The web interface provides a platform for real-time gesture recognition and translation.

* Web Interface Design: The interface is developed using VueJS to provide a responsive and accessible user experience. It features an interactive dashboard displaying live webcam input and translated text output.
* Real-Time Processing: The backend, built using Python, processes video feeds in real-time and transmits translated gestures to the front end via RESTful APIs or WebSockets.
* User Accessibility: Designed for both deaf and hearing users, the platform includes accessibility features such as text-to-speech output and customizable UI options for ease of use.
* Cross-Platform Support: The web-based approach ensures compatibility with multiple devices, eliminating the need for specialized hardware or software installations.

### 3.6.1 Wireframes

#### 3.6.1.1 Landing Page

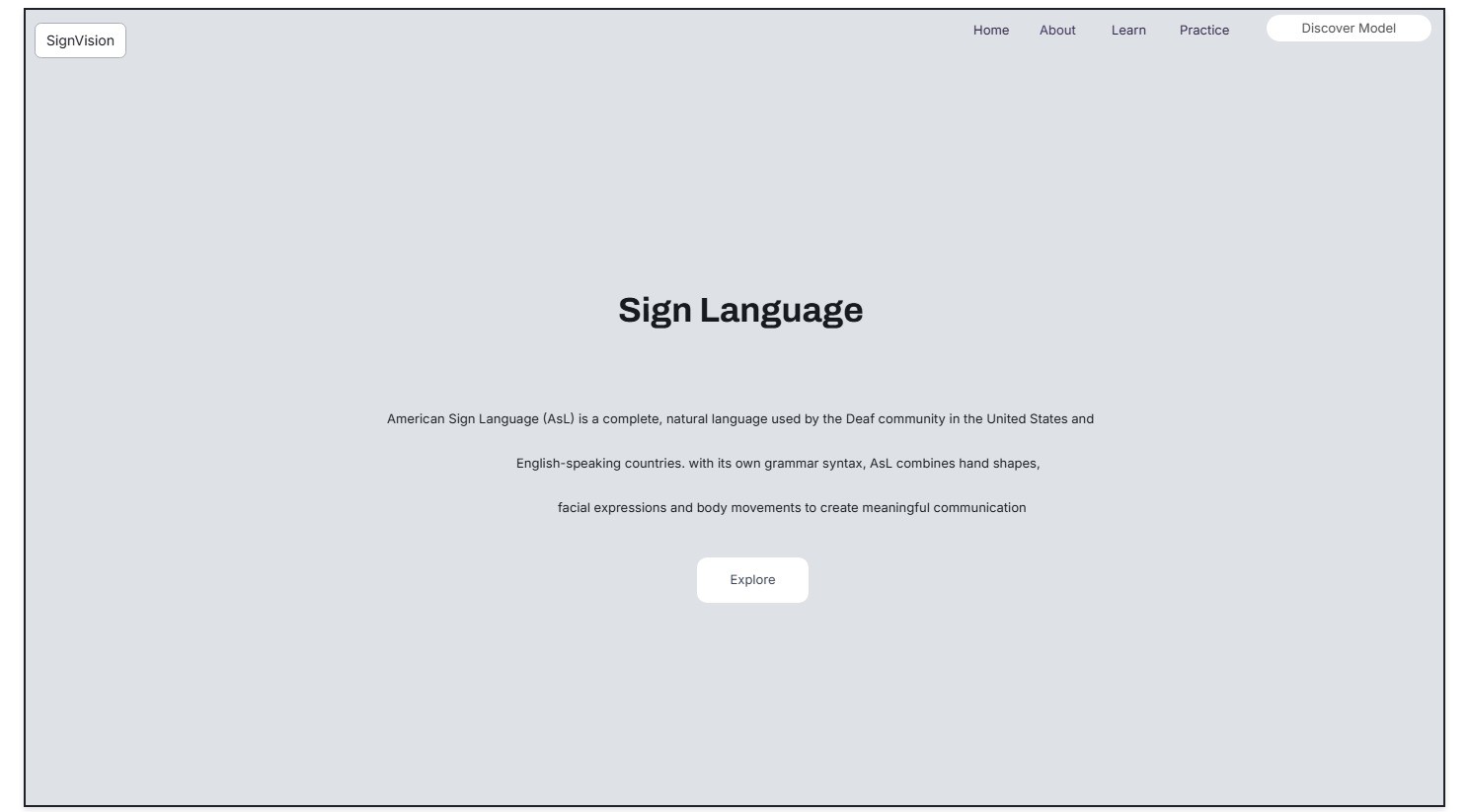


Figure 7: Landing Page

#### 3.6.1.2 About Page

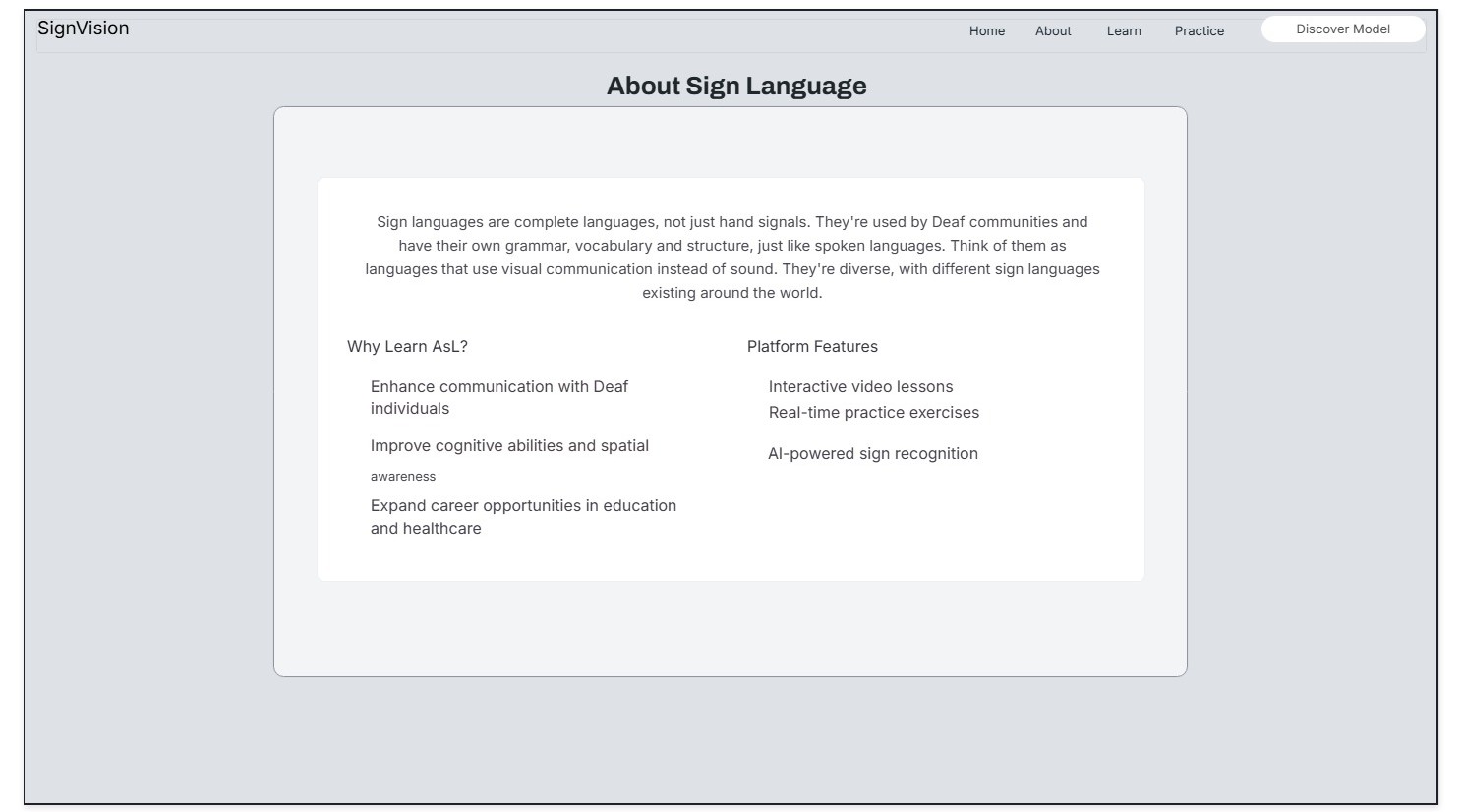


Figure 8: About Page

#### 3.6.1.2 Learn ASL Page

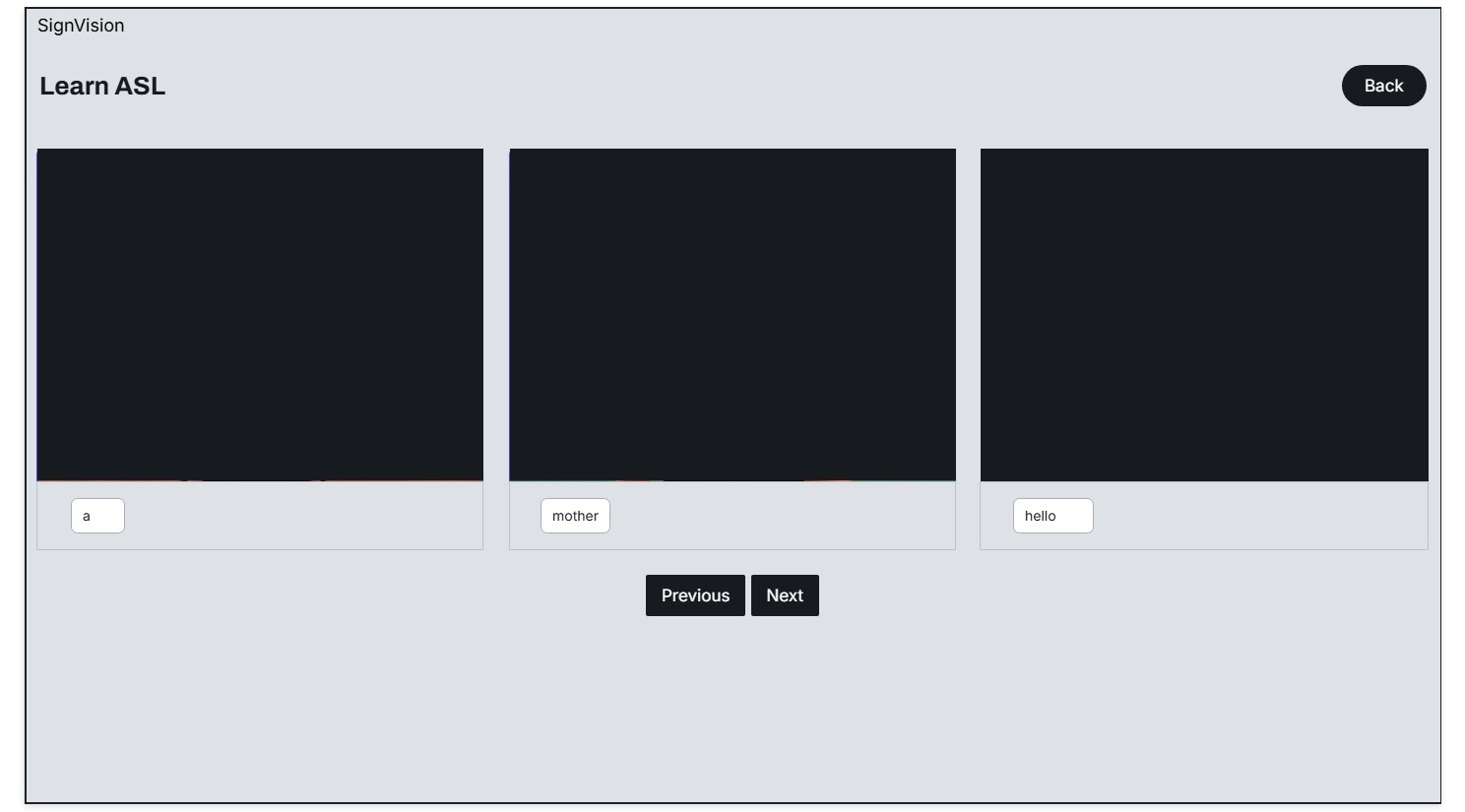
Displays videos

Figure 9: Learn ASL Page

