Received: November 22, 2022. Revised: February 5, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Effective Hybrid Convolutional-Modified Extreme Learning Machine in Early Stage Diabetic Retinopathy

Dian Candra Rini Novitasari[1,2] **Fatmawati[2]*** **Rimuljo Hendradi[2]** **Yuniar Farida[1]**
Ricky Eka Putra[3] **Rinda Nariswari[4]** **Rizal Amegia Saputra[5]**
Rr Diah Nugraheni Setyowati[6]

[1]*Mathematics Department, Faculty of Science and Technology UIN Sunan Ampel Surabaya, Surabaya, Indonesia*
[2]*Mathematics Department, Faculty of Science and Technology Universitas Airlangga, Surabaya, Indonesia*
[3]*Department of Informatics Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia*
[4]*Statistics Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia*
[5]*Information System Department, Universitas Bina Sarana Informatika, Indonesia*
[6]*Environmental Engineering Department, UIN Sunan Ampel Surabaya, Surabaya, Indonesia*
* Corresponding author's Email: fatmawati@fst.unair.ac.id*, diancrini@uinsby.ac.id+

Abstract: Diabetic retinopathy (DR) initially derives from damage to the eye blood vessels which leads to bleeding up-to permanent blindness. The severity of DR is not easily known. Therefore. it is necessary to create a system that is able to identify the severity level of DR. In this study, the identification of DR was conducted using hybrid CNN and ELM method. Hybrid CNN-ELM is useful for obtaining the most effective model in the classification system and computational time. CNN architecture is useful for extracting fundus data in image feature recognition. Several modified ELM methods (KELM, MLELM, DELM) were used to classify the severity of DR based on the results of CNN feature extraction. The classification system was tested with two datasets, namely DRIVE and Messidor. Based on the average value, the best architecture evaluation in extracting fundus data was DenseNet compared to GoogleNet, ResNet18, ResNet50, and ResNet101. Based on the computation time, the KELM method was faster than the MLELM and DELM methods, with an average time of 43 seconds. The results on the DRIVE dataset produced good evaluation values, while Messidor obtained good results on the MLELM method. It showed that the Messidor data was able to be separated well linearly. The modified CNN-MLELM method produced more stable values in both datasets with an average accuracy of 99.21%, a sensitivity of 99.29%, and a specificity of 99.21%.

Keywords: Diabetic retinopathy, CNN architecture, Feature learning, Modified ELM.

1. Introduction

Diabetic retinopathy (DR) is a microvascular complication caused by prolonged diabetes. DR is a disease that causes high rates of blindness worldwide, and it mainly occurs in the working-age population at the age of 20 to 65 [1]. DR ranks fifth as the disease that causes the most blindness worldwide with moderate to severe visual impairment. The prevalence of blindness due to DR at the standard age increased from 14.90% to 18.50% from 1990 to 2020 [2]. The high prevalence of DR is caused by the unconsciousness of DR patients in the early stages, resulting in delays in treatment [3]. The stages in DR begin with mild non-proliferative diabetic retinopathy (NPDR), which can progress to proliferative diabetic retinopathy (PDR) [4]. Increased glucose levels cause swelling or rupture of blood vessels in the retina, which can cause blindness [5, 6]. Blindness due to DR can be prevented by early identification of the severity of blood vessel damage in the retina using the computer aided diagnosis (CAD) system.

Several previous studies have implemented the CAD system in identifying the severity level of DR. The research on DR [7] made the classification of DR by looking at hard exudates, microaneurysms, hemorrhages, and cotton wool spots from the feature extraction results. The classification results obtained an accuracy of 99.00% using the support vector machine (SVM) method. DR research was also conducted using the gray-level co-occurrence matrix (GLCM)-SVM. Classifier method showed that the machine learning made had an accuracy value of 82.35% for the diagnosis of normal and DR. As for the diagnosis of NPDR and PDR, perfect accuracy was obtained, that is 100.00%, which means that all classes could be classified correctly [8]. As technology evolves, the CAD method currently being developed is the deep learning method. In several studies, deep learning can classify image data well [9]. One method in deep learning that is in great demand by researchers is the convolutional neural network (CNN). CNN is a deep learning method that classifies image data based on the features studied. In CNN, feature learning and classification are carried out in the same architecture [10].

The study to identify the severity of diabetic retinopathy using CNN was conducted by Ratul Gosh. This study tested CNN's performance on 2 (Normal and DR) and 5 (Normal, Mild, Moderate, Severe, and PDR) DR severity classes. The result was that the classification in 2 classes had a higher accuracy rate than those in 5, which is 95.00 % and 85.00% [11]. CNN has various network models, including GoogleNet [12]. GoogleNet is a type of CNN network that utilizes the use of inception modules so that it can reduce the output features of each block and can accelerate the performance of CNN [13].

Based on the previous research, it can be seen that the CNN method has good ability in terms of image classification. However, CNN also has weaknesses in the time required to conduct learning due to many layers of neurons and hidden layers in the architecture [14]. These problems can be overcome by merging CNN with other methods that have fast learning capabilities. Several studies were conducted on hybrid CNN and other classification methods, such as combining CNN with bidirectional-long short term memory (Bi-LSTM) [7]. Another research that combines the CNN learning feature model ResNet50 with the SVM classification method achieved a good and effective classification system with an accuracy of 86.76% [15]. Hybrid CNN and extreme learning machine (ELM) were carried out by Kannojia et al. [16]. The research carried out CNN feature extraction and then continued the classification process with the ELM method. The results showed a higher accuracy value compared to a single CNN.

The ELM method has undergone several developments by adding kernel functions and increasing the number of hidden layers. The ELM development method that adds kernel functions is called the kernel extreme learning machine (KELM) method [17]. Meanwhile, multi-layer extreme learning machine (MLELM) is the addition of a hidden layer to the ELM architecture [18]. The MLELM and KELM methods are included in the deep learning method known as the deep extreme learning machine (DELM) method. Several developments in ELM methods aim to improve system performance in studying data patterns and more efficient computational time.

Based on the literature studies that have been conducted, the contributions of this research are as follows:

1. Improve the accuracy of the DR severity classification system by combining the learning features of CNN and the ELM classification method.
2. It is overcoming the shortage of CNN in the problem of long computation (training) time by applying the ELM method.
3. Compare several modified ELM methods of classifying DR severity, such as KELM, MLELM, and DELM.

Some of the contributions above are carried out in several stages, such as pre-processing with changing image size and improving image quality using contrast limited adaptive histogram equalization (CLAHE), augmentation process with image rotation techniques, then studying features with CNN, and the last stage is DR severity classification using the modified ELM methods.

2. Preliminaries

2.1 Diabetic retinopathy

Diabetic retinopathy is one of the complications caused by diabetes mellitus [19]. Patients with long-term diabetes mellitus risk developing microvascular complications in the retina [20]. In patients with diabetes mellitus, high glucose levels can cause blood vessels in the retina to become blocked so that new blood vessels appear on the retina [21]. These new blood vessels are often deformed, break easily and leak [22]. Blood vessels that burst and leak due to blockage will release fluid. This fluid will accumulate on the retina so that it will result in decreased vision ability and can even lead to blindness [23].

Diabetic retinopathy is divided into several stages that indicate the patient's severity [24]. The staging is distinguished by microaneurysms, intraretinal hemorrhages or venous bleading, and

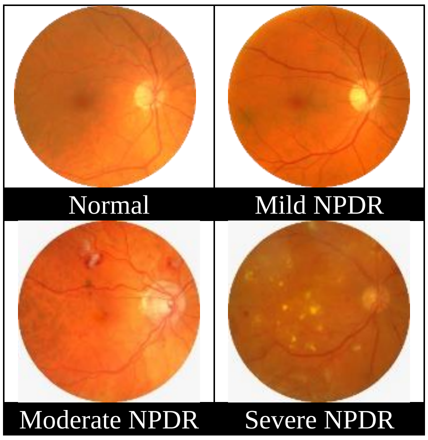

Figure. 1 Fundus image

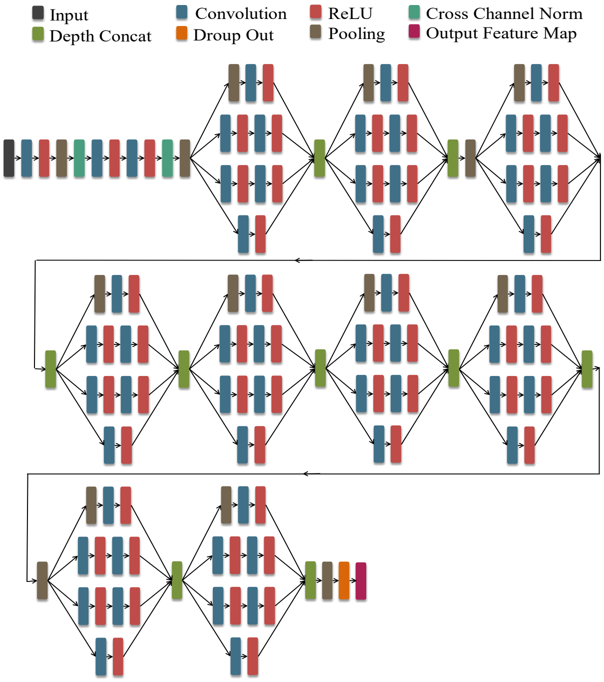

Figure. 2 Feature learning with GoogleNet

neovascularization [25]. A retinal sample with diabetic retinopathy can be seen in Fig. 1.

2.2 Convolutional learning

CNN is one of the network methods in deep learning networks [26]. The CNN method is specifically designed to overcome the problem of image data classification in MLP, which is considered having less ability to provide good results. It is due to the spatial information of the image data is not stored, and each feature in the image data is considered an independent feature. CNN consists of 2 major processes, namely feature learning and classification process.

Features learning is a CNN architecture that only takes the feature extraction process without entering the classification process. The feature extraction consists of an iterative process of convolution calculations, pooling, and the use of activation functions. CNN has several convolutional learning architectures such as ResNet, DenseNet, and GoogleNet [27]. These architectures have different learning patterns and layer depths.

2.2.1. GoogleNet

GoogleNet is one of the CNN models that won the ILSVRC competition in 2014. GoogleNet managed to classify 1000 classes of image data with an error of 6.70% [28]. Convolutional learning with GoogleNet architecture adopts the previous CNN network concept, namely inception modules. The purpose of inception is to reduce the output features in each convolution block [13]. It makes the network work in parallel and can produce high accuracy values with a more efficient computing system [29]. Convolutional learning with GoogleNet architecture consists of 9 inception modules [30].

2.2.2. ResNet

ResNet is a CNN deep learning model that won the image classification competition with 1000 ILSVRC classes in 2015. The advantage of the ResNet architecture is that an image classification system with good results can handle the vanishing gradient problem, which has been the main problem in deep learning. A special feature of the ResNet architecture is the residual block system. This residual block uses residual calculations studied in the previous layer and then added together with the results of other features. On CNN, the ResNet model has 3 different types of architecture, namely ResNet-18, ResNet-50, and ResNet-101. The naming of the three types of ResNet is based on the depth of the network. The deeper the network is, the more layers of feature learning are repeated.

2.2.3. DenseNet

DenseNet is an extension of the CNN ResNet architecture. DenseNet modifies the residual network by changing the addition process to concatenation. In ResNet, the previous feature map will be summed with the residual feature map result $f(x)$, while in DenseNet, the previous feature map will be combined or stacked with the next feature map result. DenseNet can reduce the number of parameters, strengthen feature deployment, and solve the vanishing gradient problem. DenseNet has several variants such as DenseNet-121, DenseNet-160, DenseNet-201. DenseNet variations differ based on the number of layer depths in the architecture.

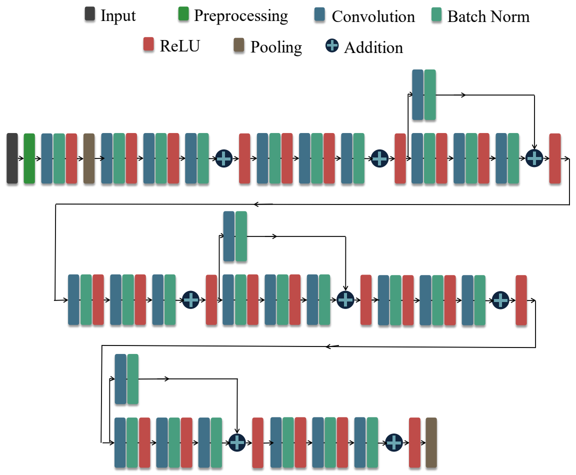

Figure. 3 Feature learning with ResNet18

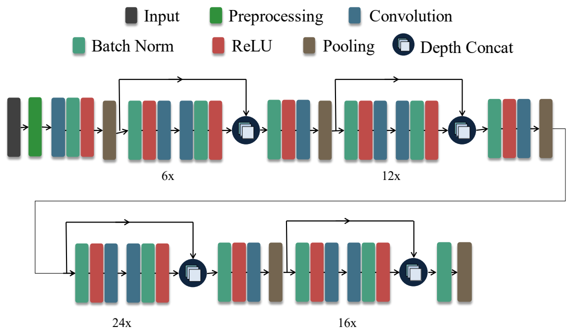

Figure. 4 Feature learning with DenseNet

The CNN method is prevalent and produces good system performance in image recognition or classification, from classifying common objects seen in everyday life to classifying high-accuracy medical data. Even though the CNN method has many types of architectures and developments, some conventional architectures are still susceptible to several problems, as follows:

1. Overfitting. Most of the CNN or deep learning methods are highly dependent on data availability. The fewer data or imbalanced classes, the more likely overfitting will occur in the CNN model [31].
2. High model complexity. Some CNN architectures require high computational power and long training times due to deep and complex learning networks [32].
3. The CNN method requires high memory. It is part of deep learning which requires high memory space to store model weights and parameters [32].

2.3 Modified ELM

ELM is a system learning method that utilizes a single hidden layer feedforward neural network

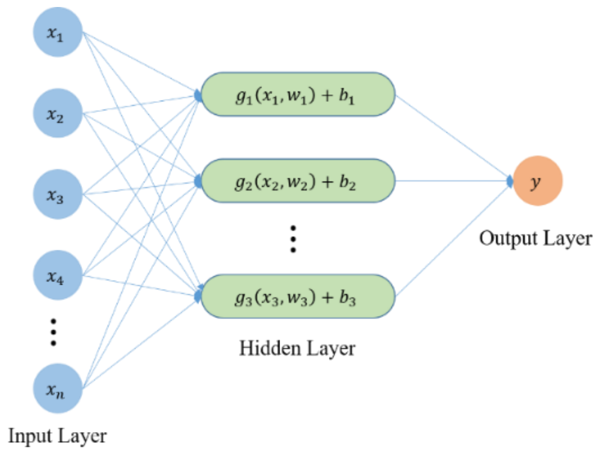

Figure. 5 ELM architecture

(SLFN) in recognizing data patterns. The conventional neural network method requires iteration and weight update to obtain the optimal weight. In contrast, the ELM method obtains the optimal weight in one step using the moore-penrose generalization method [33]. The moore-penrose generalization method optimizes the input weights determined randomly in one step, so the ELM method is relatively faster than conventional neural network methods with good performance and low computational costs [34, 35]. The architecture of the ELM method is illustrated in Fig 5.

Where w is the input weight, x is the input matrix, b is the bias of the hidden layer, and $g(x)$ is the activation function. Feature mapping with the activation function $g(x)$ which separates data linearly on a simple ELM network structure, has weak precision performance, so the ELM method is developed to handle non-linear problems through high-dimensional mapping [36]. Handling non-linear problems with the addition of kernel functions can map data into higher dimensions so that non-linear data is transformed into linear [37]. The use of the kernel function in the ELM method, known as the KELM method, is a development of the conventional ELM method that has good performance in classification and still has the advantage of the ELM method in terms of training time. Radial basis function (RBF) is a kernel function with one hyper-parameter (γ) that simplifies model configuration. RBF has a good performance in mapping data features. The RBF formula can be written in Eq. (1) [38].

$$K(x, w) = e^{-\gamma\|x-w\|^2} \quad (1)$$

The difference between the KELM and ELM methods is in the use of the mapping function in the hidden layer. The simple architecture of the ELM

Received: November 22, 2022. Revised: February 5, 2023.

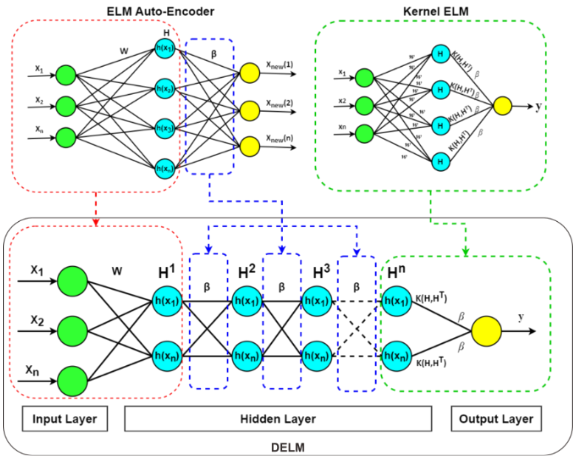

Figure. 6 DELM Architecture

method with one hidden layer and a random weight value (w) causes the ELM method to have poor performance stability [39]. The problem of poor performance stability in the ELM method can be overcome by a developed method, namely the MLELM method. MLELM develops the ELM method with two or more hidden layers [8]. The implementation of the MLELM method with more than two hidden layers makes the MLELM method more stable than the ELM method.

Applying the KELM method with the MLELM concept forms a deep learning method known as the DELM. The ELM method only has one hidden layer, while the DELM has two hidden layers in which one of them is the KELM process which is the final output of the MLELM process [40]. The DELM architecture can be seen in Fig. 6.

The DELM calculation begins by entering the data to be used and defined by **X**. Then generate the weight and bias values randomly using Eq. (2).

$$\mathbf{H} = \mathrm{g}\left(\mathbf{W}^{\mathrm{T}}\mathbf{X} + \mathbf{b}\right) \tag{2}$$

Where matrix **W** is the input weight, matrix **X** is the input matrix, vector **b** is the hidden layer's bias, and $g(.)$ is the activation function. The next process is to calculate $\boldsymbol{\beta}$ value which is determined by $\mathrm{n_i}$ and $\mathrm{n_h}$ value using Eq. (3).

$$\boldsymbol{\beta} = \begin{cases} \left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \frac{\mathbf{I}_{\mathrm{n_h}}}{c}\right)^{-1} \mathbf{H}^{\mathrm{T}}\mathbf{X}, & if\ \mathrm{n_i} = \mathrm{n_h} \\ \mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{H}^{\mathrm{T}} + \frac{\mathbf{I}_{\mathrm{n_i}}}{c}\right)^{-1} \mathbf{X}, & if\ \mathrm{n_i} \geq \mathrm{n_h} \\ \mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{X}, & if\ \mathrm{n_i} < \mathrm{n_h} \end{cases} \tag{3}$$

The next process is calculation of $\mathbf{X_{new}}$ using Eq. (4).

$$\mathbf{X_{new}} = \mathbf{X}\boldsymbol{\beta}^{\mathrm{T}} \tag{4}$$

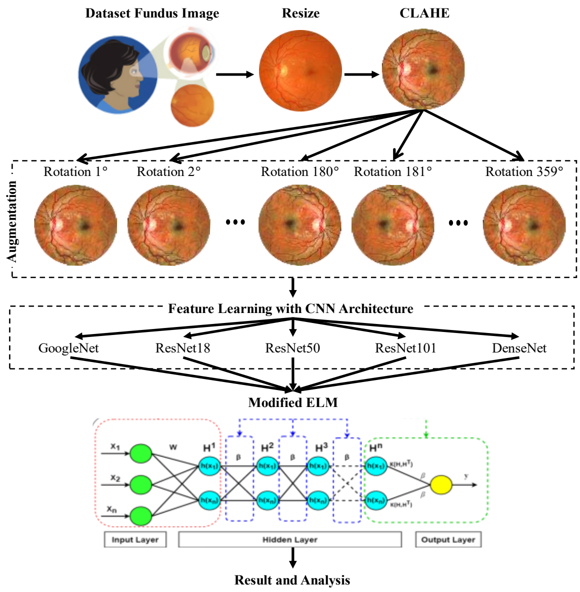

Figure. 7 Flowchart

The next step is to calculate the value of the finite layer using Eq. (5).

$$\mathbf{H^n} = g(\mathbf{H}_{n-1}(\boldsymbol{\beta}_{n-1})^T) \tag{5}$$

Repeat the calculation in step 3 until the hidden layer (n-1). Calculate the output y using Eq. (6).

$$\mathbf{y_{out}} = [(K(\mathbf{H}, \mathbf{H}^{\mathrm{T}}))^T]^{-1} \cdot \left(\frac{\mathbf{I}}{c} + \mathbf{H}\mathbf{H}^{\mathrm{T}}\right) \tag{6}$$

3. Research method

This study proposed creating a CNN hybrid model and a modified ELM method. Making this hybrid model aims to improve the performance effectiveness of the classification model. The CNN method implements convol-based feature learning, which can adequately represent image features. The application of the ELM method in the classification stage helps overcome the shortcomings of the CNN method, such as complexity and high memory requirements. This study compares two fundus image

Table 1. Data number of each DR class

Dataset	**Class DR**	**n Class**	**n Class After Augmentation**
DRIVE	Normal	10	3.590
	Mild	14	5.026
	Moderate	16	5.774
	Severe	9	3.231
Messidor	Normal	516	185.244
	Mild	153	54.927
	Moderate	247	88.673
	Severe	254	91.186

Table 2. Feature extraction in CNN Architecture

No Feature	GoogleNet	ResNet18	ResNet50	ResNet101	DenseNet
1	0.05	1.27	3.14	2.07	-0.00
2	0.22	1.79	0.05	0.00	0.00
3	0.00	0.39	0.00	0.06	0.00
4	0.56	0.24	0.12	0.00	-0.09
5	0.02	0.15	0.01	0.12	0.00
6	0.02	0.65	0.00	0.49	0.00
7	0.00	0.36	0.01	1.58	-0.00
8	0.00	1.02	0.21	0.13	0.00
9	0.03	1.38	0.00	0.36	0.00
10	0.10	0.71	0.13	0.59	0.01
⋮	⋮	⋮	⋮	⋮	⋮
Feature (n)	0.41	0.07	0.31	0.47	0.89
Total feature	**1024**	**512**	**2048**	**2048**	**1920**

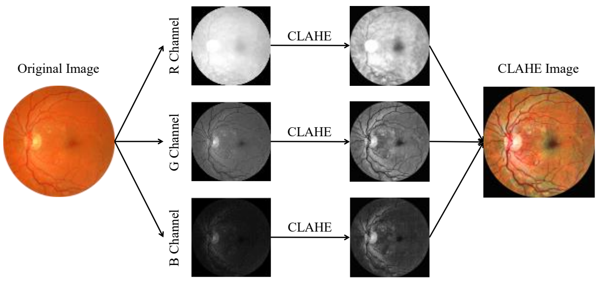

Figure. 8 CLAHE process

database sources, namely DRIVE and Messidor. The number of data in DRIVE is 44 fundus images, while the Messidor database has 1200 fundus images. The two databases were compared based on the evaluation results of the 4 DR severity class classification system (Normal, Mild, Moderate, and Severe).

The initial stage of the research is data preprocessing which consists of several steps. First, each data is resized by the CNN input size dimension of 224 × 224 pixel. Second, image quality is improved by applying the CLAHE method. Third, data augmentation is carried out by rotation of 1°. Fourth, undersampling is a solution to the problem of imbalanced data. Based on the flowchart presented in Fig. 7, the next step after preprocessing is the process of convolutional feature learning using CNN architecture. The feature learning process produces feature extraction for each fundus image data. The feature map will enter the classification process using several Modified ELM methods, namely MLELM, KELM, and DELM.

In order to obtain optimal results, several experiments are carried out with various batch sizes in the training process. After the most optimal model is obtained, the next step is to test the model on testing data to obtain the results of the classification system evaluation. Measuring the performance of the evaluation system is conducted using the calculation of the accuracy value of Eq. (7), sensitivity Eq. (8), and specificity Eq. (9).

$$Accuracy = \frac{TP_{all\_n}}{n} \times 100\% \tag{7}$$

$$Sensitivity = \frac{\sum \frac{TP}{TP+FN}}{n} \times 100\% \tag{8}$$

$$Specificity = \frac{\sum \frac{TN}{TN+FP}}{n} \times 100\% \tag{9}$$

Where n is the number of classes. A true positive (TP) is the number of correct positive prediction of DR class. A true negative (TN) is the number of correct negative prediction of DR class. A false positive (FP) is the number of incorrect positive prediction of DR class. A false negative (FN) is the number of incorrect negative prediction of DR class.

4. Results and analysis

This study was conducted based on retinal fundus image data obtained from the DRIVE database and the Messidor database with 4 DR severity classes (Normal, Mild, Moderate, and Severe). The initial pre-processing stage is cropping, resizing, and improving image quality using the CLAHE method. The results of CLAHE processing can be seen in Fig. 8. Quality improvement occurs in each RGB channel and then is recombined into RGB format. Image enhancement with CLAHE produces sharp images and can clearly show blood vessels or other small factors in the retina.

Data from the CLAHE process is then augmented with a rotation of 1°. Augmentation aims to deal with the problem of the imbalanced data. After

Received: November 22, 2022. Revised: February 5, 2023.

Table 3. Results on Evaluation Value Classsification of DR with CNN-KELM

Dataset	CNN	Batchsize	Accuracy	Sensitivity	Specificity	Time
DRIVE	ResNet-18	16	100.00	100.00	100.00	37.87
		32	100.00	100.00	100.00	38.91
		64	100.00	100.00	100.00	39.18
	ResNet-50	16	100.00	100.00	100.00	46.27
		32	100.00	100.00	100.00	44.22
		64	100.00	100.00	100.00	45.34
	ResNet-101	16	100.00	100.00	100.00	43.90
		32	100.00	100.00	100.00	45.86
		64	100.00	100.00	100.00	47.07
	DenseNet	16	100.00	100.00	100.00	43.87
		32	100.00	100.00	100.00	45.60
		64	100.00	100.00	100.00	47.28
	GoogleNet	16	100.00	100.00	100.00	39.28
		32	100.00	100.00	100.00	43.84
		64	100.00	100,00	100.00	40.90
Messidor	ResNet-18	16	94.93	94.96	94.93	36.67
		32	94.93	94.96	94.93	42.28
		64	94.93	94.96	94.93	38.87
	ResNet-50	16	97.04	97.06	97.04	47.14
		32	97.04	97.06	97.04	42.84
		64	95.51	95.59	95.51	46.38
	ResNet-101	16	97.31	97.31	97.31	44.32
		32	97.31	97.31	97.31	46.89
		64	84.64	86.49	84.64	41.74
	DenseNet	16	97.58	97.59	97.58	43.71
		32	97.58	97.59	97.58	43.91
		64	92.89	93.26	92.89	46.07
	GoogleNet	16	95.08	95.10	95.08	37.76
		32	95.08	95,10	95.08	41.11
		64	93.70	93.79	93.70	41.54

augmentation, the number of data in each class is equalized according to the minimum number of classes that are carried out randomly. The results of treating the imbalanced class can be seen in Table 1.

Based on Table 1, the difference in the data number is very significant, so it is necessary to conduct undersampling as a solution to handle the imbalanced dataset. 3240 data have been randomly selected in each class. The data of each class enter the feature learning process using several CNN architectures, namely GoogleNet, ResNet18, ResNet50, ResNet101, and DenseNet. Feature learning in each architecture produces different feature maps, which can be seen in Table 2.

A feature map is the result of extraction from each image that represents the special features of the image. The result of the feature map for each architecture becomes the input for the classification process using the Modified ELM method, namely KELM, MLELM, and DELM. The evaluation calculation is conducted using the values of accuracy, sensitivity, and specificity. This study also calculates the computational time of the classification to measure the effectiveness level of the system.

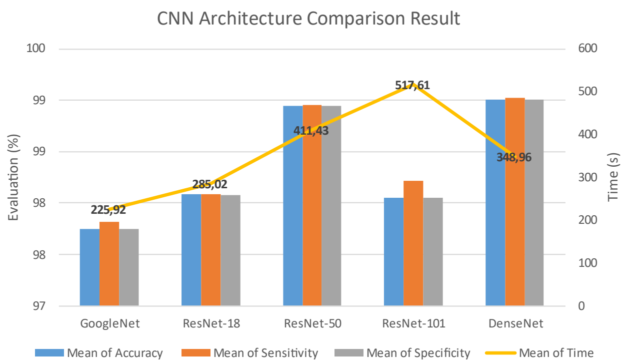

Figure. 9 Results on CNN architecture comparison

Fig. 9 compares the mean of the accuracy, sensitivity, and specificity values and the length of computation time on each CNN architecture. The overall result of the most optimal model is the DenseNet architecture. The accuracy, sensitivity, and

Received: November 22, 2022. Revised: February 5, 2023.

Table 4. Results on evaluation value classsification of DR with CNN-MLELM

Dataset	CNN	Batchsize	Accuracy	Sensitivity	Specificity	Time
DRIVE	ResNet-18	16	100.00	100.00	100.00	46.85
		32	99.73	99.73	99.73	47.99
		64	99.92	99.92	99.92	49.00
	ResNet-50	16	99.27	99.27	99.27	976.00
		32	99.73	99.73	99.73	971.00
		64	99.27	99.28	99.27	992.00
	ResNet-101	16	91.97	93.22	91.97	972.00
		32	100.00	100.00	100.00	979.00
		64	99.69	99.69	99.69	971.00
	DenseNet	16	99.96	99.96	99.96	199.00
		32	100.00	100.00	100.00	193.00
		64	99.88	99.89	99.88	197.00
	GoogleNet	16	88.29	89.43	88.29	166.00
		32	99.92	99.92	99.92	159.00
		64	99.92	99.92	99.92	157.00
Messidor	ResNet-18	16	100.00	100.00	100.00	45.41
		32	99.73	99.73	99.73	44.82
		64	99.96	99.96	99.96	44.75
	ResNet-50	16	100.00	100.00	100.00	974.00
		32	100.00	100.00	100.00	990.77
		64	100.00	100.00	100.00	1,084.00
	ResNet-101	16	99.73	99.73	99.73	950.90
		32	100.00	100.00	100.00	950.35
		64	99.85	99.85	99.85	949.84
	DenseNet	16	100.00	100.00	100.00	771.48
		32	100.00	100.00	100.00	878.12
		64	100.00	100.00	100.00	890.37
	GoogleNet	16	99.81	99.81	99.81	158.45
		32	99.81	99.81	99.81	169.23
		64	99.77	99.77	99.77	167.07

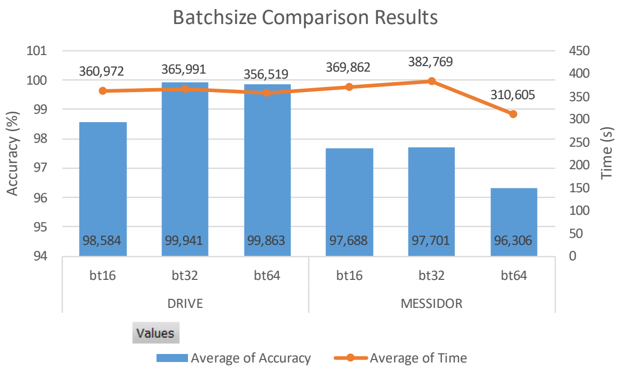

Figure. 10 Batchsize comparison results

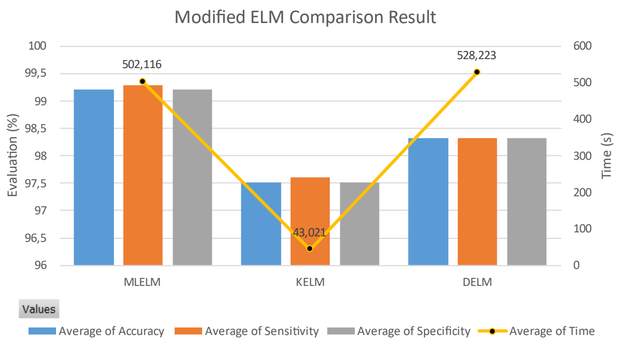

Figure. 11 Modified ELM Comparison Result

specificity results outperform all other architectures on CNN. Meanwhile, in terms of computing time, GoogleNet architecture is better, with an average time difference of 123.00 seconds compared to DenseNet. However, in the overall evaluation results, DenseNet produces the best feature extraction. It is because DenseNet is a complex architecture so that the resulting features are very representative of the features of the image dataset.

Feature extraction on the CNN architecture has tested several types of batch sizes, namely 16, 32, and 64. The comparison results for each batch size can be seen in Fig. 10. In both datasets, DRIVE and Messidor produce an optimal batch size value of 32. The DRIVE dataset obtains an average accuracy of 99.94%, while the Messidor dataset achieves an average accuracy of 97.70%.

The results of the DR classification experiment

Table 5. Results on evaluation value classsification of DR with CNN-DELM

Dataset	CNN	Batchsize	Accuracy	Sensitivity	Specificity	Time
DRIVE	ResNet-18	16	99.27	99.27	99.27	976.00
		32	99.73	99.73	99.73	971.00
		64	99.27	99.28	99.27	992.00
	ResNet-50	16	100.00	100.00	100.00	450.61
		32	100.00	100.00	100.00	479.03
		64	100.00	100.00	100.00	479.86
	ResNet-101	16	100.00	100.00	100.00	504.95
		32	100.00	100.00	100.00	484.40
		64	100.00	100.00	100.00	495.42
	DenseNet	16	100.00	100.00	100.00	528.74
		32	100.00	100.00	100.00	530.82
		64	100,00	100.00	100.00	510.25
	GoogleNet	16	100.00	100.00	100.00	383.27
		32	100.00	100.00	100.00	455.74
		64	100.00	100.00	100.00	283.18
Messidor	ResNet-18	16	93.89	93.89	93.88	534.57
		32	93.93	93.93	93.92	626.17
		64	93.55	93.55	93.53	517.18
	ResNet-50	16	97.89	97.89	97.89	275.94
		32	97.81	97.81	97.81	270.01
		64	97.54	97.54	97.54	272.45
	ResNet-101	16	98.16	98.16	98.17	696.57
		32	97.93	97.93	97.93	529.93
		64	98.27	98.27	98.28	562.07
	DenseNet	16	98.08	98.08	98.1	377.33
		32	97.93	97.93	97.94	467.24
		64	98.16	98.16	98.16	467.71
	GoogleNet	16	95.81	95.81	95.83	553.66
		32	96.43	96.43	96.45	597.88
		64	95.81	95.81	95.87	571.95

Table 6. Results on modified ELM comparison

M-ELM	**DATA**	**Accuracy**	**Sensitivity**	**Spesificity**	**Time**
KELM	DRIVE	100.00	100.00	100.00	43.51
	MESS	95.04	95.22	95.04	42.75
MLELM	DRIVE	98.50	98.66	98.50	503.31
	MESS	99.91	99.91	99.91	532.44
DELM	DRIVE	99.89	99.89	99.89	568.40
	MESS	96.75	96.75	96.75	488.04

using CNN with KELM can be seen in Table 3, that using MLELM can be seen in Table 4, and that using DELM can be seen in Table 5. Then the calculation of of the comparison of several modified ELM methods can be seen in Table 6. The results show that the KELM method has a faster computation time than other modified ELM methods. The average computation time is 43.13 seconds. The speed of computation time in KELM is much different compared to other modified ELM methods. MLELM and DELM have an average computation time of around 500.00 seconds. It shows that KELM is very effective and light in computing. Based on the average evaluation scores, the MLELM method produces the best scores outperforming KELM and DELM. The comparison of these results can be seen in Fig. 11.

In this study, the selection of the most appropriate modified ELM method was seen from the results of each dataset. There are several interesting things about this experiment. The results of the KELM and

Received: November 22, 2022. Revised: February 5, 2023.

Table 7. Other research classification DR in Messidor dataset

Author	Dataset	Num Class	Methodology	Accuracy
Srinivasan, et al, 2022 [41]	Kaggle-APTOS and IDRiD	5 class (Normal, Mild, Moderate, and Severe, PDR)	MSA-ResNetGB model	94.40%
Pundikal, Manohar, and Mallikarjun Sayabanna Holi, 2021 [42]	e-ophtha	3 class (Mild, Moderate, and Severe)	GWO-MKNN	99.10%
Putra, et al, 2020 [15]	Messidor base 12, base 13, and base 21	3 class (Mild, Moderate, and Severe)	ResNet50+Relief Feature Reduction+SVM-NB	89.12%
Gayathri, et al, 2020 [44]	Messidor	4 class (Normal, Mild, Moderate, and Severe)	Proposed CNN+J48	99.75%
Proposed Method	**Messidor**	**4 class (Normal, Mild, Moderate, and Severe)**	**CNN+MLELM**	**99.91%**

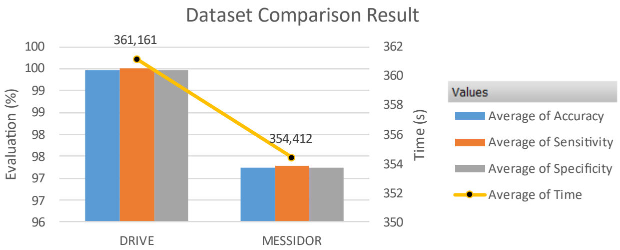

Figure. 12 Results of dataset comparison

DELM methods on the DRIVE dataset achieve a high average evaluation value, but the evaluation values on the Messidor dataset has decreased. The number of data on DRIVE is 49 (before augmentation), and Messidor has a total dataset of 1.170 (before augmentation). The difference in the data number makes the model have more ability to recognize data with a small amount. It shows that the variation of the data greatly affects the performance of the classification system.

In the MLELM method, the Messidor dataset produces better evaluation values than DRIVE, with a difference of only 1.00%. The MLELM method is different from KLEM and DELM, which utilizes kernel functions on hidden nodes. MLELM uses activation functions and stabilizes ELM performance by adding several hidden layers. This case shows that the Messidor data can be separated linearly without the need for a kernel function as in the KELM and DELM methods. In this study, the type of kernel used in KELM and DELM is polynomial.

Based on Fig. 11, the results of the MLELM method outperform KLEM and DELM, with the average accuracy value in both datasets of 99.21%, the sensitivity value of 99.29%, and the specificity value of 99,21%. In contrast, the computation time of the DR classification process is 517.88 seconds. The time is much longer than that of the KLEM method, but the time of MLELM method is faster than that of the DELM method. Therefore, the MLELM method is the most appropriate modified ELM to classify the DR level from the CNN feature extraction results.

In this study, the results of the DR classification were compared with two datasets: DRIVE and Messidor. Both data contain fundus image data showing four levels of DR severity (Normal, Mild, Moderate, and Severe). The overall results of the trials are shown in Tables 4 and 5 for the DRIVE and Messidor datasets. The results of the average evaluation scores and computational time are shown in Fig. 12. The graph shows that DRIVE achieves a good evaluation average with an accuracy of 99.46%, a sensitivity of 99.52%, and a specificity of 99.46%. Although the DRIVE dataset produces better results than the Messidor dataset, the results of the Messidor classification are considered good at determining the severity of DR. The difference in computing time required is only 6 seconds that Messidor dataset is faster than DRIVE dataset.

5. Discussion

Previous research has done hybrid CNN with classification methods based on machine learning and neural networks. Previous studies by Srinivasan [41] and Pundikal [42] its identify DR by fundus image segmentation and its obtain accuracy 94.40% and 99.10%. Previous study conducted Putra, et al. [15] that combine the CNN architecture with the SVM classification method. The data used is DR Messidor disease on certain bases (bases 12, 13, and 21). Furthermore, Gayathri, et al [43] carried out the same classification, namely DR disease using all Messidor data and increasing the number of classes to 4 classes (Normal, Mild, Moderate, and Severe). The method used is a hybrid between feature learning

CNN and the J48 classification method. This study proposes a hybrid CNN and MLELM method in the case of DR disease classification. The resulting accuracy value is better than previous studies, which is 99.91% on Messidor 4 class data. Based on trials conducted hybrid CNN and MLELM are very suitable in the problem of classification of DR disease.

DR is a dangerous eye disease. Various symptoms and signs can be identified on fundus images. In future studies, it is recommended to be able to measure and detect these signs such as microaneurysms, intraretinal or venous bleeding, blood vessel rupture, and neovascularization with the object detection method. Measurement and identification of various DR disease factors can use segmentation techniques or object detection.

6. Conclusion

In this study, the classification was carried out on four levels of DR severity (Normal, Mild, Moderate, and Severe). The data used were DRIVE and Messidor datasets which were extracted using several CNN architectures. Each architecture produced different map feature extraction values. Several experiments of the modified ELM method were carried out to determine the optimal classification system. The evaluation results were calculated using the accuracy, sensitivity, specificity, and computation time values.

Based on the average evaluation values, the best architecture in extracting fundus image features was DenseNet. The highest accuracy value was 99.00% with a relatively low computation time of 349 seconds compared to ResNet50 and ResNet101. The result of the DenseNet feature map was 1920 features from each image data.

The MLELM method is a modified ELM method that is good at classifying the severity of DR. This can be seen from the stability of the average evaluation value of the two datasets with an accuracy of 99.21%, a sensitivity of 99.29%, and a specificity of 99.21%. Based on the difference in datasets, DRIVE produced a good evaluation score compared to Messidor. It shows that the level of data variation affects the level of complexity and the results of the classification system.

Conflicts of interest

The authors declare that there is no conflict of interest.

Author contributions

Conceptualization is carried out by Dian Candra Rini Novitasari, Fatmawati, and Rimuljo Hendradi; writing is carried out by Dian Candra Rini Novitasari; data analysis is carried out by Dian Candra Rini Novitasari; supervision is carried out by Fatmawati, and Rimuljo Hendradi. Yuniar Farida is reviewer, Ricky Eka Putra is data analyst, Rinda Nariswari is data visualization, original draft preparation by Rr Diah Nugraheni Setyowati and Rizal Amegia Saputra.

References

[1] S. H. Alzahrani et al., "Awareness of diabetic retinopathy among people with diabetes in Jeddah, Saudi Arabia", *Ther. Adv. Endocrinol. Metab.*, Vol. 9, No. 4, pp. 103–112, 2018.

[2] Z. L. Teo et al., "Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis", *Ophthalmology*, Vol. 128, No. 11, pp. 1580–1591, 2021.

[3] A. Balasopoulou et al., "Symposium Recent advances and challenges in the management of retinoblastoma Globe - saving Treatments", *BMC Ophthalmol.*, Vol. 17, No. 1, p. 1, 2017, doi: 10.4103/ijo.IJO.

[4] T. Y. Wong et al., "Guidelines on Diabetic Eye Care: The International Council of Ophthalmology Recommendations for Screening, Follow-up, Referral, and Treatment Based on Resource Settings", *Ophthalmology*, Vol. 125, No. 10, pp. 1608–1622, 2018, doi: 10.1016/j.ophtha.2018.04.007.

[5] V. J. V. Potter, D. Karamichos, and D. J. Lee, "Ocular Complications of Diabetes and Therapeutic Approaches", *Biomed Res. Int.*, Vol. 2016, 2016, doi: 10.1155/2016/3801570.

[6] M. Elakkiya, K. Selvaraj, and G. Kuppuswamy, "Current and Emerging Therapies For the Management of Diabetic Retinopathy", *J. Appl. Pharm. Sci.*, Vol. 7, No. 9, pp. 243–251, 2017, doi: 10.7324/JAPS.2017.70933.

[7] S. Angadi and V. S. Reddy, "Hybrid deep network scheme for emotion recognition in speech", *Int. J. Intell. Eng. Syst.*, Vol. 12, No. 3, pp. 59–67, 2019, doi: 10.22266/ijies2019.0630.07.

[8] A. Z. Foeady, D. C. R. Novitasari, A. H. Asyhar, and M. Firmansjah, "Automated Diagnosis System of Diabetic Retinopathy Using GLCM Method and SVM Classifier", In: *Proc Electr. Eng. Comput. Sci. Informatics*, Vol. 5, No. 5, pp. 154–160, 2018.

[9] G. Marcus, "Deep Learning: A Critical

Appraisal", pp. 1–27, 2018.

[10] H. Garg and K. Kumar, "A novel exponential distance and its based TOPSIS method for interval-valued intuitionistic fuzzy sets using connection number of SPA theory", *Artif. Intell. Rev.*, Vol. 53, No. 1, pp. 595–624, 2020.

[11] R. Ghosh, K. Ghosh, and S. Maitra, "Automatic detection and classification of diabetic retinopathy stages using CNN", In: *Proc. of 2017 4th Int. Conf. Signal Process. Integr. Networks, SPIN 2017*, pp. 550–554, 2017, doi: 10.1109/SPIN.2017.8050011.

[12] W. Wang, Y. Yang, X. Wang, W. Wang, and J. Li, "Development of convolutional neural network and its application in image classification: a survey", *Opt. Eng.*, Vol. 58, No. 4, p. 40901, 2019, doi: 10.1117/1.OE.58.4.040901.

[13] C. Szegedy et al., "Going deeper with convolutions", In: *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.

[14] K. Xu, D. Feng, and H. Mi, "Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image", *Molecules*, Vol. 22, No. 12, 2017, doi: 10.3390/molecules22122054.

[15] R. E. Putra, H. Tjandrasa, and N. Suciati, "Severity classification of non-proliferative diabetic retinopathy using convolutional support vector machine", *Int. J. Intell. Eng. Syst.*, Vol. 13, No. 4, pp. 156–170, 2020, doi: 10.22266/ijies2020.0831.14.

[16] S. P. Kannojia and G. Jaiswal, "Ensemble of hybrid CNN-ELM model for image classification", In: *Proc. of 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, 2018, pp. 538–541.

[17] D. C. R. Novitasari, A. H. Asyhar, M. Thohir, A. Z. Arifin, H. Mu'jizah, and A. Z. Foeady, "Cervical Cancer Identification Based Texture Analysis Using GLCM-KELM on Colposcopy Data", In: *Proc. of 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 409–414, doi: 10.1109/ICAIIC48513.2020.9065196.

[18] A. Law and A. Ghosh, "Multi-label classification using a cascade of stacked autoencoder and extreme learning machines", *Neurocomputing*, Vol. 358, pp. 222–234, 2019.

[19] O. S. Servat, C. Hernández, and R. Simó, "Diabetic Retinopathy in the Context of Patients with Diabetes", *Ophthalmic Res.*, Vol. 62, No. 4, pp. 211–217, 2019, doi: 10.1159/000499541.

[20] A. Chawla, R. Chawla, and S. Jaggi, "Microvasular and macrovascular complications in diabetes mellitus: Distinct or continuum?", *Indian J. Endocrinol. Metab.*, Vol. 20, No. 4, pp. 546–553, 2016, doi: 10.4103/2230-8210.183480.

[21] D. S. Sisodia, S. Nair, and P. Khobragade, "Diabetic retinal fundus images: Preprocessing and feature extraction for early detection of Diabetic Retinopathy", *Biomed. Pharmacol. J.*, Vol. 10, No. 2, pp. 615–626, 2017, doi: 10.13005/bpj/1148.

[22] N. Prabhu, D. Bhoir, and N. Shanbhag, "Diabetic retinopathy screening using machine learning for hierarchical classification", *Int. J. Innov. Technol. Explor. Eng.*, Vol. 8, No. 10, pp. 1943–1948, 2019, doi: 10.35940/ijitee.J9277.0881019.

[23] E. J. Duh, J. K. Sun, and A. W. Stitt, "Diabetic retinopathy: current understanding, mechanisms, and treatment strategies", *JCI insight*, Vol. 2, No. 14, 2017.

[24] S. Qummar et al., "A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection", *IEEE Access*, Vol. 7, pp. 150530–150539, 2019, doi: 10.1109/ACCESS.2019.2947484.

[25] L. Wu, "Classification of diabetic retinopathy and diabetic macular edema", *World J. Diabetes*, Vol. 4, No. 6, p. 290, 2013, doi: 10.4239/wjd.v4.i6.290.

[26] J. Gu et al., "Recent advances in convolutional neural networks", *Pattern Recognit.*, Vol. 77, pp. 354–377, 2018, doi: 10.1016/j.patcog.2017.10.013.

[27] B. Almas, K. Sathesh, and S. Rajasekaran, "A Deep Analysis of Google Net and Alex Net for Lung Cancer Detection", *Int. J. Eng. Adv. Technol.*, Vol. 9, No. 2, pp. 395–399, 2019, doi: 10.35940/ijeat.b3226.129219.

[28] M. Ji, L. Zhang, and Q. Wu, "Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks", *Inf. Process. Agric.*, Vol. 7, No. 3, pp. 418–426, 2020, doi: 10.1016/j.inpa.2019.10.003.

[29] K. Annapurani and D. Ravilla, "CNN based image classification model", *Int. J. Innov. Technol. Explor. Eng.*, Vol. 8, No. 11, Special Issue, pp. 1106–1114, 2019, doi: 10.35940/ijitee.K1225.09811S19.

[30] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep Convolutional Neural Networks for Computer-Aided Detection : CNN Architectures, Dataset Characteristics and Transfer Learning", *IEEE Trans. Med. Imaging*, Vol. 35, No. 5, pp.

1285–1298, 2016, doi: 10.1109/TMI.2016.2528162.

[31] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprevost, "Reducing overfitting and improving generalization in training convolutional neural network (CNN) under limited sample sizes in image recognition", In: *Proc. of 2020-5th International Conference on Information Technology (InCIT)*, 2020, pp. 300–305.

[32] D. C. R. Novitasari et al., "Image Fundus Classification System for Diabetic Retinopathy Stage Detection Using Hybrid CNN-DELM", *Big Data Cogn. Comput.*, Vol. 6, No. 4, p. 146, 2022.

[33] T. Chouikh, H. Hellara, R. Barioul, F. Sandid, and O. Kanoun, "Comparative Study of Optimization Methods Proposed for the Extreme Learning Machine (ELM)", In: *Proc. of 2021 International Workshop on Impedance Spectroscopy (IWIS)*, 2021, pp. 137–140.

[34] M. Ali and R. Prasad, "Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition", *Renew. Sustain. Energy Rev.*, Vol. 104, pp. 281–295, 2019.

[35] J. Wang, S. Lu, S. H. Wang, and Y. D. Zhang, "A review on extreme learning machine", *Multimed. Tools Appl.*, pp. 1–50, 2021.

[36] F. Luo, W. Guo, Y. Yu, and G. Chen, "A multi-label classification algorithm based on kernel extreme learning machine", *Neurocomputing*, Vol. 260, pp. 313–320, 2017.

[37] N. Zhang and S. Ding, "Unsupervised and semi-supervised extreme learning machine with wavelet kernel for high dimensional data", *Memetic Comput.*, Vol. 9, No. 2, pp. 129–139, 2017.

[38] G. Tang, L. Yang, S. Ren, L. Meng, F. Yang, and H. Wang, "An automatic source code vulnerability detection approach based on KELM", *Secur. Commun. Networks*, Vol. 2021, 2021.

[39] H. Zhang, Z. Peng, J. Tang, M. Dong, K. Wang, and W. Li, "A multi-layer extreme learning machine refined by sparrow search algorithm and weighted mean filter for short-term multi-step wind speed forecasting", *Sustain. Energy Technol. Assessments*, Vol. 50, p. 101698, 2022.

[40] S. Ding, N. Zhang, X. Xu, L. Guo, and J. Zhang, "Deep Extreme Learning Machine and Its Application in EEG Classification", *Math. Probl. Eng.*, Vol. 2015, 2015, doi: 10.1155/2015/129021.

[41] V. Srinivasan and V. Rajagopal, "Multi-Scale Attention-Based Mechanism in Gradient Boosting Convolutional Neural Network for Diabetic Retinopathy Grade Classification", *Int. J. Intell. Eng. Syst*, 2022, doi: 10.22266/ijies2022.0831.44.

[42] M. Pundikal and M. S. Holi, "Microaneurysms detection using grey wolf optimizer and modified K-nearest neighbor for early diagnosis of diabetic retinopathy", *Intell. Networks Syst. Soc.*, Vol. 15, No. 1, 2021, doi: 10.22266/ijies2022.0228.13.

[43] S. Gayathri, V. P. Gopi, and P. Palanisamy, "A lightweight CNN for Diabetic Retinopathy classification from fundus images", *Biomed. Signal Process. Control*, Vol. 62, p. 102115, 2020.

[44] S. Gayathri, V. P. Gopi, and P. Palanisamy, "Diabetic retinopathy classification based on multipath CNN and machine learning classifiers", *Phys. Eng. Sci. Med.*, Vol. 44, No. 3, pp. 639–653, 2021.