

The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 2026

Volume 41, Pages 241-250

ICRETS 2026: International Conference on Research in Engineering, Technology and Science

TransFil: A Hybrid CNN–LSTM Framework for Multimodal Translation of Filipino Sign Language and Facial Expressions for Non-Signer Communication

Patrick D. Cerna

Mapua Malayan Colleges Mindanao

Jess G. Luengas

Mapua Malayan Colleges Mindanao

Luan Sebastian B. Jardin

Mapua Malayan Colleges Mindanao

Raphael M. Amorado

Mapua Malayan Colleges Mindanao

Abstract: Filipino Sign Language (FSL) serves as the national sign language of the Philippines, yet its accessibility remains limited due to widespread unfamiliarity and the dominance of American Sign Language (ASL). Existing technologies for sign language translation often neglect crucial non-manual markers such as facial expressions, which play a key role in grammar and meaning. To address this gap, this study presents the development of an AI-powered FSL recognition system that integrates both manual gestures and facial expressions, focusing on greetings and common questions. Using a video-based dataset collected in controlled conditions, the system employs Media Pipe for landmark detection, Convolutional Neural Networks (CNNs) for spatial feature extraction, and Long Short-Term Memory (LSTM) networks for temporal modeling. Overall, The facial expression recognition (FER) model, built using a CNN architecture, was trained to identify six common emotions: angry, fear, happy, neutral, sad, and surprised. It performed strongly, achieving about 97% training accuracy and 95% validation accuracy, showing that it can generalize well to new data. While, the model demonstrated reliable performance, with precision (0.95), recall (0.94), and F1-score (0.95), making it suitable for real-world use. These results align with previous studies highlighting the strength of CNNs in capturing important facial features.

Keywords: Filipino sign language, Long short-term memory (LSTM), Media pipe, Facial expression recognition, Gesture recognition, Deaf community accessibility

Introduction

Sign language is a rich and expressive way of communication that uses hand movements, facial expressions, and body gestures to convey thoughts, emotions, and ideas. Unlike spoken language, it does not use the voice but instead relies on visual signals. It is especially important for people who are Deaf, hard-of-hearing, or have difficulty speaking. In the Philippines, around 300,000 people are Deaf (Echon et al., 2021), and they rely on sign language to communicate. There are three main sign languages used in the Philippines: Filipino Sign Language (FSL), American Sign Language (ASL), and Signing Exact English (SEE). In 2018, FSL was officially declared the national sign language through Republic Act No. 11106. This law requires FSL to be used in schools,

government services, workplaces, and media. However, most Filipinos are still more familiar with ASL, which was introduced earlier in the country.

In recent years, researchers have developed and implemented various systems focused on recognizing sign language, supported by advances in different areas of technology (Tolba & Elons, 2013). Interest in sign language recognition has significantly increased (El-Bendary et al., 2010). Numerous studies have been conducted on different sign languages, including American Sign Language, Korean Sign Language, and Chinese Sign Language, among others (Brashear et al., 2006). This technology is crucial for helping individuals with speech impairments communicate, as they are unable to use spoken language (Cabalfin et al., 2012). Sign language recognition holds great promise for enhancing communication and accessibility, making it a growing field within computer science and artificial intelligence.

Historically, the development of sign language translation technologies—such as smart gloves and signing avatars—has been mostly led by hearing researchers, often without involving Deaf individuals in the process. Some of these tools do not fully match the way sign languages actually work. For example, smart gloves like SignAloud were designed to translate hand movements into speech or text, but they ignore important parts of sign language like facial expressions and body movements, which are essential for clear and correct communication. Signing avatars, which are animated characters that perform signs, also face criticism because they often look unnatural and miss the emotion, rhythm, and cultural style of real signing. Because of these issues, many Deaf community members have raised concerns about the growing use of such technologies in communication (Kusters et al., 2023). Lack of inclusion has resulted in tools that may not fully align with the linguistic and cultural needs of the Deaf community. Some members of the Deaf community have raised concerns about the growing use of technology in sign language communication. While tools like AI-based sign language translation systems can offer support in certain situations, there is worry that relying too much on these tools could reduce the value placed on learning and using sign language itself. Sign languages are not just a way to communicate—they carry cultural identity, shared experiences, and community connection. According to Desai et al. (2024), there is a real risk that AI systems might be seen as a replacement for human interpreters, which could lead to the loss of meaningful, person-to-person interaction. This raises ethical concerns about how such technologies are introduced and used.

