Real-Time Web-Based Framework for Sign Language Recognition and Baybayin Translation Utilizing YOLOv8 Model

Eirand Jan C. Barcelo
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
eirandjan.barcelo@tup.edu.ph

Jashameel Faith D. Basa
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
jashameelfaith.basa@tup.edu.ph

Rafael B. Dumaguina
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
rafael.dumaguina@tup.edu.ph

Ronald B. Laz
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
ronald.laz@tup.edu.ph

Abegail A. Lopez
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
abegail.lopez@tup.edu.ph

Camela Trisha J. Romen
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
camelatrisha.romen@tup.edu.ph

Jessica S. Velasco
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
jessica_velasco@tup.edu.ph

Lean Karlo S. Tolentino
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
leankarlo_tolentino@tup.edu.ph

Edgar A. Galido
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
edgar_galido@tup.edu.ph

Gilfred Allen M. Madrigal
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
gilfredallen_madrigal@tup.edu.ph

Romeo Jr. L. Jorda
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
romeo_jorda@tup.edu.ph

Gina B. Araojo
Department of Electronics Engineering, College of Engineering, Technological University of the Philippines
Manila, Philippines
gina_araojo@tup.edu.ph

***Abstract*—This study proposes a website enabling real-time sign language recognition and translation into Baybayin script. The researchers trained a CNN model YOLOv8, using Roboflow and Google Colab and built the website's frontend using React and TensorFlow. Hearing impaired people from the Bigay Buhay Multipurpose Cooperative participated to assess the sign language recognition accuracy. Programming experts evaluated the website's functionality. Data were gathered using checklists as a research instrument, and a modified version of the ISO quality model 25010. The study guide improved the user's sign language proficiency (score increase: 1.4666). After evaluation, YOLOv8 was shown to be a successful model for sign language detection (accuracy: 93.1%). Positive results were obtained from user assessments, with mean scores of 4.493 (expert) and 3.669 (end-user). The adjectival rating for the application was extremely good overall (4.081).**

Keywords—Sign Language, CNN, YOLOv8, Hearing Impaired

I. Introduction

The use of Filipino Sign Language (FSL) has contributed to the improvement of communication of deaf people, however, most of the population in the Philippines do not understand FSL [1]. Sign language serves as a primary mode of communication for millions of individuals worldwide, particularly for the Deaf and hard of hearing community. However, the accessibility and interpretation of sign language remain challenges, especially in web-based environments.

As the 21st century progresses, the amount of information on the internet expands exponentially, leading to advancements in information technology and data science. Artificial intelligence (AI) has emerged as a pivotal field, offering innovative solutions to societal challenges.

Incorporating AI into sign language translation provides a novel approach to addressing these challenges. Computer vision, a subset of AI, enables computers and cameras to detect and understand sign language in various contexts. This study utilizes Convolutional Neural Networks (CNNs) and a specific learning model called YOLOv8 for this purpose.

This paper addresses these challenges by proposing a novel approach for real-time web-based sign language recognition and translation, with a specific focus on Filipino Sign Language (FSL) and its connection to the traditional Baybayin script. Leveraging the YOLOv8 model, a state-of-the-art object detection system, the researchers aim to develop an accurate and efficient system capable of recognizing and interpreting sign language gestures in real-time. Furthermore, the system incorporates a feature for immediate translation of recognized gestures into the Baybayin script, thus promoting its usage and preserving cultural heritage.

II. Related Works

Learning sign language for everyday conversations presents unique challenges, as noted by Berke [2] and Notarte-Balanquit [1]. Berke highlights the difficulty of keeping up in group conversations and the challenges of interpreting which parts to sign. Notarte-Balanquit points out that borrowing signs from other languages and conversational context can influence sign language variations. Both studies underscore the challenges in learning sign language but emphasize that sign language translators can bridge communication gaps between deaf and hearing individuals.

YOLOv5 is extensively used for various computer vision tasks due to its speed and accuracy. This study presents a real-time sign language detection system using YOLOv5 and convolutional neural networks (CNN). The system captures video input via a webcam, preprocesses it with OpenCV, and then processes it using YOLOv5 and CNN to recognize hand gestures. The model achieved an average of 0.987 mAP@0.5 and 0.985 F1 scores for six hand gesture categories, demonstrating its suitability for real-time detection with a small dataset [7].

Similarly, YOLOv4, developed in 2020, enhances previous YOLO versions with improvements in speed and accuracy. The journal article describes a system for real-time American Sign Language (ASL) recognition using YOLOv4. Trained on a dataset of 8000 images across 40 classes, the model achieved a mean average precision (mAP) of 98.01% and processed video data at 28.9 frames per second, indicating its potential for accurate real-time ASL recognition [8].

The Real-Time Web-Based Framework aims to address communication barriers for the Filipino deaf community by using the YOLOv8 model for real-time Filipino Sign Language (FSL) recognition. Building on previous research [9][10], this framework enhances accessibility and inclusivity for the deaf community in the Philippines by leveraging advanced technology.

YOLOv8 improves on previous versions in both accuracy and speed. The article details the process of training YOLOv8 on a custom dataset of sign language gestures and integrating it into a Python-based application for live detection. This application can translate sign language gestures into text or spoken language in real-time, aiding communication for individuals with hearing impairments [11].

Baybayin, an ancient Philippine script, has seen renewed interest following the 2018 approval of House Bill 1022, designating it as the national writing system. Research by Maningas et al. [12] explores the historical context and decline of Baybayin during the early Spanish colonial period, providing insights into its sociopolitical significance. Additionally, Pino et al. [13] have developed an algorithm using a Support Vector Machine (SVM) classifier to transliterate Baybayin words into the Latin script from images, achieving a recognition rate of 97.9% on a new dataset.

III. METHODOLOGY

A. Research Design

This study employs a descriptive quantitative approach to evaluate the effectiveness of a real-time web-based framework for sign language recognition and Baybayin translation using the YOLOv8 model. Data will be collected through questionnaires distributed to respondents. The questionnaire, designed based on ISO 25010 standards, will assess the model's acceptability and accuracy. Respondents, selected through purposive sampling, will include deaf-mute individuals from the Bigay Buhay Multipurpose Cooperative in Novaliches, Liliw.

B. Data Collection

For the data collection process, the researchers used 1080p webcam to take pictures for the dataset. The collected images consisted of 50 common words sourced from the Deafed Sign Workbook and the entire alphabet. In total, the dataset comprised over 10,000 images, which were subsequently uploaded to Roboflow. At Roboflow, the datasets were annotated and used for training the YOLOv8 model for sign language recognition. Additionally, the Baybayin translator component was acquired from an open-source API, enhancing the versatility and functionality of the system.

C. Machine Training

The researchers used Roboflow, a platform specializing in object detection and machine learning. They utilized its annotation features for labeling datasets, its preprocessing and augmentation tools to expand datasets, and its cloud training capabilities for model training. The researchers then conducted the machine training of YOLOv8 on Google Colab.

D. Data Processing

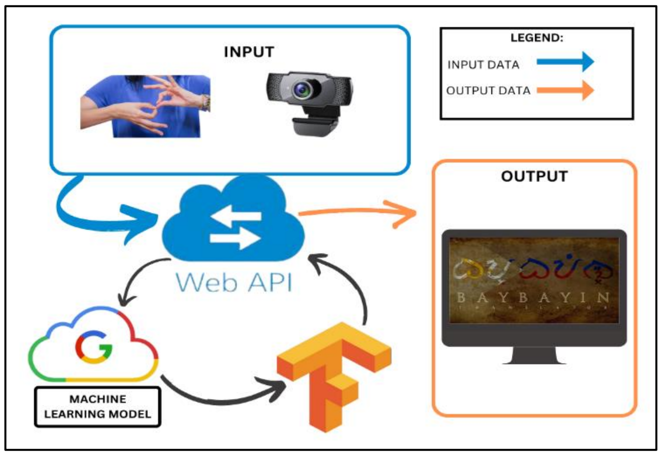

Fig. 1. Conceptual Framework of the Proposed Project

Fig. 1 illustrates the Data Processing within the system. First, the user interacts with the website application Tinig through the use of a webcam or any camera that acts as an input device. The data will then be sent to the front-end components which includes TensorFlow.js, React, and a web browser. These parts will translate a request to the backend Application Programming Interface (API). The machine learning model stored on the Cloud platform will be contacted by the requests received from the front-end components. The Filipino Sign Language (FSL) data will be processed and interpreted by TensorFlow, which will then use the Baybayin Translator to convert it into text. The information that was processed will be forwarded to the front-end component of the web application and returned back to the backend API.

E. Development

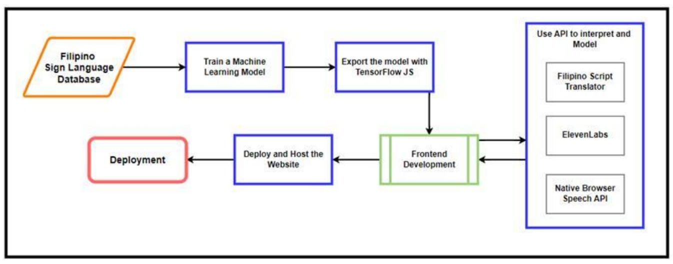

Fig. 2. Development Architecture of the Proposed Project

Fig. 2 illustrates the project development workflow. TensorFlow will be used to integrate the trained model into the website. Front-end development focuses on building the UI and incorporating TensorFlow.js features into the React framework. Various APIs, such as Baybayin Translator, ElevenLabs, and Web Speech, will bridge traditional Filipino scripts with modern communication technologies. The data object model will be hosted on Google Cloud to reduce hardware dependency. All applications will undergo thorough testing and debugging to ensure functionality before being launched on appropriate hosting platforms.

F. Accuracy Testing

This study introduced the use of the Bayesian Technique to improve accuracy by incorporating data from prior accuracy tests, providing deeper insights [14]. The Bayesian Theorem specifically determined the true accuracy of YOLOv8. Testing YOLOv8 for sign language faced challenges such as hand shape imitation difficulties, overlapping sign languages, and internet speed. The researchers employed the Bayesian Technique to address these challenges.

Formally, Bayes' theorem states:

$$P(A|B) = \frac{[P(B|A) \times P(A)]}{P(B)}$$

G. Evaluation

After collecting the requisite data, it underwent statistical treatment, specifically utilizing the arithmetic mean. The data helped ascertain the acceptability level of the website through the computed mean. The following formula was employed to calculate the data:

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Wherein:

$\overline{x}$= average (arithmetic mean)

N = number of terms in the list (the numbers being averaged)

x_i = the value of each individual item in the list of numbers being averaged

The overall statistical mean was matched with equivalent adjectival ratings to determine the overall acceptability of the website. The data was then analyzed and interpreted using a table of verbal interpretation. The table provided below illustrates this verbal interpretation:

TABLE I. TABLE INTERPRETATION

Mean Average	Verbal Interpretation
4.21 – 5.00	Excellent
3.31 – 4.20	Very Good
2.61 – 3.30	Satisfactory
1.81 – 2.60	Fair
1.00 – 1.80	Poor

IV. RESULTS AND DISCUSSION

A. Creation of Website

Figure 3 below shows the FSL-to-text page design of the website. The user's camera captures the performed sign language, which is then processed by a custom algorithm and integrated APIs such as Roboflow and Eleven Labs. This setup translates the sign language into both traditional Baybayin text and synthesized voice, creating a comprehensive communication tool. This integration bridges the gap between historical scripts and modern technology, enhancing accessibility and learning for all users.

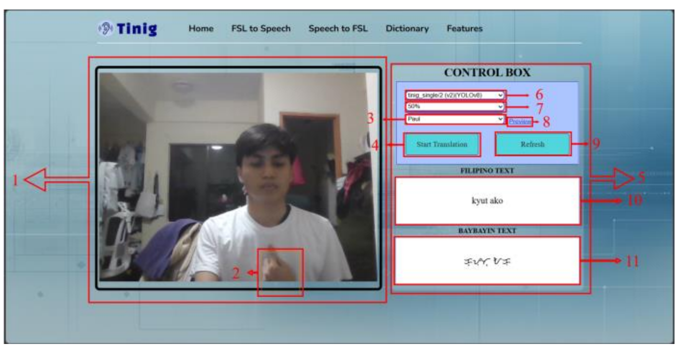

Fig. 3. Website Navigation

*1) **Canvas Displaying Captured Image:*** This area shows the live video feed from the camera where the user's sign language gestures are captured. The displayed image is crucial for the user to see and adjust their gestures accordingly to ensure accurate recognition by the system.

*2) **Detected Sign Language:*** A highlighted area within the video feed that shows the specific gesture recognized by the system. This serves to inform the user that their gesture has been detected and is being processed.

*3) **Dropdown Menu for Selecting AI Voice Model:*** This menu allows users to choose different artificial intelligence (AI) voice models for the speech synthesis part of the application, which converts text into spoken audio.

*4) **Start Translation Button:*** This button initiates the translation process of the detected sign language into text and subsequently uses the selected AI voice model to read aloud the converted text.

*5) **Control Panel Container:*** This section contains all the interactive elements and controls necessary for the operation of the system, including model selection, translation, and audio preview functionalities.

*6) **Dropdown Menu for AI Model Selection for Sign Language Detection:*** Users can select different AI models specialized in sign language detection, allowing for flexibility depending on the complexity of signs or the user's specific needs.

*7) **Dropdown Menu for Selecting Threshold for Accuracy Level:*** This menu enables users to set a threshold for the accuracy level of the sign language detection AI model, which can help in tuning the sensitivity of the gesture recognition.

*8) **Preview Button for AI Voice Model:*** Before starting the translation, users can use this button to hear a sample of the selected AI voice model, helping them decide if the tone and pitch are suitable for their needs.

*9) **Refresh Button:*** This clears out previous translations from the text display areas, allowing for a new session of translations without confusion from earlier content.

*10) **Textbox for Filipino Text:*** After translation, this textbox displays the converted Filipino text corresponding to the recognized sign language gesture.

*11) **Textbox for Baybayin Text:*** Similarly, this textbox displays the translated Baybayin text, an ancient script of the Philippines, corresponding to the recognized sign language, facilitating cultural preservation and education.

B. Training Metrics

The training metrics were provided after training the YOLOv8 model. With this the researchers were able to evaluate the performance of models in object detection.

a. Mean Average Precision (mAP)

mAP is a metric that averages precision across all classes, indicating overall detection accuracy.

TABLE II. SUMMARY OF MAP

Model	mAP
YOLOv8 - Single	99.3 %
YOLOv8 - Alphabet	99.5 %

Table II summarizes the mean average precision (mAP) of various object detection models. Each model achieved a high mAP, all above 99%.

b. Confusion Matrix

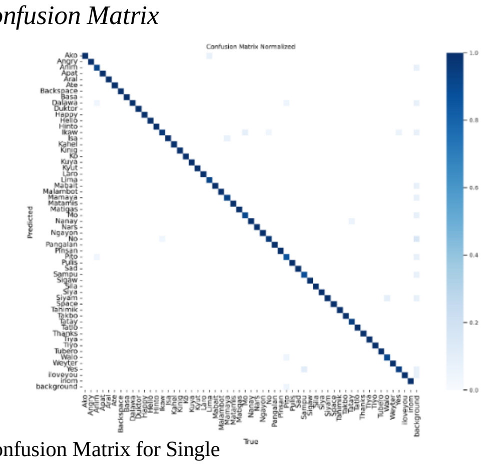

Fig. 4. Confusion Matrix for Single

Fig. 4 illustrates the confusion matrix for the Yolov8 Single model, representing the model's accuracy in identifying a single class of objects with slightly misclassified words.

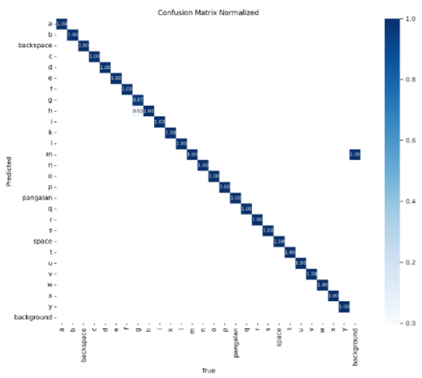

Fig. 5. Confusion Matrix for Alphabet

Fig. 5 presents the confusion matrix for the Yolov8 Alphabet model, illustrating the performance of the model in correctly classifying different alphabet characters.

c. F1 Score

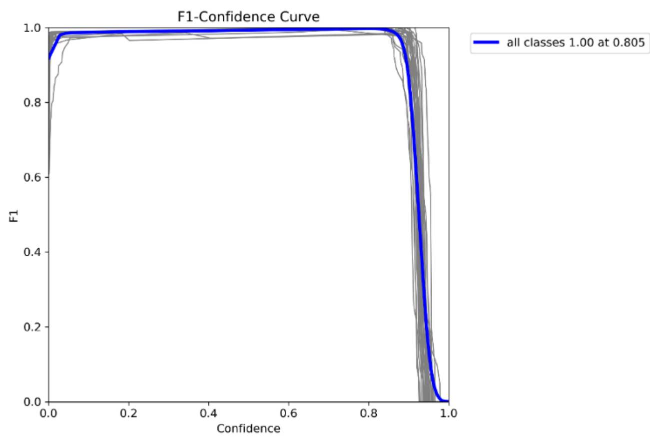

Fig. 6. F1 Curve for Single

Fig. 6 depicts an overall F1 score of 0.98 at a confidence level of 0.811, suggesting that the model achieves very high precision and recall at this confidence threshold.

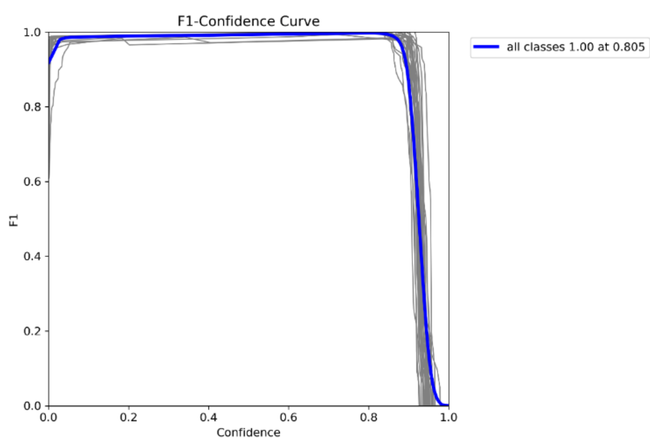

Fig. 7. F1 Curve for Alphabet

Fig. 7 presents an overall F1 score of 1.00 at a confidence level of 0.805, signifying that the model attains perfect precision and recall at this confidence threshold.

C. True Accuracy

The researchers applied the Bayesian Theorem to evaluate the speed and accuracy of the trained CNN model.

TABLE III. SUMMARY OF TRUE ACCURACY

Model	Mean Accuracy (30 respondents)	Mean Accuracy (60 respondents)	Overall Accuracy
YOLOv8	95.4%	90.8%	93.1%

Table III summarizes the accuracy results for a model called YOLOv8. It shows the mean accuracy for 30 respondents (95.4%), mean accuracy for 60 respondents (95.8%), and an overall accuracy of 93.1%.

D. Evaluation

The researchers developed a survey based on ISO 25010 standards to assess software quality. It evaluates application accessibility and user experience across reliability, portability, compatibility, maintainability, functionality, efficiency, and usability criteria.

TABLE IV. SUMMARY OF PERFORMANCE

Criteria	Mean	Adjectival Rating
A. End-User	3.669	Very Good
B. Technical Expert	4.493	Excellent
Composite Mean	4.081	Very Good

Table IV shows the summary of performance of the website. According to ISO 25010 standards, the website performed "Very Good" in functionality, efficiency, and usability based on user feedback.

V. CONCLUSION

This study has successfully developed a web-based framework that recognizes Filipino Sign Language and Baybayin translation in real time using the YOLOv8 model. It achieved 93.1% accuracy in sign language detection. User evaluations, based on ISO 25010 standards, rated the system as "Very Good". This research underscores AI's potential in assistive technologies and its role in preserving Philippine cultural heritage.

REFERENCES

[1] M. C. Ang, K. R. C. Taguibao, and C. O. Manlises, "Real-Time Web-Based Sign Language Recognition and Baybayin Translation using YOLOv8 Model," in Proc. 2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 2024, pp. 1-6, doi: 10.1109/IICAIET.2024. IICAIET55139.2022.9936801.

[2] P. T. Liza Martinez, “Borderlessness and Filipino sign language,” INQUIRER.net, https://opinion.inquirer.net/127500/borderlessness-and-filipino-sign-language (accessed Mar. 10, 2024).

[3] R. Palalimpa, “Up Student Councils push for sign language GE class,” Tinig ng Plaridel, https://www.tinigngplaridel.net/fsl/ (accessed Mar. 10, 2024).

[4] E. Lopez, “DECOLONISING MODERN LANGUAGES AND CULTURES,” Blogs, https://blogs.ncl.ac.uk/decolonisesml/tag/baybayin/ (accessed Mar. 10, 2024).

[5] J. Berke, “How hard is it to learn sign language?,” Verywell Health, https://www.verywellhealth.com/challenges-of-learning-sign-language-1049296.

[6] “Liberty Notarte-Balanquit presents Filipino sign language research,” Department of Linguistics - UP Diliman, https://linguistics.upd.edu.ph/news/liberty-notarte-balanquit-presents-filipino-sign-language-research/.

[7] D. Chiranjeevulu, J. Tejasri, I. Sasi Kala, K. Hemanth and N. Rakesh, "Real Time Sign Language Detection Using Yolov5 Algorithm," International Journal of Research Publication and Reviews, vol. 4, no. 4, pp. 2495-2499, April 2023. Available: https://ijrpr.com/uploads/V4ISSUE4/IJRPR11708.pdf

[8] Al-Shaheen, A. M., Cevik, M., & Alqaraghuli, A. (2022). American Sign Language Recognition using YOLOv4 Method. International Journal of Multidisciplinary Studies and Innovative Technologies, 6(1), 61-65. Available: https://dergipark.org.tr/en/download/article-file/2484463

[9] N. B. Linsangan et al., “Filipino sign language to text converter using K-nearest neighbor algorithm,” 2022 IEEE 14th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Dec. 2022. doi:10.1109/hnicem57413.2022.10109512

[10] M. D. Montefalcon, J. R. Padilla, and R. Llabanes Rodriguez, “Filipino sign language recognition using Deep Learning,” 2021 5th International Conference on E-Society, E-Education and E-Technology, Aug. 2021. doi:10.1145/3485768.3485783

[11] S. J. Redmon and A. Farhadi, "YOLOv8: Real-time sign language detection using object detection," FreeCodeCamp, 2024. [Online]. Available: https://www.freecodecamp.org/news/how-to-detect-objects-in-images-using-yolov8/

[12] R. M. Maningas, K. Pagapong, D. J. Santos, and J. M. Silerio, “The decline of the Baybayin script during the early Spanish period and the Spanish efforts for its...,” Medium, https://thezoomerhistorian.medium.com/the-decline-of-the-baybayin-script-during-the-early-spanish-period-and-the-spanish-efforts-for-its-de561e05eb67

[13] R. Pino, R. Mendoza, and R. Sambayan, “A baybayin word recognition system,” PeerJ Computer Science, vol. 7, Jun. 2021. doi:10.7717/peerj-cs.596

[14] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, “This electronic edition is for non-commercial purposes only. Bayesian Data Analysis Third edition.” Available: http://www.stat.columbia.edu/~gelman/book/BDA3.pdf