Comparative Analysis of Medical Diagnostic Techniques Using ANN

[1]Farooq Ayoub Dar, [2]Yashwant Singh
[1]Research Scholar, [2]Associate Professor
[1]Department of Computer Science & Technology, Central University of Jammu, Jammu & Kashmir, India

[1]aaqib4466@gmail.com, [2]yash22222k1@gmail.com

ABSTRACT

An immense and immeasurable amount of data is available to medical experts, extending from points of interest of clinical manifestations to different sorts of biochemical information and yields of imaging gadgets. Each kind of information yields data that must be assessed and relegated to a specific pathology amid the diagnostic process. To rationalize the diagnosis in every day routine and maintain a strategic distance from misdiagnosis, methods of machine learning (particularly ANNs) may be utilized. The versatile learning algorithms of machine learning may deal with several kinds of restorative heterogeneous information and classify them into various class outputs. In this paper, we concisely survey and examine the logic, capacities, and performance of ANNs in medical diagnosis of various diseases by making comparative analysis and focusing more on the medical diagnosis of Diabetes. The use of PID dataset for diagnosis is also demonstrated.

Keywords: Multi-Layer Perceptron Neural Networks (MLPNN), PID (Pima Indian Diabetes Dataset), MLFFN (Multilayer Feedforward Network), BPN (Backpropagation Network), General regression neural network (GRNN), Radial basis function (RBF).

1. INTRODUCTION

As medicinal data frameworks in present-day hospitals and restorative medical organizations end up noticeably bigger and bigger, it causes incredible challenges in extricating valuable data for decision support. The customary manual information investigation has turned out to be wasteful and techniques for proficient computer-based analysis have become the need of the hour. It has been demonstrated in the literature that the advantages of bringing machine learning into medicinal examination increase analytic exactness, diminish costs and lessen human resources. The Artificial Neural Networks serve this purpose and become the instrument to enable specialists to examine, model and comprehend complex clinical information over an expansive scope of therapeutic applications. The most utilizations of ANNs are classification problems, i.e., the undertaking task is based on the metrics of allotting the patient to one of a little arrangement of classes.

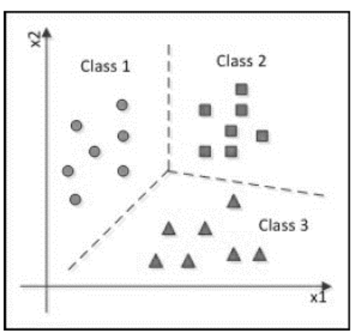
This paper is organized into sections - section 2 presents a concise survey of various algorithms for medical diagnostics, and examination of the logic, capacities, and performance of ANNs in medical diagnosis of various diseases by making comparative analysis are demonstrated in section 3. The section 4 concluded the paper.

2. LITERATURE SURVEY

The design of prediction models for various diagnosis has been an active research area for the past decade. Most of the models found in literature are based on clustering algorithms and ANNs. The ANNs are perfect in perceiving disease aliments utilizing examines or scans because there is no compelling reason to give a particular algorithm to recognize the aliment. The Neural networks learn by illustration therefore the points of interest of how to recognize an abnormality isn't required. All simulated neural network systems are separated into two learning classifications - supervised and unsupervised. In the former, the system is prepared by furnishing it with input and output training data. Amid this stage, the neural system can modify the association weights to coordinate its output with the desired response in an iterative procedure. An ANN of the unsupervised learning category, for example, the self-organizing map, the neural system is furnished just with the input training data, there are no predefined target outputs. The system must build up its own particular portrayal of the information jolts by computing the worthy association weights.

2.1. Foundation of Disease Classification Problems

One thing neural systems are better than most other classifiers at is characterizing records. The exceptionally straightforward perceptron draws a boundary limit, characterizing whether an information point has a place with some region, while a locale indicates a class. We should investigate outwardly on a x-y scatter diagram:



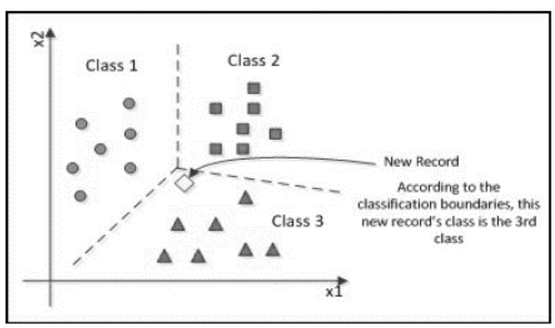


Fig 1: Decision boundary separating classes(left); Fig 2: X-Y scatter plot showing how new records are classified(right)

The dashed lines unequivocally isolate the points into classes. These points indicate information records which initially had the comparing class labels. That implies their classes were initially known, subsequently this characterization falls in the supervised learning classification. A classification algorithm looks to discover the limits between the classes in the information

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online), Web Presence: http://ijoe.vidyapublications.com

hyperspace. Once the order limits are characterized, another information point, with an obscure class, gets a class name as indicated by the limits characterized by the grouping algorithm. The figure indicates how another record is ordered. In light of the present class arrangement, the new record's class is the third class [1].

2.2. Classification Algorithms

The classification algorithms can be arranged into two kinds – MLPNN and Backpropagation Training Algorithm. In MLPNN, the literature investigation discloses a persevering use of feedforward forward neural systems. The one type of ANN component is the Multi-Layer Perceptron Neural Networks (MLPNN). In MLPNN, the solitary and essential errand of the neurons in the input layer is the division of the input signal x_i among neurons in the hidden layer. Each neuron j in the concealed layer includes its inputs signals x_i once it weights them with the strengths of the particular associations w_{ji} from the input layer and decides its output y_j as a function f of the aggregate, given as

$$Y_i = f(\sum w_{ji}\, x_i) \qquad (1)$$

On the other hand, the backpropagation algorithm can be utilized adequately to prepare or train neural systems; it is broadly perceived for applications to layered feedforward forward networks, or multi-layer perceptrons. The backpropagation learning algorithm can be isolated into two stages: propagation and weight update.

The benefits of ANNs for classification arrangement are as follows: (i) ANNs are more vigorous on account of the weights. (ii) The ANNs enhances its execution by learning. This may proceed even after the training has been stopped. (iii) The utilization of ANNs can be parallelized as determined above for better execution. (iv) There is a low blunder rate and in this manner a high level of exactness and precision once the proper training has been performed. (v) ANNs are more robust in boisterous condition. These preferences prompt the utilization of neural systems in disease diagnosis. The architecture of ANN based Diagnostic systems is shown in figure. The inputs can be the related symptoms, biochemical analysis of urine, blood etc. The outputs are diagnostic results which can be positive, negative or uncertain in some cases.

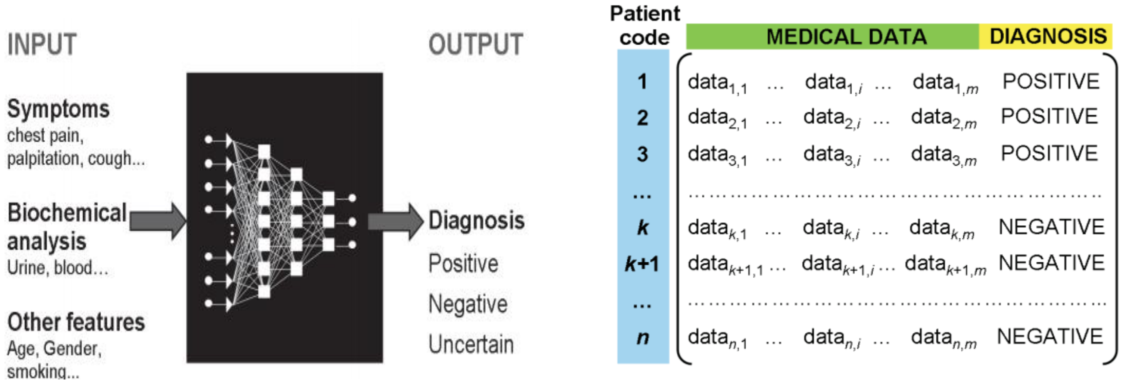


Fig. 3. Points of interest concerning ANNs-based diagnosis(Left), Example of training data(right) [6]

2.3. Related Work On Medical Diagnosis using ANN

Er. Temurtas and Yumusak [2] exhibited a relative chest ailment conclusion and that was acknowledged by utilizing a probabilistic, multilayer, generalized regression and learning vector optimization.

Turkoglu, Sengur and Das, [3] utilized a neural systems outfit based approach for diagnosis of the coronary illness. The three autonomous neural systems models were used to develop the ensemble model. The quantity of neural networks node in the ensemble model was likewise expanded but no improved achievement in execution was reported.

Therapeutic diagnosis of urological dysfunctions was assessed by Johnsson, Fernandez, Gil, Paya, and Garicia [4] out of some ANN system models as apparatuses for help. They created three network models – one supervised and two unsupervised ANN. Additionally, Altunay et al [5] analysed the uroflowmetric information and helped doctors for their diagnosis and presented a specialist diagnostic framework for assessing possible indications from the uroflow signals automatically. This framework utilized ANNs and created a pre-analytic outcome. Canaris, Heckerling, Tape, Flach, Gerber and Wigton [7] utilized ANNs combined with genetic algorithms to develop blends of clinical factors upgraded for foreseeing urinary tract disease. Juan Manuel, Francisco, Daniel and Antonio [8] built up another framework from a model based on multiagent framework in which each neuronal focus corresponds with an agent. This framework consolidates a heuristic keeping in mind the end goal to make it more strong within the sight of conceivable irregularities. The heuristic utilized depends on a neural network (orthogonal associative memory). Information through training has been added to the framework, utilizing right patterns of behaviour of the urinary tract and behaviour patterns as a result of urinary dysfunctions.

Another therapeutic application of ANN was presented by Moallem and Monadjemi [9]. The strategy of medicinal diagnosis which as a rule is utilized by doctors was analysed and changed over to a machine-implementable organization. The aftereffects of the trials and furthermore the benefits of utilizing a fuzzy approach were analysed too.

Lin [10] introduced CART (classification and regression tree) and CBR (case-based reasoning) methods to structure a smart diagnosis model planning to give an exhaustive and comprehensive expository system to raise the precision of liver ailment diagnosis.

Narender Kumar et al [11] have compared various data classifications techniques and their accuracy performance to predict chronic kidney disease. The various classifiers being classified in this paper are J48, Naive Bayes, Random forest, SUM & KNN. These classifiers are compared with the help of performance measures like ROC, kappa statistics, RMSE & MAE using WEKA tool. After experimentation it has been identified that random forest classification techniques have higher performance in terms of accuracy and prediction as compared to other techniques.

Habas, Mazurowski, Lo, Zurada, Tourassi and Baker [12] examined the impact of class imbalance in training data when creating ANN classifiers for computer-aided therapeutic diagnosis. The examination was accomplished within the sight of other characteristics that are common among restorative medical data, specifically little training set length, the extensive number of features, and correlations between features.

Xing et al (2017) [13] have proposed research which is concerned with the aim to develop a data mining algo to predict survival of CHP patients (Coronary Heart Disease). In the research, three algo's were used to develop these prediction models. These three prediction models were then compared to each other on the basis of performance. The three models are ANNs, SVMs & decision trees. Out of these three, SVM came out to be the best with 92.10% of accuracy, ANN came out to be the second with 91.10% of accuracy and Decision tree came in the last with the performance accuracy of 89.6%. In SVM, Sequential minimal optimization algorithm is used to train it. ANN, multilayer perceptron with back propagation is being used. In decision tree algo, C5 algorithm was used.

Yan, Zhang, Zhang and Zhao [14] proposed a technique for building up a completely automatic computer aided conclusion framework to assist the radiologist in recognizing and diagnosing small scale calcifications in advanced arrangement mammograms.

Sato, Higuchi, Furuse, Makuuchi, Takeda and Takamoto [15] tried a three-layered neural system investigation of phonocardiogram recordings to analyze, automatically and impartially, the state of the heart in patients with heart murmurs.

K Vishvant et al, 2014 [16] mentions the kidney abnormalities like stones, cysts, cancerous cells etc. Therefore, the importance of identification of enact and the determination of precise location of the stones in kidney is vital. It can be done manually by analyzing ultra sound image but the image is less speckle noise. In this way the ultra sound experiences preprocessing in which image restoration is first done, at that point it is applied to Gabor filtering for smoothening. The subsequent stage is the image enhancement utilizing histogram equalization. The ultrasound image in preprocessing stage is segmented utilizing level set segmentation, since it produces optimal outcomes. Level set segmentation uses two terms. In the first place term, a momentum term is utilized and second term depends on resilient propagation. Extricated area of the kidney after segmentation is connected to Symlets, Biorthogonal and Daubechies wavelet subbands to separate energy levels. These energy levels give a sign about potentiality of stone in that specific area which altogether shift from that of ordinary energy level. These energy levels are trained by MLPN and BPNN to distinguish the kind of stone. The multilayer perceptron with back propagation give high precision of 98% when contrasted with Naive Bayes classification. In this way, plainly the utilization of combinational level set segmentation, MLP with BP, wavelet filters is better approach for the identification of stone in kidney.

Fengying Xie et al (2017) [17] developed a novel method for classifying melatonin tumour as benign or malignant by analysing digital images. First they extracted elisions using a self-generating neural network. Then the features of tumour like colour, texture and border are extracted and then in the last step lesions are classified. In their work, they have designed an ensemble classifier that combines back propagation neural network with fuzzy neural network. Dr. N Ganesan et al (2010) [18] also made an attempt to utilize ANNs in the medical field (carcinogenesis) preclinical study. In this paper, they indicated how neural systems are utilized as a part of genuine clinical determination of lung tumour. In this work, the execution of neural system structure has been researched for the diagnosis of lung cancer.

2.4. Pima Indian Diabetes (PID) Database

In order to perform the survey on medical diagnosis of diabetes, PID dataset was also studied. All subjects in this PID database are Pima Indian ladies no less than 21 years of age and living close Phoenix, Arizona, USA.
The dataset contains two classes and 768 samples. The class appropriation is
Class 1: normal (500)
Class 2: Pima Indian diabetes (268)
The above samples have eight clinical features or findings - Number of times pregnant; Plasma glucose concentration (2 h in an oral glucose tolerance test); Diastolic blood pressure (mm Hg); Triceps skin fold thickness (mm); 2-h serum insulin (lU/ml); Body mass index (weight in kg/(height in m)^2); Diabetes pedigree function; Age (years).

2.4.1. Past Work on Diagnosis of Diabetes and PID Dataset

Rajeeb dev et al (2008) [19] evaluated the problem of diagnosing diabetes by treating the diagnosis as two classification problem, (that is the diabetic persons fall in class 1 and non-diabetic persons fall in class 2). For classification task, supervised MLFFN is used along with BPN learning algorithm. The coding was done in MATLAB to develop network software tool. In this they used two different ANN architectures to find result. One is single hidden layer and other is double hidden layer. Both the architectures have six input nodes and one output node because the problem is binary classification. The network was able to classify patients into diabetic and non-diabetic patients with performance of 92.50%. It is clear from the result presented that by increasing the number of hidden layers as well as neuron for a given learning rate the specific error goal has been attained in the number of epochs.

Smith et al. [20] utilized the PID dataset to assess the perceptron-like ADAPtive learning routine (ADAP). This investigation had 576 cases in the training set and 192 cases in the test set. Utilizing 576 training occurrences, the sensitivity and specificity of their calculation was 76% on the remaining 192 cases. A similar number of arbitrary training and test sets was utilized to compare the simulation results.

Kamer Kayaer and Tulay Yildirim [21] utilized three distinctive neural network structures - MLPNN, RBF and GRNN on the PID database, restorative medical data. The execution of RBF was more terrible than the MLP for every spread trials attempted. The best outcome accomplished on the test data is the one utilizing the GRNN structure (80.21%). This is near one with the most noteworthy true classification result that was accomplished by utilizing the more intricate organized ARTMAP-IC network (81%) [22]. This outcome demonstrates that, general GRNN can be a decent and functional decision to characterize a therapeutic diagnostic information.

Table 1: Disease Diagnosis by Various Authors

	Disease diagnosis	**Efficiency**	**References**
Er. N.Yamusak & Temurtas (Radial Based Function NN)	Chest (Tuberclosis,Pneumonia, Asthma, Lung cancer)	90%	[2]
Das, Turkoglue & Sengur (SAS Base Enterprise Software)	Heart (CAD)	89.01%	[3]
Gil, Johsson, Garicia, Paya	Urological Dysfunction	Approx. 90%	[4]
Altunay et. al	Analysis of uroflowmetric data	78%	[5]
Monadjemi and Moallem	Typical Disease Diagnosis	88.5%	[9]
Lin (CART & CBR)	Liver	CART=92.94 %; CBR=90.00%	[10]
Narender Kumar (Naive Bayes, Random forest, SUM & KNN)	Kidney related Diagnosis	Decision Table =99%	[11]
Mazurowski, Habas, Zaruda (Design of CAD Systems)	Computer Aided Medical Diagnosis	-	[12]
Xing et. al (SVM, ANN, Decision Trees)	Coronary Heart Disease	SVM=92.10%; ANN=91.10%; Decision Trees =89.6%	[13]
Zhang, Yan, Zhao (Features extraction, Automatic segmentation, classification, BPNN and suspicious area detection)	Radiology (Mammography)	Benign =84.10% Malignant= 80.30%	[14]
Higuchi, Sato, Makuuchi, Furuse, Takamoto and Takeda	Heart disease diagnosis (may include analysis of waveforms – ECGs, Phonocardiography, EEGs)	Showed variable accuracies for different heart diseases	[15]

K. Vishvant	Kidney abnormalities (stones, cysts, cancerous cells etc).	MLP with backpropagation=98%	[17]

Table 2: Past Work on Diabetes/Pima Indian Diabetes Database

	Disease diagnosis	**Efficiency**	**References**
Rajeeb Dev et al (Binary Classification problem)	Diabetes	92.50%	[19]
Smith et al. (Pima Indian Diabetes Database)	Diabetes	Sensitivity, specificity=76%	[20]
Tulay Yildirim & Kamer Kayaer (MLP, RBF & GRNN on PID Database)	Diabetes	GRNN=80.21 %	[21]

3. COMPARATIVE ANALYSIS

Subsequent to plunging into the literature and related work of medical diagnostic techniques using ANN, it was found that the diagnosis techniques follow some basic foundation methods for monitoring, diagnosis and classification that are common to initial diagnosis for a number of diseases. From the literature, we have got an overview of how monitoring, classification and diagnosis is accomplished using artificial neural networks. This survey shows the immense adaptability of the ANN models and their capacity, in a wide assortment of regions, to perform with critical indicative precision. Be that as it may, more work will be required before these systems may be acknowledged as technically honest and legitimate clinical guide. Quite notably, more research approving system validation of these ANN models tentatively on large number of patients are required to demonstrate that these types of network model systems do enhance the diagnostic performance.

4. CONCLUSION

In view of these Neural systems we will utilize some neural network diagnostic techniques on Pima Indian Diabetes Database, keeping in mind to achieve the maximum possible efficiency.

This monitoring, classification and diagnosis using ANN has several advantages including-(i) The capacity to process vast measure of information, (ii) Reduced probability of ignoring pertinent data, (iii) Reduction of diagnosis time. Besides, Medical Diagnosis System for analysis of ailments can be intended to analyse a disease in light of the arrangement and occurrence of symptoms. All diseases have an arrangement of symptoms related with them. The framework can reenact medicinal reasoning. The framework can be utilized as a part of remote territories where medical offices are not effortlessly accessible or not accessible in a convenient manner. The framework may give a score to each of the ailments and may likewise give a positioning or ranking to every one of the ailments in view of the arrangement of symptoms selected.

REFERENCES

[1] Yusuke Sugomori, Bostjan Kaluza, Fabio M. Soares, Alan M.F Souza: Deep Learning Practical Neural Networks with Java, 2017

[2] O. Er, N. Yumusak and F. Temurtas, "Chest disease diagnosis using artificial neural networks", Expert Systems with Applications, Vol.37, No.12, 2010, pp. 7648-7655.

[3] R. Das, I. Turkoglu and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles", Expert Systems with Applications, Vol.36, No.4, 2009, pp. 7675-7680.

[4] D. Gil, M. Johnsson, J. M. Garicia Chemizo, A. S. Paya and D. R. Fernandez, "Application of Artificial Neural Networks in the Diagnosis of Urological Dysfunctions", Expert Systems with Applications, Vol.36, No.3, 2009, pp. 5754-5760.

[5] S. Altunay, Z. Telatar, O. Erogul and E. Aydur, "A New Approach to urinary system dynamics problems: Evaluation and classification of uroflowmeter signals using artificial neural networks", Expert Systems with Applications, Vol.36, No.3, 2009, pp. 4891-4895.

[6] Filippo Amato, Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš Hamp, Josef Havel1: “Artificial neural networks in medical diagnosis”

[7] P. S. Heckerling, G. Canaris, S. D. Flach, T. G. Tape, R. S. Wigton and B. S. Gerber, "Predictors of urinary tract infection based on artificial neural networks and genetic algorithms", International Journal of Medical Informatics; Vol.76, No.4, 2007, pp. 289-296.

[8] M. P. Francisco, G. C. Juan Manuel, S. P. Antonio and R. F. Daniel, "A robust model of the neuronal regulator of the lower urinary tract based on artificial neural networks", Neurocomputing, Vol.71, No.4-6, 2008, pp. 743-754.

[9] S. A. Monadjemi and P. Moallem, "Automatic Diagnosis of Particular Diseases Using a Fuzzy-Neural Approach", International Review on Computers & Software, Vol.3, No.4, 2008, pp. 406-411.

[10] R. Lin, "An intelligent model for liver disease diagnosis", Artificial Intelligence in Medicine, Vol.47, No.1, 2009, pp. 53-62.

[11] N. Kumar and S. Khatri, "Implementing WEKA for medical data classification and early kidney disease prediction," *2017 3rd Int. Conf. Comput. Intell. Commun. Technol.*, pp. 1–6, 2017.

[12] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker and G. D. Tourassi, "Training Neural Network Classifiers for Medical Decision Making: The Effects of Imbalanced Datasets on Classification Performance", Special Issue on Advances in Neural Networks Research: IJCNN '07, 2007 International Joint Conference on Neural Networks IJCNN'07 Orlando, Florida, USA, April, 2008, Vol.21, No.3, pp. 427-436, http://ci.louisville.edu/zurada/publications/mazurowski.nn2008.pdf.

[13] Y. Xing, J. Wang, Z. Zhao, and andYonghong Gao, "Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease," *2007 Int. Conf. Converg. Inf. Technol. (ICCIT 2007)*, pp. 868–872, 2007.

[14] G. Zhang, P. Yan, H. Zhao and X. Zhang, "A Computer Aided Diagnosis System in Mammography Using Artificial Neural Networks", 2008 International Conference on BioMedical Engineering and Informatics, 2008, Vol.2, pp. 823-826.

[15] K. Higuchi, K. Sato, H. Makuuchi, A. Furuse, S. Takamoto and H. Takeda, "Automated diagnosis of heart disease in patients with heart murmurs: application of a neural network technique", Journal of Medical Engineering & Technology, Vol.30, No.2, 2006, pp. 61-68

[16] K. Viswanath and R. Gunasundari, "Design and analysis performance of kidney stone detection from ultrasound image by level set segmentation and ANN classification," *Proc.*

2014 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2014, pp. 407–414, 2014.

[17] F. Xie, H. Fan, Y. Li, Z. Jiang, R. Meng, and A. Bovik, "Melanoma classification on dermoscopy images using a neural network ensemble model," *IEEE Trans. Med. Imaging*, vol. 36, no. 3, pp. 849–858, 2017.

[18] N. Ganesan, "Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data," *Int. J. Comput. Appl. (0975*, vol. 1, no. 26, pp. 76–85, 2010.

[19] R. Dey and V. Bajpai, "Application of Artificial Neural Network (ANN) technique for Diagnosing Diabetes Mellitus," *IEEEexplorer*, no. 155, pp. 8–11, 2008.

[20] Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., Johannes, R. S., "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus", Proc. Symp. on Computer Applications and Medical Care (Piscataway, NJ: IEEE Computer Society Press), pp. 261–5, 1988.

[21] Kamer Kayaer and Tulay Yildirim, "Medical Diagnosis on Pima Indian Diabetes Using General Regression Neural Networks

[22] Carpenter, G.A., Markuzon, N., "ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases", Neural Networks, 11:323-336, 1998.