Diagnosis of liver abnormalities using Support Vector Machine

[1]Geeta Kadu, [2]Dr. Ranjana Raut, Mr. Suraj S. Gawande[3]

[1]Head of the Department, [2]Professor, [3]Application Engineer
Department of Electronics Engineering
Shri Datta Meghe Polytechnic, Nagpur, India
Maharashtra, India

ABSTRACT: **Many a times there is an incorrect diagnosis, specifically in rural areas, due to incorrect reading of clinical reports. For preliminary diagnosis Machine Learning Classification techniques are effectively used in medical field for finding abnormalities in many ailments. In this paper Support Vector Machine (SVM) a Supervised Machine Learning Classifier has been used for the demarcation of normal and diseased Liver. Since, data is not linearly separable, Non-Linear classifier is used. Kernel based SVM, such as Linear, quadratic, MLP, RBF and polynomial, with modified kernel functions are used. Indian Liver Patient Dataset (ILPD) is used for this work. Accuracy for various KFs is calculated. Finally the performance evaluation has been done where we compared the accuracy obtained from all the KF based SVM classifiers and the result is plotted.**

KEYWORDS**: Classification techniques, Supervised machine learning, Support Vector Machine (SVM), Indian Liver Patient Dataset (ILPD), Kernel Functions (KF)**

1. INTRODUCTION

Liver is a vital organ of a human body and its proper functioning is of utmost important for human survival. Due to excessive consumption of alcohol, inhaling poisonous gases due to pollution, unhealthy lifestyle liver patients are increasing day by day [1]. Liver diseases are not detected easily. Even with small liver injuries normal body functioning is not affected. Usually liver diseases are detected at an advance stage. Hence, an automated system, which can make demarcation between normal and abnormal liver, will be a boon in this case. Classification technique is a very effective automatic tool for prediction and diagnosis in medical field.

Primary diagnosis of liver diseases is done by examining the clinical reports of the patient. In many cases diagnosis varies from doctor to doctor. Classification techniques can be used as an automatic tool for demarcation between healthy and unhealthy person. Machine learning algorithms such as Decision Tree, Support Vector Machine, K-nearest-neighbor, Artificial Neural Network (ANN) etc can be implemented for classification techniques.

2. RELATED WORK

In earlier studies authors proposed various classification algorithms for the identification of liver abnormalities by working on the clinical data of liver function tests. Decision tree and SVM machine learning techniques were used by Brett A. Lidbury et.al. [2] for establishing response of various liver function tests with the level of ´Y-glutamyl transferase (GGT). Harsha Pakhale et. al, [3] discussed data mining techniques, explained in earlier papers, such as C4.5, Naive Bayes, Decision Tree, Support Vector Machine, Back Propagation Neural Network for the Classification of liver diseases. B. Yekkehkhany et al. [4] developed a framework for crop classification using SVM. Performance of different kernel functions viz. linear, polynomials and Radial Based Function (RBF) were compared. Accuracy, for classification of different crop types, was better for RBF. A survey was carried out by D. Sindhuja and R. Jemina Priyadarsini [5] for application of data mining techniques such as C4.5, Naive Bayes, Decision Tree, Support Vector Machine, Back Propagation Neural Network and Classification and Regression Tree for classification of liver disorders at an early stage. Advantages and disadvantages of various methods were discussed. Hoon Jin et. al. [6] compared six machine learning algorithms: Decision Tree, k-Nearest Neighbor, Multi-Layer Perceptron, Naïve Bayes, Logistic Regression and Random Forest. For this data mining tool WEKA was used and evaluation was done on the parameters: accuracy, precision, sensitivity and specificity. Aman Singh et. al. [7] evaluated performance of linear, non linear and decision tree classification algorithms. In linear LDA and DLDA; in non linear QDA, DQDA,NB and FFNN; and In decision tree CART were implemented. Best performance was obtained by CART model. Jankisharan Pahareeya et.al. [8] used selected classification algorithms viz. Multiple Linear Regression, Support Vector Machine, Multilayer Feed Forward Neural Network, J-48, Random Forest and Genetic Programming for separation between normal and affected liver. Data set used , ILPD, was unbalanced. Hence, undersampling and oversampling was used for balancing. Best performance was given by Random Forest Over Sampling. Anil Kumar Tiwari et. al. [9] used univariate analysis and feature selection methods for selection of attributes from the ILPD. Artificial neural network based predictive models such as BP, RBF, SOM, SVM were used for the classification.
Humar Kahramanli et. al. [10] used adaptive activation function in hidden layer and fixed sigmoid activation function in output layer for training the neural network. Then AIS (Artificial Immune Algorithm) algorithm was used for extracting rules from the trained neural network. Esraa M. Hashem et. al. [12] used SVM classification technique for liver disease diagnosis. It was

implemented on ILPD as well as BUPA liver function dataset. Performance was evaluated using parameters Accuracy, Prevalence, Sensitivity, Error rate and Specificity. Features were ranked on the basis of priority for each dataset. S. Karthik et, al. [13] used three phases for demarcation between normal and diseased liver and the identification of four types of liver diseases. In first phase MLP was employed for demarcation. In second phase accuracy was improved by extracting rules using LEM. In third phase Fuzzy rules were implemented for identification of four types of liver diseases. Naiping Li et.al. [14] implemented General Regression Neural Network (GRNN) for classification of alcoholic hepatitis. Input given to GRNN was the clinical data of patients having symptoms of alcoholic liver diseases. Result was cross checked from the experts from the medical field.

3. METHODOLOGY

The whole design & development carried out in two stages: i) Data processing and ii) Implementation of classifier. Liver data from ILPD is processed for checking missing values. SVM is used as a classifier. Block diagram of the proposed system is shown in fig. 1 below:

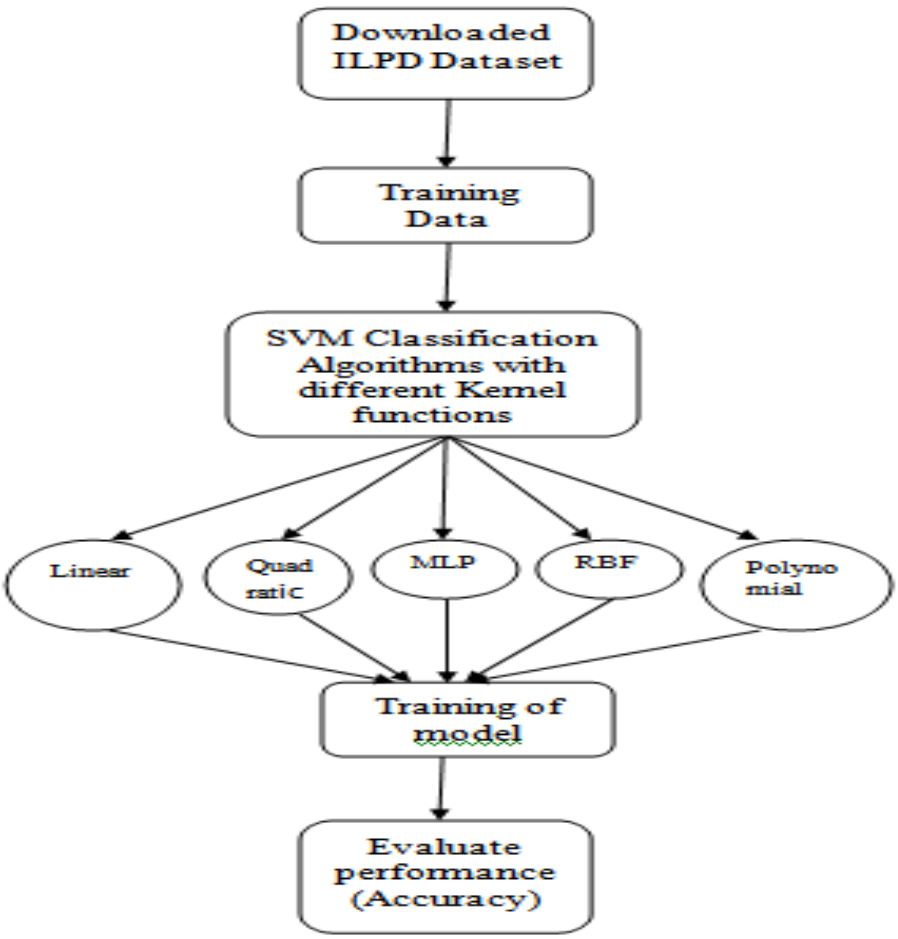

Figure 1. Block Diagram of the System

3.1 Data

ILPD data set is used in this paper. Data set was collected from north east of Andhra Pradesh, India and taken from UCI repository. There are 583 records; out of them 441 records belongs to male patients and 142 to female patients. There are 416 are liver patient records and 167 non liver patient records. Dataset contains 10 varaiables. They are age, gendrer, TB, DB, Alkphos, SGPT, SGOT, TP, ALB, A/G ratio and selector field. Attribute Information is given below:

1.Age - Age of the patient
2. Gender - Gender of the patient
3. TB - Total Bilirubin
4. DB - Direct Bilirubin
5. Alkphos - Alkaline Phosphotase
6. SGPT - Alamine Aminotransferase
7. SGOT - Aspartate Aminotransferase
8. TP - Total Protiens
9. ALB - Albumin

10. A/G Ratio - Albumin and Globulin Ratio
11. Selector field - used to split the data into two sets (liver patient or not)

3.2 Classification Algorithm

Supervised machine learning algorithms are widely used, for classification, in medical field. In this paper we have used SVM as supervised machine learning algorithm.

3.2.1 Support Vector Machines (SVM)

SVM, invented by Vapnik & Chervonenkis**,** is one of the machine learning tools which can be used both for classification as well as regression. But it is mostly used for binary classification. For more than two classifications algorithms such as "One Against All" (OAA) and "One Against One" (OAO) are used [4]. In SVM, when it is used for classification, data is segregated into two parts viz. training and testing. Each data item, selected for training, is plotted as a point in n-dimensional space. Where n is number of features you have with the value of each feature being the value of a particular coordinate. Then, to perform classification a hyper-plane is formed that differentiates the two classes. To identify the right hyper-plane following points are to be taken care of:

1. Select the hyper-plane which segregates the two classes clearly.
2. Maximize the distances between nearest data point of either class. This distance is called the margin.
3. While maximizing margin a care should be taken that there should not be a classification error.

Let us apply the above for the classification of and as shown in fig 1. Where A is the hyper plane which separates support vectors from as shown in Fig 2.

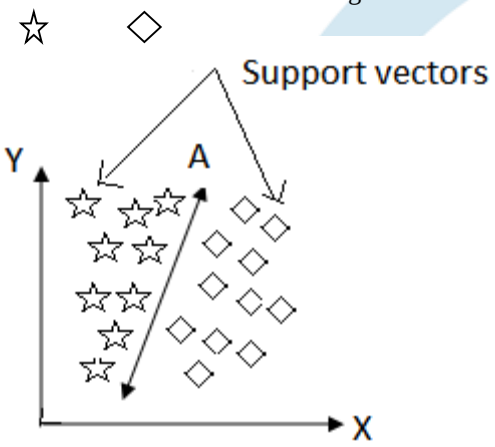

Figure 2: SVM classifier separating two classes

Above mentioned selection of hyper-plane is for linearly separable data. When data is not linearly separable we can use Non-Linear classifier. For Non-Linear classifier SVM has a technique called the kernel trick.

It helps to build a high dimensional feature space using a soft margin. Kernel functions used in SVM are linear, nonlinear, polynomial, radial basis, functions (RBF), sigmoid etc. Equations of some of them are as given below:

- Linear Kernel:
  $\kappa(x, x') = x^T x'$
- Quadratic Kernel:
  $\kappa(x, x') = (X^T x')^2$
  or
  $\kappa(x, x') = (1 + X^T x')^2$
- Polynomial Kernels: Contains polynomial terms up to degree d, for d > 0
  $\kappa(x, x') = (1 + x^T x')^d$
- Radial basis, functions (RBF
  $\kappa(x, x') = \exp(-(||x - x||)^2 / 2\sigma^2)$

Where:

$k(x_i, x_j)$ - is a kernel function
x_i & x_j - are vectors of feature space
d - is the degree of polynomial function
σ – is a free parameter.

4. EVALUATION

In this work, SVM is used as a classification algorithm for the demarcation of normal and diseased liver. Liver dataset by ILPD Medical Research Ltd., taken from UCI repository, is used. After analyzing through the dataset we found that it is not linearly separable. Hence, non-linear classifier is used. Kernel based SVM, such as Linear, quadratic, MLP, RBF and polynomial, with modified kernel functions are used.

Evaluation is done using performance parameters accuracy, mean square error (MSE), sensitivity and specificity. Best kernel function is selected after comparing performance parameters. Accuracy for various KFs is calculated. Accuracy is the percent of correct classifications.
Accuracy = (TP+TN) / (TP+TN+FP+FN)
Where:
TP - True Positive
TN - True Negative
FP – False positive
FN – False negative
Performance evaluation is done by comparing the accuracy obtained from all the KF based SVM classifiers..

5. RESULT & DISCUSSION

The developed application software efficiently performed for the classification of normal and abnormal liver. Here SVM is used for classification. Accuracy is evaluated for kernel based classifiers Linear, Polynomial, Quadratic, RBF and MLP, with modified kernel functions. Accuracy of linear kernel is 81.6781. Accuracy of Polynomial Kernel, with first and second order, is shown in table 1.

Table 1. Performance of SVM for Polynomial Kernel

Sr. No	Polynomial Order	Accuracy
1	1	81.6781
2	2	81.1644

Accuracy of Quadratic Kernel is 81.1644. Accuracy of RBF kernel, for different RBF values, is shown in table 2.

Table 2. Performance of SVM for RBF Kernel

Sr. No	RBF	Accuracy
1	1	82.5342
2	2	81.5068
3	3	81.3356
4	4	80.6507
5	5	80.137
6	6	79.6233
7	7	78.4247
8	8	77.226
9	9	76.5411
10	10	76.0274

Accuracy for MLP kernel, for different MLP parameters is shown in table 3.

Table 3. Accuracy of SVM for different MLP parameters

Sr. No	MLP Parameters	Accuracy
1	[1 -10]	76.3699
2	[2 -10]	81.3356
3	[3 -10]	78.4247
4	[4 -10]	78.2534
5	[5 -10]	78.4272
6	[6 -10]	80.4795
7	[7 -10]	79.2808
8	[8 -10]	80.137
9	[9 -10]	80.3082
10	[10 -10]	79.4521

Performance of SVM, for different kernel functions, is shown in table 4.

Table 4. Accuracy of SVM for different Kernel functions

Sr. No	Name of The Kernel Function	Accuracy
1	Linear	81.6781
2	Quadratic	81.1644
3	MLP [2 -10]	81.3356
4	rbf (1)	82.5342
5	Polynomial (1)	81.6781

6. CONCLUSION

An automated Liver disease detection system is implemented using SVM. ILPD clinical data taken from UCI repository is used. SVM is used as supervised algorithm as it is used for classification. Algorithm is used for demarcation between normal and diseased liver. To evaluate the performance of SVM different kernel functions viz. Linear, Polynomial, Quadratic, RBF and MLP are implemented. Accuracy is checked for modified kernel functions. Accuracy for linear kernel is 81.6781. In case Polynomial Kernel Accuracy is higher for first order Polynomial Kernel viz. 81.6781. Accuracy of Quadratic Kernel is 81.1644. Accuracy for RBF kernel, for different RBF values, is evaluated. Accuracy for RBF value 1 is highest and is equal to 82.5342. Highest accuracy for MLP kernel is 80.4795, for MLP parameter 6-10. SVM with RBF kernel with RBF value 1 gives highest accuracy viz. 82.5342.

.

REFERENCES

[1] Mitra Montazeri, Mahdieh Montazeri, Amin Beygzadeh and Mohammad Javad Zahedi, "Identifying efficient clinical parameters in diagnose of liver disease", HealthMED - Volume 8 / Number 10 / 2014
[2] Brett A. Lidbury, Alice M. Richardson and Tony Badrick, "Assessment of machine-learning techniques on large pathology data sets to address assay redundancy in routine liver function test profiles", DE GRUYTER, Diagnosis 2015; 2(1): 41–51
[3] Harsha Pakhale and Deepak Kumar Xaxa, "A Survey on Diagnosis of Liver Disease Classification", International Journal of Engineering and Techniques - Volume 2 Issue 3, May – June 2016
[4] B. Yekkehkhany, A. Safari, S. Homayouni and M. Hasanlou, " A Comparison Study of Different Kernel Functions for SVM-based Classification of Multi-temporal Polarimetry SAR Data", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-2/W3, 2014
[5] D.Sindhuja and R. Jemina Priyadarsini, "A Survey on Classification Techniques in Data Mining for Analyzing Liver Disease Disorder", *IJCSMC, Vol. 5, Issue. 5, May 2016, pg.483 – 488*
[6] Hoon Jin, Seoungcheon Kim and Jinhong Kim, " Decision Factors on Effective Liver Patient Data Prediction", International Journal of Bio-Science and Bio-Technology Vol.6, No.4 (2014), pp.167-178
[7] Aman Singh and Babita Pandey, "Liver disorder diagnosis using linear, nonlinear and decision tree classification algorithms", International Journal of Engineering and Technology (IJET)Vol 8 No 5 Oct-Nov 2016
[8] Jankisharan Pahareeya, Rajan Vohra, Jagdish Makhijani and Sanjay Patsariya,
" International Journal of Advanced Research in Computer Science and Software Engineering", IJARCSSE, Volume 4, Issue 2, February 2014

[9] Anil Kumar Tiwari, Lokesh Kumar Sharma and G. Rama Krishna, "Comparative Study of Artificial Neural Network based Classification for Liver Patient", Journal of Information Engineering and Applications Vol.3, No.4, 2013
[10] Humar Kahramanli and Novruz Allahverdi "Mining Classification Rules for Liver Disorders", International Journal of Mathematics and Computers in Simulation Issue 1, Volume 3, 2009
[11] Mehdi Neshat, Ali Adeli and Azra Masoumi, " A Survey on Artificial Intelligence and Expert System for Liver Disorders", ARPN Journal of Systems and Software VOL. 3, NO. 2, February 2013
[12] Esraa M. Hashem and Mai S. Mabrouk,
"A Study of Support Vector Machine Algorithm for Liver Disease Diagnosis", American Journal of Intelligent Systems 2014, 4(1): 9-14 DOI: 10.5923/j.ajis.20140401.02
[13] S. Karthik, A. Priyadarishini, J. Anuradha and B. K. Tripathy, "Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types", Advances in Applied Science Research, 2011, 2 (3): 334-345
[14] Naiping Li, Yongfang Jiang, JinMa, Bo He,Wei Tang,Mei Li, Qing Huang and Ting Yuan, "Computational Classification and Diagnosis of Alcoholic Liver Diseases Using General Regression Neural Network", Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2014, Article ID 524621