Accredited by National Journal Accreditation (ARJUNA) Managed by
Ministry of Research, Technology, and Higher Education, Republic Indonesia with Second Grade (Peringkat 2, Sinta 2)
since year 2017 to 2021 according to the decree No. 10/E/KPT/2019.

Published online on the journal's webpage: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 5 No. 5 (2021) 984 - 991 ISSN Media Elektronik: 2580-0760

Big Cats Classification Based on Body Covering

Fernanda Januar Pratama[1], Wikky Fawwaz Al Maki[2], Febryanti Sthevanie[3]
[1,2,3]School of Computing, Telkom University
[1]fernandapratama090@gmail.com, [2]wikkyfawwaz@telkomuniversity.ac.id, [3]sthevanie@telkomuniversity.ac.id

Abstract

The reduced habitat owned by an animal has a very bad impact on the survival of the animal, resulting in a continuous decrease in the number of animal populations especially in animals belonging to the big cat family such as tigers, cheetahs, jaguars, and others. To overcome the decline in the animal population, a classification model was built to classify images that focuses on the pattern of body covering possessed by animals. However, in designing an accurate classification model with an optimal level of accuracy, it is necessary to consider many aspects such as the dataset used, the number of parameters, and computation time. In this study, we propose an animal image classification model that focuses on animal body covering by combining the Pyramid Histogram of Oriented Gradient (PHOG) as the feature extraction method and the Support Vector Machine (SVM) as the classifier. Initially, the input image is processed to take the body covering pattern of the animal and converted it into a grayscale image. Then, the image is segmented by employing the median filter and the Otsu method. Therefore, the noise contained in the image can be removed and the image can be segmented. The results of the segmentation image are then extracted by using the PHOG and then proceed with the classification process by implementing the SVM. The experimental results showed that the classification model has an accuracy of 91.07%.

Keywords: CLAHE, Segmentation, PHOG, Support Vector Machines

1. Introduction

As a result of the increasing rate of human growth, it will adversely affect the number of animal populations in the world, especially rare animals that are threatened with extinction. The increasing environmental damage caused by humans causes many wild animals to lose their homes, overfishing and illegal hunting are factors that accelerate the decline in the animal population. By utilizing technological advances especially in the field of image processing and taking advantage of the increasingly high use of digital cameras, this problem can be solved by developing a classification model that can classify the image obtained from digital cameras. In this connection, the classification system can speed up taking the necessary precautions to protect endangered animals, especially animals belonging to the big cat family (Felidae) such as tigers, cheetahs, jaguars, etc. However, the classification method on images has many obstacles, such as long computational time due to large data sets and the use of high-resolution sensors, image capture involving animals with complex backgrounds, different postures, and lighting [1]. Another major problem is that animals have high intra-class variation and inter-class similarity which affects the accuracy of the system [2].

Ainuddin Faaeq et al. [3] proposed manifold learning based on non-linear dimensionality reduction for animal classification. All images are converted into vector-based representations and then the manifold learning algorithm is employed to reduce the dimensions of the vectors. Therefore, the number of features used for the classification can be reduced. Experimental results showed that four classification methods employed in the study, i.e., logistic regression, KNN, SVM, and random forest achieved different accuracy rates, i.e., 98%, 97%, 97%, and 92%, respectively. This study shows that applying manifold learning before classification can speed up the classification process and get high accuracy results.

Slavomir Matuska et al. [4] proposed SVM and local descriptors for the classification of wild animals. Image collections were extracted using feature extraction with different methods such as SIFT, SURF, OpponentSURF, and OpponentSIFT. Key points and Support Vector Machine methods are used in the classification process. The highest accuracy achieved in this study is 86% which is obtained by using a combination of SIURF

Accepted by editor: 24-07-2021 | Final Revision: 27-09-2021 | Online Publication: 31-10-2021

detector, opponentSIFT descriptor, BruteForce matcher, and the clustering process. Meanwhile, the SUSIFT detector achieves a poor accuracy, i.e. 50%.

Isha Dua et al. [5] proposed a detection system to detect elephants in areas of high human intervention. The detection system utilizes a video camera. The video frame is then identified whether there exists elephants or not. The feature extraction and classification tasks are performed by implementing the PHOG method and the SVM classifier, respectively. The experimental results achieved an accuracy of 85.29% when static images containing elephants and other objects are classified.

Annisa Suciati Salsabila et al. [6] proposed a classification of scabies in animals using the Uniform Local Binary Patterns (ULBP) method and the random forest method. Animal skin images are categorized into 2 classes, namely animals with scabies disease and animals with other diseases. Before further processing, this study applies the CLAHE method which is useful for increasing the contrast in the image and making it easier to distinguish skin wounds from animals. However, the experimental results showed that the proposed model reached an accuracy of 52% due to the insufficient number of datasets.

In this study, a classification model is developed to classify big cats based on body covering images. The model focuses on the pattern of body covering possessed by these animals. The classification model is developed by combining the PHOG feature extraction method and SVM classifier. As a descriptor of spatial shape, the PHOG feature is used since the pattern of body covering in each class has a different pattern. Therefore, it is very suitable to use the PHOG feature.

PHOG has demonstrated a good result in a smile recognition system that achieves an accuracy rate of 95.65% [7] and in a vehicle detection system that achieves an accuracy rate of 95% [8]. SVM classifier is one of the most used classification methods in classifying digital images. Also, it produces a high level of accuracy. These successes include the classification of wild animals with the highest accuracy rate reaching 86% [4] and traffic sign detection with the highest accuracy reaching 82.01% [9].

The purpose of this study is to develop a classification model that can categorize bodies covering patterns of big cats according to a predetermined class. The model can also minimize computational time and make the classification process simpler without reducing the classification accuracy. The structure of this paper is described as follows. Part II describes the proposed methodology which includes data augmentation, preprocessing, segmentation, feature extraction, classification, and performance testing. Part III explains the results and discussion. Part IV is the conclusion.

2. Research Method

The methodology proposed in this study consists of 7 stages. This study was conducted by implementing data augmentation, preprocessing, segmentation, feature extraction, training and testing data, classification, and performance testing. The processing sequence is shown in the flowchart of Figure 1.

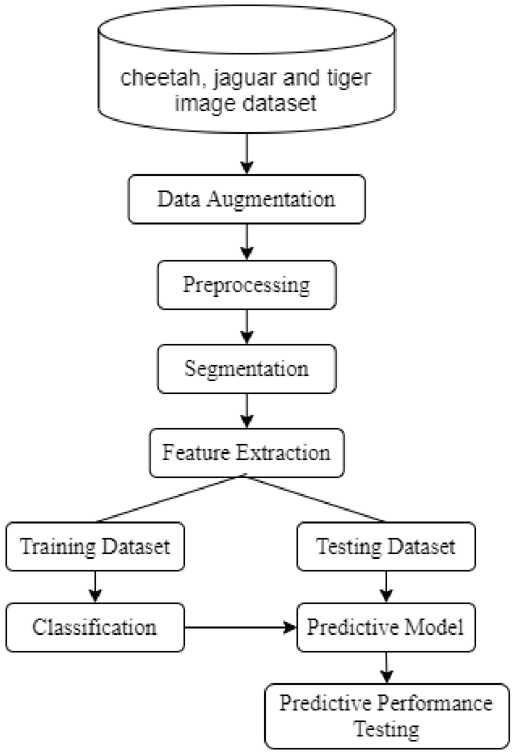

Figure 1. The architecture of the proposed system

2.1. Dataset

The dataset used was obtained from the website www.kaggle.com in the form of cheetah, jaguar, and tiger data [10]. The dataset obtained is still classified as a new dataset uploaded since May 2020. Each class consists of 900 training data images and 100 validation data images. Each image in the dataset has the size of 400 pixels x 400 pixels with RGB values lie in the range of 0-255.

In this study, we used 200 images in each class from the dataset. Therefore, the total number of images used in this study is 600 images. The images are then cropped to remove unnecessary parts so that the classification can focus on the pattern of the body covering of the big cats. The images of the body covering of the big cats are cropped randomly. The copped image has the same size, i.e. 64 x 64 pixels. Examples of cropped images in each class can be seen in Figure 2.

DOI: https://doi.org/10.29207/resti.v5i5.3328

Figure 2. Examples of animal skin images from the dataset

2.2. Data Augmentation

In this study, due to the insufficient number of samples used, we implement a data augmentation to increase the quantity of data and prevent overfitting. The most frequently used data augmentation method and has proven its effectiveness is the traditional affine and elastic transformations [11]. This method can increase image variation by manipulating the image in several ways such as rotating, flip horizontal/vertical, shift, noise, color jittering, zooming in and out.

2.3. Preprocessing

In this study, contrast stretching is used as the preprocessing method. The contrast stretching is implemented by applying Contrast Limited Adaptive Histogram Equalization (CLAHE). The CLAHE method can reduce the impact of excessive lighting or vice versa when the image is captured. CLAHE is used to increase contrast level in an image. This task is performed by stretching the intensity value to avoid excessive contrast enhancement that can occur especially when performing traditional histogram equalization [12].

The CLAHE algorithm was developed based on Adaptive Histogram Equalization (AHE). CLAHE is different from AHE in terms of contrast. The way CLAHE works is that the input image is divided into several regions called tiles. Histograms are created and calculated based on each region in the image. The histogram is then truncated according to a predetermined clip limit, the truncated histogram is redistributed to each gray level. Furthermore, the cumulative distribution function (CDF) is scaled and mapped based on each region of the image using the pixel values of the truncated histogram. Next, each region or tile is then combined into a single image using bilinear interpolation. The average number of pixels at each gray level is formulated in equation 1 [13].

$$N_{avg} = \frac{N_{CR-Xp} * N_{CR-Yp}}{N_{gray}} \quad (1)$$

Based on equation 1, N_{CR-Xp} , N_{CR-Yp} , and N_{gray} represent the number of pixels in the X direction in the contextual region, the number of pixels in the Y direction in the contextual region, and the number of gray levels in the contextual area, respectively.

CLAHE has two main parameters: clip limit (CL) and block size (BS). These parameters aim to control the improved image quality. The larger the clip limit, the brighter the image will be since the input image has a very low intensity so that a large clip limit results in a flatter histogram on the image. Larger block size can also increase the contrast in the image this is due to the larger dynamic range. The clip limit of a histogram can be calculated using equation 2 [12].

$$\beta = \frac{M}{N}\left(1 + \frac{\alpha}{100} S_{max}\right) \quad (2)$$

Based on equation 2, M is the number of pixels in each block, N is the dynamic range in this block, S_{max} is the maximum slope, and α is the clip factor.

2.4. Segmentation

Under normal circumstances, the image obtained by using a digital camera contains a large amount of noise. The noise can be caused by various conditions such as low light conditions or very slow shutter speed during the image acquisition process. Also, it can be caused by random disturbances during the image capturing process. Furthermore, the noise and random disturbances can change the characteristics of the pattern in the image to be processed. To overcome this problem, we propose a segmentation method by combining the median filter and the Otsu method. The median filter is used to filter the image whereas the Otsu method is used to separate the foreground from the background in the filtered image.

2.4.1. Image Filtering

The median filter is a filtering process that applies a non-linear filter. It is usually employed to remove noise from an image in many image processing tasks. In median filtering, the neighboring pixels of the selected window pixels are sorted in ascending order. Then, the ordered pixels are replaced by the median value. If the window size is too large, some of the desired details in the image will be erased [14]. Let us assume that an image has the dimension of 5 x 5 pixels as shown in Figure 3. If we apply median filtering using a sliding window with a dimension of 3 x 3, we will have 9 values of the image pixel to be ordered.

120	110	178	175	188
119	104	103	154	149
131	120	130	127	115
122	116	111	125	129
167	133	135	155	137

(a)

103	104	111	116	120	125	127	130	154

(b)

Figure 3. Illustration of a matrix on an image with a size of 5 x 5 (a) where the pixels in the sliding window are sorted in ascending order and calculate the median based on that order (b).

DOI: https://doi.org/10.29207/resti.v5i5.3328

Based on the illustration in Figure 3, the median value is 120. In the final step, the pixel value of 130 is replaced by 120. Figure 4 shows the results of preprocessing using median filtering.

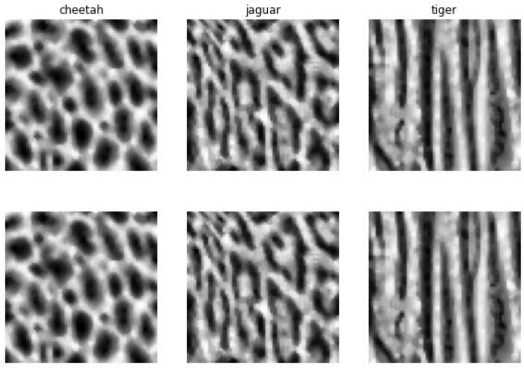

Figure 4. The results of filtering animal skin images using Median Filtering

2.4.2. Otsu Method

The Otsu method was first discovered by Nobuyuki Otsu. The Otsu method is the most used method for image segmentation. The image is converted into a black-and-white image where white represents the background pixels and black as the foreground pixels. In image binarization, thresholding has a very important role in finding the ideal threshold value and minimizing the intra-class variance of black and white pixels. The selection of the ideal threshold value is useful for highlighting objects from the background, especially in grayscale images. The mathematical equations used to determine the optimal thresholding are defined in equations 3 and 4 [15].

$$\mathrm{J(x,y)} = \begin{cases} 255 \text{ if } \mathrm{I(x,y)} \geq \mathrm{T} \\ 0 \text{ Otherwise} \end{cases} \tag{3}$$

$$\mathrm{T} = \mathrm{M}[\mathrm{x}, \mathrm{y}, \mathrm{p(x,y)}, \mathrm{I(x,y)}] \tag{4}$$

Based on equations 3 and 4, T represents the threshold, I(x, y) is the coordinate representation of the grayscale image value, and p(x, y) represents the average value of the neighborhood focusing on the I(x, y) coordinates.

Thresholds are generally divided into 2 types, namely global thresholds and local thresholds. The global threshold is a T threshold value only depending on the value of I(x, y) or the gray level. While the local threshold is a threshold value of T that does not only depend on I(x, y) but also on p(x, y). The local threshold divides the image into sub-regions and determines the threshold value for each sub-region.

The Otsu method includes a global threshold method that only depends on the gray value of the image. The Otsu method is generally widely applied because it is simple and effective in segmenting images. The results of image segmentation are shown in Figure 5.

2.5. Feature Extraction

Feature extraction is an important process where images are identified and represented for further processing. Feature extraction describes important shape details found in a pattern thereby making the formal process of identifying the pattern simple. At this stage, the PHOG descriptor is used to identify detailed patterns on the body covering of the animal.

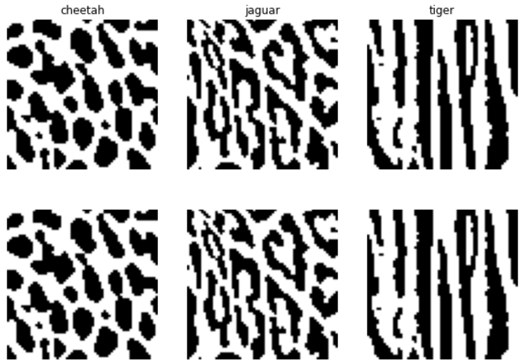

Figure 5. Binary image of animal body covering pattern using Otsu thresholding

PHOG is a spatial shape descriptor that combines the pyramid representation and Histogram of Orientation Gradients (HOG). The pattern of animal body covering in each class has a different shape. Therefore, it is very suitable to use the PHOG feature to help get a more accurate image descriptor.

First, canny edge detection is applied to a pre-processed image to extract the edge contours of the image. This task can help to get a more accurate image descriptor. Then, the image is divided into several sub-images as illustrated in Figure 6. The number of sub-images can be adjusted according to the depth of the pyramid, then the sub-images are sorted. Then, the HOG feature is calculated according to the sub-images that have been sorted. The local shape is represented by a histogram of edge orientation within a certain angle range and quantized into K bins. The gradient in the image (direction x and y) is calculated by convoluting the filter kernel on each image pixel I(x,y) using equations 5 and 6 [9].

$$G_x = [-1 \quad 0 \quad 1] * \mathrm{I} \tag{5}$$

$$G_y = [-1 \quad 0 \quad 1]^T * \mathrm{I} \tag{6}$$

The magnitude of the gradient is then calculated with the help of equations 7 and 8 [9].

$$\mathrm{G} = \sqrt{(\mathrm{G_x})^2 + (\mathrm{G_y})^2} \tag{7}$$

$$\theta = \tan^{-1}\left(\frac{\mathrm{G_y}}{\mathrm{G_x}}\right) \tag{8}$$

The orientation range is divided into k bins and used to create histograms. The histogram represents the

DOI: https://doi.org/10.29207/resti.v5i5.3328

frequency distribution of a data set based on orientation within a certain range of angles. The histograms are combined to form a long vector as a representation of an image.

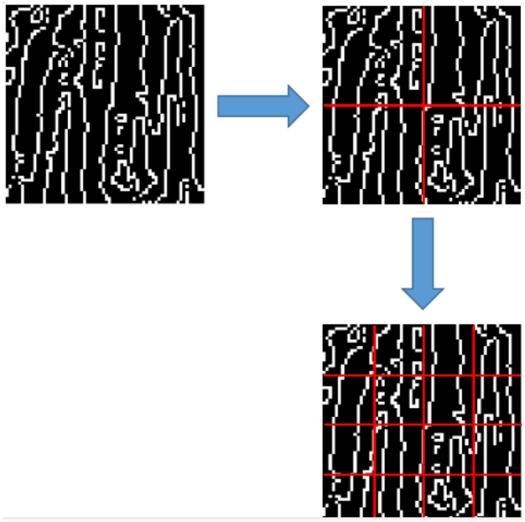

Figure 6. Illustration of PHOG feature extraction

2.6. Classification

The classification stage is carried out by implementing the multi-SVM method. Support Vector Machines (SVM) is a classification method that is closely related to supervised learning and is usually used for the classification of linear and non-linear patterns. The main goal of SVM is to find the optimal hyperplane. In this connection, the input data is converted into an n-dimensional space. Then, the SVM looks for the optimal hyperplane based on the margin distance to the closest point of each existing pattern, so that it can classify the given pattern correctly [16]. The larger the maximum margin size, the more accurately it classifies a pattern. This condition can be formulated based on the SVM linear classification optimization problem, as shown in equation 9 [16].

$$min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\lambda} t_i \quad (9)$$

Subject to:

$$y_i(wx_i + b) + t_i \geq 1 \quad (10)$$

$$t_i \geq 0, i = 1, \dots, \lambda \quad (11)$$

Based on equation 9, we want to maximize the margin between the two classes by minimizing $\|w\|^2$. t_i is a slack variable. Based on equations 10 and 11, x_i is the input data, y_i is the output of the data x_i, w, b are the parameters whose values will be searched. The hyperplane is illustrated as in Figure 7.

In this study, SVM works by dividing the input data that has been processed into 2 groups, namely training data and test data. The separation of data is an important part of the classification. Data separation using the same data can minimize data differences and can understand the characteristics of the model well.

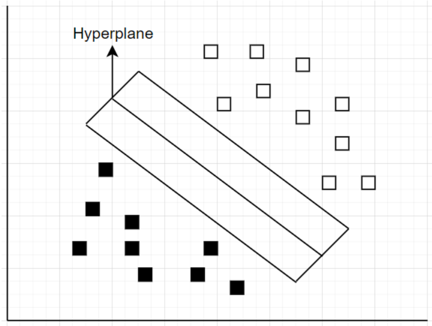

Figure 7. Illustration of hyperplane

When designing an SVM classifier, selecting optimal parameters is important to minimize errors when classifying patterns. In this connection, the parameters selected are the kernel, complexity value, and gamma value. The parameter C or the complexity parameter determines the number of misclassifications that are ignored during the process. If the value of the complexity parameter is too large, the SVM performance will decrease. In other words, the classification falls to an underfitting condition. However, if the value of the complexity parameter is too small, the classification of images by using SVM will fall to an overfitting condition. The SVM classifier generally uses kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid. The gamma parameter is represented as a boundary curve in the kernel function. If the gamma value is getting smaller, the decision boundary curve becomes smaller so that the decision boundaries are wider. Otherwise, if the gamma value is greater, the decision boundary curve becomes larger so that the decision-boundaries region has a circle-like shape.

2.7. Performance Testing

A confusion matrix is a table that represents performance results in the classification system. Each column (or row) in the confusion matrix table displays the predicted results from the classification model, while each row (or column) displays the value of the actual class. The confusion matrix can be seen in table 1. There are four metrics available in the confusion matrix table i.e. true positive (TP), false positive (FP), false negative (FN), and true negative (TN). True positive, usually called sensitivity, is the number of positive classes that are predicted correctly because they correspond to the actual class. False positive is the number of positive classes whose predictions do not match the actual class. False negative is the number of negative classes whose predictions do not match the actual class. True negative, usually called specificity, is the number of negative classes that are predicted correctly because they correspond to the actual class.

DOI: https://doi.org/10.29207/resti.v5i5.3328

Table 1. Illustration of the confusion matrix

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Based on the confusion matrix table, several methods will be used to evaluate the performance of the classification model such as classification accuracy, AUC-ROC, F1-score, Precision, and recall.

Classification Accuracy (CA) is the ability of a model to distinguish data based on each class. Classification accuracy is calculated by comparing true positive and true negative in all evaluated cases. Mathematically, it can be seen in equation 12 [16].

$$CA = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

The area under the receiver operating characteristic curve (AUC-ROC) is employed to measure how well the parameters in the classification model distinguish between two or more classes. AUC-ROC will measure the quality of the classification by calculating the area under the curve formed from the results of the calculation of the true positive rate (TPR) and false positive rate (FPR) [17]. A perfect classifier is a classifier that has an AUC value of 1.0. A classifier is considered good if the AUC value is greater than 0.5. However, if the AUC value is less than or equal to 0.5, the classifier is categorized as a poor classifier.

F1-score is a sub-contrary mean of recall and precision. F1-score is calculated by combining the results of recall and precision which have the same weight equation [16]. The formula used can be seen in equation 13 [18].

$$F1 = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (13)$$

Precision is the ratio of true positives outcomes among positive classified instances. Precision can be defined as the ratio between the number of correctly predicted images and the total number of predicted images. The formula used can be seen in equation 14 [18].

$$PR = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (14)$$

Recall is the proportion of true positives among all positive instances in the data. Recall or also known as sensitivity can be calculated as the probability that the image that is correctly predicted (true positive) is relevant. The formula used can be seen in equation 15 [18].

$$RE = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (15)$$

3. Result and Discussion

The dataset used in this study is a combination of the results of the augmentation method and the original image so that the total number of datasets becomes 15,000 images. The dataset is divided into 2 groups (training data and test data). In this connection, the total data has been divided into 12000 images for training data and 3000 images for test data.

The image on the training and test data is converted into a grayscale image. Then, the image is processed using the CLAHE method with a clip limit parameter size of 40.0 and a block size of (8, 8). Next, the image is segmented by using a combination of the median filter with a size of 2x2 and the Otsu method. After that, the image is extracted by using the PHOG feature depth 1. In this connection, we use a 9-bin orientation in the range of [0, 180] at each level, pixel_per_cell parameter size (8, 8), and using cell_per_block parameter with size (2,2) to get the feature descriptor from the image. The feature descriptor is then classified by using multi-SVM using parameters C=10, gamma=0.1, and kernel Radial Basis Function (RBF). These parameters are obtained by implementing hyperparameter tuning grid search and cross-validation methods with 10-fold. We also implemented the HOG in feature extraction. In this connection, we used 9-bin orientation in the range [0,180], pixel_per_cell parameter size (8, 8) and cell_per_block parameter with size (2, 2).

As for the model comparison, we also implemented HOG, PHOG depth 1, and PHOG depth 2 when we employed the KNN as the classifier. In this connection, we apply n_neighbors = 5 where n_neighbors is the number of neighboring parameters. To evaluate the performance of each model, we use the values of accuracy, AUC-ROC, F1, precision, and recall. Accuracy results in experiments using the SVM and KNN methods can be seen in table 2.

Table 2. Comparison results from some methods

Methods	KNN	SVM
HOG	69.43%	90.83%
PHOG depth 1	69.30%	91.07%
PHOG depth 2	69.47%	89.07%

Based on table 2, the accuracy of the SVM-based classification system where HOG, PHOG depth 1, and PHOG depth 2 were employed as the feature extraction methods reached 90.83%, 91.07%, and 89.07%, respectively. In contrast, the accuracy results using the KNN classifier method on each feature extracted HOG, PHOG depth 1, and PHOG depth 2 reached 69.43%, 69.30%, and 69.47%, respectively. The results of the comparison of the AUC-ROC, F1, Precision, and Recall values for each feature of HOG, PHOG depth 1, and PHOG depth 2 can be seen in tables 3 and 4.

DOI: https://doi.org/10.29207/resti.v5i5.3328

Table 3. Comparison of AUC_ROC, F1, Precision, and Recall using SVM classifier

Methods	HOG	PHOG depth 1	PHOG depth 2
AUC-ROC	97.99%	98.10%	97.44%
F1	90.71%	90.93%	88.86%
Precision	90.69%	90.90%	88.83%
Recall	90.76%	91.0%	88.89%

Table 4. Comparison of AUC_ROC, F1, Precision, and Recall using KNN classifier

Methods	HOG	PHOG depth 1	PHOG depth 2
AUC-ROC	89.84%	90.14%	90.08%
F1	64.06%	65.06%	65.48%
Precision	79.10%	77.18%	76.98%
Recall	70.31%	70.14%	70.28%

The experimental results show that the best result is achieved when we implemented PHOG depth 1 as the feature extraction method and SVM as the classifier.

4. Conclusion

Based on our experimental results, the classification model developed by implementing a combination of features PHOG depth 1 and SVM classifier can classify big cats body covering pattern image with the best accuracy of 91.07%. However, the classification system using a combination of features PHOG depth 2 and SVM classifier obtains an accuracy of 89.97%. The descriptor feature generated by the PHOG depth 1 feature is more suitable for the SVM classifier than the descriptor feature generated by the PHOG depth 2 feature. Compared to the results obtained from the HOG feature and the PHOG depth 2 feature, the classification system that uses the PHOG depth 1 feature is more superior. The extraction process using the PHOG feature has advantages compared to the HOG method. In this connection, PHOG uses overlapping blocks so that it can increase its durability. The results obtained from this study also prove that image classification using the SVM classifier obtains more superior results compared to that of the KNN classifier since the SVM classifier uses a kernel function. In this connection, the SVM can separate data more optimally than the KNN classifier.

From our experimental results, it is shown that there is potential to develop better models by employing more complex datasets to classify animals belonging to the big cat family. It is hoped that this research can also be implemented in organizations that protect animals so that it can make it easier to identify and take preventive actions appropriately and quickly.

References

[1] S. Taheri and Ö. Toygar, "Animal classification using facial images with score-level fusion," *IET Computer Vision*, vol. 12, no. 5, pp. 679–685, 2018, doi: 10.1049/iet-cvi.2017.0079.

[2] N. Manohar, Y. H. S. Kumar, and G. H. Kumar, "Supervised and unsupervised learning in animal classification," *2016 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016*, pp. 156–161, 2016, doi: 10.1109/ICACCI.2016.7732040.

[3] A. Faaeq, H. Guruler, and M. Peker, "Image classification using manifold learning based non-linear dimensionality reduction," *26th IEEE Signal Processing and Communications Applications Conference, SIU 2018*, pp. 1–4, 2018, doi: 10.1109/SIU.2018.8404441.

[4] S. Matuska, R. Hudec, P. Kamencay, M. Benco, and M. Zachariasova, "Classification of Wild Animals based on SVM and Local Descriptors," *AASRI Procedia*, vol. 9, no. Csp, pp. 25–30, 2014, doi: 10.1016/j.aasri.2014.09.006.

[5] A. Mittae, "A Vision based Human - Elepahant Collision," pp. 225–229, 2015.

[6] A. S. Salsabila, F. Sthevanie, and K. N. Ramadhani, "Scabies Classification in Animal Using Uniform Local Binary Patterns," *ICITEE 2020 - Proceedings of the 12th International Conference on Information Technology and Electrical Engineering*, pp. 356–361, 2020, doi: 10.1109/ICITEE49829.2020.9271720.

[7] Y. Bai, L. Guo, L. Jin, and Q. Huang, "A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition," *Proceedings - International Conference on Image Processing, ICIP*, no. 07118074, pp. 3305–3308, 2009, doi: 10.1109/ICIP.2009.5413938.

[8] H. Wang, W. Bo, and L. Sun, "Pyramid histogram of oriented gradient and particles swarm optimization based SVM for vehicle detection," *Proceedings - 2013 7th International Conference on Image and Graphics, ICIG 2013*, pp. 323–327, 2013, doi: 10.1109/ICIG.2013.70.

[9] A. Sugiharto and A. Harjoko, "Traffic sign detection based on HOG and PHOG using binary SVM and k-NN," *Proceedings - 2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2016*, pp. 317–321, 2017, doi: 10.1109/ICITACEE.2016.7892463.

[10] Kaggle, "Cheetah, Hyena, Jaguar and Tiger," 2020. https://www.kaggle.com/iluvchicken/cheetah-jaguar-and-tiger (accessed October 15, 2020).

[11] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," *2018 International Interdisciplinary PhD Workshop, IIPhDW 2018*, no. August 2019, pp. 117–122, 2018, doi: 10.1109/IIPHDW.2018.8388338.

[12] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, "Automatic Contrast-Limited Adaptive Histogram Equalization with Dual Gamma Correction," *IEEE Access*, vol. 6, pp. 11782–11792, 2018, doi: 10.1109/ACCESS.2018.2797872.

[13] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, pp. 2392–2397, 2014, doi: 10.1109/ICACCI.2014.6968381.

[14] G. George, R. M. Oommen, S. Shelly, S. S. Philipose, and A. M. Varghese, "A Survey on Various Median Filtering Techniques For Removal of Impulse Noise From Digital Image," *Proc. IEEE Conference on Emerging Devices and Smart Systems, ICEDSS 2018*, no. March, pp. 235–238, 2018, doi: 10.1109/ICEDSS.2018.8544273.

[15] E. Niharika, H. Adeeba, A. S. R. Krishna, and P. Yugander, "K-means based noisy SAR image segmentation using median filtering and Otsu method," *IEEE International Conference on IoT and its Applications, ICIOT 2017*, pp. 3–6, 2017, doi: 10.1109/ICIOTA.2017.8073630.

[16] P. Bhasker, K. Pant, and K. R. Pardasani, "Support vector machine for classification of plant and animal miRNA," *ACT 2009 - International Conference on Advances in Computing, Control and Telecommunication Technologies*, pp. 338–340, 2009, doi: 10.1109/ACT.2009.90.

[17] A. Yaseen, W. A. Abbasi, and F. U. A. A. Minhas, "Protein binding affinity prediction using support vector regression and

interfecial features," *Proceedings of 2018 15th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2018*, vol. 2018-Janua, pp. 194–198, 2018, doi: 10.1109/IBCAST.2018.8312222.

[18] M. Q. Shatnawi, M. Alrousan, and S. Amareen, "A new approach for content-based image retrieval for medical applications using low-level image descriptors," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, pp. 4363–4371, 2020, doi: 10.11591/ijece.v10i4.pp4363-4371.

DOI: https://doi.org/10.29207/resti.v5i5.3328