

AN IMPROVED MODEL FOR CLINICAL DECISION SUPPORT SYSTEM

Odikwa, Henry¹, Ugwu, Chidiebere², Inyama Hycinth³

¹Department of Computer Science, University of Port Harcourt, Nigeria

²Department of Computer Science, University of Port Harcourt, Nigeria

³Department of Computer Engineering, Nnamdi Azikiwe University, Awka, Nigeria

ABSTRACT

Misguided information in health care has caused much havoc that have led to the death of millions of people as a result of misclassification, and inconsistent health care records; hence the objective of this paper is to develop an improved clinical decision support system. This system incorporated hybrid system of non-knowledge based and knowledge based decision support system for the diagnosis of diseases and proper health care delivery records using prostate cancer and diabetes datasets to train and validate the model. The min-max method was adopted in normalizing the datasets, while genetic algorithm was deployed in initiating the training weights of the MLP. The result obtained in this paper yielded a classification accuracy of 98%, sensitivity of 0.98 and specificity of 100 for prostate cancer and accuracy of 94%, sensitivity of 0.94 and specificity of 0.67 for diabetes.

KEYWORDS

Clinical Decision Support System, Rule Base Algorithm, Multilayer Perceptron, Genetic Algorithm, Back-Propagation.

1. INTRODUCTION

Nowadays, millions of people die all over the world as a result of improper diagnosis, late detection and misclassification of diseases, especially diseases with clinical markers and related symptoms. According to Globocan report, 2002, 307,000 deaths are recorded in Africa as a result of improper disease diagnosis which results in heavy growth of diseases as shown in Fig 1 [4]. Decision support models such as knowledge-based and non-knowledge based have been developed for effective health care delivery to curb some of these misclassification and improper documentation of patients' data in health care delivery. Clinical decision support systems are powerful tools that integrate routine clinical knowledge and rationalized patient information to boost patient care [3].

Adequate medical diagnosis is an essential task to ascertain a patient's present condition of health. The medical doctors are presented with many cases and test statistics to make decisions pertaining to treatment of patient's ailment. Most importantly, in medical diagnosis, series of tests are carried out on patients and the medical doctors are faced with the challenges of analyzing and management of the diseases. In medical decisions, medical personnel rely so much on biomarkers, which are biological tools that detect the presence of diseases in a patient [1]. All

these challenges in making medical decisions pose many difficulties in the medical sector coupled with numerous medical data that are completely out of reach to the medical practitioners [6]. These challenges are of great concern and need remedy in order to alleviate the work load of these medical personnel by employing medical decision support systems. Recently, decision support system (DSS) have been employed in different facets of life, especially in medical diagnosis using machine learning algorithm models in the classification of patients' diseases. This paper builds an improved clinical decision support system that will enhance the detection and prediction accuracies of diseases.

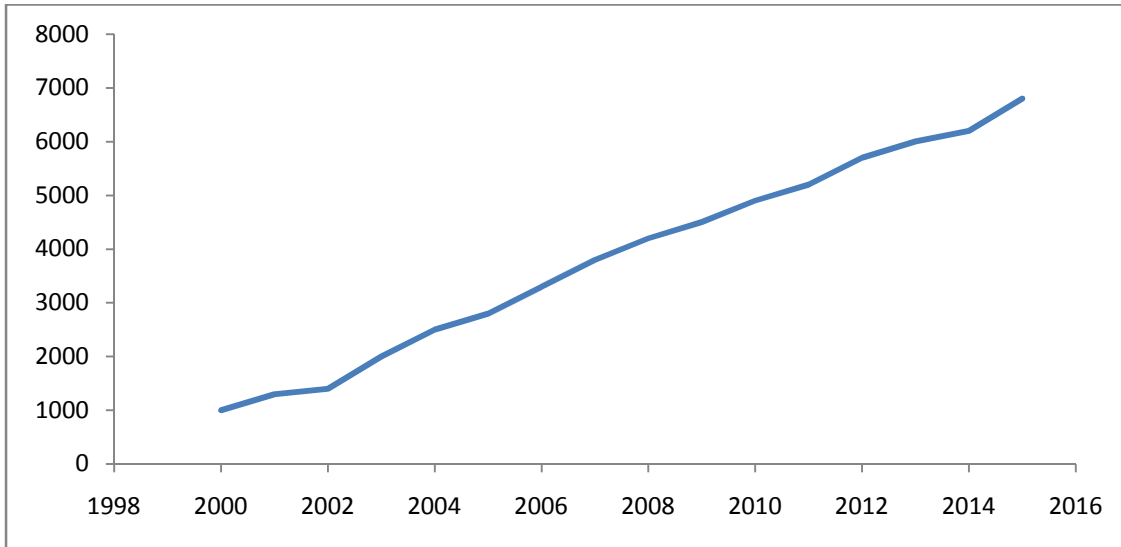


Fig. 1: Growth of Diseases in Africa 2000 to 2015 (Source: Globocan report (2002))

2. RELATED WORK

Decision support systems are clinical support that enhances health care delivery to patients and health care personnel by keeping track of the patients' medical records. The idea behind construction and building of medical support system should not be overemphasized as it is geared towards effective diagnosis of diseases using machine learning algorithms [16]. In using machine learning algorithm in developing clinical decision support system (CDSS) using knowledge base technique, [2],[10],[14] developed a case base and rule base reasoning system using rules and cases in building knowledge extraction modules for decision support system and was tested using diseases with related symptoms. This system employed a hybrid algorithm in a single knowledge based type of clinical decision support system. This system suffered from non-multifaceted problems as it was specially designed for a specific task at a time and experienced complexity of programming the cases and the rules due to vagueness. In order to eliminate the vagueness and programming complexities [13] developed a decision support system with neural network perceptron to train and validate clinical data. As a non-knowledge base decision support system, their result was more reliable in the classification of diseases and thus proved more robust than the case base and rule base reasoning involving set of rules and cases. The clinical decision support system was first built by training the datasets in a single neural network perceptron to avoid ambiguity of rules experienced in case base, whereas the trained data created a pattern

which was used in constructing the clinical decision support system. The accuracy of the result yielded 76% with a sensitivity of 0.74 and specificity of 0.80. A model for accessing patient's health care system based on real time decision support system using only case base reasoning was derived [4], an approach that characterized the use of cases in building medical decision support system, this approach was a milestone in medical support system but also suffered from programming complexities. This paper presented a different approach in building clinical decision support system by introducing a hybrid system which comprised of both knowledge-based by the use of rule reasoning and non-knowledge based in the form of artificial intelligence. This method is germane to address the problems encountered by the former machine learning algorithms in CDSS.

