Estimation of Re-hospitalization Risk of Diabetic Patients based on Radial Base Function (RBF) Neural Network Method Combined with Colonial Competition Optimization Algorithm

Mansoureh Khojandi Jazi [1], Narges Habibi [2*], Majid Harouni [3]
1- Department of Electrical Engineering, Dolatabad Branch, Islamic Azad University, Isfahan, Iran.
Email: mansourehkhojandi89@yahoo.com
2- Department of Electrical Engineering, Isfahan Branch, Islamic Azad University, Isfahan, Iran.
Email: narges.habibi@gmail.com (Corresponding Author)
3- Department of Electrical Engineering, Dolatabad Branch, Islamic Azad University, Isfahan, Iran.
Email: harouni.majid@gmail.com

Received: November 2017 Revised: February 2018 Accepted: February 2018

ABSTRACT:
Diabetes is the costliest gland disease in the world. Given the high rates of diabetic people, the necessity of reducing the costs of early re-hospitalization and increasing re-admissions within 30 days after discharge have drawn the attention of researchers and other health sector authorities to find ways to reduce potential and preventable hospital re-admissions. The objective of this paper is to estimate the risk of re-hospitalization of diabetic patients. In order to achieve this goal, the data were first pre-processed, and then, radial base function neural network combined with colonial competition optimization algorithm was used to estimate the risk of re-hospitalization of diabetic patients. Moreover, this risk was estimated using back propagation neural network algorithm and the radial base function neural network algorithm. The accuracy of the proposed method is 99.91. This method shows higher performance compared to radial base function neural network method and back propagation neural network without feature selection.

KEYWORDS: Risk of re-hospitalization of diabetic patients, Radial base function neural network, Colonial competition optimization algorithm, Back propagation neural network.

1. INTRODUCTION

Diabetes is a disorder in the body metabolism, occurs when insulin hormone (regulator of blood glucose) is not produced in the body, or its production level is less than that required by the body, or the body's cells lose their sensitivity to insulin. Insulin plays the role of a bridge between the vein and cell in the body. As a result, insulin is needed for the entry of body glucose to body cells and its consumption as energy.

This hormone is secreted from the pancreatic beta cells. In the absence of insulin or lack of the effect of insulin, glucose cannot enter the cells, resulting in increased blood sugar. The high sugar level is accumulated in the veins and causes damage to various organs of the body, such as eyes, kidney, nerves, and so on [1]. Unfortunately, no definitive treatment has been found so far for this disease. However, by observing the basic principles, we can control diabetes and prevent complications of the disease [1]. Based on global statistics, the number of people with diabetes is about 346 million people. Every 10 seconds, 2 people become diabetic. Every 8 seconds, one person loses his or her life due to diabetes. Diabetes is the cause of one million amputations per year. Ninety percent of people with diabetes are suffering from type 2 diabetes, and unfortunately, 50% of diabetic people are unaware of their diseases. Based on World Health Organization (WHO), the number of diabetic people will increase by 50% by 2030, which will impose much financial pressure on health organizations and the need for equipment and services for hospitalization of the patients [2].

Various studies have been conducted on the diabetes, such as the diagnosis and categorization of types of diabetes and investigation of psychological and social factors affecting the diabetic patients and the prediction of the status of diabetic patients.

Given the high rates of diabetic patients, the necessity of reducing the costs caused by early re-hospitalization and increasing the rate of re-admission within 30 days after discharge have drawn the attention of researchers and other health sector authorities to find the ways to reduce potential and preventable hospital re-admissions [3].

This research estimates the risk of re-hospitalization of diabetic patients using the combined method of radial base neural network and colonial competition optimization algorithm.

2. RESEARCH LITERATURE

Various studies have been conducted on diabetes (Table 1).

A: Diagnosis of diabetes disease:

Zhang et al (2017) developed a machine-based learning framework to identify type 2 diabetes through health electronic health records. In this research, the methods of random forest, nearest neighbor and Bayesian algorithm were used to diagnose type 2 diabetes, which an accuracy of 0.98 was achieved at best condition [4], [5].

This research was conducted using dataset of 123241 people collected in 2012-2014, which 300 samples were selected finally and research was conducted on them. The shortcoming of this research is that it has investigated only type 2 diabetes with limited sample (n=300) [4]. Palacios et al (2017) developed a fuzzy expert system to evaluate nephropathy control in patients with type 2 diabetes. Using the fuzzy method, they achieved an accuracy of 93.33% [6]. Hsiao et al (2016) developed a web-based model to predict liver cancer in patients with type 2 diabetes using artificial neural network and logistic regression method. In this research, they achieved an accuracy of 0.75 [7]. In 2016, Erkaymaz et al examined the effect of small neural networks in diagnosing diabetes. In this research, back propagation neural networks were used to diagnose diabetes with a sensitivity of 0.85 and an accuracy of 0.96 [8].

Ariana et al., (2016) examined the medical record of US diabetic patients. Logistic regression method was used to diagnose diabetes, and their research was based on a web database [9].

Herbert et al., (2016) examined the effect of hemoglobin glucose in diagnosis of diabetic patients. The decision tree was used to diagnose diabetes and the effect of hemoglobin glucose. They achieved an accuracy of 89.58% [10]. Mohammad et al., (2016) examined the effects of hypertension and obesity on the diagnosis of diabetes. They used k-means clustering method to diagnose diabetes and the effect of hypertension and obesity on diabetic patients, which data were finally divided into 9 clusters [11].

Yoichi Hayashi et al examined the extraction law using the recursive extraction algorithm with j48graft along with sampling techniques to diagnose type 2 diabetes. In this investigation, data set of Indian Hospital available on the Pima site was used. They achieved finally an accuracy of 83.83% [12]. In 2014, Kamadi et al examined the effect of artificial intelligence algorithms on diabetes of diagnosis. In this study, decision tree algorithms, and Gaussian and fuzzy logic and data set of the Indian hospitals available on the Pima site were used, which they achieved an accuracy of 75.8% [13]. In 2012, Karan et al used a mobile phone to diagnose diabetes. In this research, the multi-layer perceptron neural network and the web-based client-server were used [14].

B: Diabetes disease prediction:

Enomoto et al., (2017) used a logistic regression method to estimate the risk of re-hospitalization of type 2 diabetes patients over 30-day period. In this research, they concluded that diabetic patients with heart disease, infectious disease, and chest pain are more susceptible to re-hospitalization [15].

Najafian et al., (2016) predicted re-hospitalization status in 30-day period for diabetic patients underwent lower limb surgery. In this research, the effect of factors such as heart disease and the absence of heart disease were examined and the effect of heart disease on the re-hospitalization of diabetic patients was examined [16]. The shortcoming of this research was that samples were limited to those patients undergone a lower limb surgery.

In 2011, Hamil et al., evaluated the effect of heart failure, serum creatinine level, serum sodium level, and hemoglobin and systolic blood pressure on the re-hospitalization of diabetic patients. The accuracy of this method was obtained to be 0.59 [17].

Allaudeen et al., (2011) examined the effect of age, gender, and coronary artery disease on the re-hospitalization of diabetic patients. The accuracy of this method was obtained to be 0.70. [18].

Walraven et al., (2010) presented a model called LACE to predict re-hospitalization of patients, which has been considered in world due to its easiness to rank the risk factors [19].

Table 1. Previous works.

year	Reference	function type	Method	Results
2017	[4]	Diagnosis of type 2 diabetes	Random forest, Nearest neighbor and Bayesian algorithm	Accuracy of 0.98
2017	[6]	Diagnosis of type 2 diabetes	Artificial Neural Network Fuzzy Method and Logistic Regression Method	Accuracy of 93.3%
2016	[7]	Diagnosis of diabetes	Artificial Neural Network	Accuracy of 0.75

			Fuzzy Method and Logistic Regression Method	
2016	[8]	Diagnosis of diabetes	Back propagation neural networks	Sensitivity 0.85 and accuracy 0.96
2016	[9]	Diagnosis of diabetes	Logistic regression	
2016	[10]	Diagnosis of diabetes	Decision tree method	89.58
2016	[11]	Diagnosis of diabetes	k-means clustering	9 clusters
2016	[12]	Diagnosis of diabetes	Weighted fuzzy laws	Men accuracy of 83.83
2014	[13]	Diagnosis of diabetes	Multi-layer perceptron neural network	
2017	[15]	Prediction of diabetes	logistic regression	Diabetic patients with heart disease, infectious disease, and chest pain are more susceptible to re-hospitalization
2016	[16]	Prediction of diabetes	Regression	The effect of heart failure on the re-hospitalization of diabetic patients was determined
2011	[17]	Prediction of diabetes	Regression	Accuracy of this method 0.59
2011	[18]	Prediction of diabetes	Regression	Accuracy 0.67
2010	[19]	Prediction of diabetes	Regression	83.5

3. RESEARCH METHODOLOGY

In this paper, data set related to re-hospitalization of diabetic patients, which is available on the site uci.edu[1], was used. The general architecture used in this research is shown in Fig. 1, This framework includes three main sections as follows:

• Preparation and selection of appropriate features
• Modeling with radial base function neural network and colonial competition optimization algorithm
•Evaluation of results

[1] https://uci.edu/

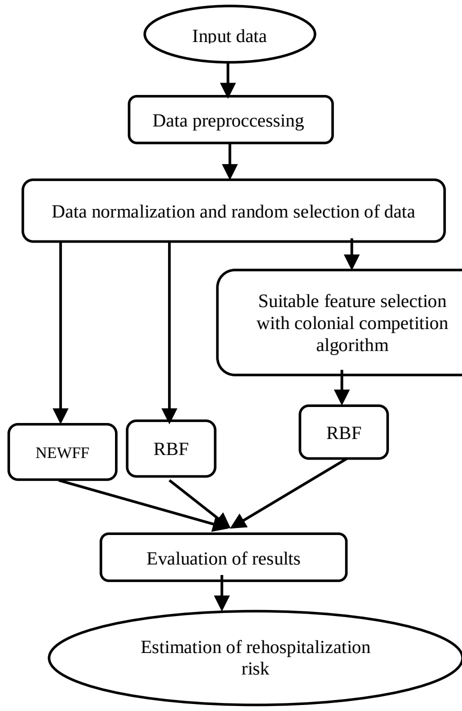

Fig. 1. The proposed method framework.

1.3. Input Data

The used data include 49 input features and a feature, which includes the outputs related to estimation of risk of re-hospitalization of diabetic patients and it includes 101766 records.

2.3. Data Preprocessing

Before the modeling stage, the following preprocessing was applied to the considered data:

• Exclusion of samples with unavailable values
• Exclusion of three features with the same values (metformin-pioglitazone, metformin-rosiglitazone, encounter_id)
•Normalization
•Data hashing

3.3. Modeling to estimate the risk of re-hospitalization in diabetic patients

In the proposed method, the feature selection was first performed on the data set, using the colonial competition algorithm. Then, using the radial base function neural network, the risk of re-hospitalization of diabetic patients was estimated. In order to show the effect of the feature selection, risk of re-hospitalization was re-estimated without applying colonial competition optimization algorithm. Then, risk of re-hospitalization of the diabetic patients was re-estimated using back propagation neural networks.

The colonial competition optimization algorithm is an evolutionary computational technique, which uses a population including potential solutions to the problem to explore in the search space. At the beginning of the colonial competition optimization algorithm in the proposed method, the initial population or solutions are generated randomly and initialized. Each population is considered as a point in the 46-dimensional space (due to having 46 input features), and by ordering this population, colonial and colonized countries are determined. Then, during consecutive repetitions and by updating the target function, it is tried to find the optimal solution leading to the selection of suitable features.

To evaluate and compare the models, the dataset division method, dividing the data to training and testing data, is used, in which 70% of the data were used for training and 30% were used for testing. The Matlab 2015a software was used to implement the proposed method.

4.3. Criteria to assess the model efficiency

To assess the model efficiency, the following criteria were used:

Mean Square Error (MSE): it is the difference between the value predicted by model or the statistical estimator and the actual value. MSE is an appropriate tool to compare the prediction errors with a dataset.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_{1,i} + Y_{2,i})^2$$

Coefficient of determination (R2): it shows the probability of correlation between the two classes of data in the future. In fact, it states approximate results of the considered parameter in the future based on the defined mathematical model, which is consistent with the available data.

$$R^2 = \left(\frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}\right)^2$$

4. RESULTS

In this section, the results of the experiments carried out to evaluate the proposed model are presented.

1.4. Radial base function neural network results without colonial competition algorithm

Modeling was first performed using the radial base function neural network without the colonial competition algorithm (Table 2). Then, the effect of using colonial competition optimization was examined (Table 3).

Radial base function neural network with 5 neurons yielded the best results. The minimum square error, with 10 different neurons is presented in Fig. 2, for training class of data and the test results are presented in Fig. 3.

Table 2. Complete results of RBF neural network.

Neuron	Mse train	R^2 train	Mse test	R^2 test
1	0.277108561087176	0.824719220761191	0.264565682328085	0.813474600350093
2	0.271397173194573	0.841691965483247	0.285937869134510	0.839408041093152
3	0.220348203627690	0.863997624109584	0.210483321835104	0.872346286620944
4	0.211289441789106	0.872783032438923	0.198169073508020	0.879432716299705
5	0.157687169220750	0.878376819228503	0.171411004876354	0.873386272042508
6	0.235854476686795	0.872386803959612	0.244655927097910	0.870372180309966
7	0.256018846400064	0.890471411273216	0.275910925846543	0.884118429945795
8	0.252724388587406	0.874501257363682	0.246551062517268	0.880930991599650
9	0.196385605988072	0.887006028563064	0.181747418711130	0.892397837909809
10	0.250347410817985	0.878928125134564	0.241451425205257	0.874435739514100

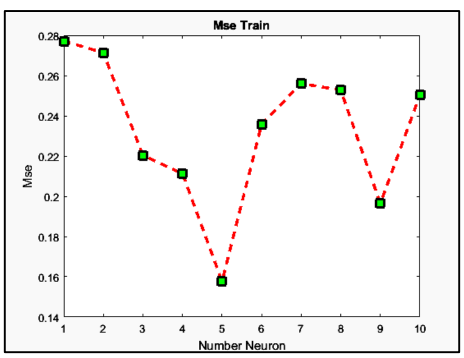

Fig. 2. MSE of training data in the RBF neural network, in the range between 1 and 10 neurons.

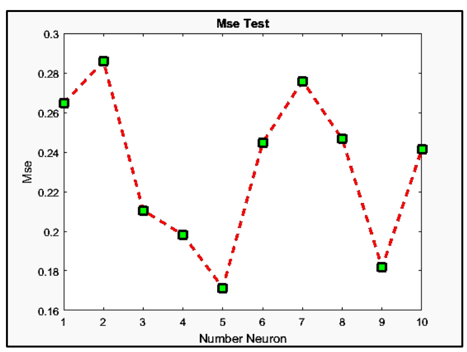

Fig. 3. MSE of testing data in RBF neural network in the range between 1 and 10 neurons.

Coefficient of determination R2 for training data in the RBF neural network in the range between 1 and 10 neurons is shown in Fig. 4, and for testing data, it is shown in Fig. 5.

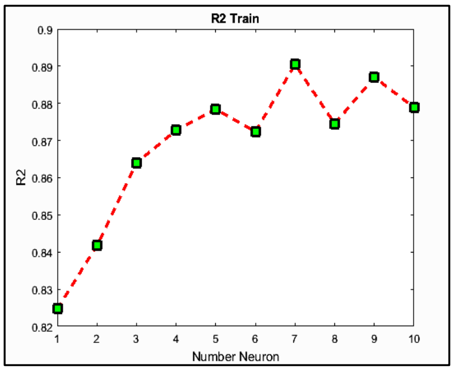

Fig. 4. R2 of training data in the RBF neural network in the range between 1 and 10 neurons.

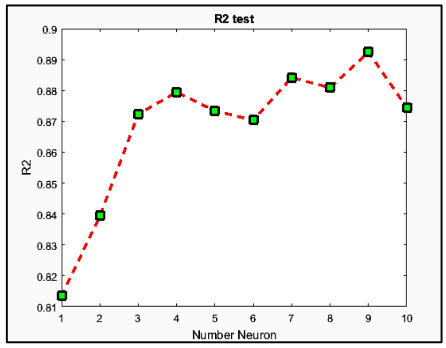

Fig. 5. R2 of testing data in the RBF neural network in the range between 1 and 10 neurons.

2.4. RBF neural network results and colonial competition optimization algorithm

The results of the RBF neural network and the colonial competition optimization algorithm are shown in Table 3, which this method with 9 neurons yielded the best result.

Table 3. Complete results of RBF neural network and colonial competition optimization algorithm.

Neuron	Mse train	R^2 train		Mse test	R^2 test
1	0.122178924641552	0.873867476636094		0.124813108821509	0.859373990969525
2	0.123822206092594	0.885015359529479		0.125301592997196	0.882560158901436
3	0.124064586018403	0.845444838936665		0.124956838656418	0.852200684948399
4	0.124990820306544	0.884915899178180		0.125558568969066	0.884277655550035
5	0.121144654600721	0.862116395027634		0.121116514501350	0.860209129399964
6	0.124275815296939	0.886847963573317		0.123711723266481	0.882009097906164
7	0.124568014945908	0.873859239147986		0.123428945518956	0.877633608124518
8	0.113555823671573	0.888017913045223		0.113467245947040	0.889131273376665
9	0.0868132242637233	0.963657928776242		0.0873815044603310	0.933744462119260
10	0.112428170905100	0.945567552521516		0.114943058002946	0.910283184950084

The results of the least square error, with 10 different neurons for the training data are shown in Fig. 6, and the testing data results are shown in Fig. 7.

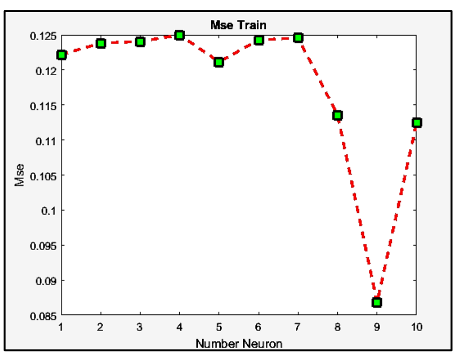

Fig. 6. MSE of training data in RBF neural network and colonial competition optimization algorithm.

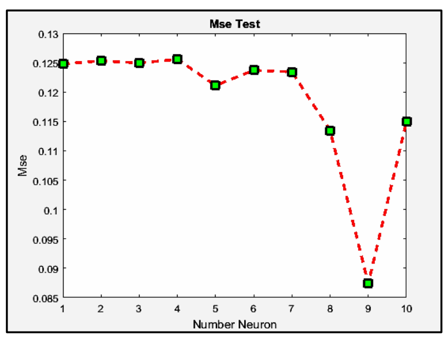

Fig. 7. MSE of testing data in RBF neural network and colonial competition optimization algorithm.

The coefficient of determination R2 for training data in the RBF neural network and the colonial competition optimization algorithm in the range between 1 and 10 neurons are shown in Fig. 8, and for the testing data, it is shown in Fig. 9.

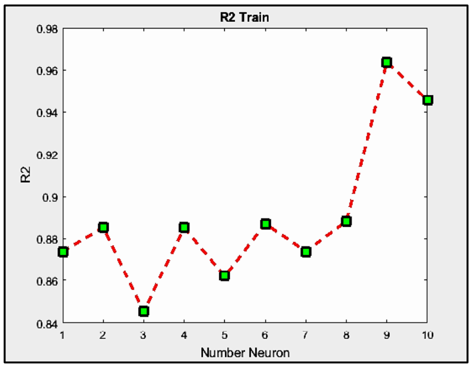

Fig. 8. MSE of training data in RBF neural network and colonial competition optimization algorithm.

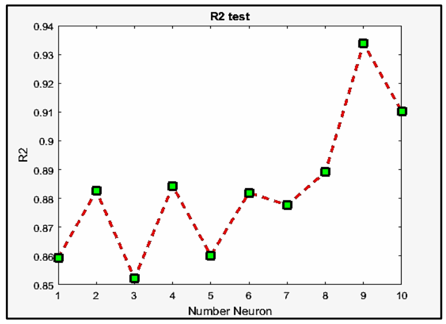

Fig. 9. R2 of testing data in the RBF neural network and colonial competition optimization algorithm.

3.4. Results of back propagation neural network

The complete results of back propagation neural network are shown in Table 4

Table 4. Complete results of back propagation neural network.

Neuron	Mse train	R^2 train	Mse test	R^2 test
1	0.203898755750366	0.702359561422528	0.296846121081952	0.704331854532943
2	0.269548476590705	0.720751400469979	0.274477303322163	0.724068136866453
3	0.317210590863107	0.719628627840569	0.299946171684056	0.720074169731394
4	0.265324623400162	0.701936997481935	0.266282029778351	0.701318850722855
5	0.212547624051097	0.851742865125929	0.212643037023726	0.835845325521894
6	0.217292621932124	0.868746700999481	0.217770371899102	0.840416476617227
7	0.300859944623472	0.721073753191069	0.293696609135207	0.722981489398773
8	0.248811089497196	0.713626781268265	0.249618307828427	0.714387958579951
9	0.275983696219863	0.840762667534088	0.378931594521114	0.816616910208510
10	0.303230770631803	0.714228911383383	0.299182456849468	0.717220918843949

The back propagation neural network with 5 neurons yielded the best result. The results of the least square error with 10 different neurons for the training dataset are shown in Fig. 10, and these results for testing dataset are shown in Fig. 11.

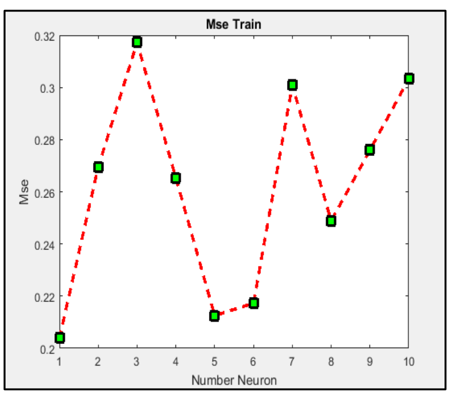

Fig. 10. MSE of training data in back propagation neural network in range between 1 and 10 neurons.

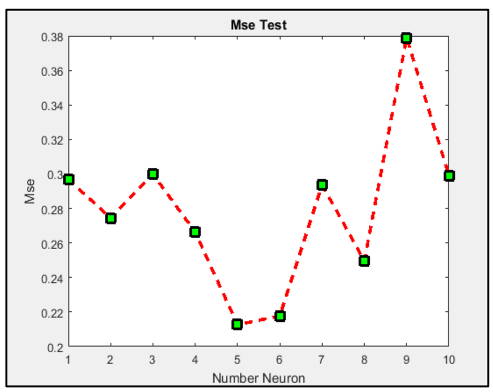

Fig. 11. MSE of testing data in the back propagation neural network in range between 1 and 10 neurons.

The coefficient of determination R2 for training data in the back propagation neural network in the range between 1 and 10 neurons is shown in Fig. 12, and it is shown in Fig. 13, for testing data.

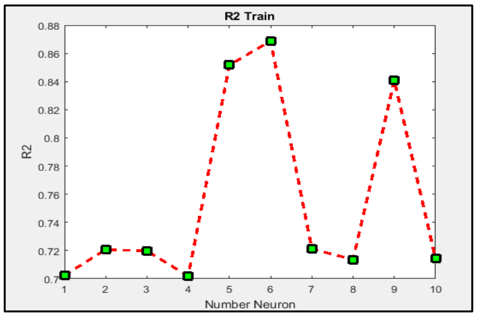

Fig. 12. Coefficient of determination R2 for training data in the back propagation neural network in the range from 1 and 10 neurons.

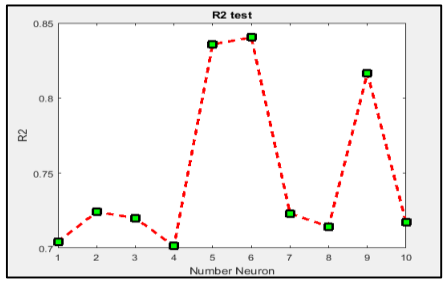

Fig. 13. Coefficient of determination R2 for testing data in the back propagation neural network in the range from 1 and 10 neurons

4.4. Evaluation of results of the proposed method

The results, including the values of the coefficient of determination R2 and the mean error of value of MSE, are shown in Tables (2, 3 and 4). Table 5 shows the comparison of three methods of RBF neural network, combined RBF neural network and colonial competition optimization algorithm, and back propagation neural network. Based on the results observed, the RBF neural network method and colonial competition optimization algorithm with 9 neurons is the most appropriate method for estimating the re-hospitalization of diabetic patients.

Table 5. comparison of algorithms examined.

Method	------------	Mse train	R^2 train	Mse test	R^2 test
Rbf	Neuron=5	0.1576	0.8783	0.1714	0.8733
Rbf+ICA	Neuron=9	0.0868	0.9636	0.0873	0.9337
NEWFF	Neuron=5	0.2125	0.8517	0.2126	0.8358

5. DISCUSSION

Based on the results of Table (4), the radial base function neural network combined with colonial competition optimization algorithm in comparison to methods such as back propagation neural network and radial base function neural network without feature selection yielded the better result.

Numerical estimates can be considered as advantages of radial base function neural network. Due to the inherent ability of radial base function in working with scattered data, using them as the basis for interpolation of multivariate data led to an acceptable efficiency.

Another advantage of radial base functions is their higher order of accuracy on a scattered distribution of points compared to other techniques.

In addition, some advantages of colonial competition optimization algorithm, compared to other evolutionary algorithms include novelty of the idea, based on human social behavior which is smarter than biological behaviors, high convergence rate and the ability to optimize functions with very high number of variables.

6. CONCLUSION AND RECOMMENDATIONS FOR FUTURE STUDIES

In previous related studies, investigating the financial status of hospitals in the area of re-hospitalization of patients has been emphasized, while the quality of hospital services and the effect of various diseases have not been considered. A few of features have been used in these studies and type 1 diabetes has been ignored.

In the method proposed in this research, after preprocessing of data, three prediction models in the area of estimating the re-hospitalization risk of diabetic patients were examined. The idea of this research was examining the differences of results of using three methods, including estimation with radial base function neural network, estimation with radial base function combined with colonial competition optimization algorithm, and estimation with back propagation neural network in order to achieve better results.

In order to expand the present research, the following items are recommended:

•Estimating the risk of re-hospitalization of diabetic patients by combining different neural networks and genetic optimization algorithms.

•Estimating the risk of re-hospitalization of diabetic patients by combining different neural networks and other evolutionary optimization algorithms.

• Estimating the risk of re-hospitalization of diabetic patients by combining support vector machine and various evolutionary algorithms.

REFERENCES

[1] Diabetes, adopted from the International Foundation for Prevention and Control of Diabetes: http://www.idpcf.com/Default.aspx?PageId=3aZm1ag4. 10-12, 2016.

[2] Statistics, adopted from the International Foundation for Prevention and Control of Diabetes: http://www.idpcf.com/Default.aspx?PageId=3aZm1ag4. 10-12, 2016.

[3] O. Hasan. **"The Role of Readmission Risk Assessment in Reducing Potentially Avoidable Rehospitalizations."** *Prescriptions for Excellence in HEALTH CARE*. 2011.

[4] W. X . Tao Zheng**. "A Machine Learning-based Framework to Identify Type 2 Diabetes through Electronic Health Records".** *International Journal of Medical Informatics.* pp. 20-127. 2017.

[5] A. Norouzi, T. Saba, M. Rahim, M. Shafry, I.M. Amin, A. Rehman. **"Visualization and Segmentation of 3D Bone from CT Images"**. *International Journal of academic research*, Vol. 4, pp. 202-206, 2012.

[6] A. A. L. Ramiro Meza –Palacios. **"Development of a Fuzzy Expert System for the Nephropathy Control Assessment in Patients with Type 2 Diabetes Mellitus" .***Expert Systems with Applications*. Pp. 335-343, 2017.

[7] C.-Y. H. Hsiao-Hsien Rau. **"Development of a Web-based Liver Cancer Prediction Model for Type II Diabetes Patients by using an Artificial Neural Network".** *Computer Methods and Programs in Biomedicine*, pp. 58-65. 2016.

[8] M. O. Okan Erkaymaz. **"Impact of Small-world Network Topology on the Conventional Artificial Neural Network for the Diagnosis of Diabetes".** *Journal of Biomedical Informatics.* No 60, pp. 162-168. 2016.

[9] E. Ariana, Anderson, W. T**. "Electronic Health Record Phenotyping Improves Detection and Screening of Type 2 Diabetes in the General United States population: A Cross-sectional, Unselected, Retrospective Study"** .*Journal of Biomedical Informatics.* Vol. 60, pp.162-168, 2016.

[10] F. Herbert and A. Jelinek. **"Data Analytics Identify Glycated Haemoglobin Co-markers for Type 2 Diabetes Mellitus Diagnosis".** *Computers in Biologyand Medicine*. Vol. 75 , pp. 90-97. 2016.

[11] G. Mohammed and M. F. Ahamad. **"Clustering as Data Mining Technique in Risk factors Analysis of Diabetes, Hypertension and Obesity .EJERS".** *European Journal of Engineering Research and Science.* 2016.

[12] N. S. Yoichi Hayashi. "**Rule Extraction using Recursive-Rule Extraction Algorithm with J48graft Combined with Sampling Selection Techniques for the Diagnosis of Type 2 Diabetes Mellitus in the Pima Indian Dataset".** *Informatics in Medicine Unlocked2*. Pp. 92-104. 2016.

[13] V.S.R.P. Kamadi. A. A. Varma. **"A computational intelligence approach for a better diagnosis of diabetic patients".** *Computers and Electrical Engineering***.** Vol. 40, pp. 1758-1765. 2014.

[14] C. B. Ogˇuz Karan. **"Diagnosing diabetes using neural networks on small mobile devices"**. *Expert Systems with Applications*. Vol. 39, pp. 54-60. 2012.

[15] M. Laura, P. Enomoto, P. Deepika, B. Shrestha, Meredith, S. Rosenthal, M. Christopher, D. Hollenbeak, and A. Robert. Gabbay. **"Risk factors associated with 30-day readmission and length of stay in"** .*Journal of Diabetes and Its Complications***.** pp. 122-127. 2017.

[16] A. Najafian. (2016). **"Thirty-day Readmission after Lower Extremity bypass in Diabetic Patients".** *Journal of Surgical Research 200***.** 2016.

[17] B. G. C. L. Hammill. **"Incremental Value of Clinical Data beyond Claims Data in Predicting 30-day Outcomes after Heart Failure Hospitalization".** *Circ Cardiovascular Quall Outcomes.* pp. 60 -67, 2011.

[18] N, S. J. Allaudeen. **"Inability of Providers to Predict Unplanned Readmissions".** *J Gen Intern Med*. 2011**.**

[19] D. I . Walraven. Derivation and validation of CMAJ. 2010.