Cascade Correlation Neural Network Model for Classification of Oral Cancer

NEHA SHARMA[1], HARI OM[2]
[1]Padmashree Dr. D.Y.Patil Institute of MCA, [2]Computer Science and Engineering,
[1]University of Pune, [2]Indian School of Mines
[1]Sector-24, Nigdi Pradhikaran, Pune, Maharashtra; [2]Indian School of Mines, Dhanbad, Jharkhand
[1]INDIA; [2]INDIA
[1]nvsharma@rediffmail.com; [2]hariom4india@gmail.com

Abstract: - The rationale of this study is to accurately classify the records of the oral cancer patient on the basis of clinical symptoms, Gross Examination, Predisposing Factor, Histopathology, various tests and treatments. In this paper, Cascade correlation neural network model has been built as it combines together the idea of cascade architecture and learning algorithm together and it is estimated to be at least 10 times faster than standard back-propagation algorithms. The records of 1025 patients described with the help of 35 attributes are analysed to predict the rate of survivability of oral cancer patients. Dataset is divided in two subgroups: training subgroup and test subgroup, in order to verify the network's ability to diagnose new cases. Performance of the model for its ability to predict is evaluated on the basis of various measures. Classification accuracy of the model is 72.10%, sensitivity is 83.05%, specificity is 64.71%, precision of the model is 61.36%, recall capacity is 83.05%, f-measure value is 0.7058 and area under ROC curve is 0.944. Lift and Gain chart also suggest that cascade correlation neural network is an effective model for predicting oral cancer.

Key-Words: - Oral Cancer, Data Mining, Artificial Neural Network, Predictive Model, Neural Network, Cascade Correlation Neural Network

1 Introduction

Artificial neural networks are regularly utilized in areas where traditionally statistical methods were used for prediction and classification [1]. The standard statistical method and neural network are usually looked at as competitive model-building approaches [2]. The primary advantage of neural network is that they can take on problems that are usually overly bewildering for conventional technology; problems that do not have an algorithmic solution or for which an algorithmic solution is too complex to be defined. By and large, neural networks are well suited to problems that human beings are exceptional at elucidation and are excellent at resolving, but computers are generally not [3]. For example, pattern recognition and forecasting, which requires the detecting the trends in data [4]. The focal point and true power of neural networks lies in their capability to characterize both linear and non-linear relationships and in their ability to learn these relationships straight from the data being modelled. Conventional linear models are basically insufficient for representing data that has non-linear characteristics [5][6]. Therefore, in this paper, we attempt is to build Cascade correlation neural network to address a critical problem of oral cancer classification. Malignancies in the oral cavity usually are escalating at an alarming rate and also have been claimed to account for approximately twenty two thousand new instances annually for men (5% of all cancers) and ninety thousand for women (2% of cancers) [7]. The late diagnosis of potential conditions of oral cancer and precancerous lesions has reported very high incidence of disease in Indian subcontinent. Oral submucous fibrosis is a precancerous condition of the oral cavity that has high potential for being malignant and may develop into squamous cell carcinoma [8]. Currently, oral cancer is a public health issue in several areas of the globe like the United Kingdom [9], South Africa [10] and several southeast Asian countries [11][12].

To create cascade correlation neural network, we began with the study of neural network model which is based on mathematics and algorithms, to understand the concepts [13] and then Hebbian learning and its variations were deliberated upon [14-16]. Also, the literature on perceptrons and back-propagation were reviewed [17], followed by review of various related work. Rosmai et al. [18] evaluated the performance of a fuzzy neural network model and fuzzy regression model to predict the likelihood of an individual in developing oral cancer based on knowledge of their risk habits and demographic profiles at Oral Cancer Research and Coordinating Centre. The result found is better than the results of oral cancer clinicians. Rosmai et al [19], once again built fuzzy logic, fuzzy neural network and fuzzy linear regression models to determine the success factors of oral cancer susceptibility prediction. For 1-input and 2-input predictor sets, all the three models were found to have 64% prediction accuracies. Nevertheless, for more numbers of inputs (for example 3-input and 4-input), the prediction accuracies of both fuzzy neural networks' and fuzzy linear regression's increased to 80%, while fuzzy logic prediction accuracy remains at 64%. Kaladhar et al. [20] used classification algorithms like CART, Random Forest, LMT, and Naïve Bayesian for predicting oral cancer survivability. 10 fold cross validation method was used for evaluating the performance of algorithm. According to the result, the Random Forest technique classified the cancer survival data set correctly. Kang et. al. [21] developed four data mining models, which included two artificial neural network (ANN) models and two classification and regression tree (CART) models, to predict hospital charges and insurance claim of cancer patients and compared their efficacies. The prediction accuracy of ANN models was found to be better than CART models. However, the CART models were efficient and effective in allocating limited medical resources. Cascade correlation has not been applied much, however few researchers have carried out comparisons to understand cascade correlation neural network [22], whereas some has developed learning algorithm for evolving the network itself [23].

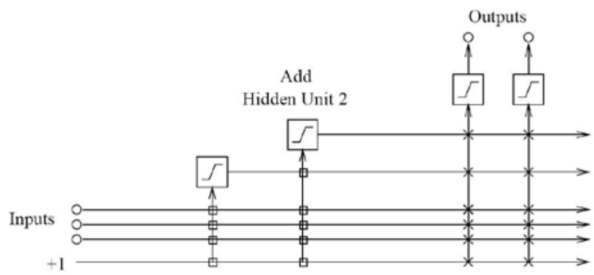
This paper is organized as follows: Section 2 discusses the artificial neural network model and section 3 covers various methods used to estimate their performance. Section 4 and section 5 presents the experimental results and discussion on the model developed, respectively. Section 6 concludes the paper.

2 Artificial Neural Networks

Neural networks adopt the functioning of the brain, which is popularly known as biological nervous system, to process information. The way human brain gains information and stores, similarly neural networks acquire knowledge through learning and store it within inter-neuron connection strengths known as synaptic weights. There are ever-increasing number of real world problems that are solved using neural networks [24][25].

2.1 Cascade Correlation Neural Network

Cascade-Correlation neural network is developed in 1990 by Fahlman and Libiere and characterized as a constructive learning rule [26]. Cascade-Correlation is a generative, feed-forward and supervised learning architecture [26]. Cascade-Correlation begins with a minimal network consisting only of an input and an output layer, then automatically trains and adds new hidden units one by one creating a multi-layer structure as shown in Fig 1. Cascade correlation combines two ideas: first is the cascade architecture and second is learning algorithm. The architecture is cascading because one hidden unit is added to the network at one point of time and it receives the inputs from all the neurons already in the network. Once the hidden unit is added to the network, it does not change and learning algorithm tries to maximize the magnitude of the correlation between the new neuron's output and the residual error signal of the network [27-29].



Where:
Frozen Connections
X Adjustable Connections
Activation Function

Fig. 1 Cascade Correlation Neural Network

This network has many advantages over other networks. It learns at least 10 times faster than standard back-propagation algorithms. The network determines its own size and topologies. It is useful for incremental learning in which new information is added to the already trained network.

2.1.1 Training algorithm for cascade correlation neural networks

Step1. Start cascade correlation neural networks that consists of only an input and an output layer.

Step2. Connect all the inputs neurons to each output neuron by a connection with adjustable weights. Multiply each input neurons's value by its respective weight and compute sum of weight input value of all input neurons. Now sent the weighted input sum along with the bias to output neurons.

Step3. Generate the candidate units which receives a connection from each input unit and from each existing hidden units. Initially, there are no connection and weights between the set of candidate units and the output units.

Step4. Adjust the candidate unit's input weights through iterative passes to maximize the degree of correlation between the candidate unit's value and the residual error at the output neuron of the network. Stop the adjustments if there is no improvement observed in the correlation score.

Step5. Identify the candidate unit with the maximum correlation, block its incoming weights and include it to the network.

Step6. Generate connection between the selected candidate unit and all the output units in order to change the candidate unit into a hidden unit. Obtain new permanent feature detector as the input weights to the new hidden unit are frozen.
Go back to step2.

Step7. Repeat the algorithm until the overall error of the network reduces to minimum.

3. Estimation of Performance

There are various measures to evaluate the effectiveness of any model and estimate its performance. The evaluation methods are significant as it provides consciousness on the quality and features of the design and facilitates to refine parameters through the iterative way of learning. Model estimation strategy also helps in identifying one of the most suitable and well designed models. Various estimation methods used are validation method, misclassification table, confusion matrix, sensitivity-specificity and lift-gain. Cross validation method with 10 folds is used for validation of the cascade correlation neural network model which is also called as rotation estimation [30] and is primarily employed to estimate the accuracy of a predictive model. A model is trained by using the known data and then tested using the unknown data [31]. The learning dataset is used for acquiring rules, the validation dataset is used for validating the rules externally and the process is repeated 10 times [32].

Misclassification summary table presents the number of rows with a particular category that were misclassified by the tree. The two sections of the misclassification table present the actual data and the misclassified data. The detailed information about the classification of data by the model is presented in terms of confusion matrix. Each category of the target variable is represented in the form of a row and column. The actual categories of the target variable are shown in the first column and the predicted categories are across the top of the table. The numbers in cells are the weights of the data with the actual category of the row and the predicted category of the column. The numbers in the diagonal cells are the weights for the correctly classified cases where the actual category matches the predicted category. The off-diagonal cells have misclassified row weights.

Sensitivity and specificity have been calculated to estimate the performance of the model. The patients who are predicted as malignant among the malignant patients are True Positive (TP) cases. The patients who are predicted as non-malignant among non-malignant patients are True Negative (TN) cases. The patients who are predicted as non-malignant among the malignant patients are False Positive (FP) cases. The patients who are predicted as malignant among the non- malignant patients are False Negative (FN) cases. The sensitivity and specificity are calculated by using TP, TN, FP and FN. Sensitivity means the probability that the algorithms can correctly predict non-malignancy, it is given by Sensitivity = TP / (TP + FN). Specificity means the probability that the algorithms can correctly predict malignant, it is defined as Specificity = TN / (FP + TN). The lift and gain table is a useful tool for measuring the value of a predictive model. The basic idea of lift and gain is to sort the predicted target values in decreasing order of purity on some target category and then compare the proportion of cases with the category in each bin with the overall proportion. In the case of a model with a continuous target variable, the predicted target values are sorted in decreasing target value order and then compared with the mean target value. The lift and gain values show how much improvement the model provides in picking out the best 10%, 20%, etc. of the cases.

4. Experimental Results

Cascade correlation neural network is an artificial neural network model that is built and its performance is analysed for estimating the effectiveness of classification of oral cancer patients. The oral cancer patient database is created in non-probabilistic and non-randomized manner and is presented in [33]. The database is then converted to comma separated values (csv) file format so that it can be used to build classification model using DTREG tool [34]. There are total 35 attributes and 1025 records. The model performs classification analysis on the target variable ‘survival’. Kernel function used for hidden neuron is sigmoid and Gaussian function and the kernel function adopted for output neuron is sigmoid function. Cross validation method with 10 folds is used for validating the model.

Cascade correlation neural network model is built with 48 neurons at internal layer, no neuron at hidden layer and 1 neuron at output layer. Minimum weight of the neuron in the network is -2.475025 and maximum weight is 2.888101. Model size summary is presented in Table 1 and Fig 2 shows model size chart for the neural network. The model size chart shows the error as a function of number of neuron in the hidden layer. The blue line shows the error rate for training data whereas red line shows the error rate for validation data. The full model was created without creating any hidden neuron.

Table 1 Cascade correlation neural network

Neurons	% Training Misclassifications	% Validation Misclassifications
0	28.5203	31.9990
1	25.4965	33.8533
2	22.8620	33.6602
3	20.6182	32.6848
4	19.0569	33.5580
5	17.8211	32.6806
6	16.3252	35.6100
7	15.6747	35.7061
8	14.3969	36.1493
9	13.9199	35.4982
10	12.9660	34.9789
11	12.7438	34.3750

Optimal Size

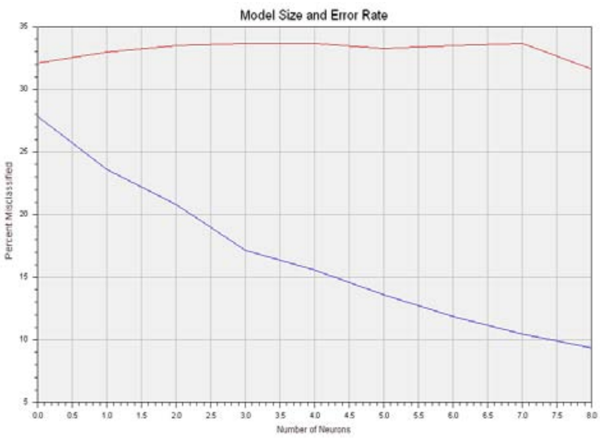


Fig 1 Model Size Chart

Misclassification statistics for cascade correlation neural network for the training and validation data is shown in Table 2 and Table 3. The table also presents the overall accuracy of the model.

Table 2 Misclassification Table for Training Data

	Actual		Misclassified			
Category	**Count**	**Weight**	**Count**	**Weight**	**%**	**Cost**
A	612	612	216	216	35.294	0.353
D	413	413	70	70	16.949	0.169
Total	1025	1025	286	286	27.902	0.279
Overall accuracy = 72.10%						

Table 3 Misclassification Table for Validation Data

	Actual		Misclassified			
Category	**Count**	**Weight**	**Count**	**Weight**	**%**	**Cost**
A	612	612	223	223	36.438	0.364
D	413	413	102	102	24.697	0.247
Total	1025	1025	325	325	31.707	0.317
Overall accuracy = 68.29%						

Confusion Matrix for training and validation data for cascade correlation neural network model is shown in Table 4.

Table 4 Confusion Matrix of the Neural Network

	Training Data		Validation Data	
Actual Category	**Predicted Category**		**Predicted Category**	
	A	D	A	D
A	396	216	389	223
D	70	343	102	311

The positive/negative ratio, true positive (TP), true negative (TN), false positive (FP), false negative (FN), sensitivity, specificity, geometric mean of sensitivity and specificity, positive predictive value (PPV), negative predictive value (NPV), geometric mean of ppv and npv, precision, recall, F-measure and area under Receiver Operating Characteristics (ROC) curve for training and validation data for the models is shown in Table 5 and Fig 2 presents Area under ROC curve for the model. ROC is a graph that demonstrates the performance of a binary classifier model. It is created by plotting true positives out of the total actual positives (TPR = true positive rate) and false positives out of the total actual negatives (FPR = false positive rate), at various threshold settings. TPR is sensitivity and FPR is specificity. The ROC is known as a relative operating characteristic curve as it is a comparison of two operating characteristics (TPR and FPR) as the criterion changes [35]. ROC analysis allows tools to choose perhaps optimal models and to discard others independently from the cost context or the class distribution.

Table 5 Sensitivity and Specificity of the Network

	Training Data	**Validation** Data
Positive/ Negative ratio	0.6748	0.6748
True positive (TP)	33.46%	30.34%
True negative (TN)	38.63%	37.95%
False positive (FP)	21.07%	21.79%
False negative (FN)	6.83%	9.95%
Sensitivity	83.05%	75.30%
Specificity	64.71%	63.56%
Geometric mean of sensitivity and specificity	73.31%	69.18%
Positive predictive value (PPV)	61.36%	58.24%
Negative predictive value (NPV)	84.98%	79.23%
Geometric mean of PPV and NPV	72.21%	67.93%
Precision	61.36%	58.24%
Recall	83.05%	75.30%
F-Measure	0.7058	0.6568
Area under ROC curve	0.779	0.731

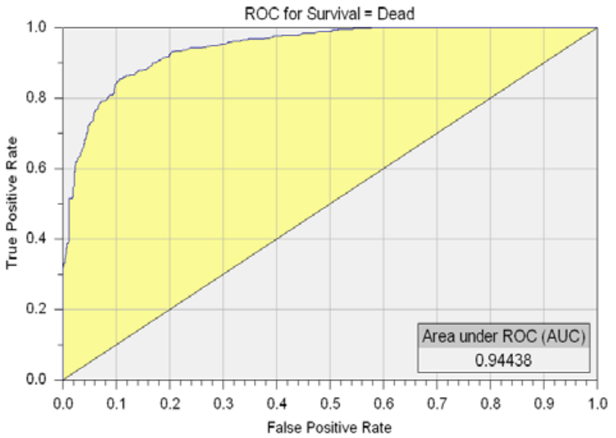


Fig 2 Area Under ROC Curve

Lift and gain for cascade correlation neural network model for Survival = D is shown in Fig 3 and Fig 4.

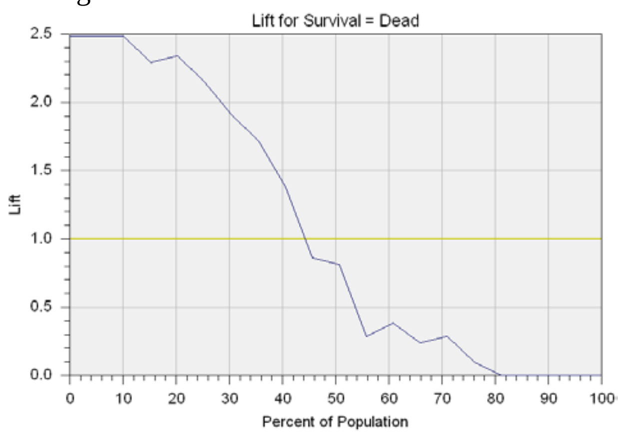


Fig 3 Lift Chart for the Model

For training data, average gain for Survival = A is 1.308 and for Survival = D is 1.426. Percent of cases with Survival = A is 59.71% and with Survival = D is 40.29%. For validation data, average gain for Survival = A is 1.262 and for Survival = D is 1.316. Percent of cases with Survival = A is 59.71% and with Survival = D is 40.29%.

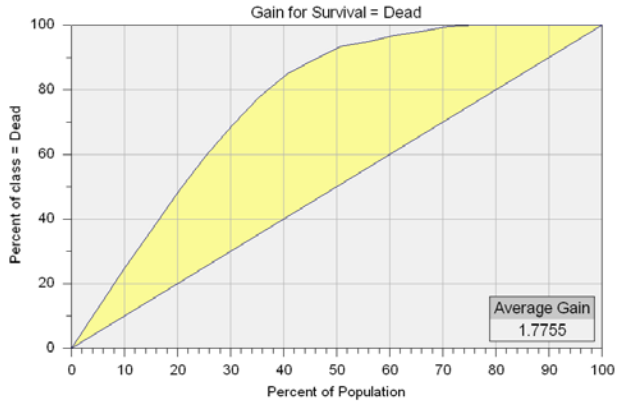


Fig 4 Gain Chart for the Model

A Sensitivity and Specificity Chart represent calculated probabilities. This chart shows that by shifting the probability threshold, sensitivity and specificity can be adjusted for classifying cases as positive or negative. Fig 5 presents the sensitivity and specificity chart for the model.

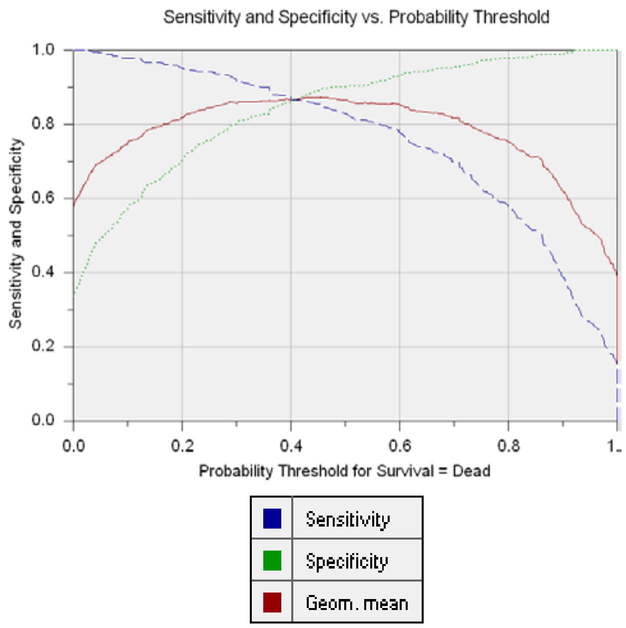


Fig 5 Sensitivity and Specificity Chart for the Model

5. Discussions

In this paper, the cascade correlation neural network, which is not very popular artificial neural network model, is built and analyzed for its effectiveness. The cascade correlation neural network is created to address classification problem of oral cancer patients. It took 00:22.51seconds to build and analyze the model, which is pretty faster that other neural networks.

Unlike other networks, the cascade correlation neural network determines its own size and topologies. It is of use for incremental learning in which new information is added to the already trained network in cascading manner. In this case, the optimal size model is created without any neuron in hidden layer. The model is analyzed for classification accuracy, positive/negative ratio, true positive (TP), true negative (TN), false positive (FP), false negative (FN), sensitivity, specificity, geometric mean of sensitivity and specificity, positive predictive value (PPV), negative predictive value (NPV), geometric mean of ppv and npv, precision, recall, F-measure, area under Receiver Operating Characteristics (ROC) curve and lift-gain. Fig 6 demonstrates the values for both training and validation data.

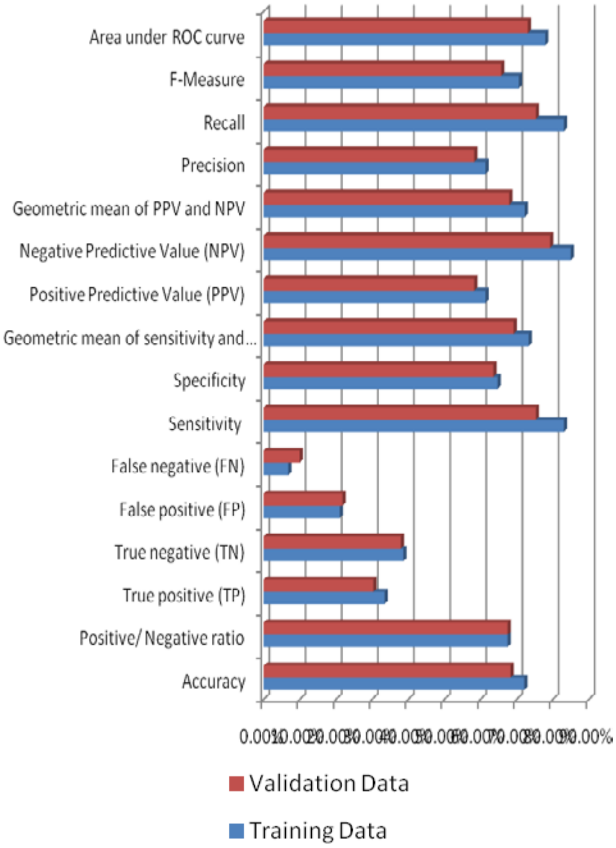


Fig 6 Analysis of the Cascade Correlation Neural Network Model

Area under ROC curve for training data is 0.779 and for validation data is 0.731. Average gain for survival = D for training data is 1.426 and for validation data is 1.316. However, Area under ROC curve for overall model is 0.944 and average gain for the model is 1.775, which suggests that the model is effective and robust.

6. Conclusion and Future Work

The analysis of the cascade correlation neural network clearly displays result and performance which confirms the model as an efficient, strong and faster model for classification of oral cancer patients. Our future work shall include developing more classification models and compare the performance with cascade correlation neural network.

References:

[1] R. P. Lippmann, An introduction to computing with neural nets, IEEE ASSP Magazine, 1987, pp. 4–22.

[2] L. Fausett, Fundamentals of neural networks: Architecture, algorithms and applications, Prentice Hall, 1994.

[3] D. E. Rumelhart, G. E. Hinton and R. J. Williams, Learning representations by back-propagating error, Nature, Vol.323, 1986, pp. 533–536.

[4] M. Smith, Neural networks for statistical modeling. New York: Van Nostrand Reinhold, 1993.

[5] J. G. Taylor, Neural networks and their applications. Wiley, 1999.

[6] H. White, Artificial neural networks: Approximation and learning theory, Oxford: Basil Blackwell, 1992.

[7] D. M. Parkin, P. Pisani, J. Ferlay, Estimates of the worldwide incidence of twenty five major cancers in 1990. Int J Cancer, Vol.80, 1999, pp827–41.

[8] S. R. Aziz, Oral submucous fibrosis: an unusual disease, J N J Dent Assoc, Vol. 68, 1997, 17–19.

[9] J. P. Canniff, W. Harvey, M. Harris, Oral submucous fibrosis: its pathogenesis and management, Br Dent J, Vol.160, 1986; pp.429–34.

[10] H. A. Seedat, C.W. van Wyk, Betel-nut chewing and submucous fibrosis in Durban, S Afr Med J, Vol.74, 1988, pp.568–71.

[11] R. Maher, A. J. Lee, K. A. Warnakulasuriya KA, Role of areca nut in the causation of oral submucous fibrosis: a case–control study in Pakistan, J Oral Pathol, Vol.23, 1994, pp.65–9.

[12] R. B. Zain, N. Ikeda, I. A. Razak, A national epidemiological survey of oral mucosal lesions in Malaysia, Community Dent Oral Epidemiol, Vol. 25, 1997, pp.377–83.

[13] W. McCulloch, W. Pitts, A Logical Calculus of Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, Vol.5, No.4, 1943, pp.115–133. doi:10.1007/BF02478259.

[14] D. Hebb, The Organization of Behavior, New York: Wiley, 1949.
[15] B. Farley, W.A. Clark, Simulation of Self-Organizing Systems by Digital Computer, IRE Transactions on Information Theory, Vol.4, No.4, 1954, pp. 76–84. doi:10.1109/TIT.1954.1057468.
[16] N. Rochester, J. H. Holland, L.H. Habit, W. L. Duda, Tests on a cell assembly theory of the action of the brain, using a large digital computer, IRE Transactions on Information Theory, Vol. 2, No.3, 1956, pp.80–93. doi:10.1109/TIT.1956.1056810.
[17] F. Rosenblatt, The Perceptron: A Probalistic Model For Information Storage And Organization In The Brain, Psychological Review, Vol.65, No.6, 1958, pp.386–408. doi:10.1037/h0042519.
[18] M. D. Rosmai, A. K. Sameemii, A. Basir, I. S. Mazlipahiv, M. D. Norzaidi, The Use of Artificial Intelligence to Identify People at Risk of Oral Cancer: Empirical Evidence in Malaysian University, International Journal of Scientific Research in Education, Vol.3, No.1, 2010, pp.10-20
[19] M. D. Rosmai, A. Basir, S. A. Kareem, S. M. Ismail, M. D. Norzaidi, Determining the Critical Success Factors of Oral Cancer Susceptibility Prediction in Malaysia Using Fuzzy Models, Sains Malaysiana, Vol.41, No.5, 2012, pp.633–640.
[20] D.S.V.G.K. Kaladhar, B. Chandana, P.B. Kumar, Predicting cancer survivability using Classification algorithms, International Journal of Research and Reviews in Computer Science, Vol.02, No.02, 2011, pp.340–343.
[21] J. O. Kang, S. H. Chung, Y. M. Suh, Prediction of Hospital Charges for Cancer Patients with Data Mining Technique, J Kor Soc Med Informatics, Vol.15, No.1, 2009, pp.13-23.
[22] T. R. Schultz and F. Rivest. Using Knowledge to Speed Learning: A Comparison of Knowledge-based Cascade-correlation and Multi-task Learning, Connection Science, Vol.13, 2001, pp. 1–30.
[23] V. Schetinin, A Learning Algorithm for Evolving Cascade Neural Networks, Neural Processing Letter, Vol.17, 2003, pp.21-31, 2003.
[24] J. A. Freeman, Simulating neural networks with mathematica. Reading, MA: Addison-Wesley, 1994.
[25] J. Hertz, A. Krogh, R. G. Palmer, Introduction to the theory of neural computation, Santa Fe Institute Studies in the Sciences of Complexity (vol. 1). Redwood City, CA: Addision-Wesley. 1991.
[26] S. E. Fahlman and C. Lebiere, The cascade-correlation learning architecture, Technical Report CMU-CS-90-100, School of Computer Science, Carnegie Mellon University, 1991.
[27] M. Hoefeld and S. E. Fahlman, Learning with limited numerical precision using the cascade-correlation algorithm, Technical Report CMU-CS-91-130, School of Computer Science, Carnegie Mellon University, 1991.
[28] S. E. Fahlman, The recurrent cascade-correlation architecture, Technical Report CMU-CS-91-100, School of Computer Science, Carnegie Mellon University, 1991.
[29] T. R. Schultz, F. Rivest, L. o Egri, J. P. Thivierge, Knowledge-based Learning with KBCC, Proceedings in the 5th International Conference on Development and Learning, 2006.
[30] S. Geisser, Predictive Inference, New York, NY: Chapman and Hall, 1993, ISBN 0-412-03471-9.
[31] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Vol.2, No.12, 1995, pp.1137–1143.
[32] P. A. Devijver, J. Kittler, Pattern Recognition: A Statistical Approach, Prentice-Hall, London, GB, 1982.
[33] N. Sharma, H. Om, Framework for early detection and prevention of oral cancer using data mining, International Journal of Advances in Engneering and Technology, Vol.4, No.2, 2012, pp.302-310.
[34] www.dtreg.com
[35] Swets, John A., Signal detection theory and ROC analysis in psychology and diagnostics : collected papers, Lawrence Erlbaum Associates, Mahwah, NJ, 1996.