Machine Learning and Artificial Intelligence
J.-L. Kim (Ed.)

doi:10.3233/FAIA230765

Apple Disease Recognition Based on Comprehensive Application of Residual Neural Network

Xiedong SONG[a, b,c,1], Vladimir Y. MARIANO[c], Xuemei SHI[d]
[a] *School of Mathematics and Computer Application Technology, Jining University, Qufu, Shandong, 273155, China*
[b] *Information Engineering Department, Yantai Vocational College, Yantai, Shandong, 264003, China*
[c] *College of Computing and Information Technologies, National University, Manila, 0900, Philippines*
[d] *School of Computer Science, Huainan Normal University, Huainan, Anhui, 232038, China*

Abstract. China is the world's largest apple growing country; Apples are affected by various diseases during the planting process, and timely detection of fruit tree diseases is of great significance for improving apple yield. With the development of artificial intelligence technology, The use of deep learning to detect the condition of fruit trees has become a research hotspot in forest science planting.ResNet50 is used to solve the bottleneck problem of recognition accuracy of traditional neural networks, and data enhancement is used to improve the quality of data samples. transfer learning is used to solve the application of small-scale data samples, and loss function is used to balance sample errors. Improve the ability to obtain important feature information by using a dual channel mixed attention mechanism. Finally, a comprehensive network model combining residual neural network, dual channel mixed attention mechanism, transfer learning and loss function is designed. The experimental results show that the effect of this comprehensive ResNet50 model is better than that of the original ResNet50 model and VGG19 model, and its classification accuracy rate can reach 96.87%. This method improves the performance in fruit disease detection, and has wider applicability in practical applications.

Keywords. Disease recognition of apple, ResNet50, CBAM, Data enhancement, Transfer learning, Loss function

1. Introduction

As a country with a large population and a long history of agricultural and forestry planting, China has been committed to solving the problem of food supply, including the promotion of fruit planting, of which apple is the most widely planted representative fruit in China. However, the invasion of diseases and pests seriously

1 Corresponding Author: Xiedong Song, a School of Mathematics and Computer Application Technology, Jining University, Qufu, Shandong, 273155, China; b Information Engineering Department, Yantai Vocational College, Yantai, Shandong, 264003, China; c College of Computing and Information Technologies, National University, Manila, 0900, Philippines; sxdmail@163.com.

affects the growth of fruit trees [1]. Therefore, it is of great significance to quickly identify and scientifically control pests in time to improve apple yield. In the past, the monitoring process of fruit tree disease mainly relied on the vision and experience of planting personnel for identification and judgment, which was slow to recognize and strong subjectivity, it has been unable to meet the needs of modern agricultural production. Pattern recognition technology has increasingly found its way into the realm of fruit cultivation and production. It has led to the rise of forestry digitization, and contribute to the enhancement of the contemporary fruit industry's growth.

Deep learning technology can effectively avoid subjectivity when manually extracting features, and has significant portability, so it has been widely applied in target classification and detection tasks [2]. This study is to use the deep learning technology to detect the common diseases of apple trees.

Because the early diseases of apple mostly occur in the leaves, This study primarily focuses on detecting leaves. By combining the ResNet50 model and an attention mechanism, the model's calculations yield greater accuracy than traditional neural networks. This enhances the practicality of the target detection algorithm for identifying diseases in fruit trees.

2. Related works

In recent years, the deep learning has gradually become a research hotspot for object recognition. This method efficiently tackles drawbacks of typical computer vision techniques, like uncertain manual feature selection. It shows better accuracy than traditional algorithms on identical image datasets [3]. For example, Bin Li et al. used ResNet50 to enhance facial expression recognition [4]. Md Salaudin Khan et al. used eight of the most popular classifiers to compare them, and it is found that XGB classifier outperforms other classifiers with unbalanced distribution in different categories in terms of good product selection [5]. Imran Iqbal et al. proposed a convolutional neural network structure incorporating transfer learning for the detection of human knee synovial fluid. Achieving high sensitivity, accuracy, and precision [6].

Scholars around the world have also conducted a large amount of research on using deep learning to identify diseases, and have modified some classic convolutional neural networks to make them more suitable for the problems to be solved, and have achieved good results. For example, Mideth Abisado and colleagues proposed a Harris Hawk optimized recurrent neural network for IoT disease detection [7]. In the optimization of training sets, the cassava disease detection captured by Ramcharan et al. in Tanzania can achieve the highest accuracy through training using DCNN and transfer learning [8]. In terms of improving the accuracy of crop detection, Alfaris sy et al. selected the Caffe framework for processing in the rice field pest classification experiment in Indonesia, and the accuracy of the experimental results reached 87% [9]. In order to accurately locate pest images, Li Weiliu proposed a method that combines CNN with Non-Maximum Suppression (NMS) Regional Suggestion Network (RPN) to eliminate overlapping detection and check the malady number in image [10].

Compared with traditional digital image processing recognition methods, deep learning-based methods automatically acquire features and achieve better classification accuracy, avoiding artificially designed feature localization.

3. The research method

3.1. VggNet

The VGGNet network was constructed and trained by Oxford Visual Geometry Group Laboratory at the University of Oxford. It is an effective network model for object recognition in convolutional neural networks. The overall architecture of VGGNet and AlexNet is similar, with the following improvements compared to AlexNet [11]:

(1) One improvement of VGGNet compared to AlexNet is the use of consecutive 3x3 convolutional kernels to replace the larger convolutional kernels in AlexNet.

(2) VGGNet's structure is simple, with consistent use of 3x3 convolutional kernels and 2x2 maximum pooling throughout.

This paper intends to use VGG19 as one of the working models for the experiment.

3.2. ResNet

The ResNet (Residual Neural Network) represents an optimization enhancement to the conventional CNN. The layer count in CNN stacking is determined by the program's requirements, the more layers of the stacked network, the richer the obtained feature information. However, when the hidden layers in traditional neural networks are stacked to a certain number, their computational accuracy reaches saturation and then begins to sharply decline, becoming increasingly difficult to train, and there is a high possibility of drawbacks such as gradient explosion and gradient vanishing [12]. The conventional method is to initialize the data or add BN blocks. Although this can solve the gradient problem, as the network deepens, another problem will arise - network degradation. For the sake of solve this problem, it was decided to use residual thinking to solve this degradation problem, while also appropriately alleviating the problem of gradient vanishing and gradient explosion, and improving the model performance [13].

According to the theory of residual neural networks, if only shallow output data is input as identity mapping to deep layers, even with an increase in the layer count within the network model, it also does not lead to network degradation. Therefore, whether to have residual modules becomes an important difference between traditional neural networks and ResNet. The residual module consists of an identity map and a residual component. Through this "short circuit connection" between the two layers, it helps to reverse the propagation of gradients during the training process and suppress network degradation [14].

The residual module in this paper uses the Bottleneck structure. There are also two main forms: a residual structure represented by a solid line and another by a dotted line. The reason why a 1x1 conv is added to the identity map branch is because the channels of the input module and the output module are not equal in some layers of the ResNet50 structure, so pixel level addition operations cannot be performed. Therefore, a 1x1 conv is added to the identity map branch to adjust the input and output channels.

This paper selects ResNet50 as the main component of the network model.

3.3. Attention mechanism

Attention mechanism has been introduced into many fields of deep learning in recent years, which is a useful supplement to neural networks. It originates from the research and simulation of human vision. Because the brain has certain bottlenecks in processing information of various complex contents, humans will selectively focus on some interesting or important information observed, and automatically ignore other information [15]. According to different attention goals, The attention mechanism can be categorized into two types: SAM (spatial attention mechanism) and CAM (channel attention mechanism). In the SAM module, the dimensions of the channel are compressed, but the spatial dimensions do not change, so the main focus lies in the spatial positioning of crucial information within the feature map. In the CAM module, the channel dimension remains unchanged and the space dimension is compressed. This module centers on the pertinent information shared among channels within every feature map [16].

The advantages and disadvantages of both attention mechanisms are evident. SAM has the advantage of enabling fast location and recognition, but it has the disadvantage of overemphasizing location importance while disregarding the information within the channel. This results in compressing all channels into a single channel, resulting in a relatively intricate feature map that complicates the identification of essential information. On the other hand, CAM has the advantage of strengthening the correlation between each channel and appropriately allocating resources to each channel. However, its disadvantage lies in potentially ignoring feature location information and compressing the feature map into 1x1 channel information.

To enhance the performance of CNN models in image classification tasks, this paper proposes a combination of the benefits of both attention mechanisms. It introduces a mixed attention mechanism called the Convolutional Block Attention Module (CBAM), which effectively distributes both location and channel information. CBAM comprises two sub-modules: CAM, which focuses on channel attention, and SAM, which focuses on spatial attention. This approach not only reduces computational overhead and parameter count but also allows for easy integration into the corresponding training model through plug-and-play functionality [17]. The structure of the CBAM is shown in Figure 1.

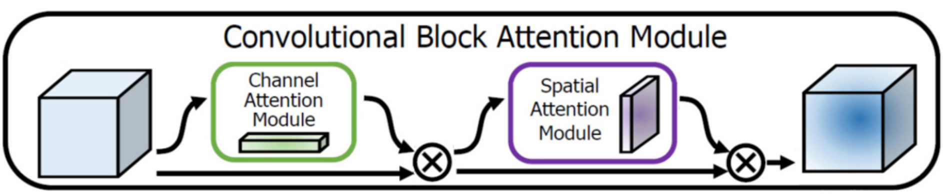

Figure 1. The CBAM architecture.

To attain the reduction of insignificant feature channels and the amplification of crucial feature channels within experimental images. As shown in Figure 2.

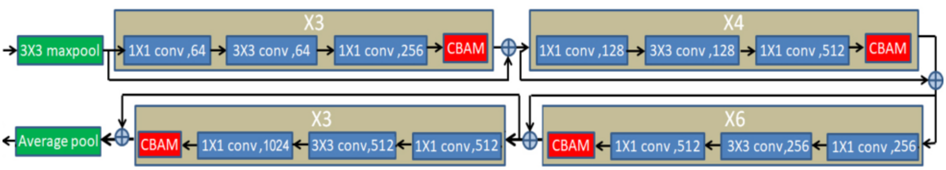

Figure 2. The Workflow of CBAM-ResNet.

3.4. Learning algorithms

In the training process of image classification, selecting the correct learning algorithm is often as important as selecting the network architecture when. Different learning algorithms are suitable for different situations. Common learning algorithms include SGDNesterov, RMSProp, AdaGrad, Adam, AdaDelta, Nadam and AdaMax. According to the research of Imran Iqbal and Gbenga Abiodun Odesanmi et al., the AdaGrad optimizer sets the learning rate to 0.0008, the number of layers to 50, the small batch size to 32, and the number of epochs to 70, The accuracy of AdaGrad algorithm on the test set is 95.9%, which is superior to other algorithms [18].

AdaGrad is a gradient descent optimization method with adaptive learning rate. It adaptively adjusts the learning rate of parameters, performing smaller updates on those parameters that are frequently used, and larger updates on those parameters that are not frequently used. Therefore, it can handle sparse data well, thereby improving the robustness of SGD. AdaGrad dynamically adapts the learning rate throughout iterations, guaranteeing individualized learning rates for each parameter within the objective function. The AdaGrad formula is:

$$G^k = \sum_{i=1}^{k} g^i \odot g^i \tag{1}$$

The AdaGrad algorithm is based on the random gradient descent method and adjusts the step size in different component directions by recording the cumulative gradient of each component. Combining the AdaGrad algorithm with the VGG19 model requires optimizing the parameters of VGG19.By combining the AdaGrad algorithm with the VGG19 model, the learning efficiency of parameters can be improved, especially for problems with sparse gradients.

3.5. Loss function

The function used to calculate the difference between the real result and the predicted result of each training in the model is called the loss function. It is a key part of optimizing the machine learning effect and can be used to measure the quality of each prediction made by the model [19]. The smaller the difference value obtained, the smaller the training loss, and the better the robustness of the learning model.

In practical applications, the selection of loss functions is constrained by many factors, such as whether there are outliers, the selection of machine learning algorithms, the time complexity of gradient descent, the difficulty of obtaining derivatives, and the confidence level of predicted values. Therefore, there is no loss function suitable for processing all types of data. The choice of loss function in the experiment mainly depends on the type of input tag data: if the input tag is an unbounded, real value, the square difference should be preferred; If the input tag is a classification flag or bit vector, cross entropy will be more appropriate. The main function used in this experiment is Cross Entropy Loss Function. The Focus Loss formula is:

$$L = \frac{1}{N}\sum_{i} L_i = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic}\log(p_{ic}) \tag{2}$$

The entire model prediction, loss acquisition, and learning process:

(1) The final layer of the neural network obtains scores for each category;

(2) The probability output of this score is obtained through the sigmoid function;

(3) The loss between the predicted category output and the real category output of the functional is calculated through the cross-entropy loss, and adjustments are made.

3.6. Transfer learning

Transfer learning is a technique useful in deep learning computations. Its purpose is to employ previously gained knowledge in different contexts. This involves repurposing a model developed for Task A to assist in creating a model for Task B. The core idea is to identify parallels between the initial and target tasks. From this resemblance, transfer learning leverages beneficial parameter transfers from the prior model to the new one. [20].

In this experiment, the backbone network undertakes the task of extracting image features. Therefore, it is necessary to first perform pre training on the ImageNet dataset to find an appropriate number of feature and parameter data. Then, in the case segmentation training process, the backbone network is initialized using the parameters obtained from the pre training, and the data set of diseased leaves of fruit trees collected in this experiment is applied to it. Through the above transfer learning method, it can improve the speed and accuracy, the overfitting phenomenon can be avoided, and the lack of training for small sample data sets can be compensated.

4. The dataset

Currently, there are not many publicly available datasets for apple disease detection. The data set used in this paper includes two sources, one from various copyright published search data sets on the network, and the other from the collection of local agricultural plantations. Finally, image data including 6371 leaves was obtained. Because of the restricted image quantity within the test dataset, to streamline training, it's essential to preprocess data images, normalize them, and employ LabelImg for image data labeling. The image data is shown in Figure.3.

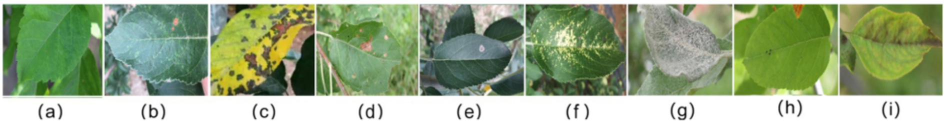

Figure 3. Dataset Of Leaf.

There are 9 different categories in the sample, include: (a) Health; (b) Alternaria leaf spot disease; (c) Brown spot disease; (d) Frog leaf spot disease; (e) Grey spot disease; (f) Mosaic disease; (g) Powdery mildew disease; (h) Rust disease; (i) Scab formation disease.

In the experiment, we performed data enhancement on the collected apple leaf images, including scaling and cropping, image enhancement, conversion to Tensor, normalization, and other operations to enhance the utilization efficiency of the dataset.

5. Experimental results analysis

5.1. Experimental evaluating indicator

To better evaluate the effectiveness of the experiment, this paper evaluates the results using three different indicators: Accuracy, precision, and Matthew's correlation coefficient.

5.2. Result analysis

In this experiment, we attempted 30 rounds of training, the ordinary ResNet50 network model, VGG19+AdaGrad model and the comprehensive application ResNet50 network model (ResNet50+CBAM+Transfer Learning + Data Enhancement) were tested respectively, and the performance is as follows:

a) Compare of Loss function:

The comparison of the test set loss function between the VGG19, ordinary ResNet50 and the comprehensive application ResNet50 is shown in Figure.4- Figure.6:

For ResNet50 model:

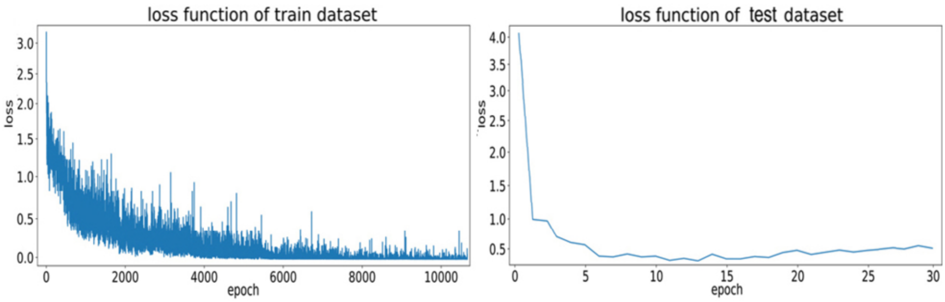

Figure 4. Loss function of train dataset and test dataset (ResNet50).

For VGG19 model:

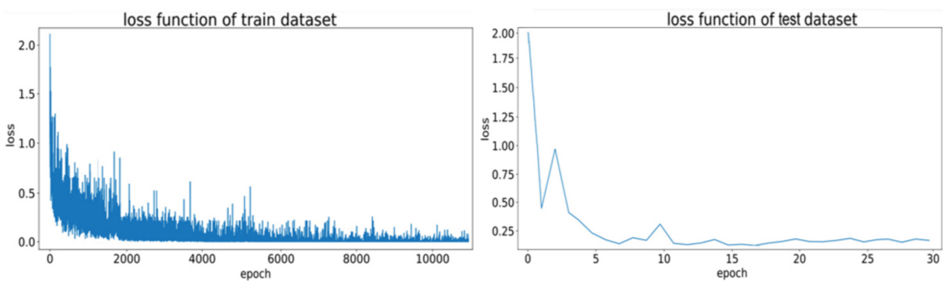

Figure 5. Loss function of train dataset and test dataset (VGG19).

For Comprehensive application of ResNet50 model:

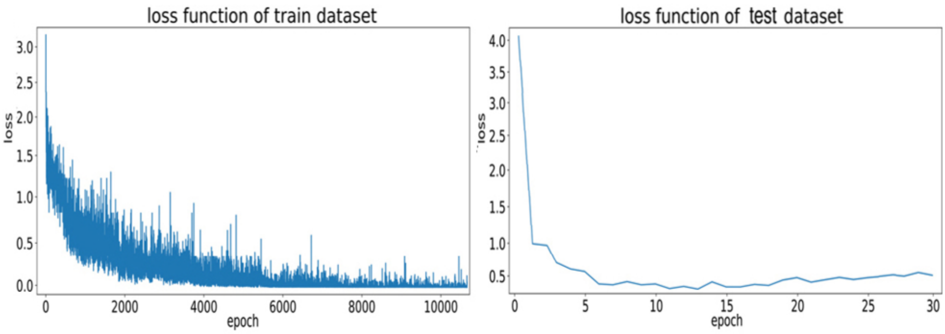

Figure 6. Loss function of train dataset and test dataset (Comprehensive Application of ResNet50).

It can be seen that the comprehensive application of the ResNet50 loss function has a smaller value and less error.

(2) Evaluation indicator comparison:

The comparison of classification and evaluation indicators between the ordinary ResNet50, VGG19 and the comprehensive application ResNet50 is shown in Table 1:

Table 1. Comparison of calculation accuracy.

Category	Accuracy	Precision	Matthew's correlation coefficient
VGG19	0.959356	0.937765	0.952011
ResNet50	0.917611	0.890105	0.909451
Comprehensive application of ResNet50	0.968762	0.947353	0.959414

Firstly, compared to the ResNet50 model, VGGNet19 has little difference in accuracy, accuracy, and various evaluation indicators of experimental results compared to the ResNet50 model. However, due to the relatively large number of parameters in VGG19, its computational complexity is higher compared to the ResNet50 model. In addition, due to the need for a large amount of convolution computation, the inference speed is slow. If speed and computing resource limitations are considered, ResNet50 is usually more efficient.

Secondly, compared to the data of the two ResNet models mentioned above, the comprehensive application of the ResNet50 model improved the recognition accuracy by 4.05%, the score by 5% compared to the ordinary ResNet50.

Based on the above situation, the comprehensive application ResNet50 used in this article achieved a classification accuracy of 96.87, with better experimental results than the other two models.

5.3. Comparison with other conventional research methods

Compared with other unilateral application research methods, the advantages of the research methods of integrated application ResNet50 model, CBAM, transfer learning, data enhancement and other technologies adopted in this paper are:

(1) Having stronger feature extraction capabilities than ordinary neural networks: The ResNet50 model has the advantages of deep and residual connectivity, allowing for learning richer and more complex features to better distinguish plant diseases and pests.

(2) Further enhance the focus of attention in image data: CBAM can help models automatically focus on important image regions and features, improving the accuracy and robustness of plant disease and pest recognition.

(3) It solves the problem of insufficient data in small sample data sets: through transfer learning, the model can be trained faster and the performance can be improved, and appropriate loss function can be used to balance the insufficient sample categories.

(4) Improved robustness of data applications: By expanding the dataset through data augmentation techniques, the model's generalization ability can be increased.

Therefore, when the above functional advantages are integrated in this experiment, not only can better feature extraction and classification accuracy be achieved, but also simpler combination efficiency and wider applicability can be achieved.

6. Conclusions

How to quickly identify diseases is an important guarantee for improving fruit tree yield. This paper comprehensively applies ResNet, transfer learning, CBAM, data enhancement, loss function, and the recognition accuracy can reach 96.87%, achieving a higher recognition accuracy than the ordinary neural network model. It has higher feature extraction efficiency and better classification recognition performance than ordinary neural networks, and solves the shortcomings of small sample datasets and improves the generalization ability of applications. This is of great significance for improving the level of forest disease control and exploring a reasonable integration mode of deep learning technology.

References

[1] Oo YM, Htun NC. Plant leaf disease detection and classification using image processing. International Journal of Research and Engineering. 2018 Apr;5(9):516-523.

[2] Geetharamani G, Pandian A. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Computers & Electrical Engineering, 2019 Jul;76:323-338.

[3] Song X, Mariano VY. Image-based apple disease detection based on residual neural network and transfer learning. ICPECA 2023:3rd International Conference on Power, Electronics and Computer Applications, Shenyang, China, 2023 Jan; 365-369.

[4] Li B, Lima D. Facial expression recognition via ResNet50. International Journal of Cognitive Computing in Engineering. 2021 Apr;2:57-64.

[5] Khan MS, Nath TD, Hossain MM, Mukherjee A, Hasnath HB, Meem TM, Khan U. Comparison of multiclass classification techniques using dry bean dataset, International Journal of Cognitive Computing in Engineering. 2023 Apr;4:6-20.

[6] Iqbal I, Shahzad G, Rafiq N et al. Deep learning - based automated detection of human knee joint's synovial fluid from magnetic resonance images with transfer learning. IET Image Processing. 2020, 14(10): 1990-1998.

[7] Abbas, S., Sampedro, G. A., Abisado, M., Almadhor, A., Yousaf, I., & Hong, S. P. Harris-Hawk-Optimization-Based Deep Recurrent Neural Network for Securing the Internet of Medical Things. Electronics, 2023 Jun;12(12):2612.

[8] Ramcharan A, Baranowski K, McCloskey P, Ahmed B, Legg J, Hughes DP. Deep learning for image-based cassava disease detection. Frontiers in Plant Science, 2017 Apr;8:1852.

[9] Alfarisy AA, Chen Q, Guo M. Deep learning based classification for paddy pests & diseases recognition. ICMAI 2018: Proceedings of 2018 international conference on mathematics and artificial intelligence, Chengdu, China, 2018 Apr; 21-25.

[10] Li W, Chen P, Wang B, Xie C. Automatic localization and count of agricultural crop pests based on an improved deep learning pipeline. Scientific Reports. 2019 Sep;9(1):7024.

[11] Jun H, Shuai L, Jinming S et al. Facial expression recognition based on VGGNet convolutional neural network. 2018 Chinese Automation Congress (CAC). IEEE, 2018: 4146-4151.

[12] Lu X, Kang X, Nishide S, Ren F. Object detection based on SSD-ResNet. CCIS 2019: IEEE 6th International Conference on Cloud Computing and Intelligence Systems, 2019 Dec; 89-92.

[13] Yun S, Bai Y, Jongnam B. Quality evaluation and satisfaction analysis of online learning of college students based on artificial intelligence. Security and Communication Networks, 2022 Oct;6(3):78-88.

[14] He K, Zhang X, Ren S, Sun J. CVPR 2016: Deep residual learning for image recognition. IEEE conference on computer vision and pattern recognition, Las Vegas,USA, 2016 Jun; 770-778.

[15] Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing, 2019 Sep;337:325-338.

[16] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing, 2021 Oct;452, 48-62.

[17] Woo S, Park J, Lee JY, Kweon CBAM: Convolutional block attention module, ECCV 2018: In Proceedings of the European conference on computer vision, Munich, Germany, 2018 May; 3-10.

[18] Iqbal I, Odesanmi GA, Wang J, Liu L. Comparative investigation of learning algorithms for image classification with small dataset. Applied Artificial Intelligence, 2021 Jan;35(10), 697-716.

[19] Christoffersen P, Jacobs K. The importance of the loss function in option valuation. Journal of Financial Economics, 2004 Oct;72(2):291-318.

[20] Ribani R, Marengoni M. A survey of transfer learning for convolutional neural networks, SIBGRAPIT 2019: IEEE 32nd SIBGRAPI conference on graphics, patterns and images tutorials, Kyoto, Japan, 2019 Oct; 47-57.