iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VI **Month of publication:** June 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Liver Disease Prognosis Based on Clinical Parameters Using Machine Learning Approach

Ayushi [1], Amit Sharma[2], Vinod Sharma[3], Dr. Rajeev Gupta[4]
[1]*Student M. Tech., Department of Computer Science & IT, University of Jammu, Jammu, India.*
[2]*Assistant Professor, Department of Computer Science & IT, Kathua Campus, University of Jammu, India.*
[3]*Professor, Department of Computer Science & IT, University of Jammu, Jammu, India.*
[4]*Assistant Professor, Guru Ram Dass Medical College, Amritsar, Punjab, India*

Abstract: ***We are living in the era of Machine Learning. The use of Machine Learning in medical diagnosis of various diseases increases many fold in recent years. In this Paper we had made an attempt to demonstrate an analytical approach for prediction of liver diseases in patients using probabilistic models of machine learning based on KSVM, SVM and KKNN. The technique used for classification and prediction are based on recognizing typical and diagnostically most important clinical features considered responsible for liver diseases. The main contributions of the research involve predicting the probability of each case against Class 'A' belonging to Non Diseased group and Class 'B' belonging to group of diseased patients. The analysis confirmed high risk and low risk patients as predicted by the probabilistic model.***
Keywords- Hepatocytes, LFT's, Clinical attributes, Probabilistic Model, High Risk, Low Risk, Analysis

I. INTRODUCTION

Liver disease is an umbrella term that encloses any type of damage or disorder that reduces the functioning of the liver. Liver diseases are usually caused by inflammation or damaged hepatocytes in the liver. Liver Diseases are not easily discovered as it is capable of maintaining normal function even when it is partially damaged. Thus, early diagnosis is one of the most important steps in liver disease treatment. LFTs are a helpful screening tool that helps in detection of liver dysfunction. Liver disease is one of the leading causes of death in India. As per the data published by the WHO in may 2014 liver disease death in India reached 216,865 or 2.44% of total [1]. According to the WHO liver disease is the tenth most commom cause of casuality in India.

II. RELATED WORK

M.Neshat et.al proposed a Fuzzy Expert System Design for diagnosis of liver disorders. Fuzzy system consist of four parts; Fuzzification, Fuzzy Inference Engine, Fuzzy Rule Base and Defuzzification. Membership functions of disease fields had been created by the expert and the dependency function formulas are created for measuring of liver disorder risk as low and high. The proposed expert system shows an accuracy of 91% [2].Rong-Ho Lin suggests an intelligent model for the diagnosis of liver diseases that integrates CART and CBR. CART is used to predict whether a patient suffers from liver disease and CBR is used to predict the type of liver diseases. The proposed methods show an accuracy of 92.94% and 90.00% respectively [3].

Mehdi Neshat et.al provides a hybrid model for the diagnosis of liver diseases. In comparison with traditional diagnoses their system is faster, more economical, more reliable and more accurate as Hopfield neural network and fuzzy Hopfield neural network diagnose liver disorders with the accuracy of 88.2% and 92% respectively [4].

Sa'diyah Noor Novita Alfisahrin et.al provide data mining techinques for optimization of classification of liver diseases. Their study aims to identify whether the patients suffer from liver disease based on the 10 important attributes of liver disease using a Decision Tree, Naive Bayes, and NBTree algorithms and the result shows NBTree algorithm has the highest accuracy; however the Naive Bayes algorithm gives the fastest computation time [5].

Dr. R.R.Janghel et.al analysed various Artificial Neural Networks models like Back Propagation Algorithm, Probabilistic Neural Networks, Competitive learning Networks, Learning vector quantization and Elman Networks for diagnosis of Hepatitis and liver disorders and the result shows ANNs in comparison with other traditional diagnostic systems is faster, more reliable and more accurate [6]

Dong Xu et.al proposed a liver disease diagnosis model based on a combination of rough set theory (RS) and LMBP neural network (RS_LMBPNN). The model first use rough set theory to eliminate redundant informantion thereby reducing the LMBP neural network training data. The results show that compared with the single LMBP neural network model, the combined model have high

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

training speed, have stronger learning ability, and also the better diagnostic accuracy [7].