#### 3.6.1.2 Practice ASL Page

### C:\Users\HP\Desktop\website designs\wireframes\Conversion output 2.jpg

Figure 10: Practice ASL

#### 3.6.1.2 Model Feedback Page

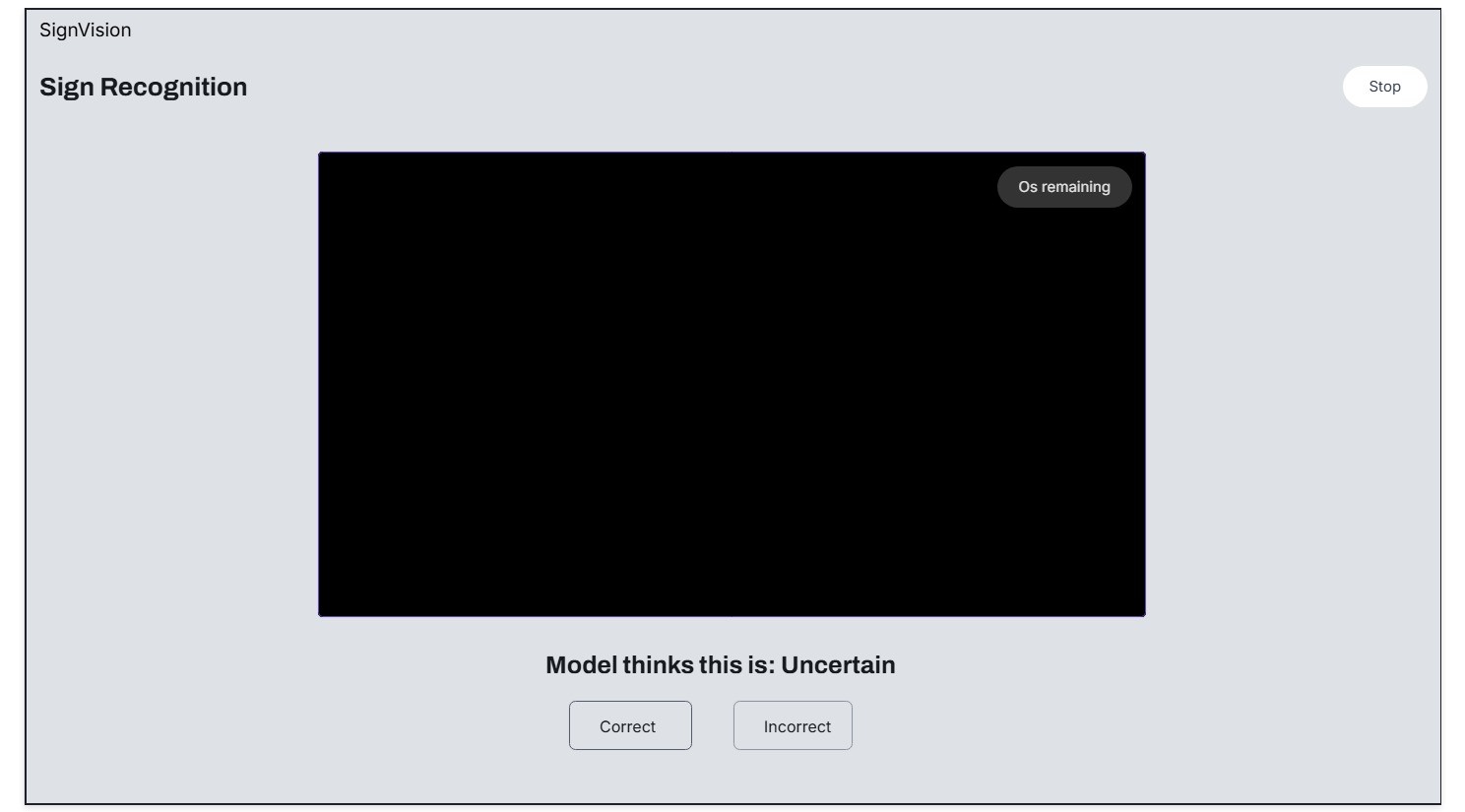


Figure 11: Model Feedback Page

## 

## 3.7 Tools and Technologies

The project uses tools to:

* Design & UI/UX: Figma to build interfaces with responsive user experience.
* Web Development: The front-end framework will be developed with HTML and CSS while the backend will be executed in Python.
* Computer Vision and Deep Learning: OpenCV serves for image processing and PyTorch frameworks serves for the development of gesture recognition models.

# Chapter 4 - Implementation

This chapter discusses how the model and web application were developed. It covers the complete pipeline from data acquisition and pre‐processing to feature extraction, model training, and the integration of the prediction system into a web-based application.

## 4.1 Model Development

### 4.1.1 Data Acquisition and Data Cleaning

#### 4.1.1.1 Dataset Overview

The primary dataset used for this project is the WLASL-Processed dataset, a structured and annotated collection of ASL video samples, sourced from Kaggle (Baskoro, 2024). It is a refined version of the original WLASL dataset.

