Identification of Diabetic Retinopathy with Retinal Fundus Imagery Using Probabilistic Neural Network

M Elveny1∗, T Anjulina2, B Sirega3r, R Syah4
1,2,3,4 Fakultas Ilmu Komputer dan Teknologi Informasi (Fasilkom-TI), Universitas Sumatera Utara, Padang Bulan 20155 USU, Medan, Indonesia
E-mail: marischaelveny@usu.ac.id

Abstract. Diabetic retinopathy is a microvascular complication of diabetes mellitus that attacks blood vessels in the retina. The main characteristics of diabetic retinopathy are microaneurysm, retinal haemorrhages, exudates, and neovascularization. One of the methods used to diagnose diabetic retinopathy is by examining the retinal fundus image. The examination is still done manually by an ophthalmologist. Manual examination requires a high level or concentration and misidentification may occur because some diabetic retinopathy characteristics are difficult to see directly, so it is needed a method that can facilitate the ophthalmologist in making decisions to identify diabetic retinopathy. The method proposed in this research is Probabilistic Neural Network to identify diabetes retinopathy. Before the identification stage is carried out, the retinal image will go through the pre-processing stage in the form of resize, green channel, contrast stretching and feature extraction using the Gray Level Co-Occurrence Matrix. After testing in this research, it was concluded that the proposed method was able to identify diabetes retinopathy with an accuracy of 86.8%.

1. Introduction
Diabetic retinopathy characterized by the emergence of points in blood vessels (microaneurysms), leakage of blood vessels, appearance of yellowish spots in the form of lipids (exudates), swelling of the retina, growth of new abnormal blood vessels and damaged nerve tissue [1][2]. Apart from the alarming statistics, research shows that at least 90% of new cases, can be reduced if there is proper care, alert and intense eye monitoring. DR can be diagnosed in 5 stages: mild, moderate, severe, proliferative or no disease [3].

Medical examination of patients with diabetic retinopathy is done in one way, namely direct observation by a doctor on the retinal image of a patient taken using a fundus camera. The results of the retinal image will be analysed by a doctor, this examination usually requires a high concentration in analysing the image [4].

Current technological developments are able to overcome the problem of identifying diabetic retinopathy, one of which is through fundal image analysis. Therefore, researchers will develop computer technology that can identify DR. Identification is done by inputting the image, then the image is processed by image processing, and information will be obtained from the processing results [5].

In this study, the authors propose the Probabilistic Neural Network method. Probabilistic Neural Network (PNN) is a neural network method that uses the principle of statistical theory, namely Bayesians Classification to replace the heuristic principle used by the Backpropagation...
algorithm (Specht, 1994). This algorithm has been widely used because of its ability to be able to process data faster than other methods [6]. On the other hand, the Probabilistic Neural Network (PNN) method was chosen because it is a type of neural network that has been proven to have a fairly high degree of accuracy in identifying, namely accuracy of 95%.

Diabetic retinopathy (DR) is a microangiopathy characterized by damage and blockage of fine vessels, including retinal precapillary arterioles, capillaries, and veins [5]. An image is expressed as a digital image if the values of x, y and the intensity value of f are limited and discrete. Digital Image Processing is a way of image processing using computer devices to be easily interpreted by humans or machines [7]. Probabilistic Neural Network (PNN) can be used for classification problems. PNN is a feed-forward neural network with a complex structure [8]. The layer on PNN consists of four layers, namely the input layer, pattern layer, summation layer, and output layer [9].

2. Methods
2.1. General Architecture
The method used to identify diabetic retinopathy there are several stages used, beginning with collecting normal image data and diabetic retinopathy which will be used for training images and test images, the pre-processing stage which consists of resizing, namely changing the image size to 800x800 size, then forming green channel which get a clearer image of blood vessels and retinal structures, improve the quality of grey images by increasing the contrast of the image with the contrast stretching method. Then the next stage is feature extraction from the results of segmentation using the Gray Level Cooccurrence Matrix (GLCM). And the last is the classification stage using Probabilistic Neural Network. After these steps are carried out, the results will be obtained identification of diabetic retinopathy.

![General Architecture](image)

**Figure 1.** General Architecture

2.2. Data Used
Image data used in this study were obtained from the Database for Dedicated to Retinal Ophthalmology (MESSIDOR) Method for Evaluating Segmentation and Indexing techniques. Messidor is a research program funded by the French Ministry of Research and Defense in the
TECHNO-VISION 2004 program. Image data to be used is divided into two datasets, namely for the training dataset and the test dataset to find out how accurate the identification process is. All data amount to 1025 images, where data sharing is shown in Table 1.

| No | Classification         | The amount of data |
|----|------------------------|--------------------|
| 1  | Normal                 | 470                |
| 2  | Diabetic Retinopathy   | 555                |

The collected images are then divided into two datasets, namely: the training dataset or training dataset which will be used as a comparison in the classification and testing of the dataset or testing dataset used to determine the accuracy of the clarification process. Training dataset accounts for 80% of the overall data and the test dataset amounts to 20% of the overall data. The division of the entire image into a training dataset and test dataset is done randomly. Details of the builder of the dataset can be seen in Table 2.

| No | Classification         | Training | Testing | The amount of data |
|----|------------------------|----------|---------|--------------------|
| 1  | Normal                 | 376      | 94      | 470                |
| 2  | Diabetic Retinopathy   | 444      | 111     | 555                |

Examples of images used in this study can be seen in Figure 2.

Figure 2. Normal Retinal Image

Figure 3. Retina Diabetic Retinopathy Image
In Figure 2 we can see a normal retinal image that is in good condition, has no symptoms of diabetic retinopathy. Whereas in Figure 3 the image looks symptomatic with the appearance of dots in blood vessels (microaneurysms), leakage of directional vessels, yellow spots appearing in the form of lipids (exudates) that indicate exposure to diabetic retinopathy.

2.3. Preprocessing
In the preprocessing stage, the retinal image is processed to be extracted by producing good characteristics. The pre-processing stage consists of resizing, forming a green channel image, and improving image quality.

2.4. Feature Extraction
After the retinal image is improved, the next step is feature extraction using the Gray Level Coccurence Matrix (GLCM). The image portion used in the calculation of the cohesive matrix is only the retina portion. To take part of the retina in the image, the concept of masking is used. Masking is a binary image that is used to take certain parts of an image depending on the pixel value that is owned by the binary value.

| No | Feature     | Direction | Feature Value          |
|----|-------------|-----------|------------------------|
| 1  | Contrast    | 0         | 28.336232790989        |
| 2  | Contrast    | 45        | 38.478492358251        |
| 3  | Contrast    | 90        | 28.60546772215         |
| 4  | Contrast    | 135       | 39.828874014921        |
| 5  | Homogeneity | 0         | 0.333285421472         |
| 6  | Homogeneity | 45        | 0.310634847895         |
| 7  | Homogeneity | 90        | 0.330938031003         |
| 8  | Homogeneity | 135       | 0.310635087345         |
| 9  | Entropy     | 0         | 6.558556719363         |
| 10 | Entropy     | 45        | 6.558556719363         |
| 11 | Entropy     | 90        | 6.568633453131         |
| 12 | Entropy     | 135       | 6.657496878819         |
| 13 | Energy      | 0         | 0.013472538352         |
| 14 | Energy      | 45        | 0.012757458086         |
| 15 | Energy      | 90        | 0.013373666261         |
| 16 | Energy      | 135       | 0.012755456307         |
| 17 | Dissimilarity | 0     | 3.527027534418         |
| 18 | Dissimilarity | 45     | 4.010886574426         |
| 19 | Dissimilarity | 90     | 3.572614205256         |
| 20 | Dissimilarity | 135   | 4.027927587833         |

2.5. Classification
The value of the feature extractions from each training data will be used as input in the testing process. Then all these values will be stored in the N matrix which is the result of the transpose of the N X X matrix, where N is the training data and X is the training vector. In this study, the training data for all images was 180, with normal divisions of 80 images and diabetic retinopathy.
of 100 images while the total test data were 45 with 20 divisions of normal images and 25 diabetic retinopathy images.

The training process in the PNN method consists of a unique step, which is to store the weight of each of them on the pattern layer formed by the vector resulting from the extraction of features from each training data. The testing process in the PNN method consists of several steps including:

1. Input layer: Enter the test data with GLCM feature extraction values from the test data.
2. Hidden layer: Test data will be calculated the proximity of the training data stored in the Data.txt file by applying the Gaussian kernel function. This process is carried out with equation 1 below.

\[ W_{ij}(X) = \frac{1}{2\pi^{d/2}\sigma^d} \exp \left[-\frac{\|x - x_{ij}\|^2}{2\sigma^2}\right] \]  

Where:
- Wi j = gaussian kernel
- d = vector dimension x
- σ = spread / smoothing parameter
- x = testing vector
- xi j = vet raining j to j from class i

3. Summation layer: then the sum of the results of the gaussian kernel function with the same class is then averaged by the amount of test data in accordance with each class. The purpose of this process is to find the probability of each class and the results of the calculations made can be seen in table 4.

| No Retina Image | Probability Value |
|-----------------|-------------------|
| 1 Diabetic Retinopathy | 0.999761458151 |
| 2 Normal | 0.000238541849 |

Output layer: The highest probability value will enter into that class. Seen in Table 3. The highest probability value is diabetic retinopathy class with a value of 0.994831018139, therefore the result of identification is diabetic retinopathy.

3. Result and Discussion

The results of this research process are identification of normal retinal fundus images and diabetic retinopathy.

3.1. System Design

At the design stage of this system in figure 4 will be explained about the design of the system menu and the design of the application interface identification of diabetic retinopathy through fundus images. This design aims to enable users to easily run applications.
3.2. Implementation
At this stage describes the implementation and testing of applications based on analysis and application design that aims to display the results of the design of applications that have been built and the application testing process in identifying diabetic retinopathy. Data testing was performed using 94 normal images and 111 diabetic retinopathy images as testing data, and 376 normal images and 444 diabetic retinopathy images as training data.

The classification testing in PNN uses different smoothing parameter ($\sigma$) values to find out what is the $\sigma$ value that has the highest accuracy. The values of $\sigma$ used are 0.1, 0.3, 0.5, 0.7 and 0.9. Testing with different $\sigma$ aims to get the value of $\sigma$ that is able to identify diabetic retinopathy with a high degree of accuracy.

Based on the test results where the x axis on the graph is the value of the smoothing parameter ($\sigma$) while the y axis is the result of the accuracy value, the smaller the value of $\sigma$, the accuracy obtained is also because the value of $\sigma$ greatly influences the value of the probability density function. While the greater the value of $\sigma$, the higher the accuracy obtained. So the best accuracy is obtained from the value $\sigma \geq 0.7$.

From this discussion, it can be seen that the lower the value of $\sigma$, the lower the level of accuracy because the value of $\sigma$ affects the total density of the probability value. So the highest accuracy is obtained with the value $\sigma \geq 0.7$. The author determines the use of $\sigma$ in this study with a value of 0.9, the highest value on the use of $\sigma$ in the hope that this value can get high accuracy. Details of the classification test results with the value $\sigma = 0.9$.

The accuracy level obtained in the classification process is 86.8%. The accuracy value can be
Figure 6. Test Accuracy Graph

obtained from equation 2.

\[
\text{Presentation accuracy} = \frac{\text{amount of correct test data}}{\text{total number of test data images}} \times 100\% = \frac{178}{205} \times 100\% = 86.8\%
\]  

From the above calculation it can be seen that the accuracy of the Probabilistic Neural Network method and Gray Level Co-occurrence in the identification of diabetic retinopathy through retinal fundus images reaches 86.8%.

4. Conclusion

Probabilistic Neural Network method is able to identify diabetic retinopathy through retinal fundus images well which has an accuracy rate of 86.8%. Based on system testing, the value of the smoothing parameter (\(\sigma\)) greatly affects accuracy. Because the smaller the value of \(\sigma\), the probability value is also smaller and the greater the value, the greater the probability value. The value of \(\sigma \geq 0.7\) is the best value of smoothing parameters to identify diabetic retinopathy by using Probabilistic Neural Network.

References

[1] Dillak, RY & Bintiri, MG 2013. Classification of Diabetic Retinopathy Image Using 3D-GLCM Projection. Proceedings of the National Seminar on Information Technology Applications (SNATI), pp. 46-51.
[2] Doshi, D., Shenoy, A., Sidhpura, D. & Gharpure, Dr. P. 2016. Diabetic Retinopathy Detection using Deep Convolutional Neural Networks. International Conference on Computing, Analytics and Security Trend (CAST) College of Engineering Pune, India, pages 261-266.
[3] Khosravianian, A. & Ayat, S. 2016. Diagnosing Breast Cancer Type by Using Probabilistic Neural Networks in Decision Support System. International Journal of Knowledge Engineering 2 (1): 73-76.
[4] Purandare, M. & Noronha, K. 2016. Hybrid System for Automatic Classification of Diabetic Retinopathy using Fundus Images. Online International Conference on Green Engineering and Technologies (IC-GET), pp. 1-5.
[5] Kumaran, Y. & Patil, CM 2016. A Brief Review of the Detection of Diabetic Retinopathy in Human Eyes Using Pre-Processing & Segmentation Techniques. International Journal of Recent Technology and Engineering (IJRTE): 310-320.
[6] Lotfi Abdelhadi and Banyetton Abdelkader. 2014. A reduced probabilistic neural network for the classification of large databases. Turkish Journal of Electrical Engineering & Computer Sciences. 979-989.
[7] Mishra, M., Jena., AR, Das, R. 2013. A Probabilistic Neural Network Approach for Classification of Vehicle. International Journal of Application or Innovation in Engineering & Management (IJAEM) 2 (7): 367 -371.

[8] Palomino, LV, Jr., VS, Net, RMF 2014. Probabilistic Neural Network and Fuzzy Cluster Analysis Methods Applied to Impedance-Based SHM for Damage Classification. Hindwai Publishing Corporation 2014 (2014): 1-12.

[9] Shahana, S, Das, DCN 2016. Probabilistic Neural Network Assisted Cell Tracking and Classification, International Research Journal of Engineering and Technology (IRJET) 3 (8): 1155-1161.