Baybáyin Character Recognition Using Convolutional Neural Network

James Arnold Nogra, Cherry Lyn Sta Romana, and Elmer Maravillas

***Abstract*—With the Philippine Congress approving the House Bill 1022 which states that the Baybáyin will be used as the national writing system of the country, the Department of Education and National Commission for Culture and Arts has vowed to reintroduce this old writing system back to the country. In order to hasten up the learning process, a convolutional neural network is designed to check the classification of hand-drawn characters. Nine convolutional neural network models were designed to check which is the best for this type of character recognition. From the 7000 hand-drawn Baybáyin characters used for training, it has found out that the best neural network for this type of classification is composed of three convolutional layers with 32 channels, 64 channels, and 128 channels respectively using 3x3 filters. The final model has also three max pooling layers right after each convolution layer with 2x2 size and two fully connected layers at the end. The number of output of this neural network model is 63 which is the same as the total number of Baybáyin characters. The model yields a 94% accuracy rate using the validation data. The other 8 CNN models also did well with accuracy rates ranging from 57% to 92%.**

***Index Terms*—Baybáyin, character recognition, convolutional neural network, deep neural network, Tensorflow.**

I. Introduction

A. Background of the Study

The Philippine Congress has approved the House Bill 1022 that states that the Baybáyin will become the national writing system of the country. The Department of Education (DepEd) and the National Commission for Culture and the Arts (NCCA), after the approval of the house bill, also showed their full support on this. Currently, there are a few institutions or groups that advocate the teaching of this ancient script like the Sanghabi Organization that aims to teach this script to anyone. There are also numerous texts and items, like vases, stones, and jars that have Baybáyin characters on it that are considered to be a national treasure like the Monreal stone found in Masbate.

Baybáyin (known as Badlit) is an ancient script used by some pre-colonial people living in the northern part of the Philippines. It is one of the few Indic scripts that were being used even before the westerners came. The Indic script is a class of script usually found within the subcontinent of India and some parts of Southeast Asia and some are even still used today. Baybáyin has been recognized as one of the deep components of the Filipino identity [1]. This script has three alphabet characters that represent the vowels A, E/I, and O/U. Baybáyin has also 15 consonants namely B, C/K, D, G, H, L, M, N, NG, P, R, S, T, W, and Y. In the pre-kudlit system, there's only as few as 17 Baybáyin characters but in order to better represent the words, the kudlit system was introduced which brought the total number of characters to 63.

Convolutional Neural Networks (CNN) are always used for character recognition because it looks for features rather than individual pixels. Even though CNN's main applications are for image recognition or classification, it has achieved high accuracy rate in the classification of the MNIST dataset. The advantage of CNN's over multilayer perceptrons is that the number of neurons is way less thus reducing the computational requirements during training.

B. Objectives of the Study

The main objective of this study is to automate the conversion of hand-drawn Baybáyin characters using deep neural networks. One way to design an automated character recognition system is to build a deep neural network that can yield high accuracy in the classification of hand-drawn characters of this ancient script. Tests will be conducted to select the best convolutional neural network model for this character recognition problem.

C. Conceptual Framework

A 28x28 pixel image drawn by a user will be fed to the convolutional neural network model and outputs the correct character classification as shown in Fig. 1. The output will either be any of the 63 classes.

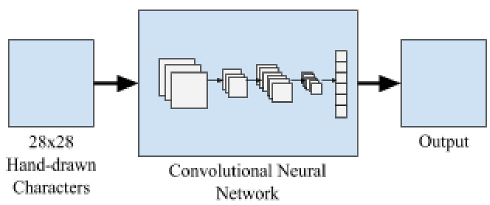

Fig. 1. Shows the conceptual framework of the entire system.

D. Significance of the Study

Because this ancient handwriting system is not used and

Manuscript received May 13, 2019; revised January 11, 2020.

J. Nogra is with the Information Technology Department at Cebu Institute of Technology – University, Philippines (e-mail: jamesnogra@gmail.com).

C. Sta Romana and E. Maravillas are with the College of Computer Studies at Cebu Institute of Technology - University (e-mail: cstaromana@gmail.com, elmer.maravillas@gmail.com).

doi: 10.18178/ijmlc.2020.10.2.930

taught in Philippine schools anymore unlike other Indic scripts of Southeast Asian countries, the new generation will not have an idea or will have difficulty on reading and writing this script once the House Bill 1022 is fully approved. Part of the House Bill 1022 states that in some public areas, there will be signages using this script aside from the English posts. This pre-colonial script is only used and found only in the Philippines so by reintroducing this to the public, this will reinvigorate the cultural identity of the Filipinos.

E. Scope and Limitation

This study will only be limited to recognizing hand-drawn characters. Conversion from Baybáyin characters in books, external images, or texts to its English alphabet word translations is not part of this research. The neural network model being made converts an image with a character on it and then classifies what type of character is it. It only classifies one character at a time, word or paragraph translations are not part of this study.

II. Review of Related Literature

The Philippine Congress has approved that the Baybáyin script will be considered as the national writing system of the country last April 2018 [1]. House Bill 1022 not only aims to introduce the writing system back to the public but also to but also educate the young. The contents of this bill also describe raising awareness and foster a wider appreciation of the importance and beauty of the ancient script [2]. As Congress approved this bill, reintroducing this ancient script to the public is probably somewhat compulsory because part of the bill states that there will be some Baybáyin scripts in public areas. The Baybáyin script characters can support or spell almost all of the vernacular words so it would still be practical that this script to be used on a regular basis.

One way to automate the conversion of Baybáyin automatically and preferably in a smartphone is through optical character recognition (OCR) algorithms. In the Philippines, there are many historical texts and items like vases and pots that have Baybáyin scripts on it. This ancient Philippine script has been recognized as one of the deep components of the Filipino identity [3]. Other nations such as Egypt, China, Greece, and the USA are investing a large effort in restoring and preserving their national historical documents [4]. In order for the new generation of Filipinos to appreciate this writing system, a character translator or converter must be implemented in mobile devices.

Implementations of character recognition usually involve machine learning where hundreds if not thousands of sample data are used to train a model. The most common machine learning technique used nowadays that has unprecedented accuracy are Convolutional Neural Networks (CNN). There are many other optical character recognition (OCR) algorithms out there but CNN is the most commonly used technique for character recognition because it has achieved more than 99% accuracy in the MNIST dataset [5]. Deep convolutional neural networks (DCNN) have achieved great success in various computer vision and pattern recognition applications, including those for handwritten Chinese character recognition (HCCR) [6]. CNN's application is usually for image classification but during its early years, this algorithm has been used to classify handwritten characters. In recent years, deep neural networks have been used in other applications such as object tracking, pose estimation, visual saliency detection, action recognition and scene labeling [7]. Another application of this type of neural network is to predict the ratings of doctors based on their reviews [8]. Another more recent application of deep convolutional neural networks is reading emotions from faces [9]. One downside of CNN's like any other neural networks is that for it to be able to recognize or classify accurately, it needs to be trained by a lot of labeled images or sample data. Also, when training these types of neural network, huge computing power is needed and they also need a long time to train. One example algorithm in which CNN has been applied to improve performance is in High efficiency video coding (HEVC) [10].

III. Methodology

In order to gather images of the Baybáyin characters that will be used in training the convolutional neural network, a separate app shown in Fig. 2 is developed that will let users draw the characters. In the app, the users were presented three sample drawn characters in which they need to follow. During the data gathering, a total of 25 people volunteered to draw the Baybáyin characters. The ages of the volunteers range from 10 to 55, males and females but most of them are 2nd-year high school students. The mobile application to gather the handwritten drawings doesn't require the participant to log in thus making them anonymous contributors. All of the participants don't have an idea or doesn't know how to read and write the Baybáyin writing system.

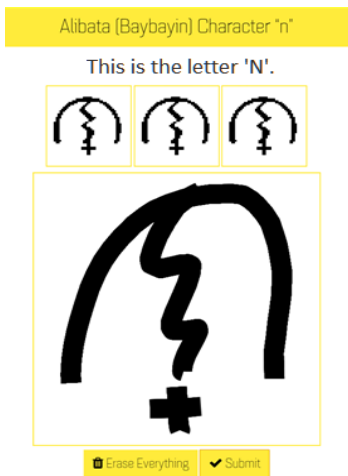

Fig. 2. Screenshots of the app that will gather sample hand drawn Baybáyin.

After two months of gathering data, there were a total of 8100 hand-drawn Baybáyin characters. Each character of the script has been ensured that it should have at least a hundred hand-drawn samples. Each handwritten drawing must be thoroughly checked if it is the correct variation of a character before it can be saved as a training or validation data. The images are then rescaled to 28×28 pixel image which is enough for the features to be still visible and differentiable.

The type of deep neural network that is employed in this study is a convolutional neural network because it has achieved great success in various computer vision and pattern recognition applications, including those for handwritten Chinese character recognition (HCCR) [7]. In order to obtain the best configuration of the convolutional neural network for this type of character recognition, various filter and channel sizes are tested.

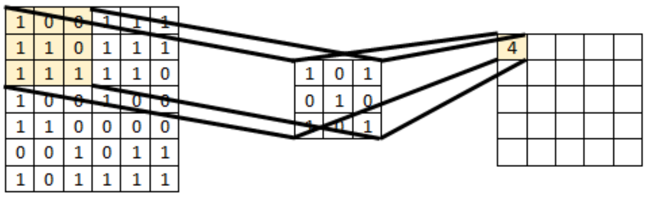

Fig. 3. Example of a convolution operation of a 7×7 input with a 3×3 filter generating a 5×5 output.

The filter sizes used for the convolution were 3×3 shown in Fig. 3, 5×5, and 7×7 while the channel sizes start at 8 channels all the way to 32 channels while the final test has an initial channel size of 32 and reaches up to 128 channels at the third convolution layer.

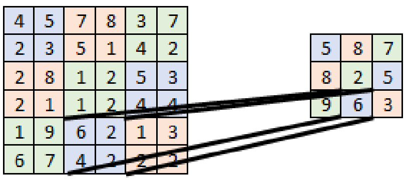

Fig. 4. Example of a max pooling operation of a 6×6 input with 2×2 filter generating a 3×3 output.

The max pooling filter used for all tests were size 2×2 shown in Fig. 4. The output size of the CNN model is 63 which is the same as the number of Baybáyin characters. Also, for all the nine tests, the number of epochs used was 50 while the learning rate is at 0.0001. The only preprocessing done before the images are used to train the neural network aside from resizing it to 28×28 pixels is converting it to a binary image (image with white and black pixels only). The model is implemented in python using the Tensorflow library for convolution, pooling, fully connected layers, dropout, and DNN methods. The dropout used in all of the test models is 80% meaning a neuron has a 0.8 probability being retained. In all of the models, the Dropout function is only done after the first fully connected layer. Dropouts have been introduced to deep neural networks to prevent a particular model from overfitting [11]. If the training loss and the validation loss values are far apart, then that is an indication that a model has overfitted. The optimizer used to train all the test models is Adam optimizer because this method is straightforward to implement, computationally efficient, has low memory usage, is invariant to the diagonal rescaling of the gradients, and is well suited for problems that are large in terms of test data [12]. The activation functions used in the models are ReLU and Softmax. ReLU activation function was used in all the three convolution operations and the first fully connected layer while the Softmax activation function is used in the last fully connected layer. The Softmax activation function at the end is important for the output because it can elevate which is the correct calculated class. To calculate the distance between what the model has been calculated and the correct classification, the cost function Cross Entropy is used shown in (1). The final structure of all the tested CNNs is shown in Fig. 5.

$$H(C, R) = -\sum_i R_i logC_i \qquad (1)$$

Equation (1) is the Cross Entropy loss function used to calculate the distance between the expected output and the calculated output. R is the actual probabilities while the C is the calculated probabilities.

The tests are done using a laptop with a CPU Intel Core i7-7700HQ and GPU Nvidia GTX 1050. The Tensorflow library used is built with GPU support which allows training the model to be a lot faster rather than the CPU version. Other systems were also used during training such as an Ubuntu 16.04 server that has 2.0GHz CPU and 2GB memory.

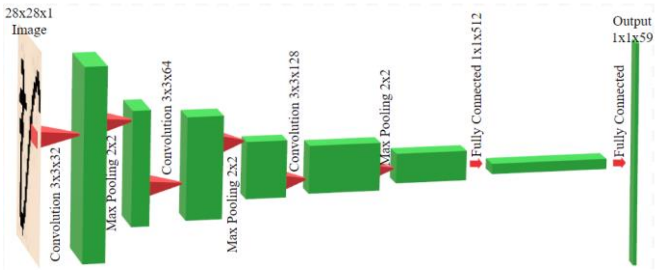

Fig. 5. Basic structure of the CNN model that will be used in the various tests.

The data recorded are the loss, accuracy (both training and validation data), and the time (seconds) it took to train the model. The loss is recorded to ensure that the models did not overfit. If the loss of a certain model, either from the training or validation data, reaches zero, then that model will be discarded as it is an indication that the model has overfitted. The accuracy recorded are for both the training data and validation data. During the training, the 8100 hand-drawn images are divided randomly into two groups. The 7000 images were considered as training data while the other 1100 images are considered as validation data. Only the 7000 images are used to train the models while the 1100 images are used to measure the accuracy of the model during training. During the data gathering and manual selection, almost 2000 images were deleted because it doesn't pass the correctness of the handwriting. Three people are tasked including one of the researcher for the manual selection of images in which will be used as training data and which are to be deleted.

IV. Results and Discussion

In order to get the best convolutional neural network model for the Baybáyin character recognition, a total of nine separate tests were done. The hyperparameters modified in each test are the filter sizes and the channel sizes of the layers. However, the filter size for the max pooling layers is all the same for all the tests which are 2×2. For the convolution operation, for all the layers, the stride used are 1 with no paddings. Every after convolution layer, a max pooling layer is applied to further reduce the dimensions of the previous layer.

Shown in Table I, the CNN model that showed promising results are the models that have 32 channels on the first convolution layer, 64 on the second convolution layer, 128 on the third convolution layer, and 256 and 512 channels on both the final fully connected layers. The model that has the least promising results are the models that start with 8 channels on the convolution layer and has 128 channels on the final fully connected layer.

TABLE I: FINAL VALIDATION ACCURACY RESULTS FROM THE NINE TESTS CONDUCTED

	3×3 Filter Size	5×5 Filter Size	7×7 Filter Size
8 > 16 > 32 > 64 > 128	0.824	0.768	0.567
16 > 32 > 64 > 128 > 256	0.892	0.894	0.835
32 > 64 > 128 > 256 > 512	0.94	0.92	0.915

TABLE II: CNN MODEL RESULTS IN TABULAR FORM WITH 3×3 FILTER SIZE

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Time (s)
10	0.45828	0.8792	0.4386	0.848	6.44
20	0.08423	0.9681	0.34958	0.901	12.11
30	1.69025	0.9163	0.39433	0.901	18.75
40	0.01021	0.9966	0.41069	0.904	24.65
50	0.01701	0.9933	0.41708	0.94	30.43

In Table II, a 3×3 filter is used with 32, 64, and 128 channels in the convolution layers 1 to 3 respectively. The fully connected layer before the output has 256 channels. After 50 epochs, the final training accuracy for this model is 99% while the validation accuracy is at 93%. The loss of using the training data is 0.02 while the loss using the validation data is 0.41 which means the model did not overfit. The training took 31 seconds to complete.

TABLE III: CNN MODEL RESULTS IN TABULAR FORM WITH 5×5 FILTER SIZE

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Time (s)
10	0.51395	0.8593	0.4343	0.847	13.32
20	0.61428	0.9414	0.35326	0.877	26.64
30	1.63052	0.9041	0.27865	0.913	39.52
40	0.01898	0.9943	0.28956	0.925	52.71
50	0.01015	0.9978	0.30605	0.92	64.49

In Table III, a 5×5 filter is used with 32, 64, and 128 channels in the convolution layers 1 to 3 respectively. The fully connected layer before the output has 256 channels. After 50 epochs, the final training accuracy for this model is 100% while the validation accuracy is at 92%. The loss of using the training data is 0.01 while the loss using the validation data is 0.31 which means the model did not overfit. The training took 65 seconds to complete.

TABLE IV: CNN MODEL RESULTS IN TABULAR FORM WITH 7×7 FILTER SIZE

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Time (s)
10	0.8409	0.7362	0.6852	0.749	38.43
20	0.37752	0.8715	0.4017	0.86	75.42
30	1.64049	0.8623	0.38462	0.868	113.46
40	0.10276	0.9643	0.29406	0.898	150.62
50	0.05722	0.9815	0.29035	0.915	187.86

In Table IV, a 7×7 filter is used with 32, 64, and 128 channels in the convolution layers 1 to 3 respectively. The fully connected layer before the output has 256 channels. After 50 epochs, the final training accuracy for this model is 98% while the validation accuracy is at 92%. The loss of using the training data is 0.06 while the loss using the validation data is 0.29 which means the model did not overfit as shown in Fig. 6. The training took 188 seconds to complete.

Shown in Fig. 7, the three models using the channel sizes 32 (Convolution 1), 64 (Convolution 2), 128 (Convolution 1), and 256 (Fully Connected 1), the model with the 3x3 filter size for the convolution operation has the highest accuracy while the model with 7×7 filter size for the convolution operation has the lowest accuracy.

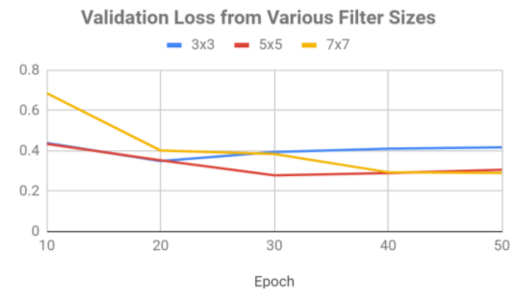

Fig. 6. Shows the graph of the loss of the model using the validation data after 50 epochs.

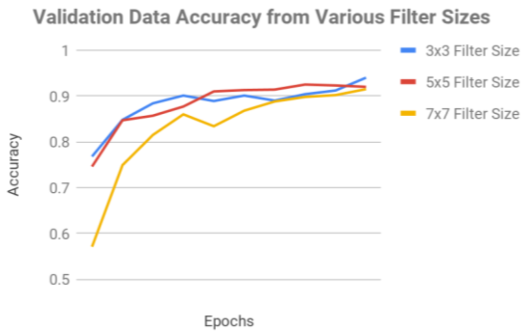

Fig. 7. Shows the graph of the accuracy of the model using the validation data after 50 epochs.

TABLE V: TRAINING DURATIONS OF THE MODELS FOR VARIOUS FILTER SIZES

Epoch	3×3	5×5	7×7
10	6.44	13.32	38.43
20	12.11	26.64	75.42
30	18.75	39.52	113.46
40	24.65	52.71	150.62
50	30.43	64.49	187.86

From the models that were tested, it can be inferred that the smaller the filter size for convolution is, the more accurate the CNN model for character recognition can get. But the smaller the filter size, the more time the model needs to be trained compared to the models that have larger filter sizes. This is because a smaller filter size for convolution can yield a slightly smaller dimension output of the original input compared to a larger filter size which can greatly reduce the dimension of the original input.

As shown in Table V and Fig. 8, the models that are faster to train are the models that have filter sizes of 3×3. One of the three models with 3×3 filter sizes also got the highest validation accuracy. The 7×7 models that were tested were the models that are the longest to train. Validation accuracies of 7×7 models is almost as high as the ones from the 3×3 and 5×5 models but these models have more parameters thus training them is slow.

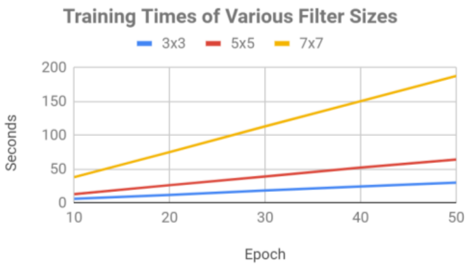

Fig. 8. Shows the graph of the training durations of the models for various filter sizes.

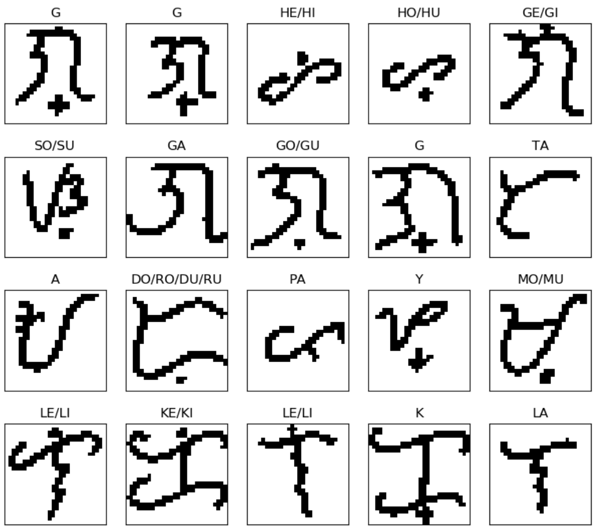

Fig. 9. Shows 20 characters from the validation data all but one (HO/HU at row 4 column 1 should be class H) being correctly classified.

Based from the nine tests, the CNN model with the channel sizes 32 (Convolution 1), 64 (Convolution 2), 128 (Convolution 3), and 256 (Fully Connected 1) with 3×3 filter size is the most accurate for the Baybáyin character recognition with an accuracy rate of 94% (see Fig 6 and Table I). Even though this model is slower to train than the other tested models, the accuracy of this trained model is more important than the training time, besides, when models including this one are used, they usually have the same execution time to classify a handwritten character. One way to achieve a higher accuracy rate for the models is to add more training data. From the calculated loss in all tests, the values of the loss function for validation and training data are almost the same which means that all the models did not overfit. Overfitting was avoided due to the fact that the models have a dropout rate of 80%. A visualized output of the final model is shown in Fig. 9. Most of the incorrect classifications are caused by these types of characters. The characters with the pattern XO/XU (po/pu, yo/yu, ko/ku, etc.) has a dot sign at the bottom while the pattern X (p, y, k, etc.) character has a plus symbol.

After the model training and selection, the model was deployed to an Ubuntu server with 2.0GHz and 2.0GB or memory. This server will be used as a backend for an eLearning app that will be developed in the future. The eLearning app will classify handwritten images of Baybáyin from users of the app. Upon testing, the classification performance of this server is always less than 1 second per character image. The performance measured includes the uploading of the image and the classification itself.

V. Conclusion

Using convolutional neural networks, Baybáyin character recognition yields an accuracy rate of 94%. This accuracy rate is good enough because each character has only less than 200 hand-drawn images of the characters that were used during the training. Adding more test hand-drawn characters for training the neural network will greatly increase the accuracy rate. This has proved that CNN models are suitable for Baybáyin character recognition.

VI. Future Work

To fully re-introduce the Baybáyin script to the public, a website or mobile app that uses the trained model must be made where users can try drawing the characters and get a response about the classification of what they have drawn. This would aid the training of the students or anyone on how to read and write this ancient script. Other recommendations would include testing other neural networks if it can achieve greater accuracy than CNN. One neural network that can probably be tested for better results is the Inception network because this type of neural network can accurately classify images even though it has different resolutions.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

James Arnold Nogra wrote the paper and conducted the experiments. He is also the Baybáyin expert as he attended workshops for this writing system. He also interviewed a couple of experts of the writing system to ensure the quality of data to be gathered. Nogra and Dr. Cherry Lyn Sta Romana worked with the design of the convolutional neural network. Sta Romana also helped with validating the contents of the paper. Dr. Elmer Maravillas helped with the validation of the results. He also helped with designing the neural network. All three authors helped in writing the research paper. All three also made sure that there are no grammatical mistakes and ideas are clearly defined.

Acknowledgment

Our thanks to Cebu Institute of Technology - University's College of Computer Studies for allowing us to undertake this research and to our families for inspiring us always. Also a big thanks to the people who volunteered to draw the hand-written samples that were used for training. A big thanks as well to the contributors who checked the validity of the drawn characters. Also thanks to Leo Castro and NCCA for providing the Baybáyin workshop.

REFERENCES

[1] L. Bataoil, *National Writing System Act. House Bill No. 1022*, Philippine Seventh Congress, First Regular Session, 2018, pp. 3-5.

[2] A. Morallo. (April 2018). House panel approves use of Baybayin as country's national writing system. *PhilStar* [Online]. p. 5. Available: https://www.philstar.com/headlines/2018/04/23/1808717/house-panel-approves-use-baybayin-countrys-national-writing-system

[3] R. Guillermo, M. Paluga, M. Soriano, and V. Totanes, *3 Baybáyin Studies*, University of the Philippines Press, 2017, pp. 7-8.

[4] S. Tangwongsan and C. Sumetphong, "Optical character recognition techniques for restoration of thai historical documents," in *Proc. 2008 International Conference on Computer and Electrical Engineering*, 2008, pp. 531-535.

[5] M. Wu and L. Chen, "Image recognition based on deep learning," in *Proc. 2015 Chinese Automation Congress (CAC)*, Wuhan, 2015, pp. 542-546.

[6] W. Yang, L. Jin, Z. Xie, and Z. Feng, "Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge," in *Proc. 2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, Tunis, 2015, pp. 551-555.

[7] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," in *Proc. 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, Beijing, 2017, pp. 721-724.

[8] R. Sharma, S. Tripathi, S. Sahu, S. Mittal, and A. Anand, "Predicting online doctor ratings from user reviews using convolutional neural networks," *International Journal of Machine Learning and Computing*, vol. 6, no. 2, pp. 149-154, 2016.

[9] W. Abdulsalam, R. Alhamdani, and M. Abdullah, "Facial emotion recognition from videos using deep convolutional neural networks," *International Journal of Machine Learning and Computing*, vol. 9, no. 1, pp. 14-19, 2019.

[10] T. Katayama, T Song, W. Shi, X. Jiang, and T. Shimamoto, "Fast CU determination algorithm based on convolutional neural network for HEVC," *International Journal of Machine Learning and Computing*, vol. 8, no. 5, pp. 442-446, 2018.

[11] N. Srivastava, G. Hinton, A. Krizhevsky, and I. Sutskeve, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, issue 1, pp. 1929-1958, January 2014.

[12] D. Kingma and J. Ba, "A method for stochastic optimization," in *Proc. 3rd International Conference for Learning Representations*, San Diego, 2014.

James Arnold Nogra is a doctorate in IT student of Cebu Institute of Technology - University. He received his master degree in computer science in 2013 in the same university. He got his bachelor degree in computer science in 2010 at the University of the Philippines. He has more than 9 years of industry experience ranging from web, mobile, and full stack development. Aside from working in the industry, he has also experience in the academe teaching courses like introduction to programming and web development.

His research interests are mostly in neural networks like CNN and LSTM. However, he is also involved in several other types of research like image compression.

Cherry Lyn Sta Romana is the current the dean of the College of Computer Studies in Cebu Institute of Technology - University. She obtained her bachelors degree and masters degree in computer science at the University of the Philippines. Before she became a dean, she was a professor of the University of the Philippines for 11 years. She has been working in the academe for more than 27 years.

Elmer Maravillas is the current chair of the Computer Science Department of Cebu Institute of Technology - University. He has been a faculty member of this university for more than 9 years. He obtained his Ph.D. in mechanical engineering in 2006 and masters degree in computer science in 1998 at De La Salle University.