Sakarya University Journal of Science 24(2), 424-431, 2020

Artificial Neural Networks Based Decision Support System for the Detection of Diabetic Retinopathy

Zehra KARAPINAR ŞENTÜRK*[1]

Abstract

Machine learning methods have been frequently used for the diagnosis of several diseases recently because of its reliability and convenience. In this paper, a comprehensive overview of the literature related to diabetes and diabetic retinopathy has been done and diagnosis of diabetic retinopathy disease is investigated. Artificial Neural Networks (ANN) method has been applied to the problem using Rapid Miner, a data mining tool. Some other methods have also adapted to the problem, but ANN based detection approach gave the best results. 88.52% sensitivity has been obtained using the features of Messidor dataset. Besides showing the success of ANN in diabetic retinopathy detection, this study also proved that Rapid Miner can be used effectively for the analysis of diabetic retinopathy.

Keywords: diabetic retinopathy, artificial neural networks, Rapid Miner.

1. INTRODUCTION

Diabetes is a critical and protracted disease which occurs depending on the insufficient releasing of insulin or the inability of human body to use insulin [1]. Main types of diabetes are type 1 and type 2 diabetes. Type 1 diabetes is the primary cause of childhood diabetes but can be seen at any age and it cannot be prevented [1]. 90% of diabetes worldwide has type 2 diabetes and it can be controlled and even get rid of it [1]. To defer farther problems caused by type 2 diabetes may be possible via regularizations in eating and work out habits when an early diagnosis is achieved, but early diagnosis is not always possible and it takes seven years on the average to diagnose this disease [2]. Diabetic retinopathy (DR) is one of the diabetic eye diseases and early diagnosis and timely treatment of DR can prevent vision disorders and blindness [1]. 21% of type 2 diabetes patients also have retinopathy once the disease is diagnosed and it is a progressive disease [3]. More than 60% of type 2 patients have retinopathy and a very often reason of recent occurrences of blindness between the ages 20 and 74 is because of diabetic retinopathy [3]. Therefore, early diagnosis of diabetes and diabetic retinopathy is crucial. Medical diagnosis of diseases is accurate, but may take a long time, costly, and psychologically disturbing for possible patients. So, computer aided decision support mechanisms which perform fast and

*Corresponding Author. zehrakarapinar@duzce.edu.tr
[1]Düzce University, Faculty of Engineering, Computer Engineering Department, 81620, Düzce, Turkey
ORCID: https://orcid.org/0000-0003-3116-1985

almost accurate diagnosis help significantly both to the doctors and the patients.

There are many studies related to the diagnosis of diabetes in the literature. In [4], the researchers applied multi-layer perceptron, radial basis function and general regression neural networks to Pima Indian Diabetes (PID) dataset and showed that general regression neural networks perform the best with 80.21% accuracy for the diagnosis of diabetes. ID3 and Decision Tree algorithms are used for diabetes detection in [5] and their accuracies were found as 80% and 72% respectively. Classification of whether a patient has diabetes or not was achieved 80.72% by AdaBoost algorithm and four base classifiers which are Support Vector Machine, Naïve Bayes, Decision Tree and Decision Stump were considered in [6]. Experimental results of the study indicate that Decision Stump supported AdaBoost classifies better than the other three base classifiers. A new distance metric, Lorentzian Distance is proposed in [7] for classification purposes and they applied this metric with k-nearest neighbor method to classification problem of diabetes. They achieved 76% success with k=10 neighbors. In [8], multi-layer perceptron (MLP) and different versions of Support Vector Machines (SVM) were used. Different kernel functions (linear, polynomial, radial) were used for diagnosis and SVM with linear kernel is decided as the most successful classifier with 77.47% accuracy.

Another machine learning based diabetes diagnosis was realized by Naïve Bayes in [9]. The authors also used Genetic Algorithm to select affective four attributes among eight and applied NB together with these selected four attributes. The performance of NB was improved to 78.69%. As a prevalent and easy to apply method, an artificial neural network was used for the classification of diabetes in [10]. Four-layer neural network with eight input neurons, 10 hidden neurons for each of two hidden layers and one neuron for output layer is trained and 92% accuracy was obtained. Different from the others, a mobile application was realized in [11] for diabetes diagnosis. Type of diabetes and the possibility of developing diabetes in the future were determined by the developed system using blood glucose level data. They developed a decision tree like algorithm based on the guidance of the health professional not a machine learning method. Random forests algorithm was used in [12] for diabetes diagnosis and they achieved 89.63% accuracy. Decision Tree, SVM, and Naïve Bayes were used for the same purpose in [13] and they showed that Naïve Bayes has the best diagnostic performance when compared to others with 76.3% accuracy. In [14], an auto encoder neural network with 24 hidden layers was proposed for classification. Average accuracy of the study after twenty trials was found as 97.3%, which is the highest classification performance related to the classification of diabetes. As seen from the above mentioned studies diabetes are diagnosed using machine learning methods satisfactorily accurate with 97.3% at most.

Although there are many studies on machine learning based diagnosis of diabetes, there is a few study related to the classification of diabetic retinopathy (DR) which affects many diabetics as mentioned above. A computer aided diagnostic system was proposed in [15]. Using a graph kernel, they proposed a multi-kernel multi-instance learning method. They first detected hemorrhages and micro aneurisms (HMA) from fundus images and extracted some features related to color, shape, texture and gradient. An IOT based diagnosis of diabetic retinopathy (DR) was studied in [16] and they determined the blood glucose level of diabetics via Dexcom G4Plantinum sensors. Readings of the sensor were send to DRapp device and the patient's diabetes type was determined as type 1 or type 2. If it was determined as type 1, then the patient was directed for some treatments and a doctor appointment is set. If the patient had type 2 diabetes, then the patient's fundus image was caught by fundus camera and retinal blood vessels were detected by the proposed image segmentation method. Accuracy of this method is 99.58% and its sensitivity is 72.51%. Face recognition methods were used for fundus images in [17] to detect exudates and their features were extracted very recently. Four classifiers were implemented by the extracted features. These classifiers are k-NN, ANN, Random Forests, and

SVM. The best classification was achieved by SVM with 80.4% accuracy.

Deep learning (DL) has been used in the solution of many medical problems. Some of them are as follows. As explained in [18], DL models have been used in many medical imaging applications. In [19], the researchers developed a tool for the ophthalmologists to help them for grading DR via DL. Retinal fundus images were graded using this tool with the support of image magnification and contrast adjustment. They obtained very high performance metrics, but their approach is not a fully machine learning based decision system. An interpretable classifier was proposed in [20] for DR via DL. The proposed approach classifies retina images indicating the severity level. They developed a method and generated visual maps to be easily interpreted by the ophthalmologists to help the diagnosis of DR.

In this study, different from the previously mentioned papers, the features of retinal fundus images of Messidor dataset [21] and artificial neural networks (ANN) model of Rapid Miner are used for the classification of possible diabetic patients whether having retinopathy since ANN provides solution to not only linear classification problems but also non-linear, multi-class problems with low computational cost. Our method provides a cheap, fast and accurate decision support system for DR classification. Contributions of the paper can be summarized as the followings:

- Features of retinal fundus images of Messidor dataset is used for diagnosis.
- Detailed tables are provided for machine learning based diagnosis of diabetes and diabetic retinopathy.
- Rapid Miner Studio is used for the first time as a tool for classification of DR so far.
- Independent Component Analysis is used to reduce the dimension of the problem.
- Sensitivity is respected as a performance metric and high sensitivity is achieved in diagnosis.

Rest of the paper is organized as follows: Section 2 gives the details of the method and the application, Section 3 provides the experimental results of the application and Section 4 concludes the paper and mentions about the future works.

2. METHODOLOGY

Early detection of DR is quite important since it is one of the main causes of blindness among the people between the ages 20 and 74 [3]. In this study, we intend to detect this serious disease by a popular and an easy approach. The details are given in the following subsections.

2.1. Dataset Description

In this study we used the dataset which includes the extracted features of Messidor dataset provided by Dr. Antal and Dr. Hajdu via UCI Machine Learning Repository [22]. For details of the features in this dataset, the explanations given in [21] can be examined. They are also summarized in Table 3. Number of samples are 1151 and 540 of them belong to the healthy people and are labeled as zero. 611 of the features are of DR patients and they are labeled as 1. The dataset contains twenty columns which represent the lesions, describe an anatomical part or is an image-level descriptor. The last column of the dataset defines the diagnostic class whether 1 in case the patient has DR or 0 if he is healthy.

2.2. Performance Assessment

Accuracy of detection of true positives, i.e. number of samples with DR and diagnosed as having DR, must especially be satisfactory for a proper decision support system. Misdetection of samples with DR by an automatic machine learning based system may mislead their users and may cause to lose the chance of coming around by early treatment. This affects the rest of the life of people with DR. That is, misclassification of healthy people and advising them to see doctor when they are not DR patient in real are not much important than misclassification of DR patients and telling them that they are healthy. In general, the performance

of a classification method is measured as given in equation (1)

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

where TP = True positive, FP = False positive, TN = True negative, FN = False negative. But, in health-based vital classification problems using true positive ratio, given in equation (2), is more convenient than other performance metrics**.**

$$\text{TPR} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

True positive ratio (TPR) is known as sensitivity or recall in the literature. Therefore, TPR will be used in this study as most of other health based classification studies instead of considering accuracy as a performance metric since our method is also a health-based application. In the experiments we will try to minimize FN and try to maximize sensitivity of the current classification problem even if it leads to a lower accuracy.

2.3. Artificial Neural Networks (ANN) Based Diabetic Retinopathy Detection

ANN is a part of machine learning (ML) methods and it is frequently used in wide range of applications nowadays such as computer vision, bioinformatics, medical image analysis, computer networks, etc. It has multiple layers: input layer, hidden layer(s) and output layer. It learns in a supervised fashion. ML based disease diagnosis is becoming more prevalent day by day since these methods provide faster and cheaper solutions to real life problems. There are many papers related to ML based detection of diseases in the literature. Machine learning based detection of diabetes as a subset of ML based diagnosis is summarized in Table I. Although there are many studies related to the diagnosis of diabetes, there are few papers related to the detection of diabetic retinopathy which seriously affects many type 2 diabetics. The details of the papers related to the diagnosis of diabetic retinopathy are given in Table II.

Table 1. Studies related to ML based detection of diabetes

Ref. No.	Method	Compared Methods	Accuracy
[4]	General regression neural networks	Multi-layer perceptron, radial basis function and general regression neural networks	80.21%
[5]	ID3	ID3 and Decision Tree	80%
[6]	AdaBoost with Decision Stump	Support Vector Machine, Naïve Bayes, Decision Tree and Decision Stump	80.72%
[7]	k-nearest neighbor with Lorentzian Distance	-	76%
[8]	SVM with linear kernel	MLP, SVM with polynomial and radial kernels	77.47%
[9]	Naïve Bayes and GA	Naïve Bayes	78.69%
[10]	Artificial Neural Networks	-	92%
[12]	Random Forests	Binary Tree, SVM, Adaptive Boosting Model, Generalized Linear Models, Neural Network Model	89.63%
[14]	Auto Encoder Neural Networks	-	97.3%

Table 2. Studies related to the detection of ML based diabetic retinopathy

Ref. No.	Method	Compared Methods	Accuracy
[15]	Multi-kernel multi-instance learning	-	91.6%
[16]	Fuzzy c-means clustering	-	99.58%
[17]	SVM	k-NN, ANN, Random Forests	80.4%
[19]	Deep learning	-	90% (TPR)
[20]	Deep learning	-	91.1% (TPR)

As seen from the tables, DR is not studied properly by ML approaches. ANN method as a compatible ML approach can be adapted to the problem of the detection of DR and may help ophthalmologists and patients. There are very few studies related to the diagnosis of DR via ANN and their results are not satisfactory enough. In this study, ANN will be adapted to the classification of DR patients using the features of an open dataset. In addition, this application will be realized using a very popular data mining tool to make an easy analysis. To further improve the performance of ANN based classification, Independent Component Analysis (ICA) has also been applied for dimensionality reduction. Details of the tool based classification are given in below section.

3. EXPERIMENTAL RESULTS

In this study, an ANN model has been created using Rapid Miner Studio, a free software for data analysis [23] . This tool has not been used yet for the detection of diabetic retinopathy so far. It provides easy and various solutions for the problem. Dimensionality reduction has also been performed to improve the performance of ANN. More number of features mean more number of input nodes in ANN model and this may complicate the learning process. Therefore, feature reduction is used. ICA method has been used for the purpose. ICA is a frequently used statistical technique and it estimates independent components by maximizing the non-Gaussianity of them, or maximizing the likelihood, or minimizing mutual information between them [24]. Weights of the attributes after ICA was applied are given in Table 3. There are 19 attributes in Messidor dataset and 14 most independent attributes have been used in the input set. Bold-written attributes in Table 3 are the top 14 attributes. These inputs were then given to ANN. 1151 samples in the dataset were divided into two as 0.8 and 0.2 splitting ratio for training and testing respectively. 920 samples were used in training the model and 230 unseen and unlabeled sample were used for testing.

Table 3. Weights of the attributes determined by ICA

Attribute No.	Weight of Attribute	Information about Attribute
1	0.003	Result of quality assessment
2	0.090	Result of pre-screening
3	10.299	Number of MAs (Micro Aneurysm) found at the confidence levels α=0.5,…,1 respectively
4	9.891	
5	9.570	
6	9.133	
7	8.166	
8	5.456	
9	-10.996	Number of exudates at the confidence levels α=0.5,…,1 respectively.
10	-2.143	
11	-1.522	
12	-0.658	
13	-0.589	
14	-0.155	
15	-0.016	
16	0.002	
17	0.001	Euclidean distance of the center of the macula and the center of the optic disc
18	0.001	Diameter of the optic disc
19	-0.384	Result of the AM/FM-based classification

Different topologies have been tried to determine the best topology of the ANN model for the problem. The highest- performance ANN model has one hidden

layer with 8 neurons, 0.01 learning rate, and 0.9 momentum coefficient. Proposed ANN model is given in Figure 1. According to the confusion matrix given in Table 4, TPR (or sensitivity) and accuracy of the established model can be calculated as given in equation (3) and (4).

$$\text{TPR} = \frac{108}{108+14} = 0.8852 \quad (3)$$

$$\text{Accuracy} = \frac{108+67}{108+67+14+41} = 0.7609 \quad (4)$$

Table 4. Confusion matrix of Diabetic Retinopathy detection model

		Actual	
		Has DR	Does not have DR
Predicted	Has DR	108	67
	Does not have DR	14	41

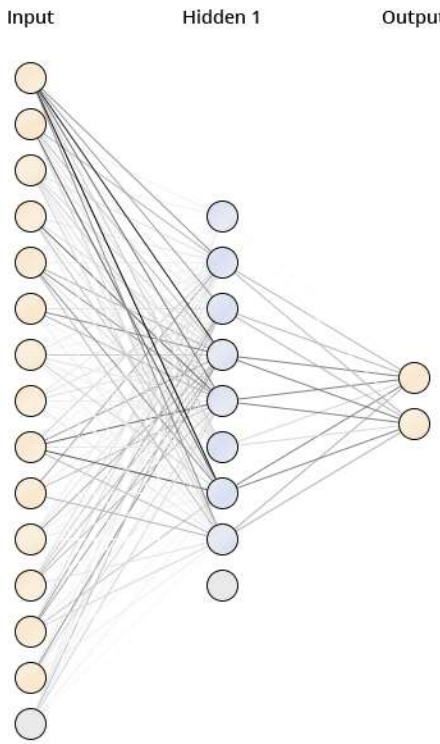

Figure 1. Proposed ANN topology

Although the accuracy of the method proposed in [16] is 99.58%, its sensitivity is low, i.e. 72.51%. This means this method classifies 72 of DR patients among 100 of them. About 28% of DR patients are predicted as normal, not having DR. That is, the method misdiagnoses DR patients considerably. Sensitivity becomes more of an issue in such health related problems. Therefore, one must regard sensitivity as a performance metric. [19] and [20] have good sensitivities but their methods are computationally harder. 88.52% sensitivity by an easy application is well enough to assist doctors for classifying their patients.

4. CONCLUSIONS AND FUTURE WORK

In this paper, ANN based DR detection method is proposed. The features of Messidor dataset have been used for the classification and quite sensitive results were obtained. Although there are many detection methods for diabetes, there is a few paper for diabetic retinopathy detection. ANN method have been frequently and easily used in many different applications and provided successful results. Applying this approach to DR classification simplifies the ML based diagnosis. Making analysis in an easier and faster way is very important for a research topic in health based problems. Rapid Miner is also introduced in this paper for the diagnosis of DR. It is proved by this research that Rapid Miner and ANN can efficiently be used in the analysis of DR data.

Early detection of DR is quite important for diabetic patients and high-sensitive and low-complexity methods must be developed for them. ML based methods proved their superior performance in many medical classification issues and DR classification can be one of these. The proposed method can be improved in the future studies. Parameter selection process can be optimized. DL can be adapted to the problem.

5. REFERENCES

[1] International Diabetes Federation, *IDF Diabetes Atlas 9th edition*. 2019.

[2] B. E. R. Gaillard and D. K. Karumanchi, "Trends in early diabetes diagnosis," *Ophthalmol. Manag.*, vol. 18, no. November 2015, pp. 28–30, 2015.

[3] M. Dansinger, "Types of Diabetes Mellitus," *WebMD Medical Reference*, 2017. [Online]. Available: https://www.webmd.com/diabetes/guide/types-of-diabetes-mellitus#3. [Accessed: 15-Jan-2019].

[4] K. Kayaer and T. Yildirim, "Medical diagnosis on Pima Indian diabetes using general regression neural networks," in *Proceedings of the international conference on artificial neural networks and neural information processing (ICANN/ICONIP)*, 2003, pp. 181–184.

[5] J. Han, J. C. Rodriguze, and M. Beheshti, "Diabetes data analysis and prediction model discovery using RapidMiner," in *2nd International Conference on Future Generation Communication and Networking*, 2008, vol. 3, pp. 96–99.

[6] V. V. Vijayan and C. Anjali, "Prediction and diagnosis of diabetes mellitus - A machine learning approach," in *IEEE Recent Advances in Intelligent Computational Systems, RAICS 2015*, 2015, no. December, pp. 122–127.

[7] H. S. Bilge and Y. Kerimbekov, "Classification with Lorentzian distance metric," in *23rd Signal Processing and Communications Applications Conference*, 2015, pp. 1–4.

[8] M. S. Kurt and T. Ensari, "Diabet diagnosis with support vector machines and multi layer perceptron," in *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, 2017, pp. 1–4.

[9] D. Choubey, S. Paul, S. Kumar, and S. Kumar, "Classification of Pima indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection," *Commun. Comput. Syst.*, pp. 451–455, 2017.

[10] O. Y. Okuboyejo, S. Misra, R. Maskeliunas, and R. Damasevicius, "A Neural Network Based Expert System for the Diagnosis of Diabetes Mellitus," in *International Conference on Information Technology Science*, 2018, vol. 724, no. February.

[11] I. Arikpo *et al.*, "Development of a Mobile Software Tool for Diabetes Diagnosis," vol. 5, no. 3, pp. 1–8, 2018.

[12] P. Samant and R. Agarwal, "Machine learning techniques for medical diagnosis of diabetes using iris images," *Comput. Methods Programs Biomed.*, vol. 157, pp. 121–128, 2018.

[13] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1578–1585, 2018.

[14] Ö. Deperlıoglu and U. Kose, "Diagnosis of Diabetes by Using Deep Neural Network," in *2nd International Symposium on Multidisciplinary Studies and Innovative Technologies*, 2018.

[15] P. Cao, F. Ren, C. Wan, J. Yang, and O. Zaiane, "Efficient multi-kernel multi-instance learning using weakly supervised and imbalanced data for diabetic retinopathy diagnosis," *Comput. Med. Imaging Graph.*, vol. 69, pp. 112–124, 2018.

[16] T. Jemima Jebaseeli, C. Anand Deva Durai, and J. Dinesh Peter, "IOT based sustainable diabetic retinopathy diagnosis system," *Sustain. Comput. Informatics Syst.*, 2018.

[17] L. B. Frazao, N. Theera-Umpon, and S. Auephanwiriyakul, "Diagnosis of diabetic retinopathy based on holistic texture and local retinal features," *Inf. Sci. (Ny).*, vol. 475, pp. 44–66, 2019.

[18] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Z. Med. Phys.*, vol. 29, no. 2, pp. 102–127, May 2019.

[19] V. Gulshan *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, p. 2402, Dec. 2016.

[20] J. de la Torre, A. Valls, and D. Puig, "A deep learning interpretable classifier for diabetic retinopathy disease grading," *Neurocomputing*, 2019.

[21] B. Antal and A. Hajdu, "An ensemble-based system for automatic screening of diabetic retinopathy," *Knowledge-Based Syst.*, vol. 60, pp. 20–27, Apr. 2014.

[22] "UCI Machine Learning Repository: Diabetic Retinopathy Debrecen Data Set Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set. [Accessed: 25-Jan-2020].

[23] "RapidMiner©." [Online]. Available: https://rapidminer.com/. [Accessed: 04-Mar-2019].

[24] A. Tharwat, "Independent component analysis: An introduction," *Appl. Comput. Informatics*, Aug. 2018.