JEEAR, Vol. 3 (1), 2024
DOI: 10.34002/jeear.v3i1.75
RESEARCH ARTICLE

WiKA: A Vision Based Sign Language Recognition from Extracted Hand Joint Features using DeepLabCut

Tanya M. Osorio*, Raineer A. Mozo, Kenned D. Sabanilla, and Noel S. Gunay

College of Engineering, Mindanao State University – General Santos,
Fatima, General Santos City, 9500, Philippines
*Corresponding author: tanya.osorio@msugensan.edu.ph

Received: November 8, 2023; Accepted: April 5, 2024; Published Online: April 12, 2024

Abstract

The use of hand gesture recognition for sign language translation to address the communication gap between the hearing majority and the deaf population has had significant breakthroughs over the years. While the contact-based approach uses wearable devices, a vision-based solution is preferred owing to the convenience it offers and since it obviates the need for complicated gears. This study presents the development of WiKA, an open-source software designed to track the joints of the hands and interpret them into their corresponding sign language counterparts. DeepLabCut, a markerless pose estimation software, was employed to develop the Hand-Joint Tracking Model through the training of a sequential Convolutional Neural Network, utilizing extracted Hand-Joint features to predict the sign language alphabets (A-Z) and numbers (1-9) based on the positioning of the joints. The developed Hand-Joint tracking model exhibited a 4.92% training error and a 5.74% test error with a p-cut-off of 60%. On the other hand, the developed sign language recognition achieved a 96.44% prediction accuracy with only 0.0356% misclass. This model can be further integrated into mobile phones for seamless conversations between the signing and non-signing populations.

Keywords: *Filipino sign language, hand-joint tracking, skeletal data, pose estimation, convolutional neural network*

Filipino Sign Language (FSL), used by deaf people in the Philippines, traces its origins back to American Sign Language (ASL) [1]. In the Philippines, the overall number of deaf, mute, or hearing-impaired people comprises roughly 1.23% of the population. FSL has proven its utility by bridging the deaf people and the hearing majority in the Philippines and elsewhere in the world [2].

Most hearing Filipinos do not understand FSL, and mastering it often requires formal instruction [3] creating a linguistic barrier and a noticeable communication gap between the deaf population and the hearing majority [4]. This gap hampers meaningful interactions, limits education opportunities, and feeds a cycle of exclusion that affects the well-being and advancement of the deaf community. The prevalence of hearing Filipinos not understanding FSL has not only formed a gap in terms of expression but has contributed to the cultural disparity and social isolation of the non-hearing population. Fostering social inclusivity and equal opportunities for the deaf community provided much-needed impetus for the researchers to explore the use of vision-based technologies by providing an FSL medium, through Automated Sign Language Recognition (ASLR), that can be interpreted and understood by the broader society [5, 6].

Automated Sign Language Recognition (SLR) innovations incorporate different techniques and algorithms for identifying visual signals and translating their meanings. Several innovations in automated SLR have been explored over the years [7] citing contact-based [8], vision-based [9] and hybrid architecture [10] as the more popular approaches in creating SLR systems. The contact-based approach uses wearable devices (such as gloves, accelerometers and bands) that send data (in the form of movement changes) to a computer. A vision-based system, on the other hand, eliminates the need for complicated gears since it only requires a camera to acquire input data. Hybrid architecture, on the other hand, incorporates features of both approaches. Major limitations of both contact-based and hybrid architecture, such as their costly development and challenging implementation, were however pointed out by several studies [11, 12].

In the Philippine setting, SLR technology has made great strides over the years. A study conducted in 2008 presented a research roadmap that envisions the Philippines as a hub for SLR and computational linguistics in the next twenty years. At the moment however, the development of a successful recognition system in the country is constrained by the need for a large amount of data and requisite expertise in diverse fields that include image processing, computer vision, natural language processing, human-computer interaction, linguistics and knowledge of Filipino deaf culture [18].

At the De La Salle University (DLSU) in Manila, studies on FSL number recognition focused on Non-maximum Suppression (NMS) feature extraction and recognition of continuous signing, as well as feature extraction and recognition from a video stream utilizing the Hidden Markov Model and Artificial Neuron Fuzzy Network. A prior program for deaf students at DLSU served as a key component for further studies that have made significant advances in sign language recognition and its applications. These studies included handshape recognition of still images, sign language modelers, translators, hand poses graphical interface, animation, teaching applications, and communications applications. Cabalfin et al. used computer vision and the Convolutional Neural Network (CNN) ResNet architecture to develop an automatic FSL recognition model based on static pictures of Filipino number signs from 1 to 10. Their research revealed that the ResNet-50 model, after approximately 15 epochs, achieved the highest validation accuracy of 86.7%, which is relevant to a similar ResNet50 model used in the current research focusing on Filipino number signs as static gestures [18]. Rivera et al explored manifold learning as a representation for detecting visual signals in FSL. This team of researchers utilized Isomap, a non-linear manifold learning technique, to train visual indicators, projecting individual signs onto a reference manifold. The study incorporated Dynamic Time Warping or Longest Common Subsequence Similarity Matching for recognition, achieving recognition rates above 80%. The research also delved into the importance of non-manual signals in Sign Language Recognition (SLR), utilizing facial recognition with Microsoft Kinect for Windows 2.0 to analyze emotions and other characteristics. The study emphasized the significance of non-manual signals in affecting the meaning of signs when combined with manual signals. Furthermore, Rivera introduced a sign language number recognition system that employed a multi-color tracking approach with a Hidden Markov Model for training and testing. Remarkably, this system achieved a notable accuracy of 85.52% in recognizing Filipino sign language numerals from a video dataset with reduced color-coded gloves compared to other methods [15].

With the previous studies described, the researchers focused on developing a vision-based static Filipino Sign Language (FSL) recognition system using DeepLabCut. The open-source toolkit used a human pose estimation algorithm enabling users to train a deep neural network using minimal training data with human-like accuracy [13, 14]. The primary objectives of the study were (1) to develop both the Hand-Joint Tracking Model and the Filipino Sign Language Recognition Model, and (2) to evaluate the performance and suitability of the system for sign language recognition. The FSL system was aptly named WiKA, meaning "language" in the Filipino dialect, as an ode to the Filipino vernacular. The system was capable of interpreting the Filipino Sign Language alphabets (A-Z) and numbers (1-9) into text and speech function embedded in the video stream [15, 16]. Leveraging DeepLabCut for Hand-Joint Tracking not only extended its utility beyond its original purpose of tracking animal behavior but also enhanced its applicability in sign language recognition. This study significantly enhanced the potential integration of Automatic Sign Language Recognition (ASLR) systems, providing an avenue for recognizing Filipino Sign Language (FSL).

Materials and Methods

The study's methodology involved the creation of a Hand-Joint Tracking Model and a Filipino Sign Language (FSL) Recognition Model. The equipment included a Nikon D3200 camera for dataset collection and specific computer hardware for training. The Hand-Joint Tracking Model was comprised of three models using RGB and grayscale videos, ResNet 50 architecture, and varied joint key points on a dataset of 49 videos. Key steps involved the labeling of points, creating a training dataset, using DeepLabCut for network training, and evaluating the model. Trajectories from the trained network were embedded into video frames, with outlier frames manually refined. Skeletal data, including bone lengths and angles, were extracted, normalized, and stored in a CSV file. Model selection was based on Test and Train Error, with and without p-cut-off. The chosen model was employed to analyze individual videos and subsequently extract poses for post-processing. The FSL Recognition Model was developed using a CNN trained with hand-joint features and involved data pre-processing, dataset acquisition, training, and symbol classification. The entire methodology ensured precise hand-joint tracking and the creation of a robust FSL recognition model.

Identified limitations of the study, however, included its focus on a symbol-level translation system, utilization of a specific camera setup, its being restricted to right-hand calibration and 2D plane plotting and its lack of real-time functionality in the Filipino Sign language system. The procedural flowchart incorporated in this study is depicted in Figure 1.

Experimental Equipment

The camera used for the acquisition of the dataset for the Hand-Joint Tracking Model and the Filipino SLR was a 60 fps Nikon D3200 with an AF-S NIKKOR 18-55mm lens and a 30hz 720p HD Integrated Camera for test data. The same equipment was used in the demonstration of the sign language algorithm. The computer hardware specifications utilized in training the Hand-Joint Tracking Model and the Filipino sign language recognition were the following: NVIDIA RTX 3050 4gb VRAM, Tesla K80 GPU from Google Colab, Ryzen 7 4800H, and 16GB RAM.

Hand-Joint Tracking Model Development

Three (3) Hand-Joint Tracking Models were developed: Model 1 was trained using 14 RGB videos from the dataset, ResNet 50, 420 frames for the network dataset, and 5 joint key joints. Model 2 was trained using 14 Grayscale converted videos from the dataset, ResNet 50, 420 frames for the network dataset, and 21 joint key joints. As for Model 3, it was trained using 38 RGB videos from the dataset, ResNet 50, 760 for the network dataset, and 21 key joints. The process flow for the Hand-Tracking Model is shown in Figure 2.

Dataset Acquisition

The dataset utilized for the construction of the hand-joint tracking model consisted of 49 videos, which involved positioning of the right hand at the center of the frame and with a plain color as background. Of this number, 26 videos were dedicated to individual hand gestures from A-Z, while 10 videos captured gestures associated with numerical values from 1-9. Additionally, the dataset included 13 videos featuring randomly signed gestures.

Extraction of Frames

The frames of the dataset videos were extracted using an automatic extraction method through a k-means algorithm with a 1 cluster step. For each video dataset, thirty (30) frames were extracted.

Labeling of Key Points

The center of the color dots represented the exact location of the plotted points. The color of the dot for the wrist was violet, shades of indigo for

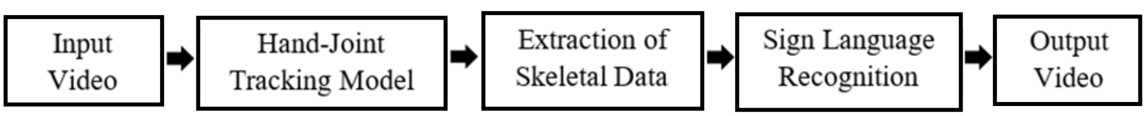


Figure 1. The procedural flowchart implemented in this study illustrating the progression starting with an input video, through a Hand-Joint Tracking Model, skeletal data extraction, Filipino Sign Language (FSL) Recognition, culminating in the output video, showcasing a systematic approach to precise hand-joint tracking and robust FSL interpretation.

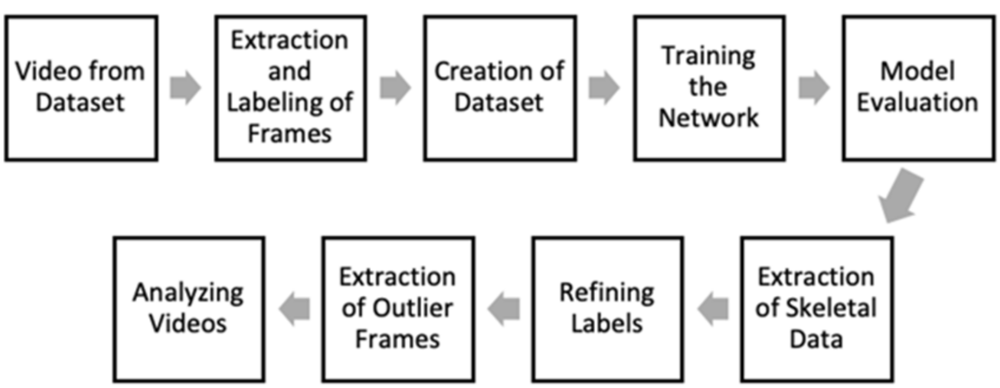


Figure 2. The comprehensive process flow for the hand-joint tracking model development, as depicted, spanned from dataset acquisition to the extraction of skeletal data. Each stage was meticulously executed, encompassing dataset preparation, frame extraction, key point labelling, training data creation, network training, model evaluation, video analysis, trajectory plotting, outlier frame correction, label refinement, and final extraction of precise skeletal data, ensuring the model's robust performance and accuracy.

the thumb, shades of blue for the index finger, shades of green for the middle finger, shades of yellow for the ring finger, and shades of red for the little finger as shown in Figure 3.

Creation of the Training Dataset

The labeled datasets derived from the videos were combined and subsequently partitioned, with 95% allocated for training data and the remaining 5% designated as test data. The training data facilitated the training process of the network, while the test data was used to evaluate the network's performance. ResNet 50 was used as the Imagenet pre-trained network weights for the network training process.

Training of the Network

Using DeepLabCut, the model underwent training for a maximum of 1,030,00 iterations. Each model was trained until the training loss and accuracy reached a plateau. Network training was executed on a Cloud GPU from Google Colab. Considering the limited runtime on Google Colab, the researchers implemented an approach of storing and saving 10 maximum snapshots throughout the entire training duration. This process involved overwriting each index at every 500th iteration, ensuring continuous training until the convergence of both training loss and accuracy was achieved.

Evaluation of the Model

The evaluation process involved an assessment of each model based on Train Error, Test Error, Train Error with p-cut-off, and Test Error with p-cut-off. Models demonstrating variations in these metrics underwent extended training iterations to refine their performance. The iteration that exhibited minimal loss and that achieved the highest accuracy was systematically saved and consequently utilized as the backbone for the Hand-Joint Tracking model.

Analyzing Videos

The trained network was used to analyze new videos by choosing a checkpoint that exhibited optimal evaluation results. The total number of frames within each video was calculated. Subsequently, the acquired labels were stored in a Multi-index Pandas Array, which contained the x,y coordinates in pixels, the time index, and the likelihood of the key point appearing throughout the entire video duration. The videos were then analyzed frame by frame, and the corresponding arrays were then stored in a Comma Separated Value (CSV) file.

Plotting Trajectories and Creating Labeled Videos

The results from the analysis of the selected videos using the trained model were used to plot the extracted poses. Through the plotted trajectories, the graphs containing the body parts vs time, likelihoods vs time, and x vs y coordinates of the joints were created. The graphs were used to assess the tracking performance of the video.

The arrays stored in the CSV file, containing x and y coordinates of labeled joint

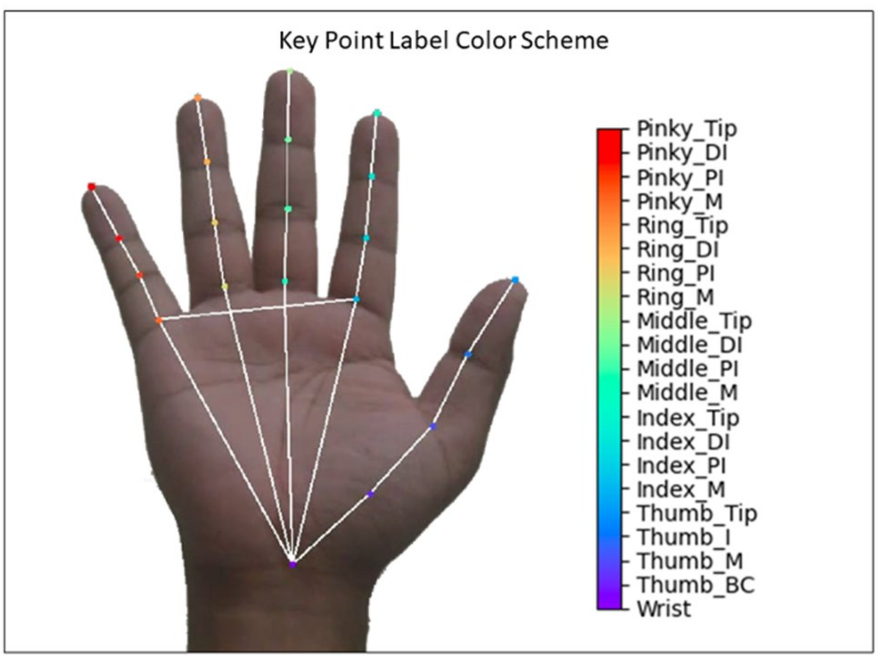


Figure 3. The representation of hand-joint locations was color-coded, where each dot indicated the exact position of the plotted points. Key points were labelled, with the wrist denoted by violet, thumb by shades of indigo, index finger by shades of blue, middle finger by shades of green, ring finger by shades of yellow, and little finger by shades of red.

points in pixels, were imported and employed to visualize the trajectories of the points within a frame. These generated colored dot points were then embedded into the video stream on a frame-by -frame basis.

Extraction of Outlier Frames

Following the training and evaluation phases of the network, frames identified as outliers were extracted and manually corrected to enhance the model's performance. This involved adjusting the plotted labels to their accurate joint locations and integrating them into the training set. The videos selected for outlier frame extraction were chosen based on the inference performance of the model from the plotted trajectories.

Refining Labels

Refinements to the labels were implemented by manually relocating them, focusing on the occluded, invalid, and inaccurately labeled hand joints to enhance the model's accuracy. Frames with these refined labels were then added to the augmented training dataset, and the model underwent retraining using the same configuration. This iterative process, beginning with the extraction of outlier frames and ending with label refinement, continued until the test loss and test accuracy reached a plateau.

Extraction of Skeletal Data

Utilizing the x, y coordinates in pixels, the precise length and orientation of each bone, as specified in the configuration file of the Hand-Joint Tracking Model, were determined. This involved employing the distance formula within a two-coordinate system, considering the pixel location of a point in a given frame. The angle necessary for the skeletal connection to return to the horizontal plane in a counterclockwise manner was then calculated. The calculation of this two-point skeletal connection relied on the application of the Tangent Trigonometric Ratio.

The skeletal data were normalized by converting the images into local coordinate systems. The computed lengths and angles, along with the probability of paired points appearing in the video stream, were stored in a CSV file.

Hand-Joint Tracking Model Selection

From the three (3) generated Hand-Joint Tracking models, the model best suited for Sign Language Recognition was selected through its Test and Train Error without p-cut-off and its Test and Train Error with p-cut-off. Each of the models' performances was also tested using a test video that contained all of the 35 classes to be used in the Filipino Sign Language Recognition system.

After selecting the best Hand-Joint Tracking model, individual videos from the dataset starting from 1-9 to A-Z were then analyzed. The poses drawn from the Hand-Joint Tracking Model were extracted for post-processing.

Filipino Sign Language Recognition Model Development

The developed FSL recognition model employed a Convolutional Neural Network (CNN) trained using the extracted hand-joint features to predict both alphabets and numbers. This involved the development of the graphical user interface (GUI), data preprocessing, dataset acquisition, training of the network, evaluation of the network and symbol classification.

Raw extracted skeletal data in CSV file from DeepLabCut underwent data cleaning using Pandas. The raw length and angle orientation of the connected points in the hand underwent value normalization for the CNN classifier network to perform well. After data cleaning and value normalization, each of the classes was then labeled from 0-35, as shown in Table 1.

Results and Discussion

Extracted Raw Skeletal Data

After extraction of the raw skeletal data from a video through DeepLabCut, a CSV file was generated. The file documented the length between

Table 1. Summary of class label assignment for numbers 1-9 and letters A-Z following data cleaning and value normalization in the development process of the Filipino Sign Language (FSL) recognition model. The model utilized a Convolutional Neural Network (CNN) trained with hand-joint features to predict both alphabets and numbers, involving steps such as GUI development, data pre-processing, dataset acquisition, network training, evaluation, and symbol classification.

Label	Class	Label	Class	Label	Class
0	1	12	D	24	P
1	2	13	E	25	Q
2	3	14	F	26	R
3	4	15	G	27	S
4	5	16	H	28	T
5	6	17	I	29	U
6	7	18	J	30	V
7	8	19	K	31	W
8	9	20	L	32	X
9	A	21	M	33	Y
10	B	22	N	34	Z
11	C	23	O		

the points, the angle formed between them, and the likelihood of the two connected points appearing throughout the entire duration of the video. This is shown in Figure 4. The total number of rows in the CSV file represented the overall number of frames in a video, while the total number of columns was for the number of points multiplied by three (length, orientation, and likelihood).

Hand-Joint Tracking Model's Performance

There were 3 models presented for searching for an optimized model. Model 1, as shown in Figure 5A reached the plateau at 37,000 training iterations. Model 2, shown in Figure 5B reached the plateau at the 113,000 training iterations mark, and Model 3 reached the plateau at the 134,500 training iterations mark.

When the models were given a confidence level of 60% (p-cut-off =0.6), Model 3, shown in Figure 5C, had 4.92 and 5.74 percent errors which was the lowest among the three models. This indicated that the model with the highest performance was the one associated with the largest video dataset. Furthermore, it could be inferred that the model with more frames and plotted points demonstrated higher accuracy and a greater number of training iterations. Hence, after considering the balance between performance, training time, and accuracy, Model 3 was selected as the Hand-Joint Tracking Model for further analysis on an unseen dataset and as the skeletal data extraction model for the FSL dataset for sign language recognition.

Filipino Sign Language Recognition Experimental Results

This presents the Model Accuracy, Model loss, Model prediction performance and Model inference time.

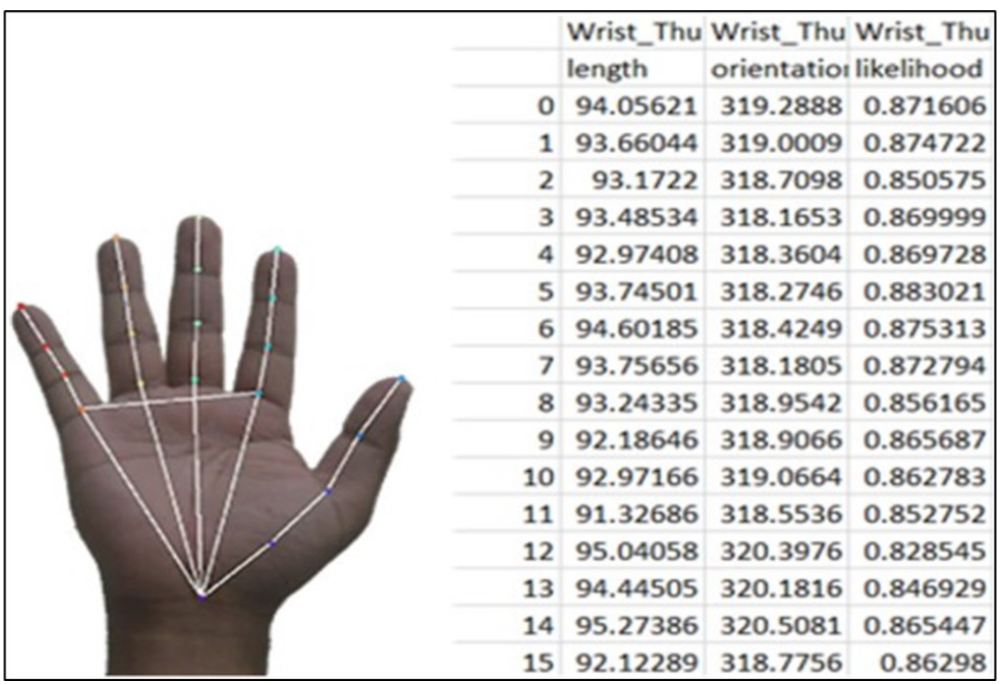
Model Accuracy and Model Loss

The model gained a 73.44% training accuracy and 96.44 validation accuracy as shown in Figure 6A. On the other hand, the model gained a 0.8422 training loss and a 0.2448 validation loss as shown in Figure 6B.

Filipino Sign Language Model Performance

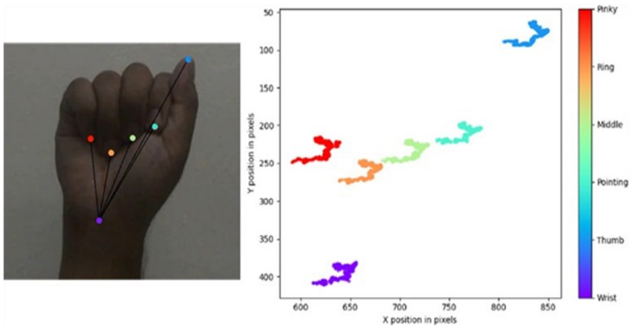
Overall, the model garnered an average prediction time of 0.000104s. This was because the input data that was being fed into the network was an array with 1 row and 42 columns. In terms of the Macro average value and Weighted Average value, the model exhibited a 97% precision, 96% Recall and 96% F1-score. The summary of the whole performance of the Filipino Sign Language recognition is shown in Table 2.

From the given results in the Hand-Joint Tracking Model, the model that exhibited the highest tracking performance was Model 3 with 35 input videos, 21 key points, and 760 dataset frames. On the other hand, the FSLR exhibited a 96.44% accuracy, 96% macro and weighted average precision, 96% macro and weighted average recall, and an 86% macro and weighted average F1-score. From the data on the WiKA software embedded with the Hand-Joint Tracking

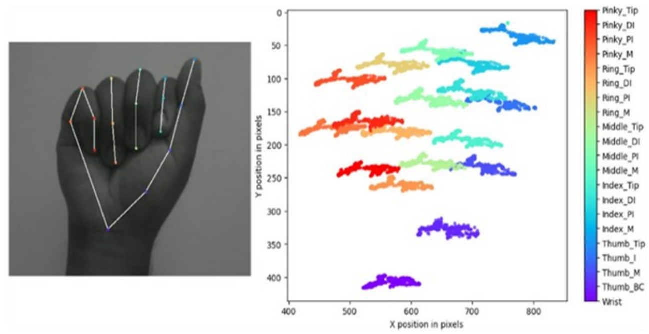


	Wrist_Thu length	Wrist_Thu orientatio	Wrist_Thu likelihood
0	94.05621	319.2888	0.871606
1	93.66044	319.0009	0.874722
2	93.1722	318.7098	0.850575
3	93.48534	318.1653	0.869999
4	92.97408	318.3604	0.869728
5	93.74501	318.2746	0.883021
6	94.60185	318.4249	0.875313
7	93.75656	318.1805	0.872794
8	93.24335	318.9542	0.856165
9	92.18646	318.9066	0.865687
10	92.97166	319.0664	0.862783
11	91.32686	318.5536	0.852752
12	95.04058	320.3976	0.828545
13	94.44505	320.1816	0.846929
14	95.27386	320.5081	0.865447
15	92.12289	318.7756	0.86298

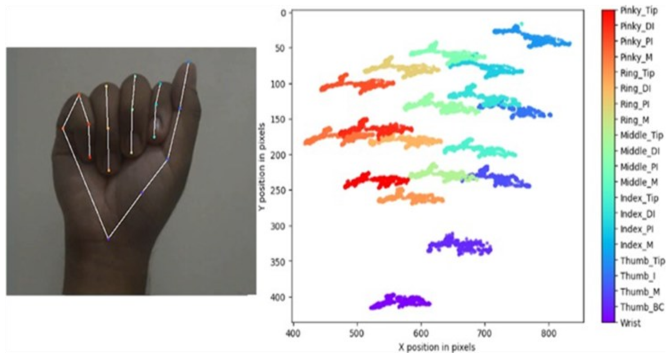
Figure 4. The extracted raw skeletal hand-joint features were visualized, displaying the total number of frames in a video (in row) and the length, orientation, and likelihood parameters (in columns).



A. Model 1 Inference to Letter A

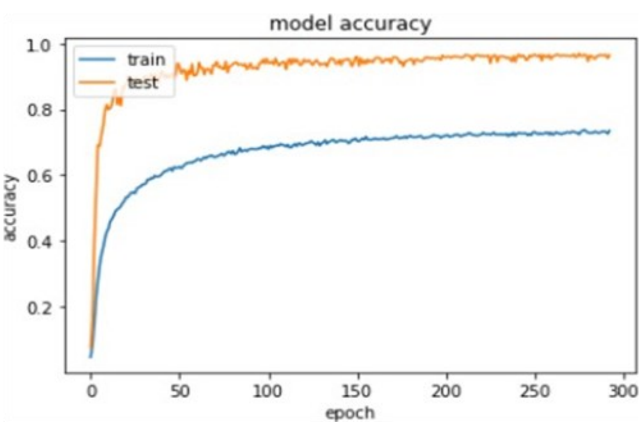


B. Model 2 Inference to Letter A

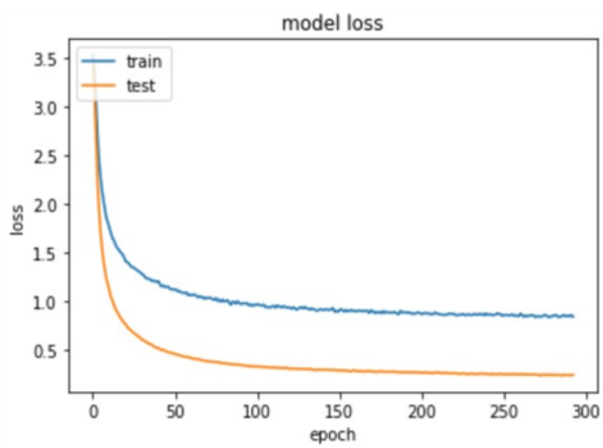


C. Model 3 Inference to Letter A

Figure 5. Three models were presented to search for an optimized model. (A) Model 1 reached a plateau at 37,000 training iterations, (B) Model 2 at 113,000 iterations, and (C) Model 3 at 134,500 iterations. When the models were assessed with a confidence level of 60% (p-cut-off = 0.6), Model 3 exhibited the lowest error rates of 4.92% and 5.74% among the three models, indicating superior performance. This suggested that the model associated with the largest video dataset performed the best, indicating the importance of dataset size. Considering the relationship between frames, plotted points, accuracy, and training iterations, Model 3 was chosen as the Hand-Joint Tracking Model for further analysis and as the skeletal data extraction model for FSL sign language recognition.



A. Model Accuracy



B. Model Loss

Figure 6. The model's training and validation accuracy, as well as training and validation loss, were depicted in the model results. Specifically, the model achieved a (A) training accuracy of 73.44% and a validation accuracy of 96.44%. Conversely, the model incurred a (B) training loss of 0.8422 and a validation loss of 0.2448.

Table 2. The table presented the performance evaluation of the Filipino-Sign Language Recognition Model using the Split Validation Dataset, assessing precision, recall, and F1-score metrics. This approach segregated a portion of the original dataset for exclusive model evaluation during training, mitigating overfitting and ensuring generalization capability. The model achieved an average prediction time of 0.000104s, processing input data as a single-row array with 42 columns. With a 97% precision, 96% recall, and 96% F1-score across both Macro average and Weighted Average values, the model demonstrated robust performance.

Letter or Number	Label	Precision	Recall	F1-score	Support
1	0	0.94	0.94	0.94	110
2	1	0.97	1	0.98	96
3	2	1	1	1	108
4	3	1	1	1	101
5	4	1	1	1	107
6	5	0.94	0.97	0.97	108
7	6	1	1	1	112
8	7	1	1	1	109
9	8	1	0.98	0.99	101
A	9	0.99	1	0.99	94
B	10	0.98	1	0.99	114
C	11	1	1	1	118
D	12	1	1	1	152
E	13	0.99	1	1	109
F	14	0.99	1	1	130
G	15	0.95	0.97	0.97	122
H	16	0.97	0.99	0.99	107
I	17	0.94	0.97	0.97	105
J	18	1	0.72	0.84	101
K	19	1	1	1	115
L	20	1	1	1	88
M	21	1	0.83	0.91	112
N	22	0.86	1	0.93	118
O	23	1	1	1	97
P	24	0.93	1	0.96	106
Q	25	0.92	0.99	0.96	111
R	26	0.68	0.82	0.74	89
S	27	0.99	1	1	124
T	28	1	1	1	101
U	29	0.85	1	0.92	106
V	30	0.99	0.99	0.99	114
W	31	1	0.93	0.97	104
X	32	1	1	1	101
Y	33	0.99	1	1	104
Z	34	0.9	0.47	0.62	100
	Accuracy			.96	3794
	Macro avg	0.97	.96	.96	3794
	Weighted avg	.97	.96	.96	3794
	Average prediction time: 0.000104s				

Model and the FSLR, the software was able to achieve 95.77% accuracy on singular videos of the alphabets A-Z and numbers 1-9. This indicated the effectiveness of WiKA in recognizing alphabets and numbers, providing a useful resource for the non-signing population to understand sign languages.

Conclusion

The researchers developed WiKA, a computer vision software designed to recognize Filipino Sign Language based on the positioning of hand joints tracked by a markerless pose estimation Hand-Joint Tracking model created through DeepLabCut. The study demonstrated the efficacy of DeepLabCut in hand joint tracking, and consequently revealed its potential not only in Sign Language recognition but also in broader applications within the field of Recognition.

The findings of this study emphasized the software's capability to specifically recognize the Alphabet from A-Z and Numbers from 1-9. The alphabets from A-Z and numbers from 1-9 were recognized with valid quantifiable results in terms of precision, recall and F1-Score. This functionality showed promise in facilitating communication between the signing and non-signing populations, providing a means for the latter to comprehend these characters without the need for Filipino Sign Language proficiency.

The development of WiKA represented a series of impactful contributions to Sign Language Recognition (SLR) and computer vision. Key achievements included an innovative Hand-Joint Tracking Model with DeepLabCut, enhancing precision in hand gesture tracking and setting an example for future applications. The FSL Recognition system, utilizing Extracted Hand Joint features, demonstrated practical applications in sign language interpretation, promising improved communication for the FSL community. Integration with Google Colab showcased advances in scalability, and a user-friendly Graphical User Interface (WiKA) captured these models, thereby solidifying the study's contributions. Overall, these developments not only enriched academic discourse but also held tangible potentials for real-world applications, fostering inclusivity and communication for the FSL community and beyond.

Building on the valuable insights gained from this study, a set of recommendations has emerged to guide future research in Filipino Sign Language (FSL) recognition systems. This includes widening the variability of video datasets and increasing training iterations for the Hand-Joint Tracking Model to improve model accuracy. Optimizing video resolution is also suggested for efficient training without compromising effectiveness, and the inclusion of diverse lighting conditions and hand shapes should be investigated. Future research is encouraged to build on the current study, transitioning to 3D Hand-Joint Tracking Models for heightened accuracy and introducing skeleton point connections to enhance the distinctiveness of skeletal features. It is further recommended that an application of this study in 3D contexts be considered, thereby promoting a more immersive and comprehensive approach to FSL recognition. Collectively, these recommendations aim to drive FSL recognition towards greater accuracy, applicability, and inclusivity, addressing the unique needs of the signing community.

Authors' Contributions

The project featured collaborative authorship, with RAM and KDS, overseeing system assembly and set-up, model development and testing, along with the drafting of the manuscript, and TMO providing the revised system set-up and testing, enhanced data visualization, enhancements in technical writing, and publication communication, under the consultation with NSG.

Acknowledgments

The authors acknowledge the Electrical and Electronics Engineering Department of the College of Engineering in Mindanao State University-General Santos City.

References

[1] Hurlbut, H.M. (2008). *Philippine Signed Languages Survey: A Rapid Appraisal*. Dallas, Texas: SIL International.

[2] Recto, R.G., Aquino IV, P.B., Legarda, L.B., Angara, J.E.M., and Binay, M.L.N.S. (2015). An Act Expanding the Benefits and Privileges of Persons with Disability (PWD). *16th Congress of the Republic of the Philippines*, Senate Bill 2890, 1 -5. Retrieved from https://legacy.senate.gov.ph/lis/bill_res.aspx?congress=16&q=SBN-2890

[3] Antonio, K.B. (2023). How the PH deaf community strives for inclusivity through Filipino Sign Language. *Rappler*, 1–6. Retrieved from https://www.rappler.com/moveph/filipino-deaf-community-strives-inclusivity-filipino-sign-language/

[4] Maiorana-Basas, M. and Pagliaro, C.M. (2014). Technology Use Among Adults Who Are Deaf and Hard of Hearing: A National Survey. *Journal of Deaf Studies and Deaf Education*, 19, 400–410. Doi: 10.1093/deafed/enu005.

[5] Gomase, K., Dhanawade A., Gurav, P. and Lokare S. (2022). Sign Language Recognition Using MediaPipe. *International Research Journal of Engineering and Technology*, 744-746. Doi: 10.1117/12.2674613.

[6] Wadhawan, A. and Kumar, P. (2021). Sign Language Recognition Systems: A Decade Systematic Literature Review. *Archives of Computational Methods in Engineering*, 28, 785–813. Doi: 10.1007/s11831-019-09384-2.

[7] Madhiarasan, D. M. and Roy, P.P. (2022). A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets. *ArXiv*, XX, 1–30. Doi: 10.48550/arXiv.2204.03328.

[8] Ahmed, M.A, Zaidan, B.B., Zaidan, A.A., Salih, M.M, and Bin Lakulu, M.M. (2018). A Review on Systems-Based Sensory Gloves for Sign Language Recognition State of the Art between 2007 and 2017. *Sensors*, 18, 2208. Doi: 10.3390/s18072208.

[9] Bantupalli, K. and Xie, Y. (2018). American Sign Language Recognition using Deep Learning and Computer Vision. *2018 IEEE International Conference on Big Data (Big Data)*, 4896–4899. Doi: 10.1109/BigData.2018.8622141.

[10] Culver, V. R. (2004). A hybrid sign language recognition system. *Eighth International Symposium on Wearable Computers*, 1, 30–33. Doi: 10.1109/ISWC.2004.2

[11] Sha, J., Ma, J., Mou, H., and Hou, J. (2020). A Review of Vision Based Dynamic Hand Gestures Recognition. *Computer Science and Application*, 10, 990–1001. Doi: 10.12677/csa.2020.105102.

[12] Jarabese, M.B.D., Marzan, C.S., Boado, J.Q., Lopez, R.R.M.F., Ofiana, L.G.B., and Pilarca, K.J.P. (2021). Sign to Speech Convolutional Neural Network-Based Filipino Sign Language Hand Gesture Recognition System. *2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC)*, 147–153. Doi: 10.1109/ISCSIC54682.2021.00036.

[13] Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., and Mathis, M.W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols,* 14, 2152–2176. Doi: 10.1038/s41596-019-0176-0.

[14] Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W. and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21, 1281–1289. Doi: 10.1038/ s41593-018-0209-y.

[15] River, J. and Ong, C. (2018). Facial Expression Recognition in Filipino Sign Language Classification using 3D Animation Units. *Proceedings of the 18th Philippine Computing Science Congress (PCSC 2018)*, 1, 44-51. Retrieved from https://www.researchgate.net/ publication/325711425_Facial_Expression_Recognition_in_Filipino_Sign_Language_Classification_using_3D_Animation_Units

[16] Naidoo, S., Omlin, C. and Glaser, M. (2002). Vision-Based Static Hand Gesture Recognition using Support Vector Machines. Retrieved from https://www.academia.edu/48575893/Vision_Based_Static_Hand_Gesture_Recognition_using_Support_Vector_Machines

[17] Ashiquzzaman, A., Lee, H., Kim, K., Kim, H. Y., Park, J. and Kim, J. (2020). Compact spatial pyramid pooling deep convolutional neural network-based hand gestures decoder. *Applied Sciences,* 10, 1–22. Doi: 10.3390/app10217898.

[18] Martinez, L. and Cabalfin, E.P. (2008). Sign language and Computing in a Developing Country: A Research Roadmap for the Next Two Decades in the Philippines. *Pacific Asia Conference on Language, Information and Computation.* Retrieved from https://www.semanticscholar.org/ paper/Sign-language-and-Computing-in-a-Developing-A-for-Martinez-Cabalfin/14c8f6104522f35806ba78add6 e3d4d9491fe145