International Journal of Computing Sciences Research (ISSN print: 2546-0552; ISSN online: 2546-115X)
Vol. 8, pp. 3252-3267
doi: 10.25147/ijcsr.2017.001.1.217
https://stepacademic.net

Short Paper*

Filipino Sign Language Hand Gesture Recognition Using MediaPipe and Machine Learning

Lorraine Kaye M. Pilare
Engineering Department, University of San Jose-Recoletos, Philippines
pilarelorraine@gmail.com
(corresponding author)

Junejay Christian A. Mahinay
Engineering Department, University of San Jose-Recoletos, Philippines
Mahinay.Junejay@gmail.com

Augustine Clein C. Degamo
Engineering Department, University of San Jose-Recoletos, Philippines
clien.degamo@gmail.com

Brill Nash C. Piner
Engineering Department, University of San Jose-Recoletos, Philippines
brillnash.piner@gmail.com

Date received: May 7, 2024
Date received in revised form: July 13, 2024; July 17, 2024
Date accepted: July 17, 2024

Recommended citation:

Pilare, L. K., Mahinay, J. C., Degamo, A. C., & Piner, B. N. (2024). Filipino Sign Language Hand Gesture Recognition Using MediaPipe and Machine Learning. *International Journal of Computing Sciences Research, 8*, 3252-3267. https://doi.org/10.25147/ijcsr.2017.001.1.217

**Special Issue on International Research Conference on Computer Engineering and Technology Education (IRCCETE). Guest Associate Editors: Dr. Roben A. Juanatas (National University-Manila) and Dr. Nelson C. Rodelas (University of East).*

Abstract

Purpose –This study develops a Filipino sign language recognition system that can recognize sign language composed of three (3) basic words and twenty-seven (27) phrases.

Method – The proposed sign language recognition method converts sign language into text through several steps. First, video input is captured using a webcam. MediaPipe then extracts features of both the left and right hands from the video. An LSTM algorithm is trained to recognize patterns in these hand features, translating sign language into text accurately. The translated text is displayed on a monitor. A total of 900 data samples were used, with an output shape of (900, 30, 128) for a video containing 30 frames and 64 hand landmarks per frame.

Results – The proponents performed test cases to confirm that the system met the necessary quality standards, using three different dataset splits. The highest training accuracy percentage was 98.41% with 70% of training and 30% samples. The highest testing accuracy percentage was 98.89% with an even 50% split between training and testing samples. The system achieved high accuracy with a 50% training and 50% testing split.

Conclusions – This study developed a sign language recognition system for Filipino Sign Language (FSL) gestures using MediaPipe and LSTM algorithm, achieving high accuracy This advancement in sign language recognition can contribute significantly to the field.

Recommendations – To enhance the FSL gesture recognition system, several suggestions could focus on improving recognition for natural signing, exploring advanced machine learning models, expanding gesture recognition scope, and integrating high-performance hardware.

Research Implications – Using machine learning frameworks like MediaPipe, this study aims to improve the accuracy and effectiveness of FSL recognition systems.

Practical Implications – The system enhances communication between the speech impaired and the wider community, improving accessibility and inclusiveness.

Keywords – machine learning, Filipino sign language, mediaPipe, LSTM

INTRODUCTION

According to the World Federation of Deaf, there are approximately 360 million people who have hearing loss and 70 million people who are speech impaired. These individuals often rely on lip-reading or sign language for communication. Sign language is the most common method of communication for speech-impaired individuals (Kausar & Javed, 2011). Still, it varies from region to region, and there are approximately 300 different sign languages in use worldwide.

Filipino sign language (FSL) is a visual-manual language used by the deaf community in the Philippines as a means of communication (Sandjaja & Marcos, 2009). It uses hand gestures, body language, and facial expressions to convey meaning, and it has its own grammar, vocabulary, and syntax. According to the Philippine Deaf Research Center and the Philippine Federation of the Deaf (2004), FSL has five components: hand shape, location, palm orientation, movement, and non-manual signals. In sign languages, information is conveyed visually through manual and non-manual means of expression (Tolba & Elons, 2013). This was the only form of communication available for speech-impaired, and it played a vital role in their ability to communicate with one another. Despite the significance of sign language, there is still a notable communication gap between speech-impaired and those who can hear due to the complexity of learning and understanding sign language (Tharwat et al., 2015). This gap creates barriers to communication and accessibility, as they may not have equal access to information and resources, highlighting the need for innovative solutions. Efforts are being made to bridge this gap by developing technologies capable of recognizing and translating sign languages.

