Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Classification of HRS using SVM

Astha Ameta[1] and Kalpana Jain[2]

[1] College of technology and engineering

Received: 11 December 2016 Accepted: 5 January 2017 Published: 15 January 2017

Abstract

The kidney diseases are one of the main causes of death around the world. Automatic detection and classification of kidney related diseases are important for diagnosis of kidney irregularities. Hepatorenal Syndrome (HRS) is a lifethreatening medical condition when kidney fails due to liver failure. The treatment to such cases is liver transplant, or dialysis for temporary basis. This paper proposed to apply the Support Vector Machine (SVM) classification for diagnosis of HRS. The results were evaluated using realistic data from hospitals. RBF kernel function is used along with SVM. The results show a significant accuracy of 95

Index terms— support vector machine, RBF kernel, cross validation, accuracy, ROC curve.

1 I. Introduction

epatorenal Syndrome (HRS) is a major complication of Cirrhosis, where approximately 8% patients with ascites are annually incident. HRS starts developing at the latest phase of disease. It is now medically proven that it is a very important determinant for showing survival rate. A majority of reviews on HRS reflect the problems in the investigation of this syndrome. On the contrary, HRS has no experimental model. Hence, many of its aspects are still poorly understood.

A high degree of predictive accuracy is needed in the healthcare sector. The predictive accuracy of any data mining/Machine learning technique is based on the data, its quantity and quality. Techniques such as classification, clustering, time series, temporal analysis, association and correlation analysis are various data mining techniques taken into consideration. Classification techniques are used to analyze data and predict labels that describe important properties of data. Many classification techniques have been developed such as Naïve Bayes, k-NN, SVM, Decision Tree induction, Back propagation, and more. Here, we propose SVM technique to be used for diagnosis of HRS.

2 II. Support Vector Machine

SVM, abbreviated as Support Vector Machine, is a class of learning methods that can be used for the purpose of classification. Many classifiers have been proposed in the literature to study classification problems. In training SVMs, decision boundaries are directly determined from training data thus maximizing its generalization ability. Hence, ability of SVM to generalize is somehow different than those of other classifiers, usually in the case of small number of training data. In its simplest or linear form, SVM is defined as a hyperplane which separates a set of negative examples from set of positive examples by using the concept of maximize the class margin. The form in which data points are provided is {(y 1 ,x 1), (y 2 ,x 2),?,(y n ,x n)}, where x i is a vector of n-dimensions and y i can either be 1 or -1, which denotes the class to which point x i belongs. For training SVM, set of x i are pre-labeled with y i components which denotes the correct classification which is required by SVM to search for a separating hyperplane.

For the case where data are linearly separable, two hyperplanes, w . x -b = -1 and w . x + b = 1 are generated which are parallel. Thus, no training sample lies in between and distance is maximized for the two planes. In the quadratic form, it can be formalized as: Min ½ ||w|| 2 Subject to y i (w . x i -b) ? 1, 1 ? i? l. This is a

convex problem. Its dual form is: min ½ ? T Q? ? e T ? subject to y T ? = 0 and ? ? 0, where Q is an l x l matrix with Q ij = y i y j x i . x j and e is the vector of all ones. Let ? be the solution to dual problem,then w = ? yi?i???? ?? ??=1

is a solution to the primal problem. Vectors x i ,which corresponds to ? i >0, lie on the margin. Such vectors are termed as support vectors (SV). Once the above equations are resolved, then new items can be classified with w .x where x is the new sample vector that is to be classified.

For the case of non-linear separable data, Cortes and Vapnik ([14]) proposed a modification to the QP formulation (namely soft margin) according to which, examples that fall on the wrong side of the decision boundary are allowed but with a penalty. Boser et al. **??**[15]) also proposed an extension to the non-linear classifiers. A generalized form of the QP problem having soft margin along with nonlinear classifier is shown below:min ½ ||w|| 2 + Cá¶?" T e, subject to y i (w ??(x i) ? b) ? 1 ? á¶?" i andá¶?" i ? 0, 1 ? i? l,

whereá¶?" shows the training error and the parameter C is used to adjust the training error and the regularization term 1/2 ||w|| 2 . The function ? maps ? n to a higher dimensional space. In practice, kernel functions are used to perform the process of mapping. The kernel functions are represented in the form of dot product as below:K(xi, xj) = ?(xi) ??(xj).

Some commonly used kernel functions include Linear:k(xi, xj) = xi ?xj Polynomial: k(x i , x j) = (x i ?x j) d Radial Basis Function (RBF): k(x i , x j) = exp(??||x i -x j || 2),?> 0 III. Proposed Methodology for

Classification of HRS using SVM

In this paper, we propose to use Support Vector Machines (SVMs) for the diagnosis of Hepatorenal Syndrome (HRS) based on clinical data. We have collected data for 100 patients from few hospitals. For each patient data, there are 14 features, including serum albumin, billirubine, creatinine, serum sodium, serum urea, urine output, urine microscopy, USG, ascites, cirrhosis, BP-systolic, diastolic, hemoglobin, urine protein. The data collected in medicine is generally collected because of patient care activity so as to benefit patients; hence data contained in medical databases is redundant, irrelevant, and inconsistent which can affect the results produced with the use of data mining techniques. Thus, data preprocessing and scaling are required so as to remove redundant as well as noisy data and to use normal forms of data. All of the data were transformed to real values with proper definition. For example, "Normal" converted to 1 and "Abnormal" to 0.

3 C

The results obtained provide good classification accuracy. Figure 1shows the architecture of our proposed work. Flowchart for proposed methodology can be described as the following phases: Phase I: a) HRS clinical data is collected and preprocessed.

Preprocessing of proposed work includes:

? Conversion of string data to numeric form: 1. Data value "Normal" is converted to 1 and "Abnormal" to 0. 2. Data value "Yes" is converted to 1 and "No" to 0. The final model obtained is tested on new or unseen data. This is known as final model evaluation. The accuracy hence obtained is considered as the accuracy of the model generated and it shows how much accurate and efficient model has been generated.

4 IV. Experimental Results and Performance Analysis

We used Support Vector Machine as the classification technique using LIBSVM -Matlab interface for our experiment. LIBSVM is an SVM package provided by Matlab. The computations involved were implemented on intel core i5 processor. The kernel function used here is Radial Basis Function (RBF) kernel, also known as sigmoid kernel.

Accuracy is evaluated using k-fold cross validation test. K-fold Cross-validation process includes dividing a dataset into k pieces, and on each piece, testing the performance of a predictor build from the remaining 90% of the data. In our work, k=5. The performance of the classification is evaluated for six parameters, namely, accuracy, sensitivity, specificity, precision, recall, f-measure. The definitions are as follows: Figure 2 shows cross-validation accuracy of 95%.This is a curve between logarithm of two important parameters, cost function C and rbf sigma, also known as gamma, represented by ?. The best value of both these factors gives the best cross validation accuracy of 95%. among all positivesamples and FPR, on the other hand, defines how many incorrect positive results occur among all negative samples. It is also known as graph between sensitivity and 1-specificity. Figure 3 shows ROC curve obtained for proposed work. The area under the ROC curve (AUC) obtained is 0.95. This value of AUC proves that the performance of classifier is good. ???????????????? = ????+???? ???? + ???? + ???? + ????

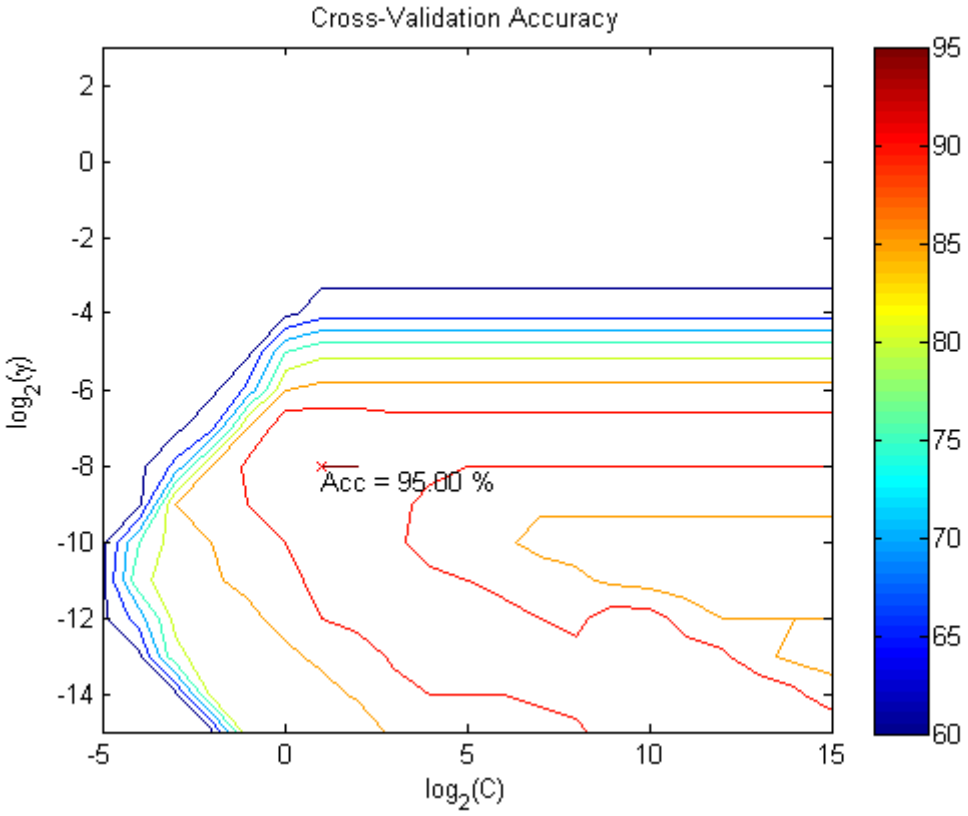
5 Global

6 V. Conclusion

In this research work, we propose to use SVM as the classification technique to diagnose HRS in patients of Cirrhosis. The performance is analyzed by comparing the predicted results with the manual results received along with data sets from hospitals. Our approach provides 95% classification accuracy and precision is recorded as 100%. It helps physician to diagnose the disease with more precision and accuracy. Sensitivity and Specificity

are computed as 90% and 100% respectively. Recall and F-Measure are measured as 90% and 94.74% respectively. Thus, SVM is proven as a good classifier for the prediction of HRS.

The proposed work can be further extended using feature selection or optimization techniques. Another extension can be application of SVM for diagnosis of similar diseases. [1]



1

Figure 1: Figure 1

[1] © 2017 Global Journals Inc. (US)

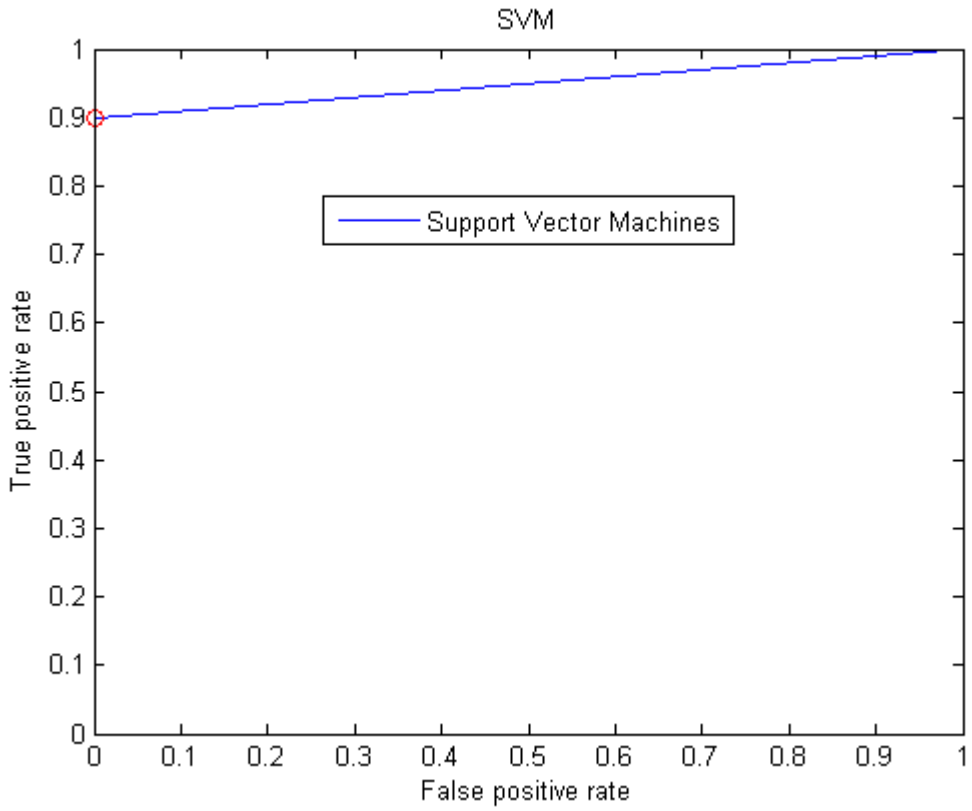
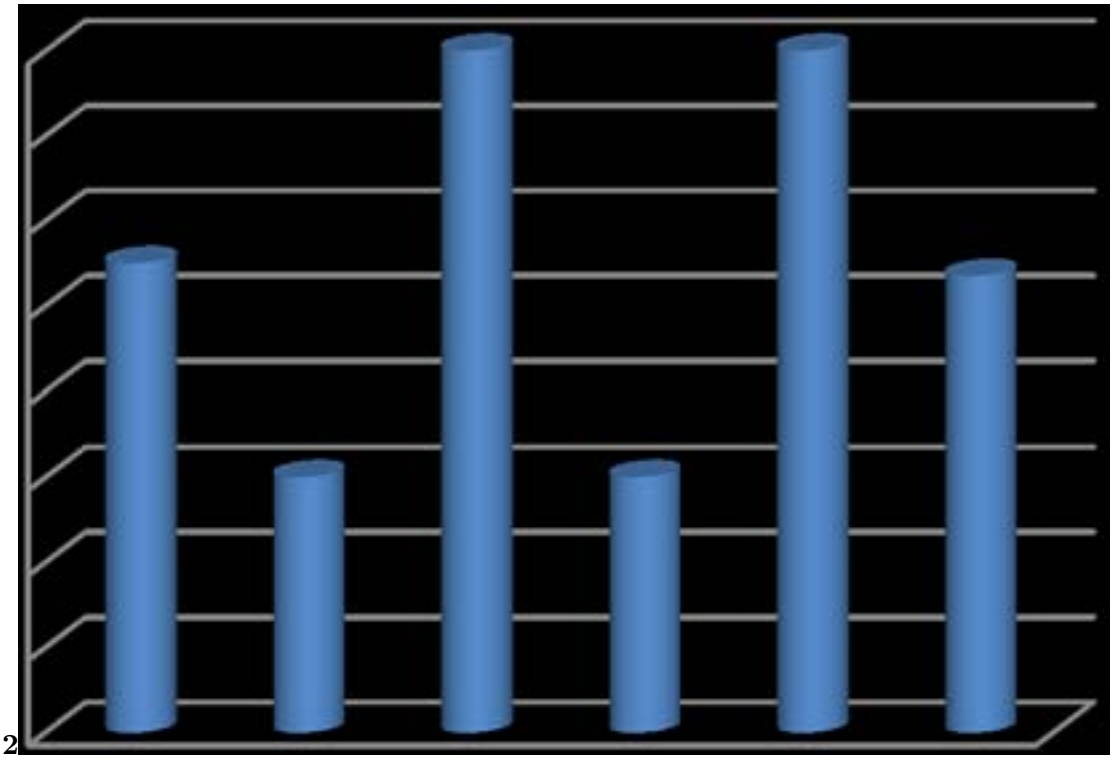


Figure 2:



2

Figure 3: Figure 2 :

.1 Global Journal of Computer Science and Technology

Volume XVII Issue I Version I 30 Year 2017 ( )

[Tomar and Agarwal ()] 'A survey on Data Mining approaches for Healthcare'. D Tomar , S Agarwal . *International Journal of Bio-Science and Bio-Technology* 2013. 5 p. .

[Boser et al. ()] 'A Training Algorithm for Optimal Margin Classifiers'. B E Boser , I M Guyon , V Vapnik . *Fifth Annual Workshop on Computational Learning Theory* 1992. ACM.

[Kousarrizi et al. ()] 'An experimental comparative study on thyroid disease diagnosis based on feature subset selection and classification'. M N Kousarrizi , F Seiti , M Teshnehlab . *International Journal of Electrical & Computer Sciences IJECS-IJENS* 2012. 12 p. 1.

[Kohli and Verma ()] *Arrhythmia classifycation using SVM with selected features, International*, N Kohli , N K Verma . 2011.

[Liu et al. (2008)] 'Breast cancer diagnosis using level-set statistics and support vector machines'. J Liu , X Yuan , B P Buckles . *30th Annual International Conference of the IEEE*, 2008. August. 2008. 2008. IEEE. p. .

[Hiesh et al. (2013)] 'Classification of schizophrenia using genetic algorithm -support vector machine (gasvm)'. M H Hiesh , Y Y L Andy , C P Shen , W Chen , F S Lin , H Y Sung , J W Lin , M J Chiu , F Lai . *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC*, 2013. July. 2013. 12 p. 1.

[Conf. Control, Automation, Robotics and Vision] *Conf. Control, Automation, Robotics and Vision*, (Singapore)

[Harb and Desuky ()] 'Feature Selection on Classification of Medical Datasets based on Particle Swarm Optimization'. H M Harb , A S Desuky . *International Journal of Computer Applications* 2014. p. 104.

[Han and Kamber ()] J Han , M Kamber . *data mining Concepts and Techniques*, 2000. Morgan Kaufmann Publishers. p. . (Classification of HRS using SVM)

[Ghumbre et al. ()] 'Heart disease diagnosis using support vector machine'. S Ghumbre , C Patil , A Ghatol . *International conference on computer science and information technology (ICCSIT')*, (Pattaya) 2011. (December))

[Jiang et al. ()] *Liver cancer Identification based on PSO-SVM Model*, H Jiang , F Tang , X Zhang . 2010. (11th Int)

[Delen et al. ()] 'Predicting breast cancer survivability: A comparison of three data mining methods'. D Delen , G Walker , A Kadam . *Artificial Intelligence in Medicine* 2005. 34 p. .

[Chen et al. (2010)] 'Support Vector Machine Methods for the Prediction of Cancer Growth. In Classification of HRS using SVM Figure 4 shows various performance parameters in the form of a bar chart with their experimental values'. X Chen , W K Ching , K F Aoki-Kinoshita , K Furuta . *Third International Joint Conference on*, 2010. May. 2010. CSO. 1 p. .

[Cortes and Vapnik ()] 'Support Vector Networks'. C Cortes , V Vapnik . *Machine Learning*, 1995. 20 p. .

[Balakrishnan et al. (2008)] 'SVM ranking with backward search for feature selection in type II diabetes databases'. S Balakrishnan , R Narayanaswamy , N Savarimuthu , R Samikannu . *IEEE International Conference on*, 2008. October. 2008. 2008. IEEE. p. . (Systems, Man and Cybernetics)