The effect of Gaussian filter and data preprocessing on the classification of Punakawan puppet images with the convolutional neural network algorithm

Kusrini[1], Muhammad Resa Arif Yudianto[1], Hanif Al Fatta[2]
[1]Magister of Informatics, Universitas AMIKOM Yogyakarta, Sleman, Indonesia
[2]Faculty of Computer Science, Universitas AMIKOM Yogyakarta, Sleman, Indonesia

Article Info

Article history:

Received Jul 1, 2021
Revised Mar 19, 2022
Accepted Mar 30, 2022

Keywords:

Computer vision
Gaussian filter
Shadow puppets
VGG16

ABSTRACT

Nowadays, many algorithms are introduced, and some researchers focused their research on the utilization of convolutional neural network (CNN). CNN algorithm is equipped with various learning architectures, enabling researchers to choose the most effective architecture for classification. However, this research suggested that to increase the accuracy of the classification, preprocessing mechanism is another significant factor to be considered too. This study utilized Gaussian filter for preprocessing mechanism and VGG16 for learning architecture. The Gaussian filter was combined with different preprocessing mechanism applied on the selected dataset, and the measurement of the accuracy as the result of the utilization of the VGG16 learning architecture was acquired. The study found that the utilization of using contrast limited adaptive histogram equalization (CLAHE) + red green blue (RGB) + Gaussian filter and thresholding images showed the highest accuracy, 98.75%. Furthermore, another significant finding is that the Gaussian filter was able to increase the accuracy on RGB images, however the accuracy decreased for green channel images. Finally, the use of CLAHE for dataset preprocessing increased the accuracy dealing with the green channel images.

Corresponding Author:

Kusrini
Magister of Informatics, Universitas AMIKOM Yogyakarta
Ringroad Utara, Condong Catur, Depok, Sleman, Yogyakarta, Indonesia
Email: kusrini@amikom.ac.id

1. INTRODUCTION

The application of computer vision has penetrated various sectors to assist humans in facilitating their needs. There are many machine learning or deep learning algorithms that can be used in the learning process of the computer vision. For example, Shin and Balasingham [1] compared machine learning-based image classification methods with deep learning algorithm. That research denoted that convolutional neural network (CNN) algorithm as the example of deep learning showed high value of accuracy 92.08%. This high accuracy is the reason why the CNN algorithm is widely used.

The use of CNN algorithm also found in several studies, for example Varshni *et al.* [2] conducted detection of pneumonia using the CNN algorithm to classify normal and pneumonia X-ray images. Umri *et al.* [3] and Morgan *et al.* [4] used CNN to detect coronavirus disease (COVID-19) in chest X-Ray images. Agastya *et al.* [5] classified pornographic images using CNN. In this study, the CNN algorithm was able to provide satisfactory classification results in classifying X-ray images affected by pneumonia and normal or healthy chest images. Another study suggested by Yadav and Jadhav [6] also discussed the implementation of the deep learning-based CNN method in classifying diseases. This classification process is very helpful for experts in diagnosing patient diseases. Besides, implementing this classification system can save time and effort in classifying diseases. From several previous studies, it clearly showed that the CNN algorithm is widely applied in the health sector.

On the other hand, this algorithm is still rarely used in cultural field. Cultural area required the implementation of this CNN algorithm in some ways to support the preservation of the cultural heritage, as the traditional culture is disappearing day by day. In Java, Indonesia, there is a magnificent cultural heritage called "Wayang Kulit", a traditional shadow puppet show, facing the end of their existence. Wayang has various types, one popular type is Wayang Kulit, made of leather or buffalo skin, chiseled, and colored with magnificent pattern and animated with stick made of buffalo horn. Wayang Kulit introduced many famous characters for the Javanese, one of them are the Punakawan character [7]. Punakawan characters consisted of Semar, Petruk, Gareng, and Bagong. The four characters are cheerful and funny, but always provide valuable life lessons through puppet scenes. Wayang Kulit performance is disappearing from the public show due to several condition: minimum support from the government, the lack of community nurturing this culture and the actors and artist from this Wayang performance decreased rapidly.

Several studies have managed to utilize deep learning-based recognition aimed to preserve the existence of this shadow puppet culture. Muhathir suggested Wayang classifications using the multi-layer perceptron (MLP) and gray level co-occurrence matrix (GLCM) algorithms that were only able to produce an accuracy value of 73.4% [8]. Sudiatmika *et al.* [9] classified Wayang using the CNN algorithm, employing the AlexNet and VGG16 architectures. The highest accuracy results are obtained on the VGG16 architecture with an accuracy value of 98%. From the two similar studies on the puppet dataset, it was found that the CNN algorithm was better than the machine learning algorithm. Furthermore, it turns out that the CNN architecture has an effect on the value of accuracy. That research indicated that VGG16 was better than the AlexNet architecture.

Cheng *et al.* [10] also compared deep learning algorithms CNN with some other machine learning algorithms on the emotional signal dataset. From this research, it was found that CNN outperformed the other two architectures with an accuracy value of 83.45%. The resulting level of accuracy is not necessarily due to the use of algorithms alone. However, it is also influenced by the treatment of the dataset before the learning process is carried out. Apollonio *et al.* [11] classified retinopathy with the CNN algorithm. Accuracy increased after adding the process of adding quality levels and images using the contrast limited adaptive histogram equalization (CLAHE) method, up to 86.76% in the number of two class datasets [11].

The various types of available CNN learning architectures encouraged researcher to know the advantages of each architecture by comparing them on the same dataset. Chowanda *et al.* [12] compared three CNN architectures on a dataset of places/landmark in Indonesia. From many experimental scenarios that have been carried out, the results show that the VGG16 algorithm is superior to VGG19 and GoogleNet with an accuracy value of 92%. This study implied that available CNN architectures cannot be used directly, because the number of classes in the fully connected layer in these architectures is 1,000 classes. Therefore, to get optimal accuracy, it is necessary to do an adjustment process, or what is often called the fine-tuning process. This fine-tuning process was applied by many researchers, for example, Apollonio *et al.* [11] applied transfer learning with VGG16 architecture to animal dataset with a total of two classes. The resulting accuracy in this study increased after the fine-tuning process was carried out, from 72.40% to 79.20%. Another method to increase the accuracy of the classification in addition to the adjustment or selection of learning architecture, is by implementing preprocessing mechanism. These methods regarded that the treatment of the dataset before the training process occurred is very significant factor that can increase the accuracy. Preprocessing is simply a process intended to make better or more suitable dataset as an input for process training. There are several treatments that can be applied on the data before the learning process is carried out, one of them is the use of filters such as Kumar and Sodhi [13]. They utilized filter by comparing a filtration method that can reduce the noise from an image.

Another factor that affected the accuracy value is the training parameter. Guo *et al.* [14] explained that in addition to architecture that affects the accuracy value, training parameters also have an effect on the level of learning. The more repetitions in the training process, the better the learning level value. With the increase in the value of the learning rate, the error value, or the error rate of the system in recognizing an object will decrease simultaneously. Algorithms also have an important role in producing high accuracy values, even though they use the same dataset, for example in the form of Wayang characters [8], [9]. However, the accuracy results obtained are very different. Deep learning algorithms are better than the machine learning algorithms under certain conditions. The condition in question is the number of datasets used, deep learning algorithms tend to require a large number of datasets. Finally, the factor that could affect the accuracy is distribution ration of training and testing dataset. The distribution of the dataset ratio randomly between training and testing data is better and more valid than the manual distribution of the dataset ratio [15].

The Punakawan puppet dataset is not publicly available, so a manual data collection process is required. In addition, due to the limited number of acquired dataset, an augmentation process is needed. This study carried out the process of doubling the image to get higher number and variation in dataset members. The use of the data augmentation method is proven to be very influential on the resulting level of accuracy, especially for deep learning algorithms such as CNN [16]. This augmentation method has several parameters, such as horizontal_flip, shear_range, and so on. Unfortunately, not all of the parameters are good and suitable to use, parameter selection must be carried out precisely and wisely because if too many augmentation parameters are used, it can decrease the accuracy value during the training process. In addition to increase the number of datasets, the use of the augmentation method can minimize the occurrence of overfitting because the dataset is less varied. In addition to enriching the number and variety of datasets used, a multi-optimizer parameter can also increase the accuracy [17]. The use of the right optimizer on certain image objects can affect evaluation and increase classification accuracy. Another study compared several image enhancement methods on iris datasets [18]. The methods compared include adaptive histogram equalization (AHE) and CLAHE. The results of the experiments that have been carried out; the results show that CLAHE can increase accuracy by 7%. Gowda and Yuan stated that the color channel influences the resulting accuracy [19]. Akagic *et al.* [20] proposed that the use of the segmentation process on data set as an input for learning could increase the accuracy of the classification in detecting the cracks.

In this study, aimed to analyze the effect of the Gaussian filter on the accuracy value. In addition, this study aimed to find the best combination of using either Gaussian filter or a non-Gaussian filter with VGG16 learning architecture in K-nearest neighbor (KNN) algorithms in several experimental scenarios. This study, by completing various experiments, found that there are two ways in reducing the noise contained in an image, first using the Gaussian method, and secondly using median filter. By using these filters, the resulting image quality level will increase. With the improvement of image quality, it is expected to increase the accuracy value of Punakawan Wayang image classification. This study employed CNN algorithm with the VGG16 learning architecture and the use of the Gaussian Filter method to evaluate the effect of the selected filter in increasing the classification accuracy. This study also carried out some scenarios where various treatment to the dataset either using a Gaussian Filter or without using a filter will be combined with VGG16 architecture.

This article is structured as follows, section one, the introduction, described the significance of this study and the position of this research among the available references. Section two described the data collection and research method used in this study. In addition, section three presented the experimental process carried out in this study and describes the experimental results obtained. Finally in part four, this article presented the conclusions and findings from the overall experiments.

2. RESEARCH METHOD

To complete the research, this study followed the research methods as depicted in Figure 1. The research began with problem identification, followed by the literature review on the particular problem. From these two initial steps, the research continued by selecting and determining the suitable methods and algorithm to solve the problem. Simultaneously, the private data collection is carried out to collect Punakawan characters and processed them in to the Punakawan dataset. The dataset development was started from collection of the Punakawan raw images by scrapping process with Google search, followed by individually labelling process.

At the same time, the study also determined the experimental scenario by combining the VGG16 architecture with the Gaussian filter method equipped with several scenarios of data treatment in the preprocessing process. From several experimental scenarios that have been carried out in the training data process, a classification model will be formed. The model of each scenario will be evaluated in the testing process using testing data to compare the combination between existing scenarios so that the best combination of experimental scenarios is obtained leading to the scenario with the highest accuracy value.

2.1. Convolutional layer

This layer is one of the most important parts of the CNN architecture. This layer is responsible for extracting the features contained in an object/image [21]. In the feature extraction process, this layer will perform a convolution operation, which is the process of multiplying the image matrix with the filter/kernel matrix. The feature extraction process can be done by several methods such as horizontal and vertical detection. Through this process, important features contained in an image can be obtained for further processing. This convolution layer consists of various neurons that are interconnected with each other. The process of calculating the matrix at the convolution layer is shown in Figure 2.

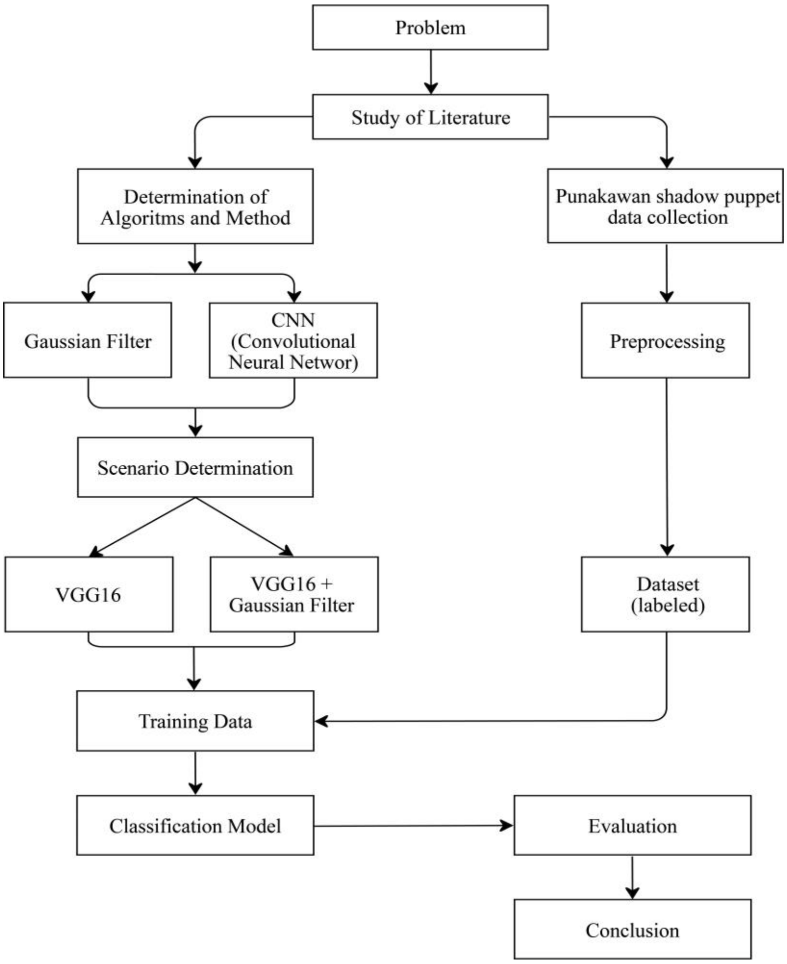


Figure 1. Research flow

The above process, as shown in Figure 2, is a dimension reduction process through the convolution layer. In this process, matrix I is multiplied with a kernel matrix of size 3×3 on matrix J. The results of the multiplication of the two matrices will be added up to produce a new value, namely the value 4 which will become the new matrix as shown in Figure 2.

2.2. Pooling layer

This layer is responsible for reducing the dimensions of the processed image on the convolution layer [21]. The purpose of using this layer is to reduce the occurrence of overfitting because there are many dimensions that do not contain features. The pooling layer has two methods, first the multiplication average pooling between two matrices and its features are determined based on the average value of a series of matrices. The second method is max pooling, which is to take the highest value from the multiplication result at the pooling layer to be taken as a feature. Illustration of the pooling process can be seen in Figure 3.

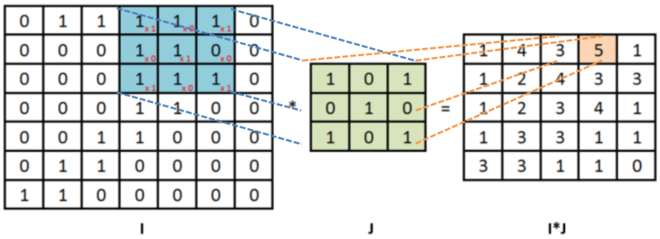


Figure 2. Convolutional layer

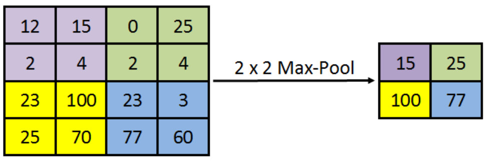


Figure 3. The pooling layer process

The illustration process that occurs in the pooling layer as shown in Figure 3 is a pooling process using the max pooling method. The illustration shows an example of the pooling process using a 2×2 kernel, where the kernel will scan the image matrix by looking for the highest value from the matrix. The highest value of the matrix will form a new matrix in the form of images from the pooling process.

2.3. Fully connected layer

After the feature extraction has been carried out in the previous two layers. The extracted features are then passed on to the fully connected layer. The extracted features are still in the form of a multidimensional array, while this layer only accepts input in the form of a 1-dimensional array [22]. Therefore, it is necessary to convert from a multidimensional array to a 1-dimensional array in the flattening process. The flatten process vector will be fed and processed with a feed-forward neural network and backpropagation for each training process with a series of epoch numbers. The output of this process can distinguish between influential and dominating features with low-level features in the image and classify them using the SoftMax classification technique.

2.4. VGG-16

The network model in VGG16 was proposed by Simonyan and Zisserman [23]. As the name suggests, this architecture consists of 16-layer blocks as feature extraction layers. The kernel used by this model consistently consists of 3×3 with one stride movement [24], [25]. The fully connected layer in the VGG16 network model has a total number of parameters of 138,357,544 parameters. The form of the neural network in the VGG16 architecture can be seen in Figure 4.

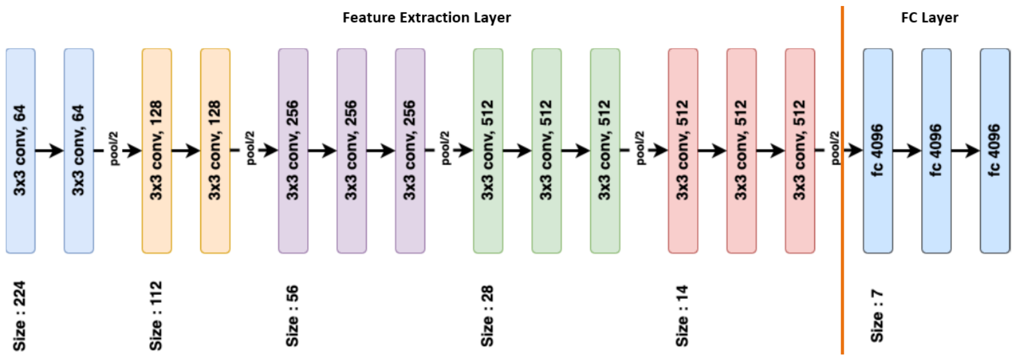


Figure 4. VGG16 architecture

2.5. Gaussian filter method

Gaussian filter is a method that serves to filter the image before the classification process. This method is a linear filter with a weighted value for each member and is selected based on the shape of the Gaussian function. This method was chosen because it can filter images by refining based on the consideration that this filter has a kernel center [26]. This filter is very effective for removing noise that is normally distributed. To calculate or determine the values of each element in the Gaussian smoothing filter that will be formed, it can be calculated through (1):

$$h(x, y) = \frac{1}{c} e^{\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

where σ is standard deviation of the Gaussian Kernel, c is the normalization constant

2.6. The Punakawan puppet dataset

The data used in this study were obtained through a scrapping process from Google with Selenium. The results of the scrapping process still need manual filtering process, due to the image duplication, so it is necessary to take one of the duplicate images out of the dataset. Punakawan character data that has been grouped based on the type of class is then labeled by storing it on Google Drive by providing the name of the folder according to the name of the Punakawan character.

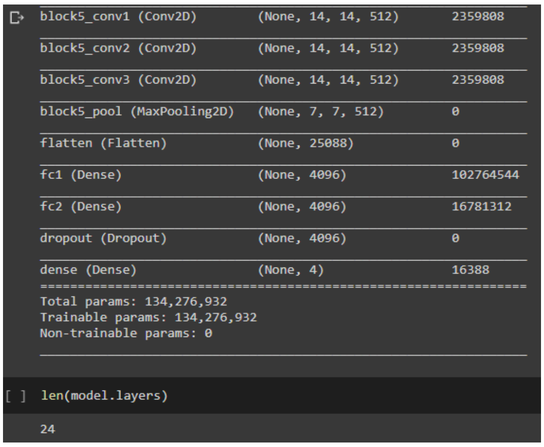
In Table 1, a sample of each of the Punakawan puppet figures is shown. The image has a variety of backgrounds, so a preprocessing process is needed to process the dataset before the training process is carried out. The total dataset used in this study is 400 data which were evenly divided into 4 classes with 100 images for each class. To increase the number and variation of images, an image augmentation process is carried out, so that it is expected to maximize the training data process and increase the accuracy value of the model.

Table 1. Sample dataset

Class	Bagong	Gareng	Semar	Petruk
Puppet Image				

2.7. Fine tuning of VGG16

Before the learning process is carried out with the VGG16 architecture, it is necessary to make an adjustment process to the dataset used in this study. This process is known as fine-tuning. As previously stated, this architecture consists of 1,000 classes in the fully connected layer. So that in order to have the optimal performance it is necessary to carry out an adjustment process. In this research, the adjustment process is carried out by trimming the fully connected layer which consists of 1,000 classes. Then add one layer, namely the dropout layer which functions to prevent the overfitting process. The last adjustment process is adding a fully connected layer with the number of classes adjusted to the number of classes in this study, namely 4 classes. The results of the fine-tuning process can be seen in the summary of the arrangement and number of layers of VGG16 in Figure 5.



```
block5_conv1 (Conv2D)        (None, 14, 14, 512)      2359808
block5_conv2 (Conv2D)        (None, 14, 14, 512)      2359808
block5_conv3 (Conv2D)        (None, 14, 14, 512)      2359808
block5_pool (MaxPooling2D)   (None, 7, 7, 512)        0
flatten (Flatten)            (None, 25088)            0
fc1 (Dense)                  (None, 4096)             102764544
fc2 (Dense)                  (None, 4096)             16781312
dropout (Dropout)            (None, 4096)             0
dense (Dense)                (None, 4)                16388
=================================================================
Total params: 134,276,932
Trainable params: 134,276,932
Non-trainable params: 0

len(model.layers)

24
```

Figure 5. The results of fine tuning

2.8. Experimental scenario

To prove that the use of a Gaussian filter affected the accuracy value, eight experimental scenarios are proposed in this study. This study used eight experimental scenarios which consisted of a combination of the types of dataset treatment in the selection of the image channel fed to CNN architecture, the use of a Gaussian filter and data preprocessing. The overall experimental scenario in the study is shown in Table 2.

Table 2. Experimental scenarios

No	Scenario	Information
1	S1	Green Channel + VGG16 + Gaussian Filter + Thresholding + CLAHE
2	S2	Green Channel + VGG16 + Gaussian Filter + Thresholding
3	S3	Green Channel + VGG16 + Thresholding + CLAHE
4	S4	Green Channel + VGG16 + Thresholding
5	S5	RGB + VGG16 + Gaussian Filter + Thresholding + CLAHE
6	S6	RGB + VGG16 + Gaussian Filter + Thresholding
7	S7	RGB + VGG16 + Thresholding + CLAHE
8	S8	RGB + VGG16 + Thresholding

2.9. Experimental setting

After completing data preprocessing and training on classification models for all experimental scenarios, several experimental scenarios were obtained. The model training process was carried out at epoch 50 with a dataset comparison ratio of 80:20. These models are a representation of knowledge on learning the image of Punakawan puppets. The CNN algorithm has succeeded in identifying the types of puppets from the Wayang data learning process based on the characteristics of each class. The image process is carried out at the convolution layer, wherein this layer the feature extraction of the image is carried out which is then done by reducing the dimensions in the pooling layer process so that the features in the image are seen to improve.

After obtaining the features in the image, the next step is the recognition and identification process on the fully connected layer which will determine the class of the Punakawan puppet image. The model is then tested using testing data to determine the level of accuracy and performance of each model from each of the experimental scenarios. The following are the test results of the classification model, which can be seen in Table 3.

Table 3. Results of the experiment

No	Scenario	Times (s)	Accuracy
1	S1	521.32	0.9167
2	S2	479.53	0.8708
3	S3	492.48	0.9667
4	S4	470.80	0.9042
5	S5	367.12	0.9875
6	S6	351.91	0.9750
7	S7	382.96	0.9625
8	S8	377	0.9292

In Table 3, the experimental results are shown in all scenarios in this study. There are two main scenarios, the use of green channel and the use of red, green, blue (RGB). In scenario 1 to scenario 4 the Punakawan image channels used are the green channel only, because according to [24], this channel has the lowest noise level compared to the other two: blue and red channels. Whereas in the scenario 5 to scenario 8 the image channel used is RGB, an image with three color signals.

3. RESULTS AND DISCUSSION

Table 3 shows the experimental data on all the scenarios in this study. The assessment parameters of this study are merely based on the accuracy value and processing time starting from the preprocessing process to the training data process. For more details, this section will describe in more detail through the diagram illustrations. The following section described the comprehensive explanation of the experimental results in this study.

3.1. Processing time comparison

This section describes the comparison of processing times in all scenarios in this study. The results of the execution time for each scenario can be seen in Figure 6. From the Figure 6, it can be seen that the scenario in the green channel image tends to require a longer processing time than the RGB image in all scenarios. In addition, scenarios that use image quality enhancement, namely using CLAHE, tend to have a longer processing time than scenarios without using CLAHE. This happens, because to improve image quality it is necessary to multiply the dimension value with the histogram value in the CLAHE method.

3.2. Comparison of all scenarios

This section will explain in more detail the accuracy value for each scenario through the graphic illustration shown in Figure 7. The accuracy results obtained in the testing process for all experimental scenarios are shown. In this graph, it can be seen that the experimental scenario using CLAHE, as seen on the scenario (S1, S3, S5, and S7) tends to show better accuracy than the scenario without the use of CLAHE. The graph in Figure 7, also indicated the use of Gaussian filters on the green channel image actually decrease the accuracy value, as seen in the S1 and S2 scenarios, while the scenarios that did not use the Gaussian filter (S3 and S4) actually had a higher accuracy value. However, in the scenario, the use of RGB image scenario combined with the Gaussian filter (S5 and S6) is showing better result than scenario S7 and S8 which does not use the Gaussian filter. Then from all the experimental scenarios, it can be seen that the S5 scenario showed the highest accuracy value compared to other experimental scenarios.

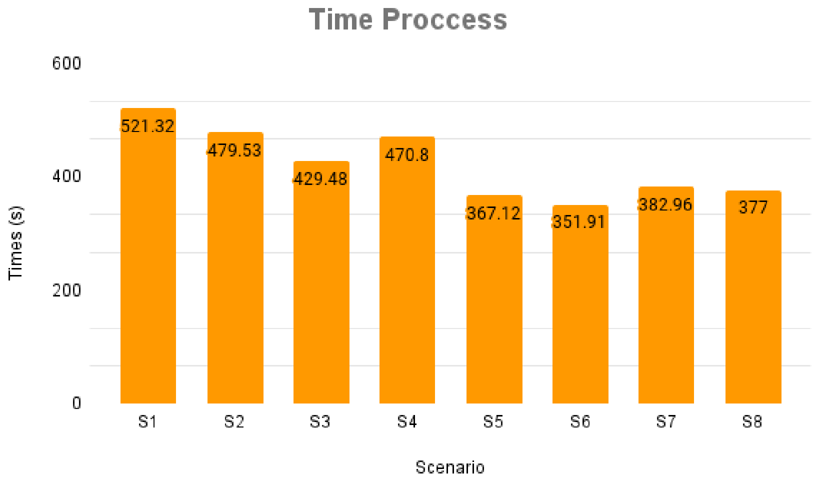


Figure 6. Comparison of processing times

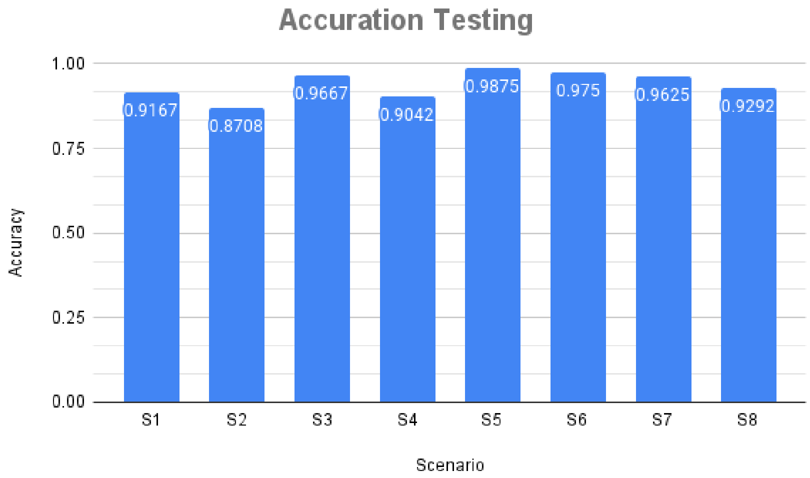


Figure 7. Results of the accuracy of all scenarios

Chowanda showed that the use of VGG16 architecture only resulted in an accuracy value of 92% because it only implemented the algorithm without any treatment of the dataset in the preprocessing process [12]. This study is succeeded in effectively increasing the accuracy at 98.75% of accuracy value by implementing several treatments such as the combination of RGB channels, Gaussian filters, thresholding, and image enhancement processes using the CLAHE method. The worst scenario in this study is (S2) because the Gaussian filter cannot be combined with an image with a green channel. After all, the green channel has the lowest noise value compared to other channels. The Gaussian filter is proven to be an effective filter to increase the classification accuracy of the RGB channel.

Furthermore, as seen on the experimental scenario, the use of Gaussian filter on the green channel image actually decreased the accuracy value. This result can be explained: the Gaussian filter reduced the noise found in an image [13]. In addition, image represented in green channel itself actually has lowest noise as the result of their noise reduction mechanism. So, when the Gaussian filter is applied on the low noise image (executing noise reduction on low noise images), it is less effective method and may result in the lower accuracy value. The feature extraction mechanism on KNN required noise, so it will not work on the image with very low of noise (noise-free images). Whereas in the RGB image the use of Gaussian can increase the accuracy value compared to scenarios (S6, S7) that do not implement the Gaussian filter. The RGB image itself consists of three channels; each channel has different noise. So that the use of a Gaussian filter is suitable when implemented in an image with an RGB channel where there is high volume of noise in the images.

4. CONCLUSION

From the result of the overall experimental scenario, it is clearly show that the Gaussian filter is not suitable filter to be used on green channel images. It may result in the lower the accuracy value. However, when Gaussian filter is implemented in RGB images, it is proven to be an effective method in increasing the accuracy value. The use of the CLAHE method in improving image quality is also effective method in increasing the accuracy of each scenario. Another finding is the processing time required in scenarios using RGB images is faster than green channel images. This study succeeded in increasing the accuracy value of 98.75% as seen on the scenario 5 (S5) after adding a Gaussian filter as a method of reducing noise levels in the image and adding the CLAHE method to sharpen and improve image quality.

ACKNOWLEDGEMENTS

The authors would like to thank Universitas AMIKOM Yogyakarta for funding this research.

REFERENCES

[1] Y. Shin and I. Balasingham, “Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2017, pp. 3277–3280, doi: 10.1109/EMBC.2017.8037556.
[2] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, “Pneumonia detection using CNN based feature extraction,” in *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Feb. 2019, pp. 1–7, doi: 10.1109/ICECCT.2019.8869364.
[3] B. K. Umri, M. Wafa Akhyari, and K. Kusrini, “Detection of COVID-19 in Chest X-ray image using CLAHE and convolutional neural network,” 2020, doi: 10.1109/ICORIS50180.2020.9320806.
[4] R. M. James, Kusrini, and M. R. Arief, “Classification of X-ray COVID-19 image using convolutional neural network,” *2020 2nd International Conference on Cybernetics and Intelligent System, ICORIS 2020*, Oct. 2020, doi: 10.1109/ICORIS50180.2020.9320828.
[5] I. M. A. Agastya, A. Setyanto, Kusrini, and D. O. D. Handayani, “Convolutional neural network for pornographic images classification,” 2018, doi: 10.1109/ICACCAF.2018.8776843.
[6] S. S. Yadav and S. M. Jadhav, “Deep convolutional neural network based medical image classification for disease diagnosis,” *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0276-2.
[7] D. R. Indah, “The symbolic meaning of ‘Punakawan Javanese Wayang’ (a value imaging study in character education at the character education course in STKIP Bina Insan Mandiri Surabaya),” *SELL Journal*, vol. 4, no. 2, pp. 99–106, 2019.
[8] M. Muhathir, M. H. Santoso, and D. A. Larasati, “Wayang image classification using SVM method and GLCM feature extraction,” *Journal Of Informatics And Telecommunication Engineering*, vol. 4, no. 2, pp. 373–382, Jan. 2021, doi: 10.31289/jite.v4i2.4524.
[9] I. B. K. Sudiatmika, Pranowo, and Suyoto, “Indonesian traditional shadow puppet image classification: a deep learning approach,” in *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Jul. 2018, pp. 130–135, doi: 10.1109/ICITEED.2018.8534776.
[10] C. Cheng, X. Wei, and Z. Jian, “Emotion recognition algorithm based on convolution neural network,” in *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nov. 2017, pp. 1–5, doi: 10.1109/ISKE.2017.8258786.
[11] F. I. Apollonio, M. Gaiani, and Z. Sun, “BIM-based modeling and data enrichment of classical architectural buildings,” *SCIentific RESearch and Information Technology*, vol. 2, no. 3, pp. 41–62, 2012, doi: 10.2423/i22394303v2n2p41.
[12] A. Chowanda and R. Sutoyo, “Deep learning for visual Indonesian place classification with convolutional neural networks,” *Procedia Computer Science*, vol. 157, pp. 436–443, 2019, doi: 10.1016/j.procs.2019.08.236.
[13] A. Kumar and S. S. Sodhi, “Comparative analysis of gaussian filter, median filter and denoise autoenocoder,” in *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*, Mar. 2020, pp. 45–51, doi: 10.23919/INDIACom49435.2020.9083712.
[14] T. Guo, J. Dong, H. Li, and Y. Gao, “Simple convolutional neural network on image classification,” in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, Mar. 2017, pp. 721–724, doi: 10.1109/ICBDA.2017.8078730.
[15] S. Fujino, T. Hatanaka, N. Mori, and K. Matsumoto, “Evolutionary deep learning based on deep convolutional neural network for anime storyboard recognition,” *Neurocomputing*, vol. 338, pp. 393–398, Apr. 2019, doi: 10.1016/j.neucom.2018.05.124.
[16] J. Shijie, W. Ping, J. Peiyi, and H. Siping, “Research on data augmentation for image classification based on convolution neural networks,” in *2017 Chinese Automation Congress (CAC)*, Oct. 2017, pp. 4165–4170, doi: 10.1109/CAC.2017.8243510.
[17] A. M. Taqi, A. Awad, F. Al-Azzo, and M. Milanova, “The impact of multi-optimizers and data augmentation on tensorflow convolutional neural network performance,” in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Apr. 2018, pp. 140–145, doi: 10.1109/MIPR.2018.00032.
[18] I. Armeni *et al.*, “3D semantic parsing of large-scale indoor spaces,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 1534–1543, doi: 10.1109/CVPR.2016.170.
[19] S. N. Gowda and C. Yuan, “ColorNet: investigating the importance of color spaces for image classification,” in *Computer Vision – ACCV 2018*, 2019, pp. 581–596.
[20] A. Akagic, E. Buza, S. Omanovic, and A. Karabegovic, “Pavement crack detection using Otsu thresholding for image segmentation,” in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2018, pp. 1092–1097, doi: 10.23919/MIPRO.2018.8400199.
[21] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, “Conceptual understanding of convolutional neural network- a deep learning approach,” *Procedia Computer Science*, vol. 132, pp. 679–688, 2018, doi: 10.1016/j.procs.2018.05.069.
[22] Y. Zhou, H. Wang, F. Xu, and Y.-Q. Jin, “Polarimetric SAR image classification using deep convolutional neural networks,” *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1935–1939, Dec. 2016, doi: 10.1109/LGRS.2016.2618840.
[23] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv:1409.1556*, Apr. 2014.

[24] M. U. Hassan, "VGG16 – convolutional network for classification and detection." neurohive.io, Accessed: Aug. 30, 2018. [Online]. Available: https://neurohive.io/en/popular-networks/vgg16/.
[25] K. Kusrini *et al.*, "Data augmentation for automated pest classification in mango farms," *Computers and Electronics in Agriculture*, vol. 179, no. 1, Dec. 2020, doi: 10.1016/j.compag.2020.105842.
[26] T. M. Khan, D. G. Bailey, M. A. U. Khan, and Y. Kong, "Efficient hardware implementation for fingerprint image enhancement using anisotropic gaussian filter," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2116–2126, May 2017, doi: 10.1109/TIP.2017.2671781.

BIOGRAPHIES OF AUTHORS

Kusrini is a professor from Universitas AMIKOM Yogyakarta Indonesia. She finished her doctoral program from Universitas Gadjah Mada Yogyakarta Indonesia in 2010. She is interested in exploring many things about machine learning and other artificial intelligence field. She also loves in doing research about decision support system and database. She is member of the IEEE and IEEE Systems, Man, and Cybernetics Society. WOS Research ID C-7787-2015, Sinta ID 153621, IEEE ID 92728506 Facebook /kusrini.iskandar, Instagram @kusriniamikom, Youtube Channel @Kusrini Kusrini, and Web: www.kusrini.com. She can be contacted at email: kusrini@amikom.ac.id.

Muhammad Resa Arif Yudianto is a student in Magister of Informatics, Universitas AMIKOM Yogyakarta. He is interested in exploring software engineering, image processing and text mining. WOS Research ID ABE-2504-2021, Sinta ID 6772336, Linkedin https://www.linkedin.com/in/muhammadresa/. He can be contacted at email: muhammadresa0203@students.amikom.ac.id, resamuhammad96@unimma.ac.id.

Hanif Al Fatta is a Ph.D candidate In Information Technology from Universiti Teknikal Malaysia Melaka, Malaysia and a lecturer in Universitas AMIKOM Yogyakarta, Indonesia. His research interest is on exploration of human computer interaction, pervasive learning, game based learning, and usability evaluation. Facebook https://id-id.facebook.com/hanif.a.fatta, Instagram@hanifalfatta, Sinta ID 6003712. He can be contacted at email: hanif.a@amikom.ac.id.