Optical Character Recognition of Baybayin Writing System using YOLOv3 Algorithm

Angel Mikaela P. Ligsay [1], John B. Rivera [2], Jocelyn F. Villaverde [3]

[1, 2] School of Electrical, Electronics, and Computer Engineering
Mapúa University
Manila, Philippines
Email: ampligsay@mymail.mapua.edu.ph [1], jbrivera@mymail.mapua.edu.ph [2]

[3] School of Electrical, Electronics, and Computer Engineering
Mapúa University
Manila, Philippines
Email: jfvillaverde@mapua.edu.ph [3]

2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET) | 978-1-6654-6837-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/IICAIET55139.2022.9936792

Abstract— **In the Philippines, Baybayin is one of its writing systems that originated in pre-Hispanic colonialism. The centuries-old writing system gained attention and popularity, which later turned into an approved bill in 2018. The recent development of research aimed at translating Baybayin characters into Alphabets, the globally recognizable writing system, uses Artificial Intelligence or A.I. Different researchers have developed an optical character recognition system for the Baybayin script but are incapable of translating multiple characters in single image and are all using object classification algorithms. Therefore, there is a need for a system using a YOLOv3 based CNN architecture capable of recognizing Baybayin scripts in word form. Using the YOLOv3 algorithm, the system was able to achieve an accuracy of 98.92%. It was observed that some of the misclassifications are due to distorted or illegible handwriting. It can be concluded that the optical character recognition of Baybayin characters using the YOLOv3 algorithm is of high accuracy when it comes to detecting and classifying Baybayin characters.**

Keywords— Baybayin, Optical Character Recognition, YOLOv3 Algorithm, Raspberry Pi, Confusion Matrix

I. INTRODUCTION

The writing system has already been an undisputed factor in the development of civilization. Past writings allow future generations to accumulate knowledge for technological advancements, history learning, and societal improvements. Literacy in a writing system also empowers people to communicate through different channels. This communication promotes understanding of different perspectives [1]. In the Philippines, Baybayin is one of its writing systems that originated in pre-Hispanic colonialism.

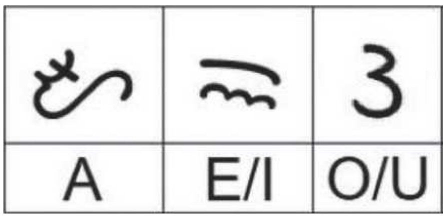

Fig. 1 Sample Baybayin Characters [2]

The centuries-old writing system gained attention and popularity, which later turned into an approved bill in 2018. The Philippine government expressed its efforts in reviving the Baybayin script by integrating Baybayin translation in signages and learning materials. The Baybayin script consists of 3 vowels and 14 consonants, forming 17 characters. Syllables are the foundation of the Baybayin consonants, and the dots of characters state change the syllable's vowel [2].

The recent development of research aimed at translating Baybayin characters into Alphabets, the globally recognizable writing system, uses Artificial Intelligence or A.I. In [3], A.I. is used to discern whether a character is Latin or Baybayin and print the Latin equivalent in the output screen. The article identifies characters on images using optical character recognition (OCR) with the support vector machine (SVM). OCR is simply a process of reading printed/handwritten characters electronically and displaying them as encoded text. [4] is also a study aimed at image classification of Baybayin characters using the Long Short-Term Memory (LSTM) neural network. The training of the model required data that they gathered using an Android app. The Android app shows a Baybayin character reference, which the person using the app will replicate. In the final stages of the paper, the LSTM model achieved 95.6% during training and 92.9% for the actual accuracy. The article [5] utilized an improved You Only Look Once (YOLO) detection network, the YOLOv3, to identify characters. The methodology includes the YOLOv3 algorithm and Faster Region-Based Convolutional Neural Network (Faster R-CNN) comparison in character recognition. In comparison to previous generic object identification algorithms, YOLOv3 is exceptionally accurate and quick because of improvements made to its predecessors, YOLOv1 and YOLOv2. YOLOv3 is made up of two components: the feature extraction layers and the detection layers. The creator of YOLOv3 employs darknet53, a brand-new network for feature extraction. It employs sequential 3 by 3 and 1 by 1 convolutional layers, as well as residual connections similar to residual neural networks, and it utilizes a total of 53 convolutional layers. The comparison resulted in YOLOv3 obtaining a higher F1-score and faster processing time. In the articles' conclusion, it stated that YOLOv3 achieved a recognition rate of 96.08%. However, these research focus only on a single character recognition which can minimize the real-world application.

Different researchers have developed an optical character recognition system for the Baybayin script [6-8]. These systems are only capable of recognizing and converting one Baybayin character only and are unable to translate multiple characters in a single image. Additionally, no researchers used the YOLOv3 model for text recognition of the Baybayin writing system. Therefore, there is a need for a system using a YOLOv3 based CNN architecture capable of recognizing Baybayin scripts in word form.

This study aims to develop an optical recognition system of the Baybayin script that will convert Baybayin text into the digitized standard script, the Latin alphabet. Specifically, this study aims to (1) develop an image capturing device via Raspberry Pi, (2) to use the YOLOv3 algorithm to convert Baybayin text to the Latin alphabet, and (3) to use a confusion matrix to determine the performance of the prototype.

This study is mainly beneficial to the National Commission for Culture and the Arts (NCCA) in achieving the objectives stated in the National Writing System Act of the Philippines, such as the promotion, protection, preservation, and conservation of Baybayin as a tool for cultural and economic development. Other than the NCCA, the research will also aid the Department of Education (DepEd) in the promotion and teaching of the Baybayin script. The study will also be beneficial to the people pursuing studies related to Baybayin, such as studying historical artifacts and documents and learning the manner of writing in the said script. The study will also aid other researchers studying the field of object recognition using YOLO and optical character recognition by providing ample amounts of data and knowledge regarding the topic.

The scope of this study includes the application of optical character recognition techniques to convert Baybayin text to digitalized text using Raspberry Pi. The scope of text recognition and conversion will range from a single Baybayin character to multiple character recognition. The algorithm to be applied in this study will be focused and limited to the YOLOv3 (You Only Look Once version 3) algorithm. The optical character recognition of Baybayin characters will be only limited to the detection of text with white background and black ink.

II. METHODOLOGY

A. *Conceptual Framework*

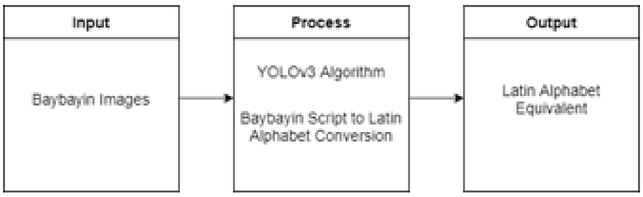

Fig. 2 Conceptual Framework

Figure 2 describes the conceptual framework of the optical character recognition system for the Baybayin script. The input will come from the captured images of the camera module. The images captured will then be processed through the system through the use of the YOLOv3 algorithm to convert this Baybayin text to its Latin alphabet equivalent. The final output would be the display of the digitalized Latin alphabet equivalent translation.

B. *Hardware Development*

The hardware block diagram shown in Figure 3 illustrates the main hardware design of the system. A camera module, a microcomputer, and an LCD display are the main components of the prototype. The camera will serve as the input source of the system which is responsible for the image acquisition process of the system. The Raspberry Pi is a mini-computer used to perform a variety of projects. Due to its affordability and performance, Raspberry Pi is used as a substitute microcontroller in prototypes. Raspberry Pi can also be used as the microcontroller for image processing or object detection projects [9-11].Therefore, the system will use Raspberry Pi for processing the images and videos. The result of the conversion of Baybayin text to the Latin alphabet will be displayed on the screen of the LCD monitor.

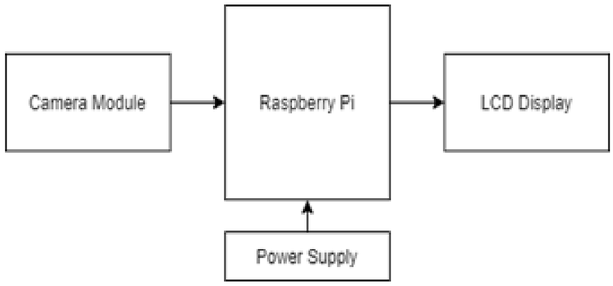

Fig. 3 Hardware Block Diagram

C. *Software Development*

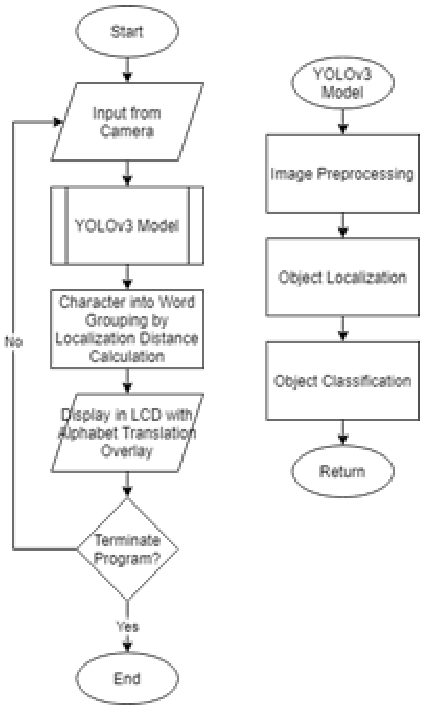

Fig. 4 Main Program Flowchart

The prototype has simple software that instructs the hardware components on what to do. The program treats the video feed from the Raspberry Pi camera as the input. The input goes to the trained YOLOv3 model, which carries on to the object detection process. The objects are localized, and a bounding box is attached to the entity. The algorithm classifies the localized objects according to what Baybayin character they are. The algorithm will also be capable of grouping the Baybayin characters into words by recognizing the spaces between the characters, which translate according to their corresponding alphabet counterpart. Finally, the alphabet translation displays on an LCD monitor as an overlay to the Baybayin words. Figure 4 below shows the complete flow of the prototype software.

1) Training Procedure: An F1-score will gauge the accuracy of the trained model in recognizing Baybayin characters. The F1 score (also called F score and F measure) is a metric derived from the confusion matrix. Precision and recall are the two indispensable factors in calculating the F1 score. Precision is simply the classifier's accuracy in pinpointing relevant entities or predicting actual positives. In the equation below, the precision is the True Positive divided by the summation of True Positive and False Positive. The addition of True Positive and False Positive corresponds to all the detections of the model.

$$\text{Precision} = TP\,/\,(TP + FP) \qquad (1)$$

Meanwhile, recall is the classifier's accuracy in detecting all ground truths. The equation below shows how to calculate the recall. The addition of True Positive and False Negative corresponds to all certainties/actual positives detected.

$$\text{Recall} = TP\,/\,(TP + FN) \qquad (2)$$

The equation below uses precision and recall to solve for the F1 score. The equation shows that the F1 score is a harmonic mean of recall and precision. F1 scores values come in from 1 to 0, where 1 denotes a perfect model, while 0 denotes model failure. A 0.5 F1 score is an average accuracy model, and above that is considered a good model. For this paper, a 0.75 F1 score threshold for the YOLOv3 model is set to conclude that the model is accurate in Baybayin character recognition.

$$\text{F1-Score}=2[(\text{Precision}*\text{Recall})/(\text{Precision} + \text{Recall})] \qquad (3)$$

D. *Data Gathering*

The datasets used for the model training and testing procedure of the system will be taken from the locally generated datasets which consist of 8496 handwritten Baybayin character images. The dataset will be divided into two parts, 80% will be used for model training and 20% will be used for model testing. The datasets will be labeled by drawing bounding boxes around the characters since the datasets from the sources are not yet labeled. Figure 5 shows some of the image samples of the dataset.

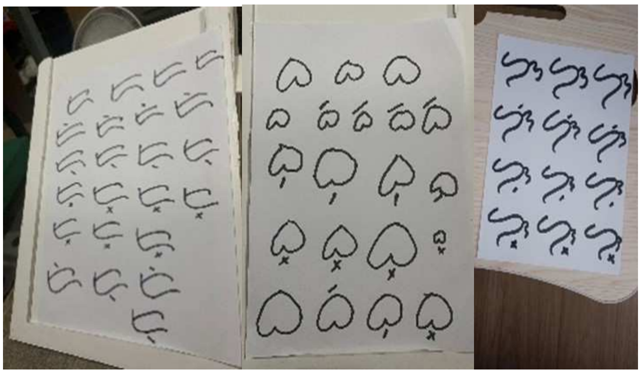

Fig. 5 Sample Images from the Training Dataset

The prototype testing dataset that will be used in testing the final prototype will come from the handwritten character datasets locally generated by the researchers. The total number of the testing datasets is 2124.

E. *Experimental Setup*

Figure 6 below, shows the initial setup of the proposed prototype. A paper tray will be placed in front of the camera. The camera would be able to have a clear view of the paper placed on the paper tray. This will be done by determining the proper position and angle of which the camera would have full coverage of the paper tray. The maximum paper size that can be placed on the paper tray is set to A4 (8.3 x 11.7 inches). The datasets will contain three different-sized images of a specific Baybayin character. The first set will be ranging from a text dimension (W x H) of 0.5x0.5 inches to 3x3 inches, the second set will contain texts with the size ranging from 4x4 inches to 6.25x6 inches, and the third set will contain texts with the size ranging from 6.25x7 inches to 6.25x9 inches. The image will then be processed through the microcomputer. The final output would be the Latin Alphabet equivalent of the Baybayin text input image. The output will be displayed on the LCD display.

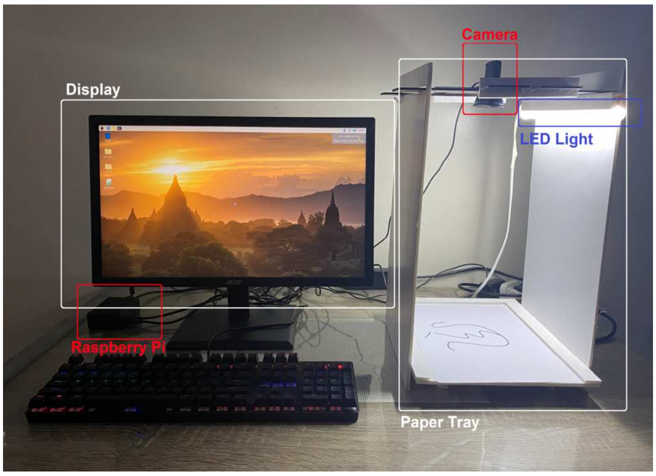

Fig. 6 Experimental Setup

F. *Statistical Analysis*

The confusion matrix will be used in the analysis of the finished model. A confusion matrix is a popular tool in evaluating the accuracy of an object detection model. Four cells with four components form the confusion matrix. The components of the matrix for this study are as follows: True Positive (TP) is when the classifier's prediction is identical to the actual object. False Positive (FP) is when the classifier incorrectly predicts the actual class of the object and predicts a different object. True Negative (TN) is when the classifier correctly predicts the non-entity as a non-entity class. False Negative (FN) is when the classifier fails to correctly predict the actual non-entity as a non-entity class.

This part describes the testing and analysis of the accuracy of the finished trained model. The test sets will consist of images containing different characters of Baybayin as stated in the data gathering section of this paper. Images that contain multiple characters are used to test the accuracy of detection per character (per "class" in the context of confusion matrix) when other characters are present within the image. From the confusion matrix, the accuracy of the finished model can be calculated. The accuracy metric based on the confusion matrix is described as the ratio between the total number of correctly predicted results over the total number of predictions made, as shown in the formula below.

$$\text{Accuracy} = [2(\Sigma TP)\,/\,(2(\Sigma TP) + \Sigma FP + \Sigma FN\,)] \qquad (4)$$

This metric will be the basis on the evaluation of the accuracy of the system in terms of converting Baybayin script. A value of 0 can be considered the lowest accuracy level while a value equal to 1 can be considered the highest accuracy level. values come in from 1 to 0, where 1 denotes a perfect model, while 0 denotes model failure. A 0.5 accuracy level is an average accuracy model, and above that is considered a good model. For this paper, a 0.75 accuracy threshold for the prototype is set to conclude that the prototype is accurate in Baybayin character recognition. The values can also be multiplied to 100 to determine the accuracy levels in terms of percentage.

III. Results and Discussion

Handwritten Baybayin characters were used as input to the Raspberry Pi 4 using a webcam and were recorded under the actual class. The Baybayin characters were detected and classified using the YOLOv3 algorithm. The bounding boxes are accompanied with the corresponding class (Baybayin character label). These labels are then recorded under the predicted class. The total number of characters used in testing is 2124 which is 20% of the total dataset used for the study. From the recorded testing data, the confusion matrix was generated.

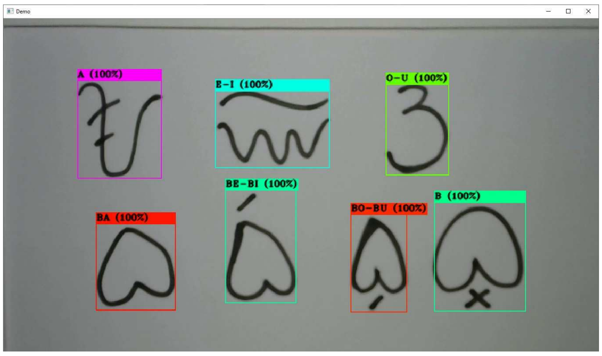

Fig. 7 Baybayin Character Recognition, Detection, and Classification Sample

For the testing procedure, a total of 2142 Baybayin characters. Figure 7 shows a sample of the Baybayin character recognition system made from the YOLOv3 algorithm. As per observation during the testing procedure, the frame rate of the real-time video detection is considerably low. The system's detection capabilities are limited to the characters present within the field of vision of the camera.

TABLE I. Testing Results Sample

Character	Actual Class	Predicted Class
Character 1	GA	O-U
Character 2	B	BE-BI
Character 3	DA	D-R
Character 4	H	HO-HU
Character 5	K	KA
Character 6	WA	WA
Character 7	WE-WI	WE-WI
Character 8	WO-WU	WO-WU
Character 9	TE-TI	TE-TI
Character 10	TO-TU	TO-TU
Character 11	T	T
Character 12	NGE-NGI	NGE-NGI
Character 13	NGO-NGU	NGO-NGU
Character 14	NG	NG
Character 15	PA	PA

Table I shows the 15 testing samples, consisting of 15 characters. As observed, the system was able to recognize the Baybayin characters correctly although there are still a few tests that produced a misclassified output which are all recorded and are reflected on the confusion matrix table.

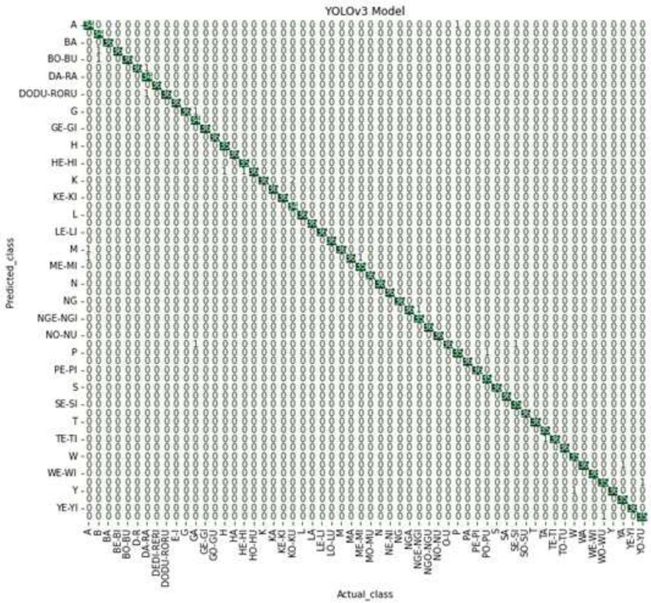

Fig. 8 Confusion Matrix Result for YOLOv3 Model Testing

Figure 8 above shows the Confusion Matrix Result for the custom YOLOv3 Model for the optical character recognition system for Baybayin characters. The results of the confusion matrix were generated from the testing procedure of this study. It is observed that some of the misclassifications are due to distorted or illegible handwriting of the writers.

TABLE II. Model Accuracy

Accuracy							
A	99.86%	**HA**	100.00%	**N**	100.00%	**SO-SU**	100.00%
B	99.90%	**HE-HI**	99.95%	**NE-NI**	100.00%	**T**	100.00%
BA	100.00%	**HO-HU**	99.90%	**NG**	100.00%	**TA**	99.95%
BE-BI	99.95%	**K**	99.90%	**NGA**	99.95%	**TE-TI**	99.95%
BO-BU	99.95%	**KA**	100.00%	**NGE-NGI**	99.95%	**TO-TU**	100.00%
D-R	99.95%	**KE-KI**	100.00%	**NGO-NGU**	100.00%	**W**	99.95%
DA-RA	99.90%	**KO-KU**	99.90%	**NO-NU**	100.00%	**WA**	99.95%
DEDI-RERI	100.00%	**L**	100.00%	**O-U**	99.90%	**WE-WI**	100.00%
DODU-RORU	99.95%	**LA**	100.00%	**P**	99.90%	**WO-WU**	99.90%
E-I	100.00%	**LE-LI**	100.00%	**PA**	100.00%	**Y**	99.95%
G	99.95%	**LO-LU**	100.00%	**PE-PI**	100.00%	**YA**	99.95%
GA	99.90%	**M**	99.95%	**PO-PU**	99.95%	**YE-YI**	100.00%
GE-GI	100.00%	**MA**	99.90%	**S**	100.00%	**YO-YU**	99.90%
GO-GU	100.00%	**ME-MI**	99.95%	**SA**	100.00%	**Overall Accuracy**	98.92%
H	99.95%	**MO-MU**	100.00%	**SE-SI**	99.95%		

IV. Conclusion

The custom trained YOLOv3 model was able to detect and classify all 59 Baybayin characters using a camera. In the testing phase of the study, the model produced an overall accuracy of 0.9892 or 98.92% via Confusion Matrix. It is observed that some of the misclassifications are due to distorted or illegible handwriting. Since an accuracy value greater than or equal to 0.75 is to be considered an accurate model, it can be concluded that the optical character recognition of Baybayin characters using the YOLOv3 algorithm is of high accuracy when it comes to detecting and classifying Baybayin characters.

V. Recommendations

To enhance the real-time detection of the optical character recognition (OCR) system of Baybayin characters, the researchers recommend utilizing a more powerful microcomputer for the deployment of the YOLOv3 model such as the Jetson Nano microcomputer which is equipped with a Graphics Processing Unit (GPU).

To create an adaptive OCR system, the use of different writing tools and images with different background colors are recommended to be included in the training dataset. Another method of increasing the accuracy of the model is to use data augmentation. Data augmentation allows the model to train in different settings and environments without the need of gathering data in set conditions. Using a new model for future studies is also highly recommended as there has been newer and faster OCR models currently. But there is a need to consider the hardware in which the OCR model will be deployed, as faster models require a stronger device. Modifying these OCR models by adding image filtering is also a valid option in further increasing the model's accuracy, however the model's preprocessing time and devices (GPU, RAM, Camera used) must be re-evaluated.

Acknowledgment

This project would not have been made possible without the constant contribution, help, and guidance from different people. Specifically, we would like to give our warmest gratitude to our thesis adviser, Dr. Jocelyn F. Villaverde, for continuing to guide us in the path that this study has taken.

Moreover, we would like to give special thanks to our batchmates from Mapúa University for aiding in the production of our dataset by writing hundreds of Baybayin characters, which has made a huge impact in the success of our study.

Lastly, we would also like to express our gratitude to our professors from Mapúa University for providing us the necessary knowledge and skill that allowed us to pursue this study.

References

[1] M. A. Myasoedova and Z. P. Myasoedova, "CORUS OF GESTURES IN WRITTEN FORM AS A TOOL FOR STUDYING THE PECULIARITIES OF THEIR FORMATION (BY THE EXAMPLE OF RUSSIAN SIGN LANGUAGE)," Международный научный журнал "Современные информационные технологии и ИТ-образование", 2018. https://doi.org/10.25559/SITITO.14.201802.426-435

[2] R. Pino, R. Mendoza, and R. Sambayan "Block-level Optical Character Recognition System for Automatic Transliterations of Baybayin Texts Using Support Vector Machine." 2022 Philippine Journal of Science. 151. 303-315.

[3] R. Pino, R. Mendoza, and R. Sambayan, "Optical character recognition system for baybayin scripts using support vector machine" 2021 PeerJ Computer Science 7:e360 https://doi.org/10.7717/peerj-cs.360.

[4] J. A. Nogra, C. L. S. Romana and E. Maravillas, "LSTM Neural Networks for Baybáyin Handwriting Recognition," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), 2019, pp. 62-66, doi: 10.1109/CCOMS.2019.8821789.

[5] Z. Liu, Z. Wang and Y. Xing, "Wagon Number Recognition Based on the YOLOv3 Detector," 2019 IEEE 2nd International Conference on Computer and Communication Engineering Technology (CCET), 2019, pp. 159-163, doi: 10.1109/CCET48361.2019.8989186.

[6] L. R. Bague, R. J. L. Jorda, B. N. Fortaleza, A. Evanculla, M. A. V. Paez, and J. Velasco, "Recognition of Baybayin (Ancient Philippine Character) Handwritten Letters Using VGG16 Deep Convolutional Neural Network Model," Int. J. Emerg. Trends Eng. Res., vol. 8, 9, pp. 5233–5237, Sep. 2020, doi: 10.30534/ijeter/2020/55892020.

[7] J. A. Nogra, C. L. S. Romana, and E. Maravillas, "Baybáyin Character Recognition Using Convolutional Neural Network," International Journal of Machine Learning and Computing, 2020, Vol. 10, 2, doi: 10.18178/ijmlc.2020.10.2.930.

[8] M. J. A. Daday, "Recognition of Baybayin Symbols (Ancient Pre-Colonial Philippine Writing System) using Image Processing," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, 1, pp. 594–598, Feb. 2020, doi: 10.30534/ijatcse/2020/83912020.

[9] C. C. Paglinawan, A. N. Yumang, L. C. M. Andrada, E. C. Garcia and J. M. F. Hernandez, "Optimization of Vehicle Speed Calculation on Raspberry Pi Using Sparse Random Projection," 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM), 2018, pp. 1-6, doi: 10.1109/HNICEM.2018.8666325.

[10] C. F. Ostia et al., "Electronic Vision System with Personalized Calibration of Color Compensation for People with Partial Color Vision Deficiency Using Raspberry Pi Digital Image Processing," 2019 5th International Conference on Control, Automation and Robotics (ICCAR), 2019, pp. 315-319, doi: 10.1109/ICCAR.2019.8813398.

[11] C. C. Hortinela et al., "Milled Rice Grain Grading using Raspberry Pi with Image Processing and Support Vector Machines with Adaptive Boosting," 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), 2020, pp. 1-6, doi: 10.1109/HNICEM51456.2020.9400102.