Original Article

Thai Amulet Recognition using Vision Transformers

Kittasil Silanon[1], Kullawat Chaowanawatee[1,*], Thitinan Kliangsuwan[1]

[1]College of Computing, Prince of Songkla University Phuket Campus,

Phuket, 83120, Thailand

* Corresponding author, Email address: kullawat.c@phuket.psu.ac.th

Abstract

This research demonstrated the application of the deep learning method's vision transformers model to categorize a Thai amulet's image. To construct a dataset, we selected famous Thai amulets that can be found easily online. Several vision transforms models are used to train the system. We conducted five experiments to determine the most effective performance model. In Experiment 1, the model will receive direct training. Additional datasets will be produced in Experiment 2 using the augmentation technique. The Test-Time data Augmentation technique will be utilized in Experiment 3 to generate modifications in test dataset images. Finally, labeled and unlabeled data (Pseudo-Labelling) were used simultaneously in each batch to train the model network in Experiment 4. The ensemble learning model combination is employed to improve model performance for Experiment 5. Furthermore, we developed a mobile application for capturing user-provided image data. Users can upload images to the server for processing and access valuable information from the database, such as the background of amulets, creation ceremonies, and associated beliefs, to learn more about Thai amulets.

Keywords: Thai amulets; deep learning; vision transformers; image recognition.

1. Introduction

Thai amulets, or "PHRA KHRUEANG," are symbols of religion that have been with the Thai people for a long time (Premjai & Matthew, 2022). Thais believe these amulets will protect them from danger, make them invincible, and bring good luck. Amulets are small Buddha images but can also feature images of monks, maestros, the Bodhisattvas, and other gods. Almost every Thai Buddhist has at least one amulet. It is expected that both young and older people will wear at least one amulet around their neck to feel closer to Buddha. In addition, amulets are considered a fusion of religion, art, and history, inspiring a significant market for collectors and scholars. However, numerous types of amulets in Thailand are challenging for individuals to understand and recognize. Therefore, many research investigations on Thai amulets have been proposed. Chomtip, Juti, Terapong, and Pimluk (2010) have introduced a system called the "Buddhist Amulet Recognition System" (BARS). A template matching technique is applied to recognize the Buddhist amulet image in the recognition process. The system precision is equal to 80% and takes 0.76 milliseconds pre-image for processing. Chomtip, Vachiravit, Pornpetch, and Nattida (2011) developed a system that can recognize Thai Buddhist sculptures. The system is called the "Thai Buddhist Sculpture Recognition System (TBuSRS)." In image recognition, the Euclidean distance technique is applied to recognize the Buddhist sculpture image. The precision rates of training and un-training datasets are 90.00 percent and 72.38 percent, respectively. The average access time of the system is around 2.72

seconds per image. Chomtip and Natdani (2013) developed the Buddhist Amulet Coin Recognition System (BACRS). The BACRS applied the rule-based technique and genetic algorithm method to recognize amulet coins. The precision rate of the system is 91.53%, and the average access time is 1.05 seconds for the pre-amulet image. Waranat and Chomptip (2014) used image processing and artificial neural networks (ANN) to recognize the digital images of Thai Buddhist amulets. The two basic ANN models employed in this research are perceptron and multi-layer perceptron. The perceptron can recognize the amulet images correctly at 97.67%, and the multi-layer perceptron can produce the best classification result at 100%. Thanachai, Chalie, Toshiako, Pished and Kaneko (2014) demonstrated how to determine the kind of amulet from the taken image. The pre-processing techniques such as grayscale conversion, filtering, Prewitt edge detection, cropping, and resizing are applied to compute the image similarity value, and then the template with the highest similarity value is identified as the recognition result. The system can identify the recognition rate at 60.6% accuracy. Weera and Thanasin (2015) proposed the development of Thai Buddha amulet identification using a simple local correlation feature. The feature is used with K-nearest neighbors for the classification process. The result showed that the proposed method gains a high recognition rate of about 89.35%. Narut and Sangthong (2020) applied the Convolution Neural Network (CNN) of the deep learning method to classify the Benjapakee Buddha amulets images. The CNN architecture is designed for suitable recognition. The efficiency of the model could correctly identify 80% of amulet images. Tanasai (2021) presented a case study of Thai amulet recognition using a geometric surface image. The geometric surface images and their color photographs are first trained with the Generative Adversarial Networks (GAN) model. The trained generator model is then used to predict the geometric surface image from the input color image. The evaluation showed that the predicted geometric surface images contain less ambiguity than their color image counterparts under different lighting conditions. Chomtip and Varin (2022) used the pre-training CNN model (ResNet50) to recognize Buddhist amulets. Furthermore, pre-training CNN models such as ResNet18 and ResNet101 are trained to compare the accuracy. The system also conducted cross-validation on an untrained dataset with accuracy, sensitivity, specificity, and precision rates of 99%, 95%, 99%, and 95.41%, respectively.

Previous studies have made significant strides in amulet recognition using various technological methods. However, there still needs to be a gap in effectively categorizing these amulets using the latest advancements in deep learning. This paper seeks to fill this gap by demonstrating the application of Vision Transformers (ViTs), an advanced deep-learning model, in categorizing Thai amulet images. Unlike previous studies, which primarily focused on traditional image processing techniques and conventional neural network architectures, our research leverages the advanced capabilities of ViTs. We selected well-known Thai amulets to form a comprehensive dataset, ensuring the system's relevance and applicability. Our contribution is twofold: First, we explore the effectiveness of ViTs in the specific context of Thai amulet recognition, a domain yet to be extensively studied with such advanced technology. Second, we conduct five experiments designed to test different aspects of the ViTs model, including direct training, data augmentation, Test-Time Augmentation (TTA), Pseudo-Labelling, and ensemble learning. This comprehensive approach aims to optimize the model's performance in recognizing the diverse range of Thai amulets. Additionally, we developed a mobile application for capturing user-provided image data. Users can upload images to the server for processing and access valuable information from the database, such as the background of amulets, creation ceremonies, and associated beliefs, to learn more about Thai amulets. We organize the paper as follows. Section 2 describes the details of the dataset and ViTs model. We present the experiment details and results in Section 3. Finally, we conclude this paper in Section 4.

2. Dataset and Vision Transformers Model

In this section, we first describe the classes and information of the dataset. Then we describe the details of vision transformers (ViTs) model and its characteristic.

2.1 Thai Amulet Dataset

Since there are many different kinds of amulets, we selected famous amulets that are simple to generate a dataset that consists of "**Phra Khun Phaen**," "**Phra Kring**," "**Phra Nang Phaya**," "**Phra Phong Suphan**," "**Phra Rod**," "**Phra Somdej**," and "**Phra Sum Kor**" (Prowd, 2021). We collected Thai amulet images from public datasets. The dataset was randomly partitioned into training and test sets, with 80:20 ratios, resulting in 1631 and 404 images in the respective subsets. For our dataset, all samples were resized to match the input shape of the models under test. Consequently, they were normalized using ImageNet normalization statistics. Figure 1 shows image of selected Thai amulet classes.

[**Figure 1**]

2.2 Vision Transformers Model

Vision transformers (ViTs) (Alexey et al., 2021) are a type of deep learning neural network architecture that was originally developed for natural language processing (NLP) tasks. However, ViTs have also been shown to be effective for various computer vision tasks, such as image classification, object detection, and semantic segmentation. ViTs work by first converting images into a sequence of patches. Each patch is then embedded into a high-dimensional vector space. The embedded patches are then fed into a transformer encoder, which learns to represent the relationships between different parts of the image. The transformer encoder is a stack of self-attention layers. Self-attention is a mechanism that allows the model to learn long-range dependencies in the data. This is important for computer vision tasks, as it allows the model to learn relationships between different parts of the image, even if they are far apart. Once the transformer encoder has processed the embedded patches, the outputs are fed into a decoder, producing a vector representing the probability distribution of the image belonging to different classes. ViTs have several advantages over traditional CNN models for computer vision tasks. Such as being more efficient in terms of both memory and computing requirements, better at learning long-range dependencies in the data, more flexible, and easily adapted to different tasks. Currently, EVA-02 (Yuxin et al., 2023) is a powerful and versatile ViTs model that can be used for various applications. EVA-02 has achieved state-of-the-art results on many vision tasks while utilizing significantly fewer parameters and computing budgets than other vision transformers models. There are four variations of the EVA-02 model, with sizes ranging from 6M to 304M parameters. These models perform magnificently in image recognition tasks (Table.1).

[**Table 1**]

3. Experiments and Results

Five experiments were conducted to assess the classification performance of Thai amulet recognition. Essential parameters for all experiments are setup when training the model. The learning rate value controls how quickly the model learns and is set to 2e^{-3}. The batch size, the number of images the model processes in each iteration, is set to 16. The number of training epochs, the number of times the model sees the entire training dataset, is set to 30. The input image to train the model is resized to 224x224 pixels. Predictive model evaluation is done using the K-Fold Cross Validation approach (Shanthababu, 2023). The dataset is divided into five folds. The model is then trained and evaluated five times, using a different fold as the validation set each time. This technique reduces the risk of overfitting and provides a more accurate estimate of the model's generalization performance.

3.1 Vision Transformers Training

In this experiment, we executed the EVA-02 model in four distinct variations (Tiny, Small, Base, and Large). Each model was directly trained using the training dataset. The performance is evaluated using accuracy and weighted F1-score. Accuracy is a simple measure of overall correctness. At the same time, the weighted F1-score is a more sophisticated metric that considers class imbalances. Testing model results are shown in Table 2. The large EVA-02 model at the fifth fold achieved the best performance at 62.35% accuracy and 61.86% weighted F1-score. As part of the class performance analytics, Figure 2 demonstrates that while the model has a reasonable recognition rate for some classes, it struggles with others. Notably, the "**Phra Sum Kor**" class has a high recognition rate, but "**Phra Rod**" and "**Phra Somdej**" appear to be frequently confused with "**Phra Sum Kor**". This suggests the model requires more distinctive features to differentiate between amulets with similar design elements. Direct training without augmentation seems insufficient for the model to learn the necessary robust features. Figure 3 illustrates some errors of recognition. However, the recognition results could be more satisfactory because of the limited number of datasets. As a result, in the following experiment, we will increase the number of training datasets to improve the model's performance.

[**Table 2**]

[**Figure 2**]

[**Figure 3**]

3.2 Data Augmentation

Data augmentation is a technique used in deep learning to increase the size and diversity of a training dataset artificially (Alexander et al., 2020). This is done by creating new data from existing data using various transformations. In order to increase the number of training datasets, this experiment used horizontal/vertical flipping, space translation, random focus, noise addition, and slight rotation. Figure 4 illustrates a dataset resulting from data augmentation. Testing model results are shown in Table 3. The base EVA-02 model at the fifth fold achieved the best performance at 66.0% accuracy and 65.35% weighted F1-score. As part of the class performance analytics, Figure 5 demonstrates that the recognition rates for most classes and "**Phra Rod**" and "**Phra Somdej**" improved, indicating that augmentation helps the model generalize better. However, there still needs to be more clarity in some classes. These errors imply that while augmentation adds robustness, it may not address the model's sensitivity to intra-class variability. Employing data augmentation increases the amount of training datasets to train the model. However, we can increase the number of test datasets to improve model performance, as described in the following experiment.

[Figure 4]

[Table 3]

[Figure 5]

3.3 Test-Time Data Augmentation

Test time data augmentation (TTA) is a technique that involves applying random transformations to test images before making predictions (Masanari, 2021). This can improve the model's performance by making it more robust to variations in the input data. TTA is similar to data augmentation, typically used during training to increase the size and diversity of the training dataset. However, TTA is applied to test images, which are not used to train the model. After the test images have been augmented, the model makes predictions on each augmented image. The final prediction is then made by averaging the predictions from the augmented images. We compared the performance results of the test dataset without TTA from the previous experiment and the test dataset using the TTA technique of models. From Table 4, the base EVA-02 model at the fifth fold using the TTA technique achieved the best performance at 67.72% accuracy and 66.81% weighted F1- score. As part of the class performance analytics, Figure 6 demonstrates the classification results for all classes. The "**Phra Somdej**" class still improved the results with an accuracy of 45%. Many others had increased accuracy in this experiment, such as the "**Phra Khun Phaen**" class, "**Phra Nang Phaya**" class, "**Phra Phong Suphan**" class, and the "**Phra Sum Kor**" class, but issues remain in others like "**Phra Rod**". This suggests that TTA helps with generalization to some extent but may also introduce noise that can lead to misclassification in cases where the model is not well-tuned to the augmented data. In the following experiment, we discussed using labeled and unlabeled datasets to improve the model's performance.

[Table 4]

[Figure 6]

3.4 Pseudo-Labelling

Pseudo-labeling (Dong-Hyun, 2013) is a semi-supervised learning technique that can be used to improve the performance of models. It uses a model trained on a labeled dataset to predict labels for an unlabeled dataset. The predicted labels (pseudo-labels) are then used to retrain a model, which can achieve better performance than the model trained on the labeled dataset alone. In this experiment, the best performance model from the previous experiment is selected to predict pseudo-labels for unlabeled data. The labels (training datasets) and pseudo-labels datasets are used to retrain the EVA-02 model. Finally, the new retrained model predicts the test dataset with the TTA technique. Table 5 shows the performance model using Pseudo labeling. The large EVA-02 model at the fourth fold using the Pseudo labeling technique achieved the best performance at 69.80% accuracy and 69.12% weighted F1-score. As part of the class performance analytics, Figure 7 demonstrates the classification results for all classes. The classes of "**Phra Khun**

Phaen," "**Phra Kring**," "**Phra Phong Suphan**," and "**Phra Rod**" are improved. However, the confusion matrix indicates persistent misclassifications, for example, between "**Phra Nang Phaya**" and "**Phra Phong Suphan**". This could imply a need for better-quality pseudo-labels or a more refined approach to integrating unlabeled data into the training process. To enhance the recognition rate, in the final experiment, we will discover how to use a combination of multiple models for better prediction results.

[**Table 5**]

[**Figure 7**]

3.5 Ensemble Learning

Ensemble learning (Jason, 2021) is a machine learning technique that combines the predictions from multiple models to produce a more accurate and robust prediction. It is based on the idea that a group of models can make better predictions than any individual model. The best model from the previous is selected as a model base (The large EVA-02 model at the fourth fold using the Pseudo labeling technique). We added one model to evaluate the recognition accuracy, and an averaging technique was utilized. Subsequently, the class with the maximum probability after averaging is selected as the final prediction. Table 6 demonstrates the recognition rate after ensemble learning. Combining the Large-Fold 4 and Large-Fold 3 models achieved the best performance, around 70.0%. As part of the class performance analytics, Figure 8 shows improved performance across most classes, suggesting that ensemble methods can effectively reduce the impact of individual model biases. Nevertheless, some misclassifications persist, indicating that the constituent models in the ensemble may still share common weaknesses, such as overfitting certain features that need to be more indicative of the correct class. The "**Phra Rod**" and "**Phra Somdej**" classes should search for additional datasets to improve the model's accuracy further.

