



Severity Classification of Non-Proliferative Diabetic Retinopathy Using Convolutional Support Vector Machine

Ricky Eka Putra^{1,2*} Handayani Tjandrasa¹ Nanik Suciati¹

¹Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

²Department of Informatics Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia

* Corresponding author's Email: rickyeka@unesa.ac.id; handatj@its.ac.id

Abstract: Diabetic retinopathy is the principle disease that can make blindness. The early stage of diabetic retinopathy is Non-Proliferative Diabetic Retinopathy, which is splitted into three levels namely mild, moderate, and severe. This research is conducted to classify data on Base21, Base 13, and Base 12 from the Messidor database into 2 classes (mild, severe) and 3 classes (mild, moderate, severe). This research is useful for minimizing the funds spent and can be a breakthrough for people who has diabetic retinopathy that lack the hospital diagnosing funds. There are five stages in this research, those are pre-processing, image enhancement, feature extraction, feature reduction, and classification. The pre-processing step consists of cropping and resizing data, then the image is enhanced using Contrast Limited Adaptive Histogram Equalization, Morphology Contrast Enhancement, and Homomorphic. The results of the image enhancement are used as the inputs to the feature extraction layer which in this study uses the GoogLeNet, ResNet18, ResNet50, and ResNet101. Then the data is reduced using the Principle Component Analysis and Relief before entering the classification layer. The Support Vector Machine - Naive Bayes is used to replace the fully connected layer in Convolutional Neural Network to speed up and to optimize the classification process. The best results from the experiments are obtained by the Homomorphic, ResNet50, and Relief before entering to the Support Vector Machine-Naïve Bayes. The Homomorphic obtains 85.87% accuracy, ResNet50 can achieve 86.76% accuracy, and the Relief can reach 89.12% accuracy.

Keywords: Non-proliferative diabetic retinopathy classification, Support vector machine, Convolutional neural network, Filtering.

1. Introduction

Diabetic retinopathy is the principle disease that can make blindness. This condition is brought about by raised blood glucose levels, which cause harm to little veins in the retina, which at that point give off an impression of being *microaneurysm* (MA) as modest red spots [1]. The effect of diabetic retinopathy is the accumulation of excess fluid, blood, cholesterol and other fats in the retina, resulting in thickening and swelling of the macula. Weak walls in blood vessels in the retina of the eye cause bleeding / hemorrhages (HA) in the retina [2].

Based on the severity of the disease, diabetic retinopathy (DR) is categorized as normal, Non-Proliferative Diabetic Retinopathy (NPDR), and

Proliferative Diabetic Retinopathy (PDR). NPDR is the initial stage of Diabetic Retinopathy. This stage consists of mild, moderate, and severe NPDR [1],[3]. Because the number of diabetic retinopathy events throughout the world continues to increase, this disease still remains an important problem because many patients with diabetic retinopathy who do not realize that they have the disease [4].

Along with the times and technological advances are very rapid, a lot of research conducted on diabetic retinopathy for early detection in patients suffering from this disease by using artificial intelligence or also known as machine learning [5]. Artificial Intelligence is modeling a system that can learn those problems like the neural network of the human brain. Some methods that are being developed in artificial

intelligence are Fuzzy Logic [6], Evolutionary Computing [7], and Machine Learning [8].

Machine Learning is a model approach to a system so that it can work as closely as possible with the neural network of the human brain. This method is the most popular because it is able to study and generalize a problem like the human brain. Some applications of this method are prediction and classification. A distinctive feature of machine learning is that there is a training and testing process. One method in Machine Learning that is often used is Neural Network [9].

This research was conducted to classify the severity in NPDR using modifications, namely by the Convolutional Neural Network and Support Vector Machine method or called Convolutional Support Vector Machine. This method uses convolutional layers in Convolutional Neural Network as feature extraction and continued with Support Vector Machine as the classifier. This research was conducted using different preprocessing and convolutional architectures. There are five stages in this study, namely image cropping, image quality improvement, image feature extraction, feature reduction, and NPDR classification. Classification of retinal images of patients with diabetic retinopathy are important because they will affect the follow-up and treatment that should be obtained in medical terms. This research is helpful to limiting the assets spent and can be a leap forward for individuals who has diabetic retinopathy that come up short on the clinic diagnosing reserves.

This paper is organized as follows. Section 2 discusses the related works about the classification severity level of NPDR in general. Section 3 presents the proposed methodology for the NPDR classification using Convolutional Support Vector machine (CSVM). Section 4 describes the experiments that have been performed and presents their results. Section 5 is about the conclusion and the future work of this research.

2. Related works

There are several researches about image classification by using Convolutional Neural Network (CNN) method and Support Vector Machine (SVM) method. Yun et al. proposed an automatic classification severity level of NPDR by using Neural Network method [10]. This method gives good performance to separate the image of diabetic retinopathy with six extracted features and the accuracy is 72%. Neural Network has some hidden layers that has some neurons on them. The more hidden layers, the better result it gives. Neural

network method that used many hidden layers is called deep learning [11]. One of multilayer perceptron in deep learning is called CNN.

Li et al. conducted classification of lungs image using high resolution computed tomography (HRCT) from interstitial lung disease (ILD) pattern. The result of the classification is good and can give automatic discriminatory feature extraction and can get the better accuracy [12]. The other research about diabetic retinopathy by using CNN with 80.000 image data and get 95% in the sensitivity and 75% in accuracy on 5.000 image validation [3].

Because of the good ability, this method is good for complex problems and big data. As a result, training process on CNN methods will need a long time [13].

The other method for classification is using SVM method [14]. Other research about SVM for attention deficit hyperactivity disorder (ADHD) classification gets 100% for the accuracy. It shows, SVM is very effective for classification of 2 classes and need less time [15]. The other research about SVM is the classification of medical datasets which result is better than Radial Basis Function Network and Naive Bayes classifier. The accuracy is 93.75% [16].

For having a higher classification result, based on the research by [17], by using Principle Component Analysis for reduction data, the study got the higher result by using SVM. That method helps to reduce high dimensional network data to provide the more informative features form the thereby decreasing the execution time for classification and increasing the classification accuracy.

Based on those researches CNN needs a long time for classification. The research by S. Dutta et al. [1] got the accuracy of 78.3% for the best result by using CNN. H. Pratt et al. [3] got the accuracy of 75% and it took 188 minutes for running the image validation. This paper proposes a new methodology that combines CNN and SVM in classifying NPDR. This methodology can shorten the processing time in carrying out the classification process based on research by [14]. This research proposes a new methodology that combines the image enhancement method, CNN architecture as feature extraction method, feature reduction method, and classification model using Support Vector Machine - Naïve Bayes (SVM-NB). There are two experiments of classification presented in this paper. First experiment is three classes classification of NPDR (mild, moderate, and severe) and the second one is two classes classification of NPDR (mild and severe).

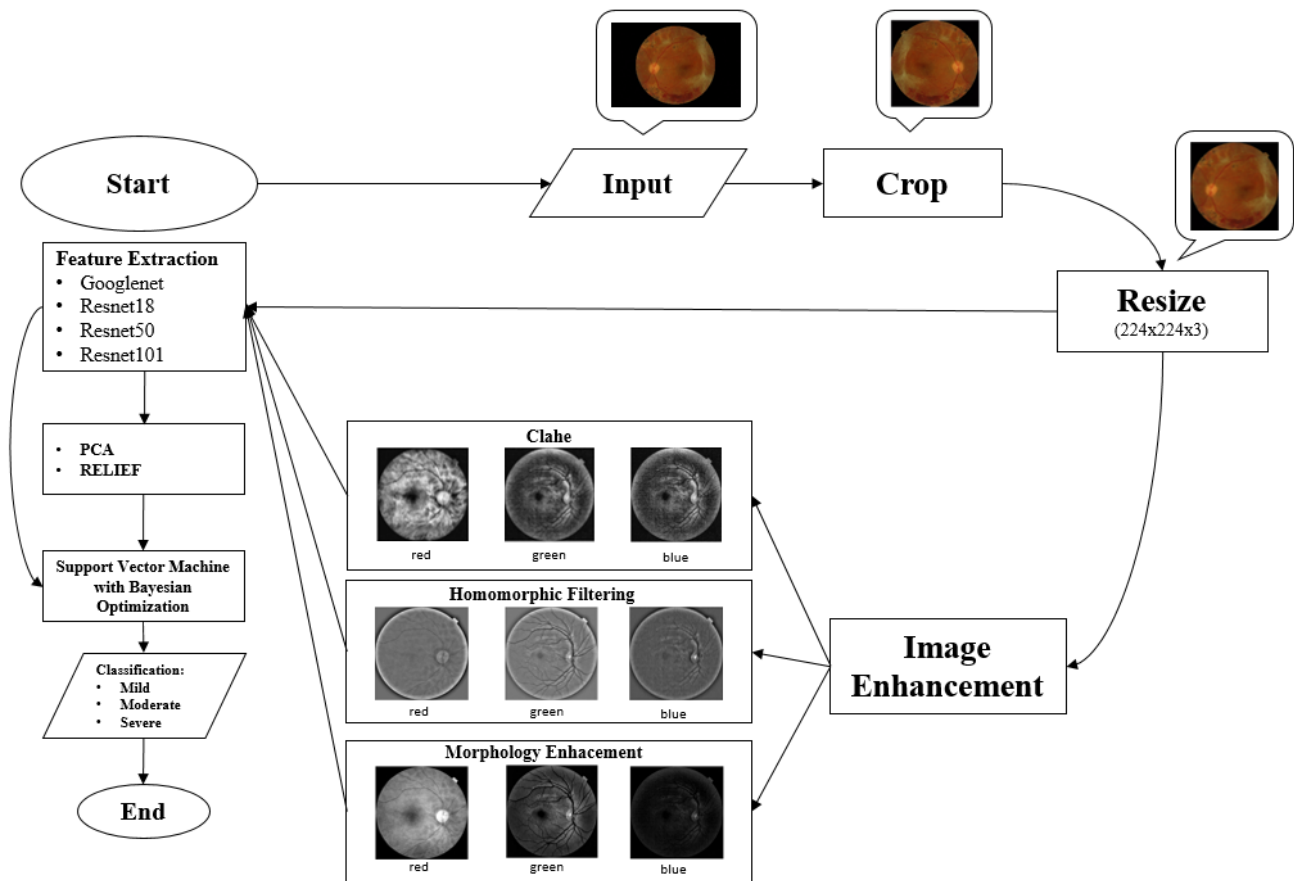


Figure. 1 System experiment model of NPDR severity classification

3. Material and methods

This research is conducted with classifying data into 2 classes (mild, severe) and 3 classes (mild, moderate, severe) on Base21, Base13, and Base12 which are subsets in Messidor Database [18]. In Base21 there are 33 retinal fundus images for 3 classes and 17 images for two classes, in Base13 there are 48 images for 3 classes and 24 images for 2 classes, and in Base12 there are 68 images for 3 classes and 49 images for 2 classes. The data is divided into 75% training data and 25% testing data.

This research has five stages including pre-processing, image enhancement process, feature extraction, feature reduction, and classification. The system experiment severity classification of NPDR model can be seen at Fig. 1.

3.1 Pre-processing

There are 2 processes in pre-processing of this research, those are cropping image and resize image. Cropping image reduces the background area of the image. This makes the retina area more dominant in the image, making the classification process better. This step takes the image from the row minimum - 2

pixels until row maximum + 2 pixels and column minimum - 2 pixels until column maximum + 2 pixels. The two pixels utilized in this step are tolerance limits of the bottom, top, left, and right edges of retina.

The second step is image resizing. In this step, the image obtained from the previous step is resized to 224 x 224 pixels. This process utilizes that size to fit the size of input layer in CNN used (GoogLeNet, Resnet18, Resnet50, ResNet101).

3.2 Image enhancement

Several image enhancements that used in this research are Contrast Limited Adaptive Histogram Equalization, Homomorphic filter, and morphological contrast enhancement. This stage aims to make the image more contrast so that the required image features become more prominent. The enhancement process applies to every channel in RGB image. So that each image is taken 3 times the enhancement process.

3.2.1. Contrast limited adaptive histogram equalization

Contrast Limited Adaptive Histogram Equalization (CLAHE) is one of picture

improvement procedures that has been normally utilized in many picture handling applications. The reason for CLAHE is to improve low-differentiate picture, as a pre-preparing of the info picture. The CLAHE method not only enhances image contrast but also results in better equilibration or histogram equalization [19].

CLAHE is designed based on splitting the picture into many non-overlapping areas of almost the same size. The pixels in this system are mapped by a linear combination of the effects of the mapping of the four closest region recommendations [20].

The steps in the CLAHE algorithm are [21]:

1. The first picture is partitioned into measured sub-pictures which is $M \times N$.
2. Calculates the histogram of each sub-image.
3. Clipped histogram for each sub-image.

The quantity of pixels in the sub-picture is appropriated at each gray degree. The normal number of pixels at each gray degree is defined in Eq. (1).

$$N_{avg} = \frac{N_{CR-Xp} \times N_{CR-Yp}}{N_{gray}} \quad (1)$$

N_{avg} is the average pixel value, N_{gray} is the number of gray degree values in the sub-image, N_{CR-Xp} is the number of pixels in the X dimension from the sub-image, and N_{CR-Yp} is the number of pixels in the Y dimension from the sub-image.

Then calculate the clip limit of a histogram using Eq. (2).

$$N_{CL} = N_{CLIP} \times N_{avg} \quad (2)$$

N_{CL} is the cliplimit and N_{CLIP} is the maximum average pixel value per gray degree value of the sub-image.

In the first histogram, the pixels will be cut if the quantity of pixels is more prominent than N_{CLIP} . The quantity of pixels equally appropriated into every degree of gray (N_d) which is defined by the total number of pixels clipped N_{TC} , the formula used in the Eq. (3).

$$N_d = \frac{N_{TC}}{N_{gray}} \quad (3)$$

The variable M expresses the area size, N states the value of grayscale and α is the clip factor which states the addition of a histogram that is between 0 and 100.

3.2.1. Homomorphic filter

Homomorphic filter is an image enhancement method that works in the frequency domain and is one of the algorithms of Discrete Fourier Transform. The point of picture upgrade is to expand the contrasts between contiguous element zones so as to help exercises, for example, ailment finding, observing and careful making arrangements for disease [22].

The first step is to perform a Discrete Fourier Transformation (DFT) to change the image from the spatial domain of the logarithmic of $f(x, y)$ to the frequency domain [23]. The formula applied to the image can be seen at Eq. (4).

$$F(u, v) = F(\ln(f(x, y))) \\ = F(\ln(i(x, y))) + F(\ln(r(x, y))) \quad (4)$$

$f(x, y)$ is an image of size $M \times N$, $F(u, v)$ is computed at $u = 0, 1, \dots, M-1$ and $v = 0, 1, \dots, N-1$.

The next step is filtering the image with Butterworth lowpass filter (BLPF). It can be seen at Eq. (5) and the filter mask is shown at Eq. (6).

$$I(u, v) = H(u, v) \cdot F(u, v) \quad (5)$$

$$H(u, v) = \frac{1}{1 + \left(\frac{D(u, v)}{D_0}\right)^{2n}} \quad (6)$$

$I(u, v)$ is the filter output in the Fourier domain, $H(u, v)$ is the transfer function of BLPF, $D(u, v)$ is the distance between point (u, v) and the center, D_0 is the cutoff frequency from the center.

Invert-transform $I(u, v)$ into spatial domain and take the exponential to obtain a filtered homomorphic image [24]. The formula utilized in this stage is defined in Eq. (7)

$$f'(x, y) = e^{F^{-1}(I(u, v))} \quad (7)$$

3.2.2. Morphological contrast enhancement

The enhancement of morphological contrast is one system used to improve picture quality dependent on the state of the article. The pixels in the picture are prepared by contrasting the neighboring pixels in the picture with the structure of the segments. The plan of the components can be picked based on the shape and size of the contiguous pixels [25]. Morphology in the computerized world can be deciphered as a method for clarifying or investigating the idea of advanced articles, one of which is picture information.

There are two fundamental morphological activities, widening and disintegration [25].

Morphological operations use two sets of inputs, namely the image and the kernel, where the kernel is a structural element of the image. The structural element is a matrix and is generally small. In this research, the structuring element used is a disk with a radius of 30 pixels.

1. Dilation operation

Dilation operation is an operation technique used to obtain a widening effect on the number of pixels that is worth 1. The widening formula for A mapping to B can be defined in Eq. (8).

$$D(A, B) = A \oplus B = \{x: \hat{B}_x \cap A \neq \emptyset\} \quad (8)$$

where \hat{B} denotes the reflection of B, that is $\hat{B} = \{w|w = -b, b \in B\}$.

The more noteworthy the size of the organizing component, the more prominent the change. The little size of the organizing component can likewise give similar outcomes with a bigger size by re-augmenting.

2. Erosion operation

Erosion operation is a technique that is the opposite of dilation, which is to reduce the structure. If the widening process results in larger objects, in the process of erosion it will produce narrowed objects. Erosion operations can be defined in Eq. (9).

$$E(A, B) = A \ominus B = \{x: B_x \subset X\} \quad (9)$$

3. Opening operation

Opening operation in morphology is a dilation operation that is coming about because of the disintegration activity of a picture. The initial activity expels a little item from the frontal area of the picture, at that point places it out of sight. Erosion operation is very useful for eliminating objects in the image, but it has the downside of reducing the size of the operation. To overcome this, widening operation can be used after erosion surgery using the same elements. The opening operation process can be defined in Eq. (10).

$$O(A, B) = A \circ B = D(E(A, B), B) \quad (10)$$

4. Closing operation

Closing operation in morphology is a combination of erosion operations and widening of an image. The erosion opening operation is preceded by the erosion opening operation preceded by the widening, although in the process of closing the expansion procedure, the erosion is followed first. The effect of the closing procedure is to widen the

outer border of the foreground object and also to close the small hole in the middle of the object, although the effect is not as high as the result of the dilation process. The closing operation can be defined in Eq. (11).

$$C(A, B) = A \cdot B = D(E(A, -B), -B) \quad (11)$$

5. Top-Hat and Bottom-Hat transformation

Top-Transformation is a process that extracts small elements and information from a given image. Top-Hat transformation is an improvement image quality by subtracting the opening process on its own. While, transformation of the Bottom-Hat is an improvement image quality on the original picture of the closing process. Top-Hat transformation can be described by Eq. (12).

$$Top - Hat(A) = A_{TH} = A - (A \circ B) \quad (12)$$

While Bottom-Hat transformation can be defined as Eq. (13).

$$Bottom - Hat(A) = A_{BH} = (A \cdot B) - A \quad (13)$$

3.3 Feature extraction

After the input image enhanced, the next process is the feature extraction of the image by using convolution of the CNN. There are several CNN architectures that can used in this research, which are GoogLeNet, ResNet18, ResNet50, and ResNet 101.

CNN is part of deep learning which is a development of multilayer perceptron. CNN was inspired by human artificial neural networks carried out by Hubel and Wiesel in visual cortex research on the senses of vision of cats [26]. In the CNN, there is a layer that has a 3D structure (width, height, depth) while the depth relates to the number of layers. Based on the form of layer, CNN can usually be divided into 2, i.e. feature extraction layer and fully connected layer [1].

The extraction layer is located after the input layer at the beginning of the architecture. Each layer is made up of multiple layers and each layer is made up of nodes connected to the previous layer. There are two types of layers in the extraction layer functionality, namely the convolution layer and the pooling layer.

The convolutional layer is classified into two, namely convolution layer 1 dimension which is used in vector-shaped data such as signals, time series, etc. [27], and layer 2 dimension convolution that is used in two dimensional data, such as images and others.

The CNN architecture of dimension 1 layer and dimension 2 layer can be seen at Fig. 2.

Pooling layer or subsampling is a reduction in matrix size. There are two types of pooling layers namely max pooling and average pooling that can be seen at Fig. 3.

a. Max pooling

Stride is a parameter that determines how many pixels a filter is shifted. The smaller the stride, the more detailed information obtained from an input, but does not always get good performance [28]. If the value of stride is 1, then the convolution process will shift by 1. Similarly, stride 2 and so on [29].

ReLU activation function is an activation layer that applies $f(x) = \max(0, x)$ where ReLU basically only creates a limit on zeros, the point is if $x \leq 0$ then $x = 0$ and if $x > 0$ then $x = x$.

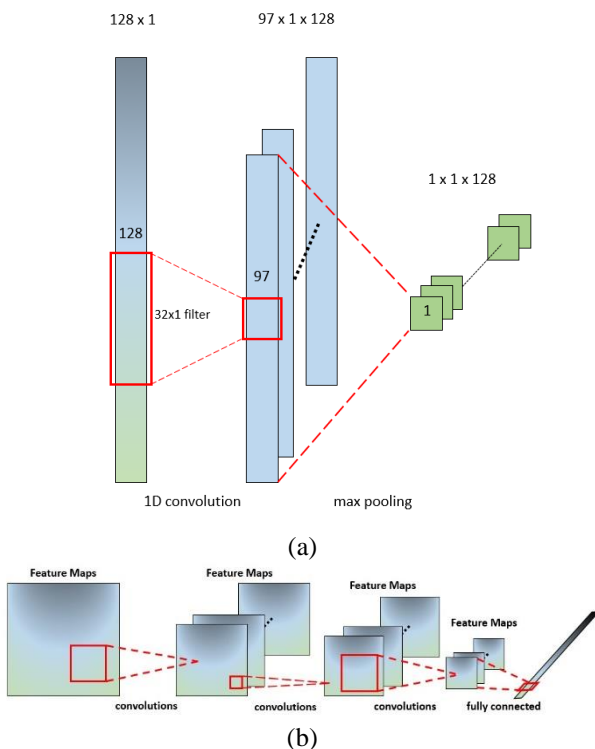


Figure. 2 Convolutional Neural Network architecture: (a) 1D and (b) 2D

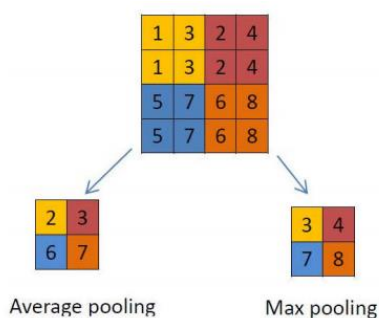


Figure. 3 Pooling layer

Max pooling is a reduction in the size of the matrix by taking the largest value or the maximum value in the sub-regions of the matrix.

b. Average pooling

Average pooling is a reduction in the size of the matrix by taking the average value in the sub-regions of the matrix.

Fully Connected Layers consist of several layers, each layer composed of nodes that are fully connected to the previous layer. The fully connected layer uses a multilayer perceptron that functions to process data so that the desired results are obtained. But, the fully connected layer is not used in this study. This layer is replaced with SVM to classify the severity level of NPDR.

The size of resulted data from this stage are 2048 features. These features are reduced at the next stage before proceeding to the classification stage.

3.3.1. GoogLeNet

GoogLeNet is one type of architecture on the CNN method created by Szegedy et al. [30]. GoogLeNet has inception modules, which carry out various convolution and unify filters for the next layer [31]. The main characteristic of this model architecture is the good utilization of computing resources in the network. The inception module of GoogLeNet can be seen in Fig. 5.

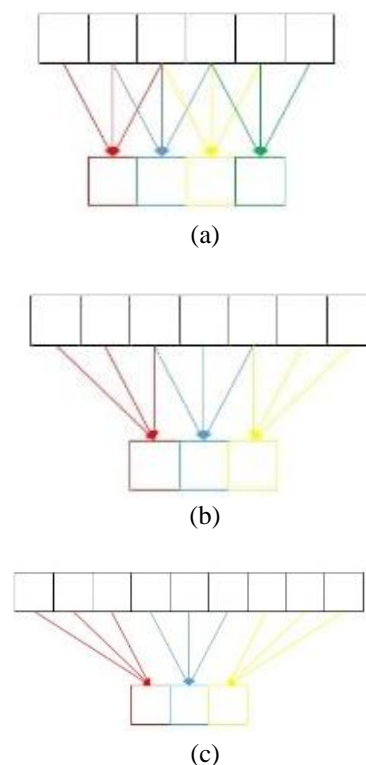


Figure. 4 (a) Stride 1, (b) stride 2, and (c) stride 3

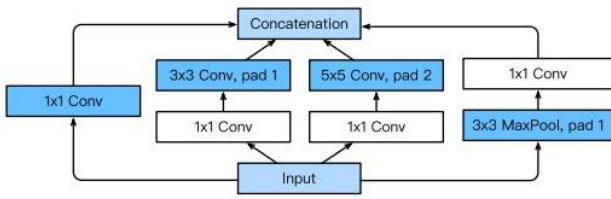


Figure. 5 The inception module of GoogLeNet

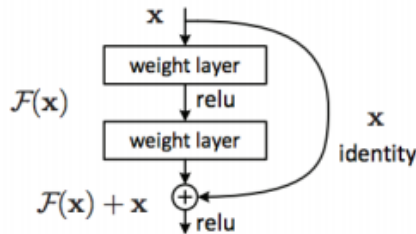


Figure. 6 Residual block in ResNet architecture [33]

The architecture of GoogLeNet uses 3 size filters, those are 1×1 , 3×3 , 5×5 . It consists of 22 layers and it lessens the number of the parameters from 60 million to 4 million. This architecture determines the best weight during network training and naturally selects the correct features [32].

3.3.2. Residual network

Residual Neural Network (ResNet) is one type of architecture in the CNN method created by Kaiming He et al. [33]. The ResNet architecture is quite revolutionary because this architecture became the state of the art at that time, namely in classification, object detection, and semantic segmentation. The difference between ResNet and other methods is that there are residual blocks as shown in Fig. 6. ResNet has several architecture such as 18, 34, 50, 101 even 152 layer [33].

The results of each filter in the ResNet architecture will pass average pooling and enter the fully connected layer network with softmax activation function to determine the classification results [34].

Softmax is an activation function commonly used to calculate probabilities that are commonly used to do multi-class classification, where softmax values are between 0 to 1 and have a number of 1 if all the elements are added using Eq. (14) [35].

$$\text{softmax}(x)_i = \frac{\exp(x_i)}{\sum_{j=1}^n \exp(x_j)} \quad (14)$$

This function is used at the end of the layer of the fully connected layer that is used to produce the probability value of an object function against an existing class.

3.4 Feature reduction

In this stage, the size of data that generated from previous stage is reduced from 2048 to 50. This reduction is the process of selecting several important and main features of each data. Some of these features have the characteristics of each data and can represent the data itself.

3.4.1. Principal component analysis

Principal Component Analysis (PCA) is a popular method for reducing dimensional features. PCA ventures information onto another space where back to back measurements contain less and less of the change of the first dataspace and packs the most significant data onto a subspace with lower dimensionality than the first space. Specific steps of PCA algorithm are:

1. Calculate the covariance matrix by using Eq. (15).

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (15)$$

Where X and Y is the data, \bar{X} and \bar{Y} is the average of the data.

2. Calculate the eigenvalues by using Eq. (16).

$$\det(C - \lambda I) = 0 \quad (16)$$

Where \det is the determinant, C is the covariance matrix, I is the identity matrix, and λ is the eigenvalue.

3. Calculate eigenvector by using Eq. (17).

$$(C - \lambda I)X = 0 \quad (17)$$

Where C is the covariance matrix, I is the identity matrix, λ is the eigenvalue, and X is the eigenvector.

4. Sort the eigenvectors based on the largest eigenvalues.
5. Obtain the transformed data using the sorted eigenvectors

3.4.2. Relief algorithm

Relief algorithm is a type of feature weighting algorithm initially proposed by Kira [36], which is utilized to classify two types of data. This method is categorized as filter-based features reduction [37]. The basic stages of Relief Algorithm are as adopts [36]:

Info: include vector set of preparing models and class label; /* N – feature dimensions, m – iterations */ [38].

Output: weight vector W which corresponds to the various parts of eigenvector.

- 1) $W[A] := 0.0$;
- 2) For $i := 1$ to m
- 3) Choose concrete examples R at random;
- 4) Find a similar nearest neighbor denoted H and a unsimilar nearest neighbor denoted M ;
- 5) For $A := 1$ to all attributes

$$W[A] := W[A] - \frac{\text{diff}(A, R, H)}{m} + \frac{\text{diff}(A, R, M)}{m}$$
- 6) End;
- 7) End;

Where: function $\text{diff}(\text{Feature}, \text{Instance}_1, \text{Instance}_2)$ is utilized to measure the difference in features between two distinct samples, which is described as:

For discrete features:

$$\text{diff}(F, I_1, I_2) = \begin{cases} 0; & \text{value}(F, I_1) = \text{value}(F, I_2) \\ 1; & \text{others} \end{cases}$$

For continuous features:

$$\text{diff}(F, I_1, I_2) = \frac{|\text{value}(F, I_1) - \text{value}(F, I_2)|}{\max(F) - \min(F)}$$

where I_1, I_2 are two samples, (F, I_1) refers to the F eigenvalue of samples I_1 .

In the wake of comprehending the significance loads W of different highlights and classes, we can sort it out. In addition, characters whose importance is more prominent than a limit esteem establish the last element subset, taking out invalid component.

3.5 Classification using SVM-NB

Support Vector Machine - Naïve Bayes (SVM-NB) is a coordinated effort of Support Vector Machine (SVM) and Naïve Bayes (NB) technique. SVM is able to find a perfect hyperplane to isolate planning tests into two groupings. In view of the objective of SVM is to boost the separation between the limits of two classes, the blending of these two limits won't show up. Because of the independence assumption, the review rate and precision of NB based arrangements are altogether influenced. To address that issue, we apply a help vector machine (SVM) based cutting strategy to dispose of tests that are ordered into wrong classes by naïve-bayes (NB) [39].

In the Support Vector Machine - Naïve Bayes (SVM-NB), the preparation tests are first procedures by the first naïve bayes calculation. For each element vector extricated from the preparation set, there will be a comparing classification created by the naïve bayes calculation. Subsequently, we will have a lot of results $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$, where $x_1 \in R^n$, $y_1 \in \{-1, +1\}$, $i = 1, 2, \dots, m$, and n define the dimension of the feature.

First, find the nearest neighbor(s) for each function. For the variable vector, if its nearest neighbor and it belongs to the same group, the vector will be held. Otherwise, the vector would be excluded from the set of arrangements. The vector can be discarded if it is placed in the wrong group, due to the dependence between itself and the nearest vector.

Given to feature vectors $\vec{u} = (u_1, u_2, \dots, u_n)$ and $\vec{v} = (v_1, v_2, \dots, v_n)$, their distance is defined as Eq. (18).

$$D(\vec{u}, \vec{v}) = \sqrt{\sum_{k=1}^n (u_k - v_k)^2} \quad (18)$$

In the algorithm, we utilize $\vec{X} = \{x_1, x_2, \dots, x_m\}$ and $\vec{Y} = \{y_1, y_2, \dots, y_m\}$ to denote the feature vectors and corresponding categories. Let $\vec{V} = \{v_1, v_2, \dots, v_m\}$ signify the grouping results created by the first naïve bayes calculation. The characterization results got from the first naïve bayes will be refined by support vector machine naïve bayes (SVM-NB) [40].

4. Experiments and analysis

This research used 3 classes which are mild, moderate, and severe NPDR. The examples of NPDR for each class can be seen in Fig. 7. The result of image enhancement can be seen in Fig.8. Fig. 8 shows the results of an enhancement process to one example of a retinal color image. For each enhancement method used, there are three result images for each channel (Red, Green, Blue) which are shown sequentially in Fig. 8.

After passing enhancement, this research utilizes feature extraction on the architecture of ResNet (ResNet18, ResNet50, ResNet101) and GoogLeNet which can be seen in the Fig. 9, while the flowchart diagram of ResNet can be seen in Fig. 10. Fig. 9 shows the running steps on GoogLeNet and ResNet so that feature extraction stage only stops at Global Average Pooling. Fig. 10 clarifies the ResNet block section utilized in Fig. 9.

Both in the ResNet and GoogLeNet architecture, before entering the fully connected layer, the feature will go through average pooling. Because here only the results of the feature extraction will be used, then it is used to cut off the average pooling layer then to the fully connected layer. Then the classification layer is replaced using support vector machine naïve bayes (SVM-NB).

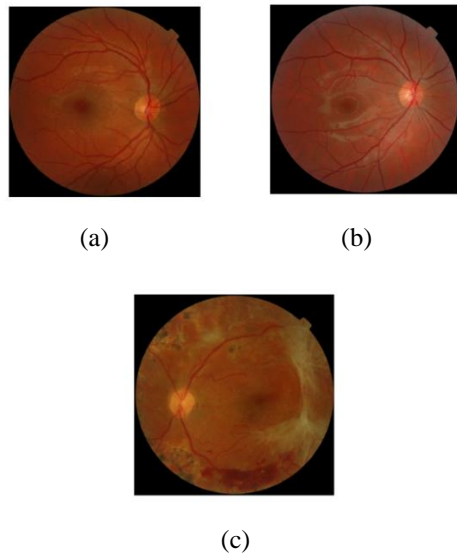


Figure. 7 Examples of image data: (a) mild, (b) moderate, and (c) severe

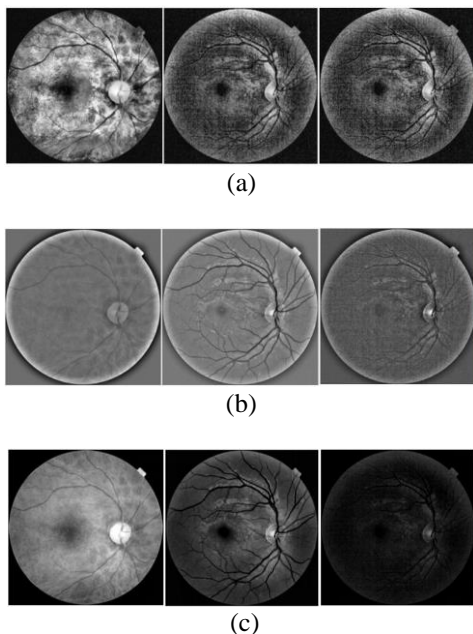


Figure. 8 The process and result of image enhancement: (a) CLAHE of mild NPDR, (b) Homomorphic of mild NPDR, and (c) morphological contrast enhancement of mild NPDR

Some of the experiments are carried out on the database involving all processes, from the cropping data, resizing, enhancement, to classifying. The whole processes can be seen in Fig. 1.

The performance of the system on the image classification is measured by accuracy. Accuracy measures the number of data that were correctly classified by the system. The formula of calculating the accuracy is shown in Eq. (19).

Based on the experiments that have been done, the system accuracy results are as shown in Tables 1 and 2. Table 1 shows the using of principle

component analysis for the classification scheme and Table 2 presents the utilizing of Relief for the classification stage. The experiments were carried out by classifying data on Base21, Base 13, and Base 12 into 2 classes (mild, severe) and 3 classes (mild, moderate, severe). The steps taken are cropping and resizing data, then image enhancement using CLAHE, morphology, and Homomorphic. The results of the image enhancement are used as input to the feature extraction layer which in this study used GoogLeNet, ResNet18, ResNet50, and ResNet101. Then the data is reduced to 50 using PCA and relief before proceeding the classification layer using SVM-NB.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

Table 1 shows that the best two average accuracy of three classes classification on Base 21 are using CLAHE and morphological contrast enhancement and principle component analysis in the feature reduction stage. They both have the same accuracy score of 75%. The two best CNN architectures used in that case are GoogLeNet for CLAHE and ResNet50 for morphological contrast enhancement. This can be seen from the high accuracy of both 87.5%. From the same table, the two perfect accuracy can be seen from all CNN architecture using morphological contrast enhancement as the pre-processing and without pre-processing steps. All of them can reach 100% accuracy. The best result for three classes classification from data on Base 13 is using Homomorphic, ResNet18 and Principle component analysis with the accuracy of 81.82% and the average accuracy using Homomorphic and all CNN architecture is 70.45%. Two classes classification on the same case results the three-best average of accuracy are using all image enhancement methods with the same accuracy score of 95%. The last experiment using principle component analysis on Base 12 results the best average accuracy obtained in three classes classification by using CLAHE as an image enhancement method. This accuracy score reaches 70.59%. The best accuracy from that case is obtained by using GoogLeNet as feature extraction method with 82.35%. Two classes classification results the best average accuracy is obtained by using Homomorphic and principle component analysis with 97.92%. There are three CNN architecture can reach the perfect accuracy (100%) in that case. They are GoogLeNet, ResNet50, and ResNet101.

Table 2 presents the best average accuracy of three classes classification on Base 21 using Relief

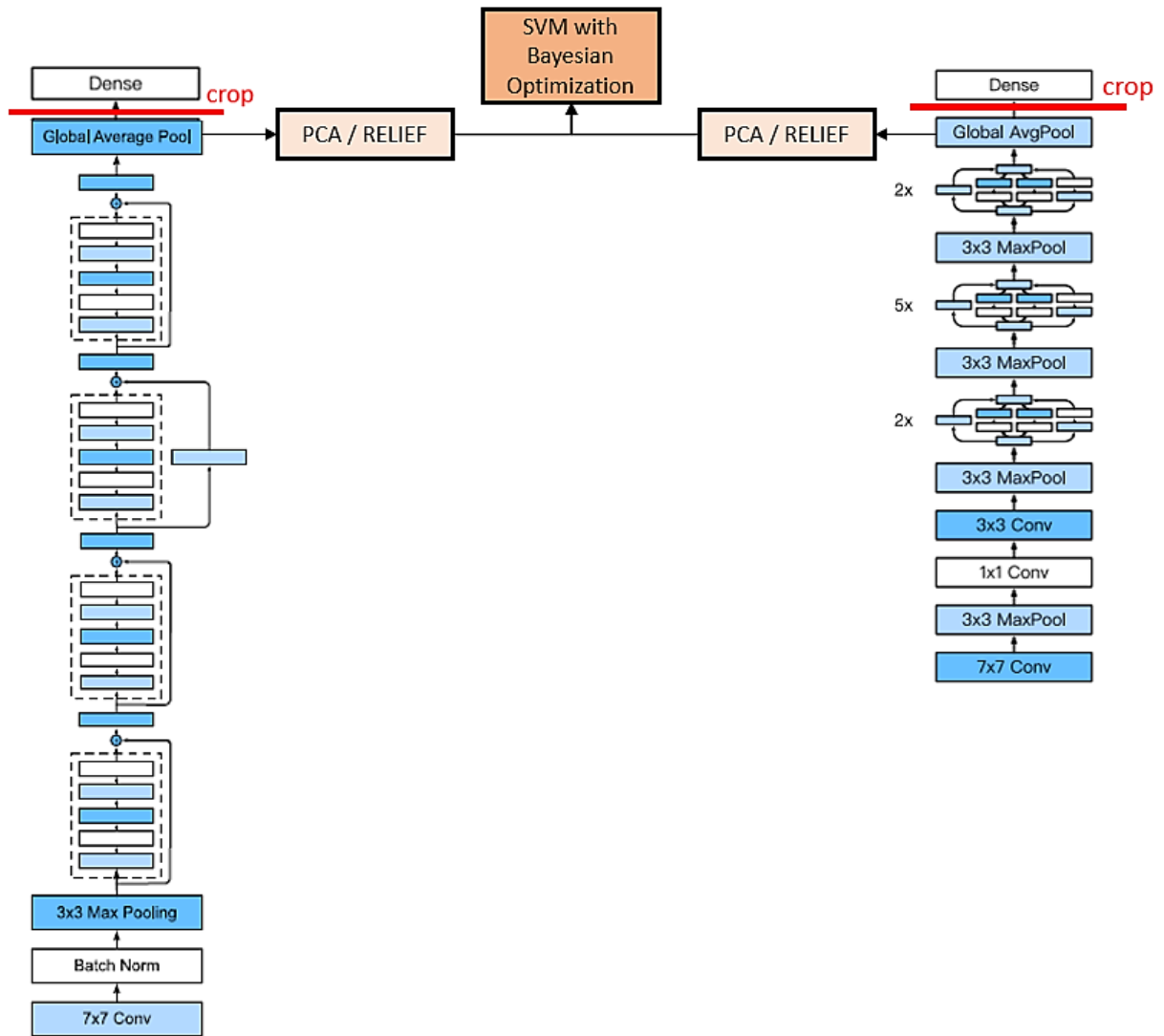


Figure. 9 Classification flowcharts

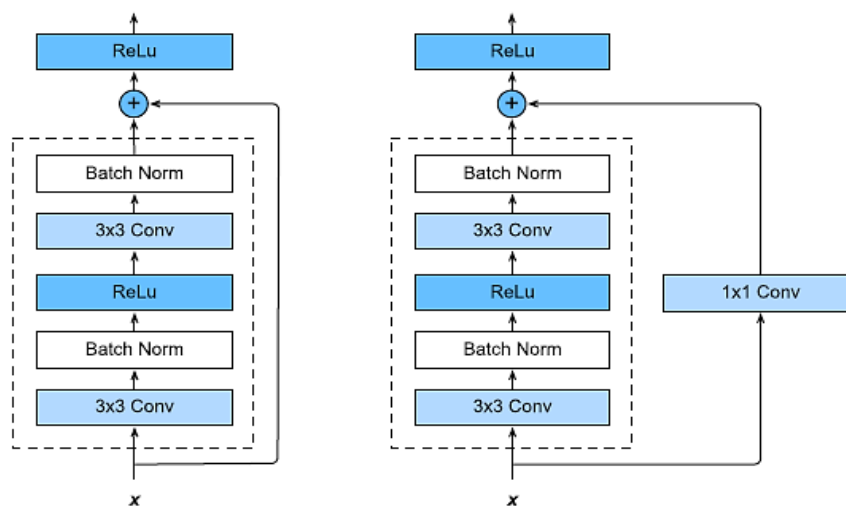


Figure. 10 ResNet block

obtained by using Homomorphic with 84.38% and the best result are shown by three ResNet (ResNet18,

ResNet50, ResNet101) architecture. They have same accuracy score of 87.50%. The perfect average

accuracy of two classes classification using Relief on the same data is achieved by using all image enhancement methods and CNN architecture. All of them can achieve 100% accuracy. The two best average accuracy of three classes classification on Base 13 using Relief are obtained by using CLAHE and without enhancement with the same accuracy score of 81.82%. The best CNN architecture utilized in this condition is ResNet50. The perfect accuracy reached by using ResNet50 for CLAHE or without enhancement. They can achieve 100% accuracy for the classification. The perfect average accuracy of two classes classification on same base and feature reduction are achieved by using CLAHE and Homomorphic with 100%.

The last research is using Relief on Base 12 results the two best average accuracy of three classes classification are obtained by morphological contrast enhancement and Homomorphic with the same accuracy. The accuracy resulted from each method is

80.88%. Homomorphic and ResNet101 present the best accuracy result for this case. These methods show the accuracy score of 94.12%.

From comparison of results in Tables 1 and 2, the Relief method looks better than principle component analysis. Those tables present the average accuracy classification using principle component analysis of 78.73% while using Relief of 89.12%. Furthermore, applying image enhancement methods produces better results than without the enhancement. This can be seen from the results of applying without image enhancement cannot surpass the results of applying the image enhancement method.

Fig. 11 presents the highest accuracy for NPDR classification reached by using Homomorphic as an image enhancement method. This accuracy achieved is 73.95% of three classes classification and 97.78% of two classes classification and the average is 85.87%.

Table 1. Results of classification using Principle Component Analysis
Base21

CSVM	Three Classes				Two Classes			
	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)
GoogLeNet	62.50	87.50	62.50	62.50	100.00	50.00	100.00	100.00
ResNet18	62.50	75.00	75.00	75.00	100.00	100.00	100.00	100.00
ResNet50	62.50	75.00	87.50	75.00	100.00	100.00	100.00	100.00
ResNet101	87.50	62.50	75.00	75.00	100.00	75.00	100.00	75.00
Average	68.75	75.00	75.00	71.88	100.00	81.25	100.00	93.75

CSVM	Three Classes				Two Classes			
	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)
GoogLeNet	54.55	54.55	45.45	72.73	80.00	100.00	100.00	100.00
ResNet18	63.64	63.60	45.45	81.82	60.00	80.00	80.00	80.00
ResNet50	72.73	72.73	72.73	63.64	80.00	100.00	100.00	100.00
ResNet101	81.82	63.64	72.73	63.64	100.00	100.00	100.00	100.00
Average	68.18	63.64	59.09	70.45%	80.00	100.00	100.00	100.00

CSVM	Three Classes				Two Classes			
	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)
GoogLeNet	52.94	82.35	76.47	41.18	100.00	83.33	75.00	100.00
ResNet18	76.47	52.94	58.82	58.82	100.00	58.33	58.33	91.67
ResNet50	70.59	70.59	76.47	76.47	91.67	66.67	58.33	100.00
ResNet101	70.59	76.47	64.71	58.82	91.67	83.33	66.67	100.00
Average	67.65	70.59	69.12	58.82	95.83	72.92	64.58	97.92

Table 2. Results of classification using Relief

Base21								
CSVM	Three Classes				Two Classes			
	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)
GoogLeNet	62.50	87.50	75.00	75.00	100.00	100.00	100.00	100.00
ResNet18	62.50	62.50	87.50	87.50	100.00	100.00	100.00	100.00
ResNet50	87.50	75.00	75.00	87.50	100.00	100.00	100.00	100.00
ResNet101	100.00	100.00	87.50	87.50	100.00	100.00	100.00	100.00
Average	78.13	81.25	81.25	84.38	100.00	100.00	100.00	100.00

Base13								
CSVM	Three Classes				Two Classes			
	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)
GoogLeNet	54.55	81.82	72.73	72.73	100.00	100.00	80.00	100.00
ResNet18	81.82	63.64	63.64	72.73	80.00	100.00	100.00	100.00
ResNet50	100.00	100.00	81.82	90.91	100.00	100.00	100.00	100.00
ResNet101	90.91	81.82	72.73	72.73	100.00	100.00	100.00	100.00
Average	81.82	81.82	72.73	77.27	95.00	100.00	95.00	100.00

Base12								
CSVM	Three Classes				Two Classes			
	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)	Without Enhancement (%)	CLAHE (%)	Morph (%)	Homomorphic (%)
GoogLeNet	76.47	94.12	82.35	76.47	100.00	100.00	100.00	100.00
ResNet18	88.24	70.59	82.35	88.24	100.00	91.67	100.00	100.00
ResNet50	76.47	76.47	76.47	64.71	100.00	100.00	100.00	100.00
ResNet101	76.47	76.47	82.35	94.12	100.00	100.00	91.67	100.00
Average	79.41	79.41	80.88	80.88	100.00	91.67	97.92	100.00

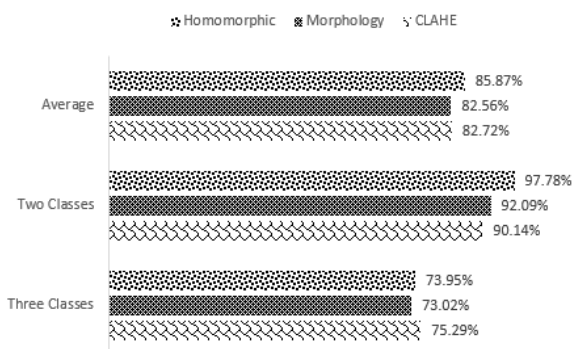


Figure. 11 Comparison result of image enhancement methods

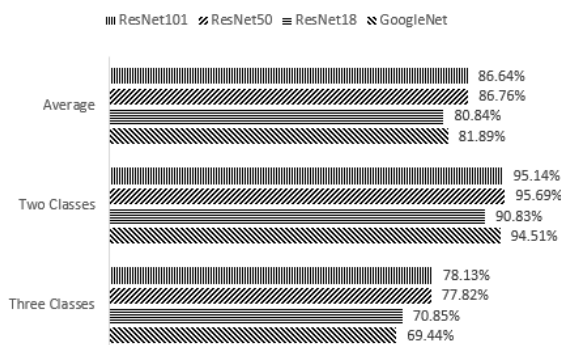


Figure. 12 Comparison result of CNN architecture

Fig. 12 shows the comparison of accuracy using each CNN architecture. The best result for three classes classification is ResNet101 and ResNet50 for two classes classification. Overall, ResNet50 has the highest accuracy for NPDR classification. The accuracy achieved is 86.76%.

Based on those researchers CNN needs a long time for classifying. The research by S. Dutta et al. [1] got the accuracy of 78.3% for the best result by using CNN, H. Pratt et al. [3] got the accuracy of 75% and it took 188 minutes for running the image validation.

In addition, this research also compares the processing time between the CNN and CSVM-NB methods by using the same device. As shown on the Table 3, the CNN method needs longer time for training and testing data than the CSVM method.

The experiment in Table 3 utilizes resnet18 architecture on base21. There is a significant difference on the execution time by using CNN, CSVM without feature reduction, and CSVM with PCA/Relief as the feature reduction. On CNN, the best execution time for training is by using homomorphic as its image enhancement. The time

Table 3. Comparison of time consuming of CNN and CSVM using base21

	CNN				CSVM		
	Number of Classes	Training Time (minutes)	Testing Time (minutes)	Accuracy (%)	Training Time (minutes)	Testing Time (minutes)	Accuracy (%)
CLAHE	2	164.8051	0.1221	50.00	0.0663	0.0106	100.00
	3	628.0797	0.1807	50.00	0.0707	0.0307	75.00
Homomorphic	2	164.2606	0.1169	100.00	0.0888	0.0310	100.00
	3	623.3857	0.1914	62.50	0.0900	0.0310	62.50
Morphology	2	226.8856	0.1575	25.00	0.0859	0.0266	100.00
	3	652.4669	0.2096	50.00	0.0887	0.0267	50.00

CSVM with Feature Reduction Methods						
	Number of Classes	Feature Reduction Methods	Training Time (minutes)	Testing Time (minutes)	Accuracy (%)	
CLAHE	2	PCA	0.0567	0.0106	100.00	
		Relief	0.0697	0.0105	100.00	
	3	PCA	0.0597	0.0105	75.00	
		Relief	0.0746	0.0106	62.50	
Homomorphic	2	PCA	0.0581	0.0310	100.00	
		Relief	0.0888	0.0312	100.00	
	3	PCA	0.0603	0.0311	75.00	
		Relief	0.0968	0.0311	87.50	
Morphology	2	PCA	0.0575	0.0266	100.00	
		Relief	0.0876	0.0266	100.00	
	3	PCA	0.0581	0.0266	75.00	
		Relief	0.0931	0.0267	87.50	

consuming recorded using that method in training is 164,261 minutes for the classification of 2 classes and 623,386 minutes for the 3 classes with an accuracy rate of 100 % for 2 classes and 62.5 % for 3 classes. The time required for this process is in stark contrast to the time consumed by the CSVM method without feature reduction and CSCM which used feature reduction.

The CSVM without feature reduction has the same image enhancement, that is homomorphic. The time consuming recorded using that method in training is 0.0888 minutes for the classification of 2 classes and 0.0968 minutes for the 3 classes. While the resulting accuracy is the same as CNN, but accuracy on CSVM with pre-processing is more excellent than CNN.

With those comparisons, CSVM with PCA and relief are more excellent in the terms of accuracy and time, the results give the better accuracy and faster classifying than CNN and CSVM without feature reduction.

5. Conclusion

NPDR classification is indispensable in making the appropriate treatment for sufferers. Therefore, an automatic NPDR classification system is needed to help the ophthalmologist and the community to give

the best treatment for patients according to their severity.

This research proposes a new methodology that combine the image enhancement method, CNN architecture as extraction feature method, feature reduction method, and classification model using SVM-NB. There are two experiments of classification presented in this paper. First experiment is three classes classification of NPDR (mild, moderate, and severe) and the second is two classes classification of NPDR (mild and severe).

The three best results from the experiments achieved by Homomorphic for image enhancement method, ResNet50 for the CNN architecture as feature extraction method, and Relief for feature reduction. The Homomorphic obtains 85.87% accuracy, ResNet50 can achieve 86.76% accuracy, and the Relief can reach 89.12% accuracy.

With those comparisons for testing, CSVM with PCA and relief are more excellent from their accuracy and time, the results give better accuracy and faster classifying than CNN and CSVM without feature reduction.

Furthermore, for the next research further segmentation needs to be added to existing lesions such as microaneurysms, hemorrhages, and exudates. This research is expected to improve the results of NPDR classification.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Ricky Eka Putra, and Handayani Tjandrasa; methodology, Ricky Eka Putra, and Handayani Tjandrasa; software, Ricky Eka Putra; validation, Ricky Eka Putra, Handayani Tjandrasa, and Nanik Suciati; formal analysis, Ricky Eka Putra, Handayani Tjandrasa, and Nanik Suciati; investigation, Ricky Eka Putra, Handayani Tjandrasa, and Nanik Suciati; resources, Ricky Eka Putra, and Handayani Tjandrasa; data curation, Ricky Eka Putra, and Handayani Tjandrasa; writing—original draft preparation, Ricky Eka Putra; writing—review and editing, Handayani Tjandrasa, and Nanik Suciati; visualization, Handayani Tjandrasa, and Nanik Suciati; supervision, Handayani Tjandrasa, and Nanik Suciati.

Acknowledgments

Messidor database is kindly provided by the Messidor program partners (see <http://www.adcis.net/en/third-party/messidor/>).

References

- [1] S. Dutta, B. C. S. Manideep, S. M. Basha, R. D. Caytiles, and N. C. S. N. Iyengar, “Classification of Diabetic Retinopathy Images by Using Deep Learning Models”, *Int. J. Grid Distrib. Comput.*, Vol. 11, No. 1, pp. 89–106, 2018.
- [2] Y. Zhao, Y. Zheng, and Y. Liu, “Intensity and Compactness Enabled Saliency Estimation for Leakage Detection in Diabetic and Malarial Retinopathy”, *IEEE Trans. Med. Imaging*, Vol. 36, No. 1, pp. 51–63, 2017.
- [3] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, “Convolutional Neural Networks for Diabetic Retinopathy”, *Procedia Comput. Sci.*, Vol. 90, No. 1, pp. 200–205, 2016.
- [4] A. D. Association, “Diagnosis and Classification of Diabetes Mellitus”, *Diabetes Care*, Vol. 36, No. SUPPL.1, pp. 67–74, 2013.
- [5] V. Raman, P. Then, and P. Sumari, “Proposed retinal abnormality detection and classification approach: Computer aided detection for diabetic retinopathy by machine learning approaches”, In: *Proc. of 2016 8th IEEE Int. Conf. Commun. Softw. Networks, ICCSN 2016*, pp. 636–641, 2016.
- [6] I. Nedeljkovic, “Image Classification Based on Fuzzy Logic”, *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, Vol. 34, pp. 1–6.
- [7] S. Cagnoni, E. Lutton, and G. Olague, “Genetic and evolutionary computation for image processing and analysis”, *Eurasip B. Ser. Signal Process. Commun.*, Vol. 8, No. 4, pp. 1–22, 2008.
- [8] L. Palagi, A. Pesyridis, E. Sciubba, and L. Tocci, “Machine Learning for the prediction of the dynamic behavior of a small scale ORC system”, *Energy*, Vol. 166, No. 2, pp. 72–82, 2019.
- [9] S. J. Lee, T. Chen, L. Yu, and C. H. Lai, “Image Classification Based on the Boost Convolutional Neural Network”, *IEEE Access*, Vol. 6, No. 3, pp. 12755–12768, 2018.
- [10] W. L. Yun, U. Rajendra Acharya, Y. V. Venkatesh, C. Chee, L. C. Min, and E. Y. K. Ng, “Identification of different stages of diabetic retinopathy using retinal optical images”, *Inf. Sci. (Ny)*, Vol. 178, No. 1, pp. 106–121, 2008.
- [11] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review”, *Neurocomputing*, Vol. 187, pp. 27–48, 2016.
- [12] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, “Medical image classification with convolutional neural network”, In: *Proc. of 2014 13th Int. Conf. Control Autom. Robot. Vision, ICARCV 2014*, Vol. 2014, No. 3, pp. 844–848, 2014.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, In: *Proc. of the 25th International Conference on Neural Information Processing Systems 2012*, Vol. 1, pp. 1097–1105, 2012.
- [14] A. Z. Foady, D. Candra, R. Novitasari, and A. H. Asyhar, “Diabetic Retinopathy: Identification and Classification using Different Kernel on Support Vector Machine”, In: *Proc. of International Conference on Mathematics and Islam 2018*, pp. 72–79, 2018.
- [15] J. C. Bledsoe, D. Xiao, A. Chaovalitwongse, and S. Mehta, “Diagnostic Classification of ADHD Versus Control: Support Vector Machine Classification Using Brief Neuropsychological Assessment”, *J. Atten. Disord.*, p. 108705471664966, 2016.
- [16] P. Janardhanan, L. Heena, and F. Sabika, “Effectiveness of support vector machines in medical data mining”, *J. Commun. Softw. Syst.*, Vol. 11, No. 1, pp. 25–30, 2015.
- [17] A. George, “Anomaly Detection based on Machine Learning Dimensionality Reduction using PCA and Classification using SVM”, *Int. J. Comput. Appl.*, Vol. 47, No. 21, pp. 5–8, 2012.
- [18] H. Tjandrasa, R. E. Putra, A. Y. Wijaya, and I. Arieshanti, “Classification of non-proliferative

- diabetic retinopathy based on hard exudates using soft margin SVM”, In: *Proc. of 2013 IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE 2013*, pp. 376–380, 2013.
- [19] S. Jenifer, S. Parasuraman, and A. Kadirvelu, “Contrast enhancement and brightness preserving of digital mammograms using fuzzy clipped contrast-limited adaptive histogram equalization algorithm”, *Appl. Soft Comput. J.*, Vol. 42, pp. 167–177, 2016.
- [20] A. M. Reza, “Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement”, *J. VLSI Signal Process. Syst. Signal Image. Video Technol.*, Vol. 38, No. 1, pp. 35–44, 2004.
- [21] G. Yadav, S. Maheshwari, and A. Agarwal, “Contrast Limited Adaptive Histogram Equalization Based Enhancement for Real Time Video System”, In: *Proc. of 2014 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2014*, pp. 2392–2397, 2014.
- [22] J. Oh and H. Hwang, “Feature enhancement of medical images using morphology-based homomorphic filter and differential evolution algorithm”, *Int. J. Control. Autom. Syst.*, Vol. 8, No. 4, pp. 857–861, 2010.
- [23] P. S. Vikhe and V. R. Thool, “Contrast enhancement in mammograms using homomorphic filter technique”, In: *Proc. of 2016 Int. Conf. Signal Inf. Process. IConSIP 2016*, Vol. 413637, 2017.
- [24] H. Tjandrasa, A. Wijayanti, and N. Suciati, “Optic nerve head segmentation using hough transform and active contours”, *Telkonnika*, Vol. 10, No. 3, pp. 531–536, 2012.
- [25] I. M. O. Widyantara, I. M. D. P. Asana, N. M. A. E. D. Wirastuti, and I. B. P. Adnyana, “Image enhancement using morphological contrast enhancement for video based image analysis”, In: *Proc. of 2016 Int. Conf. Data Softw. Eng. ICoDSE 2016*, 2017.
- [26] D. H. Hubel and T. N. Wiesel, “Receptive Fields of Single Neurones in The Cat’s Striate Cortex”, *J. Physiol.*, pp. 574–591, 1959.
- [27] Y. Kim, “Convolutional neural networks for sentence classification”, *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 1746–1751, 2014.
- [28] E. N. Arrofiqoh and H. Harintaka, “Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Resolusi Tinggi”, *Geomatika*, Vol. 24, No. 2, p. 61, 2018.
- [29] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation”, In: *Proc. of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.
- [30] C. Szegedy, W. Liu, Y. Jia, and P. Sermanet, “Going Deeper With Convolutions”, In: *Proc. of IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Vol. 7, pp. 1–9, 2015.
- [31] P. Ballester and R. M. Araujo, “On the performance of googlenet and alexnet applied to sketches”, In: *Proc. of 30th AAAI Conf. Artif. Intell. AAAI 2016*, pp. 1124–1128, 2016.
- [32] J. H. Kim, S. Y. Seo, C. G. Song, and K. S. Kim, “Assessment of Electrocardiogram Rhythms by GoogLeNet Deep Neural Network Architecture”, *J. Healthc. Eng.*, Vol. 2019, 2019.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, In: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Vol. 2016–Decem, pp. 770–778, 2016.
- [34] P. Napoletano, F. Piccoli, and R. Schettini, “Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity”, *Sensors*, Vol. 18, No. 1, 2018.
- [35] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. 2016.
- [36] K. Kira and L. A. Rendell, “The Feature Selection Problem: Traditional Methods and a New Algorithm”, *Stud. Syst. Decis. Control*, Vol. 256, pp. 129–134, 1992.
- [37] Y. Yamasari, S. M. S. Nugroho, K. Yoshimoto, H. Takahashi, and M. H. Purnomo, “Identifying Dominant Characteristics of Students’ Cognitive Domain on Clustering-based Classification”, *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 1, pp. 167–180, 2020.
- [38] L. Gao, T. Li, L. Yao, and F. Wen, “Research and application of data mining feature selection based on relief algorithm”, *J. Softw.*, Vol. 9, No. 2, pp. 515–522, 2014.
- [39] W. Feng, J. Sun, L. Zhang, C. Cao, and Q. Yang, “A support vector machine based naive Bayes algorithm for spam filtering”, In: *Proc. of 2016 IEEE 35th Int. Perform. Comput. Commun. Conf. IPCCC 2016*, 2017.
- [40] P. Sollich, “Bayesian methods for support vector machines: Evidence and predictive class probabilities”, *Mach. Learn.*, Vol. 46, No. 1–3, pp. 21–52, 2002.