Despite concerns over technological substitution, there are areas where the Deaf community has welcomed the development of sign language technologies, particularly when these tools are designed to enhance accessibility without undermining the richness of human interaction. Many in the community have expressed support for technologies that help bridge communication gaps, especially in environments where access to interpreters or Deaf professionals may be limited. As noted by De Meulder, Van Landuyt, and Omardeen (2024), AI systems that assist in real-time communication, such as automatic captions or sign language recognition tools, have been praised for promoting independence and inclusion. When developed with input from the Deaf community, these technologies are seen as empowering, providing more opportunities for Deaf individuals to participate in mainstream society while still preserving the cultural and linguistic aspects of sign language.

Although FSL was officially recognized as the national sign language of the Philippines through Republic Act No. 11106, and its use has been mandated in education, healthcare, justice, employment, media, and other public domains, awareness and understanding among the general population remain limited. Mendoza (2018) notes that many Filipinos mistakenly believe ASL is the only sign language in use, despite differences in gestures and expressions that affect clear communication. The ongoing lack of accessible communication tools continues to limit the inclusion of Deaf and hard-of-hearing individuals in essential areas such as education, employment, and public services. As noted by Petinglay (2021), bridging this language barrier is critical to promoting the rights and opportunities of Persons with Disabilities (PWDs). Current sign language technologies often overlook facial expressions, which are essential for expressing grammar and emotion in Filipino Sign Language (FSL) (Sharma et al., 2024). To address this, our study aims to develop a hybrid CNN-LSTM model that recognizes both hand gestures and facial expressions in FSL greetings, integrating them into a real-time translation system. This system will be evaluated using both quantitative performance metrics and user feedback, ensuring its effectiveness and cultural relevance for the Deaf community. This effort not only improves accessibility for non-signers but also supports those interested in learning FSL by helping them understand the meaning, grammar, and facial expression cues behind commonly used greetings.

As highlighted by Kaur and Singh (2020), deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have proven effective in detecting hand movements, facial expressions, and body posture—core elements in sign language recognition.

Furthermore, recent studies have explored multi-modal approaches, integrating gesture, facial expressions, and contextual cues to improve recognition accuracy and user experience (Pigou et al., 2018). These systems not only interpret hand signs but also capture facial expression, adding depth to the conveyed message—crucial for maintaining the nuances of human interaction. Research also shows growing interest in developing datasets and frameworks specific to various sign languages, including American Sign Language (ASL), British Sign Language (BSL), and emerging studies into Tagalog Sign Language (TSL), supporting the global effort for linguistic inclusivity in AI systems (Camgoz et al., 2020).

There are already applications which focus on FSL that help mute and the hearing-impaired people, such as the research done by students in Mapua Malayan Colleges Mindanao “A visual approach in Filipino sign language recognition using TensorFlow Lite model”. This study is not only on the development of a software in detecting expressions and translating gestures but, it also evaluates the accuracy of the model that the researchers wants to study. The main objective of this study is to develop a software that utilizes AI to recognize gestures to translate FSL (Filipino Sign Language) greetings and detect facial expressions. To achieve the main objective, the study was guided with these specific objectives:

- A. Develop a hybrid model using CNN and LSTM architectures to recognize FSL gestures that depend heavily on facial expression such as greetings using a labeled FSL video dataset and high-resolution camera for data collection.
- B. Integrate the hand gesture and facial expression recognition components into a unified translation system that converts the detected Filipino Sign Language (FSL) signs into accurate text output for real-time communication.

Evaluate the system’s performance and usability using quantitative metrics such as accuracy, precision, recall, and F1 score, along with feedback gathered from Filipino Sign Language (FSL) users.

