The Eurasia Proceedings of Science,
Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2018

Volume 2, Pages 345-353

ICRES 2018: International Conference on Research in Education and Science

A Design of Hybrid Expert System for Diagnosis of Breast Cancer and Liver Disorder

Aysegul ALAYBEYOGLU
Izmir Katip Celebi University

Naciye MULAYIM
Izmir Katip Celebi University

Abstract: It is certain that accurately and timely diagnosis of the diseases reduces the risk of morbidity and mortality of the disease. At that point, an expert system based on artificial intelligence techniques helps physicians or other healthcare professionals for diagnosis of it. In this study an expert system based on Firefly Algorithm is developed to diagnose both breast cancer and liver disorder. An experiential labour of the proposed system was managed using Indian Liver Patient Dataset and Breast Cancer Wisconsin (Original) Data Set received from UCI Machine Learning Repository sites. Standard statistical Metrics which are Negative Predictive Value, Positive Predictive Value, Specificity, Sensitivity, Precision, F_Measure and Accuracy are used to evaluate the performance of the proposed systems and simulation results show that the proposed system is 92% efficient in providing accurate diagnosis of Liver Disorder and 94.81% efficient in providing accurate diagnosis of Breast Cancer. C# programming language is used for the implementations of the system.

Keywords: Firefly algorithm, Expert system, Breast cancer, Liver disorder

Introduction

Expert System(ES)s are computer programs that are derived from AI [1] and is an intelligent interactional computer based decision tool that solves difficult and complex real life troubles and problems based upon information and notion of human experts in a particular fields [2]. Using friendly interfaces are important property of ESs and make them extremely interactional in nature and have access to complete, accurate and timely solving for these real life troubles and problems [3]. Medical practitioners can use computer devices to get help for systematizing, keeping and getting back of suitable medical knowledge to solve difficult states and these devices suggest them for favorable prognosis, diagnosis and therapeutic decision [4]. Computer technology can be used to cut down the ratio of mortality and minimize the waiting time to meet medical experts. Computer program developed by mimicking human intelligence could be used to help doctors in making timely and exact decisions considering patients' diagnosis [5].

Various intelligent systems have been developed to enhance more health service support and reduce cost of health expenses. These systems can behavior as an assistant which mimic human intelligence for patients and doctors [6- 8]. There are many studies which use artificial intelligent techniques in diagnosis of liver disorder and breast cancer. In [9], Naive Bayes classifier, Back propagation Neural Network algorithm, and Support Vector Machines Algorithms are used for liver disease diagnosis. In [10], support vector machine, a Bayesian and a k-nearest neighbour classifier is used for semi automatically segmented and quantified for diagnosis of liver disorders. In [11], a fuzzy expert system is developed for diagnosis of liver disorders. In [12], Support Vector Machine (SVM) and Naive Bayes are used to predict the liver disease. In [13], for prediction of liver disease, SVM and Random forest algorithms are compared. In [14], Artificial Neural Network (ANN) and Artificial Immune Algorithm (AIS) have been used to extract rules which have been trained for classification. Thus a set of rules is obtained for liver disorders. In [15], Breast cancer tumour is classified by using firefly algorithm to improve the parameters of local linear wavelet neural network. Support Vector Machine (SVM) is

used to detect breast cancer for classifying in [16]. Breast cancer detection is achieved by using SVM combined with the future selection technic in [17]. Lastly, Breast cancer datasets are tested with SVM algorithms and the results are compared with the results of other machine learning techniques in [18].

In this study, an expert system based on Firefly Algorithm is developed to diagnose breast cancer and liver disorder diseases. ILPD (Indian Liver Patient Dataset) has 583 patients' records and Breast Cancer Wisconsin (Original) Data Set has 699 patients' records, which are in UCI Machine Learning Repository site, are used for proposed system. Remainder of this paper is organized as follows: Section 2 presents the architecture of the proposed system, Section 3 presents the results of experimental study and the evaluations and lastly Section 4 presents the conclusion and recommendations.

