SignTalk: An Assistive Bi-Directional Real-Time Sign Language Mobile Application using MediaPipe, CNN, and GAN

Abalos, Diane Marie M[1*], Dacsil, John Chrismark B[2], Raquedan, Ainah Angeline L[3]

[1,2,3]*School of Computer Studies and Technology Colegio de San Juan de Letran Calamba, Bucal Bypass Rd, Calamba, 4027 Laguna, Philippines*

Article Information

Received: 21-11-2024
Revised: 28-11-2024
Published: 05-12-2024

Keywords
Hearing Impaired Individuals, Sign Language, Sign Talk, Special Education

****Correspondence Email:***
diane@gmail.com

Abstract

In the realm of inclusive education, the hearing-impaired community often faces communication barriers, particularly in educational settings where sign language may not be universally understood. This research study endeavors to address this issue through the development of a bi-directional sign language application tailored to the unique needs of hearing-impaired students, parents, and teachers at Paete Elementary School. Before this research, no dedicated studies or applications have been crafted yet. Employing a quantitative approach via simple random sampling as the instrumentation for the client's population, this study aims to delve into the participants' perspectives. This pioneering initiative provides valuable insights into technology-driven solutions for the communication gap in education for the hearing-impaired. Its findings are poised to significantly contribute to inclusive education and technology-assisted learning, fostering a more accessible educational environment for the hearing-impaired community. The researchers strongly encourage future studies to prioritize technological solutions that cater explicitly to the diverse needs of Special Education (SPED) students or Persons with Disabilities (PWD). Understanding their requirements and challenges is crucial in creating more inclusive and effective applications or tools. Further with research studies understanding their requirements and challenges is important in creating more inclusive and effective applications or tools. Foster interdisciplinary collaborations between technology experts, educators, therapists, and specialists in the field of disability studies. Collaborative efforts can yield comprehensive solutions that integrate technological advancements.

1. Introduction

Communication is essential for social survival, yet understanding sign language remains a challenge for non-signers. Globally, 72 million individuals are hearing-impaired, with approximately 300,000 in the Philippines. Communication barriers, particularly for families with limited sign language knowledge, often lead to frustration and isolation. While technology, such as sign language recognition apps, offers potential solutions, challenges persist with regional and stylistic language variations. Schools like Paete Elementary's SPED program face resource limitations, emphasizing the need for tools like the proposed "SignTalk" mobile app to bridge communication gaps, support Filipino Sign Language (FSL) learners, and foster inclusivity.

SignTalk aims to convert sign language gestures into text and vice versa, addressing the lack of real-time tools for hearing-impaired and non-signing individuals. The app incorporates technologies like MediaPipe, CNNs, and GANs to enhance accuracy and user experience while enabling users to contribute gesture data for updates. Its focus on ASL, with similarities to FSL, ensures broad applicability. Despite limitations such as environmental factors and single-user gesture recognition, the app seeks to create a more accessible and inclusive society by overcoming communication barriers.

1.1 Literature Review

I. Sign Language

According to Cacko (2019), Sign language, as a linguistic system, relies on manual communication and body language to convey meaning effectively. It utilizes a wide range of hand shapes, hand-arm-body movements, and facial expressions to express thoughts. Sign language is recognized as a natural language that facilitates communication, the sharing of thoughts, and the expression of emotions.

II. Filipino Sign Language

Republic Act No. 11106, also known as the Filipino Sign Language (FSL) Act, establishes FSL as the official national sign language for the Filipino deaf community. The law mandates the use of FSL in government transactions, education, media, and workplaces, promoting it as the primary mode of communication for deaf individuals while allowing alternative methods based on personal preferences. Madrid (2022) highlights FSL as a vital tool for bridging communication gaps among individuals with impairments, emphasizing its cultural and contextual significance.

Montefalcon et al. (2021) note that most Filipinos lack familiarity with FSL, underlining the need for increased education and awareness. This knowledge gap creates a divide between the deaf community and the hearing majority, exacerbated by the rapid technological advancements in Philippine society. Developing sign language recognition systems that accurately interpret and produce FSL signs can bridge this gap, fostering inclusivity and ensuring the deaf community's active participation in the digital era. This effort requires advancements in education, awareness, and technology to fully integrate FSL into Filipino society.