Related Works

Recent developments in artificial intelligence have significantly advanced the field of real-time sign language recognition. Pandey (2025) explored the potential of deep learning by combining convolutional neural networks (CNNs) for spatial feature extraction and recurrent neural networks (RNNs) for sequential modeling. The objective was to create a system that could accurately translate sign language gestures into text or speech outputs in real time. By leveraging CNNs and RNNs, the system demonstrated high recognition accuracy, making it a promising approach for improving communication for the deaf and hard-of-hearing community. This study’s findings highlight the transformative potential of deep learning in real-time applications and are highly relevant to our research, as they underscore the importance of combining different neural network architectures to optimize gesture recognition in our system. Sign language communication often relies not only on hand gestures but also on facial expressions to convey full meaning. Kean University’s team developed a system that not only recognized American and British Sign Languages (ASL and BSL) but also analyzed facial expressions to interpret emotional cues, which are essential for accurate communication. Using OpenCV and the MediaPipe library, this system was designed to capture both hand movements and facial expressions in real time. The key finding was that incorporating facial expressions significantly improved the accuracy of sign language translation. This research is pertinent to our project as we aim to embed emotional context into our sign language recognition system, making it more expressive and truer to human communication.

Zhang et al. (2023) introduced a novel approach for sign language recognition using event cameras, which are distinct from traditional cameras in that they capture motion-triggered changes instead of frame-based images. The integration of event cameras with spiking neural networks (SNNs) allowed for faster and more efficient gesture classification. This study demonstrated a 77% accuracy in recognizing gestures and pointed to the potential of event-driven sensors to enhance responsiveness and energy efficiency in real-time systems. For our project, the insights from this study could be pivotal in exploring alternative sensor technologies, especially for improving system speed and energy consumption. Stergiou and Papadopoulos (2021) conducted a comprehensive review of the state-of-the-art methods in sign language recognition, translation, and representation. Their work synthesized a wide range of technologies, from hand gesture tracking to advanced machine learning models. One of the major takeaways was the need for more integrated approaches that consider not only gestures but also contextual factors like facial expressions and emotional nuances. Their findings align closely with our research, particularly in identifying the limitations of current methods and providing a roadmap for integrating multi-modal inputs into sign language systems.

Aggarwal et al. (2023) focused on developing a real-time system that combines hand tracking with gesture classification. Their approach aimed to build a bridge for communication between hearing and non-hearing individuals, making it possible for people who use sign language to interact seamlessly with those who do not. The system achieved high confidence scores and demonstrated responsive behavior, making it a crucial step toward creating real-time, accessible sign language translation systems. This study is highly relevant to our project as it emphasizes the importance of real-time processing, which is a core feature of the system we aim to build. A recent study explored the integration of emotion recognition in sign language interpretation, highlighting the importance of facial expressions and lip movements alongside hand gestures. The researchers used machine learning, image processing, and animation tools to analyze and interpret the emotional content conveyed through sign language. This research provides crucial insights for our project, particularly in understanding how non-verbal cues like emotions can impact the effectiveness of sign language interpretation. Our goal of capturing both signs and emotions directly aligns with these findings.

In a study published in *Sensors* (2023), the researchers demonstrated the power of using the ASL Alphabet Hand Gesture Dataset to build a real-time recognition system. The primary objective was to optimize classification frameworks and model training for real-time applications, emphasizing the importance of curated datasets in improving system accuracy. The study's conclusions underscore the significance of data quality and model optimization in achieving high-performance recognition. This aligns closely with our goals, as our project will rely heavily on well-curated datasets and robust training processes to ensure efficient and accurate gesture recognition.

Research Design

This study employs a qualitative research design to deeply explore the development process and rationale behind the AI-powered Filipino Sign Language (FSL) recognition mobile application. By using this approach, the researchers aim to explain not just the outcomes, but also the underlying decisions and processes involved in building the system. As Bhandari (2020) highlights, qualitative research is particularly effective in gaining insights into how individuals experience and interpret the world—making it well-suited for understanding how users interact with assistive technologies like this sign language recognition tool.