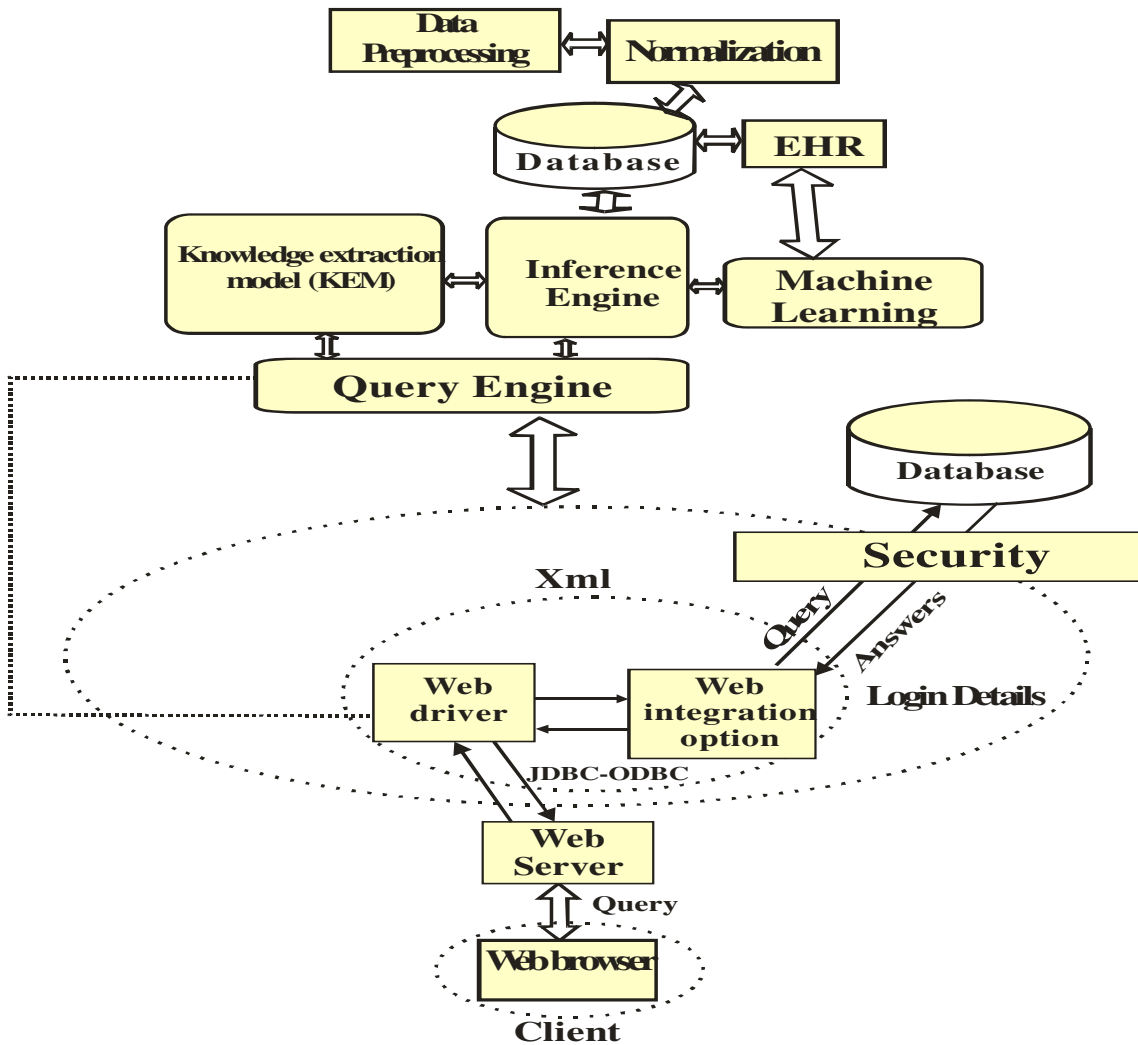


Fig. 2: Architecture of the system

3. MATERIALS AND METHODS

Fig 2 shows the architecture of the improved clinical decision support model. This new system incorporates three modules to achieve the functional capacity of the clinical decision support system. The first module is the data pre-processing phase, which prepares the datasets; the second phase comprises the training compartment, which constitutes the machine learning algorithms; and finally the web-based phase that projects the entire system into online distribution on the web.

The dataset used were 500 prostate cancer patients' repository and 100 diabetic patient's repository from 2012 to 2015 at Federal Medical Center, Umuahia, and PathConsult Nigeria Limited, both in Umuahia, Abia State, Nigeria. The datasets were normalized by applying the normalization process [12] to improve performance and reduce differences in the training results in a multilayer perceptron neural network. The data shown in Tables 1 and 2 are prostate cancer and diabetic patients' data to be normalized between 0 and 1.

3.1 The Data Normalization

Table 1: Un-Normalized Dataset for Diabetes

S/N	Age (yrs)	AIC (%)	BMI (g)	RBS(Mmol/L)	FBS(Mmol/L)	OGT(Mmol/L)
1	45	6.30	25.00	11.80	7.00	8.80
2	50	5.30	26.00	12.00	8.00	11.30
3	75	6.80	26.00	11.90	5.70	11.50
4	48	7.00	29.00	13.00	5.80	10.50
5	89	9.00	30.00	11.20	9.00	11.90
6	90	6.00	21.00	13.00	7.00	12.00
7	56	8.10	27.00	14.00	8.00	13.00
8	75	5.30	28.00	12.00	6.50	13.00
9	59	5.20	27.00	11.00	7.30	13.80
10	45	8.00	20.00	11.70	5.90	10.50
11	89	5.90	29.00	11.80	8.00	14.00
12	50	6.30	30.00	11.10	10.00	14.00
13	49	6.00	34.00	15.70	5.70	11.98
14	60	6.90	56.00	15.00	7.00	12.00
15	55	6.80	26.00	11.80	6.00	13.00
16	78	5.80	27.00	11.90	6.80	13.80
17	60	5.70	27.00	12.30	6.30	15.80
18	89	5.50	25.00	14.10	8.20	13.00
19	56	6.00	22.00	16.10	7.40	11.90
20	64	7.00	24.00	17.00	6.00	14.90

The Table 1 shows the un-normalized diabetes clinical markers dataset having five (5) features, namely glycated hemoglobin (AIC) measured in percentage, body mass index (BMI) measured in grammes, random blood sugar (RBS) measured in mol per litre, fasting blood sugar (FBS) measured in milli mole per litre and oral glucose test measured in milli mole per litre

Table 2: Un-Normalized Dataset for Prostate Cancer Data.

S/N	Age (yrs)	Prostate Weight (ng/mL)	DRE	Prostate Weight (g)	Prostate Volume (mL)
1	68	20.10	3.00	80.00	90.00
2	78	20.50	3.00	80.00	160.00
3	83	15.50	0.00	200.00	70.00
4	85	22.60	2.00	80.00	50.00
5	71	24.00	1.00	55.00	70.00
6	65	1.50	1.00	90.00	40.00
7	61	34.00	0.00	120.00	15.00
8	76	64.00	1.00	70.00	45.00
9	70	18.00	1.00	55.00	70.00
10	81	39.00	0.00	60.00	80.00
11	64	14.00	2.00	80.00	50.00
12	82	23.00	0.00	78.00	60.00
13	64	34.50	2.00	60.00	90.00
14	73	21.30	0.00	70.00	160.00
15	64	34.00	0.00	90.00	70.00
16	73	33.20	0.00	80.00	50.00
17	64	10.20	0.00	80.00	70.00
18	53	71.00	1.00	200.00	40.00
19	72	54.00	0.00	80.00	15.00

Table 2, shows the un-normalized dataset of prostate cancer clinical markers with four (4) features, namely .prostate specific antigen (PSA), measured in nanogram per milli mole; DRE which has the values from 0 to 3, 0 value means soft, the value 1 means nodular, the value 2 means firm, while the value 3 means hard. The prostate weight is measured in grams while prostate volume is measured in milli litre. The normalization of the data was done and the values range from 0 to 1. Since the various attributes (feature vectors) employed in this paper have different variable value range, and to reduce the differences in training results, normalization is necessary. The normalized data was subjected to min-max normalization method. The data normalization is determined with equation (2) [13] using the min-max value method:

$$\frac{\left(VAL. \frac{A}{F}. NORM - 0\right)}{(1 - 0)} = \frac{\left(VAL. \frac{B}{F} NORM. - MIN\right)}{(MAX. - MIN.)}$$

$$\frac{\left(VAL. \frac{A}{F}. NORM - 0\right)}{(1)} = \frac{\left(VAL. \frac{B}{F} NORM. - MIN\right)}{(MAX. - MIN.)} \quad (1)$$

$$\left(VAL. \frac{A}{F}. NORM\right) = \frac{\left(VAL. \frac{B}{F} NORM. - MIN\right)}{(MAX. - MIN.)}$$

$$X^1 = (X - Min) / (Max - Min) \quad (1)$$

Where X represents each data entry, Min is the minimum value from each row entry and max denotes the maximum value from each row entry.

Applying equation (1), the maximum value of AIC is 9 and the minimum value is 4.6; the maximum and minimum values of BMI are 20 and 56 respectively, also the minimum and maximum values of RBS are 11 and 28.1 respectively. For FBS, the minimum value is 5.6 and the maximum value is 18, and OGT has a minimum value of 8.7 and maximum value of 20 from Table 1.

$$X^1(AIC) = \frac{(6.30 - 4.60)}{(9.00 - 4.60)} = 0.40$$

$$X^1(BMI) = \frac{(25.00 - 20.00)}{(56.00 - 20.00)} = 0.10$$

$$X^1(RBS) = \frac{(11.00 - 11.00)}{(28.10 - 11.00)} = 0.00$$

$$X^1(FBS) = \frac{(7.00 - 5.60)}{(18.00 - 5.60)} = 0.10$$

$$X^1(OGT) = \frac{(8.80 - 8.70)}{(20.00 - 8.70)} = 0.00$$

These calculations represent normalization of data in Table 1 and the results of the normalized data in Table 3 for the first column and the other normalized values follow using the same process.

Table 3: Normalized Datasets for Diabetes

S/N	Age (yrs)	AIC (%)	BMI (g)	RBS (Mmol/L)	FBS (Mmol/L)	OGT (Mmol/L)
1	45	0.4	0.1	0.0	0.1	0.0
2	50	0.2	0.2	0.1	0.2	0.2
3	75	0.5	0.2	0.1	0.0	0.2
4	48	0.5	0.3	0.1	0.0	0.2
5	89	1.0	0.3	0.0	0.3	0.3
6	90	0.3	0.0	0.1	0.1	0.3
7	56	0.8	0.2	0.2	0.2	0.4
8	75	0.2	0.2	0.1	0.1	0.4
9	59	0.1	0.2	0.0	0.1	0.5
10	45	0.8	0.0	0.0	0.0	0.2
11	89	0.3	0.3	0.0	0.2	0.5
12	50	0.4	0.3	0.1	0.4	0.5
13	49	0.3	0.4	0.3	0.0	0.3
14	60	0.5	1.0	0.2	0.1	0.3
15	55	0.5	0.2	0.0	0.0	0.4
16	78	0.3	0.2	0.1	0.1	0.5
17	60	0.3	0.2	0.1	0.1	0.6
18	89	0.2	0.1	0.2	0.2	0.4
19	56	0.3	0.1	0.3	0.1	0.3
20	64	0.5	0.1	0.4	0.0	0.5

Consequently, the maximum value of PSA is 1237 and the minimum value is 0.1, the maximum and minimum values of DRE are 3 and 0 respectively; also the minimum and maximum values of Prostate weight are 250 and 25 respectively, for Prostate volume, the maximum value is 160 while the minimum value is 15 from Table 2.

$$X^1(PSA) = \frac{(20.10 - 0.10)}{(1237 - 0.10)} = 0.00$$

$$X^1(DRE) = \frac{(3.00 - 0.00)}{(3.00 - 0.00)} = 1.00$$

$$X^1(PW) = \frac{(80.00 - 25.00)}{(250.00 - 25.00)} = 0.20$$

$$X^1(PV) = \frac{(90.00 - 15.00)}{(160.00 - 15.00)} = 0.50$$

The determined normalization values of PSA, DRE, PW and PV are represented in Table 4 for the first column in Table 2 and the calculations followed thus for the other data entries normalized in the range of 0 to 1.

Table 4: Normalized Datasets for Prostate Cancer

S/N	Age (yrs)	Prostate Weight (ng/mL)	DRE	Prostate Weight(g)	Prostate Volume (mL)
1	68	0.0	1.0	0.2	0.5
2	78	0.0	1.0	0.2	1.0
3	83	0.0	0.0	0.8	0.4
4	85	0.0	0.7	0.2	0.2
5	71	0.0	0.3	0.1	0.4
6	65	0.0	0.3	0.3	0.2
7	61	0.0	0.0	0.4	0.0
8	76	0.1	0.3	0.2	0.2
9	70	0.0	0.3	0.1	0.4
10	81	0.0	0.0	0.2	0.4
11	64	0.0	0.7	0.2	0.2
12	82	0.0	0.0	0.2	0.3
13	64	0.0	0.7	0.2	0.5
14	73	0.0	0.0	0.2	1.0
15	64	0.0	0.0	0.3	0.4
16	73	0.0	0.0	0.2	0.2
17	64	0.0	0.0	0.2	0.4
18	53	0.1	0.3	0.8	0.2
19	72	0.0	0.0	0.2	0.0
20	86	0.0	1.0	0.1	0.2

This paper employed three algorithms to build a clinical decision support system with prostate clinical markers of prostate specific antigen (PSA), direct rectal examination (DRE), prostate weight (PW) and diabetes clinical markers of body mass index (BMI), fasting blood sugar (FBS), random blood sugar (RBS) and oral glucose test (OGT) in testing the CDSS. Genetic algorithm is used to evolve the initial weights of the neural network.

It is worthy to note that calculations were made based on data in Tables 2 and 3 to determine the normalized data. The Prostate Cancer feature has Minimum value of 1237, minimum value as 0.1. The direct rectal examination has maximum value of 3 and minimum value of 0; prostate weight has maximum value of 250 and minimum value of 25 while prostate volume has a maximum value of 160 and minimum value of 15 before the normalization. The diabetic patient data has the body mass index maximum value of 56 and minimum value of 20. The random blood sugar (RBS) has maximum value of 28.1 and minimum value of 11. Fasting blood sugar (FBS) has maximum value of 18 and minimum value of 5.6 while the oral glucose tolerance has maximum value of 20 and minimum value of 8.7 as initial data values from the datasets which were normalized to 0 and 1.

The normalization process is applied to the datasets in Tables 1 and 2 and the resultant normalization values are shown in Tables 3 and 4. The normalized data were stored in the database.

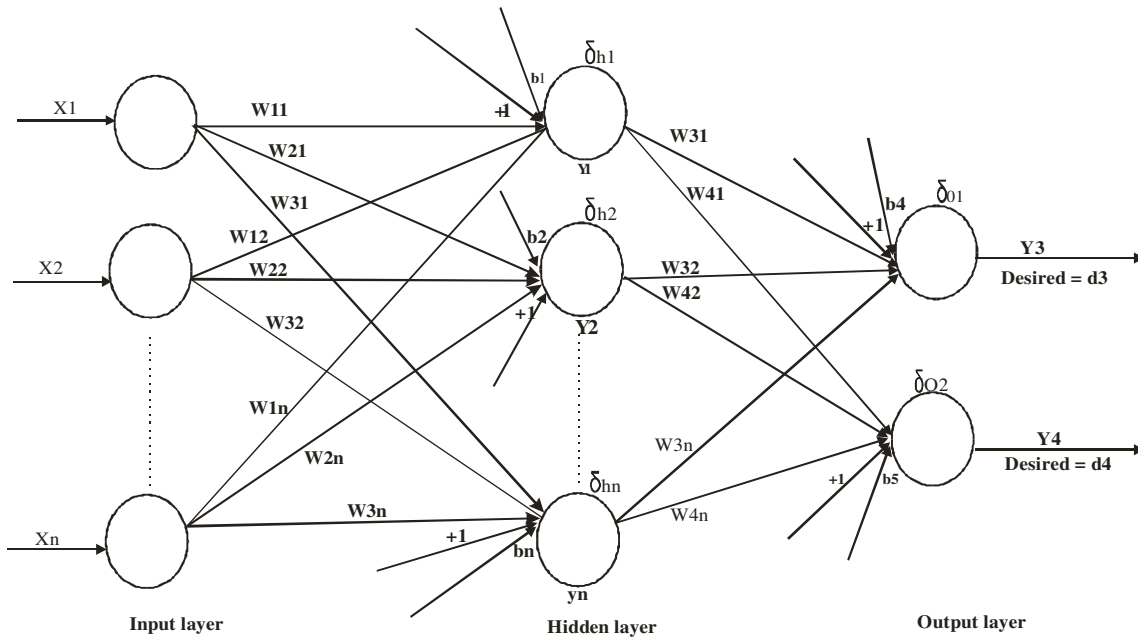


Fig. 3: The MLP Neural Network Architecture of the System

The Figure 3 shows the neural network architecture that was deployed in the multilayer neural network training using the normalized data.

3.2 Genetic algorithm

The genetic algorithm was used to initiate the neural network weights prior to the desired output. The genetic algorithm was deployed as follows with emphasis on the error calculation.

- Randomly generate initial population (weights) on a 1000, 2000 and 3000 generations.
- Calculate the fitness values by determining the error given by target - output. The fitness is given by $1/\text{error}$
- Select the individual weights based on their fitness values
- Cross the individuals based on the values of 0.7, 0.8 and 0.9 and mutated the genes with 0.01, 0.02 and 0.05.

3.3 Feed forward algorithm

The feed forward algorithm was deployed to calculate the weights in the neural network in Figure 2. The mathematical model for the feed forward algorithm is given by [6] in equation (2)

$$Input_j = x_j = \sum y_i w_{ij} \quad (2)$$

To determine the total of the input weights, hidden weights and output weights, equation (1) was expanded to equations (3), (4) and (5)

$$\Omega_1 = (x_1 * w_{11}) + (x_2 * w_{12}) + \dots (x_n * w_{1n}) \quad (3)$$

$$\Omega_2 = (x_1 * w_{21}) + (x_2 * w_{22}) + \dots (x_n * w_{2n}) \quad (4)$$

$$\Omega_3 = (x_1 * w_{31}) + (x_2 * w_{32}) + \dots (x_n * w_{3n}) \quad (5)$$

Where x_1 , x_2 and x_n are the inputs to neural network neurons and the w 's are the weights of the network.

The Output layer weights were calculated with equations (6) and (7)

$$\Phi_1 = (y_1 * w_{31}) + (y_2 * w_{32}) + \dots (y_n * w_{3n}) \quad (6)$$

$$\Phi_2 = (y_1 * w_{41}) + (y_2 * w_{42}) + \dots (y_n * w_{4n}) \quad (7)$$

In order to determine the error of the network, equation (8) was applied.

$$\delta_k = (1 - O_k) * O_k (T_k - O_k) \quad (8)$$

Where O_k is the calculated, T_k is the observed or actual output.

3.4. Back-propagation algorithm

The back-propagation algorithm is employed to adjusting the weights of the neural network that is, where genetic algorithm could not get to the actual output target, the back-propagation propagates the errors back to the neural network.

Calculate the local gradients (δ_{o1} , δ_{o2} , δ_{h1} , δ_{h2} ,... δ_{hn}) using equation (9).

$$\delta_{hn} = \varphi^1 (1 * b_3 + x_1 * w_{31} + x_2 * w_{32} + x_n * w_{3n}) * (\delta_{o1} * w_{3n} + \delta_{o2} * w_{4n}) \quad (9)$$

The weights of the network were adjusted using the learning rule depicted in equation (10)

$$W_{new} = w_{old} + \eta * \delta * input \quad (10)$$

Where w_{jk} is the weight of the connection from unit j to unit k appropriate in the next layer and δ_k is the error of unit k and η is a constant that is called the learning rate, which takes the value from 0 and 1.

The biased errors were determined with equation (11); the introduction of bias in the network was to provide every node with a trainable constant. With eqn(11), we can determine the bias errors

$$b_n(n+1) = b_n(n) + \alpha * b_n(n-1) + \eta * \delta_{h_n}(n) * 1 \quad (11)$$

The mean square error was employed to determine the errors of the network for each of the generations shown in equation (12)

$$MSE = \frac{\sqrt{(\sum_{i=1}^n (T_i - E_i)^2)}}{n} \quad (12).$$

Where T_i represents the predicted response, E_i denotes the target response and n is the number of observations (iterations).

3.5. Rule base reasoning of the CDSS

This paper deployed the following rule base reasoning algorithm in building the clinical decision support system.

For the universal set of rules [D]

Let [A] be the subset of all the rules pertaining to one prediction

Let [B] another subset comprising rules of another prediction.

For any prediction, match rules A and rules B to find out the most suitable for any prediction.

If A satisfies all the conditions in the universal set [D], then

Predict, else

Prediction is void.

Do so until all the rules have been exhausted. Then

Classify problem based on prediction A or B

End.

4. PREDICTION ACCURACY

To determine the accuracy of the system, the sensitivity and specificity were calculated with equations (13), (14) and (15), where;

$$Sensitivity(Recall) = \frac{(TruePositive)}{(TruePositive + FalseNegative)} \text{ That is,}$$

$$S = \frac{(TP)}{(TP + FN)} \quad (13)$$

Whereas true_ positive corresponds to the number of people tested positive and have the characteristics of the disease, False Negative (FN) corresponds to the number of people tested negative but have the characteristics of the disease [7], [8] also; specificity measures how well the classifier can recognize negative samples:

$$Specificity = \frac{(TrueNegative)}{(FalsePositive + FalseNegative)}$$

That is

$$SP = \frac{(TN)}{(FP + FN)} \tag{14}$$

Where true_negative corresponds to the number of people tested negative but without the characteristics, False Positive (FP) corresponds to the number of people tested positive but without the characteristics of the disease.

Precision measures how many examples classified as “positive” class are indeed “positive”:

$$Precision = \frac{(TruePositive)}{(TruePositive - FalsePositive)}$$

That is;

$$P = \frac{(TP)}{(TP - FP)} \tag{15}$$

Then the

$$(Accuracy) = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{16}$$

Table 7: Diabetic Generation for the GA

Generation	1000	2000	3000	10000	20000	30000
Training MSE	0.0394	0.0309	0.0300	0.0002	0.0012	0.0030
Testing MSE	0.0228	0.0218	0.0216	0.0006	0.0007	0.0005

Table 8: Crossover Rate and Mutation Rate of GA in Diabetes

Generations	Crossover Mutation	0.7			0.8			0.9		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
1000	Training MSE	0.00471	0.0248	0.0399	0.0557	0.0557	0.0113	0.0548	0.0667	0.0533
	Testing MSE	0.5682	0.3803	0.5511	0.6120	0.5772	0.7123	0.0615	0.0741	0.0667
2000	Training MSE	0.0289	0.0286	0.0339	0.0351	0.0357	0.0313	0.0501	0.0584	0.0526
	Testing MSE	0.5599	0.4272	0.5356	0.5752	0.6106	0.5227	0.0593	0.0709	0.0661
3000	Training MSE	0.0021	0.0229	0.0035	0.0017	0.0012	0.0012	0.0013	0.0018	0.0013
	Testing MSE	0.0016	0.0206	0.0032	0.0011	0.0010	0.0011	0.0007	0.0005	0.0002

Table 9: Crossover Rate and Mutation Rate of GA in Prostate Cancer

Generations	Crossover Mutation	0.7			0.8			0.9		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
1000	Training MSE	0.0861	0.0828	0.0765	0.0678	0.0917	0.0278	0.0781	0.0891	0.0671
	Testing MSE	0.0810	0.0890	0.0711	0.0612	0.0999	0.0500	0.0810	0.0719	0.0810
2000	Training MSE	0.0319	0.0290	0.0400	0.0610	0.0600	0.0410	0.0619	0.0612	0.0634
	Testing MSE	0.8001	0.4102	0.6123	0.6190	0.7200	0.6120	0.0610	0.0812	0.0710
3000	Training MSE	0.0312	0.0380	0.0040	0.0032	0.0200	0.0412	0.0013	0.0010	0.0020
	Testing MSE	0.0310	0.0312	0.0043	0.0034	0.8123	0.7710	0.0012	0.0011	0.0009

Table 10: Prostate Cancer Generation for the GA

Generation	1000	2000	3000	10000	20000	30000
Training MSE	0.0412	0.0400	0.0312	0.0210	0.0112	0.0100
Testing MSE	0.0612	0.0523	0.0412	0.0330	0.0289	0.0090

Table 11: Comparing Testing Time of GA and Random Weight Initialization during BPLA

Iteration	GA Weight Initialization				Random Weight Initialization			
	Prostate Cancer		Diabetes		Prostate Cancer		Diabetes	
	MSE	Time(s)	MSE	Time(s)	MSE	Time(s)	MSE	Time(s)
1000	0.0412	5	0.0394	3	0.0821	10	0.0612	8
2000	0.0400	10	0.0309	5	0.0711	12	0.0600	11
3000	0.0312	14	0.0300	10	0.0534	20	0.0543	17
10000	0.0210	21	0.0222	15	0.0501	32	0.0430	30
20000	0.0112	45	0.0111	30	0.0356	55	0.0412	50
30000	0.0100	58	0.0030	40	0.0211	2mins	0.0391	3mins

5. THE PATIENT DATABASE

The database deployed was developed into two categories; the patient database and user application database. The database houses the electronic health records of the patients as shown in Tables 5 and 6, this compartment was developed to electronically process the patients' recorded health information.

Table 12: Patient Database

S/N	Field Name	Data Type	Size	Description
1	Name	Varchar	20	Candidate Key
2	Address	Varchar	30	Candidate Key
3	Hosp_no	Varchar	5	Primary Key
4	Phone_no	Varchar	13	Candidate Key
5	DOB	Varchar	10	Candidate Key
6	State	Varchar	10	Candidate Key
7	Status	Varchar	8	Candidate Key
8	Gender	Varchar	6	Primary Key

Table 13: User Application Database

S/N	Field Name	Data Type	Size	Description
1	Name	Varchar	20	Primary Key
2	E-pin	Varchar	5	Primary Key
3	appType	Varchar	30	Application Type
4	transType	Varchar	255	Transaction Type

4. EXPERIMENTS AND RESULTS

Several experiments were conducted in building the improved model for clinical decision support system embedded as a hybrid system.

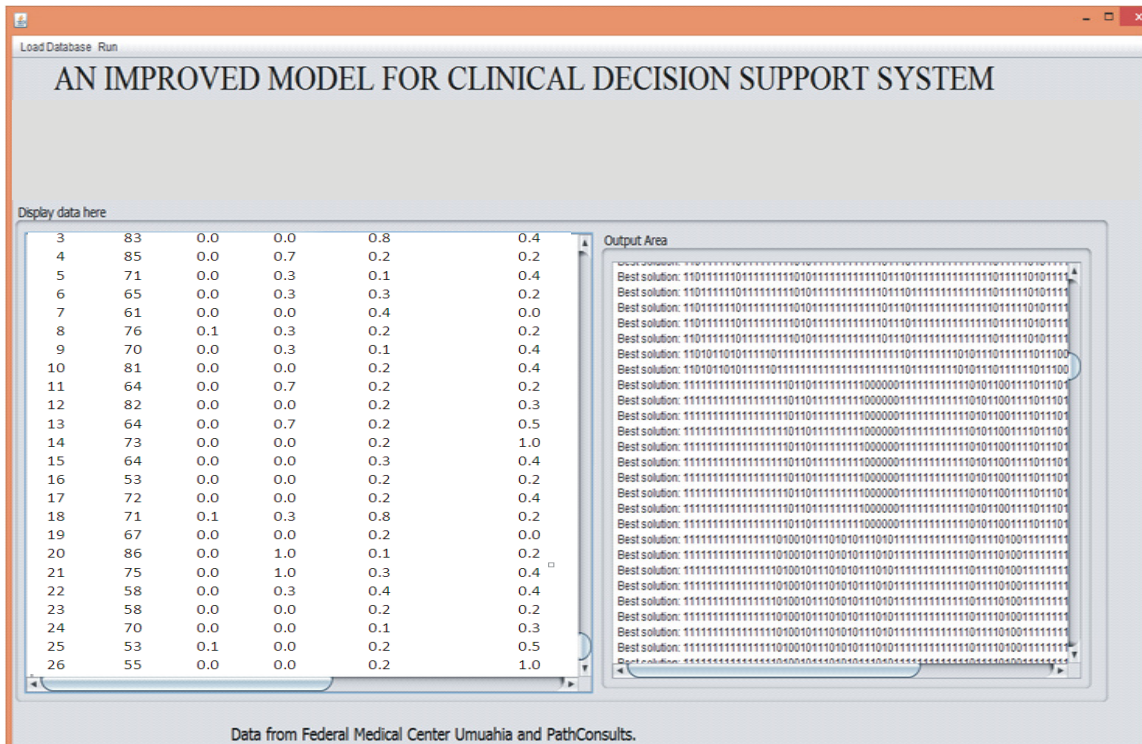


Fig 4: Weights initialization with Genetic Algorithm

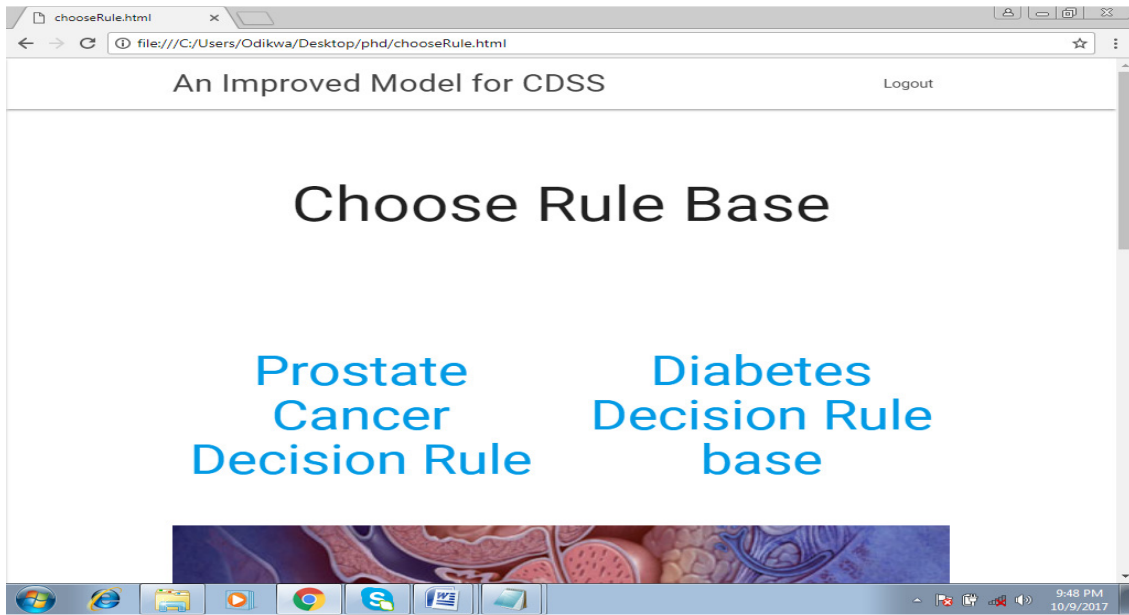


Fig. 5: Rule Base Interface of CDSS

Table 12: Result of 250 Prostate Cancer Testing Samples

Clinical Data	Diagnosis			Prognosis		Mis-classification
	Low Risk	Intermediate Risk	High Risk Cancer	Aggressive Cancer	Non-Aggressive Cancer	
PSA < 2.0 ng/mL, Prostate weight < 50g, Prostate volume < 30cm ³	2	0	0	12	0	0
PSA > 2.9 ng/mL and < 4.9ng/mL, Prostate weight > 40g but < 80g, Prostate volume > 100cm ³	2	0	0	2	0	1
PSA > 4.0ng/mL and < 10ng/mL, Prostate weight > 40g but < 100g, Prostate volume < 100cm ³ > 200cm ³	0	10	23	0	0	4
PSA > 10ng/mL, prostate weight > 40g, Prostate volume > 100cm ³	24	17	68	58	27	0
Total	28	27	91	72	27	5

Table 13: Result Analysis of Testing Diabetes Data

Clinical Data	Diagnosis		Misclassification
	Pre-diabetes	Diabetes	
BMI > 25g, RBS > 11.1m/mol, FBS > 6.9, OGT > 11.1m/mol	12	26	6
BMI < 25, RBS < 6.9, FBS < 6.9, OGT < 11.1m/mol	2	0	4
Total	14	26	10

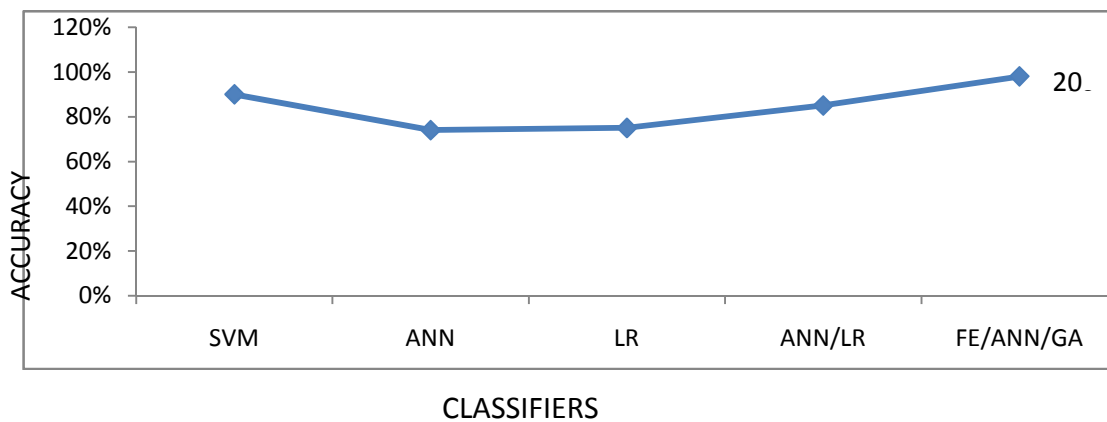


Fig. 6: Graph of accuracy, classifier and classification time

7. RESULT DISCUSSIONS

In this paper experiments were conducted that yielded promising results in the course of building clinical decision system. The system was tested with 250 prostate cancer and 50 diabetes patient datasets. Table 7 shows the crossover rate and mutation rate of GA in training MSE and testing MSE and it was found that at 3000 generations using diabetes datasets the error approached zero. Also in Table 8 of the generation for the GA in diabetes, after 30000 generations, the training MSE and testing MSE approached zero, this shows the reliability of GA in neural network weight initialization. Thus, it also follows with prostate cancer datasets, in Table 9, that the testing MSE and training MSE after 3000 individual weight generations, the errors approached zero, in Table 10, it is clear that after 30000 generations in testing and training, MSE convergence of the network was achieved as the errors approached zero. These results are attributed to the powerful nature of GA in NN weights initialization. The deployment of genetic algorithm in weight initialization minimized the global minima that would have been experienced with back-propagation. The testing time of genetic algorithm weight initialization in a multilayer perceptron was compared in Table 11 to arbitrary random weight initialization, and the GA proved superior to random weight initialization. The Table 12 shows the result analysis of the testing datasets of 250 patients and the system was able to classify 28 patients as having low risk cancer, 27 patients as having intermediate risk cancer and 91 patients with high risk cancer. With the prognostic features of the system, 72 patients were classified as having aggressive cancer while 27 patients were having non-aggressive cancer. The system misclassified 5 patients with abnormal condition as normal. Consequently, the result analysis with 50 diabetes patients' testing data in the developed system was shown in Table 13; with 14 patients classified as having pre-diabetes, 26 patients with diabetes and the system misclassified 10 patients with abnormal conditions. This result outweighed the result of [13]. Figure 4 shows the output display of weight with genetic algorithm during the training, while Figure 5 shows the clinical decision support system rule base interface for prostate cancer and diabetes. The rule base enhances the imminent diagnosis, prediction and detection of the diseases. To test the accuracy of the system, we plotted a graph of accuracy, classifier and classification time in the Fig 6, which shows that our machine learning algorithm with artificial neural network (ANN) and genetic algorithm (GA) performed better than the previous machine learning algorithms in clinical decision support system.

8. CONCLUSION

With the developed system hybrid system was employed by combining knowledge based and non-knowledge based decision support system. Genetic algorithm was successfully deployed to generate initial weights of the multilayer perceptron neural network to train the clinical markers thus the back-propagation had little work of adjusting the weights of the network, this helped us to create pattern so that any new introduced clinical markers could be classified appropriately, while the rule base algorithm successfully created rules vital to disease detection at least for diseases with related symptoms, hence, improved the accuracy and performance of clinical decision support system germane for efficient health care delivery. This paper has developed a successful technique for training data in neural network with genetic algorithm; in this case instead of employing GA alone in the whole training, it was used in weight initialization while back-propagation was used in adjusting the weights. With this method, faster convergence of the network is assured and global minima also minimized. This arrangement can be employed to train any kind of clinical data as far as it possesses clinical markers.

ACKNOWLEDGMENTS

We wish to acknowledge the efforts of Dr. O.O. Jegede, of the Pathological Department, Federal Medical Center Umuahia, Abia State, Nigeria and Dr. C.C. Chukwuegbo, of PathConsult Medical Diagnostic Center, Umuahia, Abia State, Nigeria.

We also acknowledge Professor E.O. Nwachukwu for his untiring efforts towards the accomplishment of this paper.

REFERENCES

- [1] Abbasi, M.M., & Kashiyarndi, S. 2010. Clinical decision support systems: A Discussion on different methodologies used in health care, *Journal of clinical bioinformatics*, 8(100), 200-210.
- [2] Artif, I.T. 2006. Decision support system for Lung diseases, *Journal of bioinformatics*, 16(40), 222-250.
- [3] Castareda, C., Nalley, K., Mannion, C., Bhattacharyya, P., Blake, P., Pecora, P.,...Suh, K. 2015. Clinical decision support system for improving diagnostic accuracy and achieving precision medicine, *Journal of clinical bioinformatic*. 5(4), 1-16.
- [4] Globocan report,. 2002. Growth of Diseases in Africa 2000 to 2015, July 2014 (P.4); Nigeria.
- [5] Gregory , F.C., Constantin, F.A., Richard, A., John, A., Bruce, G.B., Richard, C.,...Peter, S.(1996). An Evaluation of machine-learning methods for predicting pneumonia mortality. *Journal of artificial intelligence in medicine*, 9, 130-135.
- [6] Hsueh-Chun, L. 2012. Real-time clinical decision support system. *Journal of medical informatics*, 6(4), 112-120.
- [7] Kabari, L.(2009).Diagnosing skin disease using artificial neural network. *Journal of adaptive science and technology*, 19(34), 90-100.
- [8] Karthik, K., Binat, J.,Ramachanda, D. L., Mukund, J., and Abhay, C.(2012). Artificial neural network application in the diagnosis of disease conditions with liver ultra sound images. *Advance in ioinformatics*,20(14),120-156.
- [9] Koushat, K., and Abhishek, K. 2012. Artificial neural network for diagnosis of kidney stones diseases, *IJ Information technology and computer science*, MECS, 34(7), 400-408.
- [10] Ng, H., Muhammad, A., Meheza, I., Heama, N., & Eko, S.(2012). Prostate volume measurement using trans-abdominal ultra-sound scanning. *Advances in Environment, biotechnology and medicine*, 12(8), 1-6.
- [11] Oleg, Y., Atkov, M., Sveltana, G., Gorokhova, J., Alexandria (2014). Cronary heart disease diagnosis by artificial neural network including genetic polymorphism and clinical parameters, *Journal of Clinical Oncology*,34(16), 10-18.
- [12] Peshawa, K. 2009. Normalization process in a neural network, *Journal of institute of electrical and electronics engineering*, 4(10), 100.
- [13] Porter, MD., Christopher, R. and David, C. 2010, Combining neural network with transrectal ultra sound for the diagnosis of prostate cancer, *Oncology journal*, 11(8), 1-5.

- [14] Swati, S., & Kulothunghun, A. (2013). An Approach to improve brain disorder using machine learning techniques. IOSR Journal of Computer Engineering, 11(2), 78-81.
- [15] Tarik, E, Baris T, Anup, P, & Jurgen, F. (2014). Multiparametric MRI in prostate cancer. Biomed research international. 4(201), 3-115.
- [16] Taiwo, O.A. 2010. Types of machine learning algorithms, New advances in machine learning, 12(18), 22-34.

Authors

Odikwa Ndubuisi is pursuing his doctorate degree in computer science at University of Port Harcourt, Rivers State, Nigeria. He has M.Sc in information technology from National Open University of Nigeria and B.Tech in computer science and mathematics from Federal University of Technology, Owerri, Imo State, Nigeria. His research interest is on machine learning algorithms.



Ugwu Chidiebere is an associate professor of computer science in the University of Port Harcourt. He has doctorate degree in computer science from University of Port Harcourt, M.SC and B.SC from the same university and presently he is the ICT director of the university. His research interest is on natural language processing.



Inyama Hycinth is a professor of electrical and electronics engineering at Nnamdi Azikiwe University, Awka, Anambra State, Nigeria. His research interest is on Pattern recognition and image processing.