Nowadays, researchers have been able to propose and implement systems that target sign language recognition and use paradigm shifts in numerous technology fields (Tolba & Elons, 2013). The study of sign language recognition has recently grown in importance (El-Bendary et al., 2010). For various sign languages, such as American Sign Language, Korean Sign Language, Chinese Sign Language, and so on, several works on sign language recognition have been proposed (Brashear et al., 2006). Sign language recognition is essential for communication with the speech impaired as they cannot speak the language (Cabalfin et al., 2012). This technology has the potential to improve communication and accessibility for the speech impaired. It is an active area of research and development in computer science and artificial intelligence.

There is a growing interest in the use of computer vision and machine learning algorithms to develop sign language recognition systems that can classify and predict sign language gestures. MediaPipe and long short-term memory (LSTM) are often employed in developing sign language recognition systems due to their ability to improve the accuracy and effectiveness of the systems (Putra et al., 2022). Thus, this study develops a sign language recognition system using MediaPipe and long short-term memory (LSTM) for recognizing sign language composed of three (3) basic words and twenty-seven (27) phrases.

LITERATURE REVIEW

The study of Pramada et al. (2013), stated that speech-impaired people are frequently denied daily communication with other members of society. According to the World Health Organization (WHO), over 5% of the world's population has hearing and speaking disabilities. It is observed that they find it difficult to interact with people who can speak, with their gestures, as there are only a few people who can recognize or know how to use sign language (Pramada et al., 2013). The primary language of communication for speech-impaired people is sign language. This can be a hindrance in day-to-day communications for them. Conversion of sign language to text can be a possible solution to this obstacle (Arora & Roy, 2018).

Long Short-Term Memory

Type of recurrent neural network that can learn and retain long-term dependencies in sequential data, making it well-suited for tasks such as language translation and language modeling (Koller et al., 2017).

Speech-impaired

A speech-impaired person has a mild to severe impairment in producing speech sounds. Speech impediments are caused by a variety of factors. Stuttering, for example, could be a sign of a developmental delay, it could be hereditary, or it could occur because a child's brain is unable to coordinate the functions that allow them to speak.

MediaPipe

MediaPipe hand is a finger-tracking solution. It employs machine learning (ML) to infer 21 3D landmarks of a hand from just a single frame (Lugaresi et al., 2019).

Convolution Neural Networks (CNN)

Type of neural network that is particularly well-suited for image classification tasks, as they can learn features from input data and use them to classify the data into predefined categories (Quiroga et al., 2017).

Sign Language Recognition System

The study by Sandjaja and Marcos (2009) focused on developing a Filipino Sign Language (FSL) number recognition system. The system used a 5000-character FSL video file as input, with a frame rate of 15 frames per second and a resolution of 640 by 480 pixels. To extract essential video features, the system employed a multi-color tracking algorithm, as it did not use recursive methods. During the training and testing phases, the system utilized a hidden Markov model (HMM) to learn and recognize FSL numbers. The system achieved a tracking rate of 92.3% for all objects and an average recognition accuracy of 85.52% for FSL numbers.

The study on Filipino Sign Language (FSL) recognition using MediaPipe and machine learning presents a novel approach to recognizing FSL gestures. While there are studies on sign language recognition for various languages, including American Sign Language and Korean Sign Language, the specific focus on FSL is relatively limited. The utilization of MediaPipe for feature extraction and the LSTM algorithm for classification in the context of FSL recognition is a unique aspect of this research. This study fills a gap in the literature by addressing the communication challenges faced by speech-impaired individuals in the Philippines through FSL recognition.

METHODOLOGY

System Architecture

Figure 1 is designed specifically for detecting and recognizing hand sign language using a webcam and a laptop. The input video is collected using a Logitech c922 Pro HD Stream webcam. A laptop with an Intel Core i5 processor and 8GB of RAM is used to develop and run the system application, ensuring it has enough power to process the video data. The desktop application and a separate monitor collect data and display the generated text. An essential aspect of the system design is the accurate detection of all hand landmarks. This is why a separate monitor is included to ensure that all landmarks are correctly detected during the data collection process. The hand gestures are displayed on the monitors, allowing for the simultaneous view of both the left and right hands. The desktop application is responsible for processing the video data collected using MediaPipe and LSTM algorithms. Users can control the system through the desktop application, including training the system on new gestures, testing the system, and modifying settings.

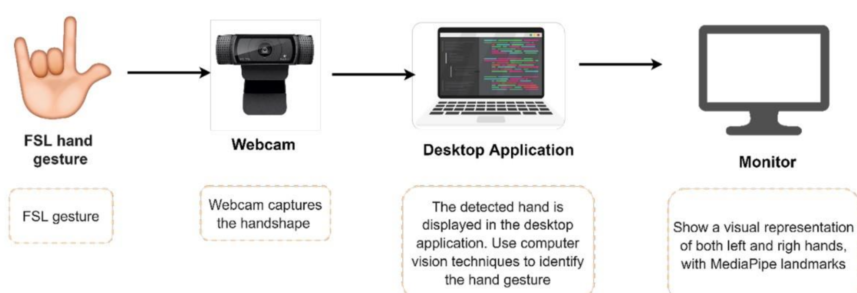

Figure 1. System Architecture

The system for detecting and recognizing Filipino sign language, as shown in Figure 2, starts by collecting data using a webcam. During this process, a speech-impaired individual performs a sign language gesture in front of the camera. The captured video is fed through MediaPipe, to obtain landmark points of both the left and right hands. These landmark points are used to identify the different sign language gestures. Once the data is collected, it is prepared for training and testing. The dataset is typically divided into two sets: a training set, which is used to train a machine learning model, and a testing set, which is used to evaluate the performance of the trained model. The LTSM recognizes the sign language gesture based on the input video. The prediction is then compared to the actual sign language gesture to determine the accuracy of the prediction. The system utilizes MediaPipe and LSTM to improve communication for speech-impaired individuals, allowing them to express themselves and better understand the world around them effectively.

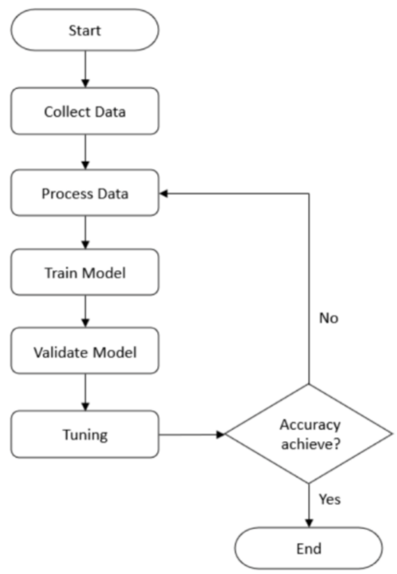

Figure 2. System Flowchart

Data Collection

The system application required a webcam and laptop. To capture the image, we used the Logitech C922 Pro HD Stream webcam as the video capture device for thirty (30) sign language words and phrases. A laptop with an Intel Core i5 processor with 8 GB RAM is used for running the system. Peripherals such as a monitor, a chair, and a ring light are used to conduct the collection of data properly. In data collection, all hand landmarks of the hands must be detected. As a response, we utilized the ring light to provide optimum illumination and avoid shadows from overlapping the essential details. The camera is placed 0.67 feet from its base, and the chair must be 1.96 feet in height and 2.75 feet away from the set table. These specifications were created to show the signer's torso and head.

Figure 3. Data Collection Set-up

Data is essential to develop machine learning, and current efforts in sign language recognition (SLR) are often limited by inadequate data. To obtain appropriate sign language recognition accuracy, a large-scale dataset is required. For Filipino sign language (FSL), there is a publicly available dataset. However, these are created to focus on static signals such as numbers and alphabets. Existing datasets are limited due to a lack of sign language expertise and the expensive cost of annotation by most annotators. Considering these factors, we decided to create our datasets. Since sign language is not an easy language to learn, we invited participants in this study who had a foundational understanding of sign language. To create the dataset, two (2) speech-impaired from the Deaf Association of Lapu-Lapu Cebu City were invited to the data collection process. The two (2) subjects were both 30-year-old women with an average height of 5.17 feet. Data collection was carried out in the computer laboratory for proper lighting and space, which can help improve the detection accuracy of the sign language.

Figure 4. Sample Image for Data Collection

Data Feature Extraction

A total of nine hundred (900) sign language samples were collected from two (2) speech-impaired individuals. Each individual contributed four hundred fifty (450) samples by performing fifty (15) different sign languages, with each sign language consisting of thirty (30) videos. Each video contains 30 frames of the sign languages performed, and each frame has extracted key points of hand landmarks which consists of 64 hand landmarks in one hand since the two hands are used in the collection, a total of 128 hand landmarks for each frame - the output shape results in (900, 30, 128). The obtained landmark points are then stored in one NumPy file.

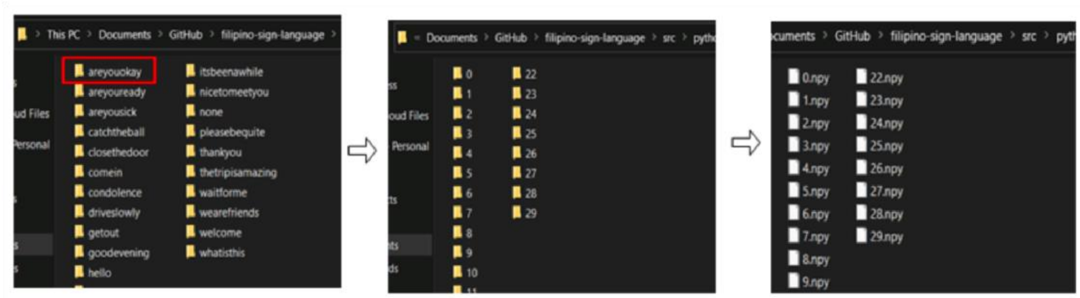

Figure 5. Corresponding Folders for Each Gesture

Data Segmentation

In this study, three cases were considered for using the data collected as a dataset for the machine learning operation. The training and testing datasets were divided into partition cases to select and compare the optimum results. Table 1.0 shows the partition cases for the training and testing dataset.

Table 1. Cases of Data Partition

Partition Cases	Number of Training Data	Number of Testing Data	Total
80% - 20%	720	180	900
70% - 30%	630	270	900
50% - 50%	450	450	900

The partition cases described involve dividing a set of samples of sign language into two sets: a training set and a testing set. The training set is used to train a model, while the testing set is used to evaluate the model's performance.

Validation Techniques

The Confusion matrix is one of the most flexible and uncomplicated metrics used for finding the accuracy and correctness of the algorithm and is also the basis of almost all the evaluation metrics (Sunasra, 2017). As illustrated by Sunasra (2017), precision, accuracy, recall or sensitivity, and specificity can be computed with the use of the confusion matrix.

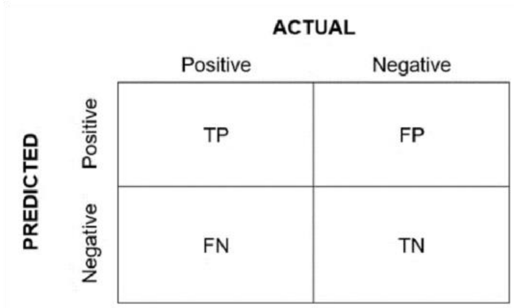

Figure 6. Confusion Matrix

Table 2. FSL Words and Phrases for Data Collection

English Translation	Filipino Translation
1. Good Evening	Magandang Gabi
2. Hello	Kumusta
3. Nice to meet you	Ikinagagalak kitang makilala
4. It's been a while	Matagal na tayong hindi nagkita
5. Welcome	Maligayang pagdating
6. How do you do?	Kamusta ka?
7. Catch the Ball	Saluhin ang bola
8. Come in	Pasok ka
9. Wait for me	Hintayin mo ako
10. Please be quiet	Manahimik ka
11. Close the door	Isaramo ang pinto
12. Drive slowly	Bagalan mo mag maneho
13. I am alone	Mag-isa lang ako
14. I am heartbroken	Heartbroken ako
15. I am sad	Malungkot ako
16. Condolence	Pakikiramay
17. I hurt myself	Sinaktan ko sarili ko
18. I'm disappointed	Bigoako
19. Are you sick?	May sakit ka ba?
20. Are you ready?	Handa kana ba?
21. Are you okay?	Ayos ka lang ba?
22. How much is this?	Magkanoito?
23. What is this?	Anoito?
24. Did you understand?	Naiintindihan mo ba?
25. I am happy	Ako ay masaya
26. I am excited	Natutuwaako
27. We are friends	Magkaibigan tayo
28. Thank you	Salamat
29. I am proud of you	Ipinagmamalakikita
30. The trip is amazing	Nakamangha ang biyahe

RESULTS

Classification Results

The following are the results of the LSTM algorithm according to the partition cases made to select better results.

For Case 1: 80% Training and 20% Testing

The model achieved a success rate of 97.22%, accurately recognizing 175 out of 180 phrases. While most phrases like "hello" and "thank you" were perfectly recognized, there were some exceptions. The model struggled with specific phrases like "come in" and "drive slowly", resulting in lower accuracies between 80% and 89.89%. To gain deeper insights into these errors, researchers can refer to the mentioned confusion matrix (Figure 7). This matrix would presumably reveal how often the model confused one phrase for another, providing valuable clues to improve the model's ability to recognize these challenging phrases. Overall, this study offers valuable data for researchers to analyze the model's strengths and weaknesses, paving the way for further development and improved performance.

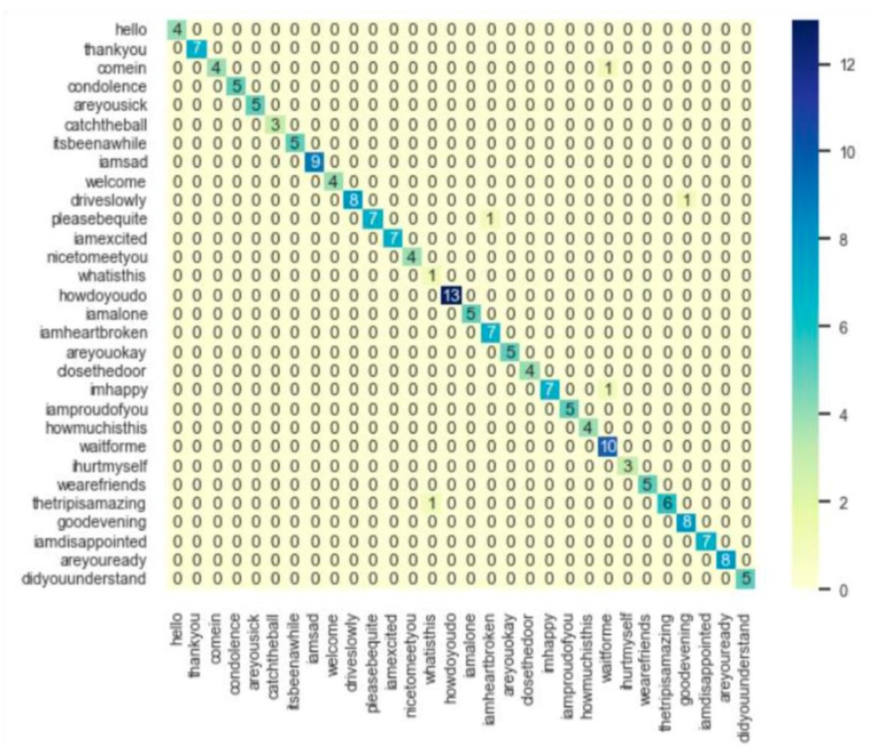

Figure 7. Confusion Matrix for Testing Samples, 20%

For Case 2: 70% Training and 30% Testing

The model achieved a high accuracy of 97.41%, correctly identifying 263 out of 270 test phrases. There was variation in performance across different phrases. Some phrases, like "hello" and "thank you," were recognized perfectly (100% accuracy). However, the model encountered difficulties with other phrases. For instance, phrases like "wait for me" had a lower accuracy (82.24%) compared to "I am proud of you" (91.77%).

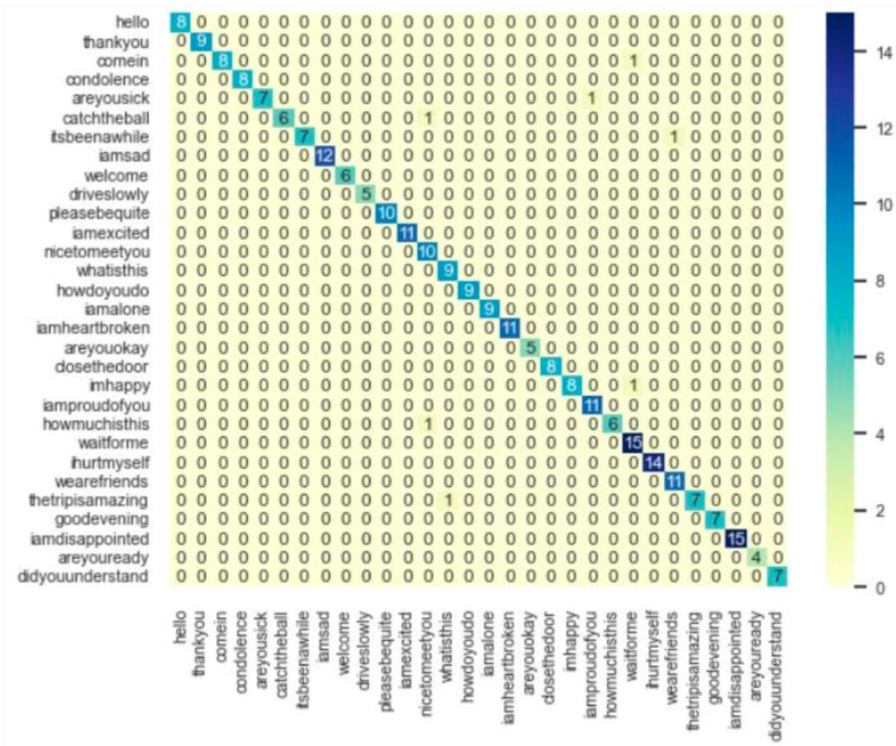

Figure 8. Confusion Matrix for Testing Samples, 30%

For Case 3: 50% Training and 50% Testing

The model achieved a remarkable accuracy of 98.89%, correctly identifying 445 phrases. Similar to previous findings, performance varied across phrases. Basic greetings like "hello" and "thank you" were recognized flawlessly (100% accuracy). The model continued to struggle with specific phrases, though to a lesser extent. For instance, "come in" had an accuracy of 88.89% compared to "good evening" at 94.27%.o "good evening" at 94.27%.

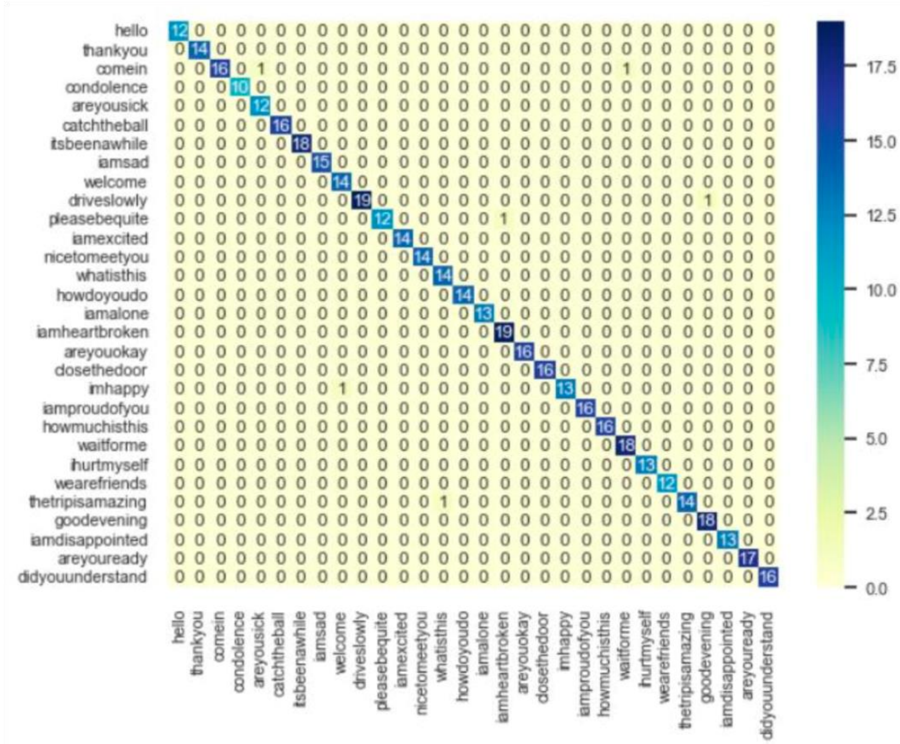

Figure 9. Confusion Matrix for Testing Samples, 50%

DISCUSSION

In Figure 7, the LSTM model achieved high accuracy, precision, recall, and F1_score for both training and testing samples, showcasing the effectiveness of the algorithm in recognizing sign language phrases. The results indicate a strong performance with an accuracy of 98.33% for training samples and 97.22% for testing samples.

In Figure 8, the LSTM model's performance is further highlighted with an accuracy of 98.41% for training samples and 97.41% for testing samples. The precision, recall, and F1_score metrics also demonstrate the model's robustness in classifying sign language phrases accurately. These results underscore the consistency of the LSTM algorithm across different sample partitions, emphasizing its reliability in recognizing sign language gestures.

Lastly, figure 9 presents the results for the 50%-50% partition of sample data, showing an accuracy of 97.56% for training samples and 98.67% for testing samples. The precision, recall, and F1_score metrics further reinforce the algorithm's proficiency in classifying sign language composed of essential words and phrases. The results indicate a high level of accuracy and effectiveness in recognizing sign language gestures using the LSTM algorithm.

CONCLUSIONS AND RECOMMENDATIONS

The study developed a Filipino Sign Language (FSL) hand gesture recognition system using MediaPipe and machine learning. The system uses a desktop application implemented in C# and Python, interfacing with a Logitech C922 Pro webcam for capturing sign language gestures. The system uses LSTM as the architecture for classification, recognizing three words and twenty-seven phrases in FSL. Three cases were tested with different training and testing sample ratios, achieving the highest training accuracy of 98.41% in case 2 and the highest testing accuracy of 98.89% in case 3. The system has potential real-world applications, including improving communication between speech-impaired individuals and the hearing population, facilitating education for speech-impaired students, and enhancing accessibility in public spaces and digital platforms. However, limitations include the need for optimal lighting conditions, requiring signers to perform gestures slowly, and relying on handshape input, which may not capture the full range of FSL gestures.

The study demonstrates promising results, but there are opportunities to enhance the system's capabilities. The system's capacity can be improved by incorporating a wider range of gestures, as the current system is limited to three words and twenty-seven phrases. Additionally, the system required signers to perform gestures slowly for optimal recognition, which may not be practical in real-world scenarios. The researchers suggest exploring the potential benefits of combining Long Short-Term Memory (LSTM) networks with other machine learning models to improve the system's performance. Furthermore, utilizing higher-specification hardware can enhance the system's performance by allowing for smoother processing and potentially increasing recognition accuracy

IMPLICATIONS

The study focuses on developing a Filipino Sign Language (FSL) recognition system using MediaPipe and long short-term memory (LSTM) to recognize basic words and phrases in FSL. It addresses the communication challenges faced by speech-impaired individuals, emphasizing the importance of converting sign language to text for improved communication. By using machine learning frameworks like MediaPipe, the study aims to enhance the accuracy and effectiveness of FSL recognition systems, catering specifically to the Filipino context. The system's scope includes classifying and predicting FSL words and phrases based on handshape input, utilizing LSTM algorithms for classification, and providing a dictionary of FSL for user reference. The significance of the study extends to benefiting Filipino speech-impaired individuals, SPED teachers, researchers, and those unfamiliar with sign language, facilitating learning and communication in the FSL domain.

ACKNOWLEDGEMENT

The *Researchers* would like to express our greatest gratitude to the following persons who made the development of this project successful. *Project Adviser*, Engr. Jullene Jane Evangelist, Engr. Eutequio III Zante L. Palermo, Engr. Shiela Cabahug, to our family and the Almighty God.

FUNDING

The study did not receive funding from any institution, which limited the scope and scale of the research project. This independent funding status underscores the researcher's commitment to conducting the study without any external influence.

DECLARATIONS

Conflict of Interest

The researcher declares no conflict of interest in this study. Disclosing any potential conflicts that could influence the study's outcomes is crucial for maintaining transparency and ethical research practices.

Informed Consent

Full consent to all participants was agreed upon before taking part in the conduct of this study.

Ethics Approval

No humans were involved and the persons in the figures are the researchers themselves. The figures exclusively represent the researchers conducting the study.

REFERENCES

Arora, S., & Roy, A. (2018). *Recognition of sign language using image processing.* International Journal of Business Intelligence and Data Mining, 13(1-3), 163-176. Inderscience Enterprises Ltd. https://doi.org/10.1504/ijbidm.2018.088428

Brashear, H., Henderson, V., Park, K. H., Hamilton, H., Lee, S., & Starner, T. (2006, October). *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). Association for Computing Machinery (ACM). https://doi.org/10.1145/1168987.1169002

Cabalfin, E. P., Martinez, L. B., Guevara, R. C. L., & Naval, P. C. (2012, November). Filipino sign language recognition using manifold projection learning. In *Proceedings of TENCON 2012 IEEE Region 10 Conference* (pp. 1-5). IEEE. https://doi.org/10.1109/tencon.2012.6412231

El-Bendary, N., Zawbaa, H. M., Daoud, M. S., Hassanien, A. E., & Nakamatsu, K. (2010, October). *Arslat: Arabic sign language alphabets translator*. In 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM) (pp. 590-595). IEEE. https://doi.org/10.1109/cisim.2010.5643519

Kausar, S., & Javed, M. Y. (2011, December). *A survey on sign language recognition.* In *2011 Frontiers of Information Technology* (pp. 95-98). IEEE. https://doi.org/10.1109/fit.2011.25

Koller, O., Zargaran, S., & Ney, H. (2017, July). *Re-sign: Re-aligned end-to-end sequence modelling with deep recurrent CNN-HMMs*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4297-4305). CVPR. https://doi.org/10.1109/cvpr.2017.364

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C., Yong, M. G., Lee, J., Chang, W., Hua, W., Georg, M., & Grundmann, M. (2019). *MediaPipe: A framework for building perception pipelines.* arXiv. https://doi.org/10.48550/arxiv.1906.08172

Philippine Deaf Research Center and Philippine Federation of the Deaf. (2004). *An introduction to Filipino sign language: (a). Understanding structure*. Philippine Deaf Resource Center, Incorporated.

Pramada, S., Saylee, D., Pranita, N., Samiksha, N., & Vaidya, M. S. (2013, February). *Intelligent sign language recognition using image processing*. *IOSR Journal of Engineering (IOSRJEN)*, *3*(2), 45-51. https://doi.org/10.9790/3021-03224551

Putra, I. A., Nurhayati, O. D., & Eridani, D. (2022, August). *Human Action Recognition (HAR) Classification Using MediaPipe and Long Short-Term Memory (LSTM). TEKNIK*, *43*(2), 190-201. https://doi.org/10.14710/teknik.v43i2.46439

Quiroga, F. M., Torrents-Barrena, J., Lanzarini, L. C., & Puig-Valls, D. (2017, December). *Invariance measures for neural networks. Applied Soft Computing*, *132*, 109817. Elsevier. https://doi.org/10.1016/j.asoc.2022.109817

Sandjaja, I. N., & Marcos, N. (2009, August). *Sign language number recognition*. In *2009 Fifth International Joint Conference on INC, IMS and IDC* (pp. 1503-1508). IEEE. https://doi.org/10.1109/ncm.2009.357

Sunasra, M. (2020, May 29). *Performance metrics for classification problems in machine learning.* Retrieved from https://medium.com/@MohammedS/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b

Tolba, M. F., & Elons, A. S. (2013, November). *Recent developments in sign language recognition systems*. In *2013 8th International Conference on Computer Engineering & Systems (ICCES)* (pp. xxxvi-xlii). IEEE. https://doi.org/10.1109/icces.2013.6707157

Tharwat, A., Gaber, T., Hassanien, A. E., Shahin, M. K., & Refaat, B. (2015, January). SIFT-based Arabic sign language recognition system. In Hassenien, A. E., Shaalan, K., & Tolba, M. F. (Eds.), *Advances in Intelligent Systems and Computing* (Vol. 334, pp. 359–370). Springer. https://doi.org/10.1007/978-3-319-13572-4_30

World Health Organization: WHO. (2018, February 2). *Deafness and hearing loss*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

Author's Biography

Lorraine Kaye M. Pilare, Junejay Christian Mahinay, Augustine Clein Degamo, and Brill Nash Piner are all fourth-year Computer Engineering students at the University of San Jose-Recoletos upon the completion of this thesis.