Dataset Preparation

The system uses two primary input modalities: hand gestures for sign language and facial expressions.



Figure 1. Sign language dataset

Raw video clips of 10 Filipino Sign Language (FSL) classes were used as the primary dataset. These videos, each lasting 3–4 seconds, were sourced from publicly available FSL research repositories, including the DLSU Animo Repository dataset (Katigbak et al., 2023) (https://animorepository.dlsu.edu.ph/etdm_ece/25/). This dataset is suitable because it specifically focuses on Filipino Sign Language (FSL), aligning with the national context of the study. It provides authentic sign samples performed by native or fluent signers, ensuring linguistic accuracy and cultural relevance.

Each FSL class contained 20–22 video samples and was expanded using data augmentation (horizontal flipping, brightness adjustment, rotation, and cropping) to increase variability and robustness. The final dataset was divided into training (80%), validation (10%), and testing (10%).

Facial Expression Dataset

Facial expression data was obtained from the FER-2013 dataset hosted on Kaggle [<https://www.kaggle.com/datasets/msambare/fer2013>]. This dataset contains thousands of labeled facial images across multiple emotional categories. For this study, the following six categories were used: happy, sad, angry, surprised, neutral, and fear (the “disgust” category was excluded due to limited representation and weaker generalizability in prior studies).

FER-2013 was chosen because it is one of the most widely used benchmarks for facial expression recognition. It contains images collected under diverse conditions—different lighting, head poses, and backgrounds—which makes it highly suitable for training models intended for real-world application. The dataset was split into 70% training, 15% validation, and 15% testing.

Table 1. List of FSL greeting words and facial expressions used in the study

FSL Words	Facial Expressions
Kamusta Ka	Angry
Kita Tayo Bukas	Fear
Walang Anuman	Happy
Salamat	Neutral
Ikinagagalak Kitang Makilala	Sad
Okay Lang	Surprised
Hello	
Magandang Umaga	
Magandang Hapon	
Magandang Gabi	

Table 1 presents the set of Filipino Sign Language (FSL) greetings and facial expressions used in this study. The selected FSL words (*Hello, Kamusta Ka, Magandang Umaga, Magandang Hapon, Magandang Gabi, Salamat, Walang Anuman, Okay Lang, Kita Tayo Bukas, Ikinagagalak Kitang Makilala*) are all greetings and polite expressions. These were chosen because greetings are among the most fundamental and frequently used communicative acts in human interaction, forming the basis of social connection and daily conversation (Firth, 1972; Coulmas, 2013). In the context of sign language research, greetings are considered high-frequency lexical items that are essential for building accessible and practical recognition systems (Jetomo & De Lara, 2025). For the facial expressions, the categories (*angry, fear, happy, neutral, sad, surprised*) were selected because they represent the set of universally recognized facial expressions. According to Ekman & Friesen (1971), these six expressions are consistently identified across cultures and are widely used benchmarks in computer vision studies (Corneanu et al., 2016). The “disgust” category was excluded due to limited dataset representation and lower classification reliability in prior research (Mollahosseini et al., 2017).

Data Augmentation

To improve the generalization capability of the models and address the limited dataset size, data augmentation was applied to both the Filipino Sign Language (FSL) greeting videos and the facial expression images. Data augmentation introduces variability into the training data, thereby reducing overfitting and enabling models to perform better on unseen samples (Shorten & Khoshgoftaar, 2019).



Figure 2. Augmentation of FSL greeting videos

All training experiments were conducted in a Google Colab Pro environment with GPU acceleration. The FSL recognition model achieved approximately 99% training accuracy and 98–99% validation accuracy, while the facial expression model reached reliable accuracy across its six classes.

