iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VII **Month of publication:** July 2017

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

Detection and Diagnosis of Pancreatic Cancer using Neuro-Fuzzy Techniques

A. S. Olufemi[1], G. Oluleye[2] , O. M. Oyeleke[3] , E. E. Theophilus[4]
[1]Department of Chemical/Petroleum Engineering, Niger Delta University, Wilberforce Island, Bayelsa State, Nigeria.
[2,4]Department of Electrical/Electronic Engineering, [3]Department of Computer Engineering,
Ogun State Institute of Technology Igbesa, Ogun State, Nigeria.

Abstract: Pancreatic Cancer is malignant growth or tumor, which forms due to an uncontrolled division of cells in the pancreas, hence, may even lead to death if not detected at the early stage. If it is Pancreas then the disease is termed as Pancreatic Cancer. This paper presents an Artificial Neural Network model to diagnose pancreatic cancer based on a set of symptoms. The ANN model was created after analyzing the actual procedure of disease diagnosis by the doctor, then the back propagation algorithm. An approach to detect various stages of cancer affected in the pancreas is presented in the paper. The result gives more than 87 % of accuracy which suggests the advantage of using the ANN model instead of manual disease diagnosis.
Keywords: Neural Network, Diagnosis, Pancreatic Cancer, Detection, Back Propagation

I. INTRODUCTION

The pancreas is a gland of 6-inch long spongy organ located behind the stomach and in front of your spine [1, 2]. The pancreas contains exocrine and endocrine glands that create pancreatic juices, hormones, and insulin. Pancreatic juices, or enzymes, made by the exocrine glands are released into the intestines by way of a series of ducts in order to help digest fat, proteins, and carbohydrates [1]. The endocrine cells are arranged in small clusters called islets of Langerhans, which release insulin and glucagon into the bloodstream. These two hormones manage levels of sugar in the blood. When they are not working properly, the result is often diabetes. Pancreatic cancer usually begins in the cells that produce the juices. The abnormal pancreas tissues continue dividing and form lumps or masses of tissue called tumor. Tumors then interfere with the main functions of the pancreas.

The number of cancer related deaths worldwide is increasing day by day and researches shows that pancreatic cancer is the eighth most common cause of cancer-related deaths worldwide and fourth worldwide. Pancreatic cancer originates of the malignant type neoplasm from transformed cells arising in tissues from the pancreas. Pancreatic cancer is hard to catch early. It doesn't cause symptoms right away. When you do get symptoms, they are often vague or you may not notice them. They include yellowing of the skin and eyes, pain in the abdomen and back, weight loss and fatigue. Also, because the pancreas is hidden behind other organs, health care providers cannot see or feel the tumors during routine exams. Doctors use a physical exam, blood tests, imaging tests, and a biopsy to diagnose it.

Because it is often found late and it spreads quickly, pancreatic cancer can be hard to treat. Possible treatments include surgery, radiation, chemotherapy, and targeted therapy. Targeted therapy uses substances that attack cancer cells without harming normal cells.

Detection of pancreatic cancer in its early stage is the key of its cure. To do this, various techniques such as X-ray (Chest Radiograph), CT, PET, MRI scan, Sputum Cytology etc. are available. However, most of these techniques are expensive and time consuming. Along with these, the new CAD (Computer - Aided - Diagnosis) techniques is increasing. These techniques help to detect the occurrence of pancreatic tumor in its early stage [3].

An Artificial Neural Network (ANN) is a mathematical representation of the human neural architecture, reflecting its “learning” and “generalization” abilities. For this reason, ANNs belong to the field of artificial intelligence. ANNs are widely applied in research because they can model highly non-linear systems in which the relationship among the variables is unknown or very complex. A review of various classes of neural networks can be found in [4, 5].

In various clinical situations which are considered difficult, ANN has been used successfully as a non-linear pattern recognition technique in making diagnostic and prognostic decisions or predictions, it is believed that it will be more widely used in biomedical systems in the next few years [6, 7].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

Nowadays, studies show that many medical diagnosis problems are being solved successfully using NN techniques. Artificial Neural Networks are applied to medicine mainly for the task which is based on the measured features to assign the patient to one of a small set of classes [8, 9]. Most applications of artificial neural networks in medicine are classification problems; that is, the task is on the basis of the measured features to assign the patient to one of a small set of classes. Several research groups are working worldwide on the applicability of neural networks in medical diagnosis [9, 10], modeling the kinetics of drug release [11], and even species determination [12-14].

