ISSN 2286-4822
www.euacademic.org

EUROPEAN ACADEMIC RESEARCH
Vol. X, Issue 6/ September 2022

Impact Factor: 3.4546 (UIF)
DRJI Value: 5.9 (B+)

Prediction of Blood Glucose Level (BGL) on Junk Food Consumers Using Artificial Neural Network and Logistic Regression Model

SAMINA NAZ
Department of Mathematics, University of Balochistan, Quetta, Pakistan
saminabugti2015@gmail.com
NAVEED SHEIKH[1]
Department of Mathematics, University of Balochistan, Quetta, Pakistan
naveed.maths@gmail.com

Abstract

Blood glucose level (BGL) or Diabetes is a prevalent condition for which there is currently no treatment. Every year, it costs millions in our state to care for impairments associated with diabetes, as anticipated by medication specially on junk food consumers. As a result, more correctly predicting the status of patients is critical, and forecasts of high accuracy and dependability must be accurate, and trustworthy techniques must be employed. The employment of artificial intelligence systems, namely neural networks, is one of these ways. Assuming the accuracy of mathematical analyses such as the logistic regression model, this research attempted to construct a new combination with the lowest error and highest consistency by integrating these mathematical techniques with neural networks. The reliability and effectiveness of the approach have been studied and appropriate comparison is made to the neural network and logistic regression techniques have been produced by using this conceptual framework and different approaches and analysing numerical data acquired. The parameters used in this study are its ability to reduce the absolute error in neural network training using a neural network in a hybrid model, with the result that the neural network's error function is equal to 0.08 and the integrated neural network model is equivalent to 0.0001.

Keywords: Blood Glucose Level (BGL), diabetes, logistic regression, artificial neural network, combined neural network

INTRODUCTION

In the previous 15 years, the global prevalence of diabetes has tripled due to junk food consuming. Around 300 million individuals are affected, resulting in a 10% increase in

[1] Dr. Naveed Sheikh received the M.Sc. degree in Applied Mathematics from the University of Balochistan, Pakistan, in 2003. M.Phil. leading to Ph.D. degree in Mechatronics Engineering from the Asian Institute of Technology, Bangkok, Thailand, in 2015. He currently works as an Assistant professor (Tenure Track System) at the Applied Mathematics Department, Control and Robotics, Department of Computer Sciences, at University of Balochistan, Quetta, Pakistan. His current research interests include the dynamics and control of robotic systems, adaptive control and social robots, Artificial Intelligence, Neural Network, Fuzzy Logic System and Mathematical Modeling.

the annual diabetic population worldwide, with even more than three million people infected in Pakistan. Mankind have always experienced from many diseases, and while doctors have been able to identify diseases and provide treatments to help patients, a failure of treatment can cause patients to stay untreated for lengthy periods of time, putting their lives at risk. The survival of the patient is in threat. As a result, several research has been conducted in the field of illness prediction to the point where modern individuals can gain advantage from decision support algorithms and sophisticated strategies of forecasting. These are some of the implementations of decision support models is in the healthcare profession, specifically in the assessment of illnesses like as diabetes. Diabetic diagnosis and prediction are delayed as a result of poor blood glucose level (BGL) management. The danger of capillaries and macrovascular complications, as well as eye illnesses and kidney failure, are all factors to consider. As a result, a model to predict diabetes was suggested, which physicians may use as a tool to anticipate diabetes. This study evaluated the association between diabetes - related complications and patient characteristics such as blood glucose level (BGL), cholesterol levels, obesity, anaemia, and body weight. (S. A. Kakar et al. 2018) used conditional logistic regression method to estimate albuminuria in type 2 diabetes employing neural networks. (A. Kakar et al. 2018) used logistic regression model to compute the odds ratio of a neuroleptic atypical type and a diabetes diagnosis in each of the age categories, adjust for demographic effects, and diagnostic. (Nielsen et al. 2014) employed fuzzy neural networks computational model to enhance diabetes forecasting (Ruczinski, Kooperberg, and Leblanc 2003), proposed composite intelligence systems for the diagnosis of microalbuminuria in type 2 diabetes patients without the need to test urine albumin.

Complex systems include a logistic regression model of novel computing approaches for machine learning, knowledge representation, and lastly using the gained information to predict output response (Jafarian and Rahimloo 2016). Every year, it takes a lot of money to treat for diabetes-related infirmities, and in this research article, we employ a hybrid neural network and logistic regression models to predict diabetes early enough to avoid complications for junk food consumptions and save money on further diabetes-related expenditures (Case et al. 2014).

2 RESOURCES AND PROCEDURE

2.1 An overview of diabetes

Diabetes, often known as diabetes mellitus, is a metabolic illness that affects the body. This illness destroys the patient's body's capacity to generate insulin, or the body develops insulin resistant, and so the released insulin cannot fulfil its usual role. The basic function of insulin is to decrease blood sugar levels through several methods. Diabetes is classified into two kinds. In category I diabetes, pancreatic beta cell damage results in reduced insulin levels, whereas type II diabetes causes increasing insulin resistance in the body, which may ultimately lead to pancreatic beta cell damage and insulin production abnormalities. It is well recognised that genetic mutations, overweight, and a sedentary lifestyle all contribute in type II diabetes.

2.2 Diabetes classifications

Diabetes is a rapidly spreading illness caused by a reduction or lack of insulin production. There are several forms of diabetes that are generally identified at diagnosis; thus, identifying the form of diabetes is interpreted within the context under which their illness exhibits. Because the former classification of diabetes was divided into two types: insulin-based and non-insulin-based, the American Diabetes Association initiated a unique categorization of diabetes: Diabetes types I and II, as well as diabetes during pregnancy and additional kinds.

2.3 Artificial Neural Network

Adaptive Artificial Neural Network is a non-parametric approach used in the medical profession to categorise individuals as diseased or normal primarily on key parameters. Artificial neural networks are used to categorise and estimate a medical symptoms focused on risk indicators (Thirugnanam et al. 2012). The heterogeneous composition of the normal brain motivates artificial neural networks. Million neurones communicate with one another (synaptic connections) to form a neural connectivity in the nervous system focused on human skills in reading, understanding, communicating, respiration, mobility, sound identification, feature extraction and problem resolution. In general, artificial neural networks replicate a portion of brain processes (Marateb et al. 2014).

2.3.1 Architecture of artificial neural network

Artificial neural network is dynamical conceptual model of sophisticated mathematical tools that have gained a significant and innovative role in scientific research in latest days as technology and digital feature of software's have accelerated, and the outcomes have been optimistic. Multilayer perceptron structures are a significant form of machine learning algorithm because they can simulate any parameter through any accuracy assuming they have a hidden layer, an appropriate activation function in the hidden layer, and enough hidden layer neurons. As a result, the subsequent arrangement of feedforward neural model development to diabetes diagnosis is proposed.

Typically, artificial neural networks have three kinds of neural regions, which are as follows:

- Input nodes: Retrieve the raw information being sent to the system.
- Hidden nodes: the operation of such layers is dictated by the signals and weights, as well as their link to the hidden nodes. When a hidden unit is triggered, its weights among input nodes and hidden units are computed.
- Output nodes: output unit performance based on hidden unit behaviour and weight, as well as the relationship between hidden layer and outcome.

The neural network constructed in this article has a 1-10-5 topology, as illustrated in Figure 1.

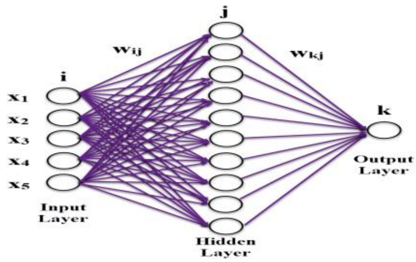

Figure 1. Artificial Neural Network Structure

2.4 Regression analysis

Regression analysis is a powerful approach for investigating the relationship among the dependent variable and the independent variable or variables, with the goal of explaining the dependent variable out from independent variable or factors. The primary distinction between regression and correlation would be that regression predicts the dependent variable using a framework in which the independent variable is one of its constituents. The correlation, on the other hand, merely indicates the strength of the link between both the independent and dependent variables (Ruczinski, Kooperberg, and Leblanc 2003).

2.4.1 The various methods of regression analysis

Five key characteristics should be highlighted in a research of regression analysis approach.

- Probit Model
- The SMC curve technique is used for regression.
- There are two types of linear regression: simple linear regression and multiple linear regression.
- Ordinal classification is a form of regression analysis that is being used to estimate a variable with an ordinal value.
- Multinomial logistic regression is used when the result has three or more potential categories which are not ordered (e.g., "disease A" vs. "disease B" vs. "disease C"). Focusing on the values of the independent variables, binary logistic regression has been used to estimate the probabilities of being a condition.

2.4.2 Statistical Regression model assumptions

- There would be a distinction between both the dependent and independent variables, as well as the two characteristics must be segregated. In other sense, should be distributed equally among independent and dependent variables.
- Theoretical models can improve the connection between variables and assumptions.
- The Gaussian distribution or Normal Distribution is a probability distribution that is symmetrical around the mean, indicating that data close to the mean occur more frequently than data distant from the mean.

- Variables with homogeneity of variance in the dependent and independent variables.

2.4.3 Logistic Regression Model

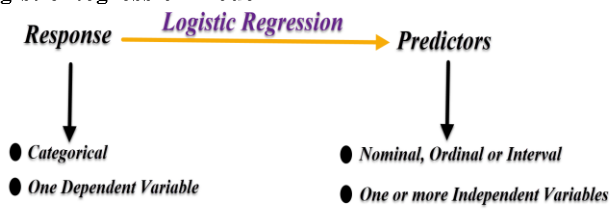

Figure 2. Logistic Regression Model

Logistic regression is a classification approach for determining the relationship between one or more categorical or continuous independent variables and the categorical dependent parameter. The regression model is illustrated in Figure 2 and defined as follows:

$$\log it\ (p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2\, x_{2,i} + \cdots + \beta_k x_{k,i} \tag{1}$$

$$p = P_r(Y_i = 1) \tag{2}$$

$$p = P_r\left(Y_i = \frac{1}{X}\right) = \frac{e^{\beta_0+\beta_1 x_{1,i}+\beta_2 x_{2,i}+\cdots+\beta_k x_{k,i}}}{1+e^{\beta_0+\beta_1 x_{1,i}+\beta_2 x_{2,i}+\cdots+\beta_k x_{k,i}}} \tag{3}$$

Except for the dependant variable, the logistic regression modelling approach is identical to linear regression in that:

Y= The odd proportion

ln(y) = Natural Logarithm

Independent Variables are $x_1, x_2, ..., x_k$

Independent variable coefficients = $\beta_0, \beta_1, \beta_2, \dots, \beta_k$

e = Error of a variable

To compute coefficients, utilise statistical models described as Maximum Likelihood, then using the logistic regression formula:

$$\ln(y') = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{4}$$

The value of y' is an approximation of the comparative odd ratio that may be computed using $\ln(y')$. That is to say:

$$e^{\ln(y')} = y' \tag{5}$$

Ultimately, by using odd ratio approximation, calculate the probability of an event as described in the following:

$$Evaluate\ the\ probability\ of\ an\ event\ occurring = \frac{y'}{1+y'} \tag{6}$$

2.5 Proposed Methodology

This study begins with a quick introduction to logistic regression and neural networks, both of which have several applications in health screening. The accurate model of blood glucose level (BGL) on junk food consumers is presented in this article by integrating neural network and logistic regression. Initially, the laboratory's essential parameters were established (See Table 1). The influence of each parameter on output and the value of the given relevant data was then calculated through logistic regression and statistical analysis parameters. The suggested model with inputs and the probability of each of the outcomes to properly forecast was created by include the possibilities of output for each of the rules as well as its effect on the intended output. The recommended approach with fewer errors displays whether or not the individual is infected with this condition, to be notified of patients in initial phases, and thus to take the appropriate steps to manage the sickness. The logistic regression findings have been implemented to the neural network and exerted a considerable influence on the neural network performance.

3 DISCUSSION AND CONCLUSION

3.1 Preparation of Data

The fifth element is examined by experienced doctors when determining the existence or absence of diabetes. Table 1 shows the statistical population of the data as well as the explanation of the parameters that used to anticipate diabetes illness. And the logistic regression model generates a result that indicates whether or not the patient is diabetic.

Table 1 Logistic regression parameters often used to diagnose diabetes

x_1 = Blood Glucose	Three hours Glucose Test	
$x_2 = \frac{Weight\ in\ Kilograms}{Height\ in\ meter\ square}$ = (BMI)	**Body Mass Index (BMI)**	**State**
	Lower than 22	Thin
	In between 22 to 25	Standard
	In between 25 to 30	Bulky
	In between 30 to 35	Obesity
	Above 35	Health Risk
x_3 = Triglycerides	**Measures of Triglycerides**	**State**
	Lower than 150 mg	Outstanding state
	In between 150 mg to 199 mg	Risky state
	In between 200 mg to 499 mg	Dangerous state
	More than 500 mg	Highly Dangerous state
x_4 = Cholesterol	**Cholesterol**	**State**
	Below 200 mg	Outstanding state
	200 to 239	Risky state
	More than 240 mg	Highly Dangerous state
x_5 = Haemoglobin A1C	**Haemoglobin A1C**	**State**
	Below 5.7	Not affected with diabetes
	5.7 to 6.4	Prediabetes
	Above 6.4	Diabetes
x_6 = Fasting Blood Sugar (FBS)	**(FBS)**	**State**
	Below 100	Not affected with diabetes
	100 to 125	Prediabetes
	Above 126	Diabetes

3.2 The output of logistic regression analysis

Logistic regression is a statistical approach for estimating and data analysis. The general form of logistic regression is as follows:

$$\log\left[\frac{-\pi}{\pi-1}\right] = \left[\sum x\beta\right] + \alpha \tag{7}$$

x would be a vector of independent factors in patients that contains body mass index, three-hour glucose levels, triglycerides, cholesterol as well as haemoglobin A1C, and β is a vector of coefficient estimates. π is likely outcomes or events to being present. Table 2 compares diabetes patients and individuals who are edge of the disease to normal individuals using parameters with predicted coefficients values and P-values.

Table 2 Predicted coefficients and the effects of prediabetes and diabetes in healthy adults.

P-Value		Predicted coefficients		Variables
Pre-diabetes	Diabetes	Pre-diabetes	Diabetes	
0.14	0.23	0.03	0.04	x_1 = Blood Glucose
0.41	0.72	-0.06	-0.21	$x_2 = \frac{\text{Weight in Kilograms}}{\text{Height in meter square}}$ = (BMI)
0.21	0.23	0.03	0.04	x_3 = Triglycerides
0.21	0.22	0.03	0.04	x_4 = Cholesterol
0.61	1.01	-0.04	0.44	x_5 = Haemoglobin A1C

Diabetes is diagnosed by 3-hour blood glucose (22%), body mass index (BMI) is (68%), triglycerides (18%), cholesterol (16%), and haemoglobin A1C (96%). In the diagnosis of pre-diabetes, 3-hour blood glucose levels are (14%), body mass index (BMI) is (40%), triglycerides (16%), cholesterol (12%) as well as haemoglobin A1C (57%). Figure 2 depicts the parameters in this model depending on the value and relevance of each parameter.

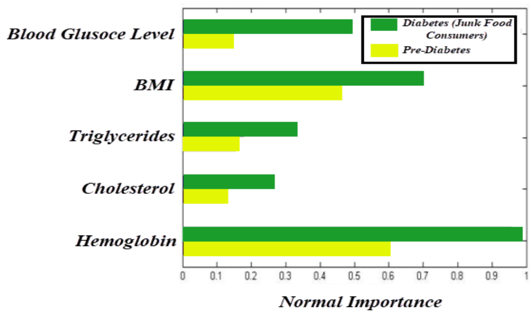

Figure 3. The logistic regression model's value of each parameter in forecasting diabetes and prediabetes

Based on the findings in Table 2, the logistic regression equations for individuals who have diabetes (P_1) with pre-diabetes (P_2) are as described in the following:

$$P_1 = 1.51x_1 - 0.92x_2 + 1.50x_3 + 1.62x_4 - 0.61x_5 \quad (8)$$

$$P_2 = 1.33x_1 - 0.50x_2 + 1.36x_3 + 1.41x_4 - 0.06x_5 \quad (9)$$

We investigated two groups in our research, one of whom were non-diabetics while the other of which would have been diabetic based on (blood glucose level (BGL) on junk food consumers). We performed one test to identify the condition on both patients, while also taking into consideration the limits of the values obtained, which range from zero to one and fluctuate to a particularly big number. Following testing, we categorise them in increasing order, indicating how much more each test result is linked to a greater disease threat.

The ROC curve for diabetes comparison to non-diabetic individuals is illustrated in Figure 4. The ROC curve area under the curve seems to be non - parametric. Despite the fact that the results of the test are numerical, and how significantly it distinguishes from both the next as well as former patients' testing results, and how significantly the patient changing demographics and whether it is reasonable or not. Such concept is extremely effective since we do not need to concern more about number of irregular curves or the design of the curves as there is only get one relevant factor, that is the area below the curve.

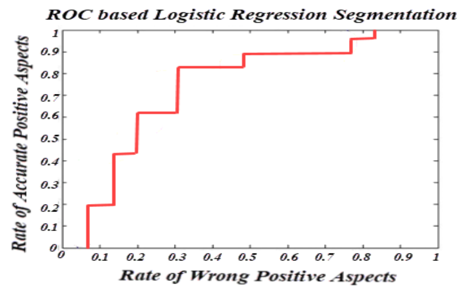

Figure 4. The variation in the area of the ROC curve between diabetes patients and non-diabetics patients.

3.3 The Findings of Combining Neural Networks with the Suggested Model

In this section, the dataset in Table 1 is used as a leading artificial neural network input, either individually or in conjunction with logistic regression. Figure 5 depicts how to minimise errors in a basic artificial neural network graph by increasing the number of repetitions and thus the error sum of squares until the highest accuracy efficiency is seen with such iterations. The red line represents the training network, the green line represents testing, and the yellow line represents network validation. This is significantly less error than the simple artificial neural network technique.

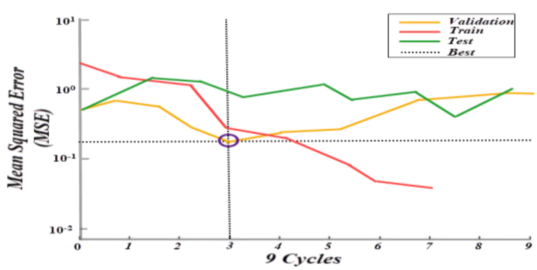

Figure 5. The neural network functionality of the artificial neural network paired with logistic regression

Figure 6 illustrates a data regression graph for the training phase, testing phase, validation phase and complete data. Here horizontal axis represents the desired output target, while the vertical axis represents the neural network's output. The best line to generate an operation is indeed a black dotted line, however the line that generated the neural network procedure in such an artificial neural network is a red line with a gradient of 0.94 0.35 calculated by multiplying the target plus 0.43, a green line with such a gradient of 0 calculated by multiplying by the target plus 0.03, a yellow line with a gradient of 0.39 calculated by multiplying the target plus 0.77, a black line with a gradient of 0.4 multiplied by the target plus 0.9.

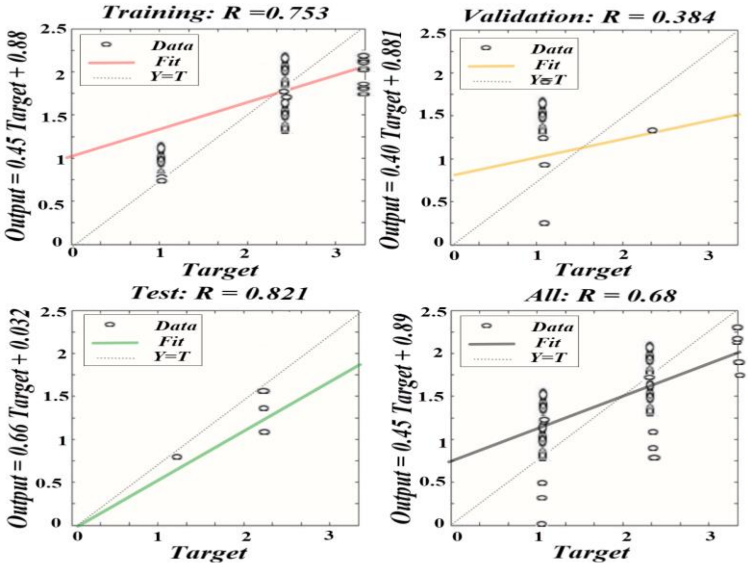

Figure 6. Regression of artificial neural network regressions regression

Figure 7 depicts the logistic regression of integrating the artificial neural system to regression models, and indeed the lines that this method may generate are red lines having slope of 1 calculated by multiplying the target plus 0.0003, green lines having a gradient of 1 calculated by multiplying the target plus 0.003, yellow lines having a gradient of 1 calculated by multiplying the target plus 0.004, and black lines having a gradient of 1 multiplied by the target plus 0.0009.

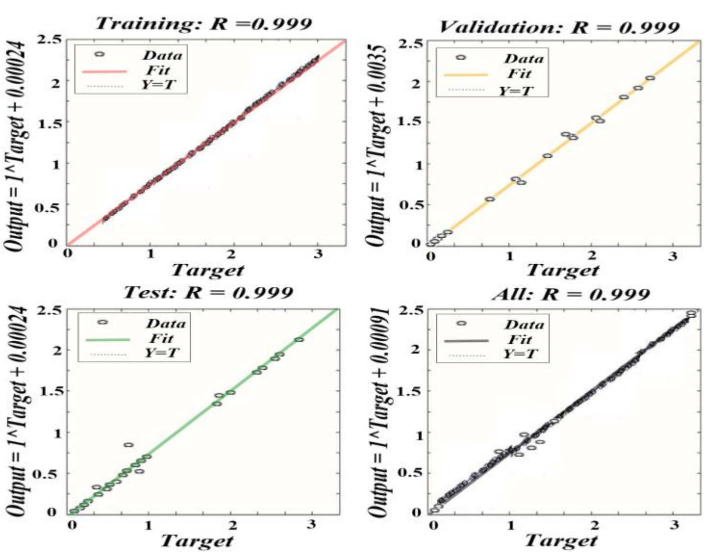

Figure 7. Regressions of a composite artificial neural network employing logistic regression

Figures 8 and 9 illustrate the histogram error of the basic artificial neural system and the histogram error of the coupled artificial neural system incorporating logistic regression, respectively.

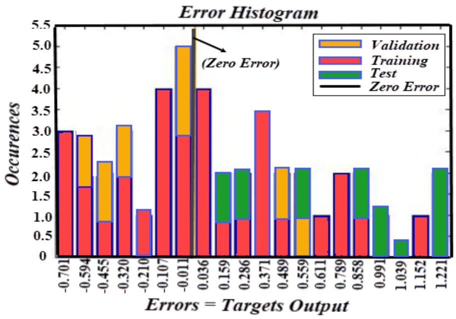

Figure 8. Histogram showing the artificial neural network's error values.

Figure 9. Histogram error levels as an outcome of combining an artificial neural system using logistic regression

CONCLUSION

In this article, a hybrid neural network model was designed to estimate diabetes by integrating an artificial neural network with logistic regression. The purpose of this study would be to develop appropriate factors that have an effect on diabetes, as well as to evaluate a neural network hybrid model using logistic method to estimate diabetes. The suggested framework is designed in the MATLAB software. The suggested technique with minimal errors requires that a person diagnosed with this condition or not seem to be conscious of it at a preliminary phase and take the appropriate steps to prevent the infection. The performance requirement, which is to set a minimum error functionality in training and testing utilizing a hybrid neural network architecture, led to the realization that the neural network's error function is 0.08 and the hybrid artificial neural network architecture is equivalent to 0.0001. Diabetes may be anticipated by using approach with the highest reliability and lowest error.

REFERENCES

1. Case, Andrew D., Ronald Byrd, Eddrena Claggett, Sandra DeVeaux, Reno Perkins, Cindy Huang, Michael J. Sernyak, et al. 2014. "Stakeholders' Perspectives on Community-Based Participatory Research to Enhance Mental Health Services." *American Journal of Community Psychology* 54 (3–4): 397–408.
2. Jafarian, Ahmad, and Parastoo Rahimloo. 2016. "Prediction of Diabetes by Using Artificial Neural Network, Logistic Regression Statistical Model and Combination of Them." *Bulletin de La Société Royale Des Sciences de Liège*, 1148–64.
3. Kakar, Azizullah, Naveed Sheikh, Bilal Ahmed, Saleem Iqbal, Abdul Rahman, Saboor Ahmad Kakar, Arbab Raza, Samina Naz, and Junaid Babar. 2018. "Systematic Analysis and Classification of Cardiac Rate Variability Using Artificial Neural Network." *International Journal of Advanced Computer Science and Applications* 9 (11): 746–50.
4. Kakar, Saboor Ahmad, Naveed Sheikh, Adnan Naseem, Saleem Iqbal, Abdul Rehman, Aziz ullah Kakar, Bilal Ahmad Kakar, Hazrat Ali Kakar, and Bilal Khan. 2018. "Artificial Neural Network Based Weather Prediction Using Back Propagation Technique." *International Journal of Advanced Computer Science and Applications* 9 (8): 462–70.
5. Marateb, Hamid R., Marjan Mansourian, Elham Faghihimani, Masoud Amini, and Dario Farina. 2014. "A Hybrid Intelligent System for Diagnosing Microalbuminuria in Type 2 Diabetes Patients without Having to Measure Urinary Albumin." *Computers in Biology and Medicine* 45 (1): 34–42.
6. Nielsen, Dennis S., Łukasz Krych, Karsten Buschard, Camilla H.F. Hansen, and Axel K. Hansen. 2014. "Beyond Genetics. Influence of Dietary Factors and Gut Microbiota on Type 1 Diabetes." *FEBS Letters* 588 (22): 4234–43.
7. Ruczinski, Ingo, Charles Kooperberg, and Michael Leblanc. 2003. "Logic Regression." *Journal of Computational and Graphical Statistics* 12 (3): 475–511.
8. Thirugnanam, Mythili, Praveen Kumar, S. Vignesh Srivatsan, and C. R. Nerlesh. 2012. "Improving, the Prediction Rate of Diabetes Diagnosis Using Fuzzy, Neural Network, Case Based (FNC) Approach." *Procedia Engineering* 38: 1709–18.