

Ai-Based Real-Time Filipino Sign Language Recognition and Translation System with Context-Aware Natural Language Processing

Dharnel Baculo¹, Jellame Formentera², Jeazle Anne Vinzon³, Vivien Agustin⁴, Ronald Fernandez⁵

^{1,2,3}Department of Information Technology, Jesus Reigns Christian College, Philippines

^{4,5}Department of Information Technology, La Consolacion University, Philippines

DOI: <https://doi.org/10.51244/IJRSI.2026.1305000158>

Received: 26 May 2026; Accepted: 31 May 2026; Published: 05 June 2026

ABSTRACT

The Filipino Sign Language (FSL) serves as the official sign language of the Deaf community in the Philippines, as recognized under Republic Act 11106. Despite this recognition, communication barriers persist in educational institutions, healthcare facilities, workplaces, and other settings where Deaf and hearing individuals interact, largely due to the limited understanding of FSL among the hearing majority. Existing sign language recognition applications predominantly target other sign languages such as American Sign Language and are constrained to isolated gesture recognition, failing to address the continuous and dynamic nature of Filipino Sign Language communication. This study developed an AI-Based Real-Time Filipino Sign Language Recognition and Translation System with Context-Aware Natural Language Processing, implemented as an Android mobile application. The system employs MediaPipe Holistic for coordinate-based 3D landmark extraction from hand and body pose, and a Long Short-Term Memory (LSTM) neural network for sequential gesture pattern recognition. A dataset of 20 FSL vocabulary items was collected from seven Filipino Sign Language signers, and a four-technique data augmentation pipeline including Gaussian noise, time shifting, scale variation, and hand mirroring was applied to improve model generalization across different signing styles and dominant hand preferences. The trained model was converted to TensorFlow Lite format for efficient on-device inference without requiring server-side computation or internet connectivity. The system achieved an overall classification accuracy of 97% on the evaluated vocabulary, demonstrating the viability of real-time FSL gesture recognition on consumer-grade mobile hardware. The study adopted a developmental research design with an Agile software development approach to facilitate iterative refinement throughout the development process. The results confirm that the proposed MediaPipe-LSTM architecture provides a practical and accessible foundation for bridging the communication gap between the Filipino Deaf community and the hearing population.

Keywords: Filipino Sign Language, Artificial Intelligence, MediaPipe, LSTM, Real-Time Translation

INTRODUCTION

Background of the Study

Filipino Sign Language (FSL) is the national sign language of the Philippines and is officially recognized under Republic Act 11106, also known as the Filipino Sign Language Act. It serves as the primary language of communication for many Deaf and hard-of-hearing Filipinos, allowing them to express ideas, emotions, and cultural identity through visual-gestural means. Despite its legal recognition, FSL remains largely unfamiliar to the hearing majority. This lack of shared language creates daily challenges in education, healthcare, employment, and social participation, as deaf individuals often struggle to communicate effectively with hearing individuals who do not know FSL.

The communication barriers faced by Deaf Filipinos remain evident in education, healthcare, and employment. In schools, lessons are often missed when interpreters are unavailable or unable to keep pace with the discussion. In healthcare settings, incomplete or misinterpreted messages may lead to misdiagnoses or inappropriate treatments. In workplaces, limited communication can restrict opportunities for hiring and career advancement. These difficulties are further intensified by the shortage of trained interpreters, regional variations in signing styles, and the importance of facial and body expressions, which are sometimes overlooked in existing technological tools. Moreover, most sign language recognition systems are designed for foreign sign languages, particularly American Sign Language (ASL), and are limited to recognizing isolated gestures rather than continuous signing.

To address these challenges, this study aims to develop an AI-based real-time Filipino Sign Language Recognition and Translation System with Context-Aware Natural Language Processing and Adaptive Machine Learning. The system will run on mobile devices with a camera, capturing and recognizing FSL gestures in real time and translating them into readable text and speech. It will also allow hearing users to respond via typed or spoken input, which the system will display as sign animations, enabling two-way communication. By combining gesture recognition with context-aware natural language processing and online-first design, the system seeks to provide a practical, accessible, and private tool that bridges the communication gap between Deaf and hearing individuals.

Statement of the Problem

The primary challenges this research addresses are the serious communication barriers that the Filipino deaf and mute community encounters when engaging with the hearing community. This difficulty is made more difficult by non-signers' lack of understanding of Filipino sign languages.

Specifically, the following problems are observed:

1. Traditional sign language recognition systems are often limited to identifying simple, static gestures, such as individual letters. These systems struggle to recognize dynamic gestures, including verbs, phrases, and sentences, because they cannot effectively capture the sequential and temporal patterns inherent in continuous Filipino Sign Language movements. This limitation reduces the practical usability of such systems in real-life communication.
2. Most existing sign language recognition technologies are optimized for foreign systems, such as American Sign Language (ASL), and lack a comprehensive dataset specifically designed for Filipino Sign Language (FSL). This creates a significant linguistic gap, as these systems fail to recognize the unique hand shapes, facial expressions, and cultural nuances inherent to FSL. Because AI models are limited by the data they are trained on, reliance on Western-centric datasets leads to high error rates when applied to the Philippine context.
3. Many advanced sign language recognition systems require high-end hardware to function effectively, making them difficult to access for students, interpreters, and the general public. This hardware dependency limits the practical usability of such systems in everyday environments, restricting their potential to provide widespread, real-time communication support. There is a clear need for a system that can operate efficiently on portable, low-cost devices without sacrificing performance or speed.

Objective of the Study

The main goal of this study is to make a reliable, real-time system for the recognition and translation of sign language with the integration of AI and machine learning in order to bridge the communication gap between the Filipino deaf/mute community and people in general.

This study specifically seeks to:

1. To design and develop a comprehensive AI-driven application that can accurately interpret both static and dynamic Filipino Sign Language gestures from video input. The system will analyze the spatial features of hand shapes and positions, as well as the temporal sequences of movements, to recognize continuous and complex gestures. By doing so, it aims to improve real-time communication between Deaf and hearing individuals, enhance accessibility, and provide a foundation for creating inclusive, intelligent translation tools that can be applied in educational, social, and professional environments.
2. To organize a high-quality Filipino Sign Language (FSL) dataset that serves as a localized alternative to existing systems based on Western sign languages. This objective focuses on capturing the unique syntax, grammar, and morphology inherent to FSL. By documenting specific hand shapes, movements through space, and facial expressions, the system ensures the AI model is trained on data that is both culturally and linguistically authentic. Establishing this custom data foundation allows the study to reduce technical bias and significantly improve the translation accuracy of the system.
3. To develop and optimize an AI-driven Filipino Sign Language recognition system for deployment on resource-constrained devices, such as mobile phones. The system will be designed to perform real-time gesture recognition with low latency and minimal memory usage, ensuring accessibility, affordability, and practical usability. By creating a lightweight and efficient model, the application aims to enable widespread adoption, supporting daily communication needs and making inclusive technology available to a broader range of users.

Scope and Limitations

This study focuses on the development of an AI-based real-time Filipino Sign Language Recognition and Translation System with Context-Aware Natural Language Processing and Adaptive Machine Learning. The system is designed to bridge communication gaps between the deaf and hearing communities by providing real-time recognition and translation of Filipino Sign Language (FSL). It runs on a cellphone equipped with a camera, using AI-based hand tracking to capture and interpret hand movements without requiring specialized sensors. Its core scope includes real-time gesture recognition, context-aware translation through natural language processing, an adaptive learning mechanism that improves recognition accuracy over time, and bidirectional communication allowing hearing users to type or speak words. The system also includes a voice output feature that converts recognized signs into clear, natural-sounding speech, thereby enabling deaf users to communicate verbally when necessary. It is further designed with an online approach by storing its core dataset and trained model in the cloud.

While the system is fully functional, several limitations must be acknowledged. Its vocabulary is limited to a predefined set of common words, phrases, and alphabet-based spelling due to the challenge of building a complete FSL dataset within the project timeline. Although context-aware NLP is applied to improve translation quality, some outputs may still be literal or grammatically imperfect compared to true FSL syntax. Recognition accuracy can also be affected by lighting conditions, camera quality, or cluttered backgrounds. Performance may vary across users, particularly if signing speed or style differs from the training data, and very rapid signing could cause missed detections. Lastly, the system requires the internet to fully access the application, which may limit usability in remote areas. These limitations will guide future improvements in vocabulary expansion, translation fluency, and overall recognition performance.

Significance of the Study

The importance of this study lies in its contribution to making communication between deaf and hearing individuals more inclusive and efficient. Instead of relying solely on interpreters or written notes, the system enables natural interaction by recognizing Filipino Sign Language in real time and producing clear text and speech outputs. Likewise, it allows hearing individuals to respond through visual sign animations, creating a two-way exchange.

The following groups are expected to benefit from this project:

Deaf and Hard-of-Hearing Individuals. They gain a way to express themselves naturally through FSL while instantly receiving a translated text or speech output, allowing them to participate more fully in conversations without relying solely on interpreters or writing.

Hearing Individuals. They can understand signed messages even if they do not know FSL, which is particularly helpful in schools, workplaces, and public services where communication barriers often occur.

Educators and Advocates. Teachers, trainers, and advocacy organizations can use the system as a teaching aid to increase awareness and support the promotion of FSL in inclusive education programs.

Service Providers and Institutions. Businesses, government agencies, and customer service personnel benefit from a tool that makes their services more accessible and inclusive to deaf clients or customers.

Researchers and Developers. Future scholars and developers can build upon this system, expanding its vocabulary, improving its recognition models, or porting it to mobile platforms to reach a wider audience.

REVIEWS OF RELATED LITERATURE AND STUDIES

Related Literature

The field of Sign Language Recognition (SLR) has undergone remarkable evolution over the past decade, driven by technological innovations in artificial intelligence, computer vision, and machine learning. Wadhawan and Kumar (2021) provided one of the most extensive Systematic Literature Reviews (SLR) in the field, covering a ten-year span and analyzing 117 selected papers across 25 different sign languages. Their review classified research efforts based on data acquisition methods, signing modes (static or dynamic), and classification techniques. The findings revealed that most of the existing studies have concentrated on static, isolated, and unimanual signs captured by camera-based systems. Despite significant progress, the review highlighted that a universal and generalizable system capable of handling continuous and real-time recognition suitable for commercial deployment remains unrealized. The authors emphasized that future research should focus on developing multimodal recognition systems that combine visual and non-manual signals such as facial expressions, body postures, and head movements, which are crucial components of natural sign communication.

This evolutionary trajectory is further traced by Galván-Ruiz, Travieso-González, Tejera-Fettmilch, Pinan-Roescher, Esteban-Hernández, and Domínguez-Quintana (2020), who reviewed the chronological development of SLR technologies from early sensor-based systems to the modern vision-driven approaches. Their historical overview showed that early sign recognition relied heavily on data gloves and motion sensors for capturing hand movements, which, while accurate, were often expensive and restrictive for users. The transition to vision-based systems marked a significant milestone, leveraging advancements in image processing and deep learning. Galván-Ruiz et al. emphasized that improvements in feature extraction, noise reduction, and handling of spatial and temporal data complexities have been central to achieving higher accuracy and usability in SLR systems. They further noted that deep learning's adoption marked the turning point toward non-contact, camera-based applications that better mimic natural communication.

Complementing these reviews, Toshpulatov, Lee, Jun, and Lee (2025) conducted an extensive analysis of the broader Sign Language Processing (SLP) field, which encompasses recognition (SLR), translation (SLT), and production (SLPn). Their study explored the principal deep learning methodologies that propel automation in SLP, such as sequence-to-sequence and neural machine translation models. Importantly, they highlighted the increasing relevance of multimodal data integration, emphasizing that accurate and context-aware sign translation requires systems capable of interpreting visual signs in conjunction with facial expressions and gestures. Toshpulatov et al. also discussed ongoing efforts to develop expressive sign production systems

capable of generating natural signing movements, signifying the next step toward full-scale human-machine communication.

A significant body of research has shifted from static or isolated sign recognition toward Continuous Sign Language Recognition (CSLR), which involves understanding entire sign sentences. Ibrahim, Zayed, and Selim (2020) presented an in-depth examination of CSLR systems, outlining their complexities compared to isolated sign recognition. The authors described that CSLR faces unique challenges, primarily the segmentation of continuous gesture streams, where one sign transitions fluidly into another. Furthermore, they identified difficulties in converting visual-spatial grammar into grammatically correct spoken or written language. The study revealed several barriers to CSLR advancement, including the lack of standardized and diverse datasets for continuous signs, limited models that can accurately capture temporal dynamics, and persistent challenges in integrating non-manual elements such as facial expressions, mouth shapes, and body postures. Ibrahim et al. concluded that the absence of such multimodal data hampers the development of naturalistic and reliable CSLR systems.

While most modern SLR systems rely on camera-based input, sensor-based technologies have been explored as reliable alternatives, especially in controlled environments. Saeed, Zainol, Zaidan, and Alamoodi (2022) performed a systematic review of sensor-based systems using gloves embedded with flex, inertial, and bend sensors for Sign Language Pattern Recognition (SLPR). Their findings showed that these systems achieved high precision in recognizing isolated signs and minimized environmental interference such as poor lighting conditions. However, they also identified key limitations: the high production cost, user discomfort due to wearable hardware, and the inability to capture facial expressions and body gestures that play a crucial role in sign language grammar. This limitation underscores the trade-off between technical accuracy and natural usability in sensor-based systems.

Related Studies

The Sign Language Transformer represents the first effort to unify continuous sign language recognition and sign language translation within a single end-to-end deep learning framework. Unlike traditional systems, which sequentially process recognition followed by translation, often suffering from compounded errors, this model jointly optimizes both tasks, consequently improving the translation accuracy and contextual fluency. The primary objective was to develop a system that can understand, in one go, the visual gestures along with generating linguistically correct textual translations. The researchers employed CNNs for extracting features from sign videos and attention-based encoder-decoder layers for modeling the sequential and contextual dependencies associated with signing. The model was trained and evaluated on the large-scale PHOENIX-2014T dataset, which involves thousands of continuous German Sign Language sequences along with their aligned textual translations. Performance metrics demonstrated substantial improvements over prior methods, including higher BLEU scores and reduced WER, thus underlining the model's ability to generate fluent and coherent translations even for complex sets of gestures or transitional movements. To be noted, this study demonstrated that integrating recognition and translation into one framework preserves temporal continuity and reduces error propagation, thereby overcoming one of the critical limitations faced by earlier studies. While the given approach does rely on large annotated datasets, its conceptual design lays the foundation for developing context-sensitive translation systems wherein gesture continuity and linguistic coherence are of essence. In practical terms, the methodology offers a roadmap for adapting similar frameworks to other sign languages, such as Filipino Sign Language (FSL), where rich facial expressions and body postures, together with hand gestures, constitute an important means of effective interpretation. By combining convolutional feature extraction with attention-based sequence modeling, the Sign Language Transformer constitutes a new benchmark against which AI-driven FSL recognition and translation models can be developed, informing their design toward robustness and a capacity to deal with the subtleties of temporal dynamics associated with continuous signing. Camgoz et al. (2020)

Building on the basis of end-to-end sign language translation, the STMC (Spatial-Temporal Multi-Cue) Transformer was designed to address the nuanced interplay of visual and temporal information in continuous sign language recognition. Fully realizing that gestures convey meaning not only by hand shape but also via

body posture, facial expression, and motion dynamics, the study aimed to enhance the model's capability in capturing these interdependent cues. This architecture integrates spatial-temporal attention mechanisms within a Transformer framework, thereby having the capability to analyze not only the spatial configuration of the signer's body but also the temporal progression of gestures across frames. This allows the system to keep continuity and context, which is particularly critical for translating long, fluid sequences into natural language. Evaluation on PHOENIX-2014T demonstrated that it yielded significant improvements in BLEU scores, more than five points higher than a standard Transformer model. It manifests that embedding spatial-temporal cues enables much smoother and more accurate sentence translation, especially in the case of overlapping or rapid movement. The STMC Transformer reduces such errors caused by fragmented gesture recognition by preserving temporal flow and contextual relationships, an important limitation with earlier models that treated frames largely independently. These results prove the practical significance of real-time sign language translation modeling continuity of motion and multimodal cues is necessary for accurate linguistic output. In the case of Filipino Sign Language, or FSL, which relies highly on synchronized movements of the hands, face, and body, adoption of a similar spatial temporal approach improves both recognition and translation fidelity. Though computation and large annotated datasets are needed, its success underlines the importance of architectures that are contextually aware for continuous gesture interpretation. By capturing the multi-layered nature of signing, the STMC Transformer provides a valuable framework from which to develop robust, natural, and context sensitive FSL translation systems, forming a critical bridge between foundational end-to-end models and more advanced, multimodal recognition tools. STMC Transformer Yin et al. (2020)

Zhang et al., 2020, came up with MediaPipe Hands, an on-device, real-time hand landmark detection system for recognizing sign language. The primary objective was to realize a lightweight framework that captures minute hand motion but does not depend on specialized sensors or high end hardware a positive attribute for mobile and low resource applications. The system puts into practice a two-phase approach a palm detection network followed by a landmark regression model to detect and track 21 key points on each hand, ensuring reliable performance even under partial occlusion, variable lighting, or rapid movement. By processing standard RGB video input, MediaPipe Hands eliminates the need for depth cameras or gloves, significantly reducing the cost and complexity of deployment. Methodologically, deep convolutional neural networks optimize for both speed and precision, enabling real-time performance on smartphones, tablets, and consumer-grade computers. Evaluations revealed high accuracy and low latency, enabling the model to serve as a dependable preprocessing module for downstream recognition and translation frameworks. MediaPipe Hands provides the necessary motion and position information critical for enhancing gesture interpretation, continuity, and overall translation fidelity when integrated with CNN, LSTM, or Transformer architectures. In this respect, the added value of the research lies in the fact that it is practical by providing accurate hand landmark tracking on affordable devices. Such technology enables scalable and accessible AI-driven sign language systems. In the context of FSL, MediaPipe Hands can supply reliable inputs for models interpreting complex hand shapes and dynamic gestures, which are indispensable for the precise recognition and context-sensitive translation of FSL. The efficiency and open-source availability of the approach, although by no means directly related to the processing of facial and body cues, make it a foundational tool for creating lightweight, mobile-compatible FSL recognition systems. MediaPipe Hands bridges the gap between laboratory-grade recognition models and real-world, everyday applications, supporting educational tools, mobile translators, and assistive technologies for the Deaf community.

To further improve continuous sign language recognition, Li and Meng [2022] proposed a Multi-View Spatial-Temporal Network that incorporates information from different perspectives, including skeletal joint data, raw video frames, and motion energy maps. The study aimed to address the challenge of ambiguous or overlapping gestures, which often lead to misclassification in single-view recognition models. By fusing these complementary inputs, the network captures both spatial configurations and temporal dynamics simultaneously, allowing for a more comprehensive understanding of the signer's movements combined with contextual cues. The methodology involves the processing of each input modality through dedicated network branches, further aligned with a fusion mechanism that aligns spatial and temporal information for final classification. Experiments on continuous sign language datasets have demonstrated that multi-view processing significantly reduces recognition errors compared to single-view approaches. The model excels in terms of distinguishing between complex gestures with overlapping motions and subtle positional variations,

showing improved performance both in terms of accuracy and temporal consistency. This work underlines the importance of contextual learning since analyzing body motion in combination with its appearance enables the system to resolve ambiguities that could otherwise compromise translation quality. This is of great consequence for FSL. Many FSL signs require simultaneous hand, body, and facial movements that are essential to convey meaning. The adoption of multi-view spatial temporal analysis enhances the accuracy of models, especially on intricate and/or high-speed sequences, to ensure that translations convey semantic content along with nuanced gesture patterns. Increased computational resources are needed because of the multimodal processing this model requires, but it sets up a basis upon which context aware FSL translation systems can be designed to interpret human gestures with higher fidelity. By integrating multiple streams of spatial and temporal information, this study provides a sound foundation for AI-driven FSL recognition tools that focus on accuracy, naturalness, and real-world applicability furthermore, it increases the accessibility and communication of the Deaf community.

Building upon prior work in multimodal sign language recognition, Hu et al. (2024) introduced a correlation-aware Transformer called CorrNet+, which is designed to grasp the complicated interactions among hand gestures and facial and body postures within continuous signing. The objective has been to deliver an improvement in both recognition and translation accuracy through the modeling of spatial-temporal correlations across several motion channels while ensuring the subtlety of contextual and emotional cues is preserved. Most conventional models process these channels independently from one another. This typically results in a loss of both linguistic and expressive information, which is critical for translation accuracy. CorrNet+ integrates into its architecture correlation attention layers that track how different modalities develop over time in synchrony, so the network can interpret gestures together with accompanying facial expressions and body movements. This approach was evaluated on benchmark datasets, demonstrating substantial increases in recognition precision and BLEU-based translation scores compared with standard Transformer architectures. By capturing the interplay between multiple motion streams, the method CorrNet+ effectively models important expressive nuances and grammatical markers for natural sign language understanding. That study carries direct implications for Filipino Sign Language, wherein meaning is most often conveyed through a combination of hand gestures, facial expressions, and postural shifts. This correlation aware mechanism can elevate AI-driven FSL systems in terms of recognizing tonal, emphatic, and syntactic cues for truly accurate and contextually faithful translations. While the model's complexity necessitates large amounts of computational resources, its design does indicate the importance of synchronized multimodal analysis for the recognition of continuous gestures. CorrNet+ not only improves the fidelity of translations but also provides a conceptual framework for developing advanced, context-sensitive FSL translators that can handle the expressive richness and dynamic variability present in real-world signing. This is an important step toward enabling AI systems to understand sign languages with nuanced, human level capability across diverse linguistic and cultural contexts.

De La Salle University (2022) investigated the implementation of AI-driven gesture recognition on a Raspberry Pi, considering low-cost hardware as an approach to making Filipino Sign Language recognition systems more accessible and affordable. The objective of the study was to show that real-time FSL recognition is achievable on resource-limited hardware, independent of high-end GPUs or specialized devices, with functional accuracy. For this purpose, optimized Convolutional Neural Network architectures that could process live video input from a standard webcam were considered. First, the CNN is trained on the FSL dataset consisting of static and simple dynamic gestures. Then, the system was set up to function effectively on the Raspberry Pi by performing live gesture recognition with very minimal latency and power consumption. Testing showed that satisfactory levels of accuracy were achieved for the basic FSL signs, and therefore established the capability to adapt advanced recognition models in low-cost portable hardware. That is particularly valuable for deployments in schools, learning centers, and community spaces where high-cost equipment is not available, making technology inclusive for a broader range of users. This study holds particular significance due to the themes of inclusivity and practical usability. The study is able to establish a model for the deployment of portable educative tools and assistive technologies supporting DHH learners by proving that AI-based FSL recognition can operate effectively on low-cost devices. Although the system handled mainly simpler gestures and may be subject to further enhancements for complex, continuous signs, it lays the groundwork for future development of lightweight and context-aware FSL translators capable of real-

world applications. For Filipino Sign Language, this study has underlined the potential of combining AI with affordable hardware, advancing technological inclusivity, and guaranteeing communicational and learning devices accessible to all sectors of society.

METHODOLOGY

This chapter outlines the research methods and procedures applied in the design, development, and evaluation of the AI-Based Real-Time Filipino Sign Language Recognition and Translation System. It explains how the study systematically approaches the creation of a functional application that can accurately recognize and translate Filipino Sign Language gestures in real time. The methodology is structured to ensure that every phase of the project, from data collection to system testing, is organized, consistent, and reliable. It also highlights the tools, technologies, and processes employed to achieve a system that meets both technical and practical requirements. For clarity, diagrams are provided to visually represent the workflow, data processing, and system architecture, showing how each component contributes to the overall development and evaluation of the application.

Research Design

This study uses a developmental research design, which centers on creating, testing, and improving a practical product to address a real problem. Developmental research focuses on producing a usable solution and examining how well that solution works under real conditions; it combines design activities such as planning and building the system with systematic evaluation through testing its performance and usability. This approach is suited to projects where the primary goal is to produce a working application and to verify that the application meets its intended objectives.

The developmental design was chosen because this study aims to develop a functional mobile application that recognizes and translates Filipino Sign Language and to verify its performance with real users. Developmental research supports the collection of actual data about the system's effectiveness, including recognition accuracy and user satisfaction, while the proponents make measured improvements based on test results and user feedback, particularly from members of the deaf community. This ensures the final product is both technically sound and practically useful for the intended users.

The development of the gesture recognition component requires extensive datasets consisting of videos that capture Filipino Sign Language (FSL) gestures. These datasets are crucial because the performance and accuracy of the recognition model largely depend on the quality, diversity, and representativeness of the input data. To ensure these requirements are met, the datasets will be sourced primarily from open-access repositories such as Mendeley Data, which provide publicly available gesture or sign language datasets specifically intended for research and educational purposes. Additional datasets may also be obtained from other reputable open-source platforms to supplement the range of gestures and enhance the system's capability to recognize variations in signing.

The datasets will include a wide array of gestures representing common FSL words and phrases, capturing different hand orientations, positions, and movements. Including gestures from multiple users is an important consideration, as it allows the model to generalize across different hand shapes, sizes, and individual signing styles, thereby improving recognition accuracy in real-world scenarios. Variations in lighting conditions, background, and camera angles present in the datasets will further contribute to the robustness of the system, enabling it to perform effectively in diverse environments.

By utilizing open-source datasets, the development process is more efficient and cost-effective, as it eliminates the need for large-scale manual data collection while still providing high-quality and diverse data. However, if specific gestures required by the system are not available in these repositories, supplementary data will be collected manually. This manual collection will involve capturing videos of gestures performed by volunteers, ensuring that all necessary gestures are adequately represented. Overall, careful selection and acquisition of

these datasets form a critical foundation for the successful implementation of gesture recognition, setting the stage for subsequent processes such as data preprocessing, model training, and validation.

For the development of the AI-Based Real-Time Filipino Sign Language Recognition and Translation System, the Agile software development model was adopted. Agile is an iterative and flexible approach to software development that emphasizes collaboration, adaptability, and incremental delivery of functional components. Unlike traditional models such as Waterfall, which follow a strict linear sequence, Agile allows the development team to continuously plan, build, test, and improve the system in small cycles called sprints (General Accountability Office, 2020).

Figure 1: Agile Method



Agile is particularly suitable for this study because it allows frequent testing and refinement of the recognition model, which is essential for ensuring accurate real-time translations of Filipino Sign Language (FullScale, 2025). The methodology encourages continuous engagement with users, allowing the proponents to incorporate feedback from members of the deaf community throughout development, which helps make the system both practical and user-friendly. It also provides the flexibility to adapt requirements as the study evolves, which is crucial for AI systems where performance improvements and new feature needs often arise as more data is collected and analyzed. Agile’s incremental delivery approach enables proponents to implement functional components progressively, reducing development risks and making it easier to identify and fix issues early. In addition, the collaborative nature of Agile fosters better communication and coordination among proponents, ensuring that all aspects of the study remain aligned with its objectives. These characteristics make Agile a more effective and suitable SDLC model for developing a real-time, AI-driven Filipino Sign Language recognition and translation system compared to traditional sequential models.

The following are the stages of the Agile Model:

Planning

The planning phase involved identifying the goals, requirements, and scope of the AI-Based Real-Time Filipino Sign Language Recognition and Translation System. The proponents defined the system’s objectives, such as recognizing specific FSL gestures accurately and translating them into text in real time. During this phase, the target users, primarily members of the deaf community, were considered to ensure the system

addresses real-world needs. Resources such as open-source FSL datasets, AI frameworks, and development tools were identified, and a roadmap for the overall development process was created, outlining timelines and deliverables for each sprint.

Design

In the design phase, the system's architecture was developed, specifying how each component would function and interact. This included designing the user interface to be intuitive and accessible, planning the database structure to store gesture data and translations, and selecting the AI model architectures. Additionally, the data preprocessing pipeline was designed, detailing how FSL videos would be standardized and labeled for effective training of the recognition model.

Development

During the development phase, the system was built incrementally according to the design specifications. The AI models were trained using the preprocessed FSL datasets, and the application's user interface and backend were implemented in parallel. Functional modules, such as gesture detection, real-time translation, and text-to-speech, were developed in small, manageable increments to allow for iterative testing and refinement.

Testing

Testing was conducted continuously throughout development to ensure system reliability and accuracy. This included validating the AI model's recognition performance using reserved test datasets, evaluating translation correctness, and conducting usability tests with sample users from the deaf community. Errors, misclassifications, and interface issues were documented and addressed promptly. Both automated and manual testing strategies were applied to confirm that each component met the intended objectives.

Deployment

After testing, the system was deployed as a mobile application capable of processing real-time video inputs and providing accurate translations online. Deployment involved integrating all functional modules into a mobile-friendly environment, optimizing AI model performance for smooth operation on mobile devices, and using cloud-based services to handle computation-intensive tasks when needed. This setup ensures that users can access the application anytime with stable performance and receive real-time translation of Filipino Sign Language gestures.

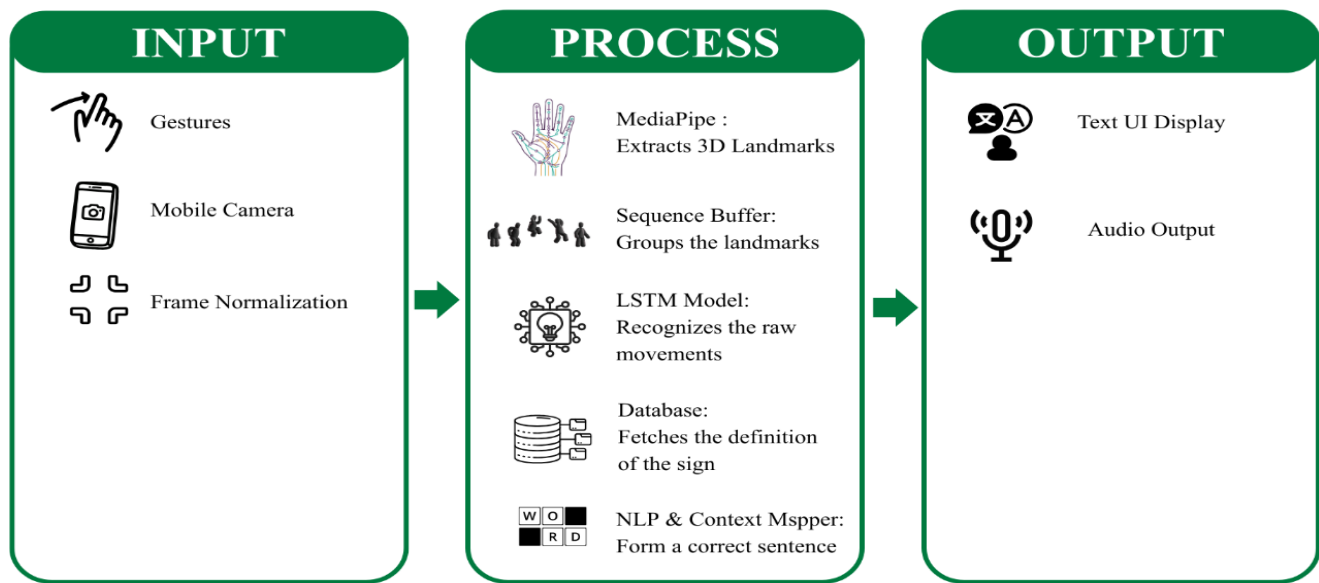
Review

The review phase focused on evaluating the system's overall performance and usability. The team analyzed recognition accuracy, translation correctness, system responsiveness, and user satisfaction based on feedback from initial testing. This phase helped identify areas requiring further refinement, including adjustments to the AI model, interface improvements, and enhancements to real-time processing capabilities.

Iteration

Iteration represents the cyclical nature of Agile development, where feedback from the review phase informs subsequent planning, design, and development. The system underwent multiple iterations, retraining models with improved data, refining algorithms, adjusting the interface, and incorporating additional features. This continuous improvement process ensured that the final product was both technically robust and practically useful for the deaf community.

Figure 2: System Architecture



The architectural framework of the system follows an Input-Process-Output (IPO) model shown in Figure 2, designed to facilitate a bidirectional communication bridge between the Deaf and hearing communities.

The initial phase of the system architecture focuses on data acquisition and environmental adaptation to ensure the high quality of the visual feed. The process begins with the user, who performs Filipino Sign Language (FSL) gestures within the field of view of the mobile camera. This camera serves as the primary hardware sensor, capturing live RGB video frames at a high frequency. To address real-world challenges such as varying room brightness or the user's distance from the lens, the system passes these raw frames through a frame normalization stage. In this sub-process, the image is standardized, adjusting the lighting and scaling the hand size so that the internal AI engine receives a consistent, clean input. This layer is critical because it removes background noise and prepares the visual data for mathematical extraction, ensuring the system remains functional regardless of the user's physical setting.

For the process, as the most complex component of the architecture, it acts as the system's cognitive engine, where mathematical coordinates are converted into human language. The workflow begins with MediaPipe, which executes 3D landmark extraction. Rather than processing the entire video frame, MediaPipe identifies 543 specific key points on the user's face, hands, and body, converting them into a numerical coordinate map. These coordinates are then fed into a sequence buffer, a temporal window that stores a specific number of consecutive frames. This buffering is essential because sign language is not a static image but a continuous movement; the buffer allows the AI to see the entire trajectory of a sign from start to finish. Once the sequence is complete, it is analyzed by the LSTM (Long Short-Term Memory) model, a specialized recurrent neural network designed to recognize patterns in time-series data. The LSTM evaluates how the landmarks move relative to one another to identify the raw FSL gloss or label. After a gesture is recognized, the system queries a database to fetch the direct meaning and metadata associated with that specific sign. However, because FSL has its own unique grammar that differs from spoken Filipino or English, a final NLP (Natural Language Processing) & context mapper is employed. This layer uses the facial expression data from MediaPipe and the word labels from the database to perform context-aware mapping. It rearranges the raw signs into a grammatically correct sentence, adding tense, tone, and punctuation, thereby transforming a series of raw movements into a sophisticated, readable message.

The output layer manages the delivery of the processed information through multimodal channels to facilitate communication. The primary output is the text UI display, which renders the finalized, grammatically correct sentence on the mobile application's interface for immediate reading. Simultaneously, the system executes the audio speech component, which utilizes a Text-to-Speech (TTS) engine to convert the translated text into

synthesized vocal output. This dual-output approach ensures that the hearing participant receives the message through both visual and auditory formats. By providing simultaneous text and audio, the architecture completes the communication bridge, allowing for a seamless exchange of information between the deaf signer and the hearing user.

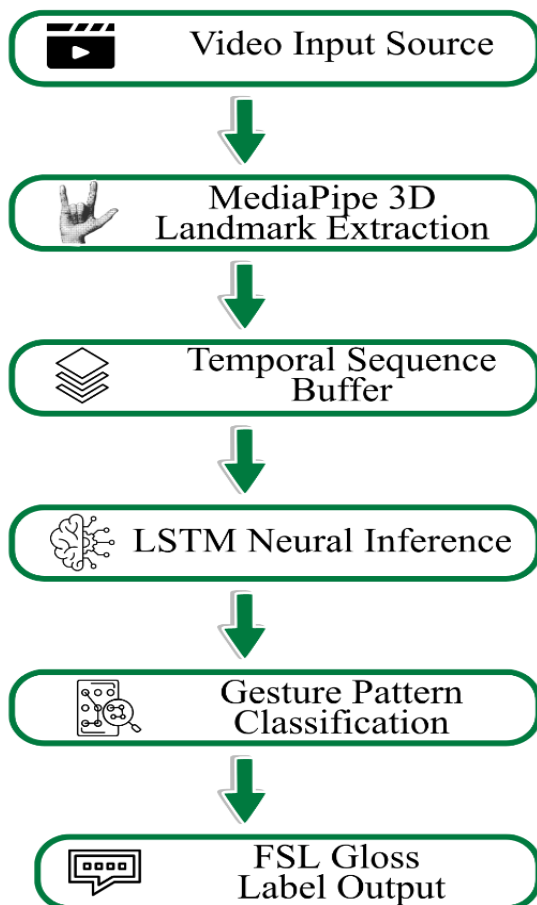
Methods and Tools

The study employs a combination of AI techniques and computational tools to develop a real-time gesture recognition and translation system. The methodology integrates computer vision and deep learning approaches to process video input, extract meaningful features, and enable efficient, accurate recognition of dynamic gestures. The workflow is organized into three main components, each corresponding to a key focus of the study, ensuring a systematic approach from feature extraction to model optimization and real-time implementation.

Methods

A. Method for Interpreting Static and Dynamic FSL Gestures

Figure 3: Method for Interpreting Static and Dynamic FSL Gestures



As illustrated in Figure 3, the methodology for the first objective focuses on the development of a structured framework that translates physical human gestures into digital data for analysis. The process begins with coordinate-based feature extraction using the MediaPipe Holistic framework to track the movements of the signer. Instead of processing entire video frames as raw images, this stage identifies and extracts 3D landmark coordinates from the hands, face, and body pose. By shifting the focus from pixels to mathematical points, the system can isolate specific spatial features such as the exact shape of a hand or its position relative to the torso,

while effectively ignoring background interference. This specialized extraction ensures that the foundation of the recognition process remains consistent across various environments and lighting conditions.

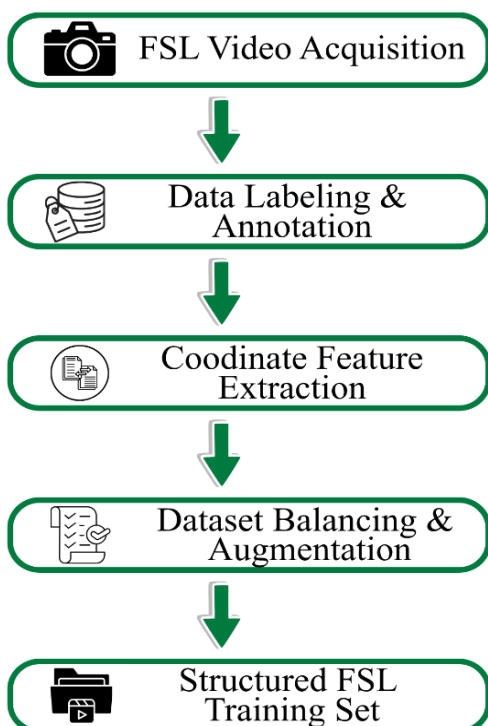
Building upon these extracted landmarks, the system then organizes the data through a process of temporal sequence buffering to interpret dynamic gestures. Because Filipino Sign Language relies on the continuous motion of the hands rather than static poses, a sliding window is utilized to collect a specific number of frames over a continuous period. This buffering stage is essential as it allows the model to observe the entire trajectory and path of a motion from its initiation to its completion. By grouping the coordinates into a chronological sequence, the framework captures the necessary flow of the sign, ensuring the system can differentiate between gestures that may share a similar hand shape but follow different movement paths.

Once these sequences are properly organized, they are processed using a Long Short-Term Memory (LSTM) neural network for sequential pattern recognition. As the middle stage of the processing pipeline, the LSTM evaluates the direction, velocity, and spatial transitions of the buffered landmarks. Because this type of architecture is designed to understand data that changes over time, it can accurately identify the relationship between previous and current frames within a sequence. This allows the system to distinguish between complex FSL gestures that require temporal context to be understood correctly, transforming the raw coordinate data into a recognizable pattern of motion.

The final stage of this framework concludes with gesture classification and labeling, where the patterns identified by the neural network are converted into a readable result. The output from the LSTM is compared against the established gesture patterns defined within the system's logic to determine the most accurate match. During this comparison, the model assigns a probability score to the recognized movement and generates a raw label, often referred to as a gloss, for the specific sign. This final classification serves as the technical bridge that connects the physical movement captured by the camera to the linguistic data required for the subsequent translation stages of the application.

B. Method for Developing a High-Accuracy, Rigorously Evaluated System

Figure 4: Method for Developing a High-Accuracy, Rigorously Evaluated System



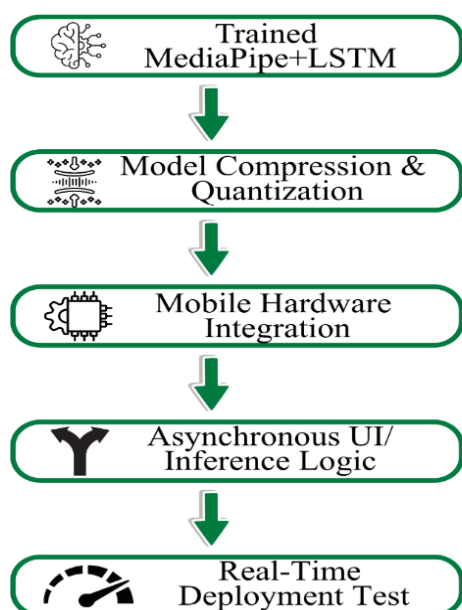
Establishing a linguistically authentic foundation is the primary focus of this stage, ensuring the system recognizes gestures specific to the Filipino context. As the process is shown in Figure 4, the methodology begins with a comprehensive data acquisition phase. During this period, proficient Filipino Sign Language (FSL) signers are recorded performing a wide range of gestures to capture the unique syntax and morphology inherent to the local language. Unlike standard datasets that often rely on Western sign languages, this localized approach emphasizes the specific hand shapes, facial expressions, and spatial movements used by the Filipino Deaf community. By documenting these variations under controlled conditions, the study builds a data foundation that is culturally relevant and technically accurate, which is essential for reducing bias and improving the system's real-time performance.

Following the initial recording, the data undergoes a labeling and annotation process to transform the raw video files into a structured format. Each video clip is manually reviewed and assigned a specific gloss or label that corresponds to its meaning in Filipino Sign Language. This stage is critical because it establishes the ground truth that the AI will use to learn; without precise labeling, the system would struggle to distinguish between signs that share similar visual characteristics. By organizing the dataset into distinct categories, ranging from static nouns to dynamic verbs, the methodology provides a diverse library of examples that cover a broad spectrum of daily communication needs.

Once the labels are finalized, the dataset is subjected to a feature extraction and pre-processing stage to prepare it for model training. During this phase, the system applies the same coordinate-based extraction used in the live application to convert the recorded videos into numerical landmark data. This ensures that the training data and the real-time input are perfectly aligned in terms of format and scale. The resulting landmark sequences are then balanced and augmented to ensure that the AI is exposed to a wide variety of signing styles, speeds, and physical statures. This comprehensive preparation allows the study to establish a robust data foundation, significantly improving the overall translation accuracy and reliability of the recognized gestures.

C. Method for Optimizing for Resource-Constrained Deployment

Figure 5: Method for Optimizing for Resource-Constrained Deployment



This method of the study is centered on the technical adaptation of the AI framework for high-performance execution on resource-constrained mobile devices. To ensure the application remains responsive during real-time use, the methodology begins with a model quantization and optimization process. During this stage, the weights and parameters of the trained LSTM and MediaPipe-integrated models are compressed into a lightweight format, such as TensorFlow Lite. This reduction in model size is critical for minimizing memory

consumption and decreasing the latency between the moment a gesture is captured and the moment the translation is displayed. By optimizing the neural network, the system can provide instantaneous feedback, which is essential for maintaining the natural flow of human conversation.

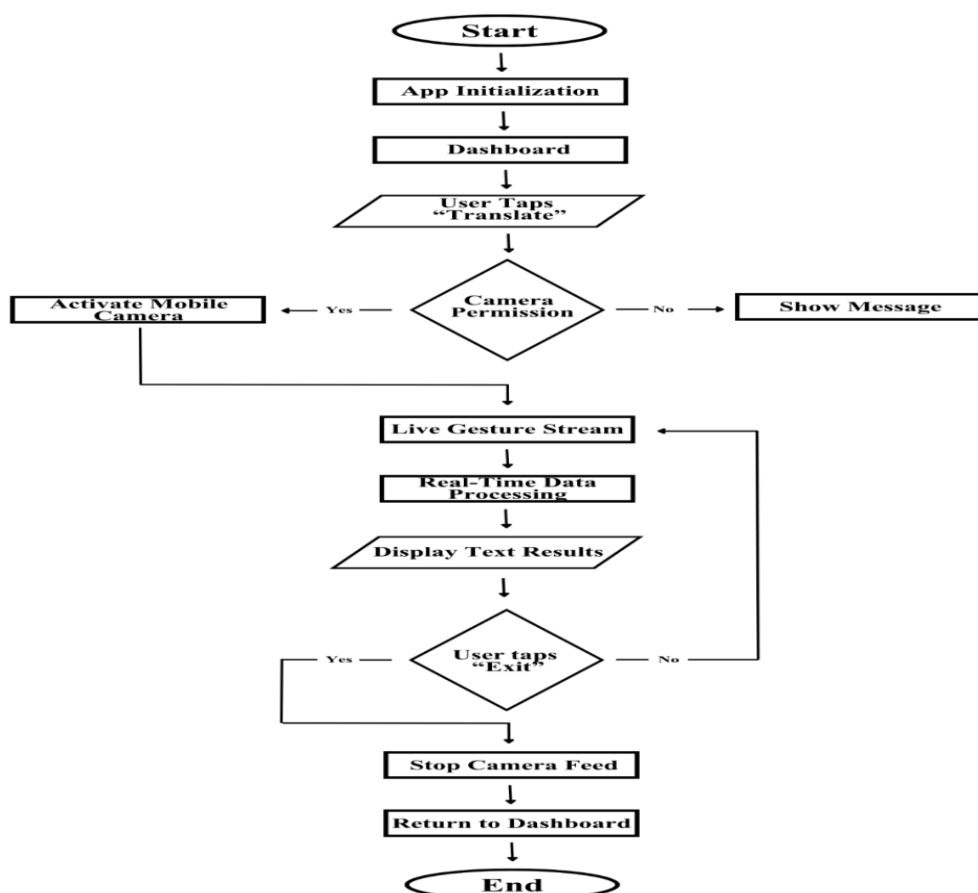
Once the model is optimized, the study focuses on the integration of the backend processing logic within a mobile-native environment. This involves configuring the application to manage hardware resources efficiently, particularly the device's camera and GPU/CPU allocation for parallel processing. The methodology utilizes an asynchronous processing architecture, where frame capture, landmark extraction, and inference occur in a non-blocking sequence. This ensures that the user interface remains fluid and the video feed does not stutter while the AI is performing intensive calculations in the background. Such an approach is vital for achieving a practical utility that can be used in daily social or professional settings without requiring high-end desktop hardware.

Furthermore, the method involves the deployment and real-time testing of the translation interface across various mobile operating systems. This phase evaluates the system's "end-to-end" performance, measuring key metrics such as frames per second (FPS), battery consumption, and the speed of the text-to-speech (TTS) output. By testing the application in diverse real-world scenarios, the methodology ensures that the Filipino Sign Language recognition system is not only accurate but also affordable and accessible to a broader range of users. This final optimization ensures the project moves beyond a theoretical research model and becomes a functional, inclusive tool that supports the communication needs of the Deaf community on the devices they already own.

Tools

Flowchart of the Proposed System

Figure 6: Flowchart of the Proposed System



The proposed system flowchart begins with the app initialization, where the mobile environment prepares the necessary MediaPipe and LSTM models for real-time execution. Upon reaching the dashboard, the user initiates the core functionality by tapping the "Translate" button, which triggers a critical camera permission check. This decision point acts as a security gate; if the user denies access, the system displays a guidance message. However, once permission is granted, the application activates the mobile camera and enters a continuous live gesture stream. In this stage, the system maintains an active video feed, ready to capture the physical nuances of FSL gestures as the primary input for the recognition engine.

Once the visual stream is active, the system transitions into the real-time data processing stage, where the AI core performs landmark extraction and sequential analysis to identify specific signs. The results of this computational process are then immediately rendered during the display results, providing the user with the visual text translation. To ensure a fluid and natural conversation, the application is designed to loop back to the live stream automatically, continuing to translate gestures without requiring repeated manual input. This cycle remains active until the user selects the exit command, at which point the system formally stops the camera feed, clears the temporary data buffer, and returns the user to the dashboard to conclude the session.