In 2004, a NN based model is proposed by Kamruzzaman et al. [15], for the diagnosis of heart diseases. In 2008, a Genetic Algorithm (GA) based technique for classifying tumor mass in the breast and to identify breast cancer has been introduced [6]. A new method for Predicting Blood Cancer and Disorder was later developed by Payandeh et al. [16]. Sumathi and Santhakumaran [17], used Artificial Neural Network to pre-diagnosis of Hypertension using Back-Propagation training algorithm. Artificial Neural Network model to diagnose skin diseases was carried out by Backpo and Kabari [18]. Similar ANN models are also developed for breast cancer detection [19], Kidney stone diseases [20], etc were researches done using ANN.

Accuracy as well as the objectivity of medical diagnosis has been increased using neural networks. In ANN, the processing element is called as neurons. An artificial Neural Network is a network of such interconnected neurons operating in parallel.

The aim of this paper is to present an approach to detect or diagnosis the various stages of pancreatic cancer disease affecting patients using artificial neural network and fuzzy logic based on a set of symptoms by some patients' condition, then the NN can be used to make an accurate prognosis of each individual as stated by [21, 22]. There is no proper automated tool use for the purpose of Pancreatic Cancer Detection diagnostic system. Hence, a self-learning intelligent system will be developed using ANN which can be used to overcome the uncertainties in the diagnosis of pancreatic cancer. Some symptoms are taken from the patient's previous medical records as well as from the doctor, and by using these data the neural network model is trained to detect the presence/absence of pancreatic cancer in that patient. To diagnose the pancreatic cancer properly using this intelligent model fuzzified symptoms values are applied.

II. METHODOLOGY

A. Mathematical Background

A neural network is formed by a series of "neurons" (or "nodes") that are organized in layers. Each neuron in a layer is connected with each neuron in the next layer through a weighted connection. The value of the weight, w_{ij} , indicates the strength of the connection between the $i-^{th}$ neuron in a layer and the $j-^{th}$ neuron in the next one.

The structure of a neural network is formed by an "input" layer, one or more "hidden" layers, and the "output" layer. The number of neurons in a layer and the number of layers depends strongly on the complexity of the system studied. Therefore, the optimal network architecture must be determined. The scheme of the ANN architecture is given in Figure 1.

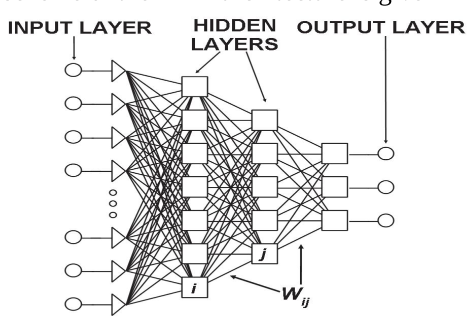


Figure 1. The neural network with two hidden layers.

The w_{ij} is the weight of the connection between the i-th and the j-th node. The neurons in the input layer receive the data and transfer them to neurons in the first hidden layer through the weighted links. Here, the data are mathematically processed and the result is transferred to the neurons in the next layer. Ultimately, the neurons in the last layer provide the network's output. The j-th neuron in a hidden layer processes the incoming data (x_i) by:

1) Calculating the weighted sum and adding a "bias" term (θ_j) according to Equation 1.

$$net_j = \sum_{i=1}^{m} x_i \times w_{ij} + \theta_j \qquad (j = 1, 2, ...n) \tag{1}$$

2) Transforming the net_j through a suitable mathematical "transfer function", and
3) Transferring the result to neurons in the next layer: Various transfer functions are available in the work of Zupan and Gasteiger [5]; however, for the purpose of this research, the sigmoid one is used:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

The mathematical process through which the network achieves *"training"* was ignored by the final user. In this way, the network can be viewed as a *"black box"* that receives a vector with *m* inputs and provides a vector with *n* outputs (Fig. 2), the ANN architecture is hidden and it is indicated as a black box.
The network *"learns"* from a series of *"examples"* that form the *"training database"* (Fig. 3), the element $data_{k,i}$ refers to the *i*-th medical data (symptom, laboratory data, age, etc.) of the *k*-th patient. Each row refers to a different patient labeled with a numerical code.
An *"example"* is formed by a vector $X_{im} = (x_{i1}, x_{i2}, ..., x_{im})$ of inputs and a vector $Y_{in} = (y_{i1}, y_{i2}, ..., y_{in})$ of outputs. The objective of the training process is to approximate the function f between the vectors X_{im} and the Y_{in} :

$$Y_{i,n} = f\left(X_{i,m}\right) \tag{3}$$

This is achieved by changing iteratively the values of the connection weights (w_{ij}) according to a suitable mathematical rule called the *training algorithm*.
The values of the weights are changed by using the steepest descent method to minimize a suitable function used as the training stopping criteria. One of the functions most commonly used is the sum-of squared residuals given by Equation (4):

$$E = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(y_{ij} - y_{ij}^{*}\right)^2 \tag{4}$$