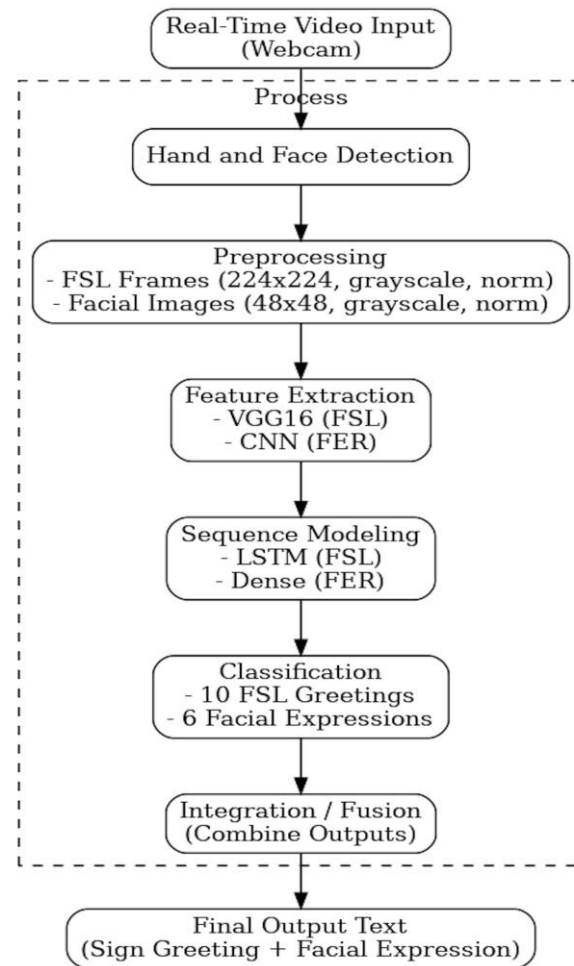


Figure 3. Conceptual framework of the study

The conceptual framework in Figure 2 illustrates the overall process flow of the proposed multimodal recognition system, beginning with real-time video input and culminating in the combined recognition of Filipino Sign Language (FSL) greetings and facial expressions.

Results and Discussion

The development of the multimodal recognition system relied on both hardware and software tools that helped with collecting data getting the data ready training the model and checking how well it works. We used a webcam as the hardware tool to capture video in real time that shows hand gestures and facial expressions. For the software part we used Google Colab Pro as the place to train the model because it has GPU acceleration, which helps with big data and complicated calculations. We built the system using Python 3.12. Made the main parts that deal with deep learning using TensorFlow and Keras. We used OpenCV to get frames from videos get images ready and connect the webcam when we tested the system in time. We also used Mediapipe to find faces and get them in the position, for the part of the system that recognizes facial expressions. We even used Graphviz to make diagrams like the picture that shows how the whole system works.

The picture in Figure 5 shows how the whole Filipino Sign Language recognition system works. It starts with a video. Ends with the system recognizing Filipino Sign Language greetings and facial expressions. The system has three parts: getting the video processing it and showing the result. The first part is getting the video from a webcam. This video has hand gestures and facial expressions that the system will recognize. The second part is where the system does all the work. It finds the hands and face in the video. Then it makes sure all the pictures are the same size and quality. For the Filipino Sign Language pictures it makes them 224×224 pixels changes them to black and white and makes sure they are all the same. It does the thing, with the facial pictures but it makes them 48×48 pixels. This helps the system focus on what's important. After that the system looks at the video. Finds the important parts. For Filipino Sign Language it uses a tool to find the important parts of the hand

gestures. For expressions it uses another tool to find the important parts of the face like the eyes, eyebrows and mouth. Then it looks at how these parts change over time. It uses a kind of computer program to understand how the hand gestures change and another kind to understand how the facial expressions change. The next step is where the system decides what it is looking at. It takes the parts it found and matches them to a list of possible greetings or facial expressions. The Filipino Sign Language system can recognize ten greetings and the facial expression system can recognize six different expressions. Then it combines the two results to get an understanding of what the person is communicating.

Finally, the system shows the result. It takes the recognized Filipino Sign Language greeting and the recognized facial expression. It shows them together. This makes the system more realistic and helpful because it understands how people use gestures and expressions when they communicate. The Filipino Sign Language recognition system is more inclusive and natural because it recognizes both non-verbal communications.

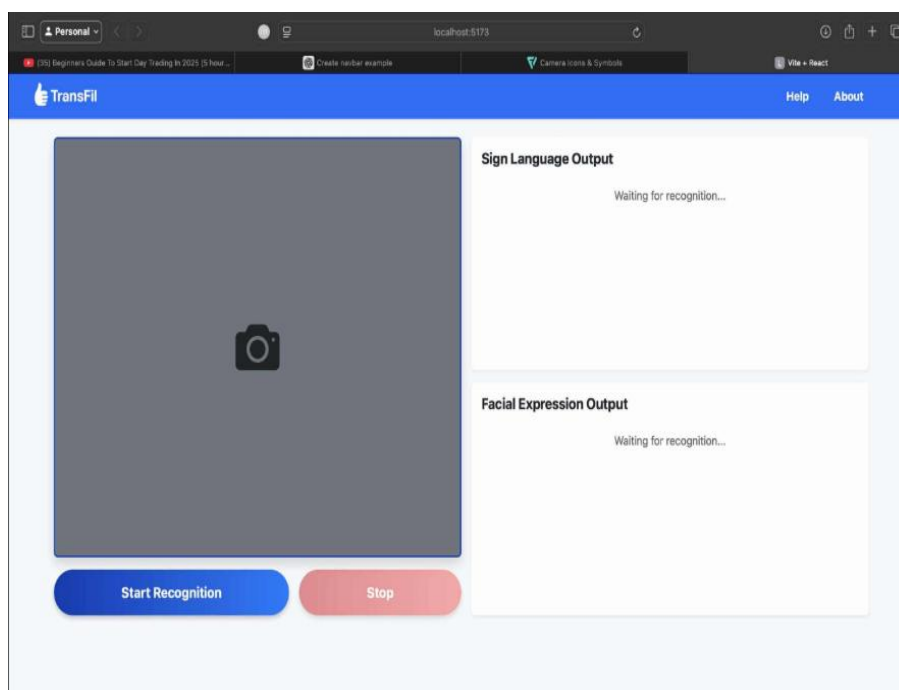


Figure 4. Filipino sign language interface for non-signers

The data we used for this study came from two places: our collections and public sources. To recognize Filipino Sign Language, we made videos of people doing ten different greeting signs. We then changed these videos in ways, such as, as flipping them making them brighter or darker turning them and cutting parts out so that our computer model could learn from many examples. For recognizing expressions, we used a set of data called FER-2013 that we found on Kaggle and we also made our own recordings. We wanted to make sure we had examples of people looking angry fearful happy, neutral, sad and surprised so Filipino Sign Language and facial expressions could be recognized properly. To see how well the models work we used a different tool. We looked at how accurate the models are and we also checked how precise they are, how well they recall things and what their F1-score is. This helps us understand how well the models can tell things apart whether it is the FSL model or the facial expression recognition model. We made tables to see how well the models can recognize each class and where they make mistakes. We also watched the models as they learned to make sure they were getting better and not just memorizing things. The models had to learn from the training data. Then do well on the validation data or it would mean they were not working right. All these tools together helped us make sure the system we proposed is good and works well

The FSL greeting recognition model was implemented using a CNN-LSTM hybrid architecture to capture both spatial and temporal features of signing sequences. The model achieved a training accuracy of 99% and a validation accuracy ranging between 98–99%. Furthermore, the validation loss stabilized at approximately 0.09, indicating strong convergence and robust generalization with minimal risk of overfitting. The confusion matrix demonstrates that all ten FSL greeting classes were recognized with high accuracy. Misclassifications were rare, with the most notable occurring between Magandang Umaga and Magandang Hapon, which share visually similar initial hand configurations and movements. Despite this, the overall classification remained reliable across all classes. The classification metrics further validate this performance, with the model achieving an average

precision of 0.98, recall of 0.97, and F1-score of 0.98. These values confirm the model's ability to maintain balance between sensitivity and predictive accuracy. The results obtained surpass the performance reported in prior studies, where typical accuracy for sign language recognition models ranges between 90–95% (Koller et al., 2019). The superior accuracy can be attributed to the combined effects of (1) extensive data augmentation, (2) preprocessing through grayscale normalization, and (3) the ability of the CNN–LSTM architecture to model both visual features and temporal dynamics in gesture sequences.

Results of Facial Expression Recognition

The facial expression recognition (FER) model was implemented using a CNN-based architecture trained on six universal facial expressions: angry, fear, happy, neutral, sad, and surprised. During training, the model achieved an accuracy of approximately 97%, while the validation accuracy stabilized at 95%, reflecting robust generalization across diverse test samples. The confusion matrix indicates that the model achieved the highest recognition rates for happy and surprised, both of which are characterized by distinctive and exaggerated facial features. In contrast, reduced performance was observed for fear and sad, which are inherently more subtle and prone to overlap in expression dynamics. Quantitatively, the FER model achieved an average precision of 0.95, recall of 0.94, and F1-score of 0.95, confirming its reliability for practical applications. These findings are consistent with existing literature that emphasizes the effectiveness of CNNs in extracting discriminative facial features (Li & Deng, 2020).

The results of this study show that the system we are talking about is really good at recognizing sign language greetings and facial expressions. This system uses a kind of pipeline that combines CNN and LSTM for recognizing sign language. It is good at understanding the order of gestures because it looks at them over time. We also used a CNN model to recognize expressions and it worked well because it can pick up on small details in peoples faces. The system we are talking about is good, at recognizing sign language greetings and facial expressions. Compared with existing approaches, the present system offers several key advantages:

- A. High Recognition Accuracy – The system gets things most of the time because it does a lot of work to get the data ready like making all the pictures the same color and it uses a lot of different examples to learn from.
- B. Balanced Multimodal Integration – The system is good at combining what it sees from gestures and facial expressions to understand what is going on. This is better than just looking at one thing.
- C. Real-Time Performance – The system can do its job quickly so it is good for things that need to happen in time, like when people are actually using it.

These findings show that learning based systems that use multiple forms of communication can really help the Deaf and Hard-of-Hearing community. Learning based systems can make things more accessible for the Deaf and Hard-of-Hearing community by breaking down communication barriers. The results of this study give us a starting point for more research on deep learning based systems. We can use this information to make learning based systems that can understand more words and handle more complicated conversations. Learning based systems have the potential to really make a difference, for the Deaf and Hard-of-Hearing community.

Conclusion

This study presented the design and implementation of a multimodal recognition system that integrates Filipino Sign Language (FSL) greeting recognition with facial expression recognition (FER). The system was developed in response to the persistent communication barriers experienced by the Deaf and Hard-of-Hearing community, with the aim of bridging gaps between signers and non-signers through deep learning–based approaches. For broader adoption, the system should be optimized for deployment on mobile and edge devices. Although the current implementation demonstrates real-time performance in a controlled environment, scaling it for everyday use requires reducing computational load while maintaining accuracy. Lightweight architectures such as MobileNet, as well as model compression and quantization techniques, could make the system more portable and efficient. This would enable deployment in smartphones, tablets, or assistive devices, significantly increasing its accessibility and impact. Overall, the study has successfully demonstrated the feasibility of a multimodal recognition system that integrates Filipino Sign Language greetings and facial expression recognition. The results validate the effectiveness of deep learning models in recognizing and interpreting multimodal communication, with applications that extend to inclusive education, accessibility technologies, and human–computer interaction. With further refinement, expansion, and user-centered development, the system holds promise as a transformative tool for bridging communication barriers and fostering inclusivity for the Deaf and Hard-of-Hearing community.

Scientific Ethics Declaration

* The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM Journal belongs to the authors.

Conflict of Interest

* The authors declare that they have no conflicts of interest

Funding

* This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Research in Engineering, Technology and Science (www.icrets.net) held in Batumi/Georgia on July 09-12, 2026.

References

- Brashear, H., Henderson, V., Park, K. H., Hamilton, H., Lee, S., & Starner, T. (2006). American sign language recognition in game development for deaf children. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 79–86). ACM.
- Cabalfin, E. P., Martinez, L. B., Guevara, R. C. L., & Naval, P. C. (2012). Filipino sign language recognition using manifold projection learning. In *TENCON 2012–2012 IEEE Region 10 Conference* (pp. 1–5). IEEE.
- Camgoz, N. C., Koller, O., Hadfield, S., & 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)* (pp. 10023–10033). IEEE.
- De Meulder, M., Van Landuyt, D., & Omardeen, R. (2024). *Lessons in co-creation: The inconvenient truths of inclusive sign language technology development*. arXiv:2408.13171.
- Desai, A., De Meulder, M., Hochgesang, J. A., Kocab, A., & Lu, A. X. (2024). Systemic biases in sign language AI research: A Deaf-led call to reevaluate research agendas. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources* (pp. 54–65). ELRA and ICCL.
- El-Bendary, N., Zawbaa, H. M., Daoud, M. S., Hassanien, A. E., & Nakamatsu, K. (2010). ArSLAT: Arabic sign language alphabets translator. In *Proceedings of the International Conference on Computer Information Systems and Industrial Management Applications* (pp. 590–595). IEEE.
- Gonzales, J. C., & Reyes, M. T. (2024). Gesture recognition of Filipino sign language using convolutional neural networks. *Knowledge*, 4(3).
- Jetomo, C. B., & De Lara, M. L. D. (2025). Filipino sign language alphabet recognition using persistent homology classification algorithm. *PeerJ Computer Science*, 11, e2720.
- Kaur, P., & Singh, M. (2020). Sign language recognition using deep learning: A review. *International Journal of Computer Applications*, 176(19), 6–11.
- Koller, O., Zargaran, S., Ney, H., & Bowden, R. (2016). Deep sign: Hybrid CNN-HMM for continuous sign language recognition. In *Proceedings of the British Machine Vision Conference (BMVC)* (pp. 136.1–136.12). BMVA Press.
- Kusters, A., Moriarty Harrelson, E., & Snoddon, K. (2023). Co-creating sign language technologies with deaf communities: Ethical reflections and future directions. In *Proceedings of the Workshop on the Representation and Processing of Sign Languages*.
- Kumar, R., Bajpai, A., & Sinha, A. (2023). *MediaPipe and CNNs for real-time ASL gesture recognition*. arXiv:2305.05296.
- Lee, Y., & Park, W. (2023). Real-time American sign language interpretation using deep learning. *Sensors*, 23(5).

Tolba, M. F., & Elons, A. S. (2013). Recent developments in sign language recognition systems. In *2013 8th International Conference on Computer Engineering & Systems (ICCES)*. IEEE

Author(s) Information

Patrick Cerna

Prof. Dr.

Mapua Malayan Colleges Mindanao

Gen. Douglas MacArthur Hwy

Matina, Davao City 8000

Davao del Sur, Philippines

Contact e-mail: pdcerna@mcm.edu.ph

ORCID iD: <https://orcid.org/0000-0002-2554-2098>

Jess G. Luengas

Mapua Malayan Colleges Mindanao

Gen. Douglas MacArthur Hwy

Matina, Davao City 8000

Davao del Sur, Philippines

ORCID iD: <https://orcid.org/0009-0007-9658-5064>

Luan Sebastian B. Jardin

Mapua Malayan Colleges Mindanao

Gen. Douglas MacArthur Hwy

Matina, Davao City 8000

Davao del Sur, Philippines

ORCID iD: <https://orcid.org/0009-0002-9043-017X>

Raphael M. Amorado

Mapua Malayan Colleges Mindanao

Gen. Douglas MacArthur Hwy

Matina, Davao City 8000

Davao del Sur, Philippines

ORCID iD: <https://orcid.org/0009-0001-9637-1599>

*Corresponding author(s) contact email

To cite this article:

Cerna, P., Luengas, J.G. Jardin, L.S.B. & Amorado, R.M. (2026). Transfil: A hybrid CNN-LSTM framework for multimodal translation of Filipino sign language and facial expressions for non-signer communication. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM)*, 41, 241-250.