ICIC Express Letters
Volume **16**, Number **4**, April **2022**

 ISSN 1881-803X
pp. **401–408**

COMPARATIVE STUDY BETWEEN ENSEMBLE AND FUSION CONVOLUTIONAL NEURAL NETWORKS FOR DIABETIC RETINOPATHY CLASSIFICATION

Pichada Saichua and Olarik Surinta*

Multi-agent Intelligent Simulation Laboratory (MISL)
Department of Information Technology
Faculty of Informatics
Mahasarakham University
Khamriang Sub-District, Kantarawichai District, Mahasarakham 44150, Thailand
pichada.sai@msu.ac.th; *Corresponding author: olarik.s@msu.ac.th

Received August 2021; accepted October 2021

Abstract. *In this paper, we have demonstrated the effectiveness of the fusion convolutional neural network (CNN) and ensemble CNN architectures for diabetic retinopathy classification. Due to the fusion and ensemble CNN architectures, we proposed to use five CNN architectures consisting of InceptionV3, ResNet50, ResNet50V2, Xception, and DenseNet121 to find the best CNN model. Two of the best CNN models were then selected for creating the fusion and ensemble CNN architectures. We also performed data augmentation techniques while training the CNN models. We found that the data augmentation technique can increase the accuracy of the CNNs. However, the data augmentation technique should not distort the retinal image. For the fusion CNNs, Xception and InceptionV3 were combined and then attached with two dense layers with the size of 1,024 units for each dense layer. Hence, we selected the optimal dropout value with 0.4. For the ensemble CNNs, the output probabilities that were calculated from the Xception and InceptionV3 models, were sent to the ensemble learning method. Using ensemble learning methods, we also compared the weighted and unweighted average methods. The results showed that the weighted average method outperformed the unweighted average method in all ensemble CNNs. From our experiments, we found that the fusion CNN architecture slightly outperformed ensemble CNN architecture.*

Keywords: Convolutional neural network, Fusion CNNs, Ensemble CNNs, Diabetic retinopathy classification, Data augmentation technique

1. **Introduction.** The World Health Organization (WHO) has listed diabetic retinopathy as one of the leading causes of blindness worldwide [1]. The main issue of diabetic retinopathy arises from consuming a main meal that contains much sugar. There are many foods that are commonly sold in convenience stores that have a high level of sweetness, such as beverages, sweets and coffee. If we receive more sweetness than what we need, the excess will be harmful to the human body, especially to the eyes. However, if the blood vessels in the retina begin to become inflamed and aneurysm develops, there will be lymphatic distribution throughout the retina. If untreated, this causes retinal ischemia and this leads to the cells that are used for vision being destroyed continuously. It eventually leads to decreased ability to see and maybe loss of vision.

Due to the increasing number of morbidities associated with diabetic retinopathy, there may be an insufficient number of ophthalmologists to treat patients. Here we report the development of a system that helps to detect and classify diabetic retinopathy from retinal images using deep learning methods. There are two types of diabetic retinopathy that we can classify from the retinal images: diabetic retinopathy and non-diabetic retinopathy.

DOI: 10.24507/icicel.16.04.401

Consequently, we can deeply classify diabetic retinopathy into five levels; these are normal, mild non-proliferative diabetic retinopathy (NPDR), moderate NPDR, severe NPDR, and proliferative DR [2,3].

In this research, we proposed a convolutional neural network (CNN) framework for diabetic retinopathy classification. First, we aimed to find the best CNN architectures: InceptionV3, ResNet50, ResNet50V2, Xception, and DenseNet121. Then we selected the two best CNN models to create the new CNN frameworks, called fusion CNNs and ensemble CNNs. Finally, we then compared the performance of these two CNN frameworks to classify the retinal images of diabetic retinopathy. In addition, the ensemble CNNs combine the weighted parameters from multiple CNN models that are classified using the softmax function, while the fusion CNNs combine the feature maps from multiple CNN models before sending to classify using the softmax function. The advantage of the fusion and ensemble CNN architectures reduced the generalization error and increased prediction performance [4].

2. **Related Work.** In this section, we survey deep learning techniques that have been proposed for recognition of diabetic retinopathy (DR) from retinal images. Yazhini and Loganathan [5] proposed a framework called integrated fusion that combined GLCM and VGG19 architecture to extract the feature vector from the retinal images. In the integrated fusion framework, the feature vector extracted by GLCM and VGG-19 was first concatenated and then sent to classify using the softmax function. The integrated fusion framework predicted the output of the DR as five levels. This integrated fusion method provided an accuracy of 71.30% and sensitivity of 50.43%.

Hattiya et al. [6] proposed to use seven CNN architectures consisting of AlexNet, ResNet50, DenseNet201, InceptionV3, MobileNet, MnasNet, and NASNetMobile for diabetic retinopathy recognition. The dataset used in the experiment was downloaded from the Kaggle website. It included 23,513 retinal images with two classes: diabetic retinopathy and non-diabetic retinopathy. They first evaluated the CNN architecture with five different color spaces: RGB, grayscale, HSV, L*a*b, and YCbCr. The result showed that training the CNN model with RGB color space gave the highest recognition accuracy. They then trained seven CNN architectures with the RGB color space. The AlexNet architecture achieved a high accuracy of 81.42%.

Vives-Boix and Ruiz-Fernández [7] proposed CNN architectures that updated the weighted parameters in the convolutional layer using the synaptic metaplasticity method. In their experiments, InceptionV3 with synaptic metaplasticity achieved an accuracy of 95.56% on a public small diabetic retinopathy dataset.

The CNN architectures can be combined to create a new framework, called ensemble CNNs that include the ensemble learning method in the last layer. It learns from the output probabilities of each CNN that is included in the ensemble CNNs. Chompookham and Surinta [8] proposed ensemble learning methods that were created from three and five deep CNN models. Further, the output probabilities that calculated from each CNN model were then transferred to the ensemble learning layer. The CNN model was trained using data augmentation techniques: height shift, vertical flip, and fill mode. In the experiment, the ensemble CNNs with the weighted average method were evaluated on plant leaf disease datasets. High accuracies above 99% were obtained from the ensemble CNN architecture. Also, an accuracy of 94.7% was obtained on a mulberry leaf dataset.

Noppitak and Surinta [9] proposed ensemble CNN architecture to enhance the efficiency of land use classification. First, they discovered the best CNN model from eight CNN architectures: InceptionResNetV2, MobileNetV2, DenseNet201, Xception, ResNet152V2, NASNetLarge, VGG16, and VGG19. Second, the data augmentation techniques, including rotation, width shift, and height shift were used while training the CNN models. Third, the ensemble CNN architecture was then generated with the 3 CNN models that were

found in the first step. For the ensemble learning method, the weighted average method was proposed. Hence, the grid-search method was proposed to optimize the weighted parameters. The experimental results showed that the ensemble CNN architecture achieved an accuracy of 92.80%.

Deepa et al. [10] proposed a multi-stage deep CNN to learn to distinguish the diabetic retinopathy image from the whole image and random patches as the input images. First, the input images were sent to the CNN architectures: InceptionV3 and Xception. Second, the probability vectors from the CNN architectures were then classified using artificial neural networks (ANNs). Third, in the ensemble classifier process, the outputs of the ANNs were classified using a support vector machine classifier. As a result, their proposed method showed an accuracy of 96.2%.

3. **Methodology.** This paper reports on the objective to improve the efficacy of CNN frameworks by applying various optimization algorithms to reducing training losses. Moreover, the comparative study methods are presented based on two CNN frameworks: 1) fusion CNNs and 2) ensemble CNNs.

3.1. Convolutional neural network architectures.

InceptionV3. InceptionV3 was proposed by Szegedy et al. [11] in 2016. InceptionV3 was modified from the previous inception architecture and focused on providing less computational cost. In the InceptionV3 architecture, first, the factorized convolutional layers and the small convolutional layers were proposed to reduce the number of parameters involved in a network and also reduce the computation cost. Second, the symmetric convolutions were replaced by asymmetric convolutions. Next, an auxiliary classifier was proposed as the regularizer. Finally, to avoid a representational bottleneck, the grid size reduction of the feature maps was proposed.

ResNet50 and ResNet50V2. ResNet and ResNetV2 were proposed by He et al. [12, 13] in 2016. In very deep networks, the numbers of stacked layers were increased. The deep networks showed the high accuracy results on the challenging ImageNet dataset. ResNet50 and ResNet50V2 include 50 parameter layers. In ResNet architecture, residual learning was proposed to allow the stacked network to jump over one or more building blocks, called shortcut connections. In ResNetV2, a new residual unit was proposed. It was shown that the new residual unit decreased the error while training around 2%.

Xception. Chollet invented an extreme version of Inception architecture, called Xception [14] in 2017. It focused on modifying the depthwise separable convolution layer, namely a depthwise convolution. In the modified depthwise convolution, the order of the depthwise convolution was pointwise convolution and then followed by a depthwise convolution. So, the number of connections is fewer and the model is lighter. As a result, the Xception architecture showed improvement in accuracy performance when compared with the InceptionV3.

DenseNet121. In 2017, Huang et al. [15] proposed densely connected convolutional networks, called DenseNet. DenseNet is like the feed-forward architecture where each dense block layer connects to all other dense block layers. For example, the first dense block layer is connected to the 2nd, 3rd, and so on until the last dense block layer. The second dense block layer is also connected to the 3rd block and so on until the last layer. If considered, the connection of the 2nd layer, the feature maps of the 1st layer and the feature maps of the 2nd layer were combined and used as inputs into the 3rd layer.

3.2. Fusion CNNs.

The proposed fusion CNNs comprise two parts. The two best CNN architectures, that were selected from Section 3.1, are chosen to create temporal features, called feature maps. We then concatenated feature maps before sending them to another layer. These feature maps were first transferred to the batch normalization layer (BN). Second, the rectified linear unit (ReLU) was proposed as a nonlinear function followed

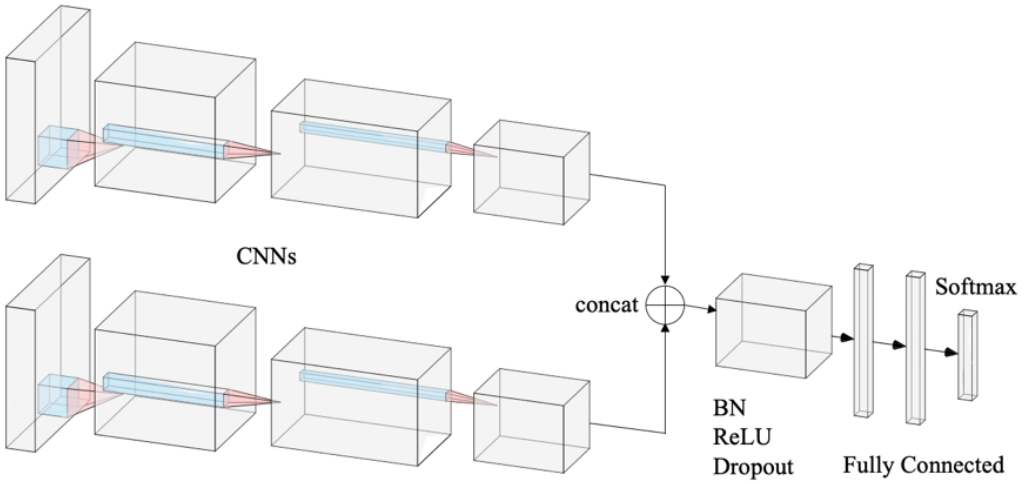


FIGURE 1. Illustration of the fusion CNN architecture

by the dropout layer to avoid overfitting. Finally, two fully connected layers and softmax activation function were attached to the network. The architecture of fusion CNNs is illustrated in Figure 1.

3.3. **Ensemble CNNs.** The ensemble CNN architecture consists of two parts. In the first part, we trained and evaluated five CNN architectures, including InceptionV3, ResNet50, ResNet50V2, Xception, and DenseNet121. After that, we chose only two CNN architectures to create the ensemble CNN framework. In the second part, the output probabilities (w) from each CNN were then calculated using ensemble learning methods [8,9]. Figure 2 illustrates the proposed ensemble CNN architecture. The ensemble learning methods are described as follows.

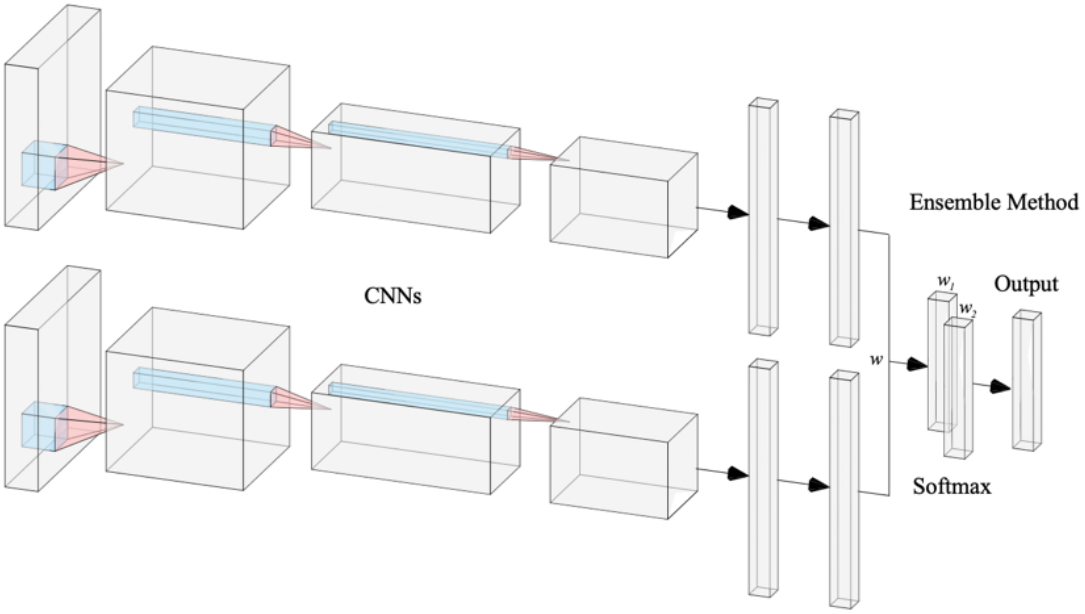


FIGURE 2. Illustration of the ensemble CNN architecture

Weighted average method. In the weighted average method, the different weight parameters are calculated with the output probabilities. We decided to calculate the higher weight with the output probabilities of the CNN model that achieved a higher classification rate. Note that the sum of all weight parameters is equal to one. The equation

of the weighted average method is given by $\hat{y} = \frac{1}{n}\sum_{i=1}^{n} \alpha\vec{y}$, where α is the weight value that multiplies with the vector of output probabilities ($\vec{y}$) and n is the number of ensemble CNN models.

Unweighted average method. In this learning method, the output probabilities of each CNN model are computed by average of the probability values. Then, the maximum value of the probabilities is selected as an output of the ensemble learning. The unweighted average method is calculated by $\hat{y} = \frac{1}{n}\sum_{i=1}^{n} \vec{y}$, where $\vec{y}$ is the vector of the output probabilities of each CNN model and n is the number of ensemble CNN models. We then used the arg max function to select the highest probability value of $\hat{y}$.

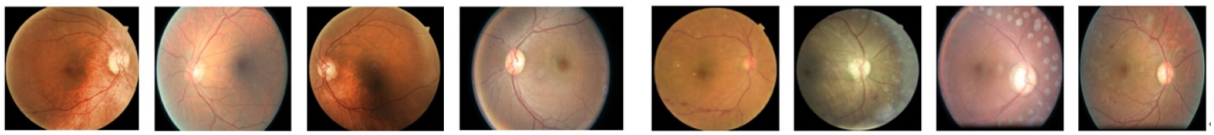
4. Experimental Setup and Results.

All experiments were evaluated in the same environment. We used the TensorFlow v2.5.0 as the deep learning framework that runs on Google Colab platform.

4.1. Diabetic retinopathy dataset.

We collected retinal images from various DR datasets that were available on the Kaggle website, including APTOS 2019 blindness detection, diabetic retinopathy detection, and diabetic retinopathy. These datasets were collected by a collaboration between the Aravind Eye Hospital and Kaggle website. The hospital screened for diabetic retinopathy of the patient from the retinal images and then created a label for each image [2,3].

The retinal images used in this study were stored in RGB color space with a JPEG format. The DR dataset included 2 classes and contained 23,510 images. The number of diabetic retinopathy and non-diabetic retinopathy images was 12,063 and 11,447 images, respectively. We then divided the DR dataset into a training set (18,808 images) and test set (4,702 images). Some examples of the DR dataset are shown in Figure 3.



(a) (b)

FIGURE 3. Illustration of the retinal images from the diabetic retinopathy dataset: (a) Non-diabetic retinopathy class and (b) diabetic retinopathy class

4.2. Experiments with CNNs and data augmentation techniques.

In this study, data augmentation techniques [16] were proposed. The data augmentation techniques used in the experiments were flip (horizontal and vertical flips), rotation (randomly with value between 0-90°), and zoom (randomly with value between $[1-0.2, 1+0.2]$) techniques. We proposed these data augmentation techniques because they did not distort the retinal images. Examples of the data augmentation techniques are shown in Figure 4.

We evaluated the performance of the CNN architectures using several factors as follows: five CNN architectures (InceptionV3, ResNet50, ResNet50V2, Xception, and DenseNet121), five optimization algorithms (SGD, Adam, Adadelta, Adagrad and Adamax), and learning rates (0.1, 0.01, 0.001, and 0.0001). In addition, our experiments evaluated several optimization algorithms because the fittest optimizer assures more reliable accuracy performance [8].

We show the results obtained with the CNN architecture with the best optimal parameters on the DR dataset in Table 1. It can be seen that Xception was the best CNN architecture in our experiments. The Xception architecture achieved an accuracy of 84.07%. It slightly outperformed InceptionV3 by about 0.6%. However, it spent more computation time on the training.

(a) (b) (c) (d) (e)

FIGURE 4. Illustration of the data augmentation techniques, including (a) original image, (b) horizontal flip, (c) vertical flip, (d) rotation, and (e) zoom

TABLE 1. The best optimizers and the accuracy (%) of each CNN model

CNNs	Image size	Optimizers	Learning rate	Training time	Test	
					Accuracy (%)	Loss
Xception	299×299	Adamax	0.001	57 min	**84.07**	0.38
InceptionV3	299×299	Adam	0.001	38 min	83.30	0.38
DenseNet121	224×224	Adadelta	0.1	33 min	82.77	0.4
ResNet50V2	224×224	Adamax	0.001	37 min	82.20	0.38
ResNet50	224×224	Adam	0.0001	38 min	81.28	0.41

The experimental results of the data augmentation techniques are shown in Table 2. It is important to emphasize that the data augmentation techniques can improve the performance of the classification of diabetic retinopathy. Xception combined with the rotation technique still outperformed all CNN architectures. It achieved 85.50% and was slightly better than training without data augmentation technique. It is quite surprising that it requires less computation time when compared with training without data augmentation techniques.

TABLE 2. The optimal data augmentation techniques of each CNN model

CNNs	Data augmentation techniques	Training time	Test	
			Accuracy (%)	Loss
Xception	Rotation	44 min	**85.50**	**0.34**
InceptionV3	Zoom	38 min	84.69	0.37
DenseNet121	Rotation	26 min	84.01	0.35
ResNet50	Flip	40 min	83.26	0.41
ResNet50V2	Flip	21 min	83.18	0.39

To summarize the results, the Xception architecture without data augmentation obtained the accuracy of 84.07%, outperforming all CNN architectures, except only the InceptionV3 with zoom technique that achieved the accuracy of 84.69%.

4.3. **Experiments with fusion CNN architecture.** In the experiment of fusion CNNs, we combined the Xception architecture with other CNNs, because we found in previous experiments that the Xception architecture provided the highest accuracy with lower loss value. The compared results of each fusion CNN are shown in Table 3.

From Table 3, our results show that the fusion CNNs between Xception+InceptionV3 obtained 86.30% accuracy with the loss value of 0.33. This fusion CNNs also required to be fully connected with two dense layers with only 1024 units of each dense layer. We can emphasize that the Xception architecture when combined with the other CNNs, still obtained an accuracy above 85%.

Table 3. Performance evaluation of the fusion CNNs

Fusion CNNs	Dense sizes	No. of dense layers	Dropout	Training time	Test	
					Accuracy (%)	Loss
Xception+InceptionV3	1024	2	0.4	53 min	**86.30**	**0.33**
Xception+DenseNet121	2048	2	0.2	33 min	85.45	0.34
Xception+ResNet50V2	4096	1	No	30 min	85.11	0.36
Xception+ResNet50	1024	1	0.1	32 min	85.07	0.4

4.4. **Experiments with ensemble CNN architecture.** For the ensemble CNNs experiments, the accuracy results of the ensemble CNNs are shown in Table 4. The performance of the ensemble CNNs was greater than the single CNN architecture. However, only ensemble CNNs between ResNet50 and ResNet50V2 (83.73%) performed with lower accuracy than the Xception architecture (84.07%). As a result, the ensemble CNNs between Xception and InceptionV3 provided an accuracy of 86.11%. When we compared the ensemble CNNs and the fusion CNNs, the fusion CNN architecture slightly outperformed the ensemble CNNs.

Table 4. Performance evaluation of the ensemble CNNs

Ensemble CNNs	Accuracy (%) of ensemble learning methods		
	Unweighted average method	Weighted average method	Weight parameters
Xception+InceptionV3	85.92	**86.11**	0.6, 0.4
ResNet50+DenseNet121	84.30	84.81	0.3, 0.7
ResNet50V2+DenseNet121	84.22	84.39	0.3, 0.7
ResNet50+ResNet50V2	83.65	83.73	0.3, 0.7

5. **Conclusions.** In this paper, we evaluated two convolutional neural network (CNN) frameworks: fusion CNN and ensemble CNN architectures, for diabetic retinopathy classification. We discover the best CNN architecture from five CNN architectures: InceptionV3, ResNet50, ResNet50V2, Xception, and DenseNet121. To optimize the parameters, we considered five optimization algorithms: SGD, Adam, Adadelta, Adagrad and Adamax. The various learning rates between 0.1 and 0.0001 were also evaluated. In the course of the training policy, the data augmentation techniques were also performed during the training, including flip, rotation, and zoom techniques. The result showed that the Xception architecture with the Adamax optimizer and a learning rate of 0.001 achieved the best accuracy performance. Interestingly, the data augmentation techniques can increase the accuracy of every CNN architecture. The fusion and the ensemble CNN architectures were compared. We also found that the combination between Xception and InceptionV3 architectures performed very well on both architectures. From the experimental results, we conclude that the performance of the fusion CNN architecture was slightly better than the ensemble CNN architecture.

In future work, to improve the performance of diabetic retinopathy classification, we will concentrate on experiments with the other CNN frameworks, such as snapshot ensemble CNN [17], Siamese network [18], and Hybrid CNN [19].

Acknowledgment. This research project was financially supported by Mahasarakham University.

REFERENCES

[1] Y. Shimizu, Eye care, vision care, vision impairment and blindness, *World Health Organization (WHO)*, https://www.who.int/health-topics/blindness-and-vision-loss, Accessed on Jun. 06, 2021.

[2] D. Doshi, A. Shenoy, D. Sidhpura and P. Gharpure, Diabetic retinopathy detection using deep convolutional neural networks, *International Conference on Computing, Analytics and Security Trends (CAST)*, pp.261-266, 2016.
[3] S. Burewar, A. B. Gonde and S. K. Vipparthi, Diabetic retinopathy detection by retinal segmentation with region merging using CNN, *The 13th International Conference on Industrial and Information Systems (ICIIS)*, pp.136-142, 2018.
[4] M. A. Ganaie, M. Hu, M. Tanveer and P. N. Suganthan, Ensemble deep learning: A review, *arXiv.org*, arXiv: 2104.02395, 2021.
[5] K. Yazhini and D. Loganathan, An integrated fusion based feature extraction and classification model for diabetic retinopathy diagnosis, *The 2nd International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp.1187-1193, 2020.
[6] T. Hattiya, K. Dittakan and S. Musikasuwan, Diabetic retinopathy detection using convolutional neural network: A comparative study on different architectures, *Mahasarakham Int. J. Eng. Technol.*, vol.7, no.1, pp.50-60, 2021.
[7] V. Vives-Boix and D. Ruiz-Fernández, Diabetic retinopathy detection through convolutional neural networks with synaptic metaplasticity, *Comput. Methods Programs Biomed.*, vol.206, pp.1-8, 2021.
[8] T. Chompookham and O. Surinta, Ensemble methods with deep convolutional neural networks for plant leaf recognition, *ICIC Express Letters*, vol.15, no.6, pp.553-565, 2021.
[9] S. Noppitak and O. Surinta, Ensemble convolutional neural network architectures for land use classification in economic crops aerial images, *ICIC Express Letters*, vol.15, no.6, pp.531-543, 2021.
[10] V. Deepa, C. S. Kumar and T. Cherian, Ensemble of multi-stage deep convolutional neural networks for automated grading of diabetic retinopathy using image patches, *J. King Saud Univ. – Comput. Inf. Sci.*, vol.33, no.6, DOI: 10.1016/j.jksuci.2021.05.009, 2021.
[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, Rethinking the inception architecture for computer vision, *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2818-2826, 2016.
[12] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770-778, 2015.
[13] K. He, X. Zhang, S. Ren and J. Sun, Identity mappings in deep residual networks, in *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, B. Leibe, J. Matas, N. Sebe and M. Welling (eds.), Cham, Springer, 2016.
[14] F. Chollet, Xception: Deep learning with depthwise separable convolutions, *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1251-1258, 2017.
[15] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, Densely connected convolutional networks, *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4700-4708, 2017.
[16] C. Shorten and T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, *J. Big Data*, vol.6, no.1, 2019.
[17] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft and K. Q. Weinberger, Snapshot ensembles: Train 1, get M for free, *The 5th International Conference on Learning Representations (ICLR)*, pp.1-14, 2017.
[18] X. Zeng, H. Chen, Y. Luo and W. Ye, Automated diabetic retinopathy detection based on binocular Siamese-like convolutional neural network, *IEEE Access*, vol.7, pp.30744-30753, 2019.
[19] Y. T. Hafiyan, Afiahayati, R. D. Yanuaryska, E. Anarossi, V. M. Sutanto, J. Triyanto and Y. Sakakibara, A hybrid convolutional neural network-extreme learning machine with augmented dataset for DNA damage classification using comet assay from buccal mucosa sample, *International Journal of Innovative Computing, Information and Control*, vol.17, no.4, pp.1191-1201, 2021.