[Table 6]

[Figure 8]

3.6 Mobile Application for Thai Amulet Recognition

The application is named "Thai Amulet Recognition" (Figure 9 (a)). Users can upload and classify Thai amulet images with this application. We used the Flutter framework to build the front-end part of our mobile application, while Python and the Flask framework were employed to develop the backend part. The classification model is stored on the server. Users can use the application's main interface to capture Thai amulet images or choose from their mobile gallery (Figure 9 (b)). Subsequently, the application uploads the image to the server. The server then prompts the trained model to predict the most likely classes of Thai amulets. The output class retrieves additional details from the database, including images, class names, and the background to learn more about Thai amulets (Figure 9 (c)).

[**Figure 9**]

4. Conclusions

In this research, we successfully developed and evaluated a series of training model experiments for Thai amulet recognition, employing a ViTs model alongside various techniques like data augmentation, test-time data augmentation, pseudo-labeling, and ensemble learning. Our comprehensive approach demonstrated enhanced performance in amulet recognition and culminated in developing a user-friendly mobile application. This application allows users to upload amulet images and receive detailed information quickly. Our methodologies have the potential to be applied in broader contexts, such as cultural heritage preservation and educational tools, and even in enhancing the tourism industry by offering interactive and informative experiences related to Thai cultural artifacts. However, the limitation that makes our work less accurate is that the number of image datasets still needs to be more significant. Moreover, the current model's performance in varying environmental conditions, such as different lighting or angles, has yet to be extensively tested. Addressing these limitations will be crucial for enhancing the practicality and robustness of the recognition system. For future work, we will collect more data and create other experiments for several types of Thai amulets in order to enhance the recognition algorithm. We will compare the ViTs model used in this study and the CNN model. Another avenue for future research is the integration of user feedback mechanisms in our mobile application. This feature would allow users to provide real-time feedback on recognition accuracy, thereby enabling continuous learning and improvement of the algorithm.

Acknowledgments

The authors would like to thank College of Computing and Prince of Songkla University, Phuket Campus, Thailand for their support and for providing a working area to test our system.

References

Iamkhorpung, P.., & Kosuta, M.. (2022). The relationship between Buddhist and animist amulets in contemporary Thailand: PHRA KHRUEANG and KHRUEANG RANG. Kasetsart Journal of Social Sciences, 43(1), 53–59. Retrieved from https://so04.tci-thaijo.org/index.php/kjss/article/view/256957

Pornpanomchai, C., Wongkorsub, J., Pornaudomdaj, T. & Vessawasdi, P. (2010). Buddhist amulet recognition system (BARS). *2010 Second International Conference on Computer and Network Technology*, Bangkok, Thailand, 2010, pp. 495-499, doi: 10.1109/ICCNT.2010.128.

Chomtip, P., Vachiravit, A., Pornpetch, I., & Nattida, P. (2011). Thai Buddhist sculpture recognition system (TBuSRS). International Journal of Engineering and

Technology. Vol. 3(4): 342-346 ISSN: 1793-8236 DOI: 10.7763/IJET.2011.V3.250

Pornpanomchai, C., & Srisupornwattana, N. (2013). Buddhist amulet coin recognition by genetic algorithm. *2013 International Computer Science and Engineering Conference (ICSEC)*, Nakhonpathom, Thailand, 2013, pp. 324-327, doi: 10.1109/ICSEC.2013.6694802.

Kitiyanan, W., & Pornpanomchai, P. (2014). Thai Buddhist amulet recognition system. *The 2014 International Conference on Informatics and Advanced Computing, Bangkok,* Thailand, 2014, pp. 9-14

Sauthananusuk, T., Charoenlarpnopparut, C., Kondo, T., Bunnum P., & Hirohiko, K. (2014). Thai Amulet Recognition Using Simple Feature. *The International Conference of Information and Communication Technology for Embedded Systems*, Ayutthaya, Thailand, 2014.

Weera, K, & Thanasin, B. (2015). Robust texture classification using local correlation features for Thai Buddha amulet recognition. Applied Mechanics and Materials (Volume 781). pages 531-53. DOI: https://doi.org/10.4028/www.scientific.net/AMM.781.531

Narut, B., & Sangthong B. (2020). Classification of Benjapakee buddha amulets image by deep learning. RMUTSB ACADEMIC JOURNAL. Vol. 8 No. 1 (2020)

Sucontphunt, T. (2021). Geometric surface image prediction for image recognition enhancement. *Smart Computing and Communication. SmartCom 2020.*

Lecture Notes in Computer Scienc, vol 12608. Springer, Cham. https://doi.org/10.1007/978-3-030-74717-6_27

Chomtip, P., Varin, P. (2022). Buddhist amulet recognition by using RestNet50. Srinakharinwirot Science Journal. Vol. 38 N.2 (2022)

Prowd, I. (2023, October 10). Types of Buddha Amulets That Bring You More Chok So You Can Live with Prosperity. Retrieved from https://thesmartlocal.co.th/buddha-amulets-thailand/

Alexey, D. el et, (2021). An image is worth 16x16 words: transformers for image recognition at scale. *ICLR*, 2021. Retrieved from https://arxiv.org/abs/2010.11929

Yuxin, F. et al. (2023) EVA-02: A Visual Representation for Neon Genesis. arXiv preprint arXiv:2303.11331, 2023.

Shanthababu,S. (2023, October 10). K-Fold Cross Validation Technique and its Essential. Retrieved from https://www.analyticsvidhya.com/blog/2022/02/k-fold-cross-validation-technique-and-its-essentials/

Alexander, B. et al. (2020). Albumentations: fast and flexible image augmentations. Information, vol. 11, no. 2, 2020, doi: 10.3390/info11020125.

Masanari, M. (2012). Understanding test-time augmentation. *In International Conference on Neural Information Processing,* 2021, pp. 558–569.

(Dong-Hyun, L. (2013) Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. *In Workshop on challenges in representation learning, ICML*, 2013, vol. 3, no. 2, p. 896.

Jason, B. (2023, October 10) . Ensemble learning. Retrieved from https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/

List of all the figures.

Figure 1 Thai amulet.

Figure 2 Confusion matrix of the large EVA-02 at 5th fold.

Figure 3 Example of error recognition.

Figure 4 Example of data augmentation.

Figure 5 Confusion matrix of the base EVA-02 at 5th fold with data augmentation.

Figure 6 Confusion matrix of the base EVA-02 at 5th fold with TTA data augmentation.

Figure 7 Confusion matrix of the large EVA-02 at 4th fold with pseudo labeling.

Figure 8 Confusion matrix of the ensemble learning of model.

Figure 9 Mobile application for Thai amulet recognition.

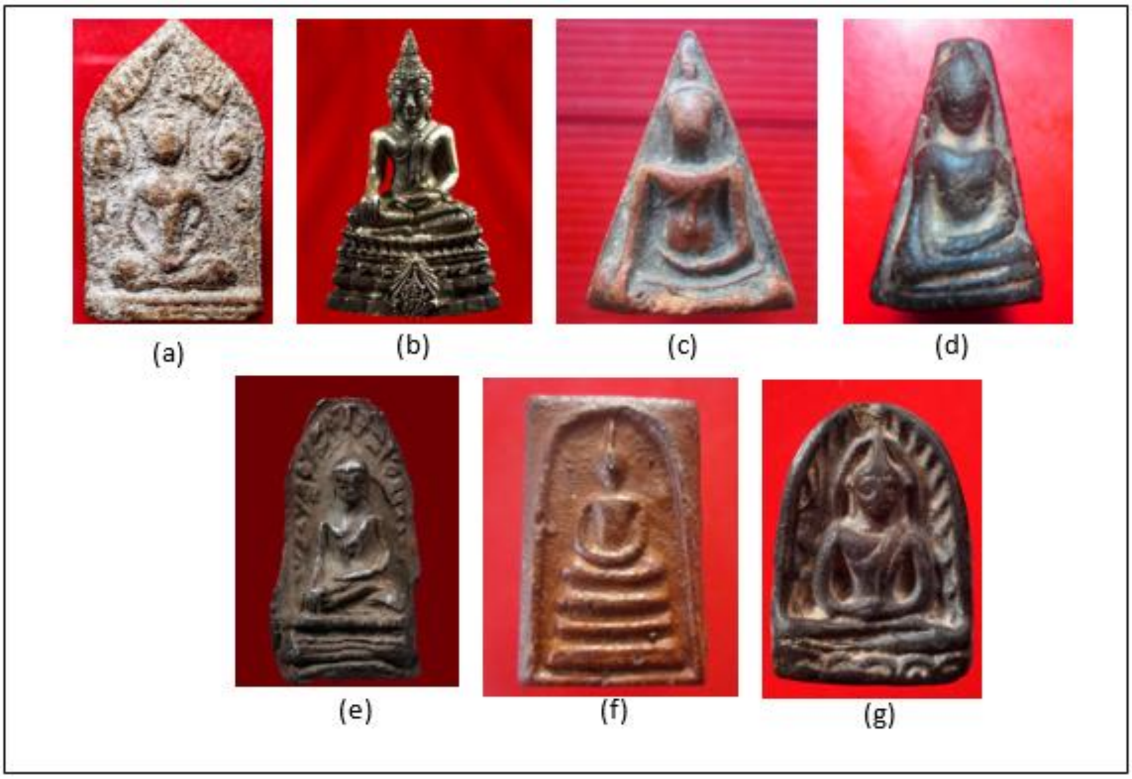
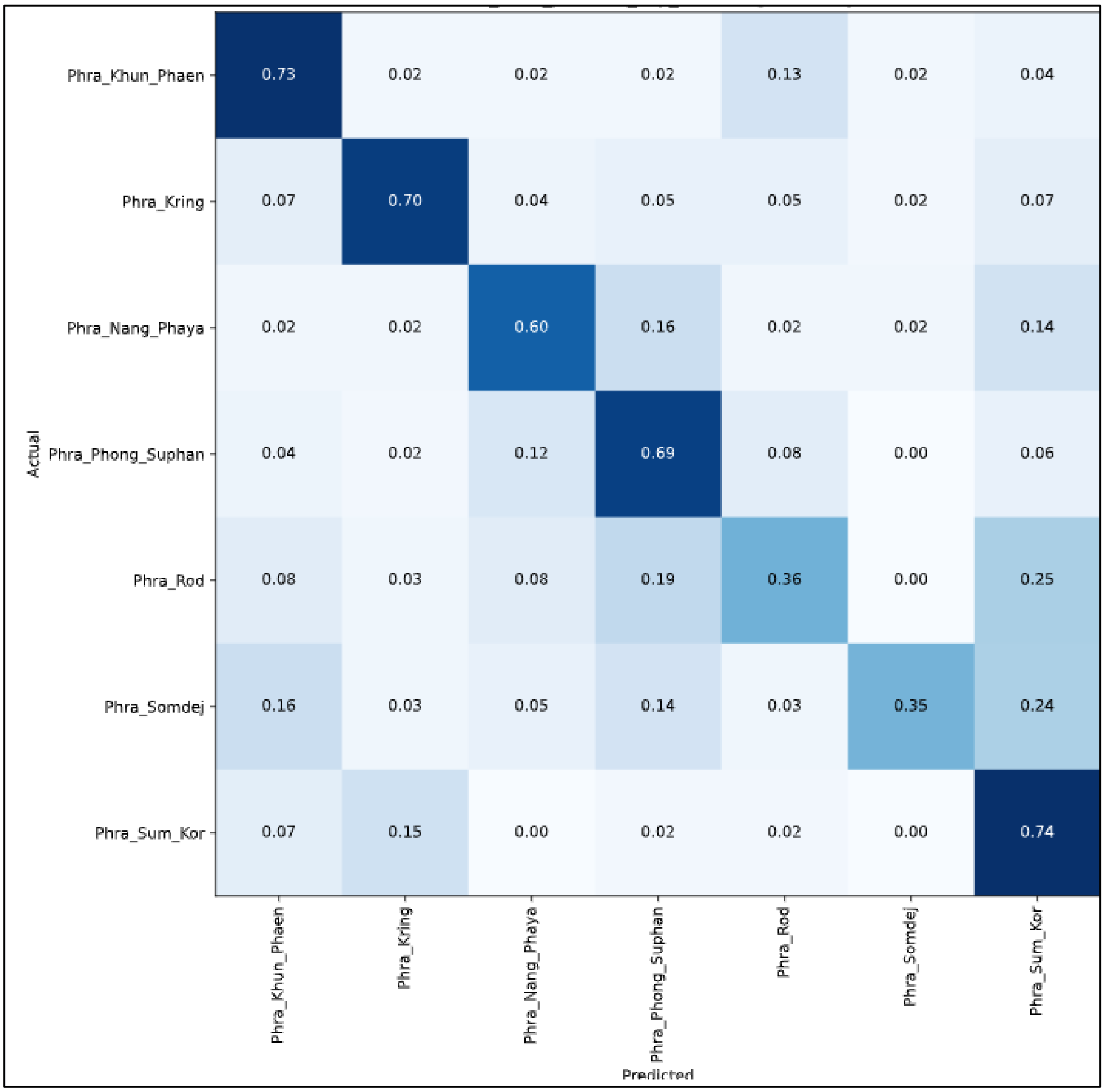


Figure 1 Thai amulet: (a) Phra Khun Phaen, (b) Phra Kring, (C) Phra Nang Phaya, (d) Phra Phong Suphan, (e) Phra Rod", (f) Phra Somdej, (g) Phra Sum Kor.



Actual \ Predicted	Phra_Khun_Phaen	Phra_Kring	Phra_Nang_Phaya	Phra_Phong_Suphan	Phra_Rod	Phra_Somdej	Phra_Sum_Kor
Phra_Khun_Phaen	0.73	0.02	0.02	0.02	0.13	0.02	0.04
Phra_Kring	0.07	0.70	0.04	0.05	0.05	0.02	0.07
Phra_Nang_Phaya	0.02	0.02	0.60	0.16	0.02	0.02	0.14
Phra_Phong_Suphan	0.04	0.02	0.12	0.69	0.08	0.00	0.06
Phra_Rod	0.08	0.03	0.08	0.19	0.36	0.00	0.25
Phra_Somdej	0.16	0.03	0.05	0.14	0.03	0.35	0.24
Phra_Sum_Kor	0.07	0.15	0.00	0.02	0.02	0.00	0.74

Figure 2 Confusion matrix of the large EVA-02 at 5th fold.

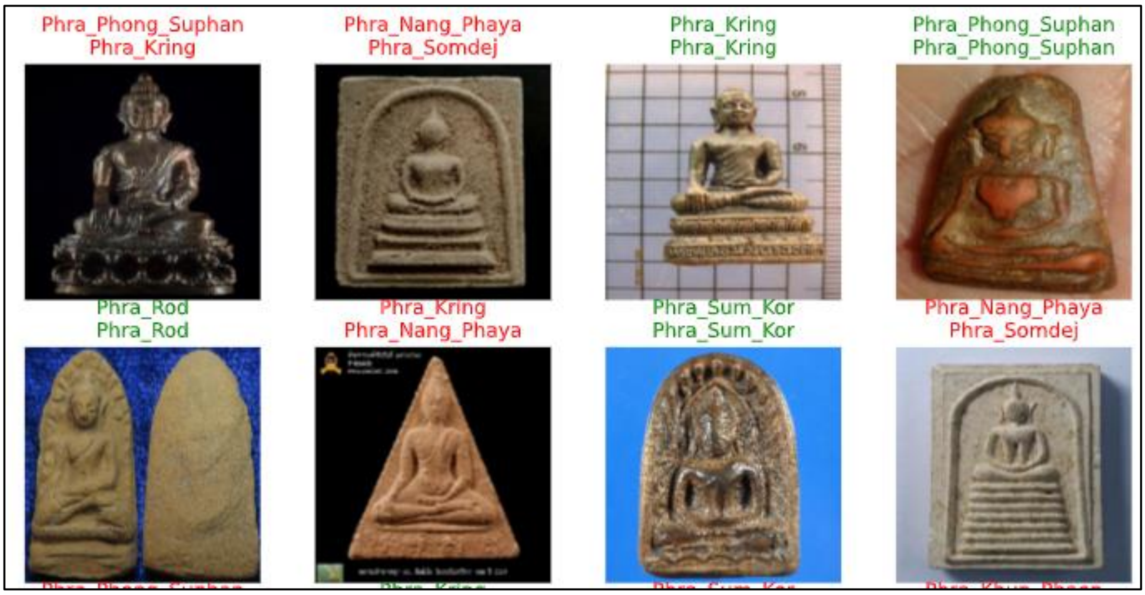


Figure 3 Example of error recognition.

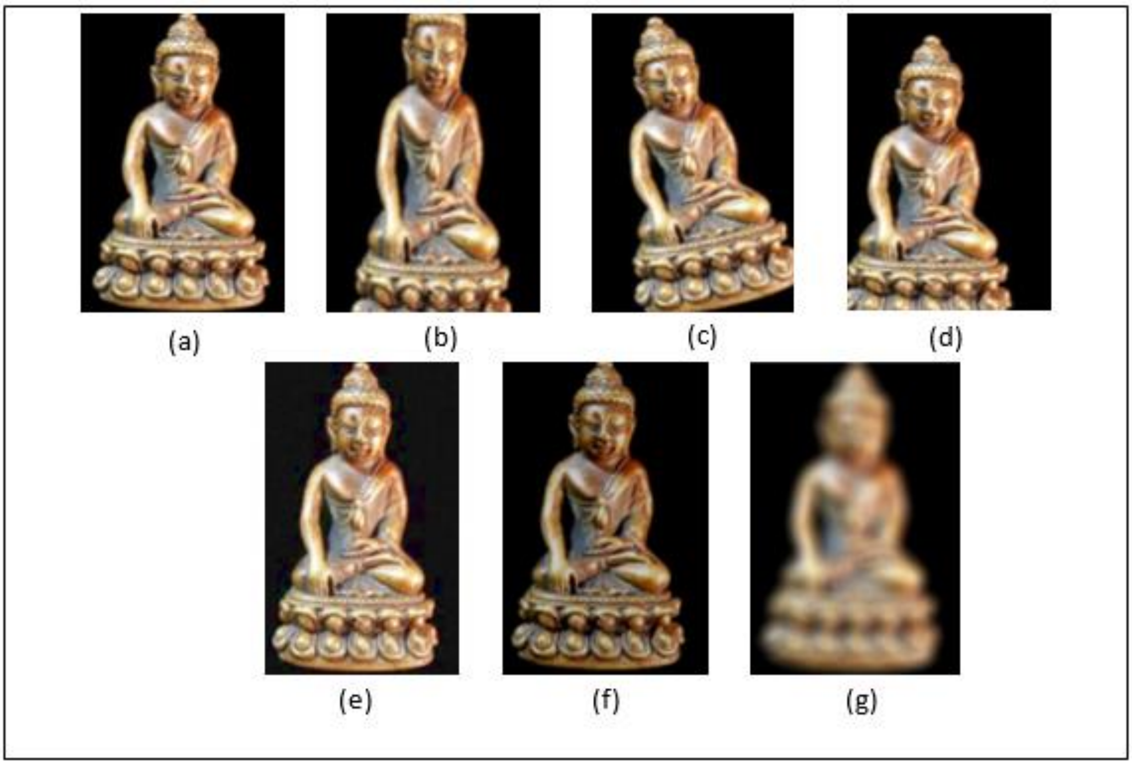


Figure 4 Example of data augmentation: (a) original, (b) scale, (c) rotate, (d) translate, (e) noise, (f) contrast, (g) defocus.

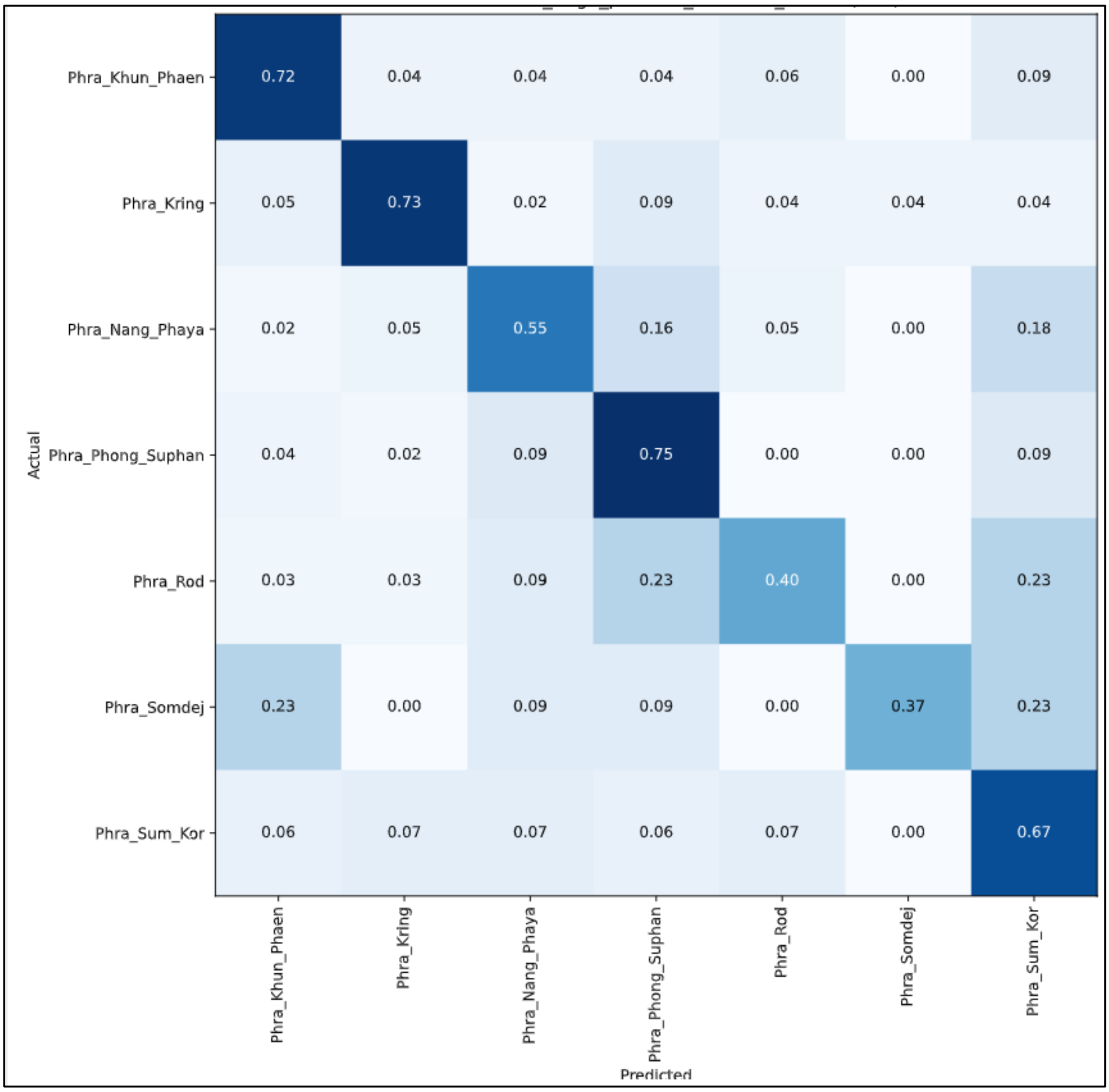


Figure 5 Confusion matrix of the base EVA-02 at 5th fold with data augmentation.

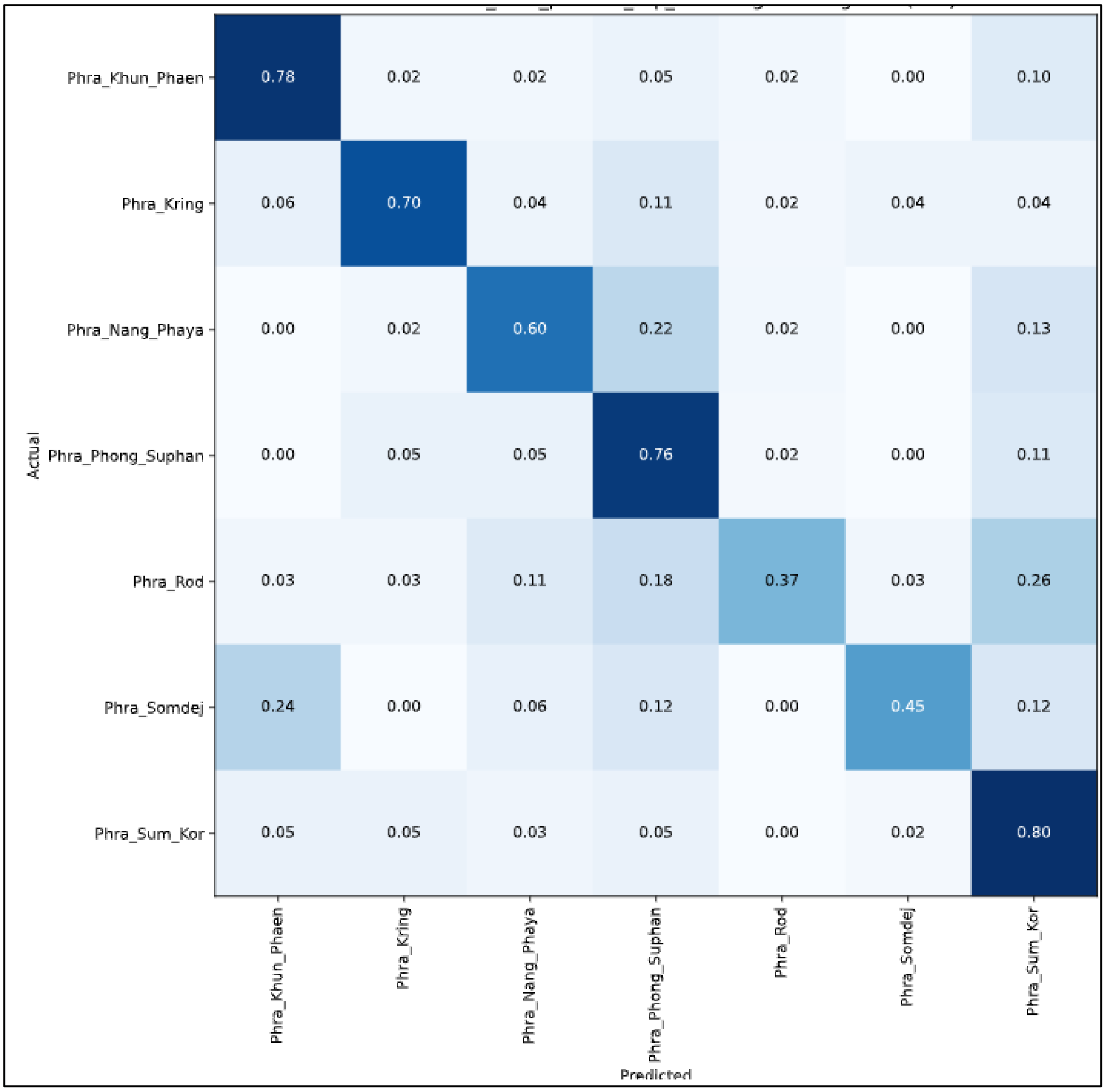


Figure 6 Confusion matrix of the base EVA-02 at 5th fold with TTA data augmentation.

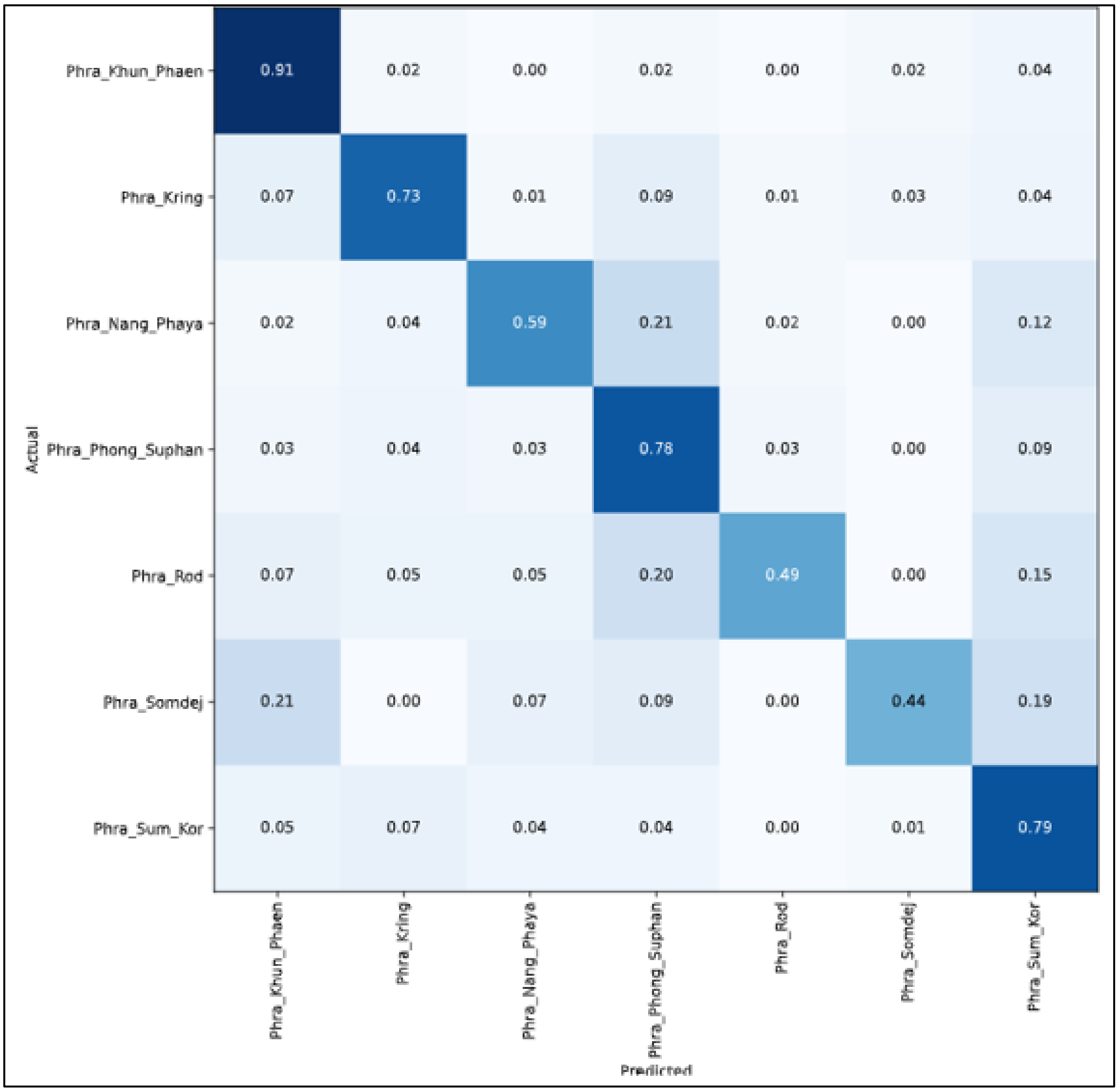
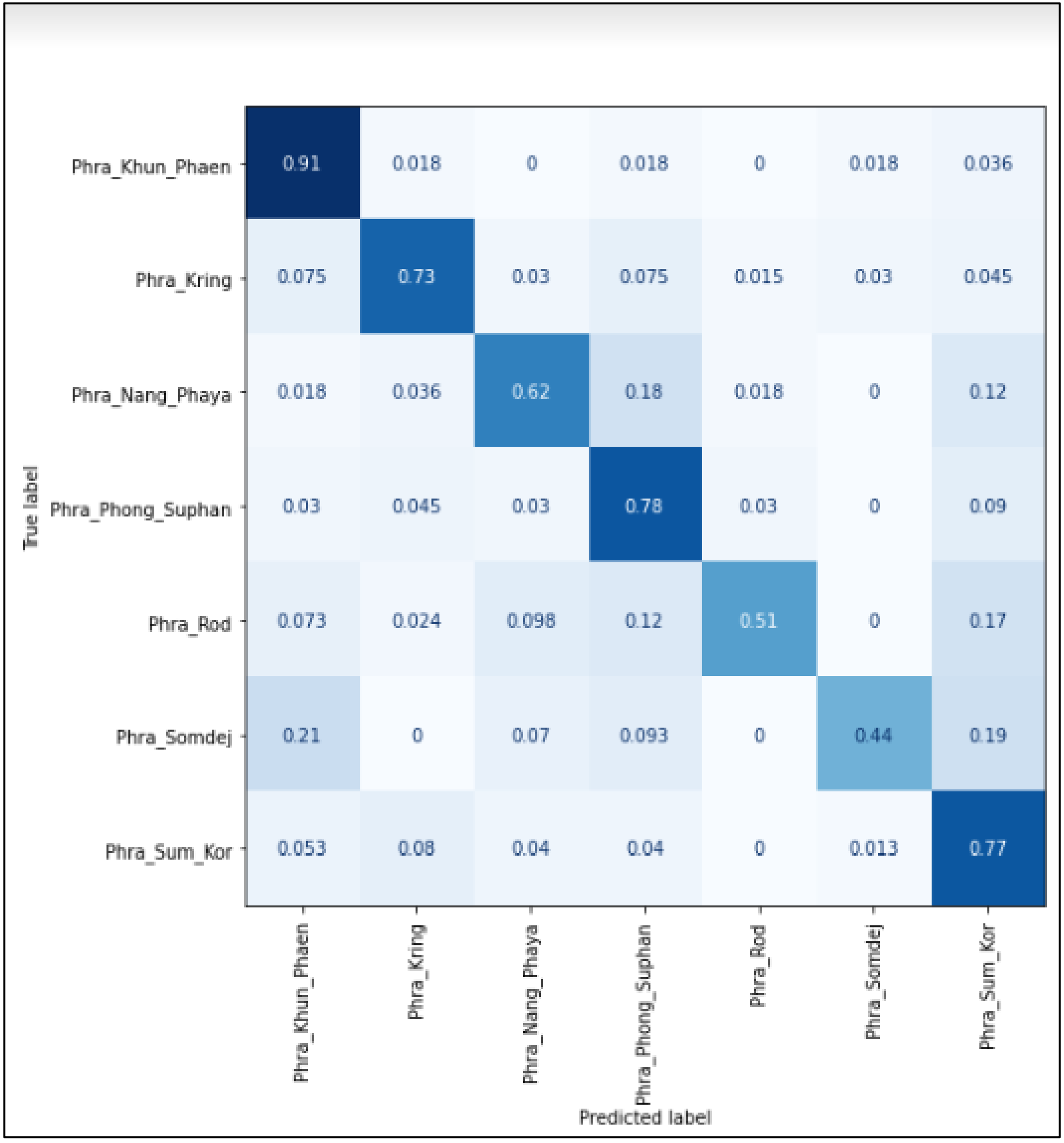
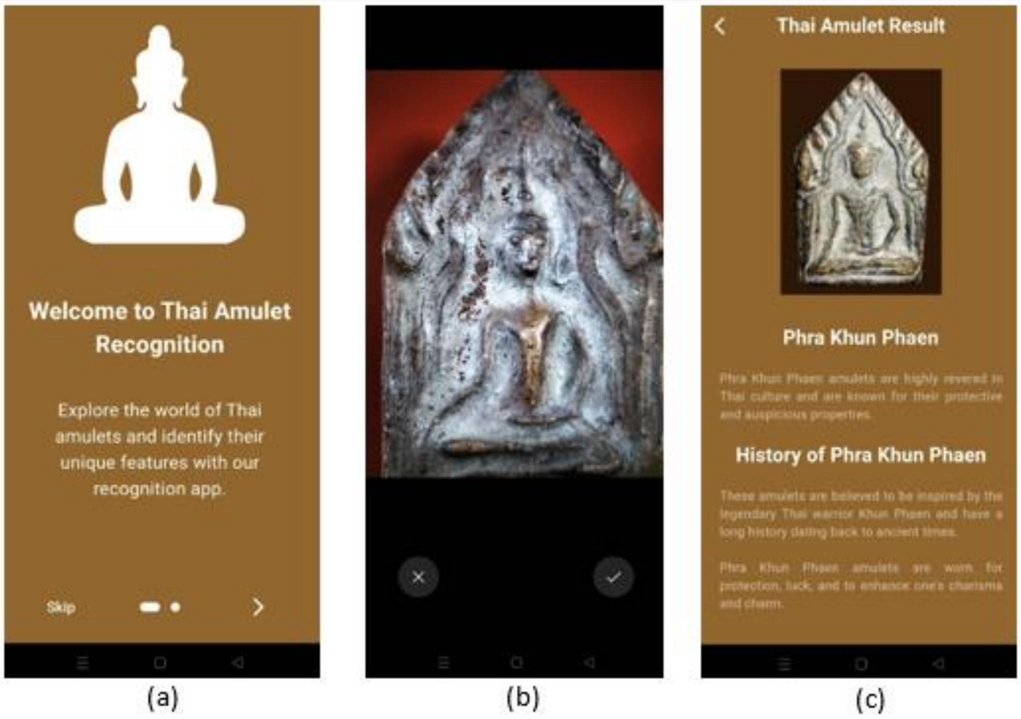


Figure 7 Confusion matrix of the large EVA-02 at 4th fold with pseudo labeling.



True label \ Predicted label	Phra_Khun_Phaen	Phra_Kring	Phra_Nang_Phaya	Phra_Phong_Suphan	Phra_Rod	Phra_Somdej	Phra_Sum_Kor
Phra_Khun_Phaen	0.91	0.018	0	0.018	0	0.018	0.036
Phra_Kring	0.075	0.73	0.03	0.075	0.015	0.03	0.045
Phra_Nang_Phaya	0.018	0.036	0.62	0.18	0.018	0	0.12
Phra_Phong_Suphan	0.03	0.045	0.03	0.78	0.03	0	0.09
Phra_Rod	0.073	0.024	0.098	0.12	0.51	0	0.17
Phra_Somdej	0.21	0	0.07	0.093	0	0.44	0.19
Phra_Sum_Kor	0.053	0.08	0.04	0.04	0	0.013	0.77

Figure 8 Confusion matrix of the ensemble learning of model.



Welcome to Thai Amulet Recognition

Explore the world of Thai amulets and identify their unique features with our recognition app.

Phra Khun Phaen

Phra Khun Phaen amulets are highly revered in Thai culture and are known for their protective and auspicious properties.

History of Phra Khun Phaen

These amulets are believed to be inspired by the legendary Thai warrior Khun Phaen and have a long history dating back to ancient times.

Phra Khun Phaen amulets are worn for protection, luck, and to enhance one's charisma and charm.

Figure 9 Mobile application for Thai amulet recognition: (a) main screen, (b) image upload, (C) amulet information.

List of all the figures.

Table 1 Four variations of eva-02 models.

Table 2 Model performance evaluation.

Table 3 Data Augmentation Model Performance Evaluation.

Table 4 Test-time data augmentation model performance evaluation.

Table 5 Pseudo-labeling model performance evaluation.

Table 6 Ensemble learning model performance evaluation.

EVA-02 Model	**Parameters (Million)**	**FLOPs (Billion)**	**Top-1 Accuracy on ImageNet (%)**
EVA-02 Tiny	10M	0.6	76.2
EVA-02 Small	31M	1.6	77.5
EVA-02 Base	101M	3.3	79.0
EVA-02 Large	304M	7.6	81.4

Table 1 Four variations of eva-02 models.

Model	**Fold**	**Test dataset**	
		Accuracy	**Weighted F1**
Tiny	**1**	0.4877	0.4852
	2	0.4938	0.4896
	3	0.5185	0.5159
	4	0.5247	0.5158
	5	0.5432	0.5362
Small	**1**	0.4506	0.4405
	2	0.4352	0.4270
	3	0.4537	0.4501
	4	0.4599	0.4543
	5	0.5000	0.4981
Base	**1**	0.5556	0.5551
	2	0.5988	0.5896
	3	0.6049	0.5991
	4	0.5895	0.5858
	5	0.5833	0.5710

Model	**Fold**	**Test dataset**	
		Accuracy	**Weighted F1**
Large	**1**	0.5957	0.5811
	2	0.5247	0.5161
	3	0.6235	0.6076
	4	0.5957	0.5906
	5	***0.6235***	***0.6186***

Table 2 Model performance evaluation.

Model	**Fold**	**Test dataset**	
		Accuracy	**Weighted F1**
Tiny	**1**	0.4846	0.4764
	2	0.4753	0.4655
	3	0.4907	0.4894
	4	0.5185	0.5173
	5	0.4784	0.4697
Small	**1**	0.4352	0.4324
	2	0.4475	0.4430
	3	0.5000	0.4948
	4	0.5031	0.4970
	5	0.5154	0.5100
Base	**1**	0.5494	0.5452
	2	0.6142	0.6061

Model	Fold	Test dataset	
		Accuracy	Weighted F1
	3	0.6142	0.6074
	4	0.6204	0.6090
	5	***0.6605***	***0.6535***
Large	**1**	0.6173	0.6118
	2	0.6574	0.6539
	3	0.5926	0.5764
	4	0.5679	0.5552
	5	0.5988	0.5961

Table 3 Data Augmentation Model Performance Evaluation.

Model	Fold	Test dataset without TTA		Test dataset using TTA	
		Accuracy	Weighted F1	Accuracy	Weighted F1
Tiny	**1**	0.4846	0.4764	0.5525	0.5497
	2	0.4753	0.4655	0.5216	0.5119
	3	0.4907	0.4894	0.5710	0.5659
	4	0.5185	0.5173	0.5586	0.5560
	5	0.4784	0.4697	0.5185	0.5079
Small	**1**	0.4352	0.4324	0.4630	0.4524
	2	0.4475	0.4430	0.5216	0.5152
	3	0.5000	0.4948	0.5340	0.5297
	4	0.5031	0.4970	0.5617	0.5548
	5	0.5154	0.5100	0.5741	0.5687

Model	**Fold**	**Test dataset without TTA**		**Test dataset using TTA**	
		Accuracy	**Weighted F1**	**Accuracy**	**Weighted F1**
Base	**1**	0.5494	0.5452	0.6296	0.6275
	2	0.6142	0.6061	0.6605	0.6499
	3	0.6142	0.6074	0.6667	0.6570
	4	0.6204	0.6090	0.6574	0.6502
	5	***0.6605***	***0.6535***	***0.6728***	***0.6681***
Large	**1**	0.6173	0.6118	0.6265	0.6130
	2	0.6574	0.6539	0.6543	0.6512
	3	0.5926	0.5764	0.6080	0.5957
	4	0.5679	0.5552	0.6204	0.6159
	5	0.5988	0.5961	0.5988	0.5986

Table 4 Test-time data augmentation model performance evaluation.

Model	**Fold**	**Pseudo-Labeling**	
		Test dataset using TTA	
		Accuracy	**Weighted F1**
Tiny	**1**	0.5421	0.5314
	2	0.5470	0.5297
	3	0.5891	0.5745
	4	0.5248	0.5147
	5	0.5569	0.5484
Small	**1**	0.5173	0.4997

Model	Fold	Pseudo-Labeling	
		Test dataset using TTA	
		Accuracy	Weighted F1
	2	0.5272	0.5174
	3	0.5569	0.5446
	4	0.5891	0.5741
	5	0.5668	0.5555
Base	**1**	0.6559	0.6524
	2	0.6683	0.6582
	3	0.6634	0.6561
	4	0.6782	0.6703
	5	0.6782	0.6688
Large	**1**	0.6782	0.6701
	2	0.6634	0.6543
	3	0.6931	0.6850
	4	***0.6980***	***0.6912***
	5	0.6287	0.6152

Table 5 Pseudo-labeling model performance evaluation.

Ensemble learning	**Accuracy**	**Weighted F1**
Large – Fold 4	-	
Add Large – Fold 3	***0.7030***	***0.7070***
Add Large –Fold3 and Large – Fold 1	0.6980	0.6914

Table 6 Ensemble learning model performance evaluation.