Detection System of Varicose Disease using Probabilistic Neural Network

Mayes M. Hoobi[1], Qaswaa A.[2]

[1]Assistant Lecturer, Computer Science Department, College of Science, Baghdad University, Baghdad, Iraq

[2]Lecturer, Computer Science Department, College of Science, Baghdad University, Baghdad, Iraq

Abstract: ***Background**: The diagnoses system of varicose disease has a good level of performance due to the complexity and uniqueness in patterns of vein of the leg. In addition, the patterns of vein are internal of the body, and its features are hard to duplicate, this reason make this method not easy to fake, and thus make it contains of a good features for varicose disease diagnoses. The proposed system used more than one type of distances with probabilistic neural network (PNN) to produce diagnoses system of varicose disease with high accuracy, in addition, this system based on veins as a factor to recognize varicose infection. **Objective**: of this research is to identify the best available evidence on the diagnosis of varicose veins of the lower limbs. The obtained results: indicate that the design of varicose diagnoses system by applying mulit-types of distances (Euclidean and Manhattan) with probabilistic neural network produced new system with high accuracy and low (FAR & FRR) as soon as possible.The results: of the proposed systemindicate that the varicose disease detection when using Euclidean distance is better than using Manhattan with probabilistic neural network.*

Keywords: Euclidean distance, Manhattan distance, Probabilistic Neural Network, Varicose

1. Introduction

Biometrics is automated methods of distinguish a person based on a physiological or behavioral characteristics. Veins are the soft, thin-walled tubes that return blood from the arms and legs to the heart, and it relies on biological information on the interior of the body [1]. The varicose vein is a little purple vein that suddenly seems to appear on legs. The diagnosis of varicose veins is made primarily by physical examination with the help of a hand-held Doppler [2].Diagnostic procedures for varicose veins rapidly evolved during the last decade. However billing codes used in Belgium are still based on the use of conventional techniques whilst new endovenous procedures are increasingly used [3]. Fordiagnostic procedures a guideline has been elaborated by the ConsiliumRadiologicum and published on the website of the Federal Public service Health, Food security and Environment. This guideline recommends the use of colour duplex Doppler ultrasound in most cases. Magnetic resonance imaging (MRI), tomography (and phlebography) might be used in exceptional cases (e.g. congenital abnormalities) before an intervention. The NIHDI subsequently sent a document to all physicians on therational use of imaging procedures to decrease the risk linked to ionizing radiation [4].The diagnosis varicose veins in legs have evolved greatly in recent years. The Doppler ultrasound has almost entirely replaced the previously procedures (e.g. volumetry, phlebography) for the imaging of varicose veins. Data from the National Institute for Health and Disability Insurance (NIHDI) show disparities in the use of billing codes in diagnostic procedures [5].The aim of this paper is examines the applicability of (Euclidean and Manhattan) distances to diagnose the varicosedisease infection. The structure of this paper as follows:Section2 presents the Probabilistic Neural Network (PNN). Subsequent to these, proposed approach using to detection varicose disease by applying Probabilistic Neural Network (PNN) with two types of distances illustrated in section 3. In section4,explains the evaluation criteria.Section 5 visualizes the results of proposed system are presented with two experiments. Finally, section 6clarifies the conclusions of proposed system.

2. Probabilistic Neural Networks (PNN)

The PNN consists of input layer, pattern layer, summation layer, and output layer. The number of nodes in input layer depends on length of image vector for specific image. The pattern layer is designed to contain one neuron (node) for each training sample available and the neurons are split into the two classes. Each neuron in the pattern layer computes a distance measure between the presented input vector and the training example represented by that pattern neuron. The summation layer contains one neuron for each class, while the output layer contains one neuron as show in figure (1).

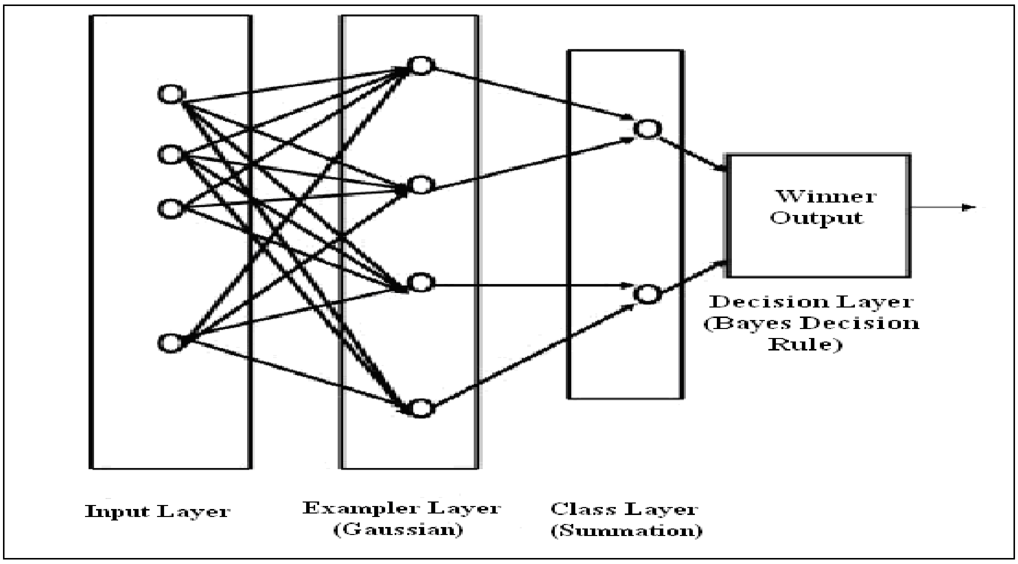

Figure 1: PNN Structure [6]

PNN algorithm is started with read image vector (X) and feed it to each Gaussian function in each class, then for each group of hidden nodes, compute all Gaussian functional values at the hidden nodes as illustrated in equation(1)[7].

$$p_{i(x)} = \exp - \left(\frac{D}{2\delta^2} \right) \quad (1)$$

Where

p_i : represents the output of a pattern node.

x : is the image vector to be assigned into class c_i .

σ : smoothing factor.

D : represents the distance between the image vector x and the sample vector (reference vector) y is computed in this paper using two types of distance measures as illustrated in the following equations [8]-[9]:

1-Euclidean Distance: is one of the most popular distance metric between two vectors x and y and is computed as in equation (2).

$$D(x, y) = \sqrt{\Sigma (x - y)^2} \quad (2)$$

2-Manhattan Distance: the Manhattan (or city block) distance between vector x and vector y is calculated as in equation (3).

$$D(x, y) = \Sigma \left| y - x \right| \quad (3)$$

After these steps, for each group of hidden nodes, feed all its Gaussian functional values to the single output node for that group as illustrated in equation (4).

$$y_i(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} p_i(x) \quad (4)$$

Where

y_i : represents the output of summation node i for $classi$

n_j : represents the number of samples in pattern layer of $classi$

p_i : represents the output of pattern node i

Finally, find maximum value of all summed functional values at the output nodes comparing the values of Y1(x) and Y2(x).

If Y1(x) > Y2(x), then is assigned to class1;otherwise class2.

3. Proposed System –Varicose Disease Diagnoses System (VDDS)

Classification is to obtain the class that is most matches to the classified sample. The proposed system VDDS focuses on vein varicose disease infection. The objective is to determine whether the proposed VDDS could recognize the infected image and which of two types of used distance is better than other. The classification via VDDS is based on the idea that each patient posse's unique vien features. This paper produces a system to obtain diagnose system of varicose disease with high accuracy, this system exploits Euclidean and Manhattan distances with PNN.The aim of this section is to apply PNN for varicose disease detection. This section, also clarifies the features that can be extracted from the patient's image based on their infection that can be employed to maximize differences between the patient cases.

3.1 Varicose Image Preprocessing and Feature Extraction Phase

At the beginning, the images of legs were obtained; the quality of leg images were bad and required several preprocessing techniques must be used to enhance the image quality. The purpose of this stage is to improve the image quality so that vein patterns can be more easily distinguished. On the other hand, the original image is in RGB format it must be converted into gray scale image, which allows faster processing, as compared to colored images. A simple type of contrast stretching is applied to enhance the details of image that increase the range of an image to cover whole brightness range. Then, the mean filter is used to reduce the noise and convert the produced gray image to binarized image (black and white). Figure (2) illustrates examples of two images preprocessing stages.

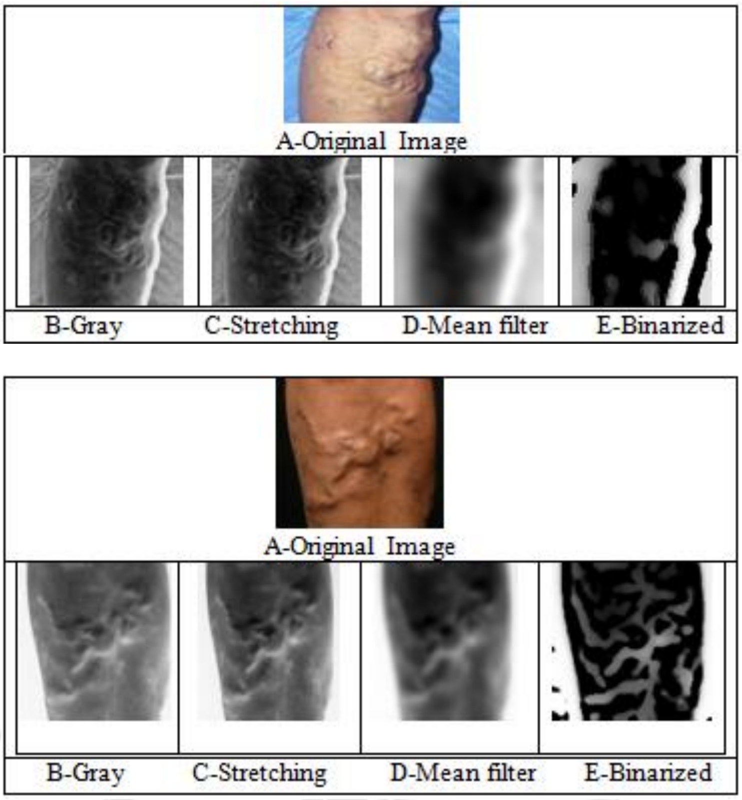

Figure 2: Preprocessing Stages

The next step is feature extraction. There are several features can be used, however, not all kinds of features are useful and widely used [10]. In this paper the following statistical features are used.

1-Mean [11]

$$Mean = \frac{Sum_i}{Np} \quad \textbf{(5)}$$

Where

$$Sum_i = \sum_{i=1}^{Np} Pi \quad \textbf{(6)}$$

Pi : Is ith picture

2-Variance [4].

$$\sigma^2 = \sum (Pix_i - Mean)^2 / Nsamp \quad \textbf{(7)}$$

Pix: is the value of the ith pixel

N: is the number of samples

3.2 Training Phase

This section illustrates how PNN was used in proposed VDDS. The number of input layer nodes represents the length of image vector according to specific feature. The number of pattern layer nodes represents the total number of training samples (patient's image) as show in figure (3).

Paper ID: ART20173435 DOI: 10.21275/ART20173435

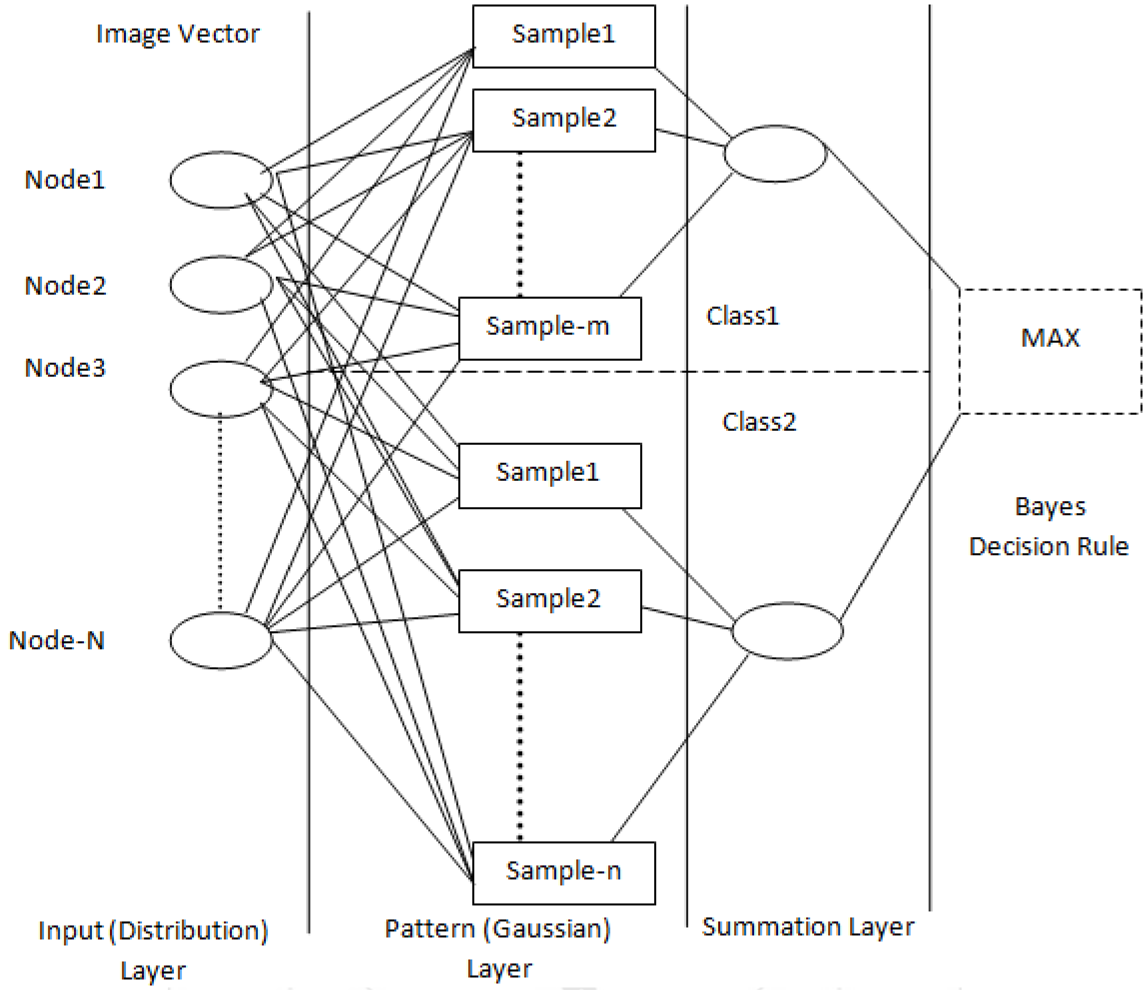

Figure 3: PNN Structure for Image Vector

Figure (4) illustrates the training phase of PNN of the VDDS, In this paper, there are two classes, the first class for infectedpatient's leg and the second one for not infected.Training with 100 samples consists of 60infected samples and 40 not infected. The main aim of training phase of PNN is to find a proper value of smoothing parameter σ. In this paper, an algorithm is proposed to determine the proper range for σ as clarified in algorithm (1). After the proper range for σ is obtained, the training samples in pattern layer nodes are presented as input information vector in input layer. Then change σ value within proper range and apply PNNalgorithm continuously until reaching to least error rates in classification. This value of σ is considered the best value for good classification .

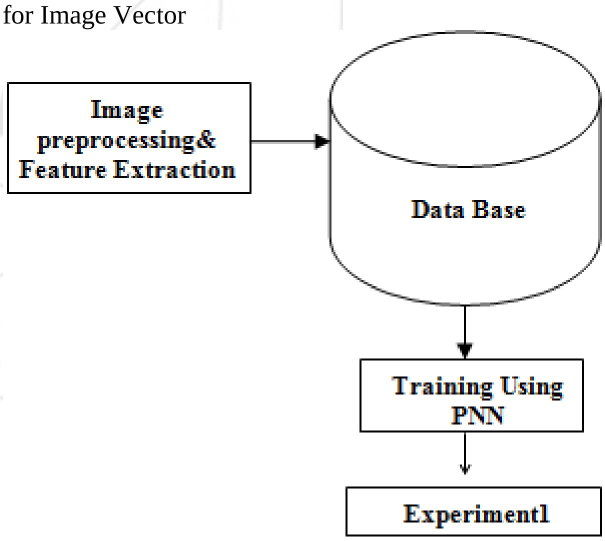

Figure 4: Training Phase of PNN (Experiment1)

Algorithm (1): Smoothing Parameter Range Determination (σ)

Input: Set of training image vectors $p_{i,j}$.

Output: Smoothing parameter (σ).

Step1: [Find Mean Vector]
Compute the mean (centeroid) vector for each class c_k , 1<= k<=2.

$$\mu_k = \frac{1}{N_k} \sum p_{i,j} .$$

Where

N : is the total number of patterns with specific class

$p_{i,j}$: is the feature number i of pattern number j

Step2: [Find Standard Deviation Vector]
Compute the standard deviation vector for each class, k.

$$std_k = \sqrt{\frac{1}{N_k} \sum \left(\mu_i - p_{i,j} \right)^2} .$$

where

μ_i : is the ith value of mean vector

Step3: [Find Range]
Find the minimum and maximum standard deviation value for each class to obtain the proper range of (σ) value.

3.3 Testing Phase (Detection of Varicose Disease Phase)

As show in figure (5),the features are extracted from the captured samples and are compared to the training model. As mentioned previously in PNN training section the optimum value of smoothing parameter σ is obtained in order to use this value in PNN classification phase, the PNN classification begins to check whether the image is infected or not. The image vector is extracted. Then the image vector is presented to the input layer nodes. At pattern layer, the similarity score between a reference template and a test data template is computed. The output score of pattern layer node has the range of [0 1]. Then this output score is put forward into summation layer. Finally, the decision is made at output layer to classify the image either as infected or not infected.

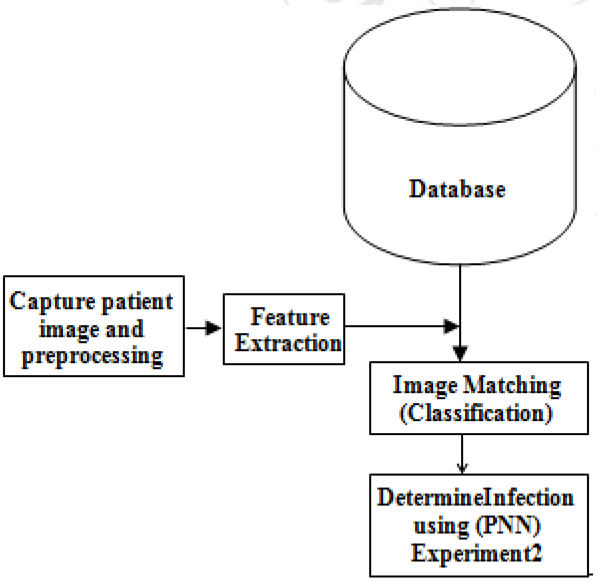

Figure 5: Testing Phase of PNN (Experiment2)

4. Evaluation Criteria

To evaluate the predictive performance of the proposed VDDS three measures are calculated. These measures are called False Reject Rate (FRR), False Accept Rate (FAR) and Accuracy. The formulas for calculating each of these measures are given as in equations (8), (9),and (10) respectively [12].

FRR: is defined as the rate at which images are infected when they could be not infected.

$$FRR = \frac{Number\ of\ infected\ images}{Total\ number\ of\ not\ infected\ images} \quad (8)$$

FAR: is defined as the rate at which images are not infected when they should be infected.

$$FAR = \frac{Number\ of\ not\ infected\ images}{Total\ number\ of\ infected\ images} \quad (9)$$

Accuracy: is defined as the proportion of true results in the population.

$$Accuracy = \frac{Number\ of\ correct\ classify}{Total\ number\ of\ image\ vector} \quad (10)$$

5. Experimental Results

Two experiments are conducted independently. In the first experiment was tested on the same sample in the training. The results of this experiment were obtained with error rates equal to 0% and Accuracy 100% with two types of distances, as shown in tables (1).

Table 1: First Experiment of Training samples

Metric	Euclidean	Manhattan
FAR%	0	0
FRR%	0	0
ACC%	100	100

The second experiment deals with new 60 samples, 35 infected and 25 not infected. This experiment includes computing the selected feature(s) for each image vector. Then the preprocessing is applied. The results of these experiments are shown in table (2).

Table 2: Second Experiment of new Testing Samples

Metric	Euclidean	Manhattan
FAR%	6	10
FRR%	0	6
ACC%	97	94

This section investigates the performance of the proposed VDDS. In testing phase the image of the patient's leg is obtained and preprocessed, then the feature vector of the adopted approach is calculated, and the resulting feature vector is compared with those stored in database and the patient's leg is recognized was infect or not by applying PNN. The PNN isappliedin twostages: training and testing, in training stage the algorithm was trained on 100 samples divided into 60 infected and 40 not infected. While in testing phase two experiments are performed with two types of distances (Euclidean and Manhattan), in the first experiment the PNN was tested on the same samples that it was trained on to ensure the good level of training, While in the second experiment 60 new samples was tested as 35 infected and 25 as not infected and also the produced results indicate to good performance of proposed diagnose system.

Paper ID: ART20173435 DOI: 10.21275/ART20173435

The PNN is used as VDDS, the results with two types of distances show the ability to distinguish infected images from not infected. The second experiment indicates that when use PNN with VDDS with Euclidean distance is better than Manhattan distance. In other wards the best result is obtained with PNN when use Euclidean distance.

6. Conclusions

From the results, one can conclude, in general, the following:

1) The accuracy of the presented work with PNN is close to meet acceptable error levels that would be required for a system with some of classification of patient's leg image.
2) When PNN with two types of distance metric (Euclidean distance, and Manhattan distance) is applied the results show that the Euclidean distance produced the best results.

References

[1] Masaki W., Toshio, E. and Morito S. and Shigeru S.2005.Palm vein authentication technology and its applications, The Biometric Consortium Conference, September 19-21, USA.

[2] A study on varicose vein treatments, Patient Reported Outcome Measures (PROMs) in England,2016

[3] Haxhe J, De Maeseneer M, Schoevaerdts J. Le traitement des varices en Belgique: oùallonsnous? Impact des nouvelles technologies sur les pratiquesmédicales.Phlébologie 2010;63(2):9-14.

[4] ConsiliumRadiologicum. Recommandationsimageriemédicale. Brussels: Service Public Fédéral Santé Publique, SécuritéAlimentaire, Environnement, Last amended October 2010.

[5] NIHDI (National Institute for Health and Disability Insurance). Billing codes for interventions on varicose veins - Period 2007-2009. Unpublished data.Brussels, 2011.

[6] Anagnostopoulos I., "Classifying Web Pages Employing A Probabilistic Neural Network", in IEEE Proceedings Software, Vol. 151, No. 3, pp. 139-150, 2004.

[7] I.Anagnostopoulos, "Classifying Web Pages Employing A Probabilistic Neural Network", in IEEE Proceedings Software, Vol. 151, No. 3, pp. 139-150, 2004.

[8] F.Monrose and A. D. Rubin, “Keystroke Dynamics as a Biometric for Authentication”, Future Gener Compute Syst Vol. 16, No.4, pp. 351– 359, 2000.

[9] R. Kenneth, "User Authentication via Keystroke Dynamics: an Artificial Immune System Based Approach", in Proceedings of 5th International Conference on Information Technology, 2011.

[10] Fazal M. and Baharum B. 2012.The Statistical Quantized Histogram Texture Features Analysis for Image Retrieval Based on Median and Laplacian Filters in the DCT Domain, IAJIT First Online Publication.

[11] Gelman, Andrew. 2008.Variance, analysis of. The new Palgrave dictionary of economics.Second Edition. Basingstoke, Hampshire New York: Palgrave Macmillan.

[12] H. Barghouthi, “Keystroke Dynamics How Typing Characteristics Differ from One Application to Another”, Msc in Information Security, Gjøvik University College, Norway, 2009.