Architecture and Interfaces of the Proposed System

The architecture of the proposed system for the diagnosis of liver disorders and breast cancer is shown in Figure 1. When the user enters knowledge of laboratory tests, the system inferences results by using knowledge base and inference engine. Based on these, the system decides whether the patient is sick or not.

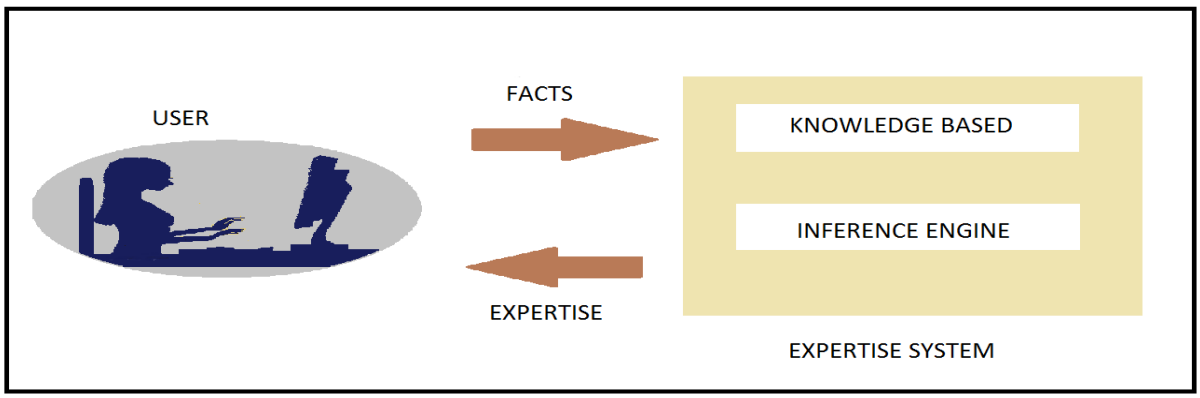
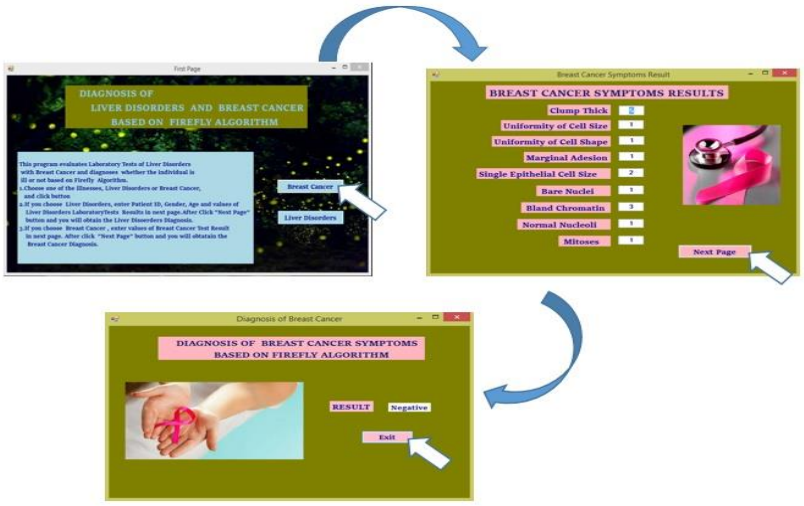


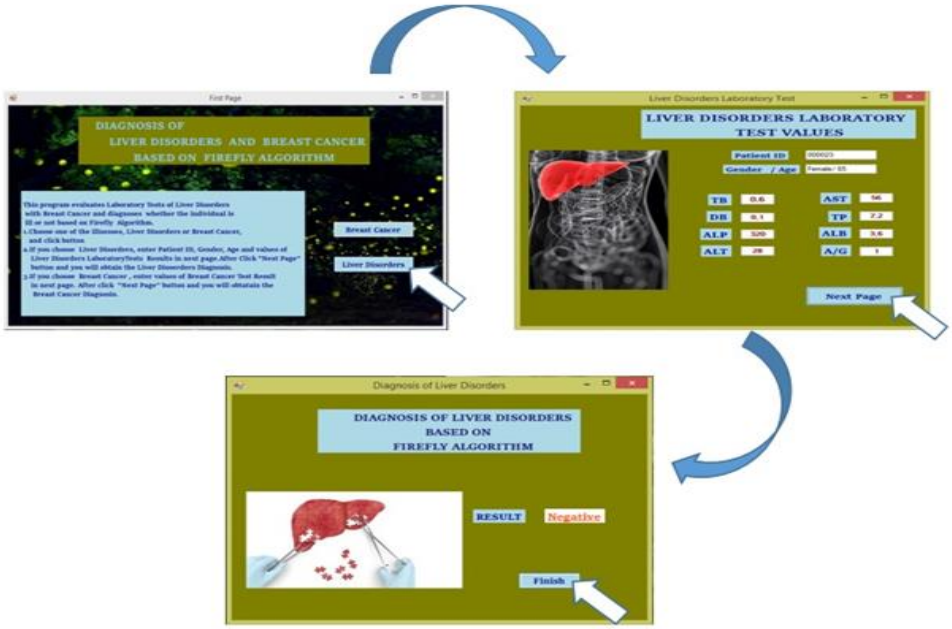
Figure 1. The architecture of the proposed system for the diagnosis of liver disorder and breast cancer

Interfaces of the Firefly algorithm based Hybrid Expert System

Designed interfaces are shown in Figure 2. First page gives information about usage of the system and presents choices for illnesses we can use in the system. If the user clicks Breast Cancer button, second page is opened and Breast Cancer Symptoms Results are entered by the user. When the Next Page button is clicked, Diagnosis of the Breast Cancer Symptoms Based on Firefly Algorithm is obtained as shown in Figure 2.a. If the user clicks Liver Disorders button, second page is opened and Patient ID, Gender, Age and Liver Disorders Laboratory Tests Results are entered as shown in Figure 2.b. When the Next Page button is clicked, Diagnosis of Liver Disorders Tests Results based on Firefly Algorithm is obtained.



(a) Firefly algorithm based expert system for Breast Cancer



(b) Firefly algorithm based expert system for Liver Disorders
Figure 2. Interfaces of the proposed system

Simulation Results and Evaluations

In the proposed expert system, firefly algorithm is used for diagnosis of liver disorder and breast cancer detection. C# programming language is used for implementations.
Some definitions of terminology and derivations from a confusion matrix and formulas are given below:

- **True positive (TP):** Sick people are accurately defined as sick,
- **False positive (FP):** Healthy people are inaccurately defined as sick,
- **True negative (TN):** Healthy people are accurately defined as healthy,
- **False negative (FN):** Sick people are inaccurately defined as healthy [19]

Table 1. Formulas and Parameters of CBR

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}$$

$$\text{Negative predictive Value} = \frac{TN}{TN + FN}$$

$$\text{Positive Predictive Value} = \frac{TP}{TP + FP}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{F\_Measure} = 2\,x\frac{precision\ x\ sensitivity}{precision + \ sensitivity}$$

Firefly Algorithm based Expert System for Liver Disorders

ILPD (Indian Liver Patient Dataset) which is in UCI Machine Learning Repository site has 583 patient records (2010). This data set comprises 416 liver patient registers and 167 non liver patient registers. In dataset, three number of patients have missing values and fourteen number of patients have the same data values. The data set is taken from north east of Andhra Pradesh, India. Chooser divides into two groups: liver patient or not. This

data set involves 441 male patient records and 142 female patient records. Any patient whose age surpasses 89 is written as being of age "90" [20]. There are 10 attributes that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos in dataset. In this study, last 8 laboratory tests results are used for diagnosis [20]. Table 2 shows attributes in ILPD Dataset and normal values of attributes.

Table 2. Attribute in ILPD Dataset and information of normal values of attributes [21].

Variable	**Information (Normal Value)**
Age	Age of the patient
Gender	Gender of the patient
TB	Total Bilirubin (0.22-1.0 mg/dl)
DB	Direct Bilirubin (0.0-0.2mg/dl)
Alkphos	Alkaline Phosphotase (110-310U/L)
SGPT	Alamine Aminotransferase (5-45U/L)
SGOT	Aspartate Aminotransferase (5-40U/L)
TP	Total Proteins (5.5-8gm/dl)
ALB	Albumin (3.5-5gm/dl)
A/G Ratio (LFT)	Albumin and Globulin Ratio (>=1)
SELECTOR	field used to split the data into two sets

In this study, 300 numbers of patient data are used to evaluate the system and to equilibrate the numbers of liver and non-liver patients. While number of positive liver disorder patient is 204, negative liver patient is 104. In Table 3, Liver Disorder Laboratory Test results can be seen according to the system based on firefly algorithm. As it can be obtained from Table 3, the system predicts accurately liver disorder test results of 188 number of patients which have liver disorders; it predicts inaccurately liver disorders test results of 16 number of patients. It can also be seen from the table that the system predicts accurately liver disorders test results of 88 number of patients which do not have liver disorders, it predicts inaccurately liver disorders test results of 8 number of them.

Table 3. Liver Disorder Laboratory Test results according to the system based on firefly algorithm

Liver Disorders Laboratory Test Results	**Condition Positive**	**Condition Negative**
Test Results Positive	**True Positive (TP) = 188**	**False Positive (FP) = 8**
Test Results Negative	**False Negative (FN) = 16**	**True Negative (TN) = 88**

As it is shown in Table 4, Negative predictive value is achieved 83.01%, Positive predictive value is achieved 95.91%, Sensitivity is achieved 92%, Specificity is achieved 84.16% and Accuracy (ACC) is achieved 92%. Positive predictive value 95.1% is significant because it gives a high confidence that its positive result is true. If sum of the sensitivity and specificity is higher than 170, it can be accepted that the system is useful and helpful as clinical investigation. In this study, the system determines condition as follows:

$$\text{Sensitivity} + \text{Specificity} = 92 + 84.16 = 176.16 > 170$$

Table 4. Simulation results and evaluations of liver disorder expert system

Negative Predictive Value	**Positive Predictive Value**	**Specificity**	**Sensitivity**	**Precision**	**F_Measure**	**Accuracy**
83.01	**95.91**	**84.61**	**92**	**95.91**	**93.91**	**92**

Figure 3 compares the values of real negative liver disorder and proposed system's negative liver disorder values in dataset. As it can be seen from the figure, the system achieves highly accurate predictions of Non-Liver Patients' Test Results values.

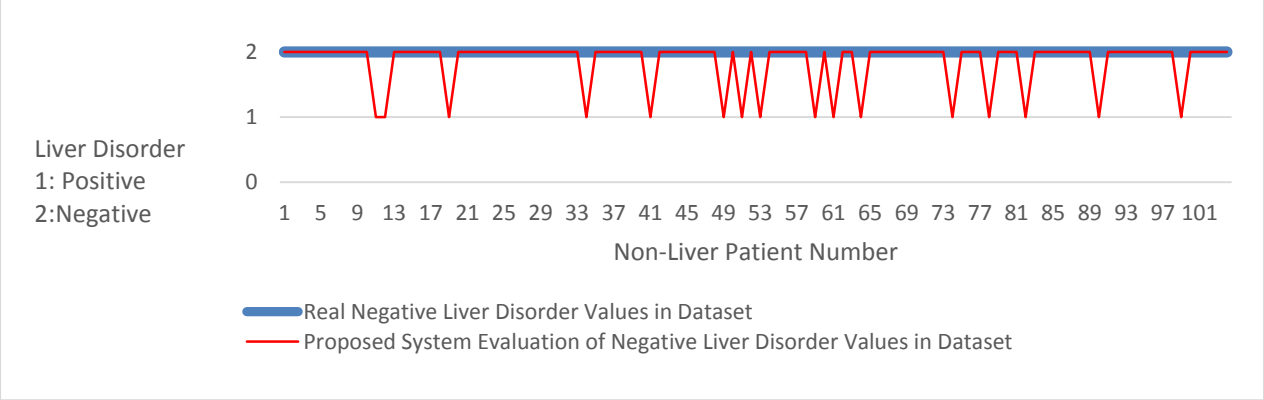


Figure 3. Comparison of real negative liver disorder values in dataset and proposed system's negative liver disorders values in dataset

Figure 4 compares real positive liver disorder values in dataset and proposed system's positive liver disorders values. It can be obtained from the simulation results that the system achieves highly accurate predictions of Liver Patients' Test Results values.

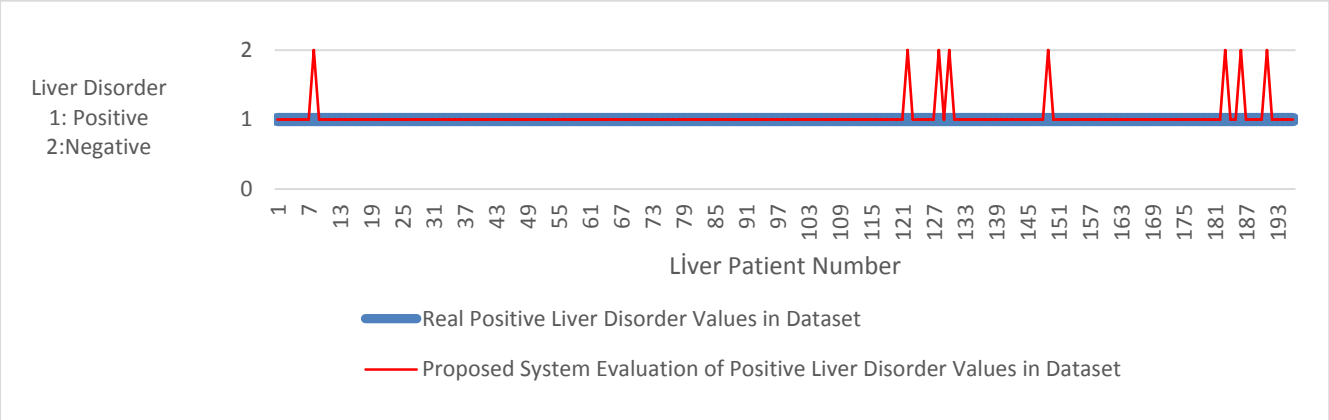


Figure 4. Comparison of real positive liver disorders value in dataset and proposed system's positive liver disorders values

Firefly Algorithm based Expert System for Breast Cancer

Breast Cancer Wisconsin (Original) Data Set has 699 patients' records which are used for evaluations of the proposed system. Specimens are obtained periodically. Therefore, the database represents this chronological grouping of the data set. Information of the dataset appears as below:

Group 1: 367 instances (January 1989)
Group 2: 70 instances (October 1989)
Group 3: 31 instances (February 1990)
Group 4: 17 instances (April 1990)
Group 5: 48 instances (August 1990)
Group 6: 49 instances (Updated January 1991)
Group 7: 31 instances (June 1991)
Group 8: 86 instances (November 1991)

Total: 699 points (as of the donated database on 15 July 1992)

This data set comprises 241 number of malignant registers and 458 non benign registers. This breast cancer database is obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg [22].

There are 10 attributes in the dataset:

	Attribute	Domain
1.	Sample code number	id number
2.	Clump Thickness	1 - 10
3.	Uniformity of Cell Size	1 - 10
4.	Uniformity of Cell Shape	1 - 10
5.	Marginal Adhesion	1 - 10
6.	Single Epithelial Cell Size	1 - 10
7.	Bare Nuclei	1 - 10
8.	Bland Chromatin	1 - 10
9.	Normal Nucleoli	1 - 10
10.	Mitoses	1 - 10
11.	Class:	(2 for benign, 4 for malignant) [22]

In this study, patient Breast Cancer dataset by ignoring repetitive data is used for evaluating proposed system to equilibrate number of breast cancer and non-breast cancer. Number of Positive Breast cancer patient is 238 and negative liver patient is 225. In Table 5, Breast Cancer Laboratory Test results are given according to the system based on firefly algorithm. As it can be seen from the table, while the system predicts accurately breast cancer test results of 226 patients which have liver disorders, it predicts inaccurately liver disorders test results of 12 of them. It can also be obtained from the table that, the system predicts accurately liver disorders test results of 213 patients which do not have breast cancer, it predicts inaccurately liver disorders test results of 12 of them.

Table 5. Proposed system's breast cancer laboratory test results

Breast Cancer Results	**Condition Positive**	**Condition Negative**
Test Results Positive	**True positive (TP) = 226**	**False positive (FP) = 12**
Test Results Negative	**False negative (FN) = 12**	**True negative (TN) = 213**

As it is shown in Table 6, Negative predictive value is achieved 83.01%, Positive predictive value is achieved 95.91%, Sensitivity is achieved 92%, Specificity is achieved 84.16% and Accuracy (ACC) is achieved 92%. Positive predictive value 95.1% can be accepted as significant because it gives a high confidence that its positive result is true. If sum of the sensitivity and specificity is higher than 170, it is said that the system is useful and helpful as clinical investigation. In this study, the system determines condition follows:

$$\text{Sensitivity} + \text{Specificity} = 92 + 84.16 = 176.16 > 170$$

Table 6. Simulation results and evaluations of breast cancer expert system

Negative Predictive Value	**Positive Predictive Value**	**Specificity**	**Sensitivity**	**Precision**	**F_Measure**	**Accuracy**
94.67	**94.95**	**94.67**	**94.95**	**94.95**	**94.95**	**94.81**

Figure 5 compares values of real negative breast cancer and the proposed system negative breast cancer values in dataset. It can be obtained from the figure that the system achieves highly accurate predictions of Non-Breast Cancer Patients' Test Results values.

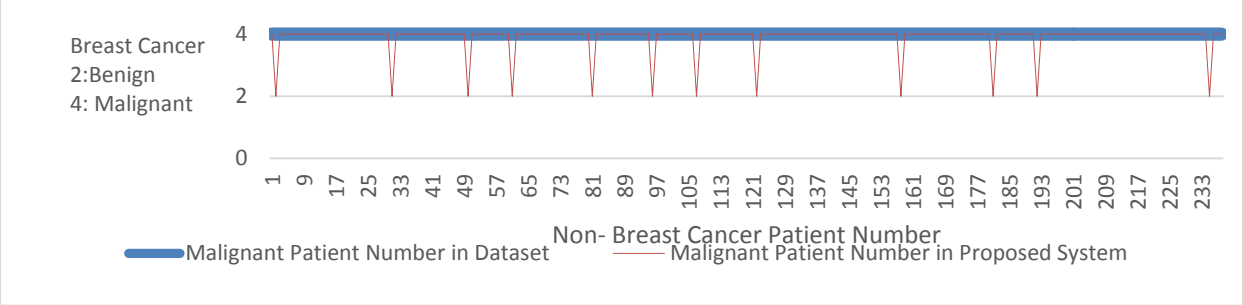


Figure 5. Comparison of real negative breast cancer values in dataset and proposed system's negative breast cancer values in dataset

Figure 6 compares the values of real positive breast cancer and the proposed system's positive breast cancer values. It can be obtained from the simulation results that the system achieves highly accurate predictions of Breast Cancer Patients.

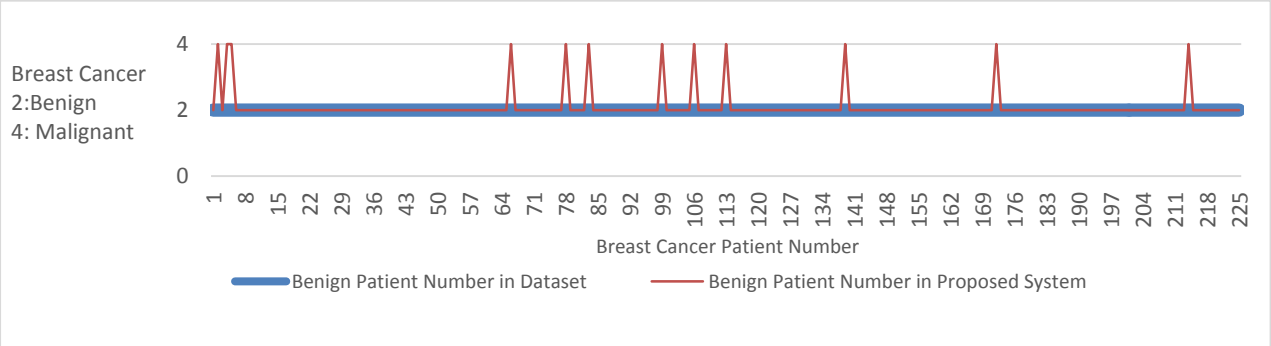


Figure 6. Comparison of the values of real positive breast cancer and proposed system's positive breast cancer.

Conclusion

In this study, an expert system is developed for the diagnosis of Breast cancer and Liver disorders. The expert system is driven by Firefly Algorithm. The system can be used by practitioners and touched health care personnel accurately and timely diagnosis of Liver Disorders and Breast Cancer irrespective of the position of the medical. The system can be useful for the hospitals which have insufficient number of doctors and medical personnel. It also can help to reduce a great number of patients who are waiting for to be checked by doctors [23]. The system which is based on firefly algorithm can be used to cut down the ratio of mortality and minimize the waiting time to meet medical experts. Computer program which is developed by mimicking human intelligence can be used to help doctors in making timely and exact decisions considering patients' diagnosis [5].

References

Chair, E. F., Friedland ,P. E., Johnson, B. B., Nii H. P., Schorr, H., Shrobe, H.& Engelmore R. S., (May 1993), Knowledge-Based System in Japan, Expert Systems and Artificial Intelligence, JTEC(Japanese Technology Evaluation Center) Panel on.

Samuel, O.W., Omisore, M.O. & Ojokoh, B.A., 2013 Elsevier Ltd., A web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever.

Durkin, J. J. (1994). Expert system design and development. New Jersey: Prentice-Hall.

Szolovits, P., Patil, R. S., & Schwartz, W. B. (1988). Artificial intelligence in medical diagnosis. Journal of Internal Medicine, 108, 80–87.

Ishak, W.H.W & Siraj ,F., (2002) . Artificial Intelligence in Medical Application: An Exploration. Health Informatics Europe Journal [Online].

Ishak, W.H.W & *Yamin*, F.M., (2001). Artificial Intelligence in Decision-Making. Presented at National Conference on Management Science: New Paradigms for the Knowledge Economy (19-20 June), Universiti Putra Malaysia, Serdang, Selangor.

Alexopoulos, E., Dounias, G. D., & Vemmos, K. (1999). Medical diagnosis of stroke using inductive machine learning. In Machine learning and applications (pp. 20–23). Chania, Greece

Bourlas, P., Giakoumakis, E., & Papakonstantinou, G. (1999). A knowledge acquisition and management system for ECG diagnosis. In Machine learning and applications: Machine learning in medical applications (pp. 27–29). Chania, Greece.

Ramana, B. V., Babu, M.S.P., & Venkateswarlu, N. B. (2011)A Critical Study of Selected Classification Algorithms for Liver Disease International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May.

Ribeiro, R., Marinho, R., Velosa, J., Ramalho, F., & Sanches, J. M.,(2011) Diffuse liver disease classification from ultrasound surface characterization, clinical and laboratorial data, http://users.isr.ist.utl.pt/~jmrs/research/publications/myPapers/2011/2011_ibPRIA_RicardoRibeiro.pdf (Accessed 2.2. 2016)

Neshat,M., Yaghobi, M., Naghibi, M.B., & Esmaelzadeh, A., (2008) Fuzzy Expert System Design for Diagnosis of liver disorders, International Symposium on Knowledge Acquisition and Modeling, 252 – 256, 978-0-7695-3488-6,IEEE.

Vijayarani, S.,& Dhayanand, Mr.S., Liver Disease Prediction using SVM and Naïve Bayes Algorithms (2015)International Journal of Science, Engineering and Technology Research (IJSETR) Volume 4, Issue 4, April .

Gulia, A., Vohra, R., &Rani,P.,(2014) Liver Patient Classification Using Intelligent Techniques, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5110-5115

Kahramanli, H., & Allahverdi,N.,(2009) , Mining Classification Rules for Liver Disorders. International Journal Of Mathematics And Computers In Simulation 1, 9-19

Akay M.F.,(2009) Support vector machines combined with feature selection for breast cancer diagnosis, Elsevier Ltd. All rights reserved, Expert Systems with Applications 36 3240–3247

Medical News Today : Your source for health news since 2003. http://www.medicalnewstoday.com/articles/37136.php?page=2 Accessed 01.02.2016.

Breast Cancer Care(2001) https://www.breastcancercare.org.uk/information-support/have-i-got-breast-cancer/what-breast-cancer

NHS Choices, http://www.nhs.uk/Conditions/Cancer-of-the-breast-female/Pages/Introduction.aspx, Accessed 02.02.2016.

Sensitivity and specificity From Wikipedia, the free encyclopedia, https://en.wikipedia.org/wiki/Sensitivity_and_specificity, Accessed 10.02.2016

The UC Irvine Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29

Accessed 12.10.2015.

M.S.Prasasd Babu & Somesh Katta , (2015), Artificial Immune Recognition Systems in Medical Diagnosis , Software Engineering and Service Science (ICSESS), 6th IEEE International Conference on, 978-1-4799-8352-0,IEEE.

The UC Irvine Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29

Accessed 12.10.2015.

Rong-Ho Lin. An intelligent model for liver disease diagnosis. Artificial Intelligence in Medicine 2009;

Author Information

Aysegul Alaybeyoglu
Izmir Katip Celebi University
Computer Engineering Department
Contact e-mail: *aysegul.alaybeyoglu@ikc.edu.tr*

Naciye Mulayim
Izmir Katip Celebi University
Biomedical Engineering Department

Appendix

Table 1.A Some of the training data set of 70 number of liver disorder patients

AGE	gender	TB	DB	TP	ALB	A/G	SGPT	SGOT	ALKPHOS	RESULT
52	Male	0.6	0.1	178	26	27	6.5	3.6	1.2	-1
66	Male	0.6	0.2	100	17	148	5	3.3	1.9	-1
55	Male	0.8	0.2	482	112	99	5.7	2.6	0.8	1
37	Male	0.8	0.2	147	27	46	5	2.5	1	1
61	Male	0.8	0.1	282	85	231	8.5	4.3	1	1
61	Male	0.8	0.2	163	18	19	6.3	2.8	0.8	-1
61	Male	0.8	0.2	192	28	35	6.9	3.4	0.9	-1
. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .

Table 2.A Some of the training data set of 60 number of breast cancer patients

Sample code number	Clump Thickness	Uniformity of Cell Size	Uniformity of Cell Shape	Marginal Adhesion	Single Epithelial Cell Size	Bare Nuclei	Bland Chromatin	Normal Nucleoli	Mitoses	Class:
1018099	1	1	1	1	2	10	3	1	1	-1
1223543	1	2	1	3	2	1	1	2	1	-1
1223793	6	10	7	7	6	4	8	10	2	1
654546	1	1	1	1	2	1	1	1	8	-1
654546	1	1	1	3	2	1	1	1	1	-1
1091262	2	5	3	3	6	7	7	5	1	1
1167439	2	3	4	4	2	5	2	5	1	1
850831	2	7	10	10	7	10	4	9	4	1
1105257	3	7	7	4	4	9	4	8	1	1
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .