

Fusion of BIFFOA and Adaptive Two-Phase Mutation for Helmetless Motorcyclist Detection

Sutikno¹

Department of Computer Science
Diponegoro University
Semarang, Indonesia

Agus Harjoko^{2*}, Afiahayati³

Department of Computer Science and Electronics
Universitas Gadjah Mada
Yogyakarta, Indonesia

Abstract—Road traffic injuries and deaths cause considerable economic losses to individuals, families, and nations as a whole. One of the strategies needed to curtail these fatalities is the surveillance of helmetless motorcyclists, which is carried out by developing an automatic detection system based on computer vision. Generally, this system consists of three subsystems, namely, moving object segmentation, motorcycle classification, and helmetless head detection. HOPG-LDB (Histogram of Oriented Phase and Gradient - Local Difference Binary) descriptor for this system produced good accuracy; however, it still has a drawback related to a large number of features. Based on these observations, this paper proposed an Adaptive Two-phase Mutation Binary Improved Fruit Fly Optimization Algorithm (ATMBIFFOA) to reduce the features. The ATMBIFFOA is a new feature selection algorithm that improved BIFFOA (Binary Improved Fruit Fly Optimization Algorithm) with an adaptive two-phase mutation algorithm. The BIFFOA produced good accuracy; however, weak in reducing feature dimension. The adaptive two-phase mutation algorithm was used to cover this weakness. The experiment results show that the proposed method can reduce the number of features and computation time effectively from BIFFOA. The proposed method produced motorcycle classification accuracy of 96.06% for the JSC1 dataset and 96.85% for the JSC2 dataset. As for helmetless head detection, the proposed method produced an average precision of 66.29% for the JSC1 dataset and 63.95% for the JSC2 dataset.

Keywords—Motorcycle classification; helmetless head detection; BIFFOA; two-phase mutation algorithm

I. INTRODUCTION

Road traffic injuries and deaths cause considerable economic losses to individuals, families, and nations as a whole. Based on the current trends, these problems are predicted to continually occur for a long period. Furthermore, World Health Organization (WHO) published that traffic accidents were the 7th leading cause of death in the world, with 1.35 million mortality cases being recorded yearly [1]. In Indonesia, the number of deaths caused by two and three-wheel motorcyclists was approximately 74% among other traffic accidents [1]. The main cause of this type of accident is the head injury sustained due to the unyieldingness of the use of helmets. WHO reported that the use of helmets reduces the risk of 69% of head injuries [1]. Most countries mandated the use of helmets; however, many motorcyclists still violate the regulation and escape the consequences, because of the difficulty of direct surveillance on the highway, which is not

monitored for a full day. Meanwhile, research in automatic detection based on computer vision has been growing rapidly, to curtail these problems.

In general, the study of detection of motorcyclists not wearing helmets was divided into two subsystems, namely motorcycle detection and helmetless head detection [2]. The feature extraction process gives an impact on the performance of both subsystems. Previous studies have used hand-crafted features and a convolutional neural network (CNN). The Histogram of Oriented Gradient (HOG) descriptor is a hand-crafted feature that results in relatively high accuracy. The author in [3] used HOG to classify vehicles in various environments and views. The author in [4] used HOG in both subsystems which resulted in good accuracy, but it still incorrectly detects distant objects. The author in [5] reported that HOG produces higher accuracy than Wavelet Transform (WT), Local Binary Pattern (LBP), and their combination in helmetless head detection. The author in [6] reported that HOG produces higher accuracy than Scale-Invariant Feature Transform (SIFT) and LBP in both subsystems.

Currently, the CNN method is popular for classification and detection in various domains. The author in [7] used CNN for motorcycle detection to overcome the problem of changing lighting and poor video quality. CNN was also used for helmetless head detection with various models, for example, AlexNet [8], VGG16 [9], VGG19, Inception V3, and MobileNets [10]. The author in [11] combined HOG and LBP for vehicle classification, and compared hand-crafted features (combination of HOG, LBP, and Haralick) and Custom CNN for helmetless head detection. The result showed that the method produces higher accuracy than HOG and LBP for motorcycle classification. For helmetless head detection, Custom CNN is superior in terms of accuracy and hand-crafted features are superior in terms of prediction time with relatively good accuracy. In addition, [12] compared several hand-crafted features (HOG, LBP, and Gabor) and CNN for vehicle classification. The result showed that the HOG produces better accuracy than other descriptors and CNN.

However, several authors stated that the HOG lacks to deal effectively with images of varying lighting [13], and different local patterns [14]. This ineffectiveness can be solved by combining HOG, Histogram of Oriented Phase (HOP), and Local Difference Binary (LDB) descriptors called Histogram of Oriented Phase and Gradient - Local Difference Binary (HOPG-LDB) [15]. The result of the experiment showed that

*Corresponding Author

the HOPG-LDB descriptor increases the accuracy of the HOG, however, it still has a drawback related to a large number of features. The author in [16] stated that one of the preprocessing techniques that reduce these numbers is feature selection. The author in [17] stated that feature selection techniques are divided into 2, namely filters and wrappers. The author in [18] reported that the wrapper method tends to provide better performance than the filter. This technique can significantly improve the selection of relevant features [19].

The author in [20] reported that Binary Improved Fruit Fly Optimization Algorithm (BIFFOA) feature selection produces a good performance. Based on the experiment of this method compared with other algorithms, namely binary Gray Wolf Optimization (bGWO), Binary Gravitational Search Algorithm (BGSA), Binary Bat Algorithm (BBA), Binary Salp Swarm Algorithm (BSSA), Binary Grasshopper Optimization Algorithm (BGOA), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Correlation-based Feature Selection (CFS), Fast Correlation-Based Filter (FCBF), F-Score, Information Gain (IG), and spectrum. The results showed that the BIFFOA has the best accuracy, however, weak in reducing feature dimensions.

A solution for reducing feature dimensions was proposed in [21], which integrated the Gray Wolf Optimizer algorithm (GWO) and two-phase mutation (TMGWO). The first phase mutation is used to reduce feature dimensions and the second attempt to increase accuracy. The experiments of this method were compared with other algorithms, namely BBA, Binary Crow Search Algorithm (BCSA), binary Gray Wolf Optimization Algorithm (bGWOA), binary Whale Optimization Algorithm (bWOA), Discrete Particle Swarm Optimization (DPSO), Flower Algorithm (FA), Multi-Verse Optimization (MVO), PSO, and Non-Linear Particle Swarm Optimization (NLPSO). The results showed that the TMGWO produces the best accuracy and second-best feature reduction compared to other methods.

Detection of motorcyclists not wearing helmets in real-time is required the high accuracy and speed. The addition of a feature selection process can improve this performance. BIFFOA feature selection produced a good performance; however, weak in reducing feature dimension [20]. This weakness can be solved by adding a two-phase mutation algorithm. This study aims to add a feature selection process to detect motorcyclists not wearing helmets. The addition of the proposed feature selection is to reduce the number of features so the computation time of motorcycle classification and helmetless head detection can be reduced. The main contributions of this paper are:

- Adaptive Two-phase Mutation Binary Improved Fruit Fly Optimization Algorithm (ATMBIFFOA) is proposed. This algorithm is a fusion of BIFFOA and an adaptive two-phase mutation algorithm.
- An algorithm of adaptive two-phase mutation that is modified from a two-phase mutation is proposed.
- The ATMBIFFOA feature selection is added after the feature extraction process to reduce features in the motorcycle classification and helmetless head detection.

This paper is organized as follows. Related work is presented in Section II that is divided into two parts: motorcycle detection and helmetless head detection. Section III explains the dataset used, the proposed algorithm, and the evaluation methods. In Section IV, we present the experimental result and discussion. Finally, the main conclusion is introduced and future work is suggested in Section V.

II. RELATED WORK

In general, this study is divided into two subsystems: motorcycle detection and helmetless head detection.

A. Motorcycle Detection

Motorcycle detection has concerned three processes: vehicle segmentation, feature extraction, and classification. The author in [22] used three shape features: length, width, and their ratios to categorize vehicles into five groups. The result received from the usage of the decision tree (DT) classifier confirmed high accuracy, however, the features had been now no longer able to differentiate bicycles, motorcycles, and tricycles. The author in [23] used the features of length, width, area, diameter, and the ratio of distance to decide the object's center of mass and its main axis length. The classification method used a multilayer perceptron (MLP) to categorize the vehicles into three categories: heavy and mild duties, and motorcycles.

The author in [24] used the area feature to categorize motorcycles and others. Meanwhile, the author in [25] proposed a way that specializes in calculating the number of motorcycles on the street in real-time. The features used are area, height, and width to categorize motorcycles and non-motorcycles.

The author in [6] compared a few descriptors: HOG, SIFT, and LBP in classifying motorcycles and non-motorcycles. The effects confirmed that the HOG descriptor has exceptional accuracy. The author in [26] compared HOG, Speeded Up Robust Features (SURF), SIFT, and LBP in motorcycles detection. It has a look at extensively utilized information of images taken from in front, besides, and at the back of motorcycles. The result confirmed that the HOG descriptor has better accuracy.

A observe through as in [27] proposed a system that categorized vehicles into four categories: cars, vans, buses, and motorcycles. The system used Intensity Pyramid-based HOG (IPHOG) descriptor and support vector machine (SVM) classifier. The outcomes confirmed that the situations of climate and light converting have decreased accuracy than the normal condition.

The author in [28] used the LBP descriptor and SVM classifier to locate motorcycles. This descriptor became as compared with SURF, HOG, and Haar Wavelet. The outcomes confirmed that the proposed method has higher accuracy than the others. The author in [5] proposed a WT descriptor that became as compared with LBP, HOG, and SURF. The outcomes confirmed that the WT descriptor has higher accuracy than the others.

The author in [29] proposed a combination of shape and color features comprising of area, the ratio of width and height,

and color deviation standard. These features served as entering for the k-nearest neighbors (KNN) classifier to decide the motorcycles and non-motorcycles. The proposed approach becomes capable of calculating the wide variety of passengers on a motorcycle. The outcomes confirmed mistakes in type due to the fact the data had been taken from afar, overlapping vehicles, and the passenger sitting too near the rider. The author in [11] concatenated HOG and LBP with sequential minimum optimization (SMO) for training the SVM classifier. The outcomes display that the combination of those descriptors produced higher accuracy than the HOG and LBP descriptors. The author in [15] concatenated HOG, HOP, and LDB descriptors with MLP classifier. The results show that the proposed descriptor has higher accuracy than HOG, HOP, LDB, HOG-HOP, HOG-LDB, and HOP-LDB descriptors. Moreover, the proposed method has higher accuracy than in [5], [6], and [11].

CNN has additionally been used for motorcycle detection. The author in [30] used CNN which specializes in jam situations. The CNN is also used in [7] to address numerous lights and bad video quality. The take a look at outcomes displays that the accuracy generated is better than hand-crafted features, however, the computation time is much longer.

B. Helmetless Head Detection

Helmetless head detection has involved three stages: ROI (region of interest) determination, feature extraction, and classification. The ROI determination pursuits to check the region round a rider's head. The heads of the rider and passenger are above the motorcycle image; therefore, the studies focused at the top a part of the object. Once the head region is known, the following steps are feature extraction and classification.

The author in [24] used the circular hough transform (CHT) descriptor for helmetless head detection. The result confirmed that it error still occurs for the detection of two or extra passengers. The author in [5] proposed the HOG descriptor, and the dataset was taken in a static environment. The assessment was performed by comparing HOG, WT, LBP, WT+LBP, WT+HOG, LBP+HOG, and WT+HOG+LBP descriptors. The result confirmed that the HOG descriptor has the best accuracy. The author in [6] compared HOG, SIFT, and LBP descriptors. The outcomes confirmed that HOG has the best accuracy. However, the data were taken on a quiet road.

Some researchers have combined shape, texture, and color features to enhance accuracy. The author in [29] used features of arc circularity, average intensity, and hue. Data were taken from three recording conditions, which include near, far, and medium. It turned into located that the greatest mistakes had been from the data recording acquired from afar.

The author in [26] used the features of arc circularity, average intensity and hue, and Center Symmetric-Local Binary Pattern (CS-LBP). These features served as entering the KNN classifier for the classification of heads with and without helmets. The technique focused on troubles with data recording

taken from special angles withinside the front, besides, and back. However, the head images were cropped manually. The author in [31] extensively utilized arc circularity, average intensity and color, and HOG.

The author in [28] used geometric, shape, and texture characteristics. The study used a combination of CHT, LBP, and HOG descriptors. CHT is used to decide the geometric form of an image. This technique has a weak point that the incapability of detecting images of low resolution. The author in [15] used the HOPG-LDB descriptor that concatenated HOG, HOP, and LDB descriptors. The results show that the HOPG-LDB descriptor has higher accuracy than HOG, HOP, LDB, HOG-HOP, HOG-LDB, and HOP-LDB descriptors. Moreover, the proposed method has higher accuracy than in [5] and [6].

The author in [32] used CNN with the YOLOv2 model to detect riders without helmets. The author in [7] used the AlexNet model on each light and heavy traffic. The author in [8] extensively utilized the AlexNet model and inaccurate detections have been made for riders placing on hats. The author in [33] used the iter_45, Inception-V3 network, and full ImageNet network models. The author in [34] proposed Faster Regions with Convolution Neural Network (R-CNN) for decreasing the computing time. However, inaccurate detections have been nevertheless made for bicycle riders. The author in [11] compared hand-crafted features (a combination of HOG, LBP, and Haralick) and Custom CNN. Custom CNN is advanced in terms of accuracy and hand-crafted features are advanced in terms of prediction time with pretty proper accuracy.

III. MATERIALS AND METHODS

A. Dataset

This study used two datasets, namely JSC1 and JSC2 taken from the rear and front of an object, respectively. Both datasets contain motorcycle and non-motorcycle images used for the input of motorcycle classification. The input of helmetless head detection used the motorcycle images. Fig. 1 shows some samples of both datasets.

The datasets were generated from the segmentation process of the video. This process consists of several stages, namely histogram equalization of video frames that have been converted to grayscale, Gaussian Mixture Model (GMM) to determine foreground, and morphological operations (opening and dilation) to remove noise. The author in [35] stated that GMM is robust to lighting changes. The first video for the JSC1 dataset was recorded on Cipinang Baru Timur Street at East Jakarta, Indonesia with a frame speed of 19.49 fps. The second video for the JSC2 dataset was recorded on Budi Raya Street at West Jakarta with a frame speed of 20 fps. Both videos have a resolution of 1280 x 720 pixels and a duration of 3 hours. Training data was generated from the first 2 hours and testing data was generated from the rest. This data division technique was also used in [6]. The number of the training and testing data are shown in Table I and Table II, respectively.



Fig. 1. Samples of Datasets.

TABLE I. THE NUMBER OF TRAINING DATA

Dataset	Motorcycle Classification Subsystem		Helmetless Head Detection Subsystem	
	Motorcycle	Non-motorcycle	Head with helmet	Head without helmet
JSC1	1602	1602	2052	1694
JSC2	4066	4066	1984	1984

TABLE II. THE NUMBER OF TESTING DATA

Dataset	Motorcycle Classification Subsystem		Helmetless Head Detection Subsystem	
	Motorcycle	Non-motorcycle	Motorcyclist with helmet	Motorcyclist without helmet
JSC1	531	587	416	115
JSC2	1390	1852	1091	299

B. Developed System

In general, the system for detecting helmetless motorcyclists is divided into 3 subsystems, namely moving object segmentation, motorcycle classification, and helmetless head detection. This study focuses on developing motorcycle classification and helmetless head detection, as shown in Fig. 2. The stage of the head detection in the helmetless head detection subsystem is begun the determination process of the

ROI of the head from the motorcycle image. The ROI limits are determined based on the minimum and maximum positions of the upper 1/3 of the blob image generated from the segmentation process. The resulting image was converted to a grayscale image and was enhanced by its contrast using CLAHE (contrast-limited adaptive histogram equalization). Fig. 3(a) is an example of a motorcycle image. Fig. 3(b) is the result of this process. The next step was to create two binary images with opposite intensities using thresholding and inverse thresholding, but some blobs still need to be filtered and fixed. Fig. 3(c) shows the result of this process. The filtering was carried out by morphological operations (opening and filling holes), removing blobs on the side and top edges, and removing too big blobs. Moreover, overly tall blobs were fixed by removing the bottom. Fig. 3(d) shows the result of this process. Edge detection of Laplace of Gaussian (LoG) is used for the next step with the results as in Fig. 3(e). After that, CHT is applied to both images, and then the results are combined on an ROI head image. An example of this result is as in Fig. 3(f). The classification in the head detection is used to classify the objects bounding box on the circular into the head and non-head. Classification in helmetless head detection is used to classify head objects into heads wearing a helmet and not wearing a helmet. Fig. 3(g) is an example of the classification of head detection. The author in [5] reported that the MLP classifier produces a good performance so this paper used it. The feature extraction process used the HOPG-LDB descriptor [15].

C. Binary Improved Fruit Fly Optimization Algorithm (BIFFOA)

The author in [20] explained that BIFFOA is developed from the Improved Fruit Fly Optimization (IFFO) algorithm for the feature selection, by converting it from continuous to binary version. The author in [36] explained that the IFFO algorithm is improved from the Fruit Fly Optimization (FFO) algorithm that is used to determine global optimization. The weakness of the FFO algorithm is that the search radius on all iterations is the same. In the IFFO Algorithm, the search radius (r) for each iteration is calculated using (1).

$$r = r_{max} \cdot \exp\left(\log\left(\frac{r_{min}}{r_{max}}\right) \cdot \frac{Iter}{Iter_{max}}\right) \quad (1)$$

where r_{max} and r_{min} are the radii of maximum and minimum, respectively. $Iter$ represents the iteration, and $Iter_{max}$ represents the maximum number of iterations. The author in [37] explained that $r_{min} = (UB-LB)/2$ and $r_{max} = 10^{-5}$, where UB (upper bound) and LB (lower bound) value 1 and 0, respectively.

The initialization parameters in the BIFFOA algorithm are PS , r_{max} , r_{min} , and $Iter_{max}$. In addition, the initial swarm location is initialized by selecting the best solution, which is determined by the agent with the smallest fitness value. The fitness function is designed as shown in (2).

$$Fitness = \alpha \gamma_R(D) + \beta \frac{|R|}{|C|} \quad (2)$$

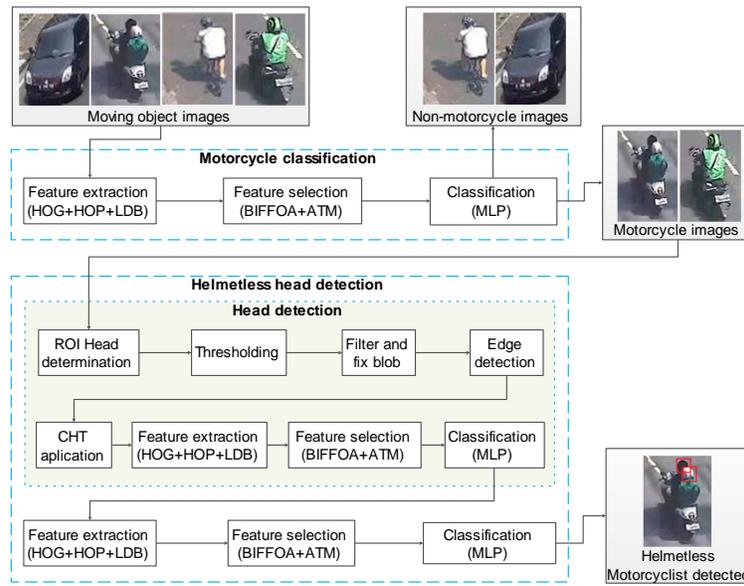


Fig. 2. A Developed System of Motorcycle Classification and Helmetless Head Detection.

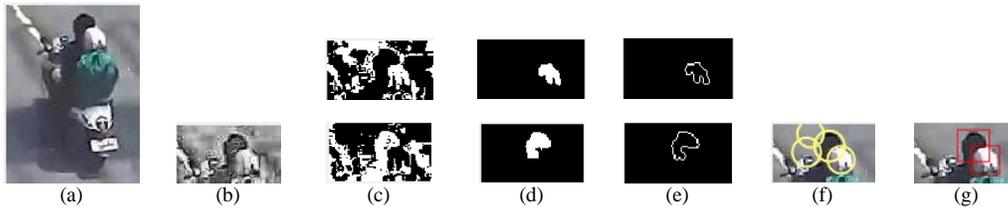


Fig. 3. Image Sample of Results from each Step: (a) Motorcycle Input (b) ROI Head Determination (c) Thresholding (d) Filter and Fix Blob (e) Edge Detection (f) CHT Application (g) Classification.

where $\gamma_R(D)$ represents the classification error rate of a given classifier. $|R|$ and $|C|$ denote the length of the selected feature subset and the total number of features, respectively. α and β represent the weight of classification accuracy and selected feature subset, respectively. The values of α and β for this study are 0.99 and 0.01, respectively. The agents used are a swarm of fruit fly positions. The initial positions of this fruit fly are binary numbers randomly generated. The position of fruit flies is updated using (3).

$$x_{i,j} = \begin{cases} 1 - \delta_j & \text{if } S(\Delta x_{i,j}) \geq \text{rand}() \\ \delta_j & \text{otherwise, } j = 1, 2, \dots, n \end{cases} \quad (3)$$

where n is the dimension length and $\text{rand}()$ is the generation of random numbers between $[0, 1]$. δ_j is the j^{th} dimension of the optimal solution. $S(\Delta x_{i,j})$ is the sigmoidal transfer function (S-shaped), as in (4).

$$S(\Delta x_{i,j}) = \frac{1}{1 + e^{-\Delta x_{i,j}}} \quad (4)$$

where $\Delta x_{i,j}$ is calculated using (5).

$$\Delta x_{i,j} = \begin{cases} \delta_j \pm r \cdot \text{rand}() & \text{if } j = d \\ \delta_j & \text{otherwise, } j = 1, 2, \dots, n \end{cases} \quad (5)$$

where r is the search radius for every iteration that is calculated using equation (1). d is a dimension index that is chosen randomly. The pseudocode of the BIFFOA is shown in Algorithm 1 [20].

Algorithm 1. The standard BIFFOA

1. **Input:** $PS, r_{max}, r_{min}, Iter_{max}$
//Initialize the BIFFOA parameter:
2. Set $PS, r_{max}, r_{min}, Iter_{max}$
3. Calculate the fitness of all agents using (2)
4. Set the best solution as swarm location
5. $Iter=0$
6. $X^*=\Delta$
7. **Repeat**
8. Calculate the search radius r using (1)
9. Calculate $\Delta x_{i,j}$ using (5)
//Ospres is foraging phase
10. **For** $i=1, 2, \dots, PS$
11. Calculate the $S(\Delta x_{i,j})$ using (4)
12. Using (3) to generate food source, $X_i=(x_{i,1}, x_{i,2}, \dots, x_{i,n})$
13. **End for**
// Vision foraging phase
14. Calculate the fitness of all agents using (2)
15. Update swarm location when there is a better solution in the population
16. **Until** $Iter=Iter_{max}$
17. **Output:** Solution X^*

D. Two-Phase Mutation Algorithm

A two-phase mutation algorithm was proposed in [21] to improve the GWO algorithm. Algorithm 2 shows the

pseudocode of the two-phase mutation [21]. The input of this algorithm is the best grey wolf (X_α) in each iteration. The X_α is mutated in two phases, the first is used to reduce features and the second is utilized in improving accuracy. The mutation is executed when the r is less than the Mutation Probability (M_p). The value of r is between 0 and 1, which is generated randomly and the M_p value is 0.5.

Algorithm 2. The standard two-phase mutation

1. **Input:** the best grey wolf X_α from each iteration
 2. $Fitness =$ calculate the fitness of X_α
//start the first phase
 3. Define vector *one_positions* to store the locations of the selected features in X_α
 4. Define $X_{mutated1} = X_\alpha$
 5. **For** $i=1$ to length of *one_positions* //for each selected feature in X_α
 6. Generate a random number r
 7. **If** ($r < M_p$)
 8. $X_{mutated1}[one_positions[i]] = 0$ while keeping the other features
 9. $Fitness_mutated =$ the fitness of $X_{mutated1}$
 10. **If** ($Fitness_mutated < Fitness$)
 11. $Fitness = Fitness_mutated$
 12. $X_\alpha = X_{mutated1}$
 13. **End if**
 14. **End if**
 15. **End for**
//start the second phase
 16. Define vector *zero_positions* to store the locations of the unselected features in X_α
 17. Define $X_{mutated2} = X_\alpha$
 18. **For** $j=1$ to length of *zero_positions* //for each unselected feature in X_α
 19. Generate a random number r
 20. **If** ($r < M_p$)
 21. $X_{mutated2}[zero_positions[j]] = 1$ while keeping the other features
 22. $Fitness_mutated =$ the fitness of $X_{mutated2}$
 23. **If** ($Fitness_mutated < Fitness$)
 24. $Fitness = Fitness_mutated$
 25. $X_\alpha = X_{mutated2}$
 26. **End if**
 27. **End if**
 28. **End for**
 29. **Output:** the improved X_α
-

E. Proposed Algorithm: Fusion of BIFFOA with Adaptive Two-Phase Mutation Algorithm

ATMBIFFOA is a new feature selection algorithm that is proposed in this paper. It improved the BIFFOA by adding an adaptive two-phase mutation algorithm that aims to reduce feature dimensions. The pseudocode of the ATMBIFFOA is found in Algorithm 3, while that of the adaptive two-phase mutation algorithm is found in Algorithm 4.

The input of the adaptive two-phase mutation algorithm is the best solution for each iteration (X^*) that is defined as shown in (6).

$$X^* = (x_1, x_2, \dots, x_n) \quad (6)$$

where x_j is the j^{th} dimension of X^* , and n is the dimension length of X^* . When each x_j values= 0, then the corresponding feature is unselected. And when each x_j values= 1, then the corresponding feature is selected. The M_p is defined as the vector as shown in (7).

$$M_p = (mp_1, mp_2, \dots, mp_n) \quad (7)$$

where mp_j is the mutation probability of the j^{th} dimension. The mp_j values are constant at the beginning of iteration (mp_j in the study is 0.5). However, when the iteration is greater than or equal to the weight iteration (I_w), then M_p is equal to the dimension weights of the best agents. The I_w is calculated using (8).

$$I_w = t_w \times Iter_{max} \quad (8)$$

where t_w is a weight threshold that values a range of [0, 1].

The dimension weights of the best agents are represented in the vector (W_i), as in (9).

$$W_i = (w_{i,1}, w_{i,2}, \dots, w_{i,n}) \quad (9)$$

where $w_{i,j}$ is the weight of the best agent in the i^{th} iteration and j^{th} dimension that is calculated using (10).

$$w_{i,j} = \frac{w_{(i-1),j} \times (i-1) + x_{i,j}}{i} \quad (10)$$

where $x_{i,j}$ is the value of the best solution in the i^{th} iteration and j^{th} dimension.

Algorithm 3. The proposed ATMBIFFOA

1. **Input:** $PS, r_{max}, r_{min}, Iter_{max}$
//Initialize the ATMBIFFOA parameter:
 2. Set $PS, r_{max}, r_{min}, Iter_{max}$
 3. Set W_0, M_p, I_w
 4. Calculate the fitness of all agents using (2)
 5. Set the best solution as swarm location
 6. $Iter = 0$
 7. $X^* = \Delta$
 8. **Repeat**
 9. Calculate the search radius r using (1)
 10. Calculate Δx_{ij} using (5)
//Ospres is foraging phase
 11. **For** $i=1, 2, \dots, PS$
 12. Calculate $S(\Delta x_{ij})$ using (4)
 13. Using (3) to generate food source, $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n})$
 14. **End for**
// Vision foraging phase
 15. Calculate the fitness of all agents using (2)
 16. Update the swarm location when there is a better solution in the population
 17. Update W_i using (9)
 18. **If** $Iter \geq I_w$
 19. $M_p = W$
 20. **End if**
//Mutation process
 21. Process of the adaptive two-phase mutation
 22. **Until** $Iter = Iter_{max}$
 23. **Output:** Solution X^*
-

The mutation process in the two-phase mutation algorithm is executed in all dimensions of one_position and zero_position vector. Therefore, this algorithm takes a long time when used in a large number of features. The adaptive two-phase mutation algorithm limited the number of mutated dimensions with the 1st Mutation Candidate Probability (P_{mc1}) and the 2nd Mutation Candidate Probability (P_{mc2}), therefore, its computation time can be reduced. The mutation position of both vectors is selected through random permutation.

Algorithm 4. The proposed adaptive two-phase mutation

1. **Input:** the best solution X^* from each iteration
//start the first phase
2. Define vector *one_positions* to store the locations of the selected features in X^*
3. Define $X_{mutated1} = X^*$
4. Define the number of mutation candidate $n_{mc} = P_{mc1} \times$ length of *one_position*
5. Define vector *one_mutation_candidate* to store the location of the mutated candidate by selecting n_{mc} random permutations in *one_position*.
6. **For** $i=1$ to length of *one_mutation_candidate*
7. Generate a random number r
8. **If** ($r < M_p[\text{one_mutation_candidate}[i]]$)
9. $X_{mutated1}[\text{one_mutation_candidate}[i]] = 0$
 while keeping the other features
10. *Fitness_mutated* = the fitness of $X_{mutated1}$
11. **If** (*Fitness_mutated* < *Fitness*)
12. *Fitness* = *Fitness_mutated*
13. $X^* = X_{mutated1}$
14. **End if**
15. **End if**
16. **End for**
//start the second phase
17. Define vector *zero_positions* to store the locations of the unselected features in X^*
18. Define $X_{mutated2} = X^*$
19. Define the number of mutation candidate $n_{mc} = P_{mc2} \times$ length of *zero_position*
20. Define vector *zero_mutation_candidate* to store the location of the mutated candidate, by selecting the n_{mc} random permutations in the *zero_position*
21. **For** $j=1$ to the length of *zero_mutation_candidate*
22. Generate a random number r
23. **If** ($r < M_p[\text{zero_mutation_candidate}[j]]$)
24. $X_{mutated2}[\text{zero_mutation_candidate}[j]] = 1$
 while keeping other features
25. *Fitness_mutated* = the fitness of $X_{mutated2}$
26. **If** (*Fitness_mutated* < *Fitness*)
27. *Fitness* = *Fitness_mutated*
28. $X^* = X_{mutated2}$
29. **End if**
30. **End if**
31. **End for**
32. **Output:** the improved X^*

F. Evaluation Methods

The parameters for measuring performance were feature number (NF), time of average classification (Time), accuracy (Acc), precision (Pre), and recall (Rec). Especially for helmetless head detection, we used average precision (AP) to measure accuracy. Equations (11), (12), and (13) are used in calculating accuracy, precision, and recall, respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

$$Recall = \frac{TP}{TP+FN} \quad (13)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively. The TP is the true detection of the ground-truth bounding box. The correct detection was measured using the intersection over union (IOU) as in (14).

$$IOU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \quad (14)$$

where B_p and B_{gt} represent the predicted and ground-truth bounding box, respectively. The detection is considered correct when the IOU is greater than or equal to the threshold. In this study, the threshold value was 0.5.

AP is the area under the precision-recall curve which has a range of [0, 1] and it is calculated as in (15) [38].

$$AP = \sum_n (R_{n+1} - R_n) P_i(R_{n+1}), \quad (15)$$

where $P_i(R_{n+1})$ is calculated using (16).

$$P_i(R_{n+1}) = \max_{\tilde{R}: \tilde{R} \geq R_{n+1}} P(\tilde{R}) \quad (16)$$

where $P(\tilde{R})$ is the precision measured at the time of recall \tilde{R} .

The experiment was carried out by selecting the best result on each process, namely a combination of cell and block sizes on the HOPG-LDB descriptor, variation of t_w value on the ATMBIFFOA, and a combination of hidden layer number, neuron number, and training algorithm on the MLP. The block size variations were 2x2 and 3x3 cells. The cell size variations in the 2x2 blocks were 4x4, 6x6, 8x8, and 12x12 pixels and the 3x3 block sizes were 4x4 and 8x8 pixels. The variations of t_w values were 0.25, 0.5, and 0.75. The variations of the number of hidden layers used are 1, 2, and 3. The number of neurons in the hidden layers (n_H) was determined by using (17) [39].

$$n_H = \sqrt{n_i + n_o} + l \quad (17)$$

where n_i is the number of neurons in the input layer, n_o is the number of neurons in the output layer, and l is an integer constant of 1 to 10. The l variation of this study was 1, 5, and 10. Finally, we used 8 variations of the training algorithm, namely the gradient descent with adaptive learning rate backpropagation (traingda), scaled conjugate gradient backpropagation (trainscg), conjugate gradient backpropagation with Powell-Beale restarts (traincgb), conjugate gradient backpropagation with Fletcher-Reeves update (traincgf), conjugate gradient backpropagation with Polak-Ribière update (traincgp), one step secant backpropagation (trainoss), gradient descent with momentum and adaptive learning rate backpropagation (traingdx), and gradient descent backpropagation (traingd). This paper used the learning rate of 0.05, the epoch maximum number of 1000, and the limit for error of 0.001 for the training. The author in [40] reported that these parameters result in a good performance.

For the ATMBIFFOA, the parameters of PS and $Iter_{max}$ are 24 and 100, respectively. The values of P_{mc1} and P_{mc2} are 0.25 and 0.01, respectively. The K-fold cross-validation (K=5) was used to separate the training and validation data in the feature selection process. Each experiment is run 5 times and the average results are used. All the experiments were carried out in Windows 10 Ultimate 64-bit operating system, with processor Intel Core (TM) i7-9750HQ CPU and 16 GB of RAM. All the algorithms were implemented in the MATLAB R2019a Software.

IV. RESULTS AND DISCUSSION

This section shows the experiment results of the proposed method for motorcycle classification and helmetless head detection.

A. Motorcycle Classification

The first experiment is to determine the best accuracy of the proposed method (ATMBIFFOA) with variations of t_w . Table III shows the results of this experiment for the motorcycle classification. From this table, we can be seen that the best accuracy reaches 96.06% for the JSC1 dataset and 96.85% for the JSC2 dataset.

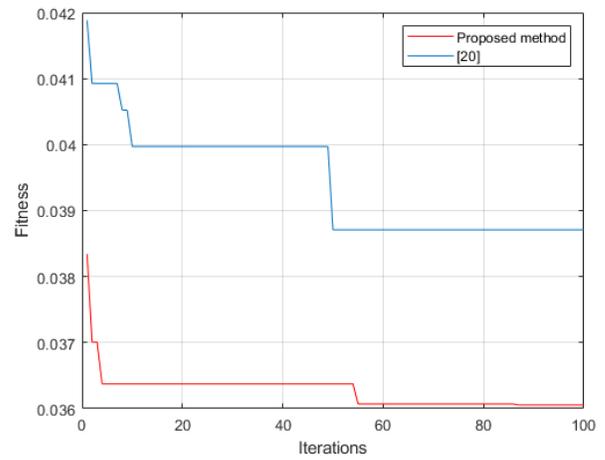
Furthermore, the proposed method is compared with the previous study, namely in Table IV. Here, [20] used BIFFOA feature selection. From this table, it can be seen that the proposed method is superior in terms of the number of features and classification time. Meanwhile, in terms of accuracy, the proposed method is superior for the JSC1 dataset, and [20] is superior for the JSC2 dataset. For this reason, we conclude that the addition of an adaptive two-phase mutation algorithm in BIFFOA can effectively reduce the number of features.

TABLE III. EXPERIMENTAL RESULTS WITH VARIATION OF t_w FOR MOTORCYCLE CLASSIFICATION

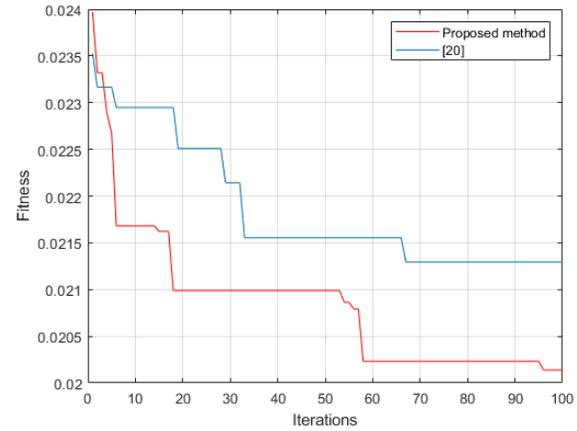
Dataset	t_w	Acc (%)	Pre (%)	Rec (%)	NF	Time ($\times 10^{-3}$ s)
JSC1	0.25	95.90	95.38	96.05	899.0	46.4321
	0.50	96.06	95.63	96.12	951.8	42.6672
	0.75	95.39	95.28	95.03	829.6	38.1412
JSC2	0.25	96.71	97.28	94.98	232.8	34.9727
	0.50	96.85	97.38	95.21	227.8	34.2448
	0.75	96.71	97.46	94.81	253.0	41.5804

TABLE IV. FEATURE SELECTION COMPARISON BETWEEN PROPOSED METHOD AND PREVIOUS STUDY

Dataset	Method	NF	Time ($\times 10^{-3}$ s)	Acc (%)	Pre (%)	Rec (%)
JSC1	[20]	1155.4	48.685	96.05	95.62	96.08
	Proposed method	951.8	42.667	96.06	95.63	96.12
JSC2	[20]	283.6	42.320	96.96	97.67	95.19
	Proposed method	227.8	34.244	96.85	97.38	95.21



(a) JSC1 Dataset.



(b) JSC2 Dataset.

Fig. 4. Convergence Curve of the Proposed Method and the Previous Study for Motorcycle Classification.

Fig. 4 displays the convergence curve of the proposed method and the previous study in [20]. The proposed method produces a better optimal solution than the BIFFOA.

B. Helmetless Head Detection

Table V shows the experimental results of helmetless head detection with variations in the value of t_w . From this table, it can be seen that the highest AP reaches 66.29% for the JSC1 dataset and 63.95% for the JSC2 dataset. Furthermore, the proposed method is compared with previous studies.

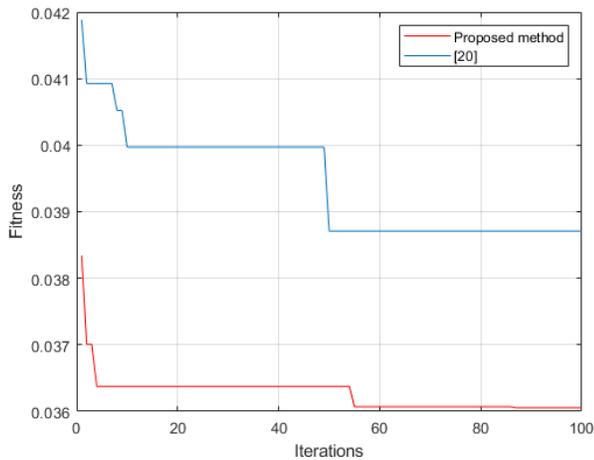
Table VI shows a comparison of the proposed feature selection method and previous study. Here, [20] used the BIFFOA feature selection. From this table, it can be seen that the proposed method is superior in terms of the number of features and classification time. In addition, the AP of the proposed method is superior for the JSC2 dataset, although it is slightly lower for the JSC1 dataset. Fig. 5 shows the comparison of the convergence curve between the proposed method and the previous study. The proposed method produces a better optimal solution than [20]. Therefore, we can conclude that the addition of an adaptive two-phase mutation algorithm to BIFFOA can reduce features effectively.

TABLE V. EXPERIMENTAL RESULTS WITH A VARIETY OF t_w FOR HELMETLESS HEAD DETECTION

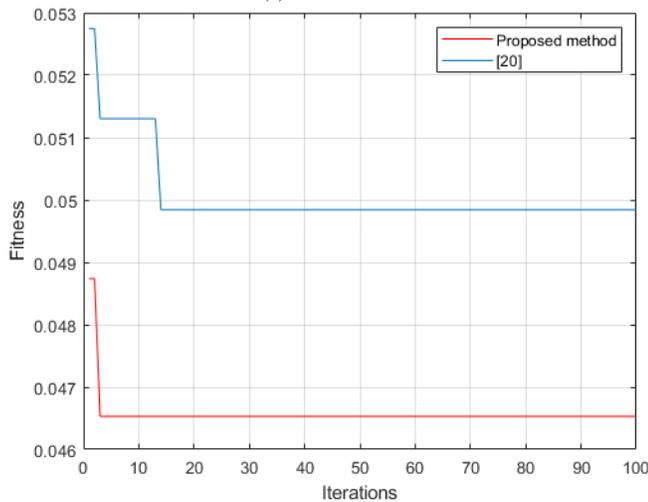
Dataset	t_w	AP (%)	Pre (%)	Rec (%)	NF	Time ($\times 10^{-3}$ s)
JSC1	0.25	66.29	57.19	76.74	138.6	4.489
	0.50	63.59	55.63	75.56	151.6	4.499
	0.75	66.09	54.90	76.74	147.6	4.487
JSC2	0.25	60.49	51.65	79.18	416.2	6.044
	0.50	63.95	52.68	81.82	391.2	4.464
	0.75	61.92	49.98	80.95	419.0	5.929

TABLE VI. FEATURE SELECTION COMPARISON BETWEEN PROPOSED METHOD AND PREVIOUS STUDY FOR HELMETLESS HEAD DETECTION

Dataset	Method	NF	Time ($\times 10^{-3}$ s)	AP (%)	Pre (%)	Rec (%)
JSC1	[20]	159.6	4.641	66.68	57.31	77.19
	Proposed method	138.6	4.489	66.29	57.19	76.74
JSC2	[20]	445.0	5.255	62.55	51.59	81.23
	Proposed method	391.2	4.464	63.95	52.68	81.82



(a) JSC1 dataset



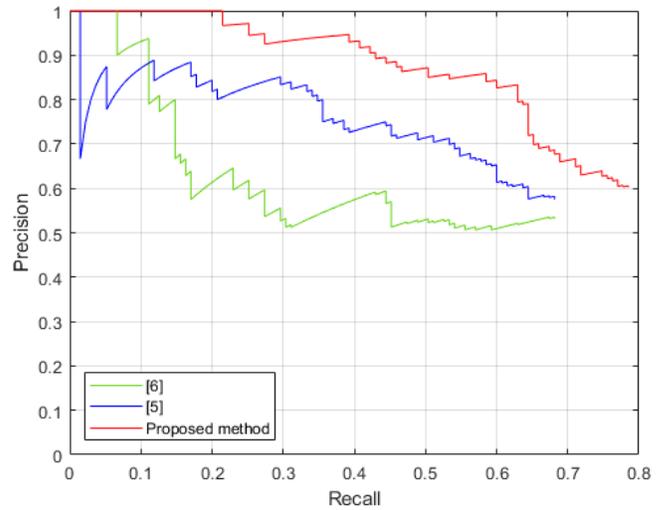
(b) JSC2 dataset

Fig. 5. Convergence Curve of the Proposed Method and the Previous Study for Helmetless Head Detection.

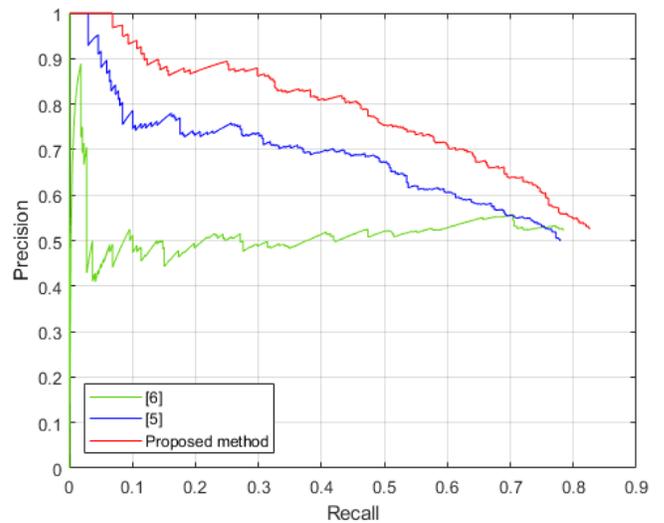
The proposed method is also compared with previous studies, as shown in Table VII. Here, [6] used a combination of HOG descriptor and SVM classifier, and [5] used a combination of HOG descriptor and MLP classifier. AP of the proposed method is superior when compared to these methods. Fig. 6 shows a comparison of the precision-recall curve of the proposed method and these methods.

TABLE VII. COMPARISON BETWEEN PROPOSED METHOD AND PREVIOUS STUDIES FOR HELMETLESS HEAD DETECTION

Dataset	Method	AP (%)	Pre (%)	Rec (%)
JSC1	[6]	43.41	53.18	68.15
	[5]	52.45	57.50	68.15
	Proposed method	66.29	57.19	76.74
JSC2	[6]	40.47	52.27	78.59
	[5]	54.30	50.00	78.13
	Proposed method	63.95	52.68	81.82



(a) JSC1 dataset



(b) JSC2 dataset

Fig. 6. The Curve of Precision-Recall of the Proposed Method and the Previous Study.

V. CONCLUSION

This paper proposed a new feature selection algorithm called ATMBIFFOA for motorcycle classification and helmetless head detection. The experiment used two datasets with different recording angles, namely the rear and front of an object. The motorcycle classification accuracy of the proposed method reaches 96.06% for the JSC1 dataset and 96.85% for the JSC2 dataset. Meanwhile, the AP of helmetless head detection reaches 66.29% for the JSC1 dataset and 63.95% for the JSC2 dataset. The proposed algorithm is more effective than BIFFOA in terms of the number of features and the time of classification. For this reason, the proposed method is more suitable for the detection of motorcyclists who do not wear helmets in real-time. However, the addition of an adaptive two-phase mutation algorithm to BIFFOA can significantly increase the feature selection time. In the future, ATMBIFFOA can be used with faster classifiers such as KNN, SVM, and DT to reduce the time consumption of feature selection.

ACKNOWLEDGMENT

The authors would like to thank the Research Directorate of Universitas Gadjah Mada in the Recognisi Tugas Akhir (RTA) 2021 schema for funding this research.

REFERENCES

- [1] W. H. O. (WHO), "Global Status Report on Road Safety 2018," Available at: <https://apps.who.int/iris/rest/bitstreams/1164010/retrieve>, March 2021.
- [2] Sutikno, Afiahayati, and A. Harjoko, "Detection on head of motorcyclist without helmet: a review," *ICIC Express Letters*, Vol. 14, pp. 605-612, June 2020.
- [3] A. S. Azim, A. Jafri, and A. Alkhaury, "Multiclass vehicle classification across different environments," *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 12, pp. 681-691, March 2021.
- [4] V. L. Padmini, G. K. Kishore, P. Durgamalleswarao, and P. T. Sree, "Real time automatic detection of motorcyclists with and without a safety helmet," *Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020)*, Trichy, India, pp. 1251-1256, September 2020.
- [5] R. R. V. E. Silva, K. R. T. Aires, and R. M. S. Veras, "Detection of helmets on motorcyclists," *Multimedia Tools and Applications*, Vol. 77, pp. 5659-5683, March 2018.
- [6] K. Dahiya, D. Singh, and C. K. Mohan, "Automatic detection of bike-riders without helmet using surveillance videos in real-time," *Proceeding of International Joint Conference on Neural Networks (IJCNN)*, Vancouver, Canada, pp. 3046-3051, July 2016.
- [7] C. Visnu, D. Singh, C. K. Mohan, and S. Babu, "Detection of motorcyclists without helmet in videos using convolutional neural network," *Proceedings of International Joint Conf. on Neural Networks (IJCNN)*, Anchorage, USA, pp. 3036-3041, May 2017.
- [8] K. C. D. Raj, A. Chairat, V. Timtong, M. N. Dailey, and M. Ekpanyapong, "Helmet violation processing using deep learning," *Proceedings of International Workshop on Advanced Image Technology (IWAIT)*, Chiang Mai, Thailand, pp. 1-4, January 2018.
- [9] Y. Kulkarni, S. Bodkhe, A. Kamte, and A. Patil, "Automatic number plate recognition for motorcyclists riding without helmet," *Proceedings of International Conference on Current Trends toward Converging Technologies*, Coimbatore, India, pp. 1-6, March 2018.
- [10] N. Boonsirisumpun, W. Puarungroj, and P. Wairotchanaphuttha, "Automatic detector for bikers with no helmet using deep learning," *Proceedings of International Computer Science and Engineering Conference (ICSEC)*, Chiang Mai, Thailand, pp. 1-4, November 2018.
- [11] L. Shine and C.V. Jiji, "Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN," *Multimedia Tools and Applications*, Vol. 79, pp. 14179-14199, February 2020.
- [12] S. Baghdadi and N. Aboutabit, "View-independent vehicle category classification system," *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 12, pp. 756-771, July 2021.
- [13] H. K. Ragb and V. K. Asari, "Histogram of oriented phase and gradient (HOPG) descriptor for improved pedestrian detection," *Proceedings of the IS & T International Conf. on Electronic Imaging: Video Surveillance and Transportation Imaging Applications*, San Francisco, USA, pp. 1-6, February 2016.
- [14] H. Wang, D. Zhang, and Z. Miao, "Fusion of LDB and HOG for face recognition," *Proceedings of the 37th Chinese Control Conf. (CCC)*, Wuhan, China, pp. 9192-9196, July 2018.
- [15] Sutikno, A. Harjoko, and Afiahayati, "Improving detection performance of helmetless motorcyclists using the combination of HOG, HOP, and LDB descriptors," *International Journal of Intelligent Engineering & System*, Vol. 15, pp. 428-440, February 2022.
- [16] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in Biology and Medicine*, Vol. 112, pp. 103375 (1-9), September 2019.
- [17] Z. Zhu, Y. Ong, and M. Dash, "Wrapper-filter feature selection algorithm using a memetic framework," *IEEE Transactions on Systems*, Vol. 37, pp. 70-76, February 2007.
- [18] J. Zhang, Y. Xiong, and S. Min, "A new hybrid filter/wrapper algorithm for feature selection in classification," *Analytica Chimica Acta*, Vol. 1080, pp. 43-54, November 2019.
- [19] K. Bouzoubaa, Y. Taher, and B. Nsiri, "Predicting DOS-DDOS attacks: review and evaluation study of feature selection methods based on wrapper process," *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 12, pp. 131-145, May 2021.
- [20] Y. Hou, J. Li, H. Yu, and Z. Li, "BIFFOA : a novel binary improved fruit fly algorithm for feature selection," *IEEE Access*, Vol. 7, pp. 781177-81194, July 2019.
- [21] M. Abdel-Basset, D. El-Shahat, I. El-Henawy, V. H. C. Albuquerque, and S. Mirjalili, "A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection," *Expert Systems with Applications*, Vol. 139, pp. 1-14, January 2020.
- [22] A. Leelasantitham and W. Wongsere, "Detection and classification of moving thai vehicles based on traffic engineering knowledge," *Proceedings of International Conference on ITS Telecommunication*, Phuket, Thailand, pp. 27-30, October 2008.
- [23] S. Fazli, S. Muhammadi, and M. Rahmani, "Neural network based vehicle classification for intelligent traffic control," *International Journal of Software Engineering & Applications*, Vol. 3, pp. 17-22, May 2012.
- [24] T. Marayatr and P. Kumhon, "Motorcyclist's helmet wearing detection using image processing," *Advanced Materials Research*, Vol. 931-932, pp. 588-592, May 2014.
- [25] Y. Dupuis, P. Subirats, and P. Vasseur, "Robust image segmentation for overhead real-time motorbike counting," *Proceedings of International IEEE Conf. on Intelligent Transportation Systems (ITSC)*, Qingdao, China, pp. 3070-3075, October 2014.
- [26] M. Ashvini, G. Revathi, B. Yogameena, and S. Saravanaperumaal, "View invariant motorcycle detection for helmet wear analysis in intelligent traffic surveillance," *Proceedings of International Conf. on Computer Vision and Image Processing*, Roorkee, India, pp. 175-185, December 2016.
- [27] Z. Chen, T. Ellis, and S.A. Velastion, "Vehicle detection, tracking and classification in urban traffic," *Proceedings of International IEEE Conf. on Intelligent Transportation Systems*, Anchorage, USA, pp. 951-956, September 2012.
- [28] R. Silva, K. Aires, T. Santos, K. Abdala, R. Veras, and A. Soares, "Automatic detection of motorcyclists without helmet," *Proceedings of Latin America Computing Conf. (CLEI)*, Caracas, Venezuela, pp. 1-7, October 2013.
- [29] R. Waranusast, N. Bundon, and P. Pattanathaburt, "Machine vision techniques for motorcycle safety helmet detection," *Proceedings of International Conf. on Image and Vision Computing New Zealand*, Wellington, New Zealand, pp. 35-40, November 2013.

- [30] C. Huynh, T. Le, and K. Hamamoto, "Convolutional neural network for motorbike detection in dense traffic," Proceedings of the International Conf. on Communications and Electronics (ICCE), Ha-Long, Vietnam, pp. 369-374, July 2016.
- [31] A.S. Talautikar, S. Sanathanan, and C.N. Modi, "An enhanced approach for detecting helmet on motorcyclists using image processing and machine learning techniques," Advanced Computing and Communication Technologies, Vol. 702, pp 109-119, July 2019.
- [32] J. Mistry, A.K. Misraa, M. Agarwal, A. Vyas, V.M. Chudasama, and K.P. Upla, "An automatic detection of helmeted and non-helmeted motorcyclist with license plate extraction using convolutional neural network," Proceedings of Seventh International Conf. on Image Processing Theory, Tools and Applications (IPTA), Montreal, Canada, pp. 1-6, November 2017.
- [33] M. A. V. Forero, "Detection of motorcycles and use of safety helmets with an algorithm using image processing techniques and artificial intelligence models," Proceedings of Joint Conf. for Urban Mobility in the Smart City, Medellin, Colombia, pp. 1-9, April 2018.
- [34] V. Mayya and A. Nayak, "Traffic surveillance video summarization for detecting traffic rules violators using R-CNN," Proceedings of International Conf. on Computer, Communication, and Computational Sciences, Kathu, Thailand, pp. 117-126, October 2017.
- [35] Y. Xu, J. Dong, B. Zhang, and D. Xu, "Background modeling methods in video analysis: a review and comparative evaluation." CAAI Transactions on Intelligence Technology, Vol. 1, pp. 43-60, January 2016.
- [36] W. Pan, "A new fruit fly optimization algorithm: taking the financial distress model as an example," Knowledge-Based Systems, Vol. 26, pp. 69-74, February 2012.
- [37] S. Mousavi, N. Alikar, and S. Niaki, "An improved fruit fly optimization algorithm to solve the homogeneous fuzzy series-parallel redundancy allocation problem under discount strategies," Soft Computing, Vol. 20, pp. 2281-2307, June 2016.
- [38] R. Padilla, S. L. Netto, and E. A. B. D. Silva, "A survey on performance metrics for object-detection algorithms," Proceedings of the International Conference on Systems, Signals, and Image Processing (IWSSIP), Niteroi, Brazil, pp. 237-242, July 2020.
- [39] X. Chen, Z. Liu, and Z. Zhang, "The measurement of planning surface roughness by neural networks based on image," Proceedings of Sixth International Conference on Natural Computation, Yantai, China, pp. 705-708, August 2010.
- [40] H. Mustafidah, S. Hartati, R. Wardoyo, and A. Harjoko, "Selection of most appropriate backpropagation training algorithm in data pattern recognition," International Journal of Computer Trends and Technology (IJCTT), Vol. 14, pp. 92-95, September 2014.