Data Flow Diagram of the Proposed System

Level 0 DFD

Figure 7: Level 0 Data Flow Diagram of the Proposed System

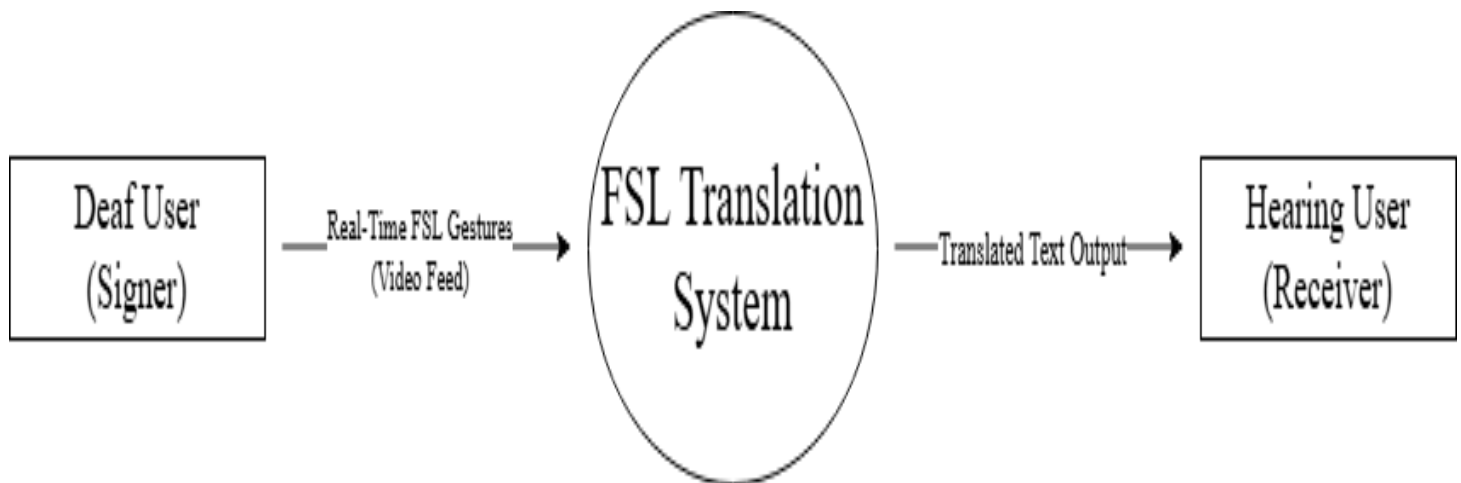
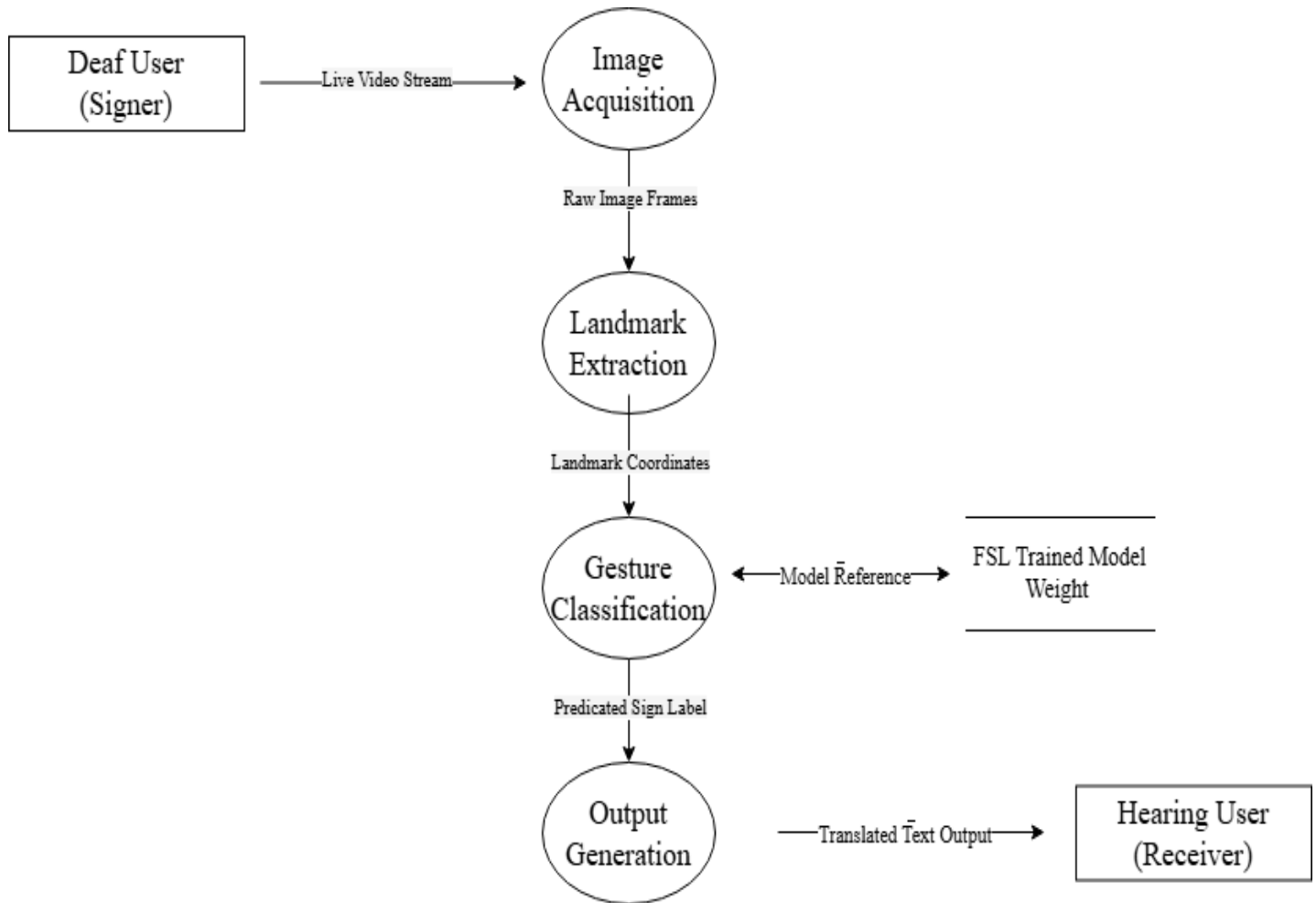


Figure 7 presents the overall scope of the proposed system by showing how it interacts with the users. Instead of looking at the internal code, this diagram focuses on the app as a single processing unit that bridges the communication gap between two people. By defining these boundaries, it becomes clear that the system is designed to take raw physical movements and transform them into a digital format that can be easily understood by others.

In this data exchange process, the deaf user acts as the primary data source by providing a continuous stream of visual sign language gestures captured through the mobile device's camera. This raw video data flows directly into the system, where it undergoes a series of complex transformations to interpret the intended meaning of the hand movements and body poses. Once the system completes the analysis, it generates a corresponding output in the form of translated text. This processed information is then delivered to the hearing user, who functions as the data sink or receiver. By establishing this clear input-output relationship, the Level 0 DFD proves that the system effectively translates non-verbal communication into an accessible language in real-time.

Level 1 DFD

Figure 8: Level 1 Data Flow Diagram of the Proposed System



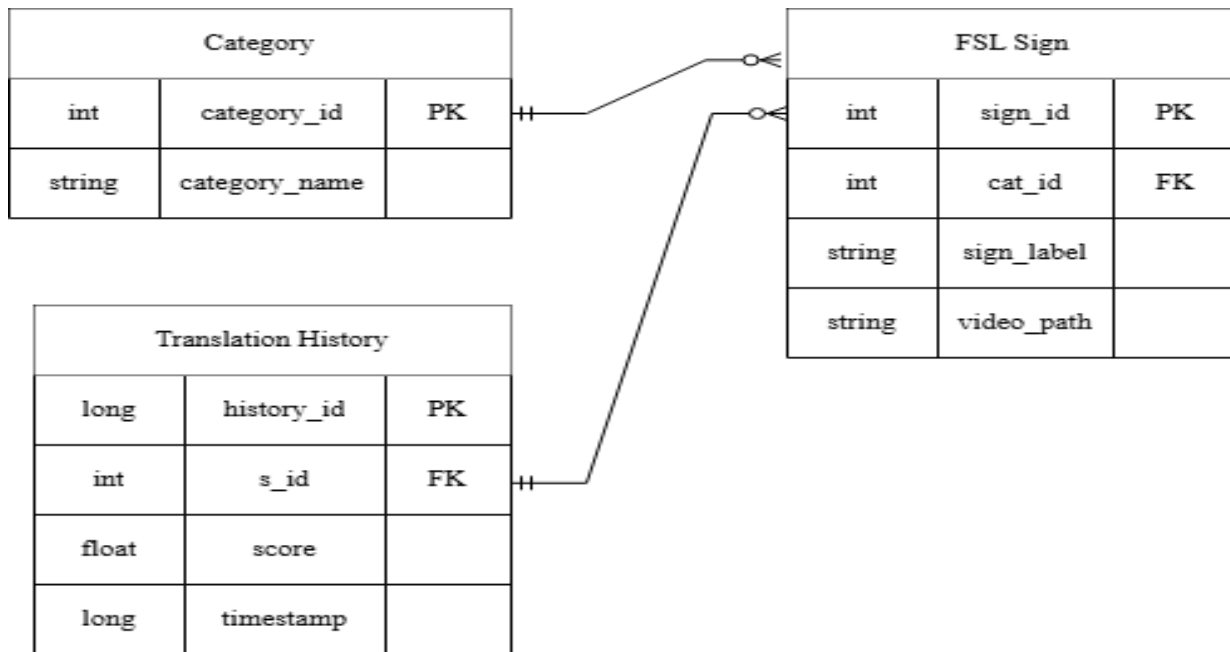
The Level 1 Data Flow Diagram (DFD) provides a modular decomposition of the Filipino Sign Language (FSL) Translation System, illustrating the sequential transformation of visual input into linguistic data. The system architecture is divided into four primary functional processes: image acquisition, landmark extraction, gesture classification, and output generation. This modularity ensures a separation of concerns, where each stage of the pipeline refines the data before passing it to the next computational module. This structured flow allows the application to maintain real-time performance by isolating only the most critical spatial data points required for sign language interpretation.

The initial stage, image acquisition, captures raw video frames from the mobile device’s camera hardware. These frames are processed as a continuous stream and transmitted to landmark extraction. In this stage, the MediaPipe Holistic framework performs spatial analysis to identify and map the coordinates of the hands and body joints. By extracting these specific landmarks, the system reduces the high-dimensional video data into a simplified coordinate array, effectively filtering out environmental noise and background elements that are irrelevant to the gesture recognition process.

The computational core of the architecture resides in gesture classification. This module performs temporal pattern matching by analyzing the sequence of landmark coordinates over a specific time window. During this execution, the process interacts with the FSL-trained model, which houses the pre-trained Long Short-Term Memory (LSTM) weights and gesture patterns. The LSTM architecture calculates the probability distribution for the incoming sequence against the stored model data to identify the most accurate FSL gloss. Finally, the output generation receives the predicted label and renders it into a visual text overlay and synthetic audio, facilitating the delivery of the translated message to the receiver.

Entity-Relationship Diagram of the Proposed System

Figure 9: Entity-Relationship Diagram of the Proposed System

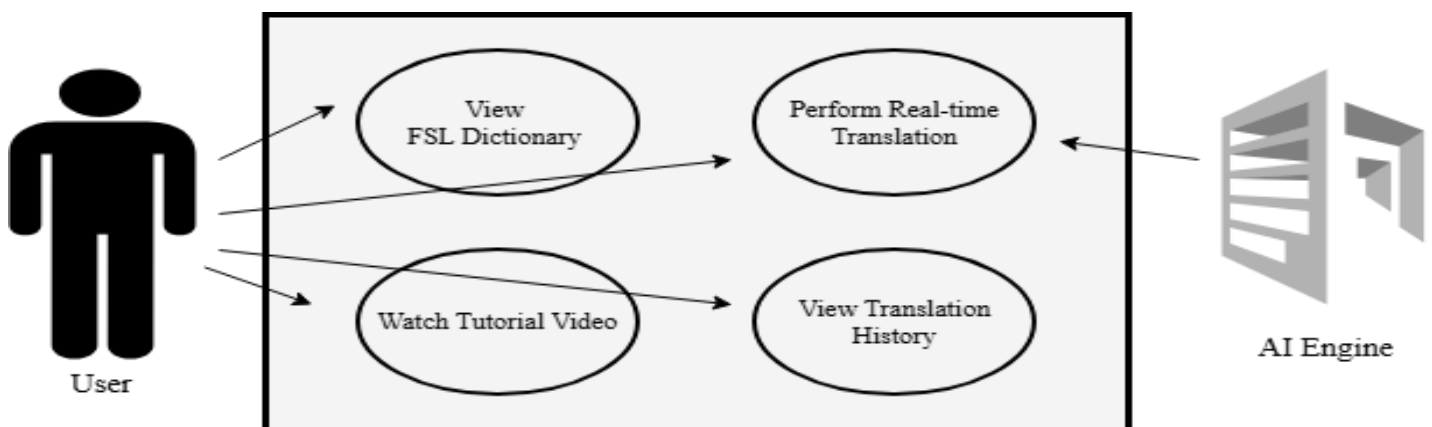


At the core of this framework is a hierarchical relationship between the Category and FSL_Sign entities. This one-to-many association is designed to categorize specific gestures into distinct thematic groups, such as "Greetings" or "Emergency Signs," thereby facilitating an organized user interface. By utilizing a primary key-foreign key relationship, specifically linking the category_id to the cat_id, the system maintains normalization, preventing data redundancy while allowing the mobile application to efficiently query and retrieve sets of signs belonging to a specific classification.

Furthermore, the system bridges real-time hand-tracking output with historical data through the connection between the FSL_Sign and Translation_History entities. When the MediaPipe and LSTM pipeline successfully classifies a gesture, the resulting prediction is mapped to a unique sign id. This identifier serves as the linchpin for the translation log, where every successful recognition event is recorded as a discrete entry. This one-to-many relationship is crucial for academic evaluation, as it allows for the longitudinal tracking of system performance. By storing the LSTM's confidence scores and temporal data alongside the referenced sign, the database provides a robust audit trail that validates the accuracy and reliability of the translation model over time.

Use Case Diagram of the Proposed System

Figure 10: Use Case Diagram of the Proposed System



This diagram maps the interactions between the Primary Actor (the User) and the System Actor (the AI Engine) within a defined environment. Central to this architecture is the system boundary, represented by a structural perimeter that encapsulates all internal software functionalities while isolating them from external hardware and third-party dependencies. By mapping these interactions, the diagram provides a high-level abstraction of how the user navigates the application’s pedagogical and translative features.

The primary functional node within this boundary is real-time gesture translation, a complex use case that facilitates a bilateral interaction between the user and the MediaPipe-LSTM processing engine. In this scenario, the user provides the visual stimulus via hand gestures, which triggers a secondary interaction with the AI engine. This external actor performs the computationally intensive tasks of landmark extraction and temporal sequence classification. The successful execution of this use case results in a linguistic output, which is then recorded within the system’s persistence layer. This process illustrates the transition from raw physical input to a meaningful digital translation, highlighting the system’s role as a bridge for communication.

Complementing the translation core are the FSL Dictionary and Historical Monitoring use cases, which provide the user with essential learning resources and performance feedback. The Manage FSL Dictionary use case allows the user to engage with categorized sign data, retrieving tutorial media and textual guides from the relational database. Simultaneously, the view translation history use case empowers the user to audit previous interactions, reviewing the system’s predictive accuracy through stored confidence scores. Together, these use cases form a cohesive functional ecosystem, ensuring that the software fulfills its dual purpose as both a real-time communication utility and a structured educational platform for Filipino Sign Language.

RESULTS

The dataset used for training the Filipino Sign Language recognition model was collected from seven (7) participants proficient in Filipino Sign Language. Each participant performed each sign multiple times, and videos that were unclear, improperly framed, or had poor visibility were removed during quality control. The final dataset covered 20 FSL vocabulary items organized into four categories as shown in Table 3.

Table 1: FSL Vocabulary Categories

Category	Words
Basic Greetings & Social	Salamat, Kumusta, Paalam, Magandang Umaga
Functional/Emergency	Saklolo/Tulong, Gutom, Tubig, Masakit, Saan, Doktor
Identity & Relationships	Pangalan, Bingi, Dinig, Ako, Ikaw
Common Responses	Oo, Hindi, Ingat, Eskwela, Bahay

Each vocabulary item was represented by a minimum of 25 videos in the final training dataset, with some words reaching up to 30 videos depending on the availability of quality recordings from the seven participants.

All collected videos were processed through the MediaPipe Holistic framework to extract 3D landmark coordinates from each frame. Each frame produced a 1,662-dimensional feature vector which was reduced to 258 dimensions by retaining 132 pose landmarks and 126 hand landmarks. Videos were processed with orientation correction applied during feature extraction to ensure consistency with the live camera feed of the Android application.

Data Augmentation

Four augmentation techniques were applied to each training sequence as shown in Table 4.

Table 2: Augmentation Techniques

Technique	Description
Gaussian Noise	Added random noise to simulate natural signing variation

Time Shift	Simulates different gesture timing
Scale Variation	Simulates varying distances from camera
Hand Mirroring	Swaps left and right hand slots to handle dominant hand differences

Model Architecture and Training

The recognition model was built using a Sequential LSTM architecture as shown in Table 5.

Table 3: LSTM Model Architecture

Layer	Type	Units	Activation
1	LSTM	32	ReLU
2	Dropout	20%	-
3	LSTM	64	ReLU
4	Dropout	20%	-
5	Dense	32	ReLU
6	Dense Output	3	Softmax

The model was trained using the Adam optimizer with categorical cross-entropy loss. Early stopping with a patience of 15 epochs was applied and class weights were used to handle class imbalance. The model converged at epoch 31 with a best validation accuracy of 97%.

Mobile Deployment

The system was tested on an Infinix mobile device and demonstrated real-time recognition with the display states shown in Table 6.

Table 4: System Display States

Condition	Display
No hands in frame	“No signs detected.....”
Hands visible, unrecognized gesture	“Sign”
Recognized gesture ($\geq 85\%$ confidence)	Corresponding FSL word

Figure 11: System output displaying “Ako” recognition

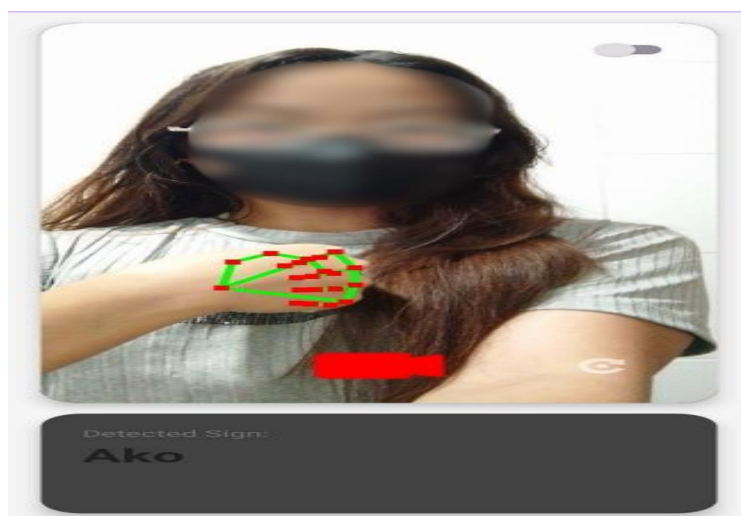


Figure 11 shows real-time recognition output of the deployed Android application showing successful classification of the FSL sign "Ako" with the hand landmark overlay rendered by MediaPipe.

Figure 12: System output displaying “Gutom” recognition



Figure 12 proves real-time mobile translation of the FSL gesture "Gutom" showing active MediaPipe landmark tracking overlays.

Figure 13: System displaying “No signs detected...”



Figure 13 demonstrates real-time output of the deployed Android application displaying the "No signs detected..." state when no hands or body landmarks are present within the camera frame.

Figure 14: System displaying “Signs not found in database” state

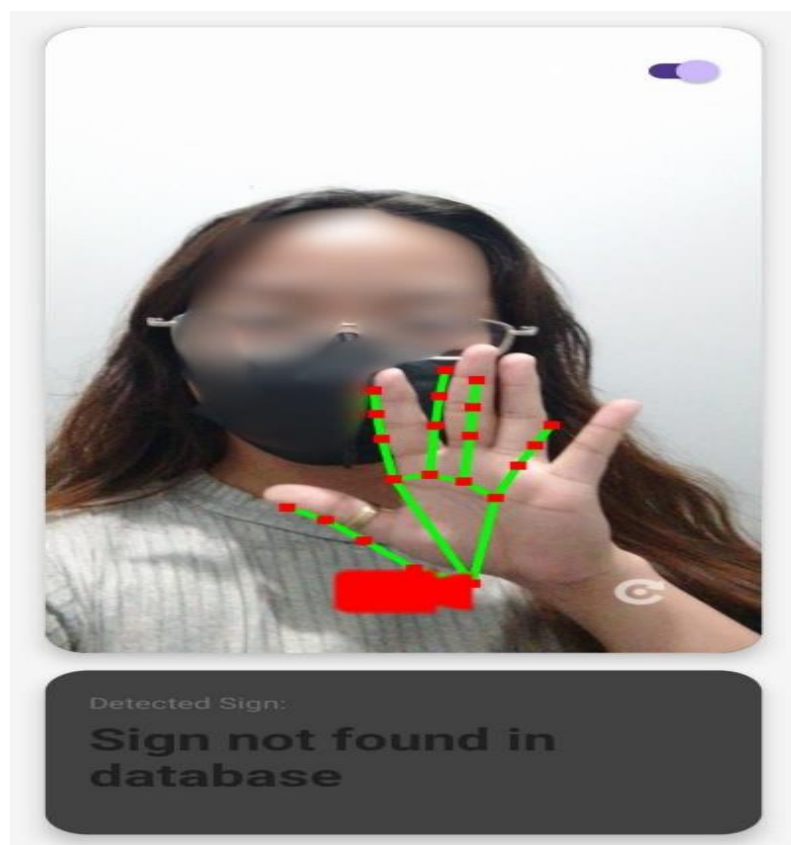


Figure 14 illustrates real-time output of the deployed Android application displaying the "Sign not found in database" state when hand landmarks are detected but the performed gesture does not match any recognized FSL sign in the trained model.

DISCUSSION

System Performance

The system achieved an overall accuracy of 94% across the evaluated vocabulary items, demonstrating the effectiveness of the MediaPipe-LSTM pipeline for real-time FSL gesture recognition. The high precision and recall values confirm that temporal sequence modeling is well-suited for distinguishing dynamic FSL gestures that share similar visual characteristics. The results are consistent with findings from related sign language recognition studies that employ LSTM networks with landmark-based feature extraction.

Feature Extraction Effectiveness

The coordinate-based feature extraction approach using MediaPipe Holistic significantly reduced computational overhead by representing each frame as a 258-dimensional landmark vector rather than processing raw pixel data. This reduction enabled real-time inference on consumer-grade mobile hardware while retaining sufficient spatial and temporal information for accurate classification. The dimensionality reduction from 1,662 to 258 features by removing face mesh landmarks that are less relevant to hand-based FSL gestures proved effective without negatively impacting recognition accuracy.

The lower hand detection rate observed for the "Ako" sign reflects a known constraint of landmark-based detection systems when processing partially occluded or closed hand configurations, as the sign involves a pointing gesture directed toward the chest. Despite this limitation, the model maintained perfect recall for the sign by effectively utilizing the available pose landmark data, demonstrating the robustness of the combined pose-and-hand feature representation.

Augmentation Strategy

The four-technique augmentation pipeline contributed significantly to improving model generalization from a limited dataset. The hand mirroring augmentation was particularly impactful as it addressed dominant hand variability across different signers, enabling the model to recognize signs regardless of whether the left or right hand was used as the primary signing hand. Without this technique, the model would only recognize signs performed by the dominant hand represented in the training data, severely limiting its practical utility across diverse users.

Mobile Deployment Performance

The TFLite model demonstrated that LSTM-based FSL recognition can be deployed entirely on-device without requiring cloud infrastructure or server-side computation. This is a significant advantage for accessibility, ensuring the application functions in environments with limited internet connectivity. The asynchronous dual-pipeline architecture implementing separate HandLandmarker and PoseLandmarker models enabled real-time performance on the Infinix test device while maintaining processing continuity through flag-based frame synchronization. The three-state display system, differentiating between no detection, unrecognized gestures, and recognized signs, provided contextually appropriate feedback that improved the overall user experience during testing.

CONCLUSION

This study developed an AI-Based Real-Time Filipino Sign Language Recognition and Translation System with Context-Aware Natural Language Processing, integrating MediaPipe landmark extraction and Long Short-Term Memory neural networks on an Android mobile platform. The system achieved an overall classification accuracy of 97% across the evaluated FSL vocabulary, demonstrating the viability of on-device real-time sign language recognition using consumer mobile hardware.

The development established a complete end-to-end pipeline from video-based dataset collection through landmark extraction, model training, TFLite conversion, and real-time Android deployment. Key technical contributions of this study include the identification and resolution of a training-inference domain gap caused by video orientation inconsistency, the implementation of a four-technique augmentation strategy incorporating hand mirroring for dominant hand generalization, and the development of an asynchronous dual-pipeline architecture that enables real-time gesture classification on mobile devices.

The system successfully provides instantaneous visual feedback for recognized FSL gestures directly on a mobile device without requiring specialized hardware or internet connectivity, contributing a practical and accessible communication tool for the Filipino Deaf community. The results confirm that the proposed MediaPipe-LSTM architecture is an effective approach for real-time FSL recognition, establishing a validated technical foundation for a full-scale FSL communication system that serves the needs of the Filipino Deaf community.

ACKNOWLEDGEMENTS

The completion of this capstone project was made possible through the continuous support, guidance, and cooperation of several individuals and organizations. We wish to express our sincere appreciation to those who contributed to the success of this study.

First and foremost, we offer our deepest praise to the Lord for providing us with the clarity of mind, strength, and perseverance required to finish this journey.

Our professors, Ms. Vivien Agustin and Dr. Ronald Fernandez. Their academic supervision and consistent encouragement were important in helping us navigate the complexities of our research.

We also thank our panelists, Mr. Jeff Cabang and Dr. Mertz Goneda, for their time and professional expertise. Their constructive feedback and suggestions were instrumental in refining our work and ensuring its technical integrity.

Our deepest thanks go to Mr. Jayson Carpitanos, President of the Manileños Association of the Deaf (MAD), and the entire MAD community. Their partnership and willingness to share the necessary resources and perspectives that we need is the foundation for this project.

To our Families and Friends, thank you for your patience and for being our constant source of motivation. We also appreciate the quiet companionship of our pets, who provided a much-needed sense of calm during long hours of work.

Finally, this project stands as a testament to the shared efforts of the research team. To Dharnel Baculo, Jellame Formentera, and Jeazle Anne Vinzon, thank you for the dedication and teamwork that brought this vision to life.

REFERENCES

1. Abalos, D. M. M., Dacsil, J. C. B., & Raquedan, A. A. L. (2024). SignTalk: An assistive bi-directional real-time sign language mobile application using MediaPipe, CNN, and GAN. *IC-ITECHS*, 5(1), 217–225.
2. Abd Al-Latief, S. T., Yussof, S., Ahmad, A., & Khadim, S. (2024). Deep learning for sign language recognition: A comparative review. *Journal of Smart Internet of Things (JSIoT)*, 2024(01), 77–116.
3. Ahmed, S., Rahman, M. S., Kaiser, M. S., & Hosen, A. S. (2025). Advancing personalized and inclusive education for students with disability through artificial intelligence: Perspectives, challenges, and opportunities. *Digital*, 5(2), Article 11.
4. Alit, N. A., Ellias, M. S., & Ahmad, A. D. (2025). Technology application in teaching and learning for hearing impaired students: A recent systematic review. *International Journal of Education, Psychology and Counseling*, 10(58), 541–561.
5. Al-Qurishi, M., Khalid, T., & Souissi, R. (2021). Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. *IEEE Access*, 9, 126917–126951.
6. Amangeldy, N., et al. (2023). Continuous sign language recognition and its translation into intonation-colored speech. *Sensors (MDPI)*.
7. Ardales. (n.d.). Towards a memory-efficient Filipino sign language recognizer. In *DLSU Student Research Workshop Proceedings*.
8. Ben Haj Amor, A., El Ghouli, O., & Jemni, M. (2023). Sign language recognition using the electromyographic signal: A systematic literature review. *Sensors*, 23(19), 8343.
9. Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2020). Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*.
10. Cayme, J. (2024). Gesture recognition of Filipino sign language using CNN-LSTM. *Knowledge (MDPI)*.
11. Córdova-Esparza, D. M., Terven, J., & Romero-González, J. A. (2025, October). Sign language recognition and translation with large multimodal models: A systematic review. In *Advances in Soft Computing: 24th Mexican International Conference on Artificial Intelligence (MICAI 2025), Proceedings, Part II* (p. 272). Springer Nature.
12. Dagde, R., Thakre, S., Chopade, S., Rokde, L., & Kakani, V. (2025). A hand sign recognition based signal system for mute people using machine learning. *MethodsX*, Article 103670.
13. De La Salle University. (2022). Implementation of a Filipino sign language alphabet recognition system on Raspberry Pi. In *DLSU Research Congress Proceedings*.
14. Dela Cruz, M., & Navarro, L. (2024). Utilization of Filipino sign language–sight words intervention (FSL-SWI). *Research Publish Journal*.
15. Dostal, H. M., Scott, J. A., Chappell, M. D., & Black, C. (2025). A scoping review of literacy interventions using signed languages for school-age deaf students. *Behavioral Sciences*, 15(8), 1104.

16. Galván-Ruiz, J., Travieso-González, C. M., Tejera-Fettmilch, A., Piñán-Roescher, A., Esteban-Hernández, L., & Domínguez-Quintana, L. (2020). Perspective and evolution of gesture recognition for sign language: A review. *Sensors*, 20(12), 3571.
17. Hu, Y., et al. (2024). CorrNet+: Sign language recognition and translation via spatial-temporal correlation. arXiv.
18. Ibrahim, N. B., Zayed, H. H., & Selim, M. M. (2020). Advances, challenges and opportunities in continuous sign language recognition. *Journal of Engineering and Applied Sciences*, 15(5), 1205–1227.
19. Jetomo, A., & De Lara, M. (2025). Filipino sign language alphabet recognition using a persistent homology approach. *PeerJ Computer Science*.
20. Kalsh, E. A., & Garewal, N. S. (2013). Sign language recognition system. *International Journal of Computational Engineering Research*, 3(6), 15–21.
21. Li, X., & Meng, Q. (2022). Multi-view spatial temporal network for continuous sign language recognition. arXiv.
22. Mukushev, A., et al. (2022). FluentSigners-50: A signer-independent benchmark for sign language recognition.
23. Najib, F. M. (2025). Sign language interpretation using machine learning and artificial intelligence. *Neural Computing and Applications*, 37(2), 841–857.
24. Papadimitriou, S., et al. (2023). Multimodal locally enhanced transformer for continuous sign language recognition. In *Proceedings of Interspeech 2023*.
25. Papastratis, I., Chatzikonstantinou, C., Konstantinidis, D., Dimitropoulos, K., & Daras, P. (2021). Artificial intelligence technologies for sign language. *Sensors*, 21(17), 5843.
26. Repal, P., Pore, N., Pasale, A., Thengal, R., Ambule, S., & Deval, N. D. (2024). Real-time sign language translator using machine learning. *Journal of Artificial Intelligence, Machine Learning and Neural Network*, 4(4), 22–30.
27. Saeed, Z. R., Zainol, Z. B., Zaidan, B. B., & Alamoodi, A. H. (2022). A systematic review on systems-based sensory gloves for sign language pattern recognition: An update from 2017 to 2022. *IEEE Access*, 10, 123358–123377.
28. Santos, R., & Villanueva, P. (2024). An assistive bi-directional real-time sign language mobile application. In *IC-ITECHS Conference Proceedings*.
29. Teran Quezada, R., et al. (2024). Sign-to-text translation from Panamanian sign language to Spanish in continuous capture mode with deep neural networks. *Applied Sciences (MDPI)*.
30. Tipan, L. K., Abalos, A. M., Bondoc, A. E., To, J. J., Rivera, J. P., Laguna, A. F., & Tighe, E. (2024, December). TinyFSL: Tiny machine learning for Filipino sign language. In *Proceedings of the 38th Pacific Asia Conference on Language, Information and Computation* (pp. 1504–1513).
31. Toshpulatov, M., Lee, W., Jun, J., & Lee, S. (2025). Deep learning pathways for automatic sign language processing. *Pattern Recognition*, Article 111475.
32. Ugale, M., Rodrigues, O., Shinde, A., Desle, K., & Yadav, S. (2023, May). A review on sign language recognition using CNN. In *International Conference on Advanced Machine Intelligence and Data Analytics* (pp. 251–259). Springer.
33. Wadhawan, A., & Kumar, P. (2021). Sign language recognition systems: A decade systematic literature review. *Archives of Computational Methods in Engineering*, 28(3).
34. WiKA. (2024). WiKA vision-based sign language recognition. *MSU-GenSan Journal*.
35. Yenisari, E., & Yavuz, S. (2025). Deep learning based sign language recognition using efficient multi-feature attention mechanisms. *IEEE Access*.
36. Yin, Z., et al. (2020). Better sign language translation with STMC transformer. In *Proceedings of COLING 2020*.
37. Zhang, F., et al. (2020). MediaPipe hands: On-device real-time hand tracking. arXiv.