Sangman Kim et.al proposed effective liver cancer diagnosis method based on neural network and fuzzy neural network. The fuzzy logic finds only the important attributes that helps in diagnosis of liver diseases and the proposed model detects liver cancer patients with an accuracy of 98~99 % [8]

Shimaa Abd Allah Ibraheem et.al proposed a Hybrid Rough-Fuzzy Classifierer for Liver Disease Diagnosis. Firstly rough sets are used to generate and reduce classification rules which are then used in fuzzy set to enhance the classification accuracy of liver diseases diagnoses. The proposed model shows 99.1 % classification accuracy in rule generation [9].

According to Parisa Tavakkoli et.al classification and clustering of liver disorders data into two healthy and ill categories is done by using five-layer ANFIS combined structure. The results show an appropriate separation between the healthy and patient data; therefore the result will be reliable and credible to detect the disease and extension for the other disease [10].

Shiladitya Saha et.al employ artificial neural network and support vector machine classifiers for classification of patients suffering from jaundice on the basis of liver condition. Several sets of MLP and SVM classifiers are combined with Decision template and Dempster-Shafer theory fusion techniques. Hybridization of two ANNs and four SVM classifiers with Dempster-Shafer algorithm gives up to 97.33% of prediction accuracy [11].

III. EXPERIMENTAL SETUP

This paper reports on methodologies and outcome of study aiming at developing robust model to classify and predict whether patient is suffering from liver disease or not using R tool and R Studio 3.4.0

Patient Selection:-The study aims to establish the relation between various clinical test features present in the data set. The data of 558 patients including 371 with normal liver grouped as Class A and 187 with diseased liver grouped as Class B are analyzed. Each case was having attributes bilirubin (BI), direct bilirubin (D(BI)), indirect bilirubin (I(BI)), total protein (TP), albumin (ALB), globulin (GLB), gender, AG ratio, aspartate amino transferase (SGOT), alanin amino transferase (SGPT) and alkaline ahosphatase (ALP) all of these are clinical in nature and the rest i,e age and gender are physiological.

Table I.

Table1: Showing patient characteristics and clicical attributes used in the model for classification

S.No.	Attributes	Short Name	Reference Value
1	Gender	GENDER	Male M Female F
2	Age	AGE	
3	Bilirubin	BI	0.1 – 1.45 mg/dl
4	Conjugated Bilirubin / Direct Bilirubin	D(BI)	0 – 0.5 mg/dl
5	Unconjugated Bilirubin / Direct Bilrubin	I(BI)	0 – 0.7 mg/dl
6	Total Protein	TP	6.1 – 8.0 gm/dl
7	Albumin	ALB	3.2-5.5 gm/dl

8	Globulin	GLB	2.3 – 3.5 gm/dl
9	AG Ratio	AG RATIO	
10	Aspartate Amino Transferase	AST/SGOT	5.0 – 46.0 IU/L
11	Alanin Amino Transferase	ALT/SGPT	5.0 – 49.0 IU/L
12	Alkaline Phosphatase	ALP	20 – 140 IU/L

Methods:-The 558 were categorized as Class 'A' (Non diseased) and Class 'B' (diseased). The probabilty based models were developed involving SVM, KSVM and KKNN. The probabilty for each case against Class 'A' and Class 'B' is predicted based on the attributes as mentioned in Table I.

A. SVM (Support Vector Machine

These classifiers are based on statistical learning theory that aims to find the hyperplanes (decision theories) that best segregates the two classes. The hyperplane is the key geometric entity that is one dimension lower than high dimensional feature space that divides that space into two regions. Mathematically hyperplane of p dimensional feature space $\vec{x} = (x_1, x_2 \ldots x_p)$ is defined as:

$$b_0 + \sum_{j=1}^{p} b_j x_j$$

$$b_0 \neq 0 gives affine palne that does not pass through origin.$$

B. KSVM (Kernal Support Vector Machine)

Kernal methods are a group of classifiers that takes user defined similarity functions (called kernels) to compute similarity over a pair of data points instead of using feature vectors. Under some condition every kernel can be represnted by a dot product in feature space. Since many machine learning algorithms can be expressed entirely in terms of dot product the need of large feature vectors for classification can be eliminated by using kernels in place of dot product. Any machine learning algorithm that can be expressed as a dot product can be written using kernels like KSVM and KKNN. Mathematically hyperplane in case of KSVM using non linear kernel is defined as:

$$K(\overrightarrow{x_i}, \overrightarrow{x_k}) = \sum_{j=1}^{p} b_{ij} x_{kj}$$

Where $\overrightarrow{x_i}, \overrightarrow{x_k} are two obsevations$ and K finds the similarity between these two observations.

C. KKNN

KKNN is a non parametric and lazy learning algorithm that classifies new cases based on a similarity function (called distance function). In K nearest neighbor's classifier euclidean distance is calculated between test data and every sample in the training data followed by classifying the test data into a class in which most of k nearest neighbors of training data belong to. Usually the value of k is a small positive integer. As the value of k increases it becomes difficult to distinguish between classes.

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

VI. RESULTS AND ANALYSIS

Clinical Attributes Analysis:-The clinical attributes considered for the liver disease analysis are BI, D(BI), I(BI), TP, ALB, GLB, ALT/SGOT, AST/SGPT and ALP. The average value for the same is shown in table (2). The relationship between clinical attributes versus class A (Non diseased) and class B (Diseased) groups are shown in figure (1) and figure (2) respectively. Table (3) shows the minimum and maximum values observed against each clinical attribute for Class A and Class B.

Table II.

Table 2:Showing the average values for clinical attributes.

Class	Age	BI	D(BI)	I(BI)	TP	ALB	GLB	AGRatio	AST	ALT	ALP
A	45.19	0.86	0.19	0.66	6.98	4.23	2.74	1.62	31.4	35.91	101.58
B	41.01	2.40	1.11	1.29	7.18	4.20	2.97	1.52	275.98	271.95	128.76

Table III.

Table 3:Showing minimum and maximum value of clinical attributes against Class A and ClassB

S.No.	Clinical Attributes	Class A		CLASS B	
		Min. Value	Max. Value	Min. Value	Max. Value
1.	BI	0.135	3.92	0.3	27.31
2.	D(BI)	0.03	2.29	0.07	17.94
3.	I(BI)	0.015	1.63	0.1	9.37
4.	TP	5.8	8.6	5.2	38.5
5.	ALB	2.1	5.5	2.6	5.9
6.	GLB	1.6	4.7	1.5	8.5
7.	AG RATIO	0.56	3.12	0.6	3.6
8.	AST	13	147	18	7450
9.	ALT	14	97	19	9750
11.	ALP	39	456	281	947

