



# Helmet Detection System for Motorcycle Riders with Explainable Artificial Intelligence Using Convolutional Neural Network and Grad-CAM

Suradej Intagorn<sup>1</sup>, Suriya Pinitkan<sup>1(✉)</sup>, Mathuros Panmuang<sup>2</sup>,  
and Chonnikarn Rodmorn<sup>3</sup>

- <sup>1</sup> Department of Mathematics, Statistics, and Computer Science, Faculty of Liberal Arts and Science, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand  
suriya.p@ku.th
- <sup>2</sup> Department of Educational Technology and Communications, Faculty of Technical Education, Rajamangala University of Technology, Thanyaburi, Pathum Thani, Thailand
- <sup>3</sup> Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

**Abstract.** Motorcycle accidents are one of the most common causes of injury and death in road users. This research has applied convolutional neural network (CNN) and explainable AI to detect motorcyclist without helmet and explain why CNN made that decision. The concept is based on deep learning and CNN principles applied to automatically detect motorcycle riders without helmet from images using three baseline classifiers: support vector machine, random forest and logistic regression to compare with deep convolutional neural network and measure accuracy with accuracy and F1-Score. The results revealed that CNN's F1-Score was 0.8326, which was the highest among all predictive models.

**Keywords:** Helmet · Explainable AI · Convolutional neural network

## 1 Introduction

Road traffic accidents are a serious problem faced by all countries, especially in urban areas where traffic congestion is a major cause of injury and loss of life and property. Motorcycle accidents are among the most common causes of injury and death among motorcyclists [1] According to the WHO 2018 report, Thailand has the highest road fatality rate in Southeast Asia at 32.7 per 100,000 population [2] This is mainly due to motorcyclists not wearing helmets and violating road signs or signals. This cause can be controlled or prevented if government agencies are strict and have systems that can be used to help detect offenders.

Currently, there are work to install warning signs and CCTV cameras around urban areas, especially in Bangkok where there are more than 60,000 CCTV cameras and the goal in the future will increase to 200,000 [3] However, no system has been developed that can access CCTV images for processing in various fields, including detecting offenders without helmets. Therefore, preventive and corrective actions should be taken. One way

to prevent and help reduce the rate of injuries and fatalities of motorcyclists is to wear a helmet while riding or riding a motorcycle. It can be seen that wearing a helmet is an effective way to reduce mortality and reduce the likelihood of head injury. According to research on the system for identifying the use of a helmet using the YOLO V5 and V3 technique, the improved algorithm's detection accuracy increased by 2.44% with the same detection rate as compared to the traditional YOLO V3 in the test set. The development of this model will improve helmet detection and guarantee safe building, which has significant practical implications. [4] From the YOLO V5 experiment findings, YOLOv5s can detect objects at an average speed of 110 frames per second. Meet all real-time detection demands. The mAP of YOLOv5x reaches 94.7% using the trainable target detector's pre-training weight, demonstrating the potency of the helmet detection-based YOLOv5. [5] Other image processing methods will be used in this study, though, to get distinct outcomes.

Based on the aforementioned problems, this research applied Image Processing techniques to process images and applied Deep Learning and Convolution Neural Network (CNN) techniques. CNN, despite its good performance, has a complex internal structure and cannot be understood by humans. Thus, this research has applied the principles of Explainable AI to help machine learning users understand the reasons why the system makes such predictions, especially when the prediction system fails. The approaches utilized are aimed at allowing the machine to learn and recognize features so that it can react quickly when it discovers someone who is not wearing a helmet and alerts the appropriate authorities in real time. As a result, this research will benefit law enforcement and reduce accidents to some extent.

## 2 Related Works

### 2.1 Helmet Detection

Motorcycle helmetless detection system research more studies are available. Since the YOLO family of algorithms, which have extremely high precision and speed, have been employed in several scene detection tasks, the majority of them are utilizing the YOLO technique. We suggest a safety helmet detection technique to create a digital safety helmet monitoring system, similar to the work of Zhou, Zhao, and Nie [5], who used a YOLOv5 model with various parameters for training and testing. The comparison and analysis of the four models. According to experimental findings, YOLOv5s's average detection speed is 110 FPS. Meet all real-time detection criteria in full. The mAP of YOLOv5x reaches 94.7% using the trainable target detector's pre-training weight, demonstrating the potency of the helmet detection-based YOLOv5. According to Dasgupta, Bandyopadhyay, and Chatterji's research [1], CNN approaches are also employed for detection. The foundation for identifying motorcycle riders who don't wear helmets is suggested in this research. The state-of-the-art method for object recognition, YOLO model, and its incremental version, YOLOv3, are used in the proposed approach's initial step to detect motorbike riders. A Convolutional Neural Network (CNN) based architecture has been suggested for the second stage of motorbike rider helmet recognition. In compared to other CNN-based techniques, the suggested model's evaluation on traffic recordings yielded encouraging results.

## 2.2 Deep Learning and Convolution Neural Network

Today, neural networks are employed in a variety of applications because they provide the notion of quick learning and can also enable deep learning approaches to reach the best validity through machine training. Based on research by Stefanie, Oliver and Frieder [6] who studied machine learning to achieve rapid classification using neural network techniques to classify Speech Recognition and Car Driving Maneuvers. In addition, Traore, Kamsu-Foguem and Tangara [7] also presented applications of deep convolution neural network (CNN) for image recognition. The results showed that on future microscopes, the categorization process could be integrated into a mobile computing solution. In pathogens diagnosis, CNN can increase the accuracy of hand-tuned feature extraction, implying some human errors. With 200 *Vibrio cholera* photos and 200 *Plasmodium falciparum* images for the training dataset and 80 images for testing data, the CNN model obtained a classification accuracy of 94%. Based on the concepts of Deep Learning and CNN in this research, image processing was applied to detect behavior of non-helmeted motorcyclists.

## 2.3 Histograms of Oriented Gradient (HOG)

HOG is represented by Dalal and Triggs [8]. It is often used in object detection and classification problems. HOG calculates the magnitude and gradient vector for each pixel, then generates a histogram for further classification features. HOG is used in a number of problems, such as its use as an extraction feature for the problem of classifying rice types in [9].

## 2.4 Object Detection

Detecting the behavior of non-helmet motorcyclists rely on the principle of Object Detection, since the motorcyclist is one of the objects in the image. Object Detection is a computer technology that uses the principles of Computer Vision and Image Processing used in AI (Artificial Intelligence) to detect a specific type of object. In general, the objective of Object Detection is at finding and classifying actual items in a single image and labeling them with rectangles to demonstrate the certainty of their existence. [10] A large number of studies have now applied the principles of Object Detection to detect objects such as research by Thipsanthia, Chamchong and Songram [11] that have applied the YOLOv3 technique to detecting and recognizing Thai traffic signs in real-time environments. The dataset was designed and distributed for existing traffic detection and recognition. With 50 classes of road signs and 200 badges in each class, a total of 9,357 images are compared across two architectures (YOLOv3 and YOLOv3 Tiny). The experiment demonstrates that YOLOv3's mean average precision (mAP) is better than YOLOv3 Tiny's (80.84%), while YOLOv3's speed is marginally better than YOLOv3's.

## 2.5 Convolutional Neural Network

Convolutional Neural Network simulates the human vision of space in small parts and merges groups of spaces together to see what is being seen. An application of CNN

to image classification found that if able to greatly enhance many CNN models On ImageNet, for example, improve ResNet-50's top-1 validation accuracy from 75.3% to 79.29%. Therefore, if it is improved image classification accuracy leads to improved transfer learning performance in other application domains such as object identification and semantic segmentation, according to the researchers [12]. In addition, Deep Convolutional Neural Networks were used to classify rice cultivars using Image Classification techniques. Fifteen hundred rice cultivars were chosen for the experiment in the photographic separation of paddy cultivars, and three Classification Algorithms methodologies were employed to compare classification efficiency and alter parameters. The results of the experiments and testing of the model performance showed that the VGG16 model had the highest accuracy of 85% [13]. Therefore, the research had to adjust various parameters to be appropriate and able to recognize the image as accurately as possible.

## 2.6 Explainable AI

Explainable AI is a concept that requires the machine to have a process of understanding so that the result can be explained and understood. In other words, they want people to understand the idea of a machine, or for a machine to think and explain human language. According to research by Pawar, O'Shea, Rea and O'Reilly [14], Explainable AI (XAI) is a field where strategies are created to explain AI system predictions and has applied XAI to employ to analyze and diagnose health data, as well as a potential approach for establishing responsibility. In the field of healthcare, transparency, outcome tracing, and model improvement are all important.

## 2.7 Grad-CAM

Grad-CAM: is to visualize what the model sees, such as checking the grain of rice to see what type of rice it is. The model serves to make predictions, which, in principle, is necessary to understand where the model is considering the correct point or not. Grad-CAM computes a heat-map  $g \in \mathbb{R}^{n \times m}$ . It shows which parts of the input image are highlighted  $x \in \mathbb{R}^{N \times M}$  have mostly influenced the classifier score in favor of the class  $c$  (upper-case letters indicate sizes that are larger than lowercase one). Let  $y_c$  denote the score the class  $c$  and  $a^k \in \mathbb{R}^{n \times m}$ ,  $k = 1, \dots, K$ , the activation maps matching to the last convolutional layer's  $k$ -th filter. A weighted average of  $a^k$ ,  $k = 1, \dots, K$ , is used to establish the Grad-CAM he class  $c$ . followed by a ReLU activation: [15]

$$g_c = \text{ReLU}\left(\sum_k \alpha_c^k a^k\right), \quad (1)$$

where the importance weights  $\{\alpha_c^k\}$  are defined as the average derivatives of  $y_c$  with respect to each pixel  $(i, j)$  in the activation  $a^k$ :

$$\alpha_c^k = \frac{1}{nm} \sum_i \sum_j \frac{\partial y_c}{\partial a^k(i, j)} \quad (2)$$

Several researches uses Grad-CAM to verify results and make the model more interpretable such as breast cancer classification [16].

### 3 Methodology

#### 3.1 Data Collection and Preprocessing

The data was collected in 24 video files from mobile phone cameras from the roadside. Each video file has a different camera angle and location. The video file is preprocessed into an image file as shown in Fig. 1 in the first row. These video files were then preprocessed using a ready-made library called ImageAI for object detection of people and motorcycles as shown in Fig. 1 in the second row. The object detection model is YOLOv3. Two custom object classes are person and motorcycle. Then automatically crop the image of the motorcyclist. If there is more than one motorcyclist in the picture, the system will extract the image according to the number of riders. The image of the motorcyclist that has been extracted is labeled Helmet, NoHelmet or UnDetermined i.e. wearing a helmet, not wearing a helmet and unable to tell. Preprocessing from ImageAI sometimes has extraction errors, i.e., motorcyclist's head is not visible, so we define them as the UnDetermined class.



**Fig. 1.** Sample images in the dataset of different classes.

In the video file, the number of helmet-wearers outnumbered the non-wearers, but in this dataset, the number of riders in the 2 classes was chosen to be approximately same. Each rider is randomly selected 1–3 images at different distances and angles.

This data is randomly divided into two parts, the training set and the test set. The training set will be 85% of the total data and the test set to 15% of the total data set. Training set is used to teach the classifier while test set is used to test the performance of the classifier Table 1 summarizes the amount of data in each class.

**Table 1.** The amount of images in the dataset.

Class	Training set (85%)	Test set (15%)	Total (100%)
Helmet	1097	194	1291
NoHelmet	1077	191	1268
UnDetermined	728	129	857

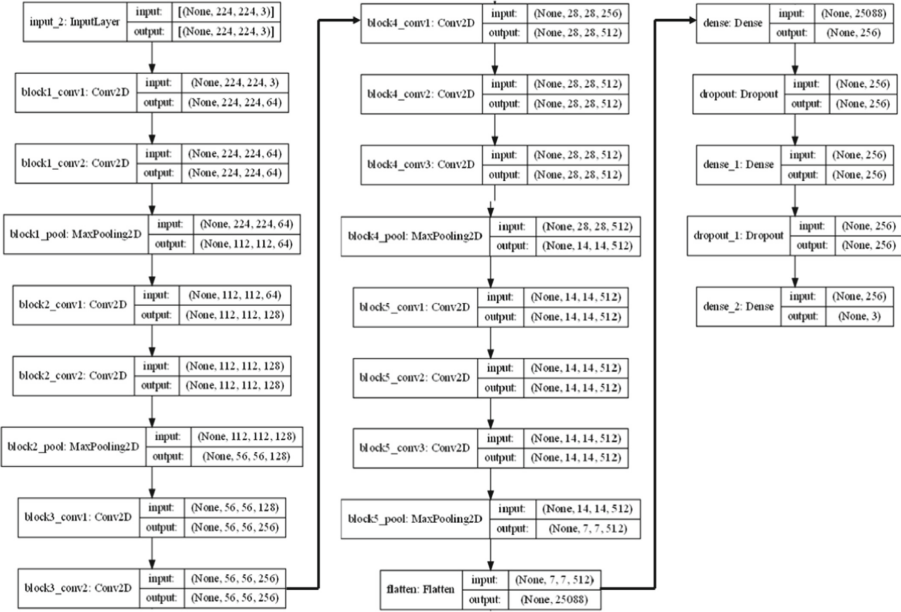
### 3.2 Deep Convolution Neural Network

Training deep neural network from scratch requires large amounts of data to avoid overfitting problem which may not be applicable to our small dataset [17]. Thus, we use a technique called transfer learning which borrows some layers from well-trained deep neural network such as [18]. The primary purpose of transfer learning is to reduce the amount of learning parameters, which in turn reduces the amount of learning data that doesn't cause overfitting.

The borrowed layers are from a CNN called VGG16 [18]. Our network borrows the first layer to the Block5\_pool layer which are convolution and pulling layers from VGG16. The main purpose of this layer set is to extract the visual features of images. In-depth details of VGG16 can be found in [18].

We used a total of two dense layers, 256 neurons each, with a ReLU (Rectified Linear Unit) activation function and a L1 regularization = 0.001. The first dropout layer is added between these two dense layers with a probability of drop out = 0.5. The regularization and dropout layer are used to reduce overfitting. Parameters of these dense layers are optimized from the training set. The main purpose of these layers is to combine visual features extracted from the previous layer with appropriate weights and be able to learn nonlinear decision boundaries for helmet detection.

The last part is also dense layer, but the activation function is softmax. The purpose of this layer is to predict the probabilities in each category. The final output is the class with the highest probability. This layer has the same number of neurons as the image class, which is 3 (Helmet, NoHelmet and UnDetermined). The deep neural network structure is visualized by using Keras utility as shown in Fig. 2.



**Fig. 2.** The deep neural network structure: The layers in the first and second columns are pre-trained layers from VGG16. The layers in the third column are trained by the training set detailed in Table 1.

## 4 Experiment Setup and Results

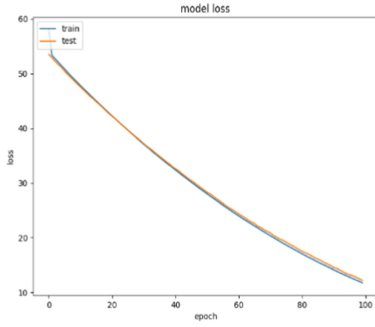
We benchmark classifiers based on accuracy and macro f1-score metrics, with the deep convolutional neural network compared to three baselines: support vector machine, random forest and logistic regression. The data for each class used is relatively balanced, except for UnDetermined which is lower than the other two classes. The data used for testing were data from the test set which was 15% of the total data set.

The deep convolutional neural network in Fig. 2 uses pretrained layers from VGG16 [18] for feature extraction and trains only the classification layers. Therefore, the image will be resized to 224x224 due to make it compatible with the dimensions of VGG16.

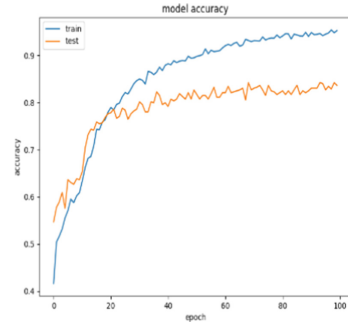
The learning curve of the CNN model are shown in Fig. 3 and Fig. 4

All preprocessing steps are same for both the proposed model and baseline except size of image and the feature extraction. The algorithm for feature extraction of all baselines is Histograms of Oriented Gradient (HOG) descriptors, images for baselines are resized to 64x128 (64 pixels wide and 128 pixels tall) according to the original paper [8]. The hyper parameters for the other HOGs are also same as the default detector in [8] (9 orientation bins  $16 \times 16$  pixel blocks of four  $8 \times 8$  pixel cells). Each image is converted to a HOG feature vector of size 3780.

All baseline classifiers and evaluations are used from the sklearn library [19]. All hyper parameters use the default library value except random\_state which is all set to 0 for reproducible result (Tables 4, 5, and 6).



**Fig. 3.** The loss curve of the CNN model



**Fig. 4.** The accuracy curve of the CNN model

**Table 2.** Accuracy, precision, recall and F-Score of classifiers.

Classifiers	Accuracy	Precision	Recall	F-Score
Deep Convolutional Network	<b>0.8365</b>	<b>0.831</b>	<b>0.8347</b>	<b>0.8326</b>
Logistic Regression	0.6945	0.6982	0.6997	0.6982
Support Vector Machine	0.7645	0.7748	0.7656	0.7695
Random Forest	0.6439	0.6872	0.6284	0.6417

**Table 3.** Confusion matrix of deep convolutional network.

True Label	Predicted Label			
	Label name	Helmet	NoHelmet	UnDetermined
Helmet		169	14	11
NoHelmet		17	155	19
UnDetermined		5	18	106

**Table 4.** Confusion matrix of logistic regression.

True Label	Predicted Label			
	Label name	Helmet	NoHelmet	UnDetermined
Helmet		140	43	11
NoHelmet		46	121	24
UnDetermined		8	25	96



**Table 5.** Confusion matrix of support vector machine.

True Label	Predicted Label			
	Label name	Helmet	NoHelmet	UnDetermined
Helmet	150	38	6	
NoHelmet	34	143	14	
UnDetermined	5	24	100	

**Table 6.** Confusion matrix of random forest.

True Label	Predicted Label			
	Label name	Helmet	NoHelmet	UnDetermined
Helmet	140	51	3	
NoHelmet	56	126	9	
UnDetermined	22	42	65	

Table 2 summarizes the accuracy and f1score of each classifiers, with Deep convolutional neural network predicting the most accurate: accuracy = 0.8365 and F1-score = 0.8326, SVM is the highest-performing baseline for, with accuracy = 0.7645 and F1-score = 0.7695.

The confusion matrix shows the number of images predicted in each class for each true label. For example, the first row in the second table states that there are 194 images in the Helmet class in the test set (169 + 14 + 11). The Deep Convolutional Neural Network (CNN) correctly predicted 169 images to Helmet class, but incorrectly predicted 14 images to NoHelmet class and 11 images to UnDetermined class. Therefore, the values on the diagonal line are the number of images correctly predicted and the values where the other position is guessing wrong. The values in Table 3 show that the number of Helmet images that CNN incorrectly predict as NoHelmet and Undertermined are approximately same which are 14 and 11 respectively. The result of NoHelmet images is same tendency with Helmet images. For UnDetermined images, the number of incorrect prediction as NoHelmet is more than as Helmet approximately three times which are 18 and 5 respectively.

#### 4.1 Visualization and Explainable AI






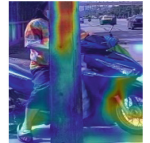



Although predictive accuracy is one of the most important aspects of classifiers, it is also important to understand why classifiers make such decisions. Accuracy = 0.8326 cannot be said whether the classifiers' decisions are reasonable or not. Although there are previous works about automated helmet detection using CNN such as [1], their system cannot give explanation about why their system make such decisions.

Deep convolutional network and three baselines make decisions based on high dimensional features that are difficult to understand for humans. In practical implementation, if we find that the classifier is making the wrong decision, the system should be able to tell the reason why it made that decision so that the developer can provide additional tutorial examples in case the decision is wrong.

Deep convolutional neural network, in addition to high accuracy, can also perform visualization, such as an algorithm called Grad-Cam [20] The result of the algorithm is to give numbers representing the importance of each pixel for predicted class. We visualize these number of those pixels in a heatmap and overlay on the input image.

Grad-Cam results should be consistent with human understanding. For example, to determine whether a helmet is being worn or not, the numbers representing importance around the helmet should be high. The Jet colormap technique for color representation of the weight numbers. Therefore, the important regions that CNN focus in the image has highlighted in red tint (Table 7).

**Table 7.** Grad-cam visualization for deep convolutional neural network.

True Label	Label name	Predicted Label		
		Helmet	NoHelmet	UnDetermined
Helmet	Helmet			
	NoHelmet			
	UnDetermined			

In the first row and first column, CNN puts the correct highlight on the helmet area. In the second row and first column, CNN highlight around head area. However, the motorcyclist worn a cap, not a helmet, so the prediction is incorrect. In the third row and first column, we cannot see the head of the riders, so it is labeled as UnDetermined, but CNN probably assumed that the storage box was a helmet and therefore predicted it as a Helmet. Therefore, CNN in this experiment accurately focuses on the helmet pixels for most images in the Helmet class, but still confuses objects with helmet-like ones, such as caps and motorcycle trunks. Therefore, adding these examples, such as a person wearing a cap, to the training set may improve accuracy in these cases.

Predicting the NoHelmet class, CNN correctly focus the head without helmet area. However, many of the images show that CNN looked at not just the head but also the rider's skin, possibly because in this dataset, the non-helmeted photos tend to be wearing short sleeves and shorts. It seems that CNN instead of just making predictions based on the riders' head without helmet, is also examining the skin pixel in the arms and legs. Therefore, although the prediction accuracy of class NoHelmet is 81% ( $155/(17 + 155 + 19)$ ), but qualitatively, we find that CNN's pixel focus is still inaccurate. Therefore, adding the samples, for example, a person wearing a helmet but wearing a short sleeve shirt and shorts to the class helmet, for example, should help CNN to better focus in the NoHelmet images.

In UnDetermined class, since the the rider's head is not visible, CNN focuses on other parts of the image such as the bike or the scene. If there is an item close to the helmet, the total weight of the neuron used to predict the Helmet class will be higher, causing the prediction to be the Helmet class instead such as the example in the third row and first column.

## 5 Conclusion

Motorcycle accidents are one of the most common causes of injury and death in road users. This is mainly due to motorcyclists not wearing helmets and violating traffic signs or signals. Based on the aforementioned problems, this research has applied Convolution Neural Network (CNN) and Grad-Cam techniques. Three baseline classifiers have been used which are support vector machine, random forest and logistic regression to compare with deep convolutional neural network. The evaluation metrics are accuracy and F1-Score. The results of the research revealed that CNN's F1-Score = 0.8326, Logistic Regression = 0.6989, Support Vector Machine = 0.7695 and Random Forest = 0.6417. The highest predictive model accuracy was CNN. Grad-CAM is also used to determined where the CNN is looking in the input image which makes the model more interpretable.

## References

1. Dasgupta, M., Bandyopadhyay, O., Chatterji, S.: Automated helmet detection for multiple motorcycle riders using CNN. In: Conference on Information and Communication Technology (CICT), pp.1–4. IEEE, Banff, AB, Canada (2019)
2. World Health Organization: Global status report on road safety 2018. Geneva, France (2018)
3. Bangkok Metropolitan Administration (BMA): Project for studying and analyzing the enterprise architecture in information technology for working processes according to the main mission of Bangkok, phase 1. BMA, Bangkok (2021)
4. Wu, F., Jin, G., Gao, M., HE, Z., and Yang, Y.: Helmet detection based on improved YOLO V3 deep model. In: 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), pp. 363–368. IEEE, Banff, AB, Canada (2019)
5. Zhou, F., Zhao, H., and Nie, Z.: Safety helmet detection based on YOLOv5, In: 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), pp. 6–11. IEEE, Shenyang, China (2021)

6. Krause, S., Otto, O., Stolzenburg, F.: Fast Classification Learning with Neural Networks and Conceptors for Speech Recognition and Car Driving Maneuvers. In: Chomphuwiset, P., Kim, J., Pawara, P. (eds.) MIWAI 2021. LNCS (LNAI), vol. 12832, pp. 45–57. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-80253-0\\_5](https://doi.org/10.1007/978-3-030-80253-0_5)
7. Traore, B.B., Kamsu-Foguem, B., Tangara, F.: Deep convolution neural network for image recognition. *Eco. Inform.* **48**, 257–268 (2018)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). vol. 1, pp. 886–893. Ieee (2005)
9. Nguyen-Quoc, H., Hoang, V.T.: Rice seed image classification based on hog descriptor with missing values imputation. *Telkomnika* **18**(4), 1897–1903 (2020)
10. Zhao, Z.-Q., Zheng, P., Xu, S.-T., Wu, X.: Object detection with deep learning: a review. *IEEE Trans. Neural Netw. Learn. Sys.* **30**(11), 3212–3232 (2019)
11. Thipsanthia, P., Chamchong, R., Songram, P.: Road Sign Detection and Recognition of Thai Traffic Based on YOLOv3. In: Chamchong, R., Wong, K.W. (eds.) MIWAI 2019. LNCS (LNAI), vol. 11909, pp. 271–279. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-33709-4\\_25](https://doi.org/10.1007/978-3-030-33709-4_25)
12. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M.: Bag of tricks for image classification with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), California, pp. 558–567 (2019)
13. Panmuang, M., Rodmorn, C., Pinitkan, S.: Image processing for classification of rice varieties with deep convolutional neural networks. In: 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP), pp. 1–6, Ayutthaya, Thailand (2021)
14. Pawar, U., O'Shea, D., Rea, S., and O'Reilly, R.: Explainable AI in Healthcare. In: International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), 1–2 Dublin, Ireland (2020)
15. Morbidelli, P., Carrera, D., Rossi, B., Fragneto, P., and Boracchi, G.: Augmented Grad-CAM: Heat-maps super resolution through augmentation. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4067–4071, Barcelona, Spain (2020)
16. Masud, M., Eldin Rashed, A.E., Hossain, M.S.: Convolutional neural network-based models for diagnosis of breast cancer. *Neural Computing and Applications* pp. 1–12 (2020)
17. Xia, S., et al.: Transferring ensemble representations using deep convolutional neural networks for small-scale image classification. *IEEE Access* **7**, 168175–168186 (2019)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)* (2014)
19. Pedregosa, F., et al.: Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
20. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Gradcam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)