The dataset is provided as a JSON file, in which each entry contains:

* The gloss (ASL word label);
* a list of video instances for that gloss; and
* metadata including frames per second (fps), bounding boxes (bbox), frame ranges, signer IDs, video source and the video URL.

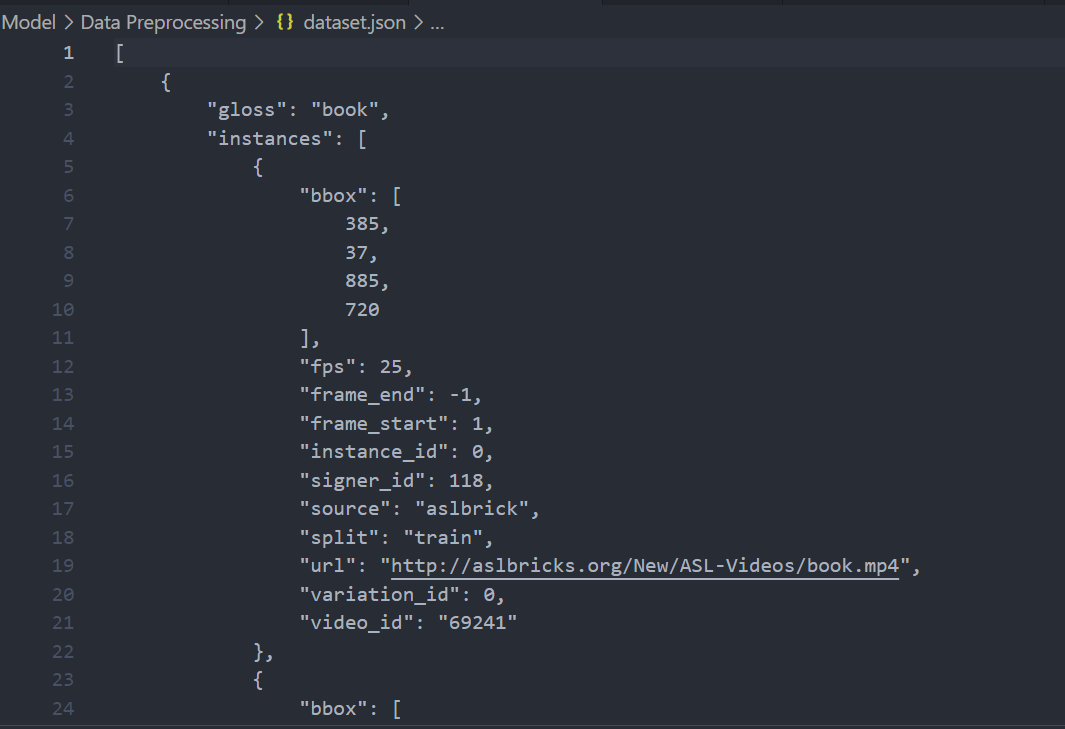


Figure 12: Dataset Overview

#### 4.1.1.2 Data Cleaning

The data cleaning process involves several stages to ensure the dataset's integrity and usability. The following steps are:

1. **Downloading videos and structuring folders**

A Python script was written to automate the download and organization of videos. For each gloss (ASL word), a corresponding directory is created and its corresponding videos are downloaded from the provided URLs. Filenames are sanitized, unsupported formats such as .swf are skipped and failed downloads can be retried up to three times. As a result, the videos are neatly structured into class-specific folders, for further processing and model training.

1. **Filtering video format**

After downloading, videos are filtered to retain only those with the .mp4 extension, as this format is supported and allows for efficient frame extraction using OpenCV.

1. **Removing corrupted videos**

Using OpenCV, each video file is opened and the first 10 frames are read. If the file cannot be opened or if the frames are unreadable, the video is flagged as corrupted and removed from the dataset. This ensures that the dataset contains only valid and readable video files, preventing interruptions during training.

1. **Class Reduction**

The original dataset containes a large number of classes with imbalanced distributions. To improve model performance and reduce noise, classes with very few samples are removed. This results in a reduced dataset with a more balanced class distribution.

1. **Class Label Extraction**

The class labels are extracted directly from the folder names of the reduced dataset. Each folder represents a specific class , a sign. The class names were cleaned by removing underscores, trimming whitespace and saved the final list to a text file. This list is used as class label reference for training and evaluation.

### 4.1.2 Data Pre-processing

#### 4.1.2.1 Video Processing and Keypoint Extraction (Feature Extraction and Classification)

After dataset is cleaned and organized, the video processing pipeline reads the ASL video samples from the organized dataset directory where each folder represents a unique sign and contains several videos.

Using MediaPipe's Holistic model, each video is processed frame-by-frame to detect and extract key landmarks for the face and hands. For every frame, 21 landmarks per hand and 468 facial landmarks are retrieved. However, since this implementation focuses on hand gestures, the emphasis is on hand landmarks and face landmarks are used to find nose region and distances are calculated between nose region and hand landmarks and these features are flattened into a fixed-size array to create consistent features. The keypoints are then normalized by computing distances from a reference facial region (typically the nose) to the hand landmarks, and the resulting values are flattened into a fixed-size array.

Each video is processed into a sequence of 30 such keypoint frames. This fixed-length sequence provides a uniform input for the ConvLSTM2D model, allowing it to capture both spatial features within each frame and temporal dynamics across the sequence. These processed sequences, along with their corresponding class names, are compiled and saved into .npy files and a CSV file for further analysis. This structured format enables effective feature extraction, serving as the input for the ConvLSTM2D model to classify signs based on spatial and temporal dynamics.

### 4.1.3 Data Augmentation Techniques

### 4.1.4 Model Architecture and Training

## 4.2 Web Application Development

### 4.2.1 Design and User Experience/Wireframes and Website Design

### 4.2.2 Learning Functionality

### 4.2.2 Practice Functionality

### 4.2.2 Integration of Model with Front-End and Feedback Collection

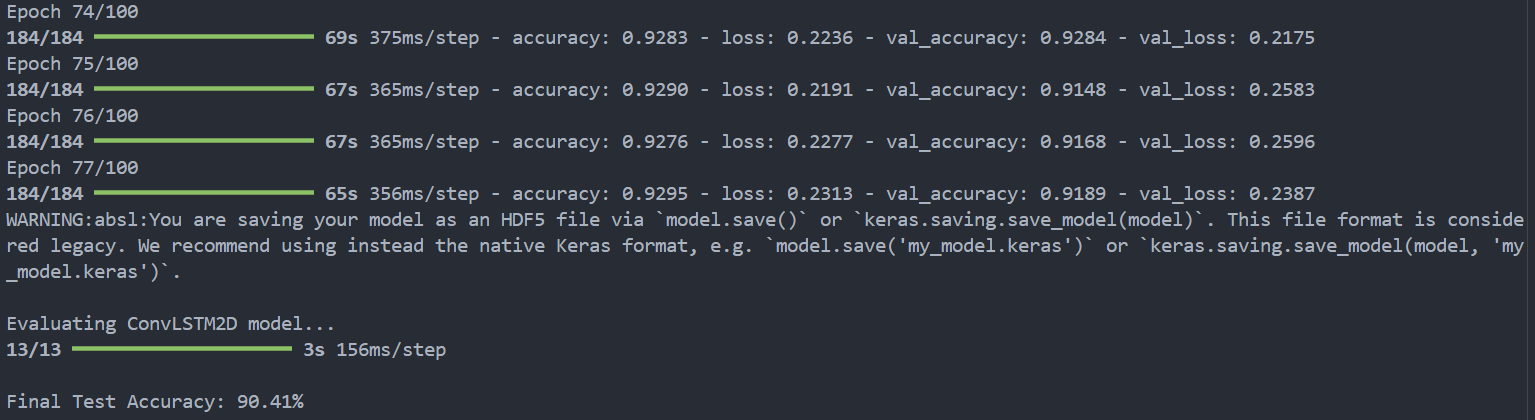
### 4.2.3 Real-Time Predictions

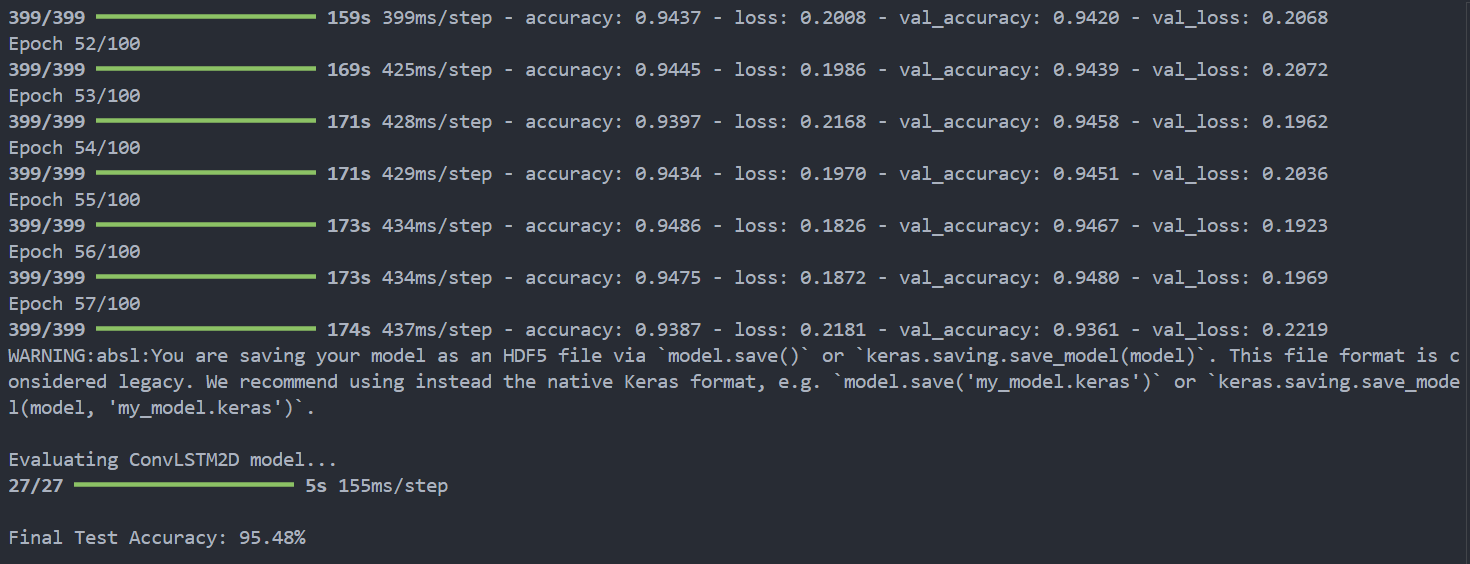
# Chapter 5 – Evaluation Method and Testing

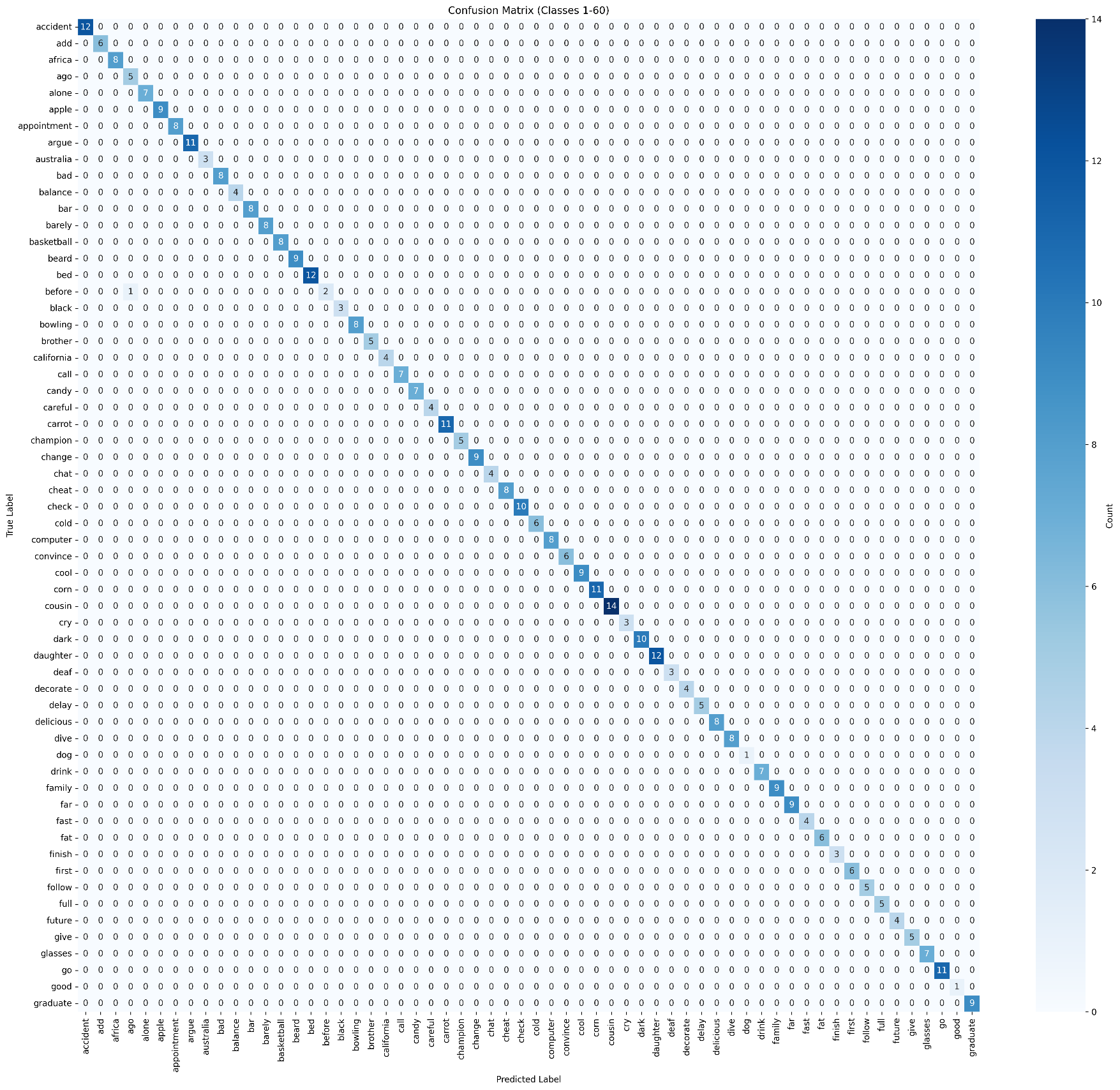
## 5.1 Model Performance Evaluation

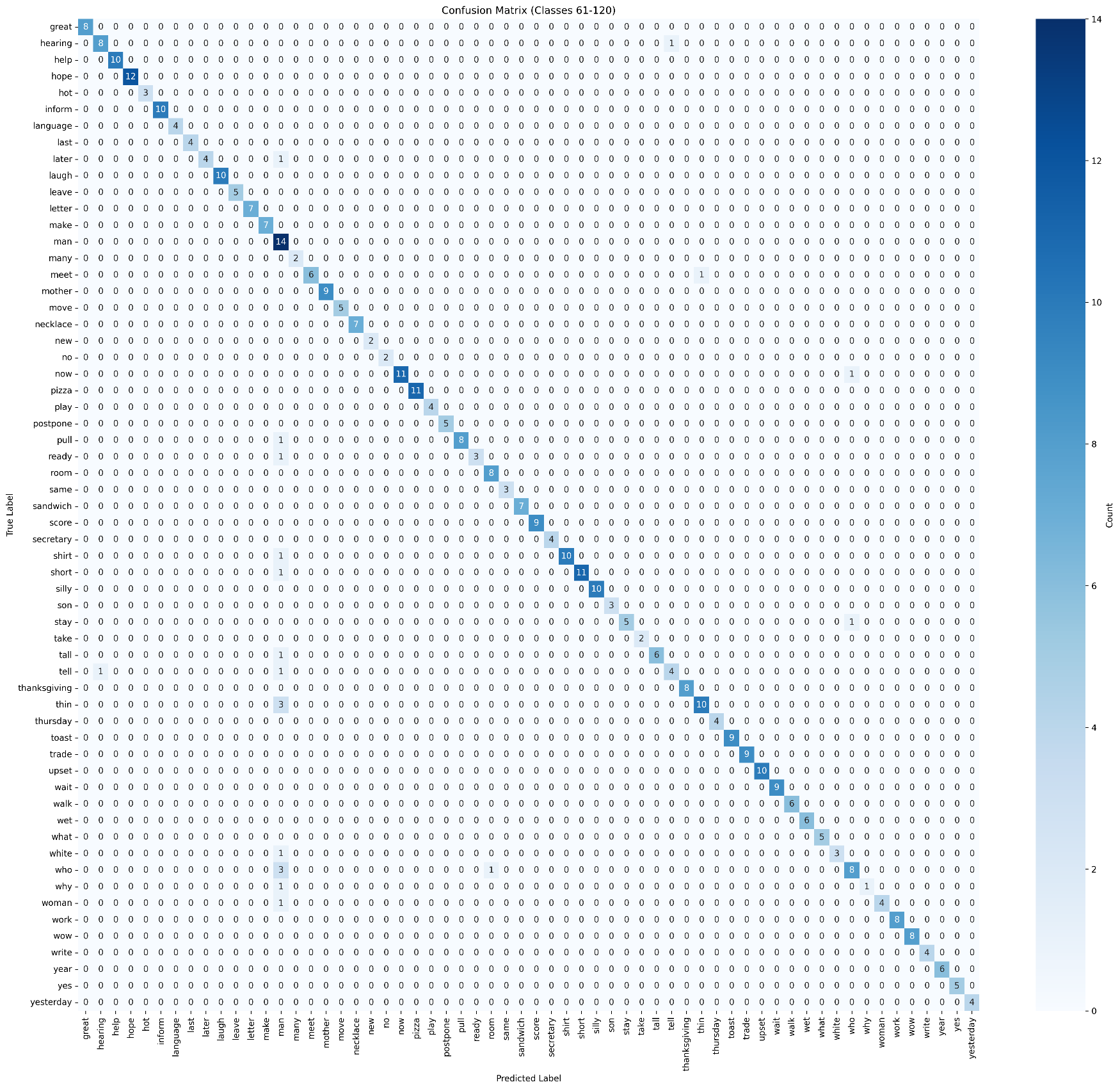
### 5.1.1 Accuracy, Precision Analysis

### 5.1.2 Confusion Matrix Analysis









# Chapter 6 – Results and Discussions

## 

## 6.2 Challenges Encountered

As the project progresses, several challenges may arise that could impact the development and implementation of the AI-driven sign language interpretation system. These potential challenges include:

* Environmental Externalities: Overall changes in lightning, cluttered backgrounds, and alteration in camera positions can make it difficult to accurately recognize gestures. Moreover, their differences in over and under the environment settings will add more issues. Background Identification, Lighting Change Normalization, and Region Isolation are pre-processing steps that can be processed to overcome these obstacles.
* Immediate Reaction Capability: For the practical scope of use, it is essential for there to be real-time processing, or else the sign language recognition model will be inefficient. However, deep learning models can put an excessive load on computational resources, which slower systems will struggle with, especially if dealing with everyday devices. The need to maintain accuracy when optimizing additional computation will force the use of model quantization, more efficient designs, and the addition of GPUs or TPUs for processing power.
* User Variability: The patterns of hand movements are unique to every individual when performing sign language. This diversity has the potential to impact the model’s adaptiveness across different users, often resulting in inaccurate interpretations. The model’s generalizability and accuracy may perhaps be best supported through thorough vetting of various inclusive and robust datasets. Besides, other aspects like hand occlusions (where one hand covers the other hand) and different signer’s angles of body position may need to be considered in the design as well.
* Dataset Limitations: It is quite possible that there is not enough data to support the claim of high-quality datasets being diverse. Most of the available datasets have inadequate variety in signer’s ethnicity, amount of light per signing scene, and signing continuity. The project's scope may need additional data coverage or at the very least, identified strategies for covering the gaps in data per signer to increase the model’s adaptability range, also known as generalization.
* Integration with Web-Based Systems: Adopting the AI sign language interpreter inside a website might bring a whole new level of technical issues like increased latency, server overload, and things like processing videos in real time. Good video hosting and interaction with the remote server tops the list. Instead of the usual HTTP, video can be streamed through WebRTC or WebSockets. On the other hand, a conventional backend might need intelligent memory management for multitasking users.
* Ethical and Cultural Considerations: Ensuring that the system respects the cultural nuances of sign language and the preferences of the deaf community may be challenging. AI-based interpreters must be designed to complement human interpreters rather than replace them. Engaging with deaf users throughout development may be essential to ensure that the system aligns with their needs and expectations.

# Chapter 7 – Conclusions and Future Works

## 7.1 Summary

This report outlines the design, development and evaluation of an AI-driven sign language interpretation system. By addressing limitations in current solutions, such as dataset diversity, reliance on specialized hardware and the challenge of continuous signing recognition, the project aims to deliver a robust, accessible tool for real-time communication.

## 7.2 Additional Features

Continuous sign language recognition (CSLR) and translation have become significant due to innovations in deep learning, particularly through multi-modal fusion techniques, Transformer-based temporal modeling and generative adversarial networks. In spite of these advances, a few challenges remain. By merging insights from the studies conducted, our future research directions can be categorized into the following five key areas:

### 7.2.1 Vocabulary Expansion

Vocabulary expansion is critical for real-world utility. Current systems focus on a limited set of gestures (such as the ASL alphabet or subsets of the MS-ASL dataset); a development is needed to fine-tune larger models on bigger and diverse datasets. This will enhance coverage of a wider vocabulary and inclusion of nuanced gestures and expressions, thereby improving real-world applicability and inclusivity. Consequently, AI-driven sign language systems would become more robust for different linguistic contexts.

### 7.2.2 Multi-Modal Fusion

The integration of multiple data modalities constitutes a crucial space for improving recognition accuracy, especially in noisy environments. In future research, a planned research may include:

* Combine skeletal key-point extraction, optical flow analysis and depth sensors to integrate all useful visual cues.
* Fuse the non-manual cues (facial expressions and body posture) with hand and arm motion modeling.
* Explore fusion strategies in various layers of the network (early, mid and late fusion) to dynamically weight the importance of each modality based on situational conditions.

### 7.2.3 Advanced Temporal Modeling

Sign language has a temporal quality, with fluid movement and overlapping gestures. Future systems should therefore integrate:

* Transformer-based architectures or attention-enhanced recurrent networks to capture long-term dependencies for improved segmentation and interpretation of continuous signing.
* Sequence learning modules, with which further experimentation using Conformer models and cross-modal relative attention could be performed, as evidenced by the work so far, to improve alignment between visual features and glosses.
* Finally, integrate unsupervised pretraining techniques to gain contextual learning, thus improving performance on cumbersome benchmarks such as Phoenix-2014 and Phoenix-2014T.

### 7.2.4 User-Centric Evaluations

The sign language interpreter should, however, primarily be at the service of the deaf community in order to be more functional:

* Extensive usability studies conducted with deaf and hard-of-hearing users to establish practical challenges.
* Regular feedback from users would allow an iterative refinement of the system's performance based on responsiveness, ease-of-use and cultural appropriateness.
* Evaluation of fairness and robustness across multiple signers and demographic backgrounds.

### 7.2.5 Adaptation and Expansion Across Languages

Whereas most currently established systems are designed with ASL in mind, huge possibilities have been opened for adapting these models to the global arena:

* Domain adaptation and transfer learning techniques to develop models applicable to other sign languages.
* Use of common datasets from particular communities (for example, AUTSL for Turkish, Auslan for Australian Sign Language) to enable more inclusive and culturally appropriate systems.
* The proposed fusion approaches are beyond the pure motivation of the two modalities and a consideration of a graph-based cross-modal information fusion would widen the underlying system use and better capture the semantic mapping between different sign language modalities.

By pursuing these integrated research directions, the AI interpreter systems achieve higher accuracy, robustness and diverse real-world applications. An integrated approach, such as through vocabulary, multi-modal integration, temporal dynamic aspects, user feedback and cross-lingual aspects, opens up possibilities for filling in the gaps while promoting cultural inclusivity.

## 7.3 Contributions and Future Work

The project contributes a cost-effective, web-deployable system that integrates state-of-the-art deep learning techniques with ethical, culturally sensitive design. Future work will focus on expanding vocabulary, enhancing multi-modal data integration and conducting user-centric evaluations to ensure the system meets the needs of its target audience.

This review highlights the progress of AI-based sign language interpretation, from early methods to advanced deep learning models. The proposed system, trained on the MS-ASL dataset, aims to recognize hand gestures using a standard webcam, ensuring accessibility without specialized hardware.

A comparison with existing solutions such as SignAll, MotionSavvy and Kinect-based systems underscores the need for real-time performance improvements, better continuous signing recognition and enhanced robustness in diverse conditions. Collaboration with deaf communities remains essential to address ethical and cultural considerations.

Future research can be focused on expanding datasets, refining AI models for continuous signing and integrating multimodal approaches to enhance recognition accuracy. By addressing these challenges, AI-driven sign language interpretation can contribute to greater inclusivity, accessibility and social integration for deaf and hard-of-hearing individuals.

The subsequent phase will focus on:

* Developing the extraction and classification modules to maximize the efficiency of the system with respect to recognition accuracy.
* Modifications in the design of the web interface to make it more user friendly without compromising speed will also be made.
* Augmenting the database with new data and additional pre processing steps for better performance in continuous sign language recognition.
* Real life usage of the system with users to obtain suggestions and improve the system.

# Bibliography

Alyami, S. and Luqman, H. (2024) ‘A Comparative Study of Continuous Sign Language Recognition Techniques’. Available at: <https://arxiv.org/abs/2406.12369> (Accessed: 15 January 2025).

Cerna, L.R., Escobedo Cardenas, E.J., Miranda, D.G., Menotti, D. and Camara-Chavez, G. (2021) *A multimodal LIBRAS-UFOP Brazilian sign language dataset of minimal pairs using a Microsoft Kinect sensor*. Available at: <https://arxiv.org/abs/2008.00932> (Accessed: 01 February 2025).

Hou, J., Liu, Y., and Wu, Y. (2019) *Watch what you sign: Real-time sign language detection on smartwatches.* Available at: <https://dl.acm.org/doi/10.1145/3300061.3300109> (Accessed: 01 February 2025).

Papastratis, I., Chatzikonstantinou, C., Konstantinidis, D., Dimitropoulos, K. and Daras, P. (2021) *Artificial Intelligence Technologies for Sign Language*. Available at: <https://www.mdpi.com/1424-8220/21/17/5843> (Accessed: 07 February 2025).

Papastratis, I., Theodorakis, S., and Maragos, P. (2020) *Adversarial training for sign language translation.* Available at: <https://ieeexplore.ieee.org/document/9191375> (Accessed: 08 February 2025).

Papastratis, I., Theodorakis, S., and Maragos, P. (2021) *Cross-modal alignment for sign language translation: A case study on Greek sign language*. Available at: <https://ieeexplore.ieee.org/document/9414275> (Accessed: 01 February 2025).

Meng, L. and Li, R. (2020) An Attention-Enhanced Multi-Scale and Dual Sign Language Recognition Network Based on a Graph Convolution Network. Available at: <https://doi.org/10.3390/s21041120> (Accessed: 26 January 2025).

Microsoft (2019) *MS-ASL American Sign Language Dataset*. Available at: <https://microsoft.github.io/data-for-society/dataset?d=MS-ASL-American-Sign-Language-Dataset> (Accessed: 18 January 2025).

Tavella, F., Galata, A. and Cangelosi, A. (2024) Bridging the Communication Gap: Artificial Agents Learning Sign Language through Imitation. Available at: <https://arxiv.org/abs/2406.10043> (Accessed: 26 January 2025).

# References

Aboaf, E. (2024) ‘Challenges in AI for Sign Language: Ensuring Cultural Sensitivity and Robustness*, Le Monde’*. Available at: <https://www.lemonde.fr/pixels/article/2024/07/31/les-ia-generatives-peuvent-elles-etre-utiles-aux-personnes-sourdes-et-malentendantes_6263205_4408996.html> (Accessed: 28 January 2025).

Adaloglou, N., Stergioulas, A., Kouris, A., Theodorakis, S., Giannakopoulos, T., and Maragos, P. (2021) *A comprehensive study on sign language recognition methods. IEEE Transactions on Pattern Analysis and Machine Intelligence*. Available at: <https://ieeexplore.ieee.org/document/9675883> (Accessed: 07 February 2025).

Aloysius, N., Geetha, M., Nedungadi, P. (2024) ‘Continuous Sign Language Recognition with Adapted Conformer via Unsupervised Pretraining’. Available at: <https://arxiv.org/abs/2405.12018> (Accessed: 25 January 2025).

Avina, V.D., Amiruzzaman, M., Amiruzzaman, S., Ngo, L.B. and Dewan, M.A.A. (2023) An AI-Based Framework for Translating American Sign Language to English and Vice Versa. Available at: <https://doi.org/10.3390/info14100569> (Accessed: 25 January 2025).

Baskoro, R. (2024) *WLASL-Processed Dataset*, Kaggle. Available at: <https://www.kaggle.com/datasets/risangbaskoro/wlasl-processed> (Accessed: 21 February 2025).

Bragg, D., Koller, O., Bellard, M. et al. (2019) Sign language recognition, generation and translation: An interdisciplinary perspective, Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS). Available at: <https://dl.acm.org/doi/pdf/10.1145/3308561.3353774> (Accessed: 25 January 2025).

Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2018). *Neural Sign Language Translation, In* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition *(CVPR).* Available at: <https://ieeexplore.ieee.org/document/8578590> (Accessed: 08 February 2025).

Chen, H. et al. (2024) SignVTCL: Multi-Modal Continuous Sign Language Recognition Enhanced by Visual-Textual Contrastive Learning. Available at: <https://arxiv.org/abs/2401.11847> (Accessed: 26 January 2025).

Chen, Y., Zhang, X. & Li, L. (2019) Real-time sign language recognition with deep learning: A review. Available at: <https://ieeexplore.ieee.org/document/8744292> (Accessed: 10 February 2025).

Dimitropoulos, K., Daras, P. et al. (2021) Deep Learning Methods for Sign Language Translation, *Sensors*. Available at: <https://www.mdpi.com/1424-8220/21/7/2437> (Accessed: 25 January 2025).

Emmorey, K. (2002) Language, cognition and the brain: Insights from sign language research. Available at: <https://staibabussalamsula.ac.id/wp-content/uploads/2024/03/LANGUAGE-COGNITION-THE-BRAIN-staibabussalamsula.ac_.id_.pdf> (Accessed: 11 February 2025).

Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., and Ney, H. (2014*) Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-Weather*. Available at: <https://www.aclweb.org/anthology/L14-1290/> (Accessed: 01 February 2025).

Galea, C. and Smeaton, A.F. (2019) *Multimodal sign language recognition using deep learning.* Available at: <https://dl.acm.org/doi/10.1145/3311823.3311874> (Accessed: 01 February 2025).

Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning, Cambridge. Available at: <https://www.deeplearningbook.org/> (Accessed: 11 February 2025).

Halvardsson, G., Peterson, J., Soto-Valero, C. and Baudry, B. (2020) Interpretation of Swedish Sign Language using Convolutional Neural Networks and Transfer Learning. Available at: <https://arxiv.org/abs/2010.07827> (Accessed: 25 January 2025).

Jaffe, D.L. (1994) Evolution of Mechanical Fingerspelling Hands for People Who Are Deaf-Blind, Journal of Rehabilitation Research and Development. Available at: <https://www.researchgate.net/publication/234456567_Evolution_of_Mechanical_Fingerspelling_Hands_for_People_Who_Are_Deaf-Blind> (Accessed: 24 January 2025).

Kumar, R., Sinha, A., Bajpai, A. and Singh, S.K. (2023) A Comparative Analysis of Techniques and Algorithms for Recognising Sign Language. Available at: <https://arxiv.org/abs/2305.13941> (Accessed: 26 January 2025).

Koller, O., Forster, J., & Ney, H. (2015). *Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems Handling Multiple Signers*, Computer Vision and Image Understanding. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1077314215000533?via%3Dihub> (Accessed: 26 January 2025).

Ladd, P. (2003) Understanding Deaf culture: In search of Deafhood, *Clevedon: Multilingual Matters*. Available at: <https://www.multilingual-matters.com/page/detail/?k=9781853595455> (Accessed: 11 February 2025).

Li, D., Rodriguez, C., Yu, X. and Li, H. (2020) Word-Level Deep Sign Language Recognition from Video: A New Large-Scale Dataset and Methods Compariso. Available at: <https://ieeexplore.ieee.org/document/9093445/citations#citations> (Accessed: 26 January 2025).

Mindbowser. (n.d) *Step by step process of Agile Scrum methodology*. Available at: <https://www.mindbowser.com/step-by-step-process-of-agile-scrum-methodology/> (Accessed: 01 February 2025).

Mittal, G., Goyal, P., and Kaur, S. (2019) *Sign language recognition using Leap Motion sensor with machine learning*. Available at: <https://www.sciencedirect.com/science/article/pii/S1877050919306623?via%3Dihub> (Accessed: 01 February 2025).

Orovwode, H., Oduntan, I.D. and Abubakar, J.A. (2023) ‘Development of a Sign Language Recognition System Using Machine Learning’, in 2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems. Available at: <https://ieeexplore.ieee.org> (Accessed: 01 February 2025).

Pu, J., Zhou, W., Li, H., and Wang, M. (2016) *Sign language recognition with multi-modal features*. Available at: <https://dl.acm.org/doi/10.1145/2964284.2964319> (Accessed: 07 February 2025).

Russell, S. and Norvig, P. (2021) Artificial Intelligence: A Modern Approach, 4th edition. London: Pearson Education. Available at: <https://www.pearson.com/en-us/subject-catalog/p/artificial-intelligence-a-modern-approach/P200000003500/9780137505135> (Accessed: 11 February 2025).

ScienceDaily (2024) ‘Breaking barriers: Study uses AI to interpret American Sign Language in real-time’, ScienceDaily. Available at: <https://www.sciencedaily.com/releases/2024/12/241216125906.htm> (Accessed: 01 February 2025).

Sincan, O.M. and Keles, H.Y. (2020) *AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods.* Available at: <https://arxiv.org/abs/2008.00932> (Accessed: 01 February 2025).

Sutton-Spence, R. and Woll, B. (1999) The linguistics of British Sign Language: An introduction, Cambridge: Cambridge University Press. Available at: <https://assets.cambridge.org/97805216/37183/sample/9780521637183web.pdf> (Accessed: 11 February 2025).

University of Surrey (2025) ‘SignGPT: AI to Improve Communication for the Deaf and Hard of Hearing’, Hearing Review, 21 January 2025. Available at: <https://www.surrey.ac.uk> (Accessed: 01 February 2025).

Vicars, W.G. (2021) ASL, American Sign Language. Available at: <https://www.lifeprint.com/asl101/pages-layout/lesson1.htm> (Accessed: 25 January 2025).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). *Advances in Neural Information Processing Systems (NeurIPS).* Available at: <https://arxiv.org/abs/1706.03762> (Accessed: 27 January 2025).

Zhang, Y. and Jiang, X. (2024) ‘Recent Advances on Deep Learning for Sign Language Recognition’, Computer Modeling in Engineering & Sciences. Available at: <https://doi.org/10.32604/cmes.2023.045731> (Accessed: 15 January 2025).

Wang, J., Zhou, W., and Li, H. (2020) *A comprehensive survey on deep learning-based sign language recognition*. Available at: <https://ieeexplore.ieee.org/document/9133333> (Accessed: 07 February 2025).

World Health Organization. (2024) Deafness and hearing loss. World Health Organization. Available at: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss> (Accessed: 11 February 2025).

# Appendices

## Research Ethics Screening Form

