International Journal of Advanced Computer Technology (IJACT)

ISSN:2319-7900

DIABETIC FOOT PREDICTION USING HYBRID ARTIFICIAL INTELLIGENT SYSTEMS

Magdy M.Aboul-Ela, Department of Computer and Information Systems, Sadat Academy for Management Science, Cairo, Egypt,
Abeer A. Amer, Department of Computer and Information systems, Sadat Academy for Management Sciences , Alexandria, Egypt,
Soha M. Ismail, Department of Computer and Information Systems, Sadat Academy for Management Sciences, Alexandria, Egypt.

Abstract

In the past few years, Artificial Intelligence (AI) field has witnessed an intensive research interest towards integrating different computing paradigms such as Expert System (ES), Fuzzy Logic (FL) and Artificial Neural Network (ANN) systems to generate more efficient hybrid systems. The field of AI is largely used in medical areas. Nowadays, several researches have been conducted to diagnose and predict diabetic mellitus and its complications such as Diabetic foot (DF). DF is one of the major public health problems in Egypt. Therefore, designing and building a hybrid intelligent system for DF prediction suggested in this study to develop an early prediction tool for diabetic people. The proposed methodology has been implemented in two stages. In the first stage, Fuzzy Expert System (FES) has been used to reduce the input from 12 inputs to 5 inputs and to deal with uncertainty of some risk factors of DF (Blood Pressure, Lipid profile, and Kidney function). In the second stage, ANN based on ES with 8 inputs and 1 output has been used for the prediction of DF by implementing three different types of learning Back-propagation algorithms (Levenberg-Marquardt (LM), Conjugate Gradient (CG), and Gradient Descent (GD)) as a learning function. Experiments results presented a comparison of the learning functions that show the LM is more accurate than the others. Furthermore, a comparative study has been conducted between all input ANN and integrated input ANN. The result shows the efficiency of using integrated inputs. The proposed approach improved a prediction accuracy measure for the recognition of DF with the rate of (98.4%).

Keywords— Artificial intelligence, expert system, fuzzy logic, artificial neural network, diabetic foot, back propagation**.**

INTRODUCTION

Diabetes mellitus is a fatal disease that results into a number of other serious illnesses such as heart attack, loss of kidney, hypertension, blindness and diabetic foot. In developing countries, there is an apparent deterioration in the quality of life of many patients due to the inadequacy of medical specialists. Providing a solution for inadequate medical services using the education of human resources requires long time and high expenses that may result in increasing the rate of morbidity of patients. Currently, a great increase occurred in the use of computer technology in the fields of medicine area diagnosis, treatment of illnesses and patient pursuit. Artificial intelligence in medicine provided numerous advantages in diagnosis, management, and prediction of highly complicated and uncertain diseases. Despite the fact that these computerized fields have very high rate of complexity and uncertainty, the use of intelligent systems such as expert systems (ES), fuzzy logic (FL), artificial neural network (ANN) and genetic algorithm (GA) have been developed in order to improve health care, minimize treatment expenses and improve the quality of life [1][2].

The main branch of artificial intelligence programs is expert systems which are highly distinguished in a specific field. A computer can represent the knowledge of an expert (human) that is passed and stored into it by a set of rules. However, the current knowledge can be accessed by users. In addition, based on the stored knowledge, the inference is used by the computer to reach to a specific result. ES is considered to be a powerful and flexible tool to find solutions to different problems that frequently cannot be solved by the classical methods. However, ES can be classified in different categories in medicine based on the aim of development. These categories are a diagnosis, predicting and prognosis, simulation, design, treatment, monitoring, teaching and instruction, management and control and planning [2] [3].

Generally, fuzzy logic is used to monitor the nonlinear systems that are difficult to model mathematically [4]. Fuzzy logic is a set of mathematical principles for knowledge representation based on degrees of membership rather than classical binary logic [5]. Different vocabulary can be used to represent fuzzy logic i.e. fuzzification, defuzzification, membership function, linguistic variables, domain, rules etc. [4]. It is one of the artificial intelligence technology tools used to handle ambiguity, uncertainties. In addition, it is basically introduced to improved tractability, robustness and low-cost solutions for most common problems. It forms a significant part in fuzzy set theory which is used in many fields such as expert system, forecasting, fuzzy control and decision making [5].

Recently, an ANN has been used to solve non-linear separable problems [6]. ANN represents a type of computing model that is similar to the human brain [7].Neural networks are made up of a series of layers of simple elements called parallel processing neurons [3]. There are mainly three types of layers in a neural network which are input layer, hidden layer and the output layer [7]. Each neuron in ANN receives a number of inputs. These inputs are exposed to an activation function which causes an output value of the neuron [3]. Neural networks are mainly characterized by their ability to learn with examples (training vectors, input and output samples of the system). The number of neurons in a layer along

ISSN:2319-7900

with the number of layers depends on the complexity of the system studied. However, ANN cannot specify the number of layers and neurons or give an interpretation of the functionality [7] [8].

A new category of systems is the Hybrid intelligent system which relies on the artificial intelligence. It depends mainly on the most important characteristics of such systems as expert system, fuzzy logic, neural network and genetic algorithms. Mixed between systems usually follow by data exchange [9]. Hybrid intelligent systems usually combine two intelligent technologies. However, recent applications tend towards the hybrid integration containing two or more intelligent technologies. The connectionist artificial intelligent approaches and the hybrid of intelligent systems can be categorized according to system structures such as fuzzy expert systems; neural network based expert systems, neuro-fuzzy systems [10].

One of the fields of artificial intelligence system used to solve problems is fuzzy expert system (**FES**). **FES** is an expert system using fuzzy logic rather than Boolean logic. It is a combination of fuzzy sets, fuzzy rules that are used to make an inference about data, it also includes many "IF-THEN" rules. In case of having difficulty setting rules for a decision making process, ANN based Expert systems appears to be very helpful. Expert system provides results only for the field of its knowledge. In other words, ES cannot find a solution for a problem out of its field of knowledge [8] [11].

Therefore, ANN based on the ES is made for building a knowledge base of the expert system. Results of expert system and neural network integration is useful since the ANN realizes numeric data processing for the ES, the ES controls the learning process of the neural network, and the ES changes the output ANN data to provide a better interpretation [9].

One of the problems encountered by a fuzzy system is the problem of finding membership functions. Besides, convenient rules are often an exhausting process of try and error. Thus, it became necessary to apply learning algorithms to the fuzzy systems that have efficient learning algorithms such as ANN, had been presented as an alternative to automate or to support the development of tuning fuzzy systems. The hybrid of ANN and FS is used to integrate both characteristics. ANN provides a good way to adjust the expert's knowledge and automatically produce additional fuzzy rules and membership function, to satisfy certain requirements and minimize design time and costs [8]. Neuro-fuzzy systems are extensively used to figure out pattern and classification recognition problems. Neural networks are highly advantageous due to their learning capabilities and their ease of implementation. Contrarily, they are considered to be disadvantageous because of the non-interpretability of their results. The fuzzy inference systems can transmit their results using a knowledge base or rule base. The cooperation of neural networks and fuzzy inference systems makes use of both techniques [12].

According to World Health Organization (WHO), Diabetes causes many acute complications such as DF. DF may occur if treatment is delayed. This delay of treatment can lead to sensory loss and damage to limbs. The importance of designing an intelligence system has been sensed in this field. Some of the works that have been done in the diabetic field using FES and ANN have been recognized by many researchers as a key study topic in ES hybrid methods. In this paper, three intelligent systems are used; expert system, fuzzy logic, and neural network techniques for prediction of DF.

The rest of this paper is organized as follows: Section 2 contains literature survey of hybrid intelligent systems in diabetic and its complications. Section 3 included the suggested methods. Section 4 introduces experimental results followed by conclusion in Section 5.

LITERATURE REVIEW

Many Intelligent system techniques have been applied to diagnose and predict diabetes and its complications. Most works reported employs: Expert system (ES) Fuzzy logic (FL), Artificial Neural Networks (ANN), a Genetic algorithm (GA) or mixed between the pervious techniques.

Work in [13] proposed the fuzzy expert system for diagnosis of diabetes using fuzzy verdict mechanism. Triangular membership functions with Madani's inference are used in fuzzy verdict mechanism. Defuzzification method is adopted to convert the fuzzy values into crisp values. The fuzzy verdict mechanism then executes rules to make a decision on the possibility of individuals suffering from diabetes and to present the knowledge with descriptions. Experimental results indicated that the method can analyze data and further transfer the acquired information into the knowledge to simulate the thinking process of humans.

Authors in [14] are forecasting for diabetic mellitus using ANN. The back-propagation algorithm has been chosen for learning and testing of 768 data whereby 268 of them are diagnosed with diabetes. The aim of training is to adjust the weights until the error measured between the desired output and the actual output is reduced. The training stops when this reaches a sufficiently low value. To analyze the data, neural network toolbox which is available in MATLAB software is used. The network with eight inputs and four outputs are then tested and results obtained are compared in terms of error. The result of this study will provide solutions to the medical staff in determining whether someone is a diabetes sufferer or not which is much easier rather than currently doing a blood test.

Authors in [15] have successfully developed an automated classifier system of diabetic retinopathy in fundus images. The system was developed based on Adaptive Neuro-Fuzzy Inference System (ANFIS) to differentiate between normal and non-proliferative eyes. ANFIS was trained with Back

ISSN:2319-7900

propagation in combination with the least squares method. ANFIS provides the best classification and can be used as a screening tool in the analysis and diagnosis of retinal images.

Researchers in [16] detected a new method on diabetes disease. They used principal component analysis (PCA) and adaptive neuro-fuzzy inference system (ANFIS) to improve the diagnostic accuracy of diabetes disease. The system has two stages. In the first stage, the dimension of diabetes disease dataset that has 8 features is reduced to 4 features using PCA. In the second stage, diagnosis of diabetes disease is conducted via adaptive neuro-fuzzy inference system classifier. The obtained classification accuracy of their system was 89.47%.

The work in [17] used fuzzy logic to improve the classification accuracy of the Pima Indian diabetes dataset. A hybrid system is designed consisting of two neural networks, an artificial neural network (ANN), and another Fuzzy neural network (FNN) that is trained using a back propagation algorithm. The inputs are divided into two groups: fuzzy-like age and blood pressure; the rest are considered as crisp data. The first stage of the proposed model is standardizing the crisp input values and feeding them to the first ANN (ANN1). After fuzzifying the fuzzy data, their values are presented to FNN and the obtained result is defuzzyfied. The attained results of ANN1 and FNN are fed to the second ANN (ANN2) to calculate the final output. If the output value is different from the actual value, weights of these networks will change. This process is repeated until satisfactory results are reached. The accuracy of this model using the k-fold cross validation is 84.24%.

The proposed hybrid intelligent system used the fuzzy expert system in addition to the neural network based on expert system. The inputs have separated into a couple of groups: fuzzy such as blood pressure and medical tests and rest are deemed to be crisp data. Fuzzy expert system is applied to integrate the fuzzy inputs then feeding them to ANN together with the crisp inputs. ANN has been used for the prediction of DF.

METHODOLOGY

Figure 1 shows the proposed methodology which is divided into three phases: collection tabulation of data, factors integrated based on fuzzy expert system and artificial neural network based on expert system for prediction of DF.

1- Data collection tabulation:

The data of 750 of diabetes and some cases with non-diabetes patients is obtained from AVC hospital, Alexandria, Egypt. The data consists of 15 variables including uncertainty and certainty data which is blood pressure. Medical tests are coded as numeric values, gender is coded as 1 or 2, and many others are coded as true or false value. The patients are both male and female.

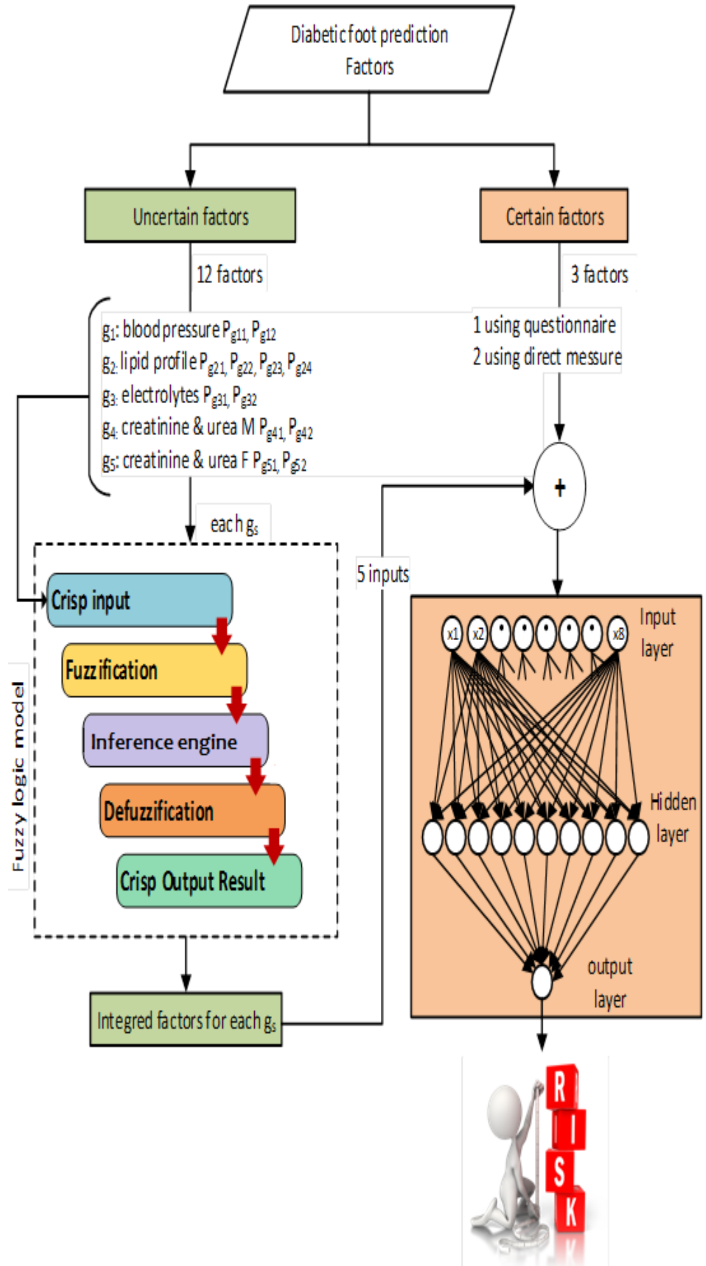

Fig. 1: Hybrid intelligent system for prediction diabetic foot

2- Factors integrated based on fuzzy logic:

In this study, a fuzzy expert system is utilized mainly to minimize the inputs of risk factors of the diabetic foot from 12 inputs to 5 inputs. These inputs included blood pressure (systolic, diastolic), Lipid profile {cholesterol (Chol.), triglycerides (Trig.), HDL, LDL} and kidney function {[part I = Electrolytes (Na, K)], [(part II_M = creatinine, urea for male)] and [(part II_F = creatinine, urea for female)]}. The trapezoidal membership function is used for all inputs, the triangular membership function is used for all outputs and

ISSN:2319-7900

Mamdani technique is used for inference engine. Firstly, the inputs and outputs will be described, after that the fuzzy inference system and defuzzification will be shown.

2.1 Input and output variables:

The input data are prepared carefully and modeled into 5 groups for further analysis. Input and output variables are organized and summarized and knowledge is generalized using fuzzy set. Fuzzy sets of inputs are:

- Blood pressure {Normal, Low risk, Moderate risk, High risk and Emergency},
- Lipid profile {Normal, Low risk, Moderate risk, High risk and V. high risk},
- Electrolytes {Low, Normal, Low risk, Moderate risk, and High risk},
- Creatinine & Urea for male {Low, Normal, Low risk, Moderate risk, and High risk},
- Creatinine & Urea for female {Low, Normal, Low risk, Moderate risk, and High risk}.

The output variables the main goal of this field is to collect the input data for each group from g1 to g5 separately. The output variables have been divided into 4 fuzzy sets (normal, low risk, moderate risk, high risk and emergency).

2.2 Fuzzification:

Fuzzification represents the process of taking a crisp input value and transforming it into the degree required by the terms. In case the form of uncertainty arises due to imprecision, ambiguity, or vagueness, the variable probably turns to be fuzzy and can be represented by a membership function [13].There are generally four types of fuzzifiers, which are used for the fuzzification process. They are: Trapezoidal fuzzifier, Triangular fuzzifiers, Singleton fuzzifier, and Gaussian fuzzifier [18]. Fuzzification of data is performed by calibrating the input parameters into the horizontal axis and projecting vertically to the upper boundary of membership function to determine the degree of membership [5]. In this study trapezoidal function as shown in equation (1) is adopted as the membership function of the inputs fuzzy number and can be expressed as the parameter set [a,b,c,d] [14]. It is used in this study because of being widely used and matching with the nature of data (risk factors of the diabetic foot). Then triangular membership functions are used for all outputs.

$$(x;a,b,c,d) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & d \le x \end{cases} \quad \textbf{(1)}$$

For instance, the fuzzy values are represented in figure 2 and 3 the input fuzzy value for blood pressure systolic and diastolic and figure 4 the output fuzzy value for blood pressure.

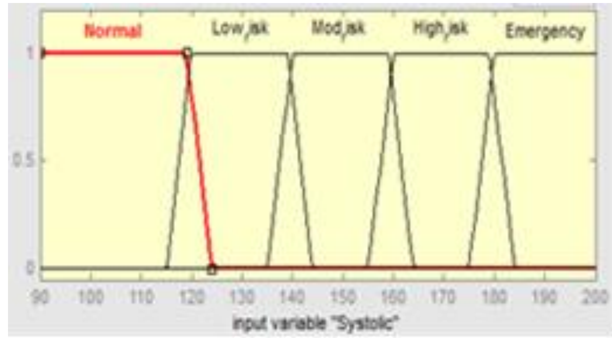

Fig. 2: Membership functions of systolic

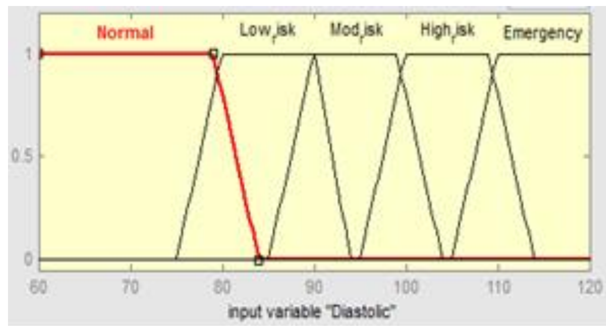

Fig. 3: Membership function of diastolic

2.3 Fuzzy inference system

Fuzzy inference is the process of formulating the mapping from a given input to an output using **FL**. **FL** is usually used for building fuzzy rules that can be easily understood by humans. Therefore, it is common to describe fuzzy variables as linguistic variables. The linguistics variables that will be used in this research are shown in table 1 and 2 for all the input and output parameters in the fuzzy model. By using those linguistic variables, fuzzy if-then rules which are the main output of the fuzzy system would be set up: generally presented in the form of if x is A then y is B where x and y are linguistic variables and A and B are linguistic values, determined by their fuzzy sets. The first part of the rule is known as the antecedent. It is made up of various parts with the operators AND or OR between them. The latter part is known as the consequent. It consists of several outputs [5] [18]. In this study, AND operator is applied for the antecedent part and the consequent part included one output. A set of 84 rules has been defined using the medical tests as well as the expert knowledge on the medical domain. Some rules will be given in table 3 for instance.

Since it gives accurate results, intuitive, and suitable for subjective input and output, Mamdani's inference engine is applied. In Mamdani's fuzzy inference method, the rules utilize the input membership values as weighting factors to

ISSN:2319-7900

determine their influence on the fuzzy output sets of the final output conclusion [5].

Table1: Classification of all inputs

Input fields	Blood pressure ranges				Fuzzy sets
	Systolic (g_{11})	*Diastolic (g_{12})*			
Blood pressure (g_1)	90-119	60-79			Normal
	120-139	80-90			Low risk
	140-159	90-99			Mod. Risk
	160-179	100-109			High risk
	SBP ≥ 180	DBP ≥ 110			Emergency
	Lipid profile ranges				
	Chol. (g_{21})	*Trig. (g_{22})*	*HDL (g_{23})*	*LDL (g_{24})*	
Lipid profile (g_2)	0-200	0-200	>60	-	Normal
	200-210	200-210	50-60	<130	Low-risk
	210-280	210-280	35-49	130-159	Mod-risk
	280-400	280-400	<35	160-189	H. risk
	-	-	-	>189	V. high risk
	Electrolytes ranges				
	Na^2 (g_{31})	*K^2 (g_{32})*			
Electrolytes (g_3)	>135	> 3.5			Low
	135-145	3.5-5.3			Normal
	146-150	5.3-5.9			Low risk
	151-160	5.9-6.5			Mod.risk
	161-170	6.5-8			H. risk
	Creatinine & urea (Male) ranges				
	Creatinine(g_{41})	*Urea (g_{42})*			
Creatinine & Urea (g_4)	< 0.9	< 17			Low
	0.9-1.3	17-43			Normal
	1.4-1.7	44-50			Low risk
	1.8-2	51-65			Mod.risk
	2.1-15	66-300			H. risk
	Creatinine & Urea (Female) ranges				**Fuzzy sets**
	Creatinine (g_{51})	*Urea (g_{52})*			
Creatinine & Urea (g_5)	< 0.6	< 15			Low
	0.6-1.1	15-36			Normal
	1.1-1.6	37-42			Low risk
	1.6-2	42-60			Mod. Risk
	2-15	61-300			High risk

2.4 Defuzzification:

Defuzzification translates the latter output into crisp values. Although, there are some defuzzification methods, such as center-of-area (gravity), center-of-sums, max-criterion and mean of maxima. The center of area (gravity) is used as shown in equation No. 2 which is the most widely used technique because, when it is used, the defuzzified values tend to move smoothly around the output fuzzy region, thus giving a more accurate representation of fuzzy set of any shape [13] [18].

$$\text{Output of } g_s = \frac{\sum_{i=1}^{n} Z_I \mu(Z_i)}{\sum_{i=1}^{I=n} \mu(z_i)} \qquad \textbf{(2)}$$

Where

Z_i means the weight for $\mu(z_i)$

$\mu(z_i)$ means the fuzzy numbers of the output fuzzy variable, gs means the integrated output groups which s = No. of groups from 1 to 5.

Table 2: Classification of all outputs fuzzy parameters

Output field	Blood pressure ranges	Fuzzy sets
Blood pressure	0-1	Normal
	1.01-2	Low risk
	2.01-3	Moderate Risk
	3.01-4	High risk
	4.01-5	Emergency
Lipid	**Lipid ranges**	**Fuzzy sets**
	0-1	Normal
	1.01-2	Low risk
	2.01-3	Moderate Risk
	3.01-4	High risk
Part_I	**Part_I ranges**	**Fuzzy sets**
	0-1	Normal
	1.01-2	Low risk
	2.01-3	Moderate Risk
	3.01-4	High risk
Part_II_M	**Part_II_M ranges**	**Fuzzy sets**
	0-1	Normal
	1.01-2	Low risk
	2.01-3	Moderate Risk
	3.01-4	High risk
Part_II_F	**Part_II_F ranges**	**Fuzzy sets**
	0-1	Normal
	1.01-2	Low risk
	2.01-3	Moderate Risk
	3.01-4	High risk

ISSN:2319-7900

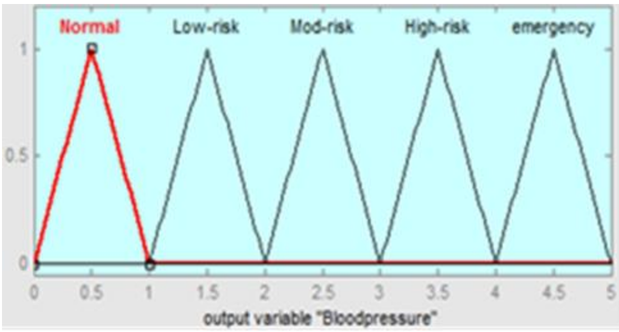

Fig. 4: Membership functions of output blood-pressure

Table 3: Examples of rules

25 rules for blood pressure are constructed like:					
	Antecedent part				**Consequent part**
IF	systolic is normal	**AND**	diastolic is normal	**THEN**	blood pressure is normal
	systolic is low_risk		diastolic is low_risk		blood pressure is low_risk
	systolic is high_risk		diastolic is high_risk		blood pressure is high_risk.

3. Artificial neural network based on expert system with integrated inputs:

In this section, inputs & output of training data set, back propagation algorithms, training functions (levenberg–Marquardt, conjugate gradient, and gradient descent), and testing the neural network will be shortly explained.

Expert system is required to have two main components: Inference Engine and Explanation facility [7]. Here ANN represents an inference engine. Thus, the expert system developed by the researcher based on using back propagation algorithm does not possess an explanation facility.

3.1 Training data set classification:

Input training data consists of blood pressure, lipid profile, kidney function part I, kidney function part II for male &female, lipid profile controlled, blood pressure controlled and examination & history of a foot. The attributes of blood pressure, lipid profile, and kidney function part I, kidney function part II is coded as a numeric value and **it represents the output of the fuzzy system.** Lipid profile is controlled, which means that (cholesterol, triglycerides, HDL and LDL) are within normal range by using medications. Blood pressure is controlled, which means that (systolic and diastolic) are within normal range by using medications. Lipid profile-is controlled and blood pressure is controlled each of attributes is coded as true = 1 or false = 0. Gender is coded as male = 1 or female =2 Examination & history of foot included previous amputation, past foot ulcer history, foot deformity (calluses or blistering), poor glycemic control, cigarette smoking, peripheral vascular disease (pulses, ABI) and peripheral neuropathy. In this paper, the risk factor of examination & history of the foot are determined according to the AVC hospital based on the AMERICAN COLLEGE OF CHEST PHYSICIANS (ACCP). Examination & history of foot had been given by questionnaire from patients as shown in table 4. Each attribute is coded as true (T) = 1 or false (F) = 0. The result of examination & history of the foot is shown in table 5. The output of the ANN model is the risk which is divided for four variables (normal, low risk, moderate risk and high risk). The input training data of risk factor of the diabetic foot will be shown in table 6 and the output training data will be shown in table 7.

3.2 Training phase:

The training of the ANN is characterized by three stages: learning, validation, and test. The proposed system has taken 750 patient records of risk factors of diabetic foot and non-diabetic foot used as patterns which is divided into the ratio of 75:15:10 for training the neural network. The selected patient's records are of same quality but with different in value and linguistic. In this study feed forward neural network architecture was used and trained with the error back propagation algorithms which are Levenberg–Marquardt, Conjugate gradient, and gradient descent.

Table 4: The input training data for examination and history of foot

Examination & history of foot	**T**	**F**	**Each Risk Factor Represents in 1 Point**
1-Previous amputation			
2-Past foot ulcer history			
3- Foot deformity			
Calluses or Blistering			
4- Poor glycemic control			
5- Cigarette smoking			
6- Peripheral vascular disease			
a- Pulses			
b- ABI			
			Each Risk Factor Represents in 2 Points
7-Peripheral neuropathy			
Total			

Table 5: Result of examination & history of foot

Result of examination & history of foot	
Total Risk Factor Score	***Risk Level***
0	No risk (Normal)
1-3	Low risk
4-6	Moderate risk
7-9	High risk

Table 6: The input of training data

No.	**Input name**	

ISSN:2319-7900

		Represents as:
1	Blood pressure	Numeric value
2	Lipid profile	Numeric value
3	Electrolytes	Numeric value
4	Creatinine & urea	Numeric value
5	Blood pressure controlled	True Or False
6	Lipid profile controlled	True Or False
7	Gender	1 = male or 2 = female
8	Examination & history of foot	Risk level (normal or low risk or moderate risk or high risk)

Table 7: The output of training data

No.	Output name	Coded as:
1	No risk (Normal)	Value = 1
2	Low risk	Value = 2
3	Moderate risk	Value = 3
4	High risk	Value = 4

3.2.1 Back-propagation Algorithms (BP)

BP is a classical domain dependent technique for supervised training. It runs by calculating the output error figuring out the gradient of this error and adjusting the ANN weights (and biases) in the descending gradient direction. Therefore, BP is a gradient descent local search method. The training process is a steady adaptation of connection weights that sends out information between simple processing units called neurons. The neurons are arranged in layers. There are connections between the neurons in one layer to those of the next [20]. BP has three layers i.e. input layer, hidden layer, and the output layer. The first layer is the input layer which identifies to the problem input variables (risk factors for diabetic foot) with one node for each input variable. The second layer is the hidden layer used to capture non-linear relationships between variables separately (10-15-20) hidden layers are used in this research. In this study, the third layer is the output layer which has only one neuron identical result (risk). The relationship between the output y_t and the input x_t is given by equation No. 3 [21]. BP has a broad used technique of ANN if a higher accuracy is required for the solution in problem domain [6]. BP algorithm is widely used in applications of the neural network in medicine [21].

$$y_t = w_0 + \sum_{j=1}^{q} w_j \cdot f\left(w_{0,j} + \sum_{i=1}^{p} w_{i,j} . x_t \right) \qquad (3)$$

Where

$w_{i,j}$ $(i = 0,1,2,\ldots\ldots, p;\ j = 1,2,\ldots.., q)$

w_i $(j = 0,1,2,\ldots q)$ are the connection weights.

p is the number of input nodes.

q is the number of hidden nodes.

f is a nonlinear activation function.

Activation function enables the system to learn nonlinear features. The symmetric saturating linear transfer function is applied in this paper. In relation to the training patterns or examples, the BP has to be trained to reach the required output. The primary aim is to discover the best training functions for prediction of the diabetic foot. For performance assessment of training functions parameters are recognition accuracy, the speed of training, correctness. In this paper, the most important parameter is used as the mean squared error (MSE) [22] which is given by equation No. 4.

$$MSE = \frac{\sum_{I=1}^{N} (y_t - o_t)^2}{N} \qquad (4)$$

Where:

N: Number of examples {input, output}.

y_t: The target output.

o_t: The observed output.

The minimization of the error mean (MSE) is guaranteed by the adjustment of the ANN's weights (w) until the error measured between the output desired and the actual output is reduced, by using one optimization algorithm through the different training functions [23]. This paper focuses on three various back propagation algorithms which are [Levenberg-Marquardt (LM), Conjugate gradient with Powell/Beale Restarts (CG), and Gradient descent GD)].

3.2.2 Levenberg–Marquardt Algorithm (trainlm)

LM is a mixture of a gauss Newton method and the steepest descent method for solving least-squares problem [23]. If a smaller value of the damping factor is used, the algorithm manner nears a Gauss-Newton. However, if a greater damping factor is chosen, then the algorithm behaves such as a gradient descent algorithm. The main aim of this algorithm is to find a minimum of a function that is a group of squares of non-linear functions. Trainlm's damping factor is set at every iteration and that generates the performance to the optimization problem. This characteristic causes trainlm to be the fastest training algorithm for networks of moderate size. Trainlm function has the problem of memory and computation overhead resulted from the calculation of the gradient and approximated Hessian matrix [22]. Trainlm algorithm sets the weight (w) of the network [23] by the following equation 5,6,7.

$$w_{k+1} = w_k - \frac{J_k e_k}{J`_k J_k + \lambda I} \qquad (5)$$

Where:

W: the weight of the network.

J: Jacobian matrix of the MSE error.

ISSN:2319-7900

e: Error between the desired and calculated network output.
k: Number of iterations.
I: Matrix identity.

Organizing strategy of the damping factor λ of Levenberg-Marquardt algorithm is made according to the following:
If the calculated error *of* ***MSE*** for w_{k+1} reduces then [23]:

If $\lambda = \lambda / 10$ **(6)**

Else $\lambda = \lambda * 10$ and $w_{k+1} = w_k$ **(7)**

Where:
λ = damping factor

3.2.3 Conjugate gradient algorithm (traincgb) with Powell/Beale Restarts

The weight has been updated by CG in the steepest descent direction (negative of the gradient). Thus, the performance function falls almost quickly. It needs only a little more storage than the other algorithms. Therefore, these algorithms are useful for networks with a big number of weights [22]. At the first iteration, traincgb is more capable; the direction of gradient descent "d" [23] is given from equations No. 8 to 11.

$$d_{old} = -2 j`_k \, e_k \quad \textbf{(8)}$$

Where
d: gradient descent
From the second iteration

$$\beta = \frac{d`_{old} \, d_{old}}{(J`_K \, e_k)` J`_K \, e_k} \quad \textbf{(9)}$$

and

$$d_{new} = -2 J`_K \, e_k + \beta d_{old} \quad \textbf{(10)}$$

$$w_{k+1} = w_k + a d_{new} \quad \textbf{(11)}$$

3.2.4 Gradient descent algorithm (traingd)

GD is updated weights and biases in the direction of the negative gradient of the performance function. It is a local search procedure, measures the output error, and calculates the gradient of the error by adjusting the weights in the descending gradient direction [21] [22]. The equation of traingd is given by (12).

$$\Delta w_k = -a_k . J'_k \quad \textbf{(12)}$$

Where:

Δw_k is a vector of weights changes

J'_k is the current gradient

a_k is the learning rate that determines the length of the weight update.

Table 8: The performance for different training algorithms in NN with 8 inputs

EXPERIMENTAL RESULTS

The algorithm implemented using C#, visual studio 2010, and MATLAB R 2011 b language. FL toolbox in MATLAB language (version 7.01) is used; all training functions are coded in MATLAB using ANN toolbox. C sharp language is used for user interface.

Experiment design: The proposed system has been simulated in the (HP) PC machine which has the following features: Intel (R) Core (TM) i5-2450M CPU @ 2.50GHz, and 4.00GB of RAM, 64-bit Windows 7.

To escape probable bias in the presentation order of the sample patterns to the ANN these sample sets were randomized. Symmetric saturating linear transfer function is utilized for the hidden layer. Initially, random weights are designated for all the layers. The learning process of the algorithm will stall when maximum numbers of iterations are completed. When learning is finished, final weights are stored for testing the test data set. The final weights are stored for testing data. 250 patient's records risk factors of diabetic foot and non-diabetic foot were used.

Basic system training parameters are max_epochs=1000, show=5, performance goal=0, time=Inf, min_grad=1e-010, max_fail=6 are fixed for each training function. The parameters for comparison between training functions LM, CG and GD are MSE based on the number of hidden neurons (10, 15 and 20). The network is trained until the least of mean squared error.

The first experimental result as shown in table 8 shows that LM algorithm reaches to the lower value of MSE, by increasing the number of hidden layers better over than others. The ability of LM algorithm to achieve the lower rate of the MSE among CG and GD because LM inherits from Gauss-newton. This algorithm turns to be rapid and successful when the network weights are perfectly refined by the steepest descent algorithm. CG algorithm acts like steepest decent, it had a slow convergence around the optimal network weighs, which justifies for CG algorithm not to recognize a minimal error like LM algorithm [23]. GD algorithm behavior specifies the learning rate which is multiplied times the negative of the gradient to decide the changes to the weights and biases that explains why GD does not amount to least error like LM algorithm [22]. The best performance plots of levenberg– Marquardt algorithm at 20 hidden layers are shown in fig.5.

Increasing the number of hidden layers in the network does not affect performance or gives precise results for all the training functions because the numbers of hidden layers depend on the number of inputs, the number of outputs, the number of training cases, the training algorithm and the amount of noise in the targets [24].

No. of layers	Training Algorithm	Performance			
		Train data set	*Validation data set*	*Test data set*	*Performance*
10	LM	6.4903e-002	6.2011e-002	6.0078e-002	6.3985e-002
	CG	1.7568e-002	1.8587e-002	1.6237e-002	1.7588e-002
	GD	2.3039e-001	2.4029e-001	2.467e-001	2.3351e-001
15	LM	3.8927e-002	3.5935e-002	3.4586e-002	3.8042e-002
	CG	2.401e-002	2.7969e-002	2.4457e-002	2.4651e-002
	GD	1.0396e-001	1.1292e-001	1.0969e-001	1.0588e-001
20	LM	8.859e-003	1.1094e-002	1.0022e-002	9.3121e-003
	CG	3.9598e-002	5.0888e-002	5.0001e-002	4.2339e-002
	GD	6.7655e-002	6.2226e-002	6.1932e-002	6.6265e-002

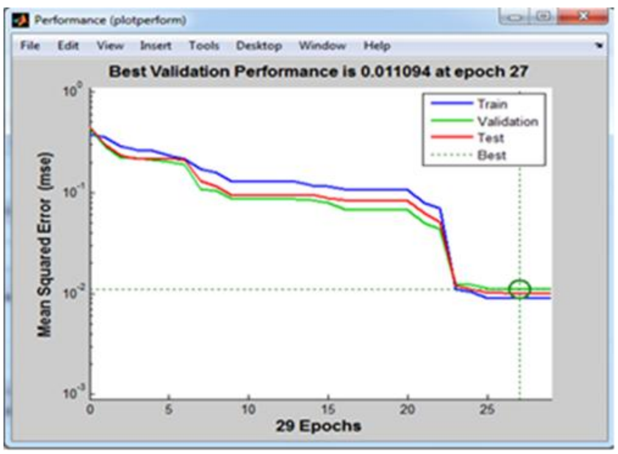

Fig. 5: The best performance plot of levenberg–Marquardt algorithm at 20 hidden layers

The second experiment evaluates the performance of neural network using all inputs based on MSE as a parameter for comparison. LM algorithm presents a lower value of MSE at 10 hidden layers comparing with CG and GD algorithms. Networks simulated using various training functions are affected according to the number of neurons in their hidden layer. Functions trainlm and traincgb are not much affected by increasing no of neurons in hidden layers [22] the result of the training sample is illustrated in table 9.

It has been observed that overtraining does not give more accuracy or efficiency for all training functions. The neural network works depending on the data being trained to the network. In case of more data training to network, it will make the network more intelligent [22]. The best performance plot of levenberg–Marquardt algorithm at 10 hidden layers is shown in fig. 6.

The last experimental result is shown in table 10. Based on the prediction accuracy Ac, a comparison is drawn between the two different ANN. Accuracy is the measuring scale for the performance of this experiment. Ac computed using the ratio of the total number of correct predictions patterns to the total predictions patterns. Ac can be determined from the equation No. 13. ANN with integrated inputs produces the correct predictions with 98.4% prediction accuracy [7] [13].

$$A_c = \frac{CP}{TP} \times 100 \qquad \textbf{(13)}$$

Where

CP = the total number of correct patterns.

TP = the total patterns.

Experimental results explain that the ANN training optimization depends on the number of input when we used the same number of training patterns.

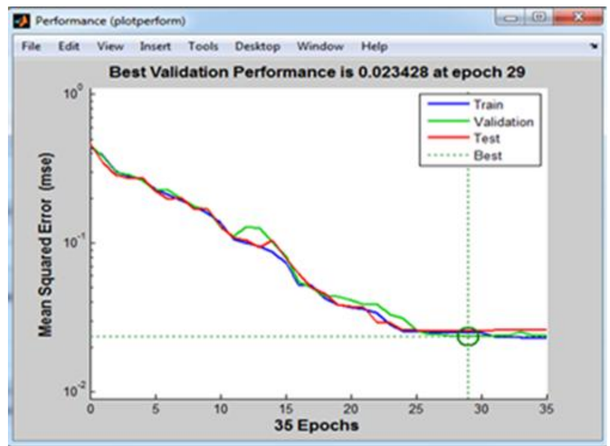

Fig. 6: The best performance plot of levenberg–Marquardt algorithm at 10 hidden layers

ACKNOWLEDGMENT

The authors would like to express their cordial thanks to staff of AVC hospital, Specially Dr. Mohamed Saad the manager of the hospital, Dr. Tarek Hafez (internal medicine and diabetes consultant), Dr. Hannaa El Nagar (internal medicine resident), Dr. Amir Sabbah (vascular surgery resident), and Dr. Rana Hamed (specialist at medical analysis).

CONCLUSION

This paper presents hybrid artificial intelligent techniques like **ES**, **FL** and **ANN** for prediction of DF, which is one of the most common complications of diabetes mellitus. The first approach relies on using FES to merge the inputs of the risk factors of the diabetic foot. Therefore, a fuzzy system that uses an inference engine to make decisions has been successfully developed. Input for the knowledge base is gathered through real-time data from hospitals and the inference engine uses a series of If-then statements to obtain more accurate intelligent results. For the second procedure, ANN based on expert system is employed with merged inputs, after that ANN is introduced to have the same feature

ISSN:2319-7900

but different in the number of inputs to calculate the risk of the diabetic foot. The novel prediction method can be applied for different back propagation train algorithms such as (LM, CG, and GD) in the method that provides high accuracy; the number of hidden layers of a neural network plays an important role for detecting the relevant features. The highest performance is obtained when the network with merged inputs used LM training algorithm at 20 hidden layers, while the output layer indicates the risk of the diabetic foot.

Table 9: The performance for different training algorithms in NN with all inputs

No. of layers	Training Algorithm	Performance			
		Train data set	Validation data set	Test data set	Performance
10	LM	2.4766e-002	2.3428e-002	2.5539e-002	2.4641e-002
	CG	4.9406e-002	5.1063e-002	4.8866e-002	4.9602e-002
	GD	2.1661e-001	2.1465e-001	2.2156e-001	2.1681e-001
15	LM	6.6592e-002	6.4436e-002	6.7359e-002	6.6344e-002
	CG	6.852e-002	6.3086e-002	6.8712e-002	6.772e-002
	GD	1.5101e-001	1.5873e-001	1.4973e-001	1.5205e-001
20	LM	7.5401e-002	7.9498e-002	7.9872e-002	7.6465e-002
	CG	7.0486e-002	7.1059e-002	6.9458e-002	7.047e-002
	GD	1.3089e-001	1.3533e-001	1.3371e-001	1.3184e-001

Table 10: The predictions accuracy of the two different ANN

No. of layers	Training Algorithm	Prediction accuracy off ANN with integrated inputs	Prediction accuracy off ANN with all inputs
10	LM	77.6%	60.4%
	CG	94.8%	58.8%
	GD	40%	38.9%
15	LM	84.8%	54.2%
	CG	94.8%	55.3%
	GD	53.2%	44%
20	LM	98.4%	45%
	CG	86%	53.1%
	GD	81.2%	56.1%

References

[1] K. Rajeswari, and V. Vaithiyanathan, "Fuzzy Based Modeling for Diabetic Diagnostic Decision Support using Artificial Neural Network", International Journal of Computer Science and Network Security, Vol.11, No. 4, pp. 126-130, April 2011.

[2] M. Zolnoori, M. Zarandi, and M. Moin, "Fuzzy Rule-Base Expert System for Evaluation Possibility of Fatal Asthma", Journal of Health Informatics in Developing Countries, Vol. 5, No. 1, pp. 171-184, Dec. 2010.

[3] Z. Abrishami, and H. Tabatabaee, "Design of a Fuzzy System and a Multi-layer Neural Network System for Diagnosis of Hypertension", Bulletin of Environment, Pharmacology and Life Sciences, Vol. 4, No. 11, pp. 138-145, Oct. 2015.

[4] F. Baig, M. Khan, Y. Noor, and M. Imran, "Design Model of Fuzzy Logic Medical Diagnosis Control System", International Journal on Computer Science and Engineering, Vol. 3, No. 5, pp. 2093-2108, May 2011.

[5] U. Uduak A., and N. Mofon M., "A Proposed Fuzzy Framework for Cholera Diagnosis and Monitoring", International Journal of Computer Applications, Vol. 82, No. 17, pp. 1-10, Nov. 2013.

[6] M. Pradhan, and R. Sahu, " Predict the Onset of Diabetes Disease using Artificial Neural Network (ANN)", International Journal of Computer Science & Emerging Technologies, Vol. 2, No. 2, pp. 303-311, April 2011.

[7] V. Nagori, and B. Trivedi, "Expert System Based on Back Propagation Network for Evaluating Motivational Strategies on Human Resources", International Journal of Applied Information Systems, Vol. 2, No.3, pp. 27- 33, May 2012.

[8] J. Vieira, F. Dias, and A. Mota, "Neuro-Fuzzy Systems: a Survey", Proceedings of the 5th WSEAS International Conference on Neural Networks and Applications, pp. 1-6, Udine, Italy, 2004.

[9] G. Setlak, M. Dbrowski, W. Szajnar, M. Piróg-Mazur and T. Kozak, "Artificial Intelligence Approach to Diabetes Diagnoses", International Journal Information Technologies and Knowledge, Vol.3, No. 13, pp. 209-214, 2009.

[10] S. Sahin, M. Tolum, and R. Hassanpour, " Hybrid Expert Systems: A Survey of Current Approaches and Applications", Expert Systems with Applications, Vol.39, No. 4, pp. 4609 – 4617, March 2012.

[11] M. Hasan, K. Sher-E-Alam, and A. Chowdhury, "Human Disease Diagnosis using a Fuzzy Expert System", Journal of Computing, Vol. 2, No. 6, pp. 66-70, June 2010.

[12] B. Fatima1 and C. Amine, "A Neuro-Fuzzy Inference Model for Breast Cancer Recognition", International Journal of Computer Science & Information Technology, Vol. 4, No. 5, pp. 163 -173, Oct. 2012.

[13] M. Kalpana, and A. Kumar "Fuzzy Expert System for Diabetes using Fuzzy Verdict Mechanism" Int. J. Advanced Networking and Applications, Vol. 3, No. 2, pp. 1128-1134, 2011.

[14] S. Jaafar and D. Ali, "Diabetes Mellitus Forecast using Artificial Neural Network (ANN)", Proceedings of the International Conference on New Techniques in Pharmaceutical and Biomedical Research, pp. 135-139, Kuala Lumpur, Malaysia, 5-7 Sept. 2005.

ISSN:2319-7900

[15] S. Kavitha, and K. Duraiswamy, "Adaptive Neuro-Fuzzy Inference System Approach for the Automatic Screening of Diabetic Retinopathy in Fundus Images", Journal of Computer Science, Vol. 7, No.7, pp. 1020-1026, 2011.

[16] K. Polat, S.Güneş, "An Expert System Approach Based on Principal Component Analysis and Adaptive Neuro-Fuzzy Inference System to Diagnosis of Diabetes Disease", Digital Signal Processing, Vol. 17, No. 4, pp.702-710, July, 2007.

[17] H. Kahramanli, and N. Allahverdi, "Design of a Hybrid System for the Diabetes and Heart Diseases," Expert Systems with Applications, Vol. 35, No.1, pp. 82-89, 2008.

[18] X. Djam, G. Wajiga, Y. Kimbi, and N. Blamah, "A Fuzzy Expert System for the Management of Malaria", International Journal of Pure and Applied Sciences and Technology", Vol. 5, No. 2, pp. 84-108, 2011.

[19] J. Zhao, and B. Bose, "Evaluation of Membership Functions for Fuzzy Logic Controlled Induction Motor Drive", Proceedings of the IEEE 2002 28th Annual Conference on Industrial Electronics Society, pp. 229-234, Sevilla, Spain, 5-8 Nov. 2002.

[20] T. Jayalakshmi, and A. Santhakumaran, "A Novel Classification Method for Diagnosis of Diabetes Mellitus using Artificial Neural Networks", Proceeding of the International Conference on Data Storage and Data Engineering, pp. 159-163, Bangalore, 9-10 Feb. 2010**.**

[21] S. Lahmiri, "A Comparative Study of Backpropagation Algorithms in Financial Prediction", International Journal of Computer Science, Engineering and Applications, Vol. 1, No. 4, pp. 15-21, Aug. 2011.

[22] B. Sharma, and P. Venugopalan,"Comparison of Neural Network Training Functions for Hematoma Classification in Brain CT Images", IOSR Journal of Computer Engineering, Vol. 16, No.1, pp. 31-35, Jan. 2014.

[23] F. Dkhichi, and B. Oukarfi, "Levenberg-Marquardt and Conjugate Gradient Training Algorithms of Neural Network for Parameter Determination of Solar Cell", International Journal of Innovation and Applied Studies, Vol. 9, No. 4, pp. 1869-1877, Dec. 2014.

[24] S. Xu, and L. Chen, "A Novel Approach for Determining the Optimal Number of Hidden Layer Neurons for FNN's and Its Application in Data Mining", Proceeding of the 5th International Conference on Information Technology and Applications , pp. 683 – 686, Cairns, Queensland, Australia, 23-26 June 2008.