International Journal of Current Research and Modern Education (IJCRME)
Impact Factor: 6.925, ISSN (Online): 2455 - 5428
(www.rdmodernresearch.com) Volume 3, Issue 1, 2018

A COMPARATIVE STUDY OF ARTIFICIAL NEURAL NETWORK ARCHITECTURES AND SUPPORT VECTOR MACHINE USING CANCER DATA

S. Prathap*, V. Vallinayagam & P. Venkatesan*****
* Department of Mathematics, Jeppiaar Engineering College, Chennai, Tamilnadu
** Department of Mathematics, St. Joseph's College of Engineering, Chennai, Tamilnadu
*** Department of Statistics, Sri Ramachandra Medical College and Research Institute, Sri Ramachandra University, Chennai, Tamilnadu

Cite This Article: S. Prathap, V. Vallinayagam & P. Venkatesan, "A Comparative Study of Artificial Neural Network Architectures and Support Vector Machine Using Cancer Data", International Journal of Current Research and Modern Education, Volume 3, Issue 1, Page Number 63-67, 2018.

Abstract:

Accurate diagnosis of cancer plays an importance role in order to save human life. From the literature, it has been found that Artificial Intelligence (AI) machine learning classifiers such as an Artificial Neural Network (ANN) and Support Vector Machine (SVM) can help doctors in diagnosing cancer more precisely. Both of them have been proven to produce good performance of cancer classification accuracy. The purpose of this study is to compare the performance of the Artificial Neural Network and Support Vector Machine for breast cancer dataset. The performance of both models is evaluated using four different measuring tools which are sensitivity, specificity, accuracy and Area Under receiving operating characteristic Curve (AUC). The comparative result shows that the Support Vector Machine classifier outperforms Artificial Neural Network classifier.

Key Words: Support Vector Machine, Artificial Neural Network, Classification, Breast Cancer & Accuracy

1. Introduction:

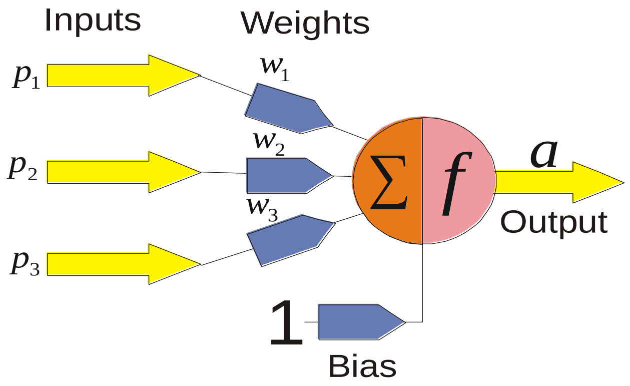
Breast cancer is the most widely recognized tumor in ladies in numerous nations. Most breast cancers are distinguished as a lump on the breast, or through mammography (DeSilva et al., 1994). Screening mammography is the best apparatus accessible for distinguishing malignant injuries previously clinical manifestations show up. The greater part of the scientists these days had proposed Artificial Intelligence (AI) characterization procedures for cancer determination. These methods have been demonstrated in helping the specialists to encourage their basic leadership process (Polat et al., 2007). There are many sorts of AI order procedures have been utilized for cancer determination, for example, Genetic Algorithm, Fuzzy Set, Artificial Neural Network, Support Vector Machine and Rough Set. Classifiers are utilized to group the disease information as benign tumors or malignant tumors. As of late, ANN and SVM are the classifiers that have been generally utilized by the analysts for growth order because of their great arrangement precision execution (Keyvanfard, 2011; Cinar et al., 2008). The objective of this paper is to compare the performance of Artificial Neural Network (ANN) and Support Vector Machine (SVM) using breast cancer dataset.

2. Materials and Methods:

2.1 Artificial Neural Network (ANN):

Artificial neural networks, at first developed to imitate basic biological neural systems, the human brain particularly, are composed of a number of interconnected simple processing elements called neurons or nodes. Each node receives an input signal which is the total information from other nodes or external stimuli, processes it locally though an activation or transfer function and produces a transformed output signal to other nodes or external outputs. Each individual neuron implements its function rather slowly and imperfectly, collectively a network can perform a surprising number of tasks quite efficiently Reilly and Cooper [15]. This information processing characteristic make Artificial Neural Networks a powerful computational device and able to learn from examples. First studies of neural networks were done in 1942 by McCullough and Pitts. After sometime, Rosenblatt conceived in 1959 the first learning algorithm, creating a model known as the preceptor, which was then only a solution to simple linear problems. In 1974, Werbos reported, the first non-linear processing capabilities of Artificial Neural Networks.

In Multilayer preceptor, the weighted sum of the inputs and bias terms are passed to activation level through a transfer function to produce the output, and the units are arranged in a layered feed –forward topology called Feed Forward Neural Network (FFNN). The diagrammatic representation of FFNN is given in Fig.1. An Artificial neural network has three layers: input layer, hidden layer and output layer. The hidden layer increases the learning power of the Multi layer preceptor. The activation or transfer function of the network modifies the input to give a desired output. The transfer function is selected by the algorithm requires a response function with a continuous, single-valued with first derivative existence. Choice of the number of hidden layers, hidden nodes and type of activation function acting an essential position in model building, Hecht-Nielsen [9] and White [4].



$$a = f(p_1 w_1 + p_2 w_2 + p_3 w_3 + b) = f\left(\sum p_i w_i + b\right)$$

Figure 1: Feed forward neural network

2.2 Support Vector Machine (SVM):

The Support Vector Machines algorithm developed by Vapnik is based on the structural risk minimization (SRM) principle is one of the most widely use supervise learning algorithm, and achieves superior generalization performance for both classification and regression problems (Christopher J. C. Burges, 1998; Smola and Alex J. Smola and Bernhard Scholkopf, 2004; Vapnik, 1995). SVM deal with the classification problem by finding the hyperplane in the feature space that achieves maximum sample margin when the training samples are separable. SVM has successful applications in many complex, real-world problems such as handwriting recognition, object recognition, data mining, bio informatics, medicines, financial forecasting an stock market.

2.2.1 Linear SVM:

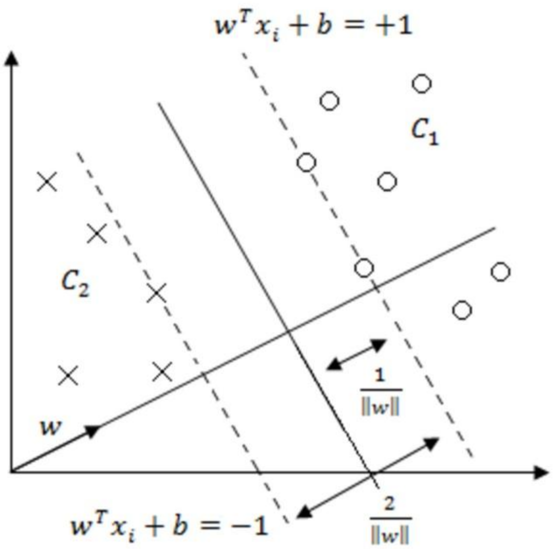


Figure 2: Linear SVM

Consider the training set D be (x_i, y_i), $i = 1,2,..n$, where $x_i \in R^n$ and the output label $y_i \in \{+1, -1\}$ for the hyperplane $w^T.x + b = 0$, which separates into two classes by satisfying the constraints $(w^T.x_i + b) \geq 1, \forall\, i$ where $y_i = 1$ and $(w^T.x_i + b) \leq -1, \forall\, i$ where $y_i = -1$. Combining both the constraints, $y_i(w^T.x_i + b) \geq 1, \forall\, i = 1,2,..n$. The distance between two hyperplanes, $w^T.x_i + b = 1$ and $w^T.x_i + b = -1$ is $\frac{2}{\|w\|}$ and is known as margin of the classifier. Hence optimization problem which maximizes the margin

$$\text{Minimize} = \frac{1}{2}\| w \|^2 \text{ subject to } \quad y_i(w^T.x_i + b) \geq 1, \forall\, i = 1,2,..n.$$

With Lagrangian multiplier α_i, the objective function becomes

$$\text{Minimize} = \frac{1}{2}\| w \|^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^T.x_i + b) - 1) \text{ subject to } \alpha_i \geq 0$$

The Lagrangian dual formulation of the above is

$$\text{Maximize } L_D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=0}^{n}\sum_{j=0}^{n} \alpha_i \alpha_j y_i\, y_j {x_i}^T x_j \text{ subject to } \alpha_i \geq 0 \text{ and } \sum_{i=1}^{n} \alpha_i y_i = 0.$$

The decision function is $f(x) = sign(w^T.x + b)$. Optimization problem () will not have a solution if D is not linearly separable. To deals with such cases, the slack variables ξ_i's are introduced which measure the degree of misclassification. The optimization problem for soft margin SVM is

$$\text{Minimize} = \frac{1}{2}\parallel w \parallel^2 + C\sum_{i=1}^{n}\xi_i \text{ subject to } \quad y_i(w^T.x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall\, i$$

Where C is a regularization parameter that determines the tradeoff between the margin size and training error. Let α_i's be the Lagrangian multipliers for $y_i(w^T.x_i + b) - 1 + \xi_i \geq 0$ and μ_i's be the Lagrangian multipliers for $\xi_i \geq 0$. The Lagrangian primal objective function is

$$\text{Minimize } L_P(\alpha) = \frac{1}{2}\parallel w \parallel^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i[y_i(w^T.x_i + b) - 1 + \xi_i] - \sum_{i=1}^{n}\mu_i\xi_i$$

The dual of this is

$$\text{Maximize } L_D(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=0}^{n}\sum_{j=0}^{n}\alpha_i\alpha_j y_i\, y_j x_i^T x_j \text{ subject to } 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{n}\alpha_i y_i = 0.$$

Which is same as () except the upper bound for the Lagrangian multipliers. In this case, the value of w is same and b can be found using Karush Kuhn Tucker conditions for the primal. Therefore optimal hyperplane $w^T.x + b = 0$ has been constructed. The decision function is

$$f(x) = sign(w^T.x + b) = sign(\sum_{i\in SV} \alpha_i y_i\, x_i.x + b).$$

2.2.2 Non Linear SVM:

To solve a non-linear classification problem with a linear classifier all we have to do is to substitute $\phi(x)$ instead of x everywhere x appears in the optimization problem. The mapping is done through the functions called kernels.

$$\text{Maximize } L_D(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=0}^{n}\sum_{j=0}^{n}\alpha_i\alpha_j y_i\, y_j\, K(x_i.x_j) \text{ subject to } 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{n}\alpha_i y_i = 0.$$

Where $K(x_i.x_j) = \phi(x_i^T)\,\phi(x_j)$ and the decision function is

$$f(x) = sign(w^T.x + b) = sign(\sum_{i\in SV} \alpha_i y_i\, K(x_i.x) + b).$$

2.3 Performance Evaluation Tools:

There are a several measuring tools to assess the execution of the classifiers that have been proposed. They are sensitivity, specificity, accuracy and area under receiving operating characteristic curve (AUC). Formula for measuring tools are given below (Keyvanfard F., M. A. Shoorehdeli, and M. Teshnehlab , 2011; Ren J, 2012; Subashini T. S., V. Ramalingam, and S. Palanivel , 2009; Azmi M. S., and Z. C. Cob 2010)

$$Sensitivity\ (\%) = \frac{TP}{FN + TP} \times 100$$

$$Specificity\ (\%) = \frac{TN}{TN + FP} \times 100$$

$$Accuracy\ (\%) = \frac{TP + TN}{TP + FN + TN + FP} \times 100$$

$$AUC\ (\%) = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \times 100$$

3. Application to Cancer Data:

Breast cancer dataset are obtained from the UCI Machine Library Database. The breast cancer dataset which is Wisconsin Breast Cancer Database (WBCD) is given by W. Nick Street (1995) from University of Wisconsin. The dataset comprise of 683 samples. The execution of the proposed method was tested and assessed utilizing Wisconsin Breast Cancer Database. Constructed models efficiency was evaluated by sensitivity, specificity, accuracy and Area Under receiving operating characteristic Curve (AUC) for datasets. The execution of the classifiers, Matlab Neural Network Toolbox is utilized to create ANN classification model and LIBSVM package is implemented in Matlab to build up the SVM classification model.

4. Results and Discussion:

In this study, two distinctive classification models have been built by utilizing two different classifiers which is ANN and SVM. The best classifiers for the dataset are resolved in view of four measuring tools, sensitivity, specificity, accuracy, and AUC. The obtained results are shown in Table 1 and Figure 1.

Table 1: Summary of the results obtain on WBCD

Measuring Tools	ANN	SVM
Sensitivity (%)	99.24	99.26
Specificity (%)	97.23	100.00
Accuracy (%)	98.53	99.52
AUC (%)	98.23	99.64

Accuracy is utilized to estimate how viable the classifier is by demonstrating the level of the true value of the class label. For this situation, the accuracy of SVM is better (99.52%) than ANN (98.53%). It implies that SVM classifier could correctly classified more data than ANN classifier. ANN and SVM accomplished estimation of sensitivity 99.24% and 99.26%. SVM accomplished 100% in specificity contrasted with ANN

(97.23%). AUC assesses the capacity of the classifiers to effectively characterize the true positive (benign tumors) and true negative (malignant tumors) classes. It can be seen that the AUC estimation of SVM is vastly improved than ANN with level of 99.64%. The AUC estimation of ANN is just 98.23%. Based on the AUC esteems acquired by the two classifiers, one might say that SVM classifier could decide the class of the data superior to ANN.

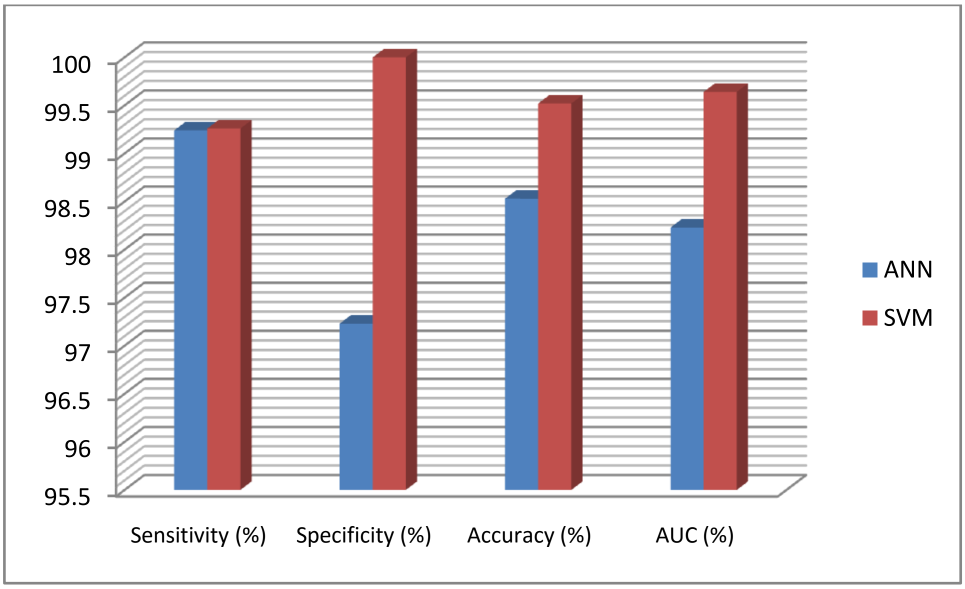


Figure 1: Performance of Classifiers on WBCD

5. Conclusion:

In this research, the execution of ANN and SVM has been analyzed in characterizing the Wisconsin Breast Cancer Database (WBCD). Results indicate that SVM classifier gives preferred execution over ANN for WBCD as far as sensitivity, specificity, accuracy an AUC value; 99.26%, 100% 99.52% and 99.64%. This work demonstrates that SVM can be adequately used to help the medical specialists to analyze breast cancer. In the future study, different training rules can be used for training ANN while SVM classifier can also be train by using different kernel functions in order to improve the performance of the classifiers

6. References:

1. Alex J. Smola and Bernhard Scholkopf (2004), A tutorial on Support Vector Regression, Statistics and Computing, 14, 199-222.
2. Azmi M. S., and Z. C. Cob (2010), Breast Cancer Prediction Based on Backpropagation Algorithm. Student Conference on Research and Development (SCOReD). 164–168.
3. Cinar, M. Engin, E. Z. Engin, and Y. Z. Atesci (2008), Early prostate cancer diagnosis by using artificial neural networks and support vector machines, Expert Systems with Application, vol. 36, pp. 6357-6361.
4. DeSilva and Choong (1994), “A Network Architecture for Maximum Entropy Estimation”, 1994 IEEE International Conference on Neural Networks, Orlando, Florida.
5. Hecht-Nielsen R (1990), Neuro Computing. Addisson-Wesley, Reading MA.
6. Keyvanfard, M. A. Shoorehdeli, and M. Teshnehlab (2011), Feature selection and classification of breast cancer on dynamic magnetic resonance imaging using ANN and SVM, American Journal of Biomedical Engineering, vol. 1, pp. 20-25.
7. Nick Street W. (1995), UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
8. Polat, S. Sahan, H. Kodaz, and S. Gunes (2007), Breast cancer and liver disorders classification using artificial immune recognition system (AIRS) with performance evaluation by fuzzy resource allocation mechanism, Expert System with Applications, vol. 32, pp. 172-183.
9. Prathap, S., Vallinayagam, V. and Venkatesan, P. (2016), A Comparative Study of Artificial Neural Network Architectures and Proportional Hazard Model using Heat Attack Data, Journal of Chemical and Pharmaceutical Sciences, Vol.9, No. 3, 2339-2343.
10. Ren J. (2012), ANN vs. SVM: Which One Performs Better in Classification of MCCs in Mammogram Imaging. Knowledge-Based Systems. 26: 144–153.
11. Reilly D. L and Copper, L. N. (1990), An Overview of neural network: Early models to real world system, in An Introduction to neural and Electronic Networks, Academic Press, San Dieg, p.p. 227-24.
12. Subashini T. S., V. Ramalingam, and S. Palanivel (2009), Breast Mass Classification Based on Cytological Patterns using RBFNN and SVM. Expert Systems with Applications. 36: 5284–5290.

13. Vallinayagam, V., Prathap, S., and Venkatesan, P. (2014), Parametric Regression Models in the Analysis of Breast Cancer Survival Data, International Journal of Science and Technology, Vol. 3, No. 3, 163-166.
14. Vallinayagam, V., Prathap, S., and Venkatesan, P. (2014), Non-Linear Regression Models for Heart Attack Data – An emprical comparison, Indian Journal of Applied Research, Vol.4, No. 6, 332-334.
15. Vapnik, V.N. (1995), The Nature of Statistical Learning Theory, 1st ed., Springer-Verlag, New York. White. H (1992), Artificial Neural Networks. Approximation and Learning Theory, Blackwell, Cambridge, MA.