Where y_{ij} and y_{ij}^{*} are the actual and network's *j*-th output corresponding to the *i*-th input vector, respectively. The current weight change on a given layer is given by Equation (5):

$$\Delta w_{ij} = -\eta \frac{dE}{dw_{ij}} \tag{5}$$

Where η is a positive constant called the *learning rate*. To achieve faster learning and avoid local minima, an additional term is used and Eq. 5 becomes:

$$\Delta w_{ij}^{k} = -\eta \frac{dE}{dw_{ij}} + \mu \Delta w_{ij}^{k-1} \tag{6}$$

Where μ is the "momentum" term and w_{ij}^{k-1} is the change of the weight w_{ij} from the (*k-1*)-th learning cycle. The learning rate controls the weight update rate according to the new weight change and the momentum acts as a stabilizer, being aware of the previous weight change.

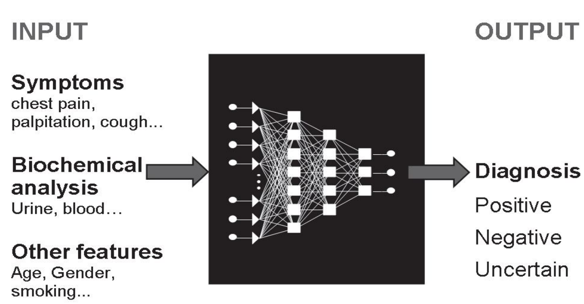


Figure 2. Details of input and output items concerning the ANNs-based diagnosis

Patient code	MEDICAL DATA					DIAGNOSIS
1	$data_{1,1}$	…	$data_{1,i}$	…	$data_{1,m}$	POSITIVE
2	$data_{2,1}$	…	$data_{2,i}$	…	$data_{2,m}$	POSITIVE
3	$data_{3,1}$	…	$data_{3,i}$	…	$data_{3,m}$	POSITIVE
…	…	…	…	…	…	…
k	$data_{k,1}$	…	$data_{k,i}$	…	$data_{k,m}$	NEGATIVE
k+1	$data_{k+1,1}$	…	$data_{k+1,i}$	…	$data_{k+1,m}$	NEGATIVE
…	…	…	…	…	…	…
n	$data_{n,1}$	…	$data_{n,i}$	…	$data_{n,m}$	NEGATIVE

Figure 3. Training of the ANN database structure.

Once a network has been structured for a particular application, that network is ready to be trained. To start this process the initial weights are chosen randomly. Then the training or learning begins. The ANN has been trained by exposing it to sets of existing data (based on the follow up history of various cancer patients) where the outcome is known. Multi-layer networks use a variety of learning techniques; the most predominant is back–propagation algorithm. Information flows from the direction of the input layer towards the output layer. The hidden neurons are able to learn the pattern in data during the training phase and mapping the relationship between input and output pairs. Each neuron in the hidden layer uses a transfer function to process data it receives from input layer and then transfers the processed information to the output neurons for further processing using a transfer function in each neuron.

The number of iterations of the training algorithm and the convergence time varies due to the weight initialization. After repeating this process for a sufficiently large number of training cycles, the network converges to some state where the error of the calculations is small.

B. Data Set

The data obtained from the Niger Delta University Teaching Hospital, (NDUTH), Okolobiri, Bayelsa State were used for this study. Initially 11 pancreatic cancer patients' data has been collected from the hospital and trained with the neural networks. The present algorithm is fast, taking only few seconds of execution time. After training and verification, the network can be used in practice to predict the diagnosis. Finally, the predicted diagnosis is evaluated by a clinical specialist. The disease name in their outcomes and the corresponding label assigned for each of them are mentioned in Table 1. On the basis of fuzzy set, this study defines each symptom by its membership value, $\mu_A(x)$.

Nominal variables are used to represent the input values in the nodes of the input layer. Nominal variables may be two-state or many-state. A two-state nominal variables is easily represented by transformation into a numeric value. For example, Male = 0, Female = 1. Many-state nominal variables are more difficult to handle. They can be represented using an encoding known as one-of-N encoding. In this number of numeric variables represent the single nominal variable. The number of numeric variables equals the number of possible values; one of the N variables is set and the others are cleared. E.g. age < 35 = {1, 0, 0}, age > = 35 < = 55 = {0, 1, 0}, age> 55 = {0, 0, 1}. Similarly the other input values are represented. Neural networks has facilities to convert both two-state and many-state nominal variables.

Data set used for the diagnosis of the pancreatic cancer is shown in Table 2. This data set consist of 11 possible symptoms and 3 categories of possible pancreatic disease outcomes. The outcome is purely dependent on the significance of symptoms for a particular patient. The total dataset consists of measured features of 120 patients in which 90 samples were used for training the network and the remaining 30 samples, for testing purpose. The set of symptoms represented as S = {Jaundice, Loss of Appetite, Weight Loss, Pain in Upper Abdomen, Irritability, Gall Bladder Enlargement, Swelling Lymph, Diabetes Mellitus, Deep Venous Thrombosis, Acholic Stool & Steatorrhea and Fatty Tissue Abnormalities}.

The set of possible outcomes of diagnosis represented as D = {Disease Detected, Disease might be Detected and Disease not Detected}.

The basic block diagram that explains about the operational procedure is shown in Figure 4. Building of the database and "learning" represents the left half (green) and its application for the diagnosis is the right part (blue).

TABLE 1
Assigned Labels And Outcome

Disease	Assigned Label
PC Detected	1
PC Might be detected	2
PC Not detected	3

TABLE 2
Symptoms Significance And Result

Weight loss	Jaundice	Irritability	GB enlargement	DVT	Swelling Lymph	DM	Loss of appetite	Pain in upper abdomen	AS & Steatorrhea	Faulty tissue Abnormalities	Result
0.60	0.45	0.30	0.10	0.55	0.35	0.15	0.26	0.80	0.72	0.18	1
0.62	0.18	0.17	0.78	0.14	0.82	0.50	0.00	0.59	0.36	0.47	2
0.32	0.73	0.25	0.55	0.61	0.13	0.20	0.69	0.33	0.49	0.86	2
0.28	0.24	0.39	0.36	0.23	0.70	0.55	0.63	0.08	0.17	0.63	3
0.43	0.59	0.73	0.29	0.70	0.37	0.13	0.68	0.75	0.55	0.43	1
0.56	0.44	0.57	0.63	0.34	0.19	0.41	0.53	0.69	0.72	0.38	1
0.72	0.52	0.60	0.19	0.42	0.40	0.21	0.63	0.30	0.39	0.52	3
0.63	0.13	0.11	0.47	0.69	0.23	051	0.12	0.24	0.10	0.63	2
0.33	0.63	0.49	0.72	0.77	0.62	0.24	0.38	0.21	0.43	0.40	1
0.78	0.24	0.54	0.41	0.64	0.78	0.43	0.55	0.30	0.18	0.23	1

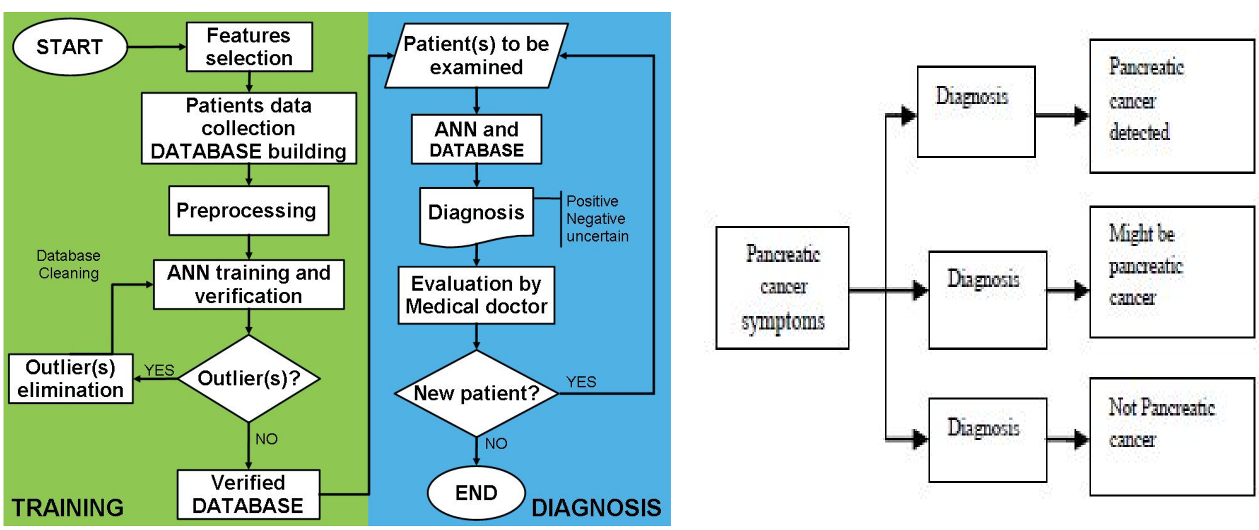


Figure 4. The block diagrams of operational steps in ANNs-based pancreatic cancer diagnosis

III. RESULTS AND DISCUSSION

The Neural Network toolbox from Matlab R2016b is used for performance estimation of the proposed networks where a three (3) layer, eleven 11 number feed forward network of inputs, sigmoid hidden neurons and linear output neurons is suggested. This study used the approach of Levenberg-Marquardt back propagation algorithm to train the network where training automatically stops when generalization stops improving, as indicated by an increase in the mean square error (MSE) of the validation samples. The proposed neural network is shown in Figure 5.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

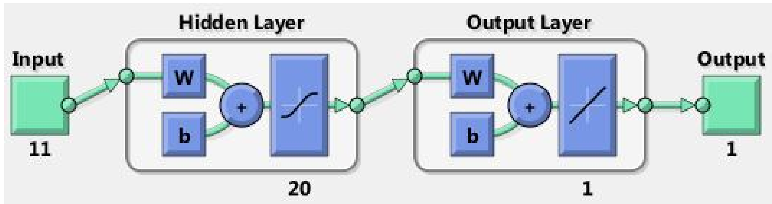


Figure 5. Proposed Neural Network

This study has used a membership based fuzzification scheme on our dataset to convert it to a fuzzified set of symptoms. A linear membership function was selected for each symptom again after an interview with physicians. Normally three to five linguistic variables were assigned to each symptom, then the classification tests were repeated.

The experimental results of implementing the new ANN methodology to distinguish between Pancreatic Cancer affected and non-affected patients based upon specified symptoms represents good capabilities of the network to learn the training patterns corresponding to symptoms of the patients. The results are found to be better using back propagation algorithm. It gives more than 87% of accuracy. The experimental setup is shown in Figure 6.

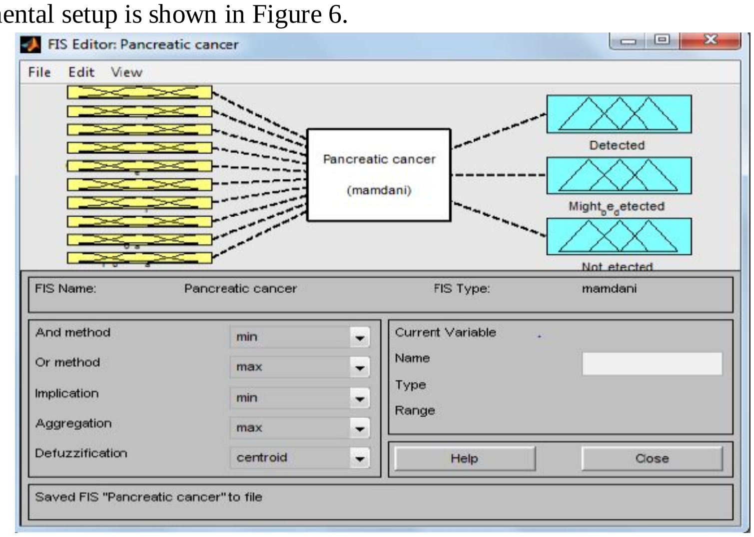


Figure 6. FIS Editor

A triangular membership function editor in MatLab is used for fuzzifying the inputs. This Membership function editor is shown in the Figure 7.

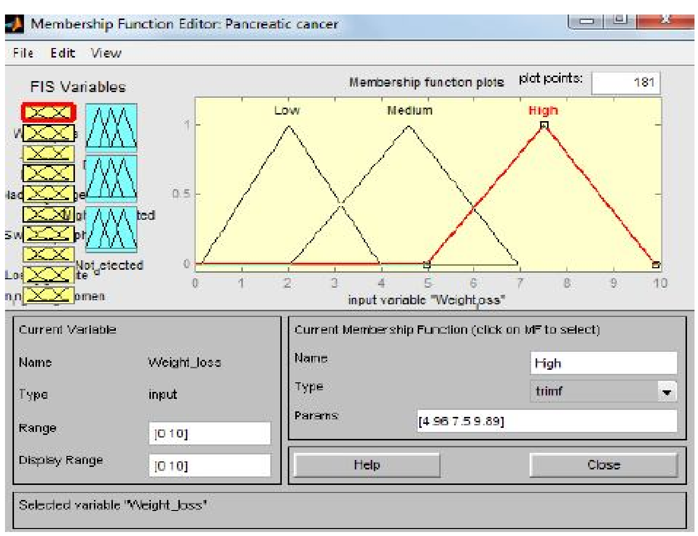


Figure 7. Membership Function Editor

The performance graph plotted based on the results obtained is shown in Figure 8.

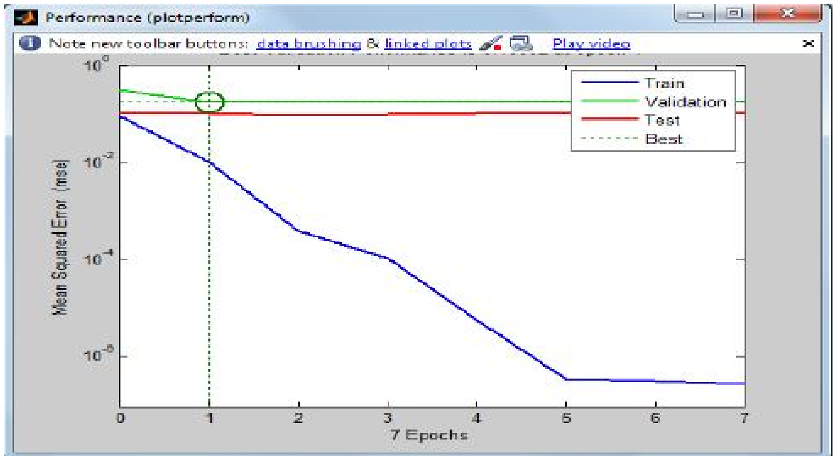


Figure 8. Performance Plot

The results of applying the artificial neural networks methodology to distinguish between PC affected and non-affected region of the pancreas based upon selected symptoms showed very good abilities of the network to learn the patterns corresponding to symptoms of the person.

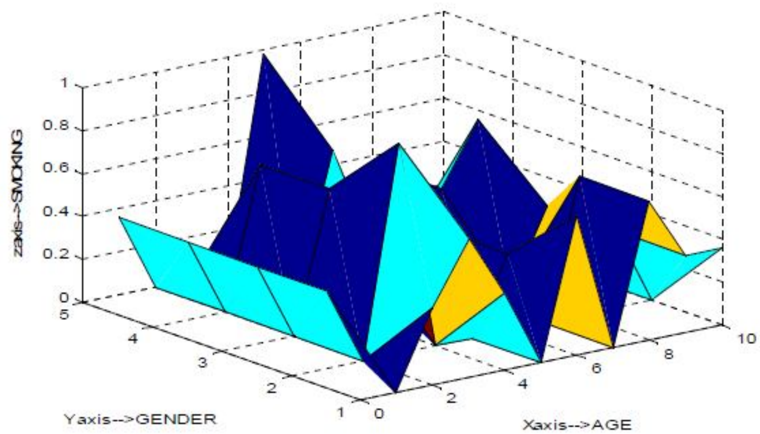


Figure 9. Pancreatic cancer detection (Age,Gen,Smoke)

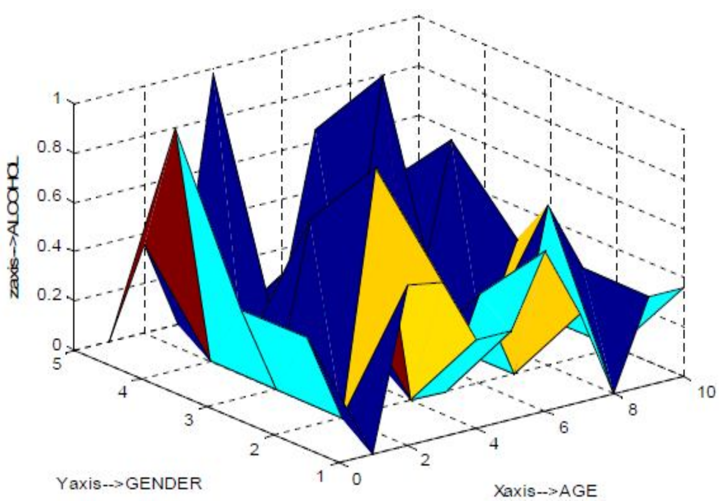


Figure 10. Pancreatic cancer detection (Age,Gen,Alcohol)

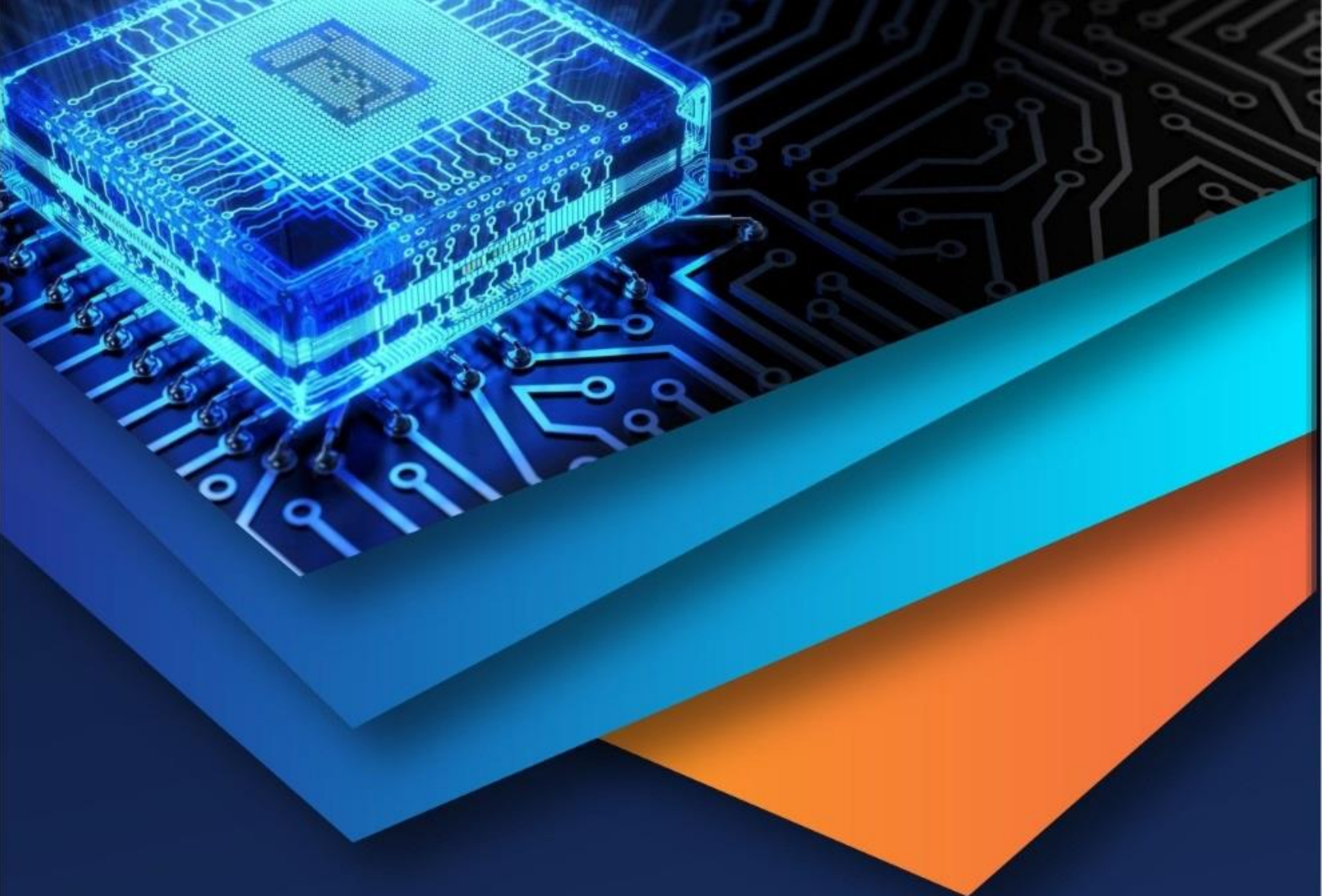
IV. CONCLUSIONS

This study has presented the development of an artificial neural network (ANN) for medical diagnosis. The research aimed to evaluate an ANN in pancreatic cancer diagnosis. The feed-forward back propagation neural network with supervised learning is proposed to diagnose pancreatic cancer. The ANN provide a powerful tool to help medical staff to analyze, model and make sense of complex clinical data across a broad range of medical applications. Pancreatic cancer detection in its early stage is the key of its cure. The automatic diagnosis of pancreatic cancer is an important, real-world medical problem. In this paper it has shown how neural networks are used in actual clinical diagnosis of pancreatic cancer.

The real procedure of medical diagnosis which usually is employed by physicians was analyzed and converted to a machine implementable format. The ANN showed significant results in dealing with data represented in symptoms. The results showed that the proposed diagnosis neural network could be useful for identifying the affected person. Artificial neural networks with the ability to learn by example are provided a very flexible and powerful tool in medical diagnosis offering very useful applications to modern medicine. Future works can be extended for other similar disease detection comprising complex and related datasets with similar or better accuracy.

REFERENCES

[1] www.medlineplus.gov. (2017, 26-06). Pancreatic Cancer. Available: https://medlineplus.gov/pancreaticcancer.html
[2] N. C. Institute, What You Need To Know About Cancer of the Pancreas. U.S.A.: National Cancer Institute, 2010.
[3] S. Singh, R. Vijay, and Y. Singh, "Artificial Neural Network and Cancer Detection."
[4] I. Aleksander and H. Morton, An introduction to neural computing vol. 3: Chapman & Hall London, 1990.
[5] J. Zupan and J. Gasteiger, Neural networks in chemistry and drug design: John Wiley & Sons, Inc., 1999.
[6] A. Das and M. Bhattacharya, "GA based neuro fuzzy techniques for breast cancer identification," in Machine Vision and Image Processing Conference, 2008. IMVIP'08. International, 2008, pp. 136-141.
[7] S. Ahmed and K. D. Potheri, "Medical Diagnosis using Neural Networks," 2014.
[8] D. Aha and D. Kibler, "Instance-based prediction of heart-disease presence with the Cleveland database," University of California, 1988.
[9] J. W. Smith, J. Everhart, W. Dickson, W. Knowler, and R. Johannes, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," in Proceedings of the Annual Symposium on Computer Application in Medical Care, 1988, p. 261.
[10] S. M. Patil and R. Mudholkar, "An osteoarthritis classifier using back propagation neural network," International Journal of Advances in Engineering & Technology, vol. 4, 2012.
[11] Y. Li, A. M. Rauth, and X. Y. Wu, "Prediction of kinetics of doxorubicin release from sulfopropyl dextran ion-exchange microspheres using artificial neural networks" European Journal of Pharmaceautical Science vol. 24, pp. 401–410, 2005.
[12] P. Fedor, I. Malenovsky, J. Vanhara, W. Sierka, and J. Havel, "Thrips (Thysanoptera) identification using artificial neural networks," Bull. Entomol. Res., vol. 98, pp. 437–447, 2008.
[13] V. Michalkova, A. Valigurova, M. Dindo, and J. Vanhara, "Larval morphology and anatomy of the parasitoid Exorista larvarum (Diptera: Tachinidae), with an emphasis on cephalopharyngeal skeleton and digestive tract" Journal of Parasitology vol. 95, pp. 544–554, 2009.
[14] N. Murarikova, J. Vanhara, A. Tothova, and J. Havel, "Polyphasic approach applying artificial neural networks, molecular analysis and postabdomen morphology to West Palaearctic Tachina spp. (Diptera, Tachinidae).," Bull. Entomol. Res. , vol. 101, pp. 165–175, 2010.
[15] S. Kamruzzaman, A. R. Hasan, A. B. Siddiquee, M. Mazumder, and E. Hoque, "Medical diagnosis using neural network," arXiv preprint arXiv:1009.4572, 2010.
[16] M. Payandeh, M. Aeinfar, V. Aeinfar, and M. Hayati, "A New Method for Diagnosis and Predicting Blood Disorder and Cancer Using Artificial Intelligence (Artificial Neural Networks)," International Journal of Hematology-Oncology and Stem Cell Research, vol. 3, pp. 25-33, 2009.
[17] B. Sumathi and A. Santhakumaran, "Pre-diagnosis of hypertension using artificial neural network," Global Journal of Computer Science and Technology, 2011.
[18] F. Bakpo and L. Kabari, "Diagnosing Skin Diseases Using an Artificial Neural Network," in Artificial Neural Networks-Methodological Advances and Biomedical Applications, ed: InTech, 2011.
[19] B. Pandey, T. Jain, V. Kothari, and T. Grover, "Evolutionary Modular Neural Network Approach for Breast Cancer Diagnosis," International Journal of Computer Science Issues, vol. 9, pp. 219-225, 2012.
[20] K. Kumar and B. Abhishek, "Artificial neural networks for diagnosis of kidney stones disease," 2012.
[21] I. A. Agboizebeta and O. J. Chukwuyeni, "Application of neuro-fuzzy expert system for the probe and prognosis of thyroid disorder," International Journal of Fuzzy Logic Systems (IJFLS) Vol, vol. 2, 2012.
[22] J. Obi and A. Imianvan, "Interactive neuro-fuzzy expert system for diagnosis of leukemia," global journal of computer science and technology, 2011.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY