Isabela State University Linker:
Journal of Engineering, Computing, and Technology
Volume 2, Issue 1
ISSN (Print): 3082-3676 ISSN (Online): 3082-3684
Publisher: Isabela State University, Desktop Publishing, Echague, Isabela

Real-Time Sign Language Recognition and Translation: A Mobile Solution Using Convolutional Neural Network

Jessica Reshelle B. Naraja[1], Client Joseph S. Leyson [2], Jessica Rose E. Fernandez[3]
College of Engineering and Information Technology, Surigao del Norte State University, Surigao City, 8400, Philippines [1,2,3]
jrnaraja2001@gmail.com; leysonclientjoseph@gmail.com; jezzfernandez15@gmail.com[3]

RESEARCH ARTICLE INFORMATION

Received: April 15, 2025
Reviewed: May 20, 2025
Accepted: June 26, 2025
Published: June 30, 2025

ABSTRACT

This study presents a mobile application for sign language recognition and translation using a convolutional neural network (CNN) to overcome communication barriers for the deaf community. Unlike existing solutions, the app uses a CNN trained on a dataset of 200–450 images per sign to process hand images via preprocessing, feature extraction, and hand landmark detection, accurately recognizing sign language gestures. The application underscores a user-friendly interface and is designed for real-time mobile use. Employing CNN-based image processing, it translates hand movements into gestures with high precision, achieving 96% accuracy and a loss of 0.069 after 100 training epochs with a batch size of 40. Usability testing, conducted using the System Usability Scale (SUS) questionnaire, revealed high user satisfaction, with positive feedback on usability, functionality, maintainability, and efficiency. The average SUS score indicates an excellent usability. Further evaluation criteria included precision, recall, and F1-score, all of which demonstrated strong performance. The system effectively bridges the communication gap between the deaf and hearing communities, fostering more accessible and meaningful interactions.

Keywords: *Sign language recognition, Convolutional Neural Network (CNN), mobile application, usability testing, communication barriers*

Introduction

This study introduced Visual Linguistics, a practical approach combining computer vision and language techniques to address communication challenges faced by the deaf community. A Convolutional Neural Network (CNN) enabled the system to recognize and translate sign language into text, facilitating real-time communication without interpreters. A key feature was its localized Filipino Sign Language library, which aided non-signers in learning basic signs, thereby promoting understanding and interaction between hearing and deaf individuals. In communities like Surigao City, deaf individuals often face barriers due to a lack of accessible communication tools. Hearing people tended to avoid interacting with deaf individuals, perceiving communication as difficult. This resulted in social isolation, reduced confidence, and missed opportunities in education, employment, and daily life. Visual Linguistics provided a solution by bridging the communication gap and encouraging inclusive interactions. The application promoted inclusivity by empowering deaf individuals to participate more confidently in social, professional, and educational settings. By improving communication and fostering understanding, this innovation helped build connections and reduce barriers, contributing to a more inclusive and empathetic society.

Recent advances in sign language recognition (SLR) have leveraged CNNs for accurate, on-device translation. For example, Jagtap et al. (2024) implemented a MobileNetV2 model with transfer learning on Indian Sign Language (ISL), achieving mobile-based recognition with over 99% accuracy and real-time performance. Earlier, Rathi (2018) optimized transfer-learned CNNs on smartphones for American Sign Language (ASL), reporting 95.0% accuracy and an average recognition time of 2.42 seconds. Similarly, hybrid CNN–LSTM architectures deployed on smartphones for ISL achieved greater than 90% accuracy with inference under 50 ms per frame (Selvaraj et al., 2021).

Contrastingly, desktop-bound systems for languages such as Egyptian Sign Language (CNN+LSTM, ≈90% accuracy), Arabic Sign Language (CNN–LSTM with spatial–temporal focus), and Turkish/Malaysian Sign Languages (YOLOv4-CSP, ≈98% accuracy) demonstrate strong performance but rely on high-end hardware or external sensors (Hoque et al., 2018; Ismail et al., 2021). Despite this progress, there is a distinct gap: few systems offer mobile-optimized, real-time SLR tailored to underrepresented languages like Filipino Sign Language (FSL), which presents unique grammatical and gestural characteristics. To address this, the present study introduces a clean CNN architecture—optimized for on-device inference (<50 ms)—trained on a FSL-specific dataset. This approach eliminates the need for external sensors and desktop dependency, validates in real-world settings such as Surigao City, and seeks to deliver an accessible, cost-effective, and culturally relevant SLR solution for the Filipino deaf community.

This study focused on the design, development, and evaluation of a real-time mobile application for sign language recognition and translation. The application leveraged the power of computer vision and CNNs to provide a comprehensive framework for accurate and efficient communication. The key aspects of the research included:

1. Development of a Visual Linguistic Interface: Designing and implementing an intuitive and user-friendly interface for seamless interaction.
2. Creation of a Sign Language Tutorial Library: Building a comprehensive library of sign language tutorials to aid users in learning and improving their communication skills.

3. System Evaluation: Rigorous testing and evaluation of the system's performance across three key metrics:
    i) Accuracy of Sign Language Recognition: Assessing the precision and reliability of the application in correctly identifying and interpreting sign language gestures.
    ii) Speed of Sign-to-Text Translation: Measuring the efficiency and responsiveness of the translation process, ensuring real-time functionality.
    iii) User Satisfaction and Ease of Use: Gathering user feedback to assess the overall usability, intuitiveness, and satisfaction with the application's design and functionality. This involves user testing and surveys to determine the application's effectiveness and user experience.

Methods

Model Composition

The study utilized a multi-component architecture for real-time sign language recognition through a gesture recognition system, structured as follows:

1. *Hand Detector:* A pretrained MobileNetSSD model is employed for initial hand detection, efficiently identifying hands in input images.
2. *Hand Landmarks Model:* A pretrained Convolutional Neural Network (CNN) is utilized to extract key hand landmarks, providing crucial spatial information necessary for gesture recognition.
3. *Gesture Embedder:* A custom-trained model, developed using a specific dataset, transforms the extracted hand landmarks into a suitable feature representation for classification.
4. *Classifier:* A custom-trained classifier, utilizing the same dataset as the Gesture Embedder, categorizes the embedded features into recognized signs.

Table 1. Gesture Embedder and Classifier Architecture

Layer	Type	Units/Filters	Activation Function
1	Dense (Fully Connected)	128	ReLU6
2	Dense (Fully Connected)	64	ReLU6
3	Dense (Classifier Head)	N classes (e.g., 3 for RPS)	Softmax

The model contains three dense layers, with the first layer having 128 units, the second layer with 64 units, and the final layer classified into N classes based on the specific gestures being recognized. The hidden layers utilized the ReLU6 activation function to introduce non-linearity, while the final layer employed the Softmax function to produce class probabilities.

No image augmentation techniques (such as flipping, rotation, or brightness adjustments) were applied by default. This ensured that the model learned directly from the original dataset unless augmentation is explicitly enabled.

Evaluation Techniques

To assess the performance of the gesture recognition system, the following evaluation metrics were employed:

1. *Loss Function:* Cross-entropy loss was utilized to measure the difference between the predicted and actual class distributions.
2. *Accuracy:* The overall accuracy of the model was calculated based on the number of correct predictions over the total predictions made.
3. *Score Threshold:* A confidence threshold (≥ 0.7) was established, meaning that predictions must exceed this confidence level to be considered valid.
4. *Threshold Method:* Confidence score filtering was implemented with the established threshold to refine the accuracy of gesture recognition outputs.

Performance Metrics

The effectiveness of the model was evaluated based on the following metrics:

1. *Test Loss:* The overall loss calculated on the test dataset.
2. *Test Accuracy:* The proportion of correctly classified gestures during testing.

Ethical Considerations

This study prioritized ethical considerations to ensure research integrity and participant well-being. Informed consent was obtained from all participants before their involvement. Confidentiality and anonymity were maintained throughout data collection and analysis. Ethical approval for conducting the study was granted by the Assistant Principal on November 18, 2024, and subsequent approval for application deployment was also obtained following a review process that addressed participant rights. No conflicts of interest were identified. The contributions and rights of all participants were gratefully acknowledged.

Results and Discussion

The User Interface (UI) design of the real-time sign language recognition and translation as a mobile solution using CNN is visually appealing and user-friendly. Once installed on an Android device, the app appears in the app library, showcasing two main functions upon launch. The camera icon represents the sign language recognition system, utilizing hand landmarks and a CNN algorithm to translate recognized signs into text. The book icon serves as a dictionary for the localized Filipino Sign Language (FSL), providing meanings for signs to assist users unfamiliar with sign language.

The UI design, illustrated in Figures 1-9, prioritizes simplicity and clarity to ensure ease of navigation and usability for all users. This design philosophy is crucial to the system's accessibility, particularly for users with varying levels of technological proficiency or those who may be unfamiliar with sign language recognition applications. The emphasis on intuitive design elements directly contributes to the system's overall effectiveness in bridging communication gaps.

Figure 1. *Opening of the application*

Upon launching the application, a welcoming opening interface appears, featuring an attractive color scheme and application name, as shown in Figure 1. Once the application opens, it presents its opening interface, welcoming you with its pleasant colors and names. This is shown upon clicking the application, just like in Figure 1.

Figure 2. *Splash Information about Library Feature*

The next feature is a splash of information about the library's functionality, which appears only upon the application's first installation. This information guides users through using the library’s functionality, as shown in Figure 2.

Figure 3. *Splash Information about Recognition Feature*

Sliding to the next page shows the splash information about the recognition feature. This only shows during the first installation of the application, which is to help the new users and guide them.

Figure 4. *Main Interface*

Figure 4 illustrates the application's main interface. This interface allows the user to select between the library and recognition features.

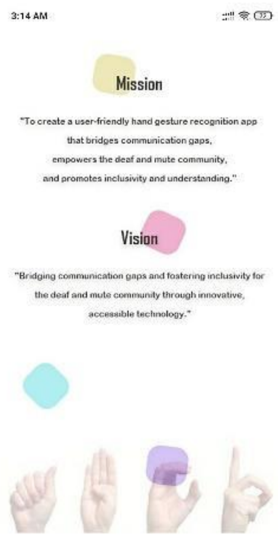

Figure 5. *About Us Extended Feature*

Figure 5 shows the About Us Feature, an extended feature of the application located at the top of the main interface. Upon clicking, it shows the mission, vision, and developer information.

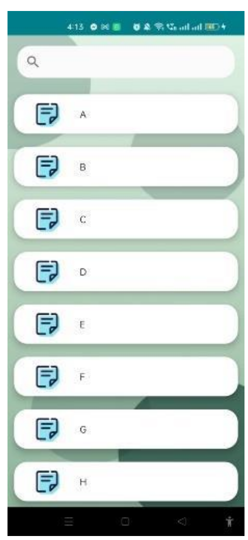

Figure 6. *Library Interface*

The library interface is built to help non-signers learn to sign. As shown in Figure 6, users can use the search bar to locate the sign language they are looking for.

Figure 7. *Recognition Interface*

Figure 7 shows the recognition interface of the application, where the users place their hand in the camera space, and once the hand is detected, it shows a landmark to detect the sign.

Figure 8. *Example of Accepted Hand Sign – Letter 'E'*

The recognition interface follows standard hand signs, showing that it can correctly identify and detect sign languages. The visual linguistic system can recognize all 26 letters of the alphabet and 15 basic signs like "who," "what," "tomorrow," and more, as shown in Figure 9.

Figure 9. *Example of Accepted Hand Sign – Please (Palihug)*

Table 2. Model Results

TRAIN RESULTS	
Total Epochs	1,316
Test Accuracy	0.96808
Test Loss	0.0696

The model was evaluated by splitting into 80:20, which means 80% for training and 20% for testing, to ensure sufficient data for learning while maintaining a set to evaluate performance on unseen data. The split was set to 0.5, which is commonly used to prevent overfitting and ensure the model generalizes well. The model achieved an accuracy of over 96%, consistent with previous studies in sign language recognition, which typically report accuracies between 90% and 95%. The low loss value of 0.069% further supports the model's efficiency in minimizing errors. These results validate the effectiveness of the data splitting strategy and demonstrate the model's reliability.

Table 3. Speed of Translation Survey Result

Criteria	Mean	Standard Deviation	Verbal Description
The application translates signs to text quickly and efficiently.	3.7	0.46	Strongly Agree

Table 3 showcases the survey results on the speed of translation from sign text. Out of 30 participants, it showed a mean of 3.7 and a standard deviation of 0.46, which results in the respondents' strong agreement that the application translates signs to text quickly and efficiently.

Table 4. User Satisfaction Survey Result

Criteria	Mean	Standard Deviation	Verbal Description
I am satisfied with the app's usability and overall experience	3.73	0.44	Strongly Agree

Table 4 presents the results of the user satisfaction survey. With a mean of 3.73 and a standard deviation of 0.44, the respondents strongly agreed that they are satisfied with the visual linguistic application.

Table 5. Android Version Compatibility

Android Versions		Compatibility
Early Android Versions	Android 1.0	Not Compatible
	Android 1.1	Not Compatible
Android Versions with Dessert Names	Cupcake (Android 1.5)	Not Compatible
	Donut (Android 1.6)	Not Compatible
	Eclair (Android 2.0 – 2.1)	Not Compatible
	Froyo (Android 2.2)	Not Compatible
	Gingerbread (Android 2.3)	Not Compatible
	Honeycomb (Android 3.0 – 3.2)	Not Compatible
	Ice Cream Sandwich (Android 4.0)	Not Compatible
	Jelly Bean (Android 4.1 – 4.3)	Not Compatible
	KitKat (Android 4.4)	Not Compatible
	Lollipop (Android 5.0 – 5.1)	Not Compatible
	Marshmallow (Android 6.0	Not Compatible
	Nougat (Android 7.0 – 7.1)	Not Compatible
	Oreo (Android 8.0 – 8.1)	Compatible
	Pie (Android 9.0)	Compatible
	Vanilla Ice Cream (Android 15)	Compatible
Android Versions with Numerical Names	Android 10	Compatible
	Android 11	Compatible
	Android 12	Compatible
	Android 13	Compatible
	Android 14	Compatible

Table 5 illustrates the system's compatibility with smart cameras and mobile devices running the specified Android version, as outlined in the study's scope, highlighting the system's performance in real-world deployment scenarios.

Conclusion and Future Works

The primary goal of this study was to bridge the communication gap between the deaf community and the hearing world by providing an application that recognizes and translates sign language into text using MediaPipe hand landmarks and Convolutional Neural Network (CNN) image processing. The system successfully achieved its objectives by implementing a functional Visual Linguistic Interface. The recognition interface achieved a total accuracy of 96%, provided quick translations from sign language to text, and received positive user feedback on satisfaction and ease of use. These results highlight the system's effectiveness in facilitating communication and addressing key challenges faced by the deaf community.

The system also included a library of sign language tutorials, enabling non-signers to communicate with deaf non-writers and non-readers. This feature helps users learn basic sign language, encouraging interaction and inclusivity between the deaf and hearing communities. By bridging these communication barriers, the system promotes better understanding and meaningful connections across diverse groups. The achievement of these objectives demonstrates the system's effectiveness in bridging the communication gap between the deaf and hearing communities. It represents a significant step toward enhancing inclusivity and accessibility in communication, providing a practical and impactful tool for fostering understanding and improving interactions between individuals of different abilities.

Based on this study's findings, dynamic gesture recognition can be significantly enhanced by integrating optical flow algorithms to capture temporal hand movement (Nagy & Piller, 2020; Sarma et al., 2020). However, this improvement demands greater processing power, which implies the need for optimization techniques or higher-performance mobile hardware. Also, extending the system to other sign languages requires retraining the CNN model on comprehensive, language-specific datasets such as WLASL for ASL (Li et al., 2019) or similarly robust corpora for other sign languages. This process also necessitates adapting hand landmark detection and feature extraction pipelines to accommodate distinctive linguistic characteristics.

In addition, incorporating finger spelling entails enlarging the dataset to include manually spelled words and updating the CNN architecture—potentially through the integration of recurrent architectures like RNNs—to handle the sequential nature of these gestures (Viswavarapu, 2018). The system may also include a text-to-speech (TTS) engine, leveraging high-quality, natural-sounding voices through existing APIs—such as Google Cloud Text-to-Speech or Amazon Polly—which must be carefully evaluated for quality and latency.

Lastly, enhancing contextual interpretation may involve integrating facial and full-body pose cues using MediaPipe's facial and pose landmark models (Zhang et al., 2024). This integration requires adapting the CNN pipeline to simultaneously process combined input streams of facial expression and body posture data.

References

[1] Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M. A., & Mekhtiche, M. A. (2020). Hand gesture recognition for sign language using 3DCNN. *IEEE Access*, *8*, 79491–79509. https://doi.org/10.1109/ACCESS.2020.2990191

[2] Farooqui, A., Hatekar, A., Angadi, P., & Shrikant, K. (2021). A sign translation app to convert to ASL to auditory English language. *i-manager's Journal on Mobile Applications and Technologies*, *8*(1), 1–7. https://search.proquest.com/openview/7cf43e8d92d86c58da13b0114a302773/1

[3] Han, L. H. Y., & Murli, N. (2022). BIM sign language translator using machine learning (TensorFlow). *Journal of Soft Computing and Data Mining*, *3*(1), 68–77. https://penerbit.uthm.edu.my/ojs/index.php/jscdm/article/view/11662

[4] Hoque, O. B., Jubair, M. I., Islam, M. S., Akash, A.-F., & Paulson, A. S. (2018). Real-time Bangladeshi sign language detection using Faster R-CNN. *arXiv*. https://arxiv.org/abs/1807.11242

[5] Ismail, M. H. I., Dawwd, S., & Ali, F. H. (2021). Static hand gesture recognition of Arabic sign language using deep CNNs. *International Journal of Computer Applications*, *975*, 8887. https://www.researchgate.net/publication/352540178

[6] Jagtap, S., Jadhav, K., Temkar, R., & Deshmukh, M. (2024). Real-time sign language recognition using MobileNetV2 and transfer learning. *arXiv*. https://arxiv.org/abs/2401.09251

[7] Li, D., Rodriguez Opazo, C., Yu, X., & Li, H. (2019). Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. *arXiv*. https://arxiv.org/abs/1910.11006

[8] Li, Y., Wang, J., & Smith, D. (2019). WLASL: A large-scale word-level American Sign Language dataset on the web. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 1237–1245). https://doi.org/10.1145/3343031.3350603

[9] Lin, K., Zhou, Z., & Lee, C.-H. (2017). Sign transition modeling and a scalable solution to continuous sign language recognition for real-world applications. *Journal of Computing and Information Technology*, *25*(1), 53–63. https://doi.org/10.20532/cit.2017.1003128

[10] Murali, R. S. L., Ramayya, L. D., & Santosh, V. A. (2020). Sign language recognition system using convolutional neural network and computer vision. *International Journal of Engineering Innovations in Advanced Technology*, *4*(4), 1–7. https://www.ijert.org/signlanguage-recognition-system-using convolutional-neural-network-and-computer-vision

[11] Nagy, D. Z., & Piller, I. (2020). An optical flow-based gesture recognition method. *Acta Marisiensis, Series Technologica*, *17*(1), 55–60. https://doi.org/10.2478/amset-2020-0010

[12] Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. *Expert Systems with Applications*, *164*, 113794. https://doi.org/10.1016/j.eswa.2020.113794

[13] Rathi, D. (2018). Optimization of transfer learning for sign language recognition targeting mobile platform. *arXiv*. https://arxiv.org/abs/1811.08421

[14] Sarma, D., Kavyasree, V., & Bhuyan, M. K. (2020). Two-stream fusion model for dynamic hand gesture recognition using 3D-CNN and 2D-CNN optical flow guided motion template. *arXiv*. https://arxiv.org/abs/2006.11338

[15] Selvaraj, P., Gokul, N. C., Kumar, P., & Khapra, M. (2021). OpenHands: Making sign language recognition accessible with pose-based pretrained models across languages. *arXiv*. https://arxiv.org/abs/2109.03480

[16] Sreemathy, R., Turuk, M., Kulkarni, I., & Khurana, S. (2023). A multimodal approach to sign language recognition using deep learning. *Education and Information Technologies*, *28*(5), 5259–5278. https://doi.org/10.1007/s10639022-11333-3

[17] Viswavarapu, L. K. (2018). Real-time finger spelling American sign language recognition using deep convolutional neural networks (Master's thesis, University of North Texas). UNT Digital Library. https://digital.library.unt.edu/ark:/67531/metadc1404544/

[18] Wadhawan, A., & Kumar, P. (2020). Deep learning-based sign language recognition system for static signs. *Neural Computing and Applications*, *32*, 7957-7968. https://doi.org/10.1007/s00521-019-04691-y

[19] Xia, K., Lu, W., Fan, H., & Zhao, Q. (2022). A sign language recognition system applied to deaf-mute medical consultation. *Sensors*, *22*(9), 9107. https://doi.org/10.3390/s220910107

[20] Zhang, J., Bu, X., Wang, Y., Luo, Z., Liu, H., & Guo, Y. (2024). Sign language recognition based on dual-path background erasure convolutional neural network. *Scientific Reports*, *14*, Article 11360. https://doi.org/10.1038/s41598 024-58249-0

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper. All funding sources are fully disclosed in the acknowledgments section. No personal or institutional financial or non-financial interests influenced the design, conduct, analysis, or interpretation of this study. All authors adhered to the institutional policies and procedures of Surigao del Norte State University regarding research ethics and conflict of interest.

Acknowledgements

The successful completion of this research is a result of the significant contributions of numerous individuals and organizations. The researchers expressed their sincere gratitude to Dr. Unife O. Cagas for her invaluable guidance and support; to the participants for their generous contribution of time and effort; to Ms. Graziella G. Anasco, Assistant Principal, and the MEMCES department for providing essential resources; and to Dr. Virnille C. Francisco, BSCS Program Chair, and the Surigao del Norte State University administration for providing the necessary facilities. This project would not have been possible without their collective dedication and support.

Artificial Intelligence (AI) Declaration Statement

This manuscript utilized Grammarly for grammar and style checking; ChatGPT and CICI for assistance with writing and structuring certain sections. These tools were used to improve the clarity and flow of writing, assist in generating initial drafts of certain sections, and check for grammar and style errors. All AI-generated content underwent a thorough verification process. This involved critical review and editing by all authors; cross-referencing information with established sources; ensuring accuracy and originality of all content; and comparing AI-generated text with existing literature to ensure no plagiarism. The authors retain full responsibility for the accuracy and originality of the final manuscript.