Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018)
IEEE Conference Record: # 42666; IEEE Xplore ISBN:978-1-5386-3570-4

# *Accuracy Prediction using Machine Learning Techniques for Indian Patient Liver Disease*

***L. Alice Auxilia***
*School of Information Technology and Engineering*
VIT University, Vellore, India
aliceauxilia@gmail.com

***Abstract* - The utilization of medicinal datasets has pulled in the consideration of specialists around the world. Machine Learning methods have been broadly utilized as a part of creating choice emotionally supportive networks for ailments forecast through an arrangement of therapeutic datasets. Grouping systems have been broadly utilized as a part of the restorative field for exact order than an individual classifier. Liver malady (additionally called hepatic infection) is a sort of harm to or illness of the liver. There are in excess of a hundred various types of liver ailment. In this task, I have taken the datasets of general Indian liver ailment patient's records to help basic leadership. Indian Liver Patient's datasets demonstrate that proposed technique amazingly enhances the illnesses expectation precision.**

***Keywords* – Liver Disease; Pearson Correlation; Classification; Decision Tree; Naive Bayes; Support Vector Machine; Random Forest; Artificial Neural Network.**

## I. INTRODUCTION

The point of this task is to some degree diminish the time delay caused because of the superfluous forward and backward transporting between the healing centre and the pathology lab. Truly, work has been done in recognizing the beginning of sicknesses like coronary illness, Parkinson's from different highlights. For this situation, a machine learning calculation will be prepared to foresee a liver ailment in patients. This is by all accounts a great case of managed learning. We have been given a settled number of highlights for every datum point, and our point will be to prepare an assortment of Supervised Learning calculations on this information, so that, when another information point emerges, our best performing classifier can be utilized to classify the information point as a positive case or negative.

To analyze the conceivable issue in view of the test report esteems. The specialists can enter the patient's report as information. Utilizing the framework we are foreseeing if the patient is having a liver infection or not by examining the accessible dataset. The datasets are taken from UCI vault. We are catching the test report contribution from UI and store it in a table. We have made a table with the current dataset which has every one of the records. We have utilized Decision Tree calculation, Naïve Bayes, Random Forest, Support Vector Machine, Artificial Neural Network to make a model. In light of the model and the test information we are anticipating if the patient can have Liver Disease and the precision of the ailment.

## II. LITERATURE SURVEY

Ashwani Kumar, Neelam Sahu [1] has broke down Classification is one of the great techniques to dissect liver and non liver patient in awesome exactness. They utilized the C4.5, Random Forest, CART, Random Tree and REP tree as classifier for order of liver information and accomplished 79.22% exactness in Random Forest utilizing 80-20% information segment with 6 highlights. They suggested Random woodland is better classifier for characterization of liver illness among all.

Chandrasegar Thirumalai, Rashad Manzoor [2] has broke down its relationship coefficient with different characteristics, by utilizing Pearson strategy. From these coefficient esteems, the properties having values which are more prominent than 0.55, are mulled over. Further, machine learning technique is utilized for discovering the minimum cost capacity and its relating theta esteem. By utilizing straightforward direct relapse strategy on the diminished properties, we have gotten direct conditions. Their proposed display one can take astute choices with the assistance of convenient electronic gadgets like advanced cells and tabs.

Harsha Pakhale, Deepak Kumar Xaxa [3] has examined the individual model does not fulfill the arrangement exactness of model so they have gathering the individual models to build up the hearty classifier. They have additionally connected the data pick up include determination strategy to computationally expand the execution of model.

Dr. S. Vijayarani, Mr.S.Dhayanand [4] has finished up the SVM classifier is considered as a best calculation due to its most elevated grouping exactness. Then again, while looking at the execution time, the Naïve Bayes classifier needs least execution time.

Anju Gulia, Dr. Rajan Vohra, Praveen Rani [5] presents an approach that will be utilized for half breed display development of group wellbeing administrations. These arrangement calculations can be actualized for other overwhelming sicknesses likewise like cardiovascular and diabetes forecast and grouping. An another extension is to seeing climate by applying new calculations will made any

upgrades over strategies which are utilized as a part of this paper in future.

Kalyan Nagaraj and Amulyashree Sridhar [6] say an endeavour is made for forecast of liver infection in patients utilizing information mining systems. A cross breed Neuro-SVM show was produced for characterization of liver patients in view of their organic parameters utilizing fake neural system. The cross breed display is sent as a graphical UI (GUI) in R. The GUI can be utilized as a screening instrument by specialists for expectation of liver illness in patients in future.

Jankisharan Pahareeya, Rajan Vohra, Jagdish Makhijani , Sanjay Patsariya [7] closes Throughout the examination ten times cross approval is performed. The proposed Random Forest over testing (200%) show beat every single other method. The Random Forest over examining (200%) display is the generally best indicator among every single other strategy.

Bendi Venkata Ramana, Surendra. Prasad Babu. M, Venkateswarlu. N.B [8] has broke down Modified pivot timberland calculation was proposed with multi layer observation grouping calculation and irregular subset include determination technique for UCI liver informational collection. Changed pivot woodland calculation was proposed with closest neighbour arrangement calculation and connection based element determination strategy for INDIA liver dataset.

Kotsiantis. S.B [9] advises to build the forecast exactness of the straightforward Bayes demonstrate. The idea of joining classifiers is proposed as another bearing for the change of the execution of individual classifiers. he utilized a discretization strategy and expelled excess highlights utilizing a channel include determination technique. At long last, he played out a substantial scale correlation with different endeavours that have attempted to enhance the precision of the basic Bayes calculation and additionally other best in class calculations and he took better exactness as a rule utilizing less time for preparing, as well.

## III. MACHINE LEARNING TECHNIQUES

Classification assumes vital part for order of information. In this examination work we have utilized different classification systems for order of liver patient information. They are portrayed beneath:

*A. Decision Tree* - A decision tree [23] is a stream diagram like structure, where each interior (non-leaf) hub signifies a test on a property, each branch speaks to the result of a test, and each leaf (or terminal) hub holds a class name. The highest hub in a tree is the root hub. There are numerous particular decision tree calculations [10][11].

*B. Naive Bayes* - In machine learning, Naive Bayes classifiers are a group of basic probabilistic classifiers in view of applying Bayes [24] hypothesis with solid (Naive) autonomy suspicions between the highlights. Naive Bayes classifiers are very versatile, requiring various parameters straight in the quantity of factors (highlights/indicators) in a learning issue. Most extreme probability preparing should be possible by assessing a shut frame articulation, which takes direct time, instead of by costly iterative guess as utilized for some different kinds of classifiers. Naive Bayes is a basic strategy for developing classifiers: models that dole out class names to issue cases, spoke to as vectors of highlight esteems, where the class names are drawn from some limited set. It isn't a solitary calculation for preparing such classifiers, yet a group of calculations in light of a typical guideline: all Naive Bayes [12][13] classifiers expect that the estimation of a specific component is autonomous of the estimation of some other element, given the class variable.

*C. Support Vector Machine (SVM)* - In machine learning, bolster vector machines (SVMs, likewise support vector systems) are managed learning models with related learning calculations that break down information utilized for order and relapse investigation. Given an arrangement of preparing illustrations, each set apart as having a place with either of two classes, a SVM preparing calculation constructs a model that doles out new cases to one classification or the other, making it a non-probabilistic twofold direct classifier (despite the fact that techniques, for example, Platt scaling exist to utilize SVM in a probabilistic characterization setting). A SVM [14][15][25] display is a portrayal of the cases as focuses in space, mapped with the goal that the cases of the different classifications are partitioned by an unmistakable hole that is as wide as could reasonably be expected. New illustrations are then mapped into that same space and anticipated to have a place with a classification in view of which side of the hole they fall. Not-withstanding performing direct arrangement, SVMs can productively play out a non-straight grouping utilizing what is known as the bit trap, certainly mapping their contributions to high-dimensional element spaces. At the point when information are not marked, regulated learning isn't conceivable, and an unsupervised learning approach is required, which endeavours to discover normal bunching of the information to gatherings, and after that guide new information to these shaped gatherings. The help vector bunching calculation made by Hava Siegelmann and Vladimir Vapnik, applies the measurements of help vectors, created in the help vector machines calculation, to order unlabeled information, and is a standout amongst the most broadly utilized grouping calculations in mechanical applications.

*D. Random Forest* - Irregular forests [16][17] or random decision forests are a group learning strategy for arrangement, relapse and different assignments, that work by developing a large number of decision trees at preparing time and yielding the class that is the method of the classes (characterization) or mean expectation (relapse) of the individual trees. Irregular decision forests amend for decision trees' propensity for over-fitting to their preparation set.

*E. Artificial Neural Network (ANN)* - An ANN depends on an accumulation of associated units or hubs called fake neurons (an improved adaptation of natural

neurons in a creature mind). Every association (an improved rendition of a neurotransmitter) between counterfeit neurons can transmit a flag starting with one then onto the next. The fake neuron that gets the flag can process it and after that flag simulated neurons associated with it. In like manner ANN, usage the flag at an association between fake neurons is a genuine number and the yield of each fake neuron is computed by a non-direct capacity of the whole of its sources of info. Fake neurons and associations ordinarily have a weight that modifies as learning continues. The weight increments or declines the quality of the flag at an association. Manufactured neurons may have an edge with the end goal that exclusive if the total flag crosses that edge is the flag sent. Regularly, counterfeit neurons are sorted out in layers. Distinctive layers may perform various types of changes on their data sources. Signs go from the main (contribution), to the last (yield) layer, potentially in the wake of navigating the layers numerous circumstances. The first objective of the ANN [18][19] approach was to take care of issues similarly that a human cerebrum would. Be that as it may, after some time, consideration concentrated on coordinating particular assignments, prompting deviations from science. ANNs have been utilized on an assortment of undertakings, including PC vision, discourse acknowledgment, machine interpretation, informal organization sifting, playing board and computer games and restorative determination.

## IV. PEARSON CORRELATION

Table I. Pearson Correlation of Indian Liver Patient Dataset

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| A | 1 | 0.696 | 0.764 | -0.03 | -0.08 | -0.22 | -0.22 | -0.16 |
| B | 0.69 | 1 | 0.20 | -0.03 | -0.07 | -0.22 | -0.22 | -0.224 |
| C | 0.764 | 0.206 | 1 | -0.03 | -0.022 | 0.017 | 0 | 0.041 |
| D | -0.03 | -0.03 | -0.03 | 1 | 0.909 | 0.269 | 0.234 | 0.137 |
| E | -0.08 | -0.07 | -0.02 | 0.90 | 1 | 0.324 | 0.288 | 0.113 |
| F | -0.22 | -0.22 | 0.017 | 0.26 | 0.324 | 1 | 0.845 | 0.254 |
| G | -0.22 | -0.22 | 0 | 0.23 | 0.288 | 0.845 | 1 | 0.211 |
| H | -0.16 | -0.22 | -0.04 | 0.13 | 0.113 | 0.254 | 0.211 | 1 |

*A– Albumin, B – Albumin_and_Globulin_Ratio, C – Total_Proteins, D– Alamine_Aminotransferase, E – Aspartate_Aminotransferase, F – Direct_Bilirubin, G – Total_Bilirubin, H – Alkaline_Phosphotase.*

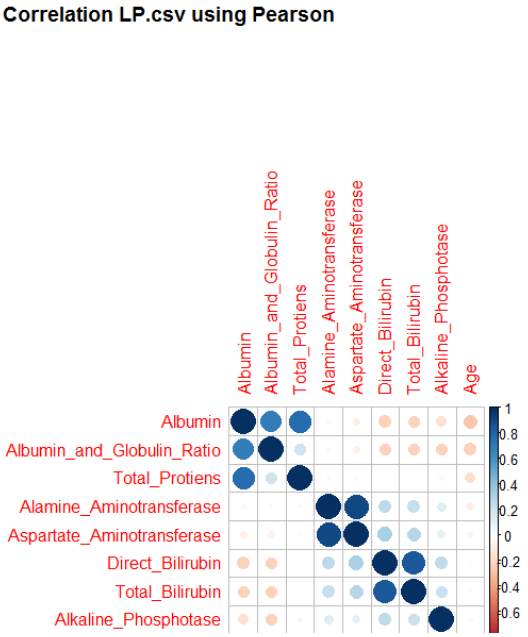


Figure1.*Pearson Correlation*

## V. PERFORMANCE MEASUREMENT

Execution of every classifier can be assessed by utilizing some exceptionally surely understood factual measures like precision, affectability and specificity. These measures are characterized by genuine positive (TP), genuine negative (TN), false positive (FP) what's more, false negative (FN). With the assistance of disarray network we can compute the execution measures[20].

Order Accuracy measures the extent of amend expectations thinking about the positive and negative inputs. It is computed as takes after:
Grouping precision = (TP+TN)/N (1)

Sensitivity [21] measures the extent of the genuine positives, that is, the capacity of the framework on foreseeing the right values in the cases displayed. It is computed as takes after:
Sensitivity=TP/(TP+FN) (2)

Specificity [22] measures the extent of the genuine negatives, that is, the capacity of the framework on foreseeing the right values for the cases that are the inverse of the coveted one. It is figured as takes after:
Specificity =TN/(TN +FP) (3)

## VI. NUMERIC RESULTS

This experiment is done in RStudio environment with R programming language, under classification algorithms of machine learning

Table II. Numeric Results of Indian Liver Patient Dataset

| | DT | NB | RF | SVM | ANN |
|---|---|---|---|---|---|
| ACC | 81% | 37% | 77% | 77% | 71% |
| CI | 0.77,0.84 | 0.33,0.41 | 0.72,0.54 | 0.70,0.60 | 0.65,0.55 |
| NIR | 0.75 | 0.75 | 0.76 | 0.70 | 0.74 |
| P- VAL | 0 | 1 | 1 | 2 | 2 |
| KAPPA | 0.39 | 0.05 | 0.07 | 0.11 | 0.09 |
| MC-VA | 7.3 | <2e | 0.07 | 0.05 | 0.1 |
| SENS | 0.37 | 0.87 | 0.88 | 0.74 | 0.88 |
| SPECI | 0.95 | 0.22 | 0.60 | 0.85 | 0.75 |
| PP VAL | 0.72 | 0.26 | 0.80 | 0.72 | 0.79 |
| NP VAL | 0.82 | 0.84 | 0.84 | 0.82 | 0.65 |
| PREV | 0.24 | 0.24 | 0.70 | 0.68 | 0.72 |
| DR | 0.09 | 0.21 | 0.72 | 0.65 | 0.82 |
| DP | 0.12 | 0.80 | 0.80 | 0.75 | 0.71 |
| BA | 0.66 | 0.54 | 0.69 | 0.63 | 0.83 |
| PC | LP | LP | LP | LP | LP |

*DT – Decision Tree, NB – Naive Bayes, RF – Random Forest, SVM – Support Vector Machine, ANN – Artificial Neural Network, Acc – Accuracy, NIR – Number of Information Rate, SENS – Sensitivity, SPECI – Specificity, PP – Positive Prediction Value, NP – Negative Prediction Value, PREV – Prevalence, DR – Dectection Rate, DP – Detection Prevalence, BA – Balance Accuracy, PC – Positive Class, LP – Liver Patient*

### *A. Decision Tree*

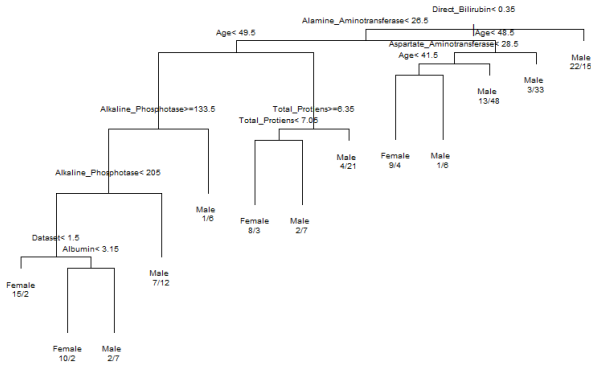


Figure2. *Decision Tree*

### *B. Random Forest*

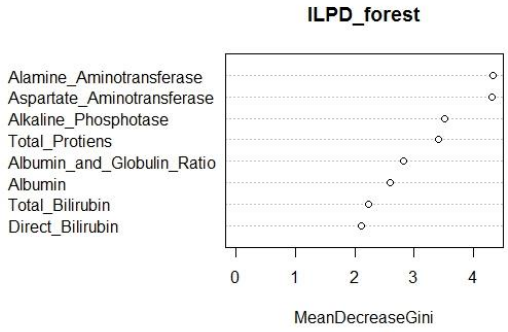


Figure3. *Random Forest*

### *C. Support Vector Machine*

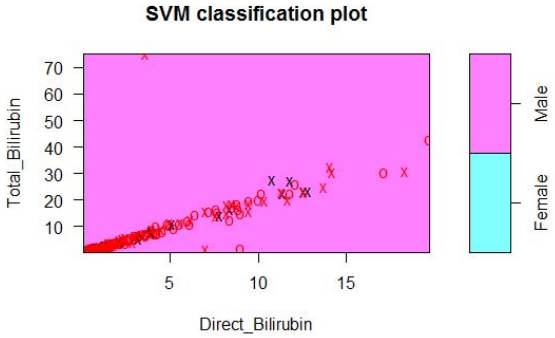


Figure4. *Support Vector Machine*

### *D. Artificial Neural Network*

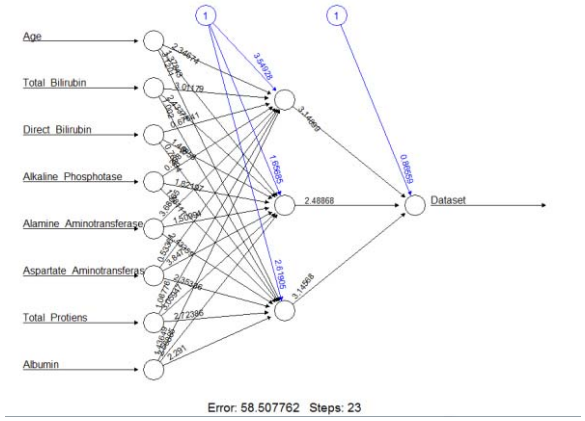


Figure5. *Artificial Neural Network*

## VII. CONCLUSION

Distinguishing proof of liver contamination at preparatory stage is imperative to battle the recurrence and seriousness condition of patients in India. The patients must be screened in view of beginning indications for improvement of customized treatment. In this investigation, an endeavor is made for forecast of liver sickness in patients utilizing machine learning techniques. It has been seen that decision tree gives better outcomes contrast with other order classification algorithms

## REFERENCES

[1]. Ashwani Kumar, Neelam Sahu, "Categorization of Liver Disease Using Classification Techniques", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 5 Issue V, May 2017, IC Value: 45.98 ISSN: 2321-9653.
[2]. Chandrasegar Thirumalai, IEEE Member, Rashad Manzoor, "Cost Optimization using Normal Linear Regression Method for Breast Cancer Type I Skin", International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.
[3]. Harsha Pakhale, Deepak Kumar Xaxa, "Development of an Efficient Classifier for Classification of Liver Patient with Feature Selection", International Journal of Computer Science and Information Technologies, Vol. 7 (3), 2016, 1541-1544.
[4]. Dr. S. Vijayarani, Mr.S.Dhayanand, "Liver Disease Prediction using SVM and Naïve Bayes Algorithms", International Journal of Science, Engineering and Technology Research (IJSETR) Volume 4, Issue 4, April 2015.
[5]. Anju Gulia, Dr. Rajan Vohra , Praveen Rani "Liver Patient Classification Using Intelligent Techniques", International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5110-5115.
[6]. Kalyan Nagaraj and Amulyashree Sridhar, "NeuroSVM: A Graphical User Interface for Identification of Liver Patients", IJCSIT. 5(6): 8280-8284 (2014).
[7]. Jankisharan Pahareeya, Rajan Vohra, Jagdish Makhijani , Sanjay Patsariya, "Liver Patient Classification using Intelligence Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February 2014.
[8]. Bendi Venkata Ramana, Surendra. Prasad Babu. M, Venkateswarlu. N.B, "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis", International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011 page no 101-114.
[9]. Kotsiantis. S.B, "Increasing the Classification Accuracy of Simple Bayesian Classifier", AIMSA, pp.198-207, 2004.
[10]. V.Kirubha, S.Manju Priya, "Survey on Data Mining Algorithms in Disease Prediction", International Journal of Computer Trends and Technology (IJCTT) – Volume 38 Number 3 - August 2016.
[11]. Bendi Venkata Ramana, Prof. M. S. Prasad Babu and Prof. N. B. Venkateswarlu, â€œA "Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysisâ€", International Journal of Computer Science Issues, ISSN: 1694 -0784, May 2012.
[12]. Database collected from ILPD(Indian Liver Patient Dataset)Dataset using UCI machine LearningRepository:https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset).
[13].A.S.Aneeshkumarand C.JothiVenkateswaran, "Estimating the Surveillance of Liver Disorder using Classification Algorithms", International Journal of Computer Applications (0975 – 8887), Volume 57– No.6, November 2012.
[14]. S. Karthik A, Priyadarishini, J. Anuradha and B. K. Tripathy (2011). "Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types", Advances in Applied Science Research.
[15]. Esraa M Hashem, Mai S Mabrouk (2014). "A Study of Support Vector Machine Algorithm for Liver Disease Diagnosis", American Journal of Intelligent Systems. 4(1): 9-14.
[16]. Ebenezer Obaloluwa Olaniyi khashman Aadnan, "Liver DiseaseDiagnosisBasedon Neural Networks" , Advances in Computational Intelligence, Proceedings of the 16th International Conference on Neural Networks (NN '15), November 7-9, 2015.
[17]. P. Thangarajul,R.Mehala, Performance Analysis of PSO-KStar Classifier over Liver Diseases, International Journal of Advanced Research in Computer Engineering, 2015.
[18]. P. Mazaheri, A. Narouzi and A. Karimi (2015), Using Algorithms to Predict Liver Disease Classification, Electronics Information and Planning. 3 :255-259.
[19]. H. Jin , S. Kim and J. Kim (2014), Decision Factors on Effective Liver Patient Data Prediction, International Journal of Bio-Science and Bio-Technology.
6(4): 167-178.
[20]. Reetu andN. Kumar (2015), Medical Diagnosis for Liver Cancer Using Classification Techniques, International Journal of Recent Scientific Research. 6(6): 4809-4813.
[21]. R. Sathya, Annamma Abrahim (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. International Journal of Advanced Research in Artificial Intelligence. 2(2): 34-38.
[22]. P.Rajeswari, G.Sophia Reena (2010). Analysis of Liver Disorder using Data Mining Algorithm. Global Journal of Computer Science and Technology. 10(14): 48-52.
[23]. Jankishran Pahariyavohra, Jagdeesh makhijani and sanjay patsariya. Liver patient classification using intelligence techniques. International journal of advanced research in computer science and software engineering. 4(2): 295-299.
[24]. Prasad Babu et. al. (2014). An implementation of hierarchical clustering on Indian Liver Patient Dataset. International Journal of Emerging Technologies in Computational and Applied Sciences. 8(6): 543-547.
[25]. S. Karthik A, Priyadarishini, J. Anuradha and B. K. Tripathy (2011). Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types. Advances in Applied Science Research.

L. Alice Auxilia studying MS (Software Engineering) at VIT University. She writes about Machine Learning Techniques that has been applied in the real time system particularly with the dataset of Indian Liver Patient. She has published papers in the international journals and presented research papers in international and national conferences. You can contact her through mail aliceauxilia@gmail.com.