III. Deaf in the Philippines

A study by the Philippine Deaf Resource Center (PDRC) and the University of the Philippines (UP) College of Law identifies several barriers to the inclusion of persons with disabilities (PWDs) in the Philippines. Dr. Erwin Alampay from UP NCPAG highlights limited accessibility in transportation, public spaces, and ICT as key challenges, emphasizing the need for inclusive policies to ensure equal opportunities (Yang, 2018). CNN Philippines reports that Deaf Filipinos often face a lack of representation and accessibility, such as having to personally arrange and pay for interpreters in hospitals. Efforts to address these challenges include initiatives like the development of sign language recognition apps for the Deaf community. These steps are crucial for fostering inclusivity. However, Bual stresses that creating an inclusive society requires collective responsibility from all sectors, not just the Deaf community.

IV. Sign Language Recognition Technology

The implementation of sign language recognition technology is vital in overcoming challenges faced by sign language users. As most communication technologies are designed for spoken or written languages, they unintentionally exclude sign language users. Rastgoo (2021) emphasizes that creating a sign language recognition system capable of translating signs into text or voice is a key hurdle in facilitating communication between the deaf and hearing communities. Current research primarily uses isolated datasets of individual signs, but real-time conversation systems are necessary for practical communication.

Neiva (2018) notes that deaf individuals use hand gestures and facial expressions to communicate, but inclusion is hindered by hearing people's limited understanding of sign language. Innovations in mobile technology, such as gesture recognition on smartphones, have helped address this challenge. Some solutions have used sensor-based gloves or skin-tone recognition, employing classification methods like Support Vector Machines and Neural Networks. While static gestures have accuracy levels above 80%, dynamic motions achieve accuracy above 90%, though testing is often conducted in controlled environments with minimal motion and specific lighting.

V. Machine Learning

As described by Ayodele (2018), machine learning involves automated computational processes relying on logical or binary operations to acquire knowledge from a set of examples. The primary goal of Machine Learning is to generate simple classification models that are easily understandable to humans. These models aim to mimic human reasoning effectively, providing insights into the decision-making process. While the development of machine learning models may involve leveraging background knowledge, the expectation is that these models operate autonomously without requiring human intervention. Like statistical approaches, machine learning aims to achieve autonomous functionality while maintaining the ability to provide valuable insights.

VI. Image Processing

Kamal Preet Kour et al. (2019) propose a novel method for recognizing sign language gestures using image processing techniques. Their approach involves segmenting hand gestures from the background and extracting features such as color histograms, shape, and contour features. They introduce a new ensemble classifier combining decision tree and k-nearest neighbor (KNN) algorithms, achieving an impressive average recognition rate of 92.3% across 26 sign language gestures. Their method outperforms existing ones in terms of accuracy. The authors suggest future improvements focusing on robustness to lighting and background variations and the use of deep learning techniques for feature extraction and classification.

Pramada et al. (2018) highlight the need to improve sign language recognition accuracy, especially for users with varying hand deformities or postures. They propose combining image processing, machine learning, and AI to create an intelligent system capable of accurately recognizing sign language gestures.

VII. MediaPipe

In India, over one million individuals with hearing and speech impairments rely on sign language for communication. However, regional dialects, such as Assamese Sign Language, have received little attention in sign language recognition (SLR) systems. To address this gap, Bora et al. (2023) introduced a deep learning-based SLR system focusing on recognizing key letters from the Assamese alphabet. The system uses Google's open-source MediaPipe for hand tracking, which identifies 21 specific points on the hand, gathers them as coordinate data, and normalizes it into a CSV file for analysis.

Grishchenko (2020) explains that MediaPipe's Holistic framework combines machine learning models with computer vision to predict facial, hand, and body positions in real-time from video streams. This toolkit allows for the creation of applications that understand human gestures and motions, including sign language, and is adaptable for various uses like augmented reality and fitness monitoring. It has made significant contributions to machine learning research by offering publicly available solutions for Python and JavaScript developers.

Kellinger (2021) highlights the potential of preprocessing techniques, such as pose estimation, to improve real-time sign language detection. These methods track key points on the human body, enabling simpler hardware configurations and efficient identification. MediaPipe's low latency and high accuracy in real-world applications make it a strong candidate for sign language recognition, and its Holistic component, which integrates pose, hand, and facial tracking, plays a crucial role in this technology's effectiveness.

VIII. MediaPipe Framework

The MediaPipe framework, developed by Lugaresi et al. (2021), helps machine learning practitioners create production-ready applications, share code, and prototype technologies. It simplifies the process of building perception pipelines by connecting individual perception models and addressing challenges in constructing and deploying technology. These pipelines help infer sensory data and produce the desired outcomes, with components easily reused across platforms.

Kavana (2022) highlights MediaPipe's use in hand tracking through its MediaPipe Hands model. By comparing it to OpenPose's skeletal estimation model, MediaPipe reduces computation by using regression analysis to determine finger coordinates based on the palm's position. This results in more accurate finger shape estimation and enables hand tracking with fewer resources, making it ideal for mobile platforms.

IX. Mobile Sign Language Translation

The development of technology has significantly helped individuals with disabilities, particularly in improving communication for the deaf. Mindaña et al. (2021) introduced a mobile app for real-time Filipino Sign Language (FSL) emergency sign detection and basic FSL clips, enabling better communication for hearing-impaired individuals. This mobile-based solution addresses the shortage of sign language interpreters and helps bridge the communication gap between deaf and hearing people.

Masbate et al. (2020) also noted that using technology like mobile applications makes it easier to reach and communicate with the deaf community. Sign Language Recognition (SLR) systems, which translate sign language into text or speech, play a crucial role in enhancing communication. Jiang and Ahmad (2019) developed a vision-based system using a USB camera and a skin tone algorithm to detect hand gestures, achieving a 99.4% success rate in recognizing five common American sign language gestures.

Further research by Empe et al. (2020) found that mobile and web applications effectively support deaf students' learning. Nasereddin (2018) highlighted the potential of mobile apps to provide modern information and communication to distance learners. Additionally, de Castro et al. (2023) suggested that automatic sign language translation software could reduce communication barriers, allowing deaf and hearing individuals to communicate without the need for interpreters. While many mobile apps help non-signers learn sign language, Tolentino et al. (2019) pointed out that these apps often require significant data, internet connectivity, and storage.

X. User Interaction Method

Goyal (2023) highlights that technological advancements have largely overlooked the unique needs of the Deaf and Hard of Hearing (DHH) community, despite significant improvements in communication and daily living for others. The study focuses on developing technology to address these needs, exploring static signs and gesture-based signals in sign language. By using Long Short-Term Memory (LSTM) for temporal relationships and Convolutional Neural Networks (CNN) for static signal recognition, the research aims to create more accurate and effective Sign Language Recognition (SLR) models, aiding communication for speech and hearing-impaired individuals.

Stoll (2020) presents an innovative deep-learning approach to convert spoken language into sign language videos, using a combination of a Neural Machine Translation (NMT) network and a Motion Graph (MG) to generate human posture sequences. This method offers a practical, cost-effective solution for translating spoken languages into sign language, with minimal annotation required. Though not as detailed as avatar-based systems, it can generate realistic sign language. The study suggests further exploration of multi-signer systems, high-definition capabilities, and system improvements to handle diverse environments and users of different sign language competency levels. The research emphasizes the importance of addressing the needs of hearing-impaired individuals and the potential for mobile-based sign language recognition systems to break down communication barriers and promote social inclusion.

XI. Storing Datasets

According to Bali (2022), in Stage 1, multi-hand landmarks were extracted from photos and preprocessed using MediaPipe. This framework included a variety of body identification and tracking algorithms that were honed using the sizable dataset provided by Google. MediaPipe created normalized 3D coordinates and a skeleton model by recording key locations on various body regions. In Stage 2, cloud storage was used to speed up data processing and provide enough storage space for the many photos used in the model's training. As a result, sign language databases like those for American Sign Language or Indian Sign Language might be quickly retrieved and used while still maintaining security. Two sections of cloud storage were created: one for training data and the other for test/validation data. In Stage 3, incorrect items were eliminated as part of the data cleaning and normalization process, assuring the correctness and objectivity of the prediction model. The dataset was then split into training and validation sets. The sign language recognition model was constructed in Stage 4 using an LSTM (Long Short-Term Memory) model, which allowed the capturing of sequence dependencies. Finally, in Stage 5, the trained model could be kept for later use without the requirement for retraining, and the test accuracy of the model was assessed using a different test dataset. Re

XII. Real-Time Sign Language Recognition

Real-time Sign Language Recognition has been a research focus for over two decades, with various classifiers such as linear classifiers, neural networks, and Bayesian networks being explored to enhance the systems. Garcia and Viesca (2018) proposed a real-time American Sign Language (ASL) recognition system using Convolutional Neural Networks (CNN). This system is designed to convert ASL signs from video recordings into text, addressing the challenges of capturing video input, classifying each frame, and generating the most probable word as output. In a different area of research, Dorado and Villanueva (2023) developed a Filipino children's speech recognition system using Deep Neural Networks, aiming for low-latency recognition and achieving a low word mistake rate. Their system uses a dataset of raw speech files from Filipino children, collected from the University of the Philippines' Digital Signal Processing Laboratory. Meanwhile, Das et al. (2022) focused on bridging the communication gap between the deaf and hearing populations by creating a real-time ASL fingerspelling translator. This technology aims to enhance communication, particularly in emergency situations where textual communication may not be effective. The system enables users to view videos with sign language interpreters and understand finger-typed messages. Lastly, Garcia et al. (2021) advanced ASL recognition by integrating CNN with a multi-view augmentation technique, allowing the system to process 3D depth images for more accurate real-time classification of ASL alphabets.

XIII. ISO/IEC 25010:2023

In the context of adopting agile methodologies, various forms of content, such as data, information, and knowledge, play a vital role in the creation, communication, and consumption of information within an organization. Effectively managing this content is essential not only during the development phase but also throughout the deployment and beyond. The primary objective of the software development process is to successfully deliver the final "product" in the form of software implementation. This product may attract potential investors who are interested in the future services it will provide. In the evaluation of software products, software processes, and potential enhancements, ISO/IEC 25010:2023 holds significant importance. This standard provides essential guidelines and criteria for assessing and improving software products and processes. By adhering to ISO/IEC 25010:2023, organizations can ensure that their software development efforts meet international standards and can deliver high-quality and reliable products that meet customer requirements and expectations.

XIV. Synthesis

Gesture recognition has gained significant attention in recent years, driven by its potential applications in areas like human-computer interaction, virtual reality, and assistive technologies for individuals with disabilities. Researchers are focused on developing real-time gesture recognition systems that can handle challenges like varying lighting conditions, background noise, occlusions, and complex hand movements. A key area of this research is sign language recognition for the hearing impaired, where machine learning techniques improve the accuracy of hand gesture predictions. Murillo et al. (2021) applied the Mediapipe framework with recurrent neural networks for real-time gesture detection using webcams or mobile devices, achieving notable success in sign language recognition. Saleem et al. (2023) introduced a cost-effective system that allows non-deaf individuals to communicate with the speech- and hearing-impaired using a Leap Motion Device and Convolutional Neural Networks (CNN) to process hand gestures. This system transforms hand motion data into speech, facilitating communication without requiring users to learn sign language. While current sign language recognition applications mainly focus on one-way communication, where sign language is translated for non-signers, there is a growing emphasis on developing two-way communication platforms. These systems aim to enable sign language users to express themselves fluently, while non-signers can respond seamlessly, with the app translating their responses into sign language.

XV. Conceptual Framework

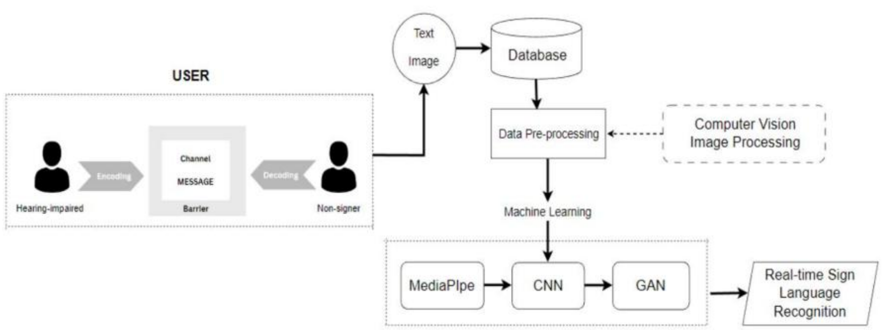

Fig.1 Conceptual Framework

2. Research Methods

The study at Paete Elementary School in Laguna, Philippines, explored communication challenges faced by the Special Education Department (SPED) when interacting with hearing-impaired individuals. Using a Quantitative descriptive research design, the researchers aimed to understand specific communication barriers and identify areas for improvement. The study emphasized strategies aligned with ISO 25010 to promote inclusive and effective communication within the educational setting.

SignTalk Respondents

Respondents	Population Size	Sample Size
Hearing Impaired Students	30	26
Guardian of Hearing Impaired	30	26
(Margin of error) ***e = 0.05***	*(Total Population)* ***N = 60***	***n = 52.1 ≈ 52***

Fig. 2 Table SignTalk Respondent

The researchers employed the Agile methodology for their software development life cycle. Agile methodology employs an iterative approach, where each phase is repeated until all sprints are completed. It consists of six key phases: requirements, design, development, testing, deployment, and review. This ensures a more favorable outcome for the system.

The development of the mobile application was conducted using Python within the Android Studio environment, with database management handled through MySQL. Documentation related to the project was created using Microsoft Word or Google Docs. For hardware, the application supported devices with cameras ranging from 720p to 9216p resolution, ensuring robust hand recognition and high-quality image capture. Testing was conducted on mobile devices with specifications including Android 11 on a Helio G95 Dual chip CPU with 11GB RAM, or Android 12 with MagicUI 6.1 running on a Qualcomm Snapdragon 695 CPU with 32GB RAM. The app required front and rear cameras with minimum resolutions of 16 megapixels and 108 megapixels respectively, though devices with at least an 8-megapixel camera, digital zoom, expandable storage via microSD, Wi-Fi, and standard connectivity options were also suitable. An alternative configuration featuring a triple rear camera setup (64MP main, 5MP ultra-wide, and 2MP macro) and a 16MP front camera ensured effective hand identification and excellent image capture during testing.

Input-Process-Output Model

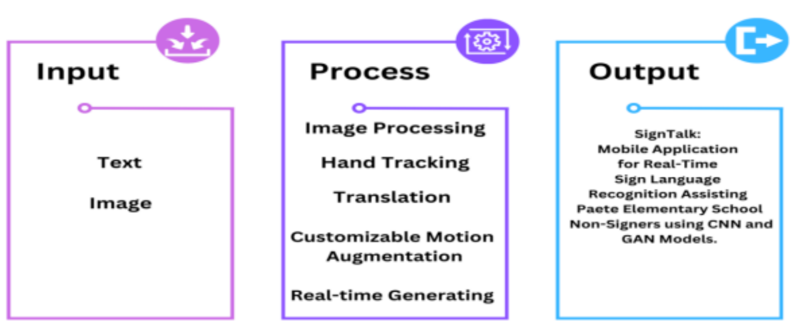

Fig.3 Input Proses Output Model

3. Result and Discussion

The application successfully facilitates two-way communication by translating words into sign language videos and vice versa, supported by a user-friendly tutorial for beginners and children. Its inclusivity and accessibility contribute to a more empathetic society. Advanced machine learning algorithms and computer vision techniques enable precise recognition of hand signs, with testing involving ten volunteers and diverse datasets achieving a rapid 6.2-second recognition speed.

Users can expand the application's database by contributing new gestures, which are validated by administrators to ensure accuracy, fostering continual improvement. Comprehensive evaluations aligned with ISO/IEC 25010:2023 standards received overwhelmingly positive feedback from students, parents, teachers, and IT experts, validating the app's effectiveness. The application bridges communication gaps between hearing-impaired students, parents, and teachers, enabling meaningful interactions and improving educational and personal development. Its role in enhancing communication highlights its potential to foster inclusivity and stronger connections in educational settings.

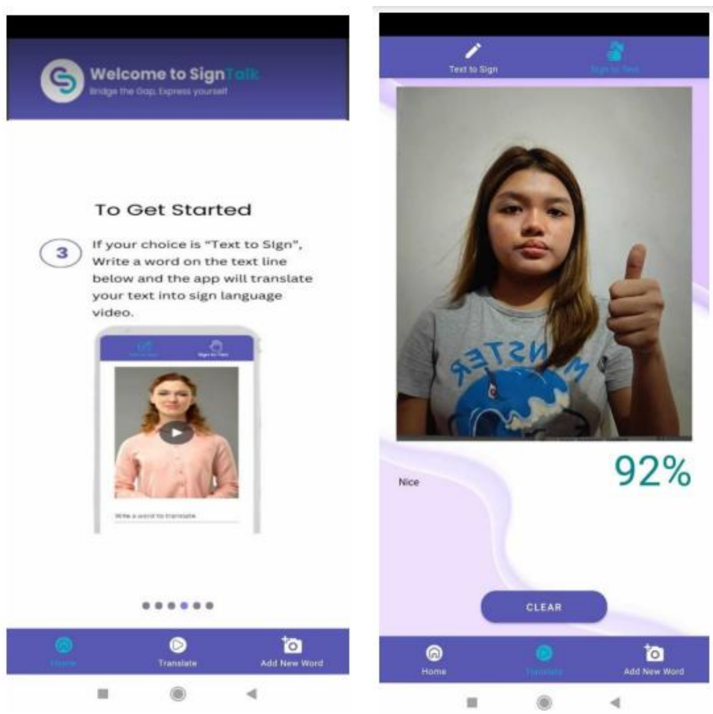

Fig. 4 Demo

4. Conclusions

The study concluded that the developed application effectively facilitates two-way communication between hearing-impaired individuals and non-signers through bi-directional translation, promoting inclusivity and accessibility. Its advanced algorithms achieved high precision in recognizing and translating sign language gestures, ensuring seamless and effective communication. User engagement played a crucial role in enhancing the app, as contributions of new gestures and feedback enabled continuous updates and improvements. With a rapid 6.2-second recognition speed, the app demonstrated efficiency and adaptability. Additionally, the application holds significant potential to positively impact the lives and education of hearing-impaired children by fostering improved communication, self-expression, and a supportive learning environment.

5. Reference

Ayodele (2018). New Advances in Machine Learning. International Journal of Machine Learning and Cybernetics, p. 9-10. https://books.google.com.ph/books?hl=en&lr=&id=XAqhDwAAQBAJ&oi=fnd&pg=PA9&dq=machine+learning+overview&ots=r2NraVylHm&sig=zaKdsz7SlzxLgAaJ6KB6Rtbkyl4&redir_esc=y#v=onepage&q&f=false

Bautista et al. (2021, July). Hand Alphabet Recognition for Dactylology Conversion to English Print Using Streaming Video Segmentation. In Proceedings of the 9th International Conference on Computer and Communications Management (pp. 46-51). https://dl.acm.org/doi/abs/10.1145/3479162.3479169

Bali (2022, April 30). Sign Language Detection for Deaf using Deep Learning, MediaPipe and OpenCV. Medium. https://medium.com/@mayank.bali/sign-language-detectionfor-deaf-using-deep-learning-mediapipe-u-opencv-4c5151e2374c

Bora et al. (2023). Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning and Deep Learning. International Conference on Machine Learning and Data Engineering. https://doi.org/10.1016/j.procs.2023.01.117

Cacko (2019). Sign Language. Retrieved from: https://www.academia.edu/19240544/Sign_language

Das et al. (June 2022) REAL TIME SIGN LANGUAGE RECOGNITION Retrieved from: https://www.jetir.org/papers/JETIRFM06048.pdf

De Castro et al. (2023). Automatic translation of sign language with multi-stream 3D CNN and generation of artificial depth maps. https://www.sciencedirect.com/science/article/abs/pii/S0957417422024125

Dorado, and Villanueva(2023, May). Development of Low-Latency and Real-Time Filipino Children Automatic Speech Recognition System using Deep Neural Network. In 2023 11th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/10131755/