Comparative Analysis of Machine Learning Algorithms for Binary Classification

SARISH ABID *, BASHARAT MANZOOR, WAQAR ASLAM, SAFEENA RAZAQ

Department of Computer Systems Engineering, Mirpur University of sciences and technology, Mirpur, AJK
**Corresponding author's email: sehrish.cse@must.edu.pk*

Abstract

Machine learning algorithms are applied in all domains to achieve classification tasks. Machine Learning is applicable to several real life problems. Aim of this paper is highly accurate predictions in test data sets using machine learning methods and comparison of these methods to select appropriate method for a particular data set for binary classifications. Three machine learning methods Artificial Neural Network (Multi-Layer Perceptron with Back Propagation Neural Network), Support Vector Machine and K-Nearest Neighbor are used in this research work. The data sets are taken from UCI website. A comparative study is carried out to evaluate the performance of the classifiers using statistical measures e.g. accuracy, specificity and sensitivity. These results are also compared with previous studies. Experimental outcomes show that the Artificial Neural Network method provides better performance, and it is strongly suggested that the Multi-Layer Perceptron with Back Propagation Neural Network method is reasonably operational for the task of binary classification followed by Support Vector Machine and K-Nearest Neighbor.

***Keywords*:** Artificial Neural Network, Classification algorithms, K-Nearest Neighbor, Machine Learning, Binary classification, Support Vector Machines

INTRODUCTION

Machine learning is a type of learning in which a system can learn with some input data and tries to improve its performance that is measurable with the help of some measures. The motivation of using machine learning is as follows: There are many tasks that require an adaptive system that can learn e.g. handwriting recognition, speech recognition etc. Learning is also useful as an alternative to hand-coding a program. For example, if one wants to develop a program which can play the game of chess, he can hand code all the rules or conditions necessary to play chess game. An alternative way of writing such programs would be to provide a system with the database of chess game and their outcomes. The system can apply learning methods to learn to play a good game of chess without explicitly telling the system which move to take in which situation. But provision of a large database is necessary from which the system can decide. Provision of database is usually easier than hand-coding the rules. So one can save a lot of manual effort if system is able to learn itself. Different types of machine learning methods can be used in variety of ways e.g. supervised learning, unsupervised learning and reinforcement learning. In our research work, we have used supervised learning in which a system is given labeled training examples (inputs+ outputs).

Classification, also called concept learning, consists of learning a description of a class of objects. This description is typically used to predict whether new objects fit the class or not. It has been of keen interest for researchers of computer engineering. In mammogram analysis: we are given a mammogram, and we want to classify the mammogram as normal, cancerous or pre-cancerous and in document understanding: we are given a rectangular region from a scanned image (region) and we want to be able to say whether this is a text region or graphic region.

Numerous classification techniques have been used in past for classification of SPECT heart, diabetes, banknote authentication, liver disorder, and radar signals returned from ionosphere data in the literature. For diabetes detection, Deng and Kasabov used 10-fold cross-validation (FC) using ESOM and attained classification accuracy of 78.4% [1]. Aslam and Nandi used genetic programming (GP) and a variation of genetic programming and gained classification accuracy of 78.5±2.2% [2]. The classification algorithms Naïve Bayes classifier, C4.5, BPNN algorithm, and SVM for the classification of some liver patient datasets were evaluated by Ramana [3]. Gulia *et al.*, also evaluated the classification algorithms Multi-Layer Perceptron, SVM, Random Forest and Bayesian Network for the classification of liver Patient datasets [4]. SVM algorithm is considered as the better performance algorithm, because it gives higher accuracy irrespective of other classification algorithms before applying feature selection. Random Forest algorithm outperformed all other techniques with the help of feature selection with an accuracy of 71.8696%. Asadi *et al.*, used a new Supervised Feed Forward Multi-layer Neural Network (SFFMNN) model [5]. In SPECTF Heart, the accuracy of Clip3 and Clip4 was 77%. Shao *et al.*, obtained 81.3874.59% accuracy [6]. Radar target identification is a very rarely explored domain. Samb achieved prediction rate of 86.29% on ionosphere dataset [7]. Pujari and Gupta obtained a greater accuracy of 93.84% using ensemble model with feature selection [8]. Mohamad *et al.*, used the technique of Artificial Neural Network using back-propagation algorithm for classification of banknote authentication dataset [9]. Ghazvini *et al.*, used Bank Notes dataset and compare Naïve Bayes and Multilayer Perceptron using the classification technique [10]. The results in his study were obtained using Naïve Bayes and Multilayer Perceptron with accuracy of 87.43 and 95.21% respectively.

Durodola *et al.*, used Artificial Neural network algorithm for the prediction of damage caused by random fatigue loading and get very encouraging results [11]. To model roadway traffic noise Hamad *et al.*, used ANN algorithm and get good performance than other existing methods so far in the literature [12]. Evaluation of exhaust emission of the engine was carried out by Celebi *et al.*, using ANN for the prediction of sound pressure level and vibration of the engine [13]. Results showed that generated model was capable for the estimation of parameters with high accuracy. Liu *et al.*, used SVM algorithm for multi class sentiment classification and concluded that this method is significantly better than the others for multi class sentiment classification [14]. Cholette and Borghesani addresses the problem of estimating continuous boundaries between acceptable and unacceptable engineering design parameters in complex engineering applications and used SVM algorithm for this purpose and get very promising results [15]. In order to predict the different types of road surfaces based on tire cavity sound acquired under normal vehicle operation Masino *et al.*, applied SVM algorithm that is comprehensive and inexpensive provides good accuracy [16]. Chen and Hao used both SVM and KNN algorithm for stock market indices prediction [17]. Mohammed *et al.*, used KNN for the solution of vehicle routing problem and get optimal route result [18]. Faziludeen and Sankaran used Evidential K nearest neighbour for ECG beat classification and proved that EKNN system outperforms [19]. Munisami *et al.*, used KNN algorithm based recognition system capable of identifying plants by using the images of their leaves and get accuracy of 87% [20].

This research has been conducted with the objective to examine and evaluate the existing machine learning techniques and find out which technique provides better results for which problem dataset. It presents comparative analysis to help researchers to solve the real world complex binary classification problems more efficiently. Complexity is also calculated and presented in terms of training time. The machine learning classifiers used in this research are evaluated by applying on real world classification problems of diabetes, liver and heart diseases detection, ionosphere data and banknote authentication classification. Some problem datasets had no missing values, and none of the feature values was categorical. Although several comparisons are available for these algorithms in previous studies but to the best of our knowledge these algorithms are being used for the first time for the problem datasets. These problems are very challenging in near future. More than several million people around the world are suffering from diabetes, liver disorders and heart disease. In immediate future large number of physicians would be needed if this rapid rise of diabetes carries on. Now, it is the need of time to use classifier system in medical diagnosis. If some attributes of a patient are available a classifier system won't need physician to figure that person is affected or not. If assessments made by physician for past patients having same conditions are stored then a classifier system can be developed to make use of the stored conditions according to the stored assessments. This will benefit physician greatly keeping the significance of the expert opinion in disease diagnosis. In addition to these, counterfeiting has a past history and will continue in the future as well. All of us are being affected by the counterfeiting of banknotes. The reprographic technologies have been developed increasing the threat of counterfeiting. So a classifier system is also needed for banknote authentication. Machine learning algorithms used in this research are introduced briefly in next sections.

MACHINE LEARNING ALGORITHMS

K-Nearest Neighbor rule

In 1950, a new classification algorithm K nearest neighbor was introduced by Fix and Hodges [21]. This algorithm is pretty different from other algorithms because there is no training case at all, so the algorithm gets the training and testing set at the same time. Input is a set of training examples $\{x_i, y_i\}$ where x_i is the set of attribute value pairs from the i^{th} instance and y_i is the label and if we are doing the classification it is the class label {ham or spam} or if we are recognizing the digits it's the value between 0 to 9. Testing point is x that we want to classify. The algorithm works by taking the point x and computing the distance $D(x, x_i)$ to every training example x_i. Out of those training examples it picks k instances which are closest to x_{i1} to x_{ik} and looks at their labels y_{i1} to y_{ik} and picks the label which is most frequent in that set of labels.

Support Vector Machines

In 1968, Vapnik and Chervonenkis introduced the concept of Support Vector Machines (SVMs) [22] . This algorithm can also be used for classification. Let's say, for linearly separable binary sets we have a two dimensional plane with two classes of objects and we want to put a border between them. The goal of SVM is to design a hyperplane that classifies all training vectors in two classes. A hyperplane can be represented by a normal vector and a scalar. Normal vector determines the orientation. The bias, on the other hand controls the displacement from the origin. The margin can be described by using two hyperplanes, by changing the angle of normal vector, we can rotate the margin or if we want to shift, we can increase or decrease the bias. The best choice will be the hyperplane that leaves the maximum margin from both classes. The margin is the distance between the hyperplane and the closest elements from the hyperplane. so the hyperplane for which margin is higher is selected [23].

Artificial Neural Networks

Neural Network (NN) is a system which has been inspired biologically. In 1943, McCulloch and Pits are generally recognized as the designer of the first neural network [24]. First learning rule for NN was devised in 1949 by Hebb [25]. Minsky and Papert published a paper in 1969 in which they highlighted the computational limitation of Perceptron unit [26]. This leads to a virtual decline in the research work in NN. Fortunately, in 1980's called re-emergence of interest in NN & many researches came up with more complex architecture in the form of multi-layer networks. That overcame the limitation of Perceptron unit. Today research in the area of NN is active and these are being used in a variety of applications.

In Artificial Neural Network (ANN), we have a network of simple processing elements which are connected to each other via weighted links. Inputs are fed to the input unit and as a result computations done in this unit and the outputs are produced. NN has variety of applications. It has been used for recognizing hand-written letters, for predicting online the quality of welding spots, for identifying relevant documents with in a carpus (large no of documents), for visualizing high-dimensional of space then tracking online the position of robot arms.

It can be divided into three main types depending upon how an ANN partitions the data into different classes: Multi-Layer Perceptron (MLP), Radial Basis Function Network (RBF) and Probabilistic Neural Network (PNN). In this research only the first type is used and interested readers can read details about other types in [27] .

Back propagation Neural Networks (BPNNs) is based upon a technique that works using supervised learning called Back propagation learning. Commonly it is named as the Feed Forward Back Propagation Neural Network (FFBPNN). With respect to architecture it is mainly a Multi-layer Perceptron. The BPNN was the gemstone that charmed and fascinated researchers and revealed the true influence of NN. It opened research flaps with never-ending opportunities in numerous fields of sciences, engineering and statistics; and it is computationally efficient. But on the darker side the Back propagation NN has also been named as the 'black box' (as we cannot interpret easily the rules that NN learns) as it has a fixed algorithmic operation only with no fixed topology (number of neurons and nodes used) for it. Irrespective of all these aspects overall the BPNN is relatively accurate and easy to work with respect to other neural networks.

EXPERIMENT

This section firstly describes the datasets used for experimental work. Secondly it explains the experiments conducted to solve the problems.

Dataset Description

Table of examples or instances are used to represent the data in supervised machine learning task. Fixed numbers of measurements, or features, are used along with a label that denotes its class to describe each instance. Features which are also called attributes are of two types namely nominal and numeric. Nominal attributes are unordered categories. Numeric data consists of real numbers. Two sets of examples are required for application of a machine learning algorithms i.e. training and test examples. Learned concept descriptions are produced by use of training examples set. To evaluate the accuracy, test examples set is required. Class labels are missing in testing phase. A class label is produced as an output when the algorithm is applied with the test example as input. In this paper, all the problem datasets are taken from UCI machine learning repository Blake and Merz [28] and details are given in Table 1.

Table 1: Dataset Details

Dataset Name	Total Samples	Training Samples 70%	Testing Samples 30%	Features
Ionosphere	1352	246	105	34
PIMA	768	538	230	8
SPECT Heart	267	187	80	44
Banknote Authentication	1372	960	412	4
ILPD	583	409	175	10

Performance Measurement

Three statistical measures are used to calculate the performance of each classification method i.e. Specificity, sensitivity and accuracy. (TN) true negative, (TP) true positive, (FN) false negative and (FP) false positive cases are used to describe these measures. Suppose, we take a test of some people for verification of some disease. The term true positive is for the case if the test results are positive and the resultant people have the disease. False negative is for the case when some of them are infected with the disease but test results show they are clear. The term true negative is for the case if the test results are negative and the resultant people are not affected with that disease**.** At last the people who are not affected with the disease and are healthy but test results is positive, is termed as false positive. FN, FP, TN, TP cases are shown in Table 2.

Table 2: Confusion Matrix for Actual and Predicted Cases

	P'(predicted)	**N'(predicted)**
P(Actual)	True Positive	False Negative
N(Actual)	False Positive	True Negative

Specificity

The capability of the system of predicting the accurate values for the cases that are the opposite of the desired one is called as specificity. In short, it measures the proportion of the true negatives. Specificity can be calculated using the equation below:

$$SPEC = Negative\ hits\ /\ Total\ negatives = \mathrm{TN}/\ (\mathrm{FP} + \mathrm{TN})\ (1)$$

Sensitivity

The capability of the system on predicting the accurate values in the cases presented is called as sensitivity. In short it may be defined as the measures the proportion of the true positives. Sensitivity can be calculated using the equation below.

$$SENS = Positive\ hits\ /\ Total\ Positives = TP\ /\ (FN+TP)(2)$$

Classification Accuracy

Considering the positive and the negative inputs classification accuracy measures the proportion of correct predictions. Classification accuracy is dependent on the data set distribution, which can lead to incorrect conclusions regarding the system performance. Classification Accuracy can be calculated using below equation:

ACC=Total Hits /Total Number of entries in the set = (TP+TN)/ (P+N)(3)

RESULTS AND DISCUSSION

The experimental work is carried out on core i3 with 2GB RAM on windows platform using MATLAB R2011. In order to divide the data of the dataset we used 70: 30 % ratios respectively for training and testing. Training data is used to train the classification algorithms. Testing data is used to calculate the strength or proficiency of classification algorithms. Outcomes taken from classification algorithms are matched with true classes to distinguish true positives, true negatives, false positive and false negative values. We compute these values to build the confusion matrix [29]. Each cell contains the row number of samples classified for the corresponding combination of desired and actual model output. While studying the performance of each classification algorithm, all results are calculated over 100 runs. Training time is averaged for these classifiers. We checked the results for ANN for different sizes of hidden layer. We used Euclidean distance in order to compute the distance in K-nearest neighbor algorithm. For support vector machine algorithm, we checked the results for three different kernel functions like Polynomial, RBF, MLP with varying parameter values like polynomial order, value of sigma in RBF and MLP parameter.

Comparative Analysis

Tables 3, 4 and 5 show the results of classifiers:

Table 3: ANN Results

Dataset Name	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ionosphere	100	100	100
PIMA	91.73	87.7	94.28
SPECT Heart	100	100	100
Banknote Authentication	100	100	100
ILPD	90.85	93.79	82.60

Table 4: SVM Results

Dataset Name	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ionosphere	92.38	97	87
PIMA	75.21	62	81
SPECT Heart	81.25	77	90
Banknote Authentication	100	100	100
ILPD	66.66	75	45

Table 5: KNN Results

Dataset Name	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ionosphere	78.095	60	92
PIMA	78.26	48	92
SPECT Heart	77.50	100	14
Banknote Authentication	100	100	100
ILPD	72	90	27

From Table 3, 4 and 5, it is clear that all classifiers achieve 100% accuracy for banknote authentication dataset. These results are also consistent with previous findings [9]. It may be because of smaller feature set containing all the information required to discriminate genuine and forged banknotes. In addition to banknote authentication dataset, ANN also achieves 100% accuracy for SPECT Heart and Ionosphere dataset.

SVM is the second best for SPECT Heart and ionosphere dataset and it achieves 81.25% and 92.4% accuracies respectively while for the same datasets, KNN is able to achieve only 77.5% and 78.11% accuracies respectively. Both of these datasets have large number of features and the results shows that ANN handles these large datasets quite easily while SVM and KNN fail to do that.

ANN is still the best classifier for PIMA and ILPD datasets as it outperforms the other two classifiers. The performance of other two classifiers is not the same, where KNN performs better than SVM. The reason for low performance of PIMA and ILPD datasets for all the classifiers are that, these algorithms do not handle missing values effectively.

Moreover, ANN not only achieves better accuracy compared to other classifiers, it provides desirable stability between sensitivity and specificity while other classifiers fail to maintain that. These results prove that ANN is the best classifier for all these datasets.

General Comparison

An overall comparison of classifiers is presented as bar graph in Figures 1-3. According to this comparison, ANN is optimal classifier for binary classification problems in all aspects including accuracy, sensitivity and specificity irrespective of any particular problem dataset. Thus we can say that ANN is the best classifier for binary classification problems.

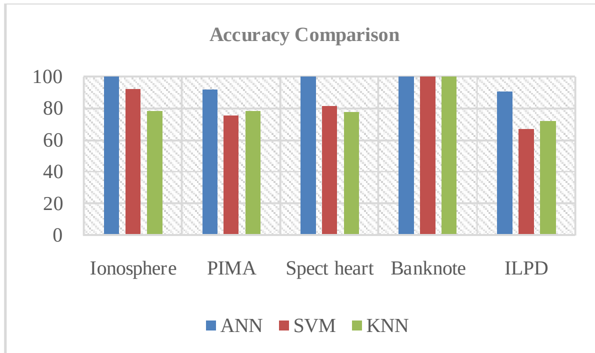

Figure 1: Accuracy Comparison for ANN, SVM and KNN

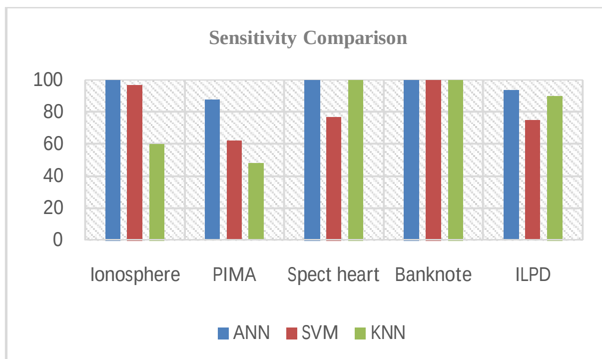

Figure 2: Sensitivity Comparison for ANN, SVM and KNN

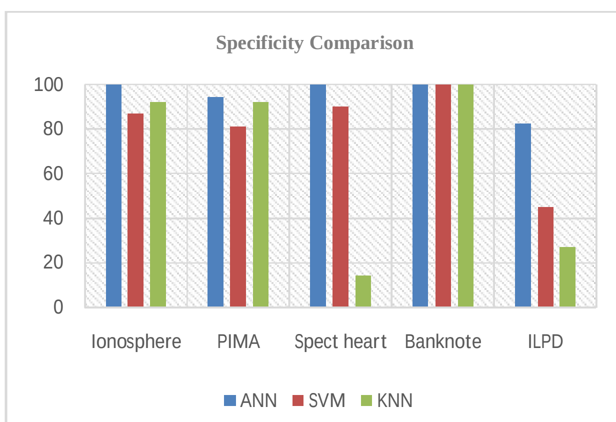

Figure 3: Specificity Comparison for ANN, SVM and KNN

In Table 6, complexity of classifiers in terms of training time is presented. Results show that classifiers used in this research are not complex. Based on training time ANN is complex and KNN is the simplest among these classifiers while SVM shows very high complexity only for liver patient dataset and for rest of dataset, its complexity is low.

Table 6: Complexity Comparison

Dataset	Average time for ANN(s)	Average time for SVM(s)	Average time for KNN(s)
Ionosphere	0.56	0.138	0.01
PIMA Indian Diabetes	4.87	0.30	0.027
Spect Heart	1.36	0.08	0.01
Banknote Authentication	0.5	0.39	0.022
ILPD	1.02	4.55	0.024

There are other methods which have been used for classification of these five datasets in the past. We compare our results with the results obtained so far in the previous literature. Table 7 gives the classification accuracies of previous methods where classification accuracy represents the percentage of instances correctly classified using test data.

Table 7: Comparison with Previous Work

Dataset	Algorithm	Accuracy (%)	Ref. No
PIMA Indian Diabetes	GP	78.5	[2]
ILPD	Random forest	71.86	[4]
SPECT Heart	Clip 3,4	77	[5]
Ionosphere	Ensemble	93.8	[8]
Banknote Authentication	Naive bayes	87.9	[10]
Our Results Accuracy (%)	**ANN**	**SVM**	**KNN**
PIMA Indian Diabetes	91.73	75.21	78.26
ILPD	90.85	66.66	72
SPECT Heart	100	81.25	77.50
Ionosphere	100	92.38	78.095
Banknote Authentication	100	100	100

It is clear from Table 7, KNN algorithm produced similar results as in previously reported datasets i.e., PIMA Indian Diabetes, Indian liver Patient and SPECT Heart datasets. SVM algorithm also produces similar results but only for ionosphere dataset. Apart from this, ANN results are much better than the results produced by SVM and KNN algorithms.

CONCLUSION

On the basis of several experiments, SVM and KNN algorithm show variation in results for different problem datasets due to size and attributes. Algorithm which performs better in terms of sensitivity and accuracy rate over a problem dataset has been considered as the best classification algorithm for that problem dataset. From results, it can be concluded that ANN is suitable for given tasks. ANN Classifier is optimal classifier for classification of all the datasets. Overall, ANN has achieved remarkable performance with highest accuracies followed by SVM and KNN. In all respects, ANN performs better, hence ANN is recommended for binary classification irrespective of any problem dataset.

As far as complexity is concerned, as mentioned in Table 6, on average ANN takes maximum five seconds, SVM takes less than five seconds and KNN takes only one second. It can be concluded that KNN performs faster and is not as complex as ANN and SVM. ANN's complexity increases with increasing hidden layer size that's why it takes more time in training. Overall training time for these classifiers is five seconds which is negligible.

RECOMMENDATIONS

For future work, it is suggested that due to excellent performance on all five datasets in this research, one can apply ANN on other binary classification problems and strengthen our conclusion after evaluation. The parameter response is not clear and can be further explored. While working with ANN, it is found that all the datasets perform differently for specific hidden layer size so we can't say that which size is good for all datasets. This area can be further explored. SVM is good for binary classification. We used three kernel functions in this research, so other functions can be explored. Euclidean distance is used in our study, so we suggest that other distances can be used in future research work.

REFERENCES

[1] D. Deng and N. Kasabov, "On-line pattern analysis by evolving self-organizing maps," in *Proceedings of the Fifth Biannual Conference on Artificial Neural Networks and Expert Systems (ANNES 2001)*, 2001, pp. 46-51. doi: 10.1016/S0925-2312(02)00599-4

[2] M. W. Aslam and A. K. Nandi, "Detection of diabetes using genetic programming," in *18th European Signal Processing Conference (EUSIPCO-2010)*, Aalborg, Denmark, August 23-27, 2010, pp. 1184-1188

[3] B. V. Ramana *et al.*, "A Critical study of selected classification algorithms for liver disease diagnosis," *Int. J. Database Manag. Syst.*, vol. 3, no. 2, pp. 101-114, May 2011. doi: 10.5121/ijdms.2011.3207

[4] A. Gulia *et al.*, "Liver patient classification using intelligent techniques," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 4, pp. 5110-5115, 2014. doi: 10.1.1.443.8662

[5] R. Asadi *et al.*, "New Supervised Multi Layer Feed Forward Neural Network Model to Accelerate Classification with High Accuracy," *Eur. J. Sci. Res.*, vol. 33, no. 1, pp. 163-178, 2009. doi: 10.1.1.454.8820

[6] Y. H. Shao *et al.*, "Least squares recursive projection twin support vector machine for classification," in *Pattern Recognit.* vol. 45, ed, 2012, pp. 2299-2307.

[7] M. L. Samb *et al.*, "A novel RFE-SVM-based feature selection approach for classification," *Int. J. Adv. Sci. Technol.*, vol. 43, pp. 27-36, 2012. doi: 10.1.1.641.826

[8] P. Pujari and J. B. Gupta, "Improving classification accuracy by using feature selection and ensemble model," *Int. J. Soft Comput. Eng.*, vol. 2, no. 2, pp. 380-386, 2012. doi: 10.1.1.650.2904

[9] N. S. Mohamad *et al.*, "Banknote authentication using artificial neural network," *Sci. Int.*, vol. 26, no. 5, pp. 1865-1868, 2014. Available: https://www.researchgate.net/publication/279205560

[10] A. Ghazvini *et al.*, "Comparative analysis of algorithms in supervised classification: A case study of bank notes dataset," *Int. J. Comput. Trends Technol.*, vol. 17, no. 1, pp. 39-43, 2014. doi: 10.14445/22312803/IJCTT-V17P109

[11] J. F. Durodola *et al.*, "A pattern recognition artificial neural network method for random fatigue loading life prediction," *Int. J. Fatigue*, vol. 99, pp. 55-67, June 2017. doi: 10.1016/j.ijfatigue.2017.02.003

[12] K. Hamad *et al.*, "Modeling roadway traffic noise in a hot climate using artificial neural networks," *Transp. Res. Part D Transp. Environ.*, vol. 53, pp. 161-177, June 2017. doi: 10.1016/j.trd.2017.04.014

[13] K. Celebi *et al.*, "Experimental and artificial neural network approach of noise and vibration characteristic of an unmodified diesel engine fuelled with conventional diesel, and biodiesel blends with natural gas addition," *Fuel*, vol. 197, pp. 159-173, June 2017. doi: 10.1016/j.fuel.2017.01.113

[14] Y. Liu *et al.*, "A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm," *Inf. Sci.*, vol. 394-395, pp. 38-52, July 2017. doi: 10.1016/j.ins.2017.02.016

[15] M. E. Cholette and P. Borghesani, "Using support vector machines for the computationally efficient identification of acceptable design parameters in computer-aided engineering applications," *Expert Syst. Appl.*, vol. 81, pp. 39-52, 15 September 2017. doi: 10.1016/j.eswa.2017.03.050

[16] J. Masino *et al.*, "Road surface prediction from acoustical measurements in the tire cavity using support vector machine," *Appl. Acoust.*, vol. 125, pp. 41-48, 2017.

[17] Y. Chen and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Expert Syst. Appl.*, vol. 80, pp. 340-355, 2017. doi: 10.1016/j.eswa.2017.02.044

[18] M. A. Mohammed *et al.*, "Solving vehicle routing problem by using improved K-nearest neighbor algorithm for best solution," *J. Comput. Sci.*, 19 April 2017. doi: 10.1016/j.jocs.2017.04.012

[19] S. Faziludeen and P. Sankaran, "ECG beat classification using evidential K-Nearest Neighbours," *Procedia Comput. Sci.*, vol. 89, pp. 499-505, 2016. doi: 10.1016/j.procs.2016.06.106

[20] T. Munisami *et al.*, "Plant leaf recognition using shape features and colour histogram with k-nearest neighbour classifiers," *Procedia Comput. Sci.*, vol. 58, pp. 740-747, 2015. doi: 10.1016/j.procs.2015.08.095

[21] E. Fix and J. Hodges, "Discriminatory analysis-nonparametric discrimination: consistency properties," Technical report, USAF School of Aciation Medicine 1951, doi: 10.2307/1403797.

[22] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," Measures of Complexity: Springer, 2015, pp. 11-30. [Online]. doi: 10.1007/978-3-319-21852-6_3

[23] M. W. Aslam, "Pattern recognition using genetic programming for classification of diabetes and modulation data," Doctoral dissertation, University of Liverpool, 2013. doi: 10.1.1.427.1195

[24] W. S. McCulloch and W. Pitts, "A Logical Calculus of the Idea Immanent in Nervous Activity," *Bull. Math. Biophys.*, vol. 5, pp. 115-133, 1943. doi: 10.1007/BF02478259

[25] D. O. Hebb, The Organization of Behavior: A Neuropsychological Theory, New York: Wiley & Sons, 1949. Available: https://books.google.com.pk/books?id=ddB4AgAAQBAJ

[26] M. L. Minsky and S. Papert, Perceptrons: An Introduction to Computational Geometry, 2nd ed. Cambridge: MIT Press, 1972. Available: https://books.google.com.pk/books?id=U-O9BHlTvJIC

[27] S. S. Haykin, Neural Networks: A Comprehensive Foundation: Macmillan, 1994. Available: https://books.google.com.pk/books?id=M5abQgAACAAJ

[28] C. L. Blake and C. J. Merz. UCI repository of Machine Learning Databases [Online]. Available: http://archive.ics.uci.edu/ml/

[29] J. Han *et al.*, Data Mining: Concepts and Techniques: Elsevier Science & Technology, 2011. Available: https://books.google.com.pk/books?id=pQws07tdpjoC