Cultural Events Classification using Hyper-parameter Optimization of Deep Learning Technique

Feng Zhipeng[1]
School of Culture Creativity
Hangzhou Normal University
Hangzhou, Zhejiang, P.R. China

Hamdan Gani[2]
Graduate School of Computer Systems
STMIK Handayani
Makassar, Indonesia

Abstract—**Through digitization, maintaining and promoting cultural heritage is being strengthened. Concerning this background, this study presents a new Indonesia cultural events dataset and automatic image classification for cultural events. The dataset was developed using the Flickr image platform, and the five cultural events image was collected including the Baliem Festival, Jember Fashion Festival, Nyepi Festival, Pacu Jawi, and Pasola Festival. Further, Convolutional Neural Networks (CNN) was developed for the classification method. A comparison of CNN models (VGG16 and VGG19) using several optimization configurations was performed to get the best model. The results showed that the VGG16 with image augmentation and dropout regularization technique performed best with 94.66% accuracy. This study hoped to support the heritage's digital documentation process and preserve Indonesia's cultural heritage.**

Keywords—Cultural events; convolutional neural network (CNN); very depth convolutional network (VGG); multi-class classification

I. INTRODUCTION

Cultural events are created based on social systems or cultural wisdom passed from one generation to another [1], [2]. They have historical roots, customs, values, and beliefs influenced by many aspects such as region, social, and culture [3]. Hence, each specific ethnic group can be recognized based on their traditional cultural events. Understanding cultural values benefits maintaining cultural heritage [4]. As stated by The United Nations Educational, Scientific, and Cultural Organization (UNESCO) mission, every country is encouraged to approve the World Heritage Convention and ensure the identification, protection, and preservation of its cultural heritage [5]. Therefore, it is essential to sustain the cultural heritage in the face of rapid globalization in line with UNESCO's mission.

With the rapid development of technologies, the effort to preserve cultural heritage is being supported. By implementing digital documentation, cultural heritage can be quickly promoted and maintained. Several benefits of digital cultural heritage can make possible (a) transformation of heritage objects into the digital form [6], (b) quickly access to digital heritage [7], (c) indexation of historical heritage contents and extraction of their information [8], and (d) permanent preservation of digital objects [9]. Concerning all of those benefits, classification methods play an essential part. Classification refers to developing the classification model that will recognize the instances into categories or classes based on

the training data (named supervised learning). The classification model learns from a training dataset and implements the achieved knowledge to classify new data. Therefore, documentation and classification of cultural heritage are essential since each country must save and preserve its cultural heritage.

In Indonesia's context, several issues have been discussed concerning Indonesian cultural heritage, such as foreign country claims to Indonesian regarding the cultural heritage, lack of the inter-generational transfer of knowledge in education, and lack of recognition from the local government [10]. In recent times, several efforts have been implemented to preserve and promote Indonesia's cultural heritage. For example, the Indonesian government strengthened the cultural heritage curricula in education, especially for the young generation. The Indonesian government also promoted the tagline "visit Indonesia" that has the goal to spread Indonesia's cultural events across the globe and targeted to attract visitors to Indonesia [11]. However, the preservation and promotion of cultural heritage are challenging. Indonesia is the world's largest archipelago nation, and it has one of the most varied cultural heritages with more than 300 distinct ethnic groups. Each ethnic group in Indonesia has cultural identities. Because of Indonesian culture's richness, the number of recognizable cultural events is also quite large. Thus, it needs documentation efforts to save and maintain the original cultural heritage of Indonesia.

To the best of our knowledge, no Indonesian cultural events documentation or dataset is available that describes a specific region's cultural events. One study investigated Indonesia's cultural heritage [12]. However, that study did not present Indonesia's cultural events rather than architectural heritage. Consequently, no specific Indonesian cultural events database is publicly available. Also, The Ministry of Tourism and Creative Economic of the Republic of Indonesia (Kemenparekraf RI) struggled to promote cultural events using the website https://www.indonesia.travel/, which primarily focuses on promoting various destinations in Indonesia for domestic and international tourism. However, that website does not promote cultural events.

Therefore, to support cultural heritage preservation, this study aims to present a new Indonesia's cultural events dataset and automatic image recognition for classification cultural events. Several CNN models for multi-class image classification were tested to achieve better accuracy. This paper

has several contributions, specifically: (i) this study presents a new dataset of Indonesia's cultural events, and it has been made openly available to replicate this work (see link on the section availability of data and materials). The dataset would also be advantageous for researchers to consider adding a new image class to achieve the large dataset. (ii) The methodology of CNN with different hyper-parameter techniques is presented, and the results of the practical comparison are shown. Those results can be used as a benchmark for future researchers to improve the multi-class classification algorithm. In general, the proposed dataset and automatic classification system hoped can enhance an essential part of the heritage's digital documentation process and support an effort to preserve the cultural heritage.

This paper is organized as follows. Section 2 explains the materials and methods used in this study. Section 3 describes this study's results, followed by the discussion in Section 4. Finally, Section 5 explains the study's conclusions and future work.

II. RELATED WORKS

Several studies have developed cultural heritage documentation and implemented different methods to classify cultural heritage [13]. For example, in the study of architectural heritage, the authors proposed the image dataset of more than 10.000 images classified into ten classes, i.e., different architectural heritage types such as columns, domes, gargoyles, and vault [14]. This study compared the deep learning algorithms to categorize cultural heritage images. Specifically, several convolutional neural networks (CNN) were implemented, AlexNet, Inception V3, ResNet, and Inception-ResNet-v2. They achieved good accuracy on the complete training data; ResNet obtained a higher accuracy. In the fine-tuning configuration, the best accuracy was achieved for the Inception-ResNet-v2.

An early study [15] was investigated on a dataset containing 1.227 images dataset of 12 cultural heritage memorials and Pisa landmarks. The image classification was compared by using the k-nearest neighbor (k-NN) classification with different types of the feature extraction, namely Scale Invariant Feature Transform (SIFT), Speed up Robust Feature (SURF), Oriented FAST and Rotated BRIEF (ORB), and Binary Robust Invariant Scalable Keypoints (BRISK). They obtained that the local feature-based classifier achieved good accuracy; on the other hand, the best performance was reached using SIFT concerning the features.

Another study proposed 100 cultural heritage of wall painting images of reflected light for the image classification task [14]. The image dataset was involved in the reflected image, such as visible light, ultraviolet light, infrared light, and visible fluorescence. The authors used Dense SURF, spectral information, and a support vector machine algorithm. They concluded that the higher accuracy was the image in reflected ultraviolet light. Simultaneously, the dense integrating SURF and spectral information obtained the best accuracy than executing them individually.

The previous study [14] proposed an Indonesian cultural heritage dataset for image, audio, and video classification. The dataset includes 100 images, 100 audios, 100 videos, and 100 text files separated into five classes. The deep learning algorithms, Convolutional Neural Network (CNN) and Recurrent Neural Networks (RNN), were executed to classify Indonesian architectural heritage. The CNN was executed for image, audio, and video classification, while RNN was executed to classify text. The results revealed that RNN obtained the best performance regarding the accuracy, classifying 92% of the text data. Concerning CNN, the higher accuracy (76% each) reached for image and video classification, and audio acquired 57% accuracy.

In the study of archaeological sites [16], the author collected the cultural heritage dataset that included 150 images and categorized them into three classes (50 images each category): archaeological sites, frescoes, and monasteries. This study aimed to classify images using several decision tree algorithms such as J48, Hoeffding tree, random tree, and random forest. The authors determined that the random forest algorithm achieved the best performance.

Although several studies have shown advances in researching the cultural heritage for image classification, mainly those research only focused on the architectural building classification as explained in the above. A study involving cultural events or ceremonies was not still profoundly studied. Thus, studying cultural events are necessary to support the way to protect and promote cultural heritage for future generations.

III. MATERIAL AND METHODS

A. Materials

The scrapping image method was performed to collect Indonesia's cultural events using Python programming. The Flickr image service was used for collecting the image datasets. The Flickr images under Attribution-Noncommercial License were collected. Five cultural events were collected: Baliem Valley Cultural Festival, Jember Fashion Carnival, Nyepi the Day of Silence, Pacu Jawi, and Pasola Festival. The dataset consisted of 1.500 images and was divided into 300 images in each class. The image samples from the five classes are shown in Fig. 1. The images represent (a) Pasola, (b) Pacu Jawi, (c) Nyepi, (d) Jember Fashion, and (e) Baliem.

B. Methods

The experiments were performed in Python v.3.7 environments, and the CNN (VGG) model was developed with the Keras library. The experiments were performed under Windows 10 platform, 16 GB Graphical Processing Unit (GPU), 256 GB SSD storage, Core i7 processor 1.80 GHz, and 8 GB of RAM. All images were converted into 200 x 300 image pixels. The distribution of image classes is shown in Fig. 2.
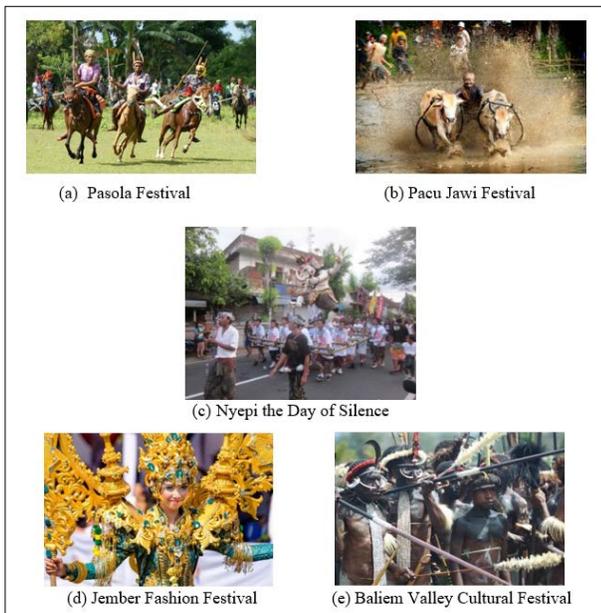
Fig. 1. Examples of Five Image Classes: (a) Pasola Festival, (b) Pacu Jawi Festival, (c) Nyepi the Day of Silence, (d) Jember Fashion Festival, and (e) Baliem Valley Cultural Festival.
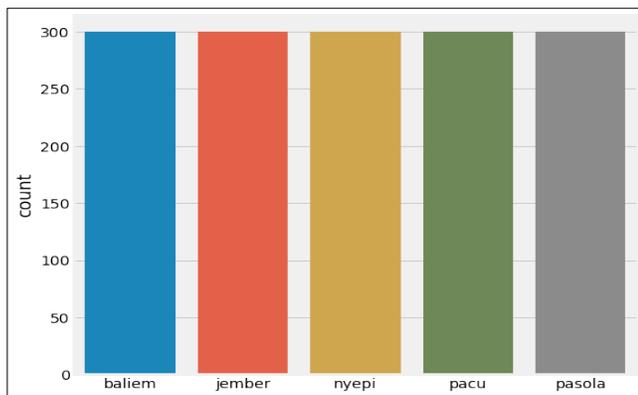

Fig. 2. Distribution of different Classes of the Dataset.

1) Convolutional Neural Network (CNN): CNN is usually applied for computer vision, as it captures images as inputs and extracts features from the images. The CNN typically contains convolutional layers (each involving several kernel sizes and filters). The convolutional layer is along with the pooling layer, decreasing data dimensionality. There are two types of pooling: max-pooling and average pooling. Max-pooling uses the maximum value from the image related to the kernel size, while average pooling utilizes the average of all the values. Once image processing is accomplished across these layers, the features from a two-dimensional matrix are converted into a vector with a flatten layer, and the achieved output is transmitted to the fully connected layer or dense layer [17].

2) Very Depth Convolutional Network (VGG): VGG's name belonged to their lab's name, the Visual Geometry Group at Oxford, and the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014 competition winner

[18]. This architecture was designed for deep convolutional network learning. The VGG architecture is created by 3 x 3 Convolutional and MaxPooling layers, with a fully connected block at the end. In the original paper, VGG architecture has shown the depth network's effect on its performance in the large-scale image database. VGG used small (3 x 3) convolution filters in the complete architecture and presented a considerable improvement in the configurations using the depth to 16-19 weight layers [18]. Another advantage of VGG architecture is that they used many filters. The number of filters grows with the depth of the model. They start at 64 and continually increase across 128, 256, and 512 filters at the end of the model's feature extraction. VGG was named for the number of layers: the VGG16 for 16 layers and the VGG19 for 19 learned layers. The architecture for VGG16 and VGG19 is presented in Fig. 3(a) and (b). The summary of the architecture of VGG models as follows: (i) apply small convolutional filters, e.g., 3 x 3 and 1 x 1, (ii) apply max pooling with a size of 2 x 2, (iii) the stacking of convolutional layers concurrently before applying a pooling layer to identify a block, (iv) dramatic reiteration of the convolutional-pooling block pattern (v) development of intense models (16 and 19 layers). The architecture of VGG16 and VGG19 are depicted in Fig. 3. In this work, the VGG 16 and VGG19 models were used for the experiment.
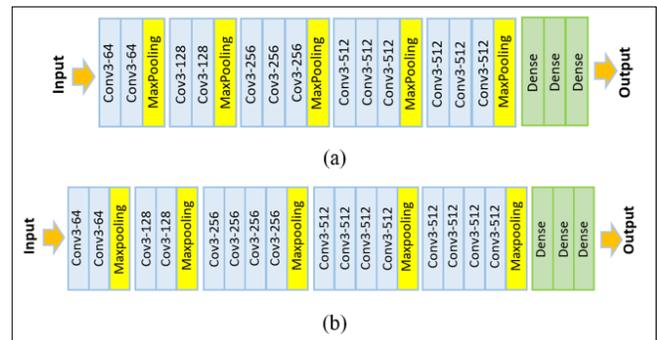

Fig. 3. The Architecture of VGG16 (a) and VGG19 (b).

3) Image augmentation optimization: The image augmentation works to create a new and unique training example. Image augmentation transforms the images' versions in the training dataset corresponding to the same class as the initial image [19]. That transformation involves several image manipulation processes, such as shifts, flips, and zooms. The current deep learning algorithms, such as CNN, can quickly learn the image features. The augmentation technique can improve the algorithm's learning process, and it is usually implemented to the training dataset, not to the test or validation dataset. In this work, the Keras deep learning library was used. That library offers several image augmentations functions via the ImageDataGenerator class. Several augmentation functions were used, such as the horizontal and vertical flip of image (randomly flip training image), rotation (randomly rotate images in the range of degrees), brightness, and zoom (randomly zoom image).

4) Dropout regularization: Dropout is an effective technique to maintain the neural network from overfitting during the training [20]. Dropout is applied by only saving a neuron active with a certain probability *p* and locating it to 0 otherwise. This condition pushes the network not to learn redundant information [20]. Consequently, this method significantly decreases overfitting and presents significant neural network improvements in supervised learning [21]. In this study, the Keras deep learning library was used for importing the dropout regularization class in the VGG model. The drop out was set on 0.5.

5) Transfer learning: In machine learning, transfer learning refers typically to a method where a model trained on a specific problem is implemented in other problems, which is a related problem [22]. Transfer learning has the advantages of reducing the training time for an algorithm model and can produce lower generalization errors. The weights in re-used or latest layers can be used as the initial point for the training process and implemented to answer the new problem. Transfer learning can be helpful when the first associated problem has many labeled data. Several high-performing models have been created for image classification on the annual ILSVRC, such as ZFNet, VGG, GoogleNet, and ResNet [17]. This competition has produced several innovative models in CNN architecture and can be implemented to transfer learning in computer vision applications. Those models have learned over 1.000.000 images for 1.000 classes and achieved state-of-the-art performance. In this study, all VGG model was developed using transfer learning and directly downloaded using the Keras library function into our python environment.

6) Configuration of VGG model: The VGG architecture included 19 weight layers, 16 convolutional layers, and 3 fully connected layers [18]. The channel number of convolutional layers starts from 64 and increases by a factor of 2 after each max-pooling layer until obtaining 512. Finally, SoftMax activation was used in the dense layer. The architecture of VGG16 is similar to VGG19; only the difference is the total number of layers (16 for VGG16). A Method for Stochastic Optimization (Adam) was implemented as an optimizer. In order to avoid overfitting, the early stopping function was implemented with configuration patience 5 and verbose 1. The number of epochs was adjusted to 30. The VGG model used 80% for training data and 20% for validation. The 150 x 150 image pixels were used as an input image for VGG models. The detailed configuration of all VGG models is presented in Table I.

7) Model evaluation: After classification was performed, several evaluation metrics were performed. Specifically, the accuracy, precision, recall, F1 score, and ROC area were used to choose the best model. The accuracy is the percentage of correct instances classified by the algorithm. Precision is the number of instances that fit the respected class and the calculated instances categorized to that class, while recall or sensitivity explains the true positive rate of prediction. The F1 score or F-measure explains the classification accuracy regarding the average precision and recall values. The F1 score values closer to 1 show a better classification accuracy. The measurement methods are calculated as follows in Eq. (1), (2), (3), and (4):

$$Accuracy = \frac{Correctly\ classified\ data}{Total\ data} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$Precission = \frac{TP}{TP+FP} \tag{3}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

Where TP is a true positive, FN is a false negative, and FP is a false positive. Finally, the region under the ROC curve shows the proportion of true positives and false positives. This value must be close to 1, indicating a perfect prediction, as the values under 0.5 imply a random guess [23].

TABLE I.   CONFIGURATION OF VGG ARCHITECTURE MODELS

	Activation Function	Number of Epoch	Weight	Batch Size	Loss Function	Optimizer	Image Augmentation	Drop Out	Early Stopping	Transfer Learning
VGG16 + Augmentation + Drop Out	Softmax	30	ImageNet	32	Categorical Cross entropy	Adam	Yes	Yes (0.5)	Yes (Patience 5, verbose 1)	Yes
VGG16 + Augmentation	Softmax	30	ImageNet	32	Categorical Cross entropy	Adam	Yes	Yes	Yes (Patience 5, verbose 1)	Yes
VGG16 Baseline	Softmax	30	ImageNet	32	Categorical Cross entropy	Adam	No	No	Yes (Patience 5, verbose 1)	Yes
VGG19 + Augmentation + Drop Out	Softmax	30	ImageNet	32	Categorical Cross entropy	Adam	Yes	Yes (0.5)	Yes (Patience 5, verbose 1)	Yes
VGG19 + Augmentation	Softmax	30	ImageNet	32	Categorical Cross entropy	Adam	Yes	Yes	Yes (Patience 5, verbose 1)	Yes
VGG19 Baseline	Softmax	30	ImageNet	32	Categorical Cross entropy	Adam	No	No	Yes (Patience 5, verbose 1)	Yes
Hyper-Parameter	Activation Function	Number of Epoch	Weight	Batch Size	Loss Function	Optimizer	Image Augmentation	Drop Out	Early Stopping	Transfer Learning

IV. Experimental Results

This section explains the analysis of the obtained algorithm's performance on the cultural events dataset. Several algorithms with hyper-parameter were evaluated using the accuracy, precision, recall, F1 score, and ROC area. The performance of the algorithm can be seen in Fig. 4.

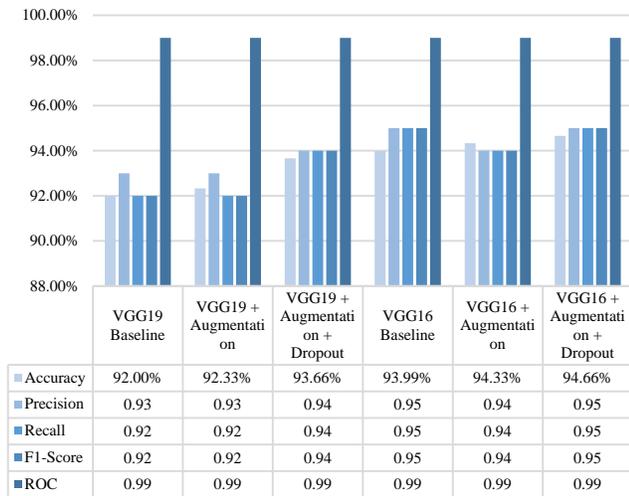|	VGG19 Baseline	VGG19 + Augmentation	VGG19 + Augmentation + Dropout	VGG16 Baseline	VGG16 + Augmentation	VGG16 + Augmentation + Dropout
Accuracy	92.00%	92.33%	93.66%	93.99%	94.33%	94.66%
Precision	0.93	0.93	0.94	0.95	0.94	0.95
Recall	0.92	0.92	0.94	0.95	0.94	0.95
F1-Score	0.92	0.92	0.94	0.95	0.94	0.95
ROC	0.99	0.99	0.99	0.99	0.99	0.99

Fig. 4. The Performance of each Algorithm

Based on the results, it showed that VGG16 performed better than VGG19. The combination of "VGG16 + Augmentation + Dropout" performed the best with 94.66% of correctly classified images, followed by "VGG16 + Augmentation" with 94.33%, and "VGG16 Baseline" with 93.99%. On the other hand, the combination of "VGG19 + Augmentation + Dropout" performed with 93.66% of correctly classified images, followed by "VGG19 + Augmentation" with 92.33%, and "VGG19 Baseline" with 92.00%. It also showed that the precision, recall, F1-score, and ROC area were better for the "VGG16 + Augmentation + Dropout" than the other VGG configurations.

After implementing the hyper-parameter optimization, the algorithms performed better than the baseline. The model performance showed a very slight increase in the model's mean accuracy, 92.33% in "VGG19 + Augmentation" compared to 92.00% with the VGG19 baseline model. It also confirmed that dropout regularization performed well. There was a very slight rise in the model's accuracy, 93.66% in "VGG19 + Augmentation + Dropout" compared to 92.33% with the "VGG19 + Augmentation".

A similar improvement is presented in the VGG16 model. The estimated performance of the "VGG + Augmentation + Dropout" model indicated a possible increase in performance compared to the baseline from 93.99% to 94.66%. This study's findings confirm that a hyper-parameter configuration (image augmentation and dropout regularization) can improve the models in line with previous works [24].

In order to detect which of the classes classified correctly, the confusion matrices were performed. The results of the confusion matrices are presented in Table II. The diagonal numbers show the correctly classified images (blue background), while other numbers in rows describe the misclassifications of images. It showed that the "VGG19 Baseline" most accurately classified the Pacu Jawi and Pasola images, while the "VGG19 + Augmentation" and "VGG19 + Augmentation + Dropout" most correctly classified the Jember and Pacu Jawi images. The "VGG16 Baseline" and "VGG16 + Augmentation" accurately classified the Pacu Jawi, Pasola, and Jember images, while the "VGG16 + Augmentation + Dropout" most correctly classified the Pacu Jawi images.

TABLE II. THE CONFUSION MATRICES FOR EACH ALGORITHM

Algorithm	Baliem	Jember	Nyepi	Pacu Jawi	Pasola	Classified as
VGG19 Baseline	52	0	3	0	1	Baliem
	4	56	5	0	0	Jember
	3	0	44	0	2	Nyepi
	0	0	0	65	1	Pacu Jawi
	1	0	3	1	59	Pasola
VGG19 + Augmentation	48	0	5	2	1	Baliem
	3	58	4	0	0	Jember
	2	0	47	0	0	Nyepi
	0	0	0	66	0	Pacu Jawi
	1	0	3	2	58	Pasola
VGG19 + Augmentation + Dropout	52	0	3	1	0	Baliem
	3	59	3	0	0	Jember
	3	0	46	0	0	Nyepi
	0	0	0	66	0	Pacu Jawi
	1	0	3	2	58	Pasola
VGG16 Baseline	48	1	4	2	1	Baliem
	1	62	2	0	0	Jember
	1	0	45	0	3	Nyepi
	1	0	0	65	0	Pacu Jawi
	0	1	1	0	62	Pasola
VGG16 + Augmentation	52	0	2	1	1	Baliem
	1	62	2	0	0	Jember
	4	0	43	0	2	Nyepi
	2	0	0	64	0	Pacu Jawi
	1	1	0	0	62	Pasola
VGG19 + Augmentation + Dropout	55	0	0	1	0	Baliem
	2	62	1	0	0	Jember
	5	0	41	0	3	Nyepi
	2	0	0	64	0	Pacu Jawi
	0	1	1	0	62	Pasola

In terms of misclassification, "VGG19 + Augmentation + Dropout" mostly misclassified the Nyepi images, while the rest of the other algorithms misclassified the Nyepi and Baliem images. Based on the results, Pacu Jawi images were most correctly classified among all algorithms, while Nyepi images were most misclassified. Finally, to fully show these algorithms' performance, the accuracy and loss model

presented in Fig. 5. As for simplicity only "VGG16 + Augmentation + Dropout" and "VGG19 + Augmentation + Dropout" models are presented.
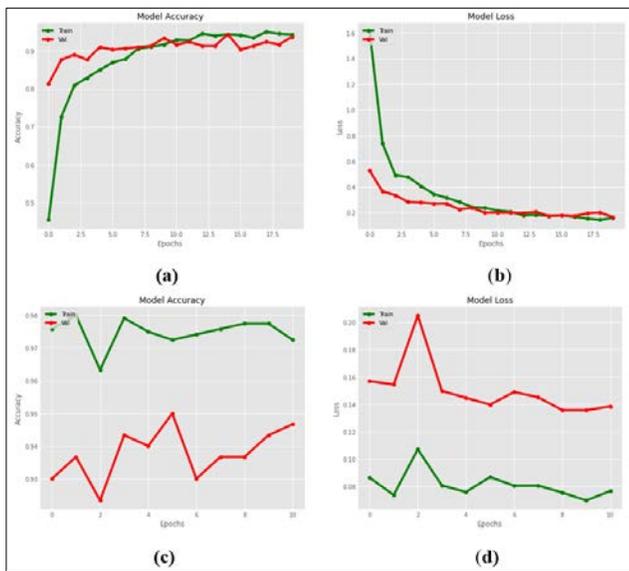

Fig. 5. (a) Accuracy for VGG19, (b) Loss Results for VGG19, (c) Accuracy for VGG16, and (d) Loss Results for VGG16.

As observed in Fig. 5, the VGG19 model performed well, with an accuracy of 93.66 % (a). Both training and validation loss had a reducing trend achieving the value of 0.20 for training and validation in the last epoch (b). On the other hand, VGG16 with 94.66% accuracy seemed to perform well and tended to increase to fit (train and validation line) further after the last epoch (c). The model showed a decreasing trend from the value of 0.16 to 0.14 (d) concerning the validation loss. The model's epoch finished on 18 epochs for VGG19 and 10 epochs for VGG16, regarding no further improvement on accuracy and avoiding the overfitting.

In a survey study, image augmentation has been proved to improve the performance of the models and enhance the limitation of datasets to take advantage of significant data capabilities [19]. In agreement with our results from performance accuracy in Fig. 4, it displayed that model using the image augmentation technique (in both VGG19 and VGG16 models) was better than the baseline model. In their experimental study, Nandini et al. [25] compared the dropout technique for image classification using several algorithms using three different image datasets. They found that the dropout regularization technique accomplished the best results in image classification compared with other classification algorithms. This finding is also in line with our results in Fig. 4 that display the dropout technique's performance, and image augmentation performs best compared to other configuration models.

V. DISCUSSION

This study presents a new Indonesia cultural events dataset and automatic image recognition for classification cultural events. Several findings obtained from this study: (i) the convolutional neural network, with VGG16 architecture,

performed well when classifying images compared to the other models (despite that other algorithms accuracy were relatively good accuracy); (ii) all classifiers performed better after adding hyper-parameter configuration (image augmentation and dropout regularization); and (iii) algorithms most correctly classified the Jember and Pacu Jawi images, while Nyepi images most frequently misclassified.

As shown in the previous section, the "VGG16 + Augmentation + Dropout" performed well in all performance measures. It obtained the highest classification accuracy compared to the other algorithms. Also, it performed the best in the other evaluation measures, such as precision, recall, F1 score, and the ROC. Furthermore, the CNN with VGG architecture model showed excellent classification accuracy. Specifically, CNN mainly performs better than other non-neural network algorithms applied for image classification tasks [17], [25]–[27].

Moreover, CNN is suitable for a large dataset. Regarding the overfitting problems, hyper-parameter tuning has been implemented with an early stopping function to avoid the overfitting problem. The Keras library's early stopping function was used to stop training when the training accuracy gets a specific threshold. Hence, the optimal model weights can be achieved and save computation time and power. Although CNN is computationally intensive, it can achieve good performance using several hyper-parameter configurations. Thus, based on this study's results, hyper-parameter configurations are a promising way to improve the algorithm's performance, especially CNN with VGG architecture.

This study could develop cultural events image classification models that could be more efficient and computationally solid. As multi-class image classification usually includes many images and needs substantial computational resources, it must be operated correctly and reliable. Therefore, developing optimized models and reliable classification methods is essential for current and future studies.

As the limitation of this study, the proposed dataset used a balanced class distribution. Thus, our proposed VGG configuration needs to test in an imbalanced classification problem to show the validity of our proposed configuration models.

As future work, this study plans to enhance the proposed image datasets, including different cultural events images from different Indonesia regions. Also, generally, the image classification generally includes large data sets, further work needs to develop large datasets of Indonesia culture's image dataset.

VI. CONCLUSION

This study presents a new Indonesia cultural events dataset and automatic image recognition for classification cultural events. This study compared several configurations of hyper-parameter configurations of CNN for multi-class image classification. In particular, CNN architecture, such as VGG19 and VGG16, were tested before and after the hyper-parameter configuration. Overall, the VGG19 and VGG16 achieved a good performance, but considering the hyper-parameter

optimization, the VGG16 using image augmentation and dropout regularization achieved the best classification with 94.66% accuracy. Other algorithms achieved less than 94.33% accuracy. Despite that, the accuracy of other algorithms was relatively good. This study confirms that CNN with VGG architecture is a better choice for multi-class image classification, and they offer good performance for classification tasks. Finally, this study's findings hoped to support the heritage's digital documentation process and maintain cultural heritage.

ACKNOWLEDGMENT

AVAILABILITY OF DATA AND MATERIALS

https://www.dropbox.com/l/scl/AAAO7Es-r_8y3Xtv0ZFkFjh4hjqTl-1GViQ

REFERENCES

[1] J. M. Hernández-Mogollón, P. A. Duarte, and J. A. Folgado-Fernández, "The contribution of cultural events to the formation of the cognitive and affective images of a tourist destination," Journal of Destination Marketing & Management, vol. 8, pp. 170–178, 2018, doi: https://doi.org/10.1016/j.jdmm.2017.03.004.

[2] B. Djibat, S. Deni, and Z. Saing, "The culture of Makayaklo in North Maluku Society: Teaching the values of building solidarity and social integration," International Journal of Critical Cultural Studies, vol. 17, no. 1, pp. 43–54, 2019, doi: 10.18848/2327-0055/CGP/v17i01/43-54.

[3] M. R. Abdulla, "Culture, Religion, and Freedom of Religion or Belief," Review of Faith and International Affairs, vol. 16, no. 4, pp. 102–115, 2018, doi: 10.1080/15570274.2018.1535033.

[4] T. Lussetyowati, "Preservation and Conservation through Cultural Heritage Tourism. Case Study: Musi Riverside Palembang," Procedia - Social and Behavioral Sciences, vol. 184, no. August 2014, pp. 401–406, 2015, doi: 10.1016/j.sbspro.2015.05.109.

[5] P. Alasuutari and A. Kangas, "The global spread of the concept of cultural policy," Poetics, vol. 82, no. March, p. 101445, 2020, doi: 10.1016/j.poetic.2020.101445.

[6] N. A. Khan, S. M. Shafi, and H. Ahangar, "Digitization of Cultural Heritage," Journal of Cases on Information Technology, vol. 20, no. 4, pp. 1–16, Oct. 2018, doi: 10.4018/JCIT.2018100101.

[7] V. Borissova, "Cultural heritage digitization and related intellectual property issues," Journal of Cultural Heritage, vol. 34, pp. 145–150, 2018, doi: https://doi.org/10.1016/j.culher.2018.04.023.

[8] G. Colavizza, M. Ehrmann, and F. Bortoluzzi, "Index-Driven Digitization and Indexation of Historical Archives," Frontiers in Digital Humanities, vol. 6, p. 4, 2019, doi: 10.3389/fdigh.2019.00004.

[9] Z. Ognjanović, B. Marinković, M. Šegan-Radonjić, and D. Masliković, "Cultural heritage digitization in Serbia: Standards, policies, and case studies," Sustainability (Switzerland), vol. 11, no. 14, 2019, doi: 10.3390/su11143788.

[10] R. Asfina and R. Ovilia, "Be Proud of Indonesian Cultural Heritage Richness and Be Alert of Its Preservation Efforts in the Global World," Humanus, vol. 15, no. 2, p. 195, 2017, doi: 10.24036/jh.v15i2.6428.

[11] N. Amanah, I. Atjo, and M. Dewi, "Indonesia Tourism Communication Strategy in Event ITB Berlin 2016," The 4th Conference on Communication, Culture and Media Studies Indonesia, no. October, pp. 10–11, 2017.

[12] R. A. Kambau, Z. A. Hasibuan, and M. O. Pratama, "Classification for multiformat object of cultural heritage using deep learning," Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018, pp. 1–7, 2018, doi: 10.1109/IAC.2018.8780557.

[13] R. Janković, "Machine learning models for cultural heritage image classification: Comparison based on attribute selection," Information (Switzerland), vol. 11, no. 1, 2020, doi: 10.3390/info11010012.

[14] J. Llamas, P. M. Lerones, R. Medina, E. Zalama, and J. Gómez-García-Bermejo, "Classification of architectural heritage images using deep learning techniques," Applied Sciences (Switzerland), vol. 7, no. 10, pp. 1–26, 2017, doi: 10.3390/app7100992.

[15] G. Amato, F. Falchi, and C. Gennaro, "Fast image classification for monument recognition," Journal on Computing and Cultural Heritage, vol. 8, no. 4, 2015, doi: 10.1145/2724727.

[16] R. Janković, "Classifying cultural heritage images by using decision tree classifiers in WEKA," CEUR Workshop Proceedings, vol. 2320, pp. 119–127, 2019.

[17] J. Gu et al., "Recent advances in convolutional neural networks," Pattern Recognition, vol. 77, pp. 354–377, 2018, doi: 10.1016/j.patcog.2017.10.013.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, pp. 1–14, 2015.

[19] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," Journal of Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[20] P. Baldi and P. Sadowski, "The dropout learning algorithm," Artificial Intelligence, vol. 210, no. 1, pp. 78–122, 2014, doi: 10.1016/j.artint.2014.02.004.

[21] G. S. Nandin, A. P. S. Kumar, and C. K, "Dropout Technique for Image Classification Based On Extreme Learning Machine," Global Transitions Proceedings, pp. 0–7, 2021, doi: 10.1016/j.gltp.2021.01.015.

[22] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, A survey of transfer learning, vol. 3, no. 1. Springer International Publishing, 2016.

[23] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.

[24] D. Zhao, G. Yu, P. Xu, and M. Luo, "Equivalence between dropout and data augmentation: A mathematical check," Neural Networks, vol. 115, pp. 82–89, 2019, doi: 10.1016/j.neunet.2019.03.013.

[25] M. P. Véstias, "A survey of convolutional neural networks on edge with reconfigurable computing," Algorithms, vol. 12, no. 8, 2019, doi: 10.3390/a12080154.

[26] R. D. Yogaswara and A. D. Wibawa, "Comparison of Supervised Learning Image Classification Algorithms for Food and Non-Food Objects," in 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM), Nov. 2018, pp. 317–324, doi: 10.1109/CENIM.2018.8711387.

[27] F. Siraj, M. A. Salahuddin, and S. A. M. Yusof, "Digital Image Classification for Malaysian Blooming Flower," in 2010 Second International Conference on Computational Intelligence, Modelling and Simulation, Sep. 2010, pp. 33–38, doi: 10.1109/CIMSiM.2010.92.