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

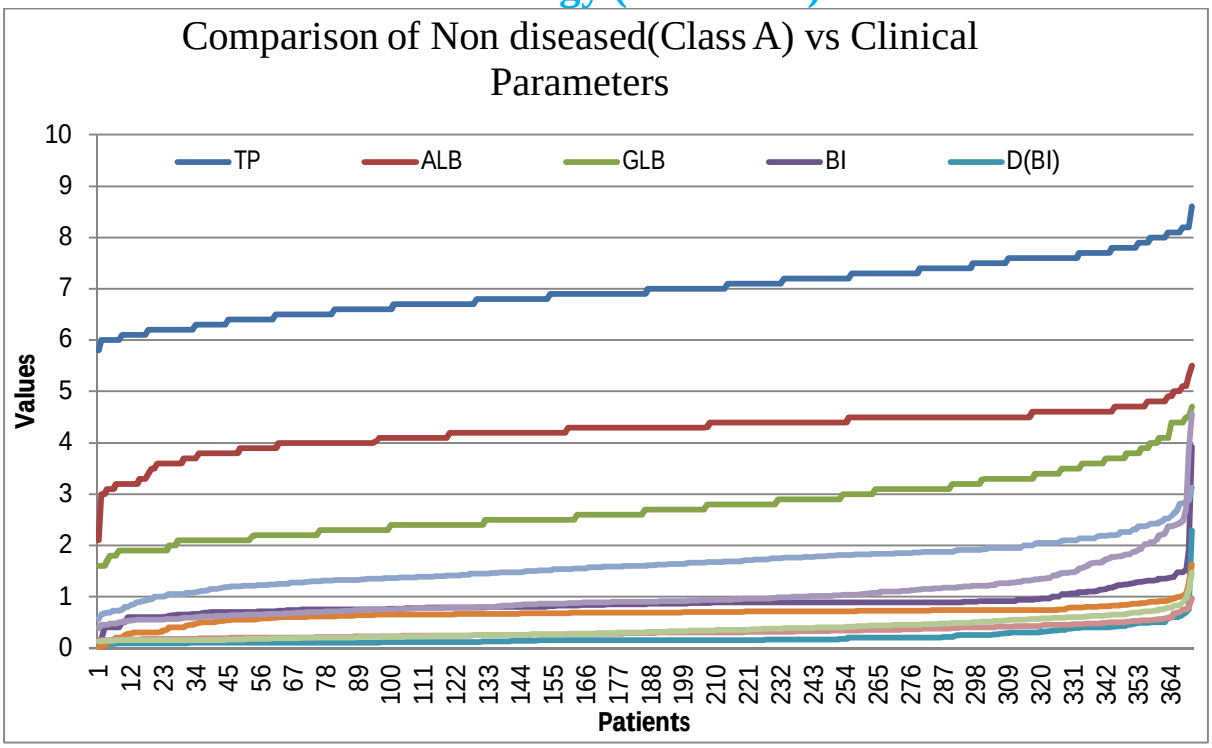


Figure1: Showing relationship between Non-diseased (Class A) Patients and Clinical Attributes

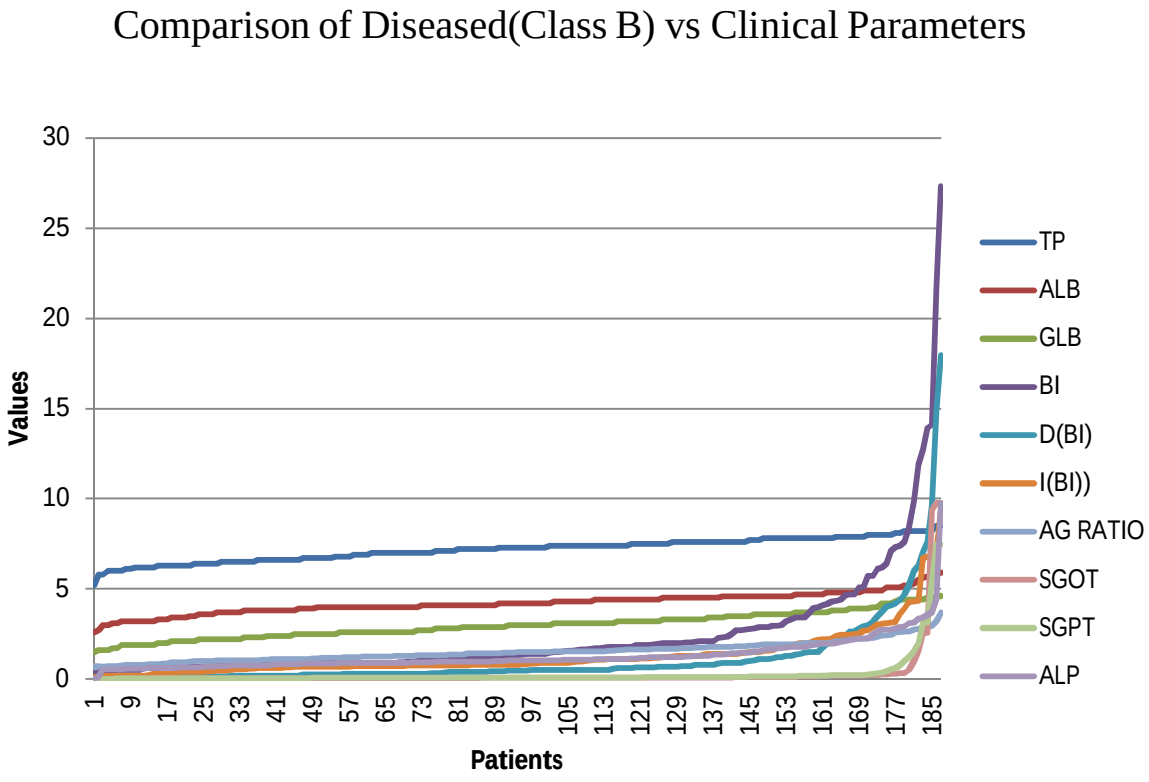


Figure 2: Showing relationship between Diseased (Class B) Patients and Clinical Attributes

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Classifiers Analysis:-Three classifiers SVM, KSVM and KKNN are applied on the same dataset and the results were observed and anlyzed in terms of accuracy, percision and specificity as shown in Table IV and figure 3.

Table IV.

Table 4: Showing Accuracy, Precision and Specificity of Classifiers

Classifier	Accuracy	Precision	Specificity
SVM	83.16	96.03	98.92
KSVM	85.66	98.19	99.46
KKNN	84.41	98.07	99.46

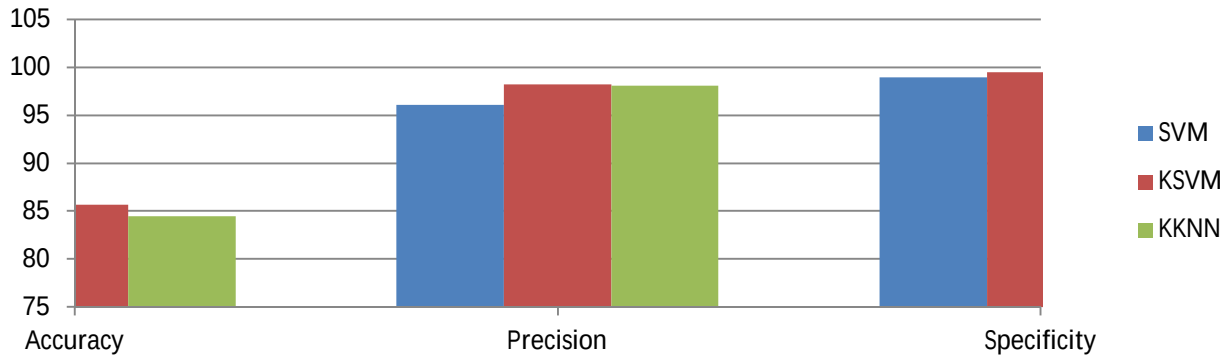


Figure 3: Showing accuracy, precision and specificity of candidate classifiers

The predicted probabilty of each classifier against Class A and Class B is shown in Table V

Table V

Table 5: Showing predicted proability of each classifier against Class A and Class B

S.NO.	Machine Learning Algorithm	Predicted Probability				Groups			
		Class A		Class B		Class A		Class B	
		Min. Value	Max. Value	Min. Value	Max. Value	Under Risk	Non Risk	Under Risk	Non Risk
1.	SVM	0.004	0.96	0.03	0.99	14	357	112	75
2.	KSVM	0.003	0.95	0.04	0.99	12	359	118	62
3.	KKNN	0	1	0	1	1	370	165	22

Volume 5 Issue VI, June 2017
ISSN: 2321-9653

VII. CONCLUSION

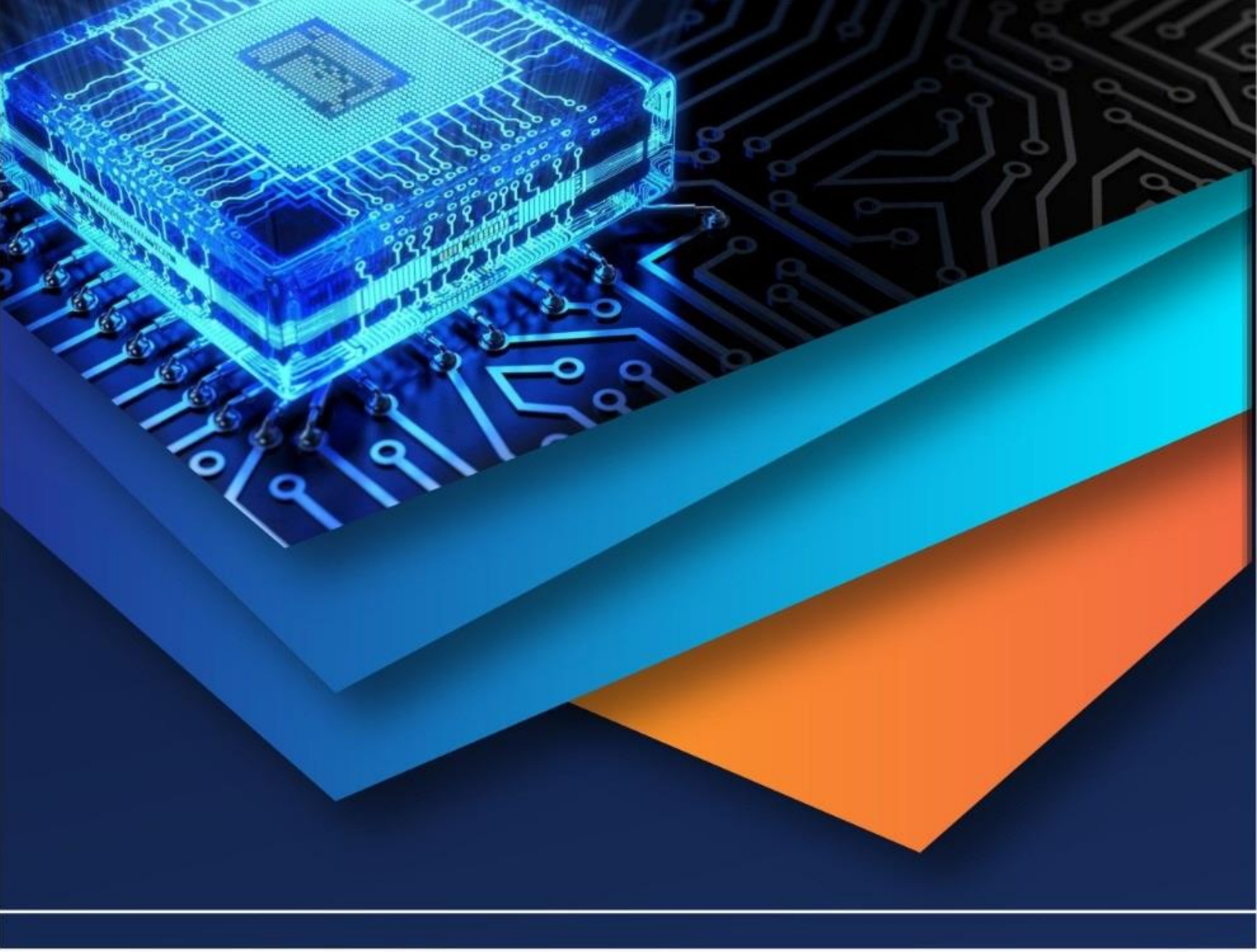
A. The predicted probability >0.48 in SVM, KSVM and KNN for Class 'A' included 357, 359 and 370 cases. In other words these cases were predicted to be out of risk group compared to the 14, 12 and 1 case belong to low risk group category as shown in Table

B. The predicted probabilty >0.48 in SVM, KSVM and KKNN for Class 'B' included 114, 118 and 165 cases. In other words these cases are confirmed with severly infected disease group while 73, 69 and 22 were cases of Class 'B' predicted with <0.48 indicating the group belonging to disease with no severity as shown in table (5).

C. The observed accuracy against classifiers SVM, KSVM and KKNN are 83.16, 85.66 and 84.40. The maximum accuracy is observed in case of KSVM as shown in table (4) and figure (3).

REFERENCES

[1] WHO, "World Health Statistics 2014," 2014 Global Health Observatory (GHO) data WHO, 2014.
[2] M. Neshat, M. Yaghobi, M. B. Naghibi, and A. Esmaelzadeh, "Fuzzy Expert System Design for Diagnosis of Liver Disorders," 2008 International Symposium on Knowledge Acquisition and Modeling, 2008.
[3] R.-H. Lin, "An intelligent model for liver disease diagnosis," Artificial Intelligence in Medicine, vol. 47, no. 1, pp. 53–62, 2009.
[4] M. Neshat and A. E. Zadeh, "Hopfield neural network and fuzzy Hopfield neural network for diagnosis of liver disorders," 2010 5th IEEE International Conference Intelligent Systems, 2010.
[5] S. N. N. Alfisahrin and T. Mantoro, "Data Mining Techniques for Optimization of Liver Disease Classification," 2013 International Conference on Advanced Computer Science Applications and Technologies, 2013.
[6] R. Janghel, A. Shukla, and K. Verma, "Soft computing based expert system for Hepatitis and liver disorders," 2016 IEEE International Conference on Engineering and Technology (ICETECH), 2016.
[7] D. Xu, H. Fu, and W. Jiang, "Research on Liver Disease Diagnosis Based on RS_LMBP Neural Network," 2016 12th International Conference on Computational Intelligence and Security (CIS), 2016.
[8] S. Kim, S. Jung, Y. Park, J. Lee, and J. Park, "Effective liver cancer diagnosis method based on machine learning algorithm," 2014 7th International Conference on Biomedical Engineering and Informatics, 2014.
[9] Shimaa Abd Allah Ibraheem, Hatem Mohamed Abd Elkader, Ibraheen Selim, and Reda Hussein," Hybrid Rough-Fuzzy Classifier for Liver Disease Diagnosis" 2016 I. J. of Electronics and Information Engineering (IJEIE), 2016.
[10] P. Tavakkoli, D. M. Souran, S. Tavakkoli, M. Hatamian, A. Mehrabian, and V. E. Balas, "Classification of the liver disorders data using Multi-Layer adaptive Neuro-Fuzzy inference system," 2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2015.
[11] S. Saha, S. Saha, and P. P. Bhattacharyya, "Classifier fusion for liver function test based Indian jaundice classification," 2015 International Conference on Man and Machine Interfacing (MAMI), 2015
[12] Prakash B. Desai, Vinayak V. Hull, Avinash A K. Math, Sonal N. Vernekar, and Shruthi S. Kulkarni, "A review on laboratory liver function tests," Pan African Med ical Journal, vol3 page17,2009.
[13] K.-S. Kang, "Abnormality on Liver Function Test," Pediatric Gastroenterology, Hepatology & Nutrition, vol. 16, no. 4, p. 225, 2013.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY