

Using Artificial Neural Network and Machine Learning Algorithms to Scrutinize Liver Diseases

Tanu Sharma

Techno India NJR Institute of Technology, 313001 Udaipur, Rajasthan , India

Gaurav Kumawat

Techno India NJR Institute of Technology, 313001 Udaipur, Rajasthan , India

Prasun Chakrabarti

Techno India NJR Institute of Technology, 313001 Udaipur, Rajasthan , India

Sandeep Poddar

Lincoln University College, 47301 Petaling Jaya, Selangor D. E., Malaysia

Tulika Chakrabarti

Sir Padampat Singhania University Udaipur 313601, Rajasthan, India

Ali-Mohammad Kamali

Department of Neuroscience, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran

Babak Kateb

Director, National Center for NanoBioElectronics; Director, Brain Technology and Innovation Park; President, Brain Mapping Foundation; Chairman and CEO, Society for brain mapping and Therapeutics and, Chairman of Neuroscience20-G20 summit, Los Angeles, CA, USA

Mohammad Nami (✉ mtneurosci2@gmail.com)

Department of Neuroscience, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran

Research Article

Keywords: Classification, SVM, Neural Network, C5.0, Random Tress

Posted Date: March 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-324049/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Using Artificial Neural Network and Machine Learning Algorithms to Scrutinize Liver Diseases

Tanu Sharma¹, Gaurav Kumawat¹, Prasun Chakrabarti¹, Sandeep Poddar², Tulika Chakrabarti³,
Ali-Mohammad Kamali⁴, Babak Kateb^{5*}, Mohammad Nami^{4, 6,7*}

¹Techno India NJR Institute of Technology, 313001 Udaipur, Rajasthan, India

²Lincoln University College, 47301 Petaling Jaya, Selangor D. E., Malaysia

³Sir Padampat Singhania University Udaipur 313601, Rajasthan, India

⁴Department of Neuroscience, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran

⁵Director, National Center for NanoBioElectronics; Director, Brain Technology and Innovation Park; President, Brain Mapping Foundation; Chairman and CEO, Society for brain mapping and Therapeutics and, Chairman of Neuroscience20-G20 summit, Los Angeles, CA, USA

⁶Neuroscience Center, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT AIP), City of Knowledge 084301103, Panama

⁷Inclusive Brain Health, Swiss Alternative Medicine, Geneva, Switzerland

***Corresponding authors:** Mohammad Nami MD, PhD, Department of Neuroscience, School of Advanced Medical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran. torabinami@sums.ac.ir and, Babak Kateb MD, Society for Brain Mapping and Therapeutics (SBMT), Los Angeles, CA, USA babak.kateb@worldbrainmapping.org

Abstract

In this recommendation, liver incurable record is investigated for structure gathering replica to foresee liver disorder. This proposal actualized a component replica development and near investigation for refining the forecast exactness of Indian liver incurable in three stages. In the primary stage, the min-max standardization calculation is put into the unique liver incurable record accumulated from the UCI document. In liver incurable conjecture the second stage, by the use of PSO characteristic decision, a portion of the liver incurable record from the sum normalized liver incurable record is gained which includes simply enormous impute. In the third stage, portrayal figuring's are applied to the educational list. In this paper, an introduction evaluation between various estimation: Random Trees, Neural Network, eXtreme Gradient Boosting, Support Vector Machine (SVM), C5.0. The rule objective is to survey the rightness in social affair information concerning the benefit and plausibility of every calculation with respect to the accuracy, precision, affectability, and unequivocally. Exploratory outcomes show that the Neural Network gives the most fundamental accuracy (93.48 %) with the least blunder rate. All assessments are executed inside a redirection environment and drove in SPSS information mining device.

Keywords: Classification, SVM, Neural Network, C5.0, Random Tress.

1. Introduction

The liver is the biggest mety and strong organ that plot on the correct flank of the tummy. Around 3 pounds in weight, rosy earthy colored in shading and rubbery in nature. The liver has two significant segments which are known as the correct projections and the left flaps and are additionally viewed as an organ since it makes and secretes bile. Bile is a liquid that contains water, synthetic substances, and bile acids which are produced using put away cholesterol in the liver. Bile is put away in the gallbladder and when food enters the duodenum (the chief piece of the small digestive system) it causes our body to separate food and ingest fats from that food. In the upper right segment of the mid-region is the liver which is fenced by the ribcage. Liver sicknesses are messes in liver capacity which brings about disease. Hepatic infection is likewise called liver illness. The liver does numerous basic capacities inside the body and on the off chance that it gets broken, it prompts harm in the body. The liver infection causes all the issues that can prompt liver disappointment. Usually, if more than 75% or three-quarters of liver tissues are affected, the dysfunction occurs and puts the liver functions in jeopardy[1].

2. Issue Statement and Background Knowledge

Rajeswari *et al.*, (2010) has put forth that the information grouping depends on the liver problem. The preparatory dataset is created by gathering information from the UCI vault, which comprises of 345 occurrences with 7 unique credits. This paper achieves the area of fact portrayal acquire by Naïve Bayes calculations. FT tree figurings, and K Star estimations and by large execution made known by FT Tree computation when taken a stab at liver disorder record, time is taken to sprint the record for the outcome is rapid when diverged from various counts with a precise of 97.10%. Construct on the preliminary outcomes the game plan precision is found to be finer using FT Tree estimation stand out from various counts [2].

Alfisahrin *et al.*, (2013) have anticipated seeing if the victim has the liver illness exposed to the “ten” immense characteristics of liver infirmity utilizing a Decision Tree, Naive Bayes, and NB Tree assessments. The outcome appears NB Tree check is the most exact technique; at any rate, the Naïve Bayes calculation puts the speediest assessment decade. For the impending evaluation, the introduction of the NB Tree check would be the objective of the advancement of precision by searching the basic facts in perceiving liver difficulty victim. Later assessment, the presentation of the NB Tree assessment will be the objective of the advancement of precision by knowing the ruling fact in seeing liver illness victim [3].

Dhamodharan (2014) has anticipated that numerous liver issues need clinical consideration of the doctor. They foresee 3 significant liver infections, for example, liver malignant growth, cirrhosis, hepatitis with the assistance of unmistakable side effects. The essential objective is to anticipate the class types from classes, for example, liver malignancy, cirrhosis, hepatitis, and "no infections". In this paper, Naïve Bayes and FT tree calculation precision are looked at and the outcome is gotten. The outcome reasons that the exactness of the Naïve Bayes calculation is far superior to different calculations. [4]

Seker *et al.* (2014) have anticipated and implement the information mining strategies, for example, KNN, SVM, MLP, or choice trees on a remarkable dataset, which is gathered from 16,380 examination outcomes over a decade. This investigation can be valuable for additional examinations like lessening the quantity of examination since the forecast can be associated and besides the connection can be used for recognizing the peculiarity of the investigation [5].

Kumar *et al.*, (2015) have anticipated portrays the order of liver problem through component determination and fluffy K-implies grouping. Different liver problems additionally share the same property estimations and it needs more exertion to group liver issue type accurately with essential ascribes. So Fuzzy based arrangement gives finer execution in these confounding classes and accomplished over 94 rate precision for every kind of liver problem [6].

Thangarajul *et al.*, (2015) have anticipated examining the information of liver infections utilizing molecule swarm streamlining calculation (PSO) with K Star order. In two perspective for ordering the presence of infection are not. The anticipated calculation upgraded the exhibition of exactness when contrasted being characterization calculations. The PSO-Star calculation is the best reasonable calculation for the order of liver issues as it improved the exhibition in forecast exactness as examined before. The PSO-KStar calculation is considered is one of the great information mining calculations for understandability, changeability, and 100% precision. [7]

Gregory (2015) have anticipated two genuine liver victim record were explored for erection characterization pattern to foresee liver determination. Eleven information excavate characterization calculations were implement to the record and the exhibition of all assorted are looked at opposite one another as far as exactness, accuracy, and review. Because of the test results, the arrangement exactness is discovered to be better utilizing FT Tree calculation contrast with different calculations., it additionally shows the upgraded

presentation as indicated by the qualities furthermore, it gives 38.2% specificity, 86.4% sensitivity, 77.5% precision, and 78.0% accuracy results independently. [8]

Vijayarani (2015) have anticipated liver illnesses utilizing order calculations in their examination slog. The calculations applied in this slog are Naïve Bayes and backing support vector machine. Examinations of these calculations are tender and the exhibition factors for the characterization are precision, implementation time. From the outcomes, this work draws the derivation that the SVM classifier is contemplated as the finest reckoning calculation because of its mainhighassembleto the point esteems. Once again, while contrasting the implementation time, it was seen that the Naïve Bayes classifier needs the smallestimplementation time and from the usage results it is inferred that the SVM is an unrivaled Classifier for anticipating the liver contaminations and taking a gander at the execution time, the Naïve Bayes classifier require smallest execution time. [9]

Olaniyi *et al.* (2015) implement back proliferation neural organization and outspread premise work neural organization are intended to analyze these illnesses and furthermore forestall misdiagnosis of the liver lines issue victim. The calculations were contrasted and the C4.5, CART, Naïve Bayes, Support Vector Machine and presumed that the outspread premise work the neural organization is the ideal way since it has an acknowledgment pace of 70% which has demonstrated much precise and effective than different calculations. [10]

Baitharua *et al.*, (2016) implement center around the part of curing analysis by knowing design by the gathered information of liver issue to create clever clinical choice emotionally supportive networks to support the doctors. In this paper the utilization of a few arrangement calculations to characterize these infections and look at the viability, rectification rate among them. In this paper, a similar examination of information grouping precision utilizing liver problem information in various situations is introduced. The prescient exhibitions of famous classifiers are looked at perceptible. By investigating the outcomes multifacetedapproach gives the general finest arrangement conclusion with the precision of 71.59% than any other algorithm [11].

Liver cancer diagnosis using IBK, NNge and simple cart algorithms based on BUPA liver disorder dataset was demonstrated by Tiwari *et al.*[12]

Gulia *et al.*, (2014) implement hybrid replica turn of events close to assessment for better figure precision of liver victim in three stages. In the main stage, portrayal figurings are implemented to the primary liver victimrecord accumulated from the UCI storage facility. In the subsequent stage, by the usage of incorporate decision, a group of the liver victim from sum liver victim record is acquired which includes figuratively speaking basic characteristics and thereafter applying picked portrayal computations on gained, an immense subset of credits. SVM count is appraised as the finer execution computation since it gives high precision in individual to other portrayal figurings preceding smearing incorporate assurance. Be that as it may, Random Forest computation is considered as the better execution estimation aftersmearingcharacteristic assurance. In the third stage, the results of portrayal estimations with and without incorporated assurance are differentiated and each other. The conclusion procured from our tests show that Random Forest figuring beat all various systems with the aid of characteristic assurance with a precision of 71.8696% [13]

.3. Methodology

Patients with the Liver illness have been consistently expanding a direct result of unreasonable utilization of liquor, breathe in destructive gases, admission of defiled food, pickles, and medications. This dataset was utilized to assess expectation calculations with an end goal to lessen trouble on specialists. This informational file holds 167 non-liver patient data, 416 liver victim sets accumulated from Andhra Pradesh, India. The "Data" area is a category name used to isolate social events into the liver victim or not. This enlightening file holds 441 male, 142 female patient records. This dataset was downloaded from the Kaggle.

Table 1: Record columns

| S.No. | Column | explanation |
|-------|--------------------------|--|
| 1. | Age | Patient age |
| 2. | Gender | Patient Gender |
| 3. | Total_Bilirubin | Total Bilirubin |
| 4. | Direct_Bilirubin | Direct Bilirubin |
| 5. | Alkaline_Phosphotase | Alkaline Phosphotase |
| 6. | Alamine_Aminotransferase | Alamine Aminotransferase |
| 7. | Aspartate_Aminotr | Aspartate Aminotransferase |
| 8. | Total_Protiens | Total Protiens |
| 9. | Albumin | Albumin |
| 10. | Albumin_and_Globu | Albumin and Globulin Ratio |
| 11. | Record | field used to split the data into two sets (patient with liver disease, or no disease) |

3.1 Random Trees

Arbitrary woods are an AI relapse technique for grouping that is driven by developing liver information into countless decision trees at getting ready time and yielding the class that is the technique for the categoryrelentthrough solitary sapling.Its yields characterize adequately on the vast liver dataset. It can dispense with a piece of extraordinary numerous data attributes without variable deletion. It offers a gauge of what factors are of the most extreme significance in the characterization. Arbitrary Forests utilizes numerous grouping trees. The backwoods pick the characterization having the main extreme number counts.

3.2 C5.0

While there is various usage of choice trees, one of the most notable is the C5.0 calculation. The C5.0 calculation has become the business standard for delivering choice trees since it does well for most kinds of issues straightforwardly out of the case. Contrasted with further developed and refined AI models (for example Neural Networks and Support Vector Machines), the choice trees under the C5.0 calculation, for the most part, perform almost also yet are a lot more obvious and convey. One of the advantages of the C5.0 calculation is that it is obstinate about pruning; it deals with a considerable lot of the choices naturally utilizing genuinely sensible defaults. Its general technique is to defer the tree. It does this by first growing an enormous tree that overfits the preparation information. Subsequently, hubs and branches that have little impact on the characterization mistakes are taken out.

3.3 Neural Network

A receptive association isa progression of computations that attempt to see covered associations in a lot of data through a cycle that impersonates the working of the human cerebrum. In this sense, neural associations insinuate structures of neurons, either common or phony. Neural associations can adjust to developing the information; along these lines, the association creates the best result without hoping to refresh the yield models.

3.4 SUPPORT VECTOR MACHINE (SVM)

It has pulled in a ton of thought in the latest decagon, also, adequately attempted to numerous zones petition. SVMs are commonly worn for the swotting plan, backslide, or situating work. SVM relies upon quantifiable learning speculation and fundamental risk minimization head and has the objective of choosing the territory of conclusion restricts in any case called hyperplane that produces the ideal division of category. SVM is the main generous and cautious plan approach, there are various issues. The record assessment in SVM relies upon bent boxy programming, and it is cipheringexpensive, as clarifying boxy programming procedures need gigantic framework exercises similarly to dreary numerical figuring.

3.5 XGBoost (eXtreme Gradient Boosting)

The execution of the estimation was intended for figuring time and memory resources. An arrangement objective was to use open resources to arrange the model. Some key figuring use features incorporate:

- Sparse mindful use with the modified treatment of missing data regards.
- Block structure to help the parallelization of tree advancement.
- Continued priming so you'll add support an all-around fitted model on new information

3.7 Gini Coefficient

The Gini coefficient was proposed by Gini as an extent of the uniqueness of payor wealth. The Gini coefficient measures the unevenness among assessments of a repeated transport (for example, levels of pay). A Gini coefficient of zero imparts stunning equilibrium, where all characteristics are the same (for example, where everyone has a comparable compensation). A Gini coefficient of one (or 100%) conveys maximal uniqueness among values (e.g., for innumerable people where simply a solitary individual has all the compensation or usage and all others have none, the Gini coefficient will be very nearly one).

3.8 AUC

AUC signifies "Locale under the ROC Curve". That is, AUC checks the whole two-dimensional locale underneath the whole ROC turn (think indispensable math) from (0,0) to (1,1). AUC gives a full-scale degree of execution overall conceivable social occasion limits. One procedure for interpreting AUC is as the likelihood that the model positions an inconsistent positive model more altogether than an abstract negative model.

3.9 Precision

An assessment about lay, accuracy is the density of the assessments to a specific benefit, while precession is the closeness of the assessments to each other.

3.10 Recall

In plan affirmation, information recuperation, and portrayal, precision is the piece of relevant cases among the recuperated events, while an audit is the bit of the total amount of appropriate models that were truly recuperated.

3.11 Classification Matrix

An arrangement framework sorts all cases from the model into orders, by choosing if the foreseen worth composed the genuine worth. All the cases in each class are then checked, and the totals have appeared in the framework. The gathering system is a standard instrument for the evaluation of real models and is now and again suggested as a disordered matrix. The layout that is made when you pick the Classification Matrix decision investigates genuine to foreseen characteristics for each foreseen express that you show. The lines in the structure address the foreseen characteristics for the model, however, the fragments address the veritable characteristics. The classes used in the examination are fake positive, authentic positive, counterfeit negative, and real negative. A gathering system is a huge contraption for assessing the delayed consequences of the figure since it makes it clear and speaks to the effects of wrong desires.

3.12 Classification Accuracy

Grouping precision is fundamentally the movement of right game plans, either for a self-governing test set, or using some assortment of the cross-endorsement thought.

3.14 Classification Error

The course of action Error happens although $Y \neq g(X)$. The finest assorter g^* , known as the Bayes assorter, is the one that limits the likelihood of approach mess up.

4. Result

The dataset has been divided into two parts Training and Testing. IBM SPSS 18.2.1 has been used for finding the result of the model and used 80% training data and 20% testing data. In the wake of applying the pre-dealing with and status methodologies, we endeavor to separate the data ostensibly and figure out the dissemination of characteristics concerning suitability and profitability.

Table 2: Dataset Observation

| Models | Precision | Recall | F-Measure | AUC | Gini Coefficient | Accuracy |
|------------------------|-----------|--------|-----------|------|------------------|----------|
| Random Trees | 0.86 | 0.86 | 0.86 | 0.92 | 0.85 | 84.78% |
| Neural Network | 0.95 | 0.95 | 0.95 | 0.93 | 0.86 | 93.48% |
| Discriminant | 0.88 | 0.54 | 0.67 | 0.75 | 0.50 | 71.18% |
| Support Vector Machine | 0.78 | 0.69 | 0.73 | 0.77 | 0.53 | 72.90% |
| XGBoost Tree | 0.92 | 0.88 | 0.90 | 0.95 | 0.90 | 89.29% |
| C5.0 | 0.90 | 0.92 | 0.91 | 0.98 | 0.96 | 90.61% |

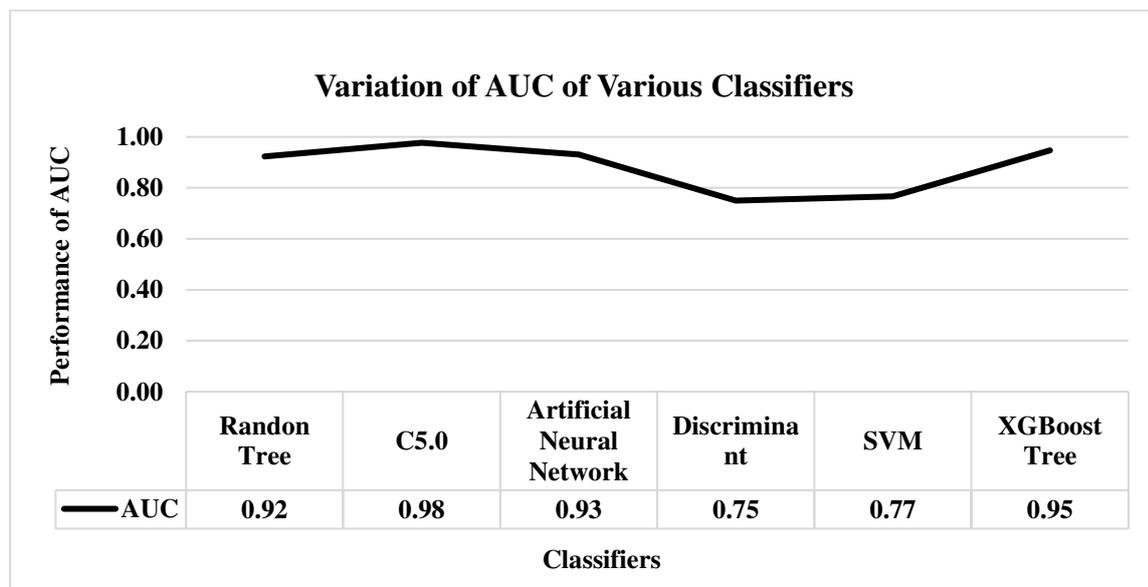


Figure 1. AUC grid.

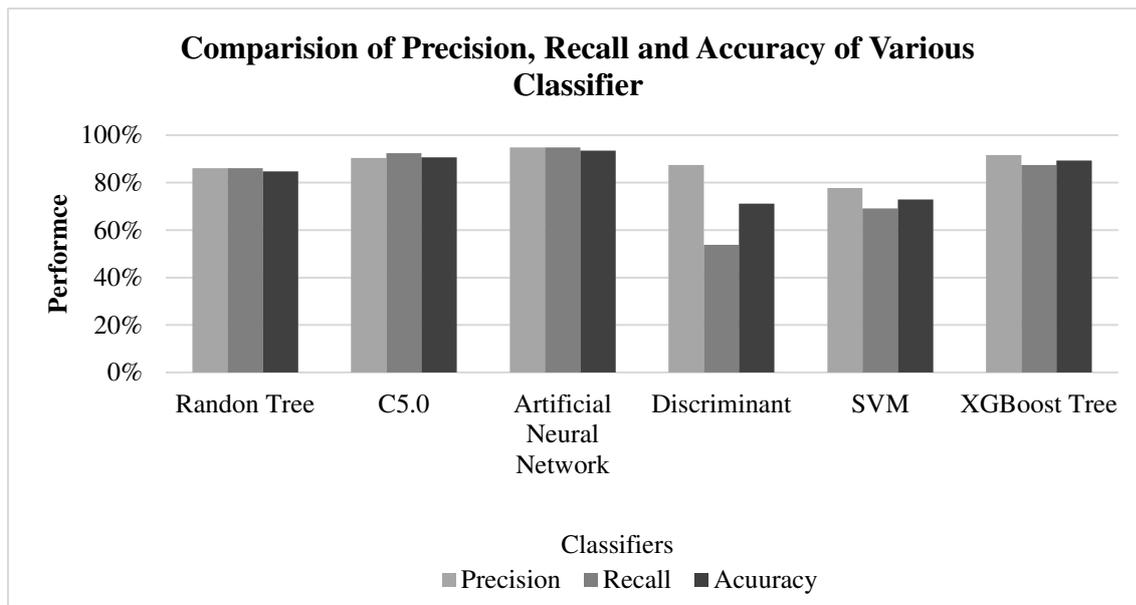


Figure 2. Precision, Recall and Accuracy Graph.

5. Conclusion

The present article plans to foresee the liver illnesses dependent on substance fusion here in the person structure. Trial such as SGOT, SGPT the result referenced endure individual shows restraint i.e. should be analyzed or not. In this perspective, six classifiers have been applied viz SVM, Random trees, C5.0, XGBosst tree, Neural Network, Discriminant tree. It has been seen that the Neural Network classifier is giving the most elevated precision with 93.48 %.

6. Future Work

There is an extension to further diminish scan space for better liver arrangement precision if upgraded choice and change methodology is being utilized. The future procedure is utilized to break down the liver area into detachable compartments for example liver and so on, in any case, the technique requires further improvement generally concerning include a choice of the dwellertowardnumerous segments: bladderlayer, bladdersection, bladdergist, bladderbreastbone. Separated through, it is wanted to grow the information base on which the framework will be tried. furthermore, strategy in this postulation can be utilized for recognizing the heart illnesses laterthrough heart record and order of the infections.

7. Acknowledgment

The authors are thankful to Techno India NJR Institute of Technology, Udaipur, Rajasthan, India for providing necessary resources and infrastructure and encouragement to conduct this project work and to Lincoln University College, Malaysia for academic support.

8. References

- [1] https://www.medicinenet.com/liver_disease/article.htm
- [2] Rajeswari P, Reena GS. Analysis of liver disorder using data mining algorithm. Global journal of computer science and technology. 2010 Nov 25.
- [3] Alfi Sahrin SN, Mantoro T. Data mining techniques for optimization of liver disease classification. In 2013 International Conference on Advanced Computer Science Applications and Technologies 2013 Dec 23 (pp. 379-384). IEEE.

- [4] Dhamodharan S. Liver disease prediction using bayesian classification. In4th National Conference on Advanced computing, applications & Technologies 2014 May (pp. 1-3).
- [5] Seker SE, Unal Y, Erdem Z, Kocer HE. Ensembled correlation between liver analysis outputs. arXiv preprint arXiv:1401.6597. 2014 Jan 25.
- [6] Aneeshkumar AS, Venkateswaran CJ. A novel approach for Liver disorder Classification using Data Mining Techniques. Engineering and Scientific International Journal. 2015 Mar;2(1):15-8.
- [7] Thangaraju P, Mehala R. Performance analysis of PSO-KStar classifier over liver diseases. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). 2015 Jul;4(7).
- [8] Priya MB, Juliet PL, Tamilselvi PR. Performance analysis of liver disease prediction using machine learning algorithms. International Research Journal of Engineering and Technology (IRJET). 2018 Jan;5(1):206-11.
- [9] Vijayarani S, Dhayanand S. Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR). 2015 Apr;4(4):816-20.
- [10] Olaniyi EO, Adnan K. Liver disease diagnosis based on neural networks. Advances in Computational Intelligence. 2013:48-53.
- [11] Baitharu TR, Pani SK. Analysis of data mining techniques for healthcare decision support system using liver disorder dataset. Procedia Computer Science. 2016 Jan 1;85:862-70.
- [12] Tiwari M., Chakrabarti T., Poddar S., Laxkar P., Sharma M., Majumder R. , Chakrabarti P. Liver cancer diagnosis using IBK, NNge and simple cart algorithms based on BUPA liver disorder dataset .The Patent Office Journal No. 13/2020 Dated 27/03/2020 page 15796
- [13] Gulia A, Vohra R, Rani P. Liver patient classification using intelligent techniques. International Journal of Computer Science and Information Technologies. 2014;5(4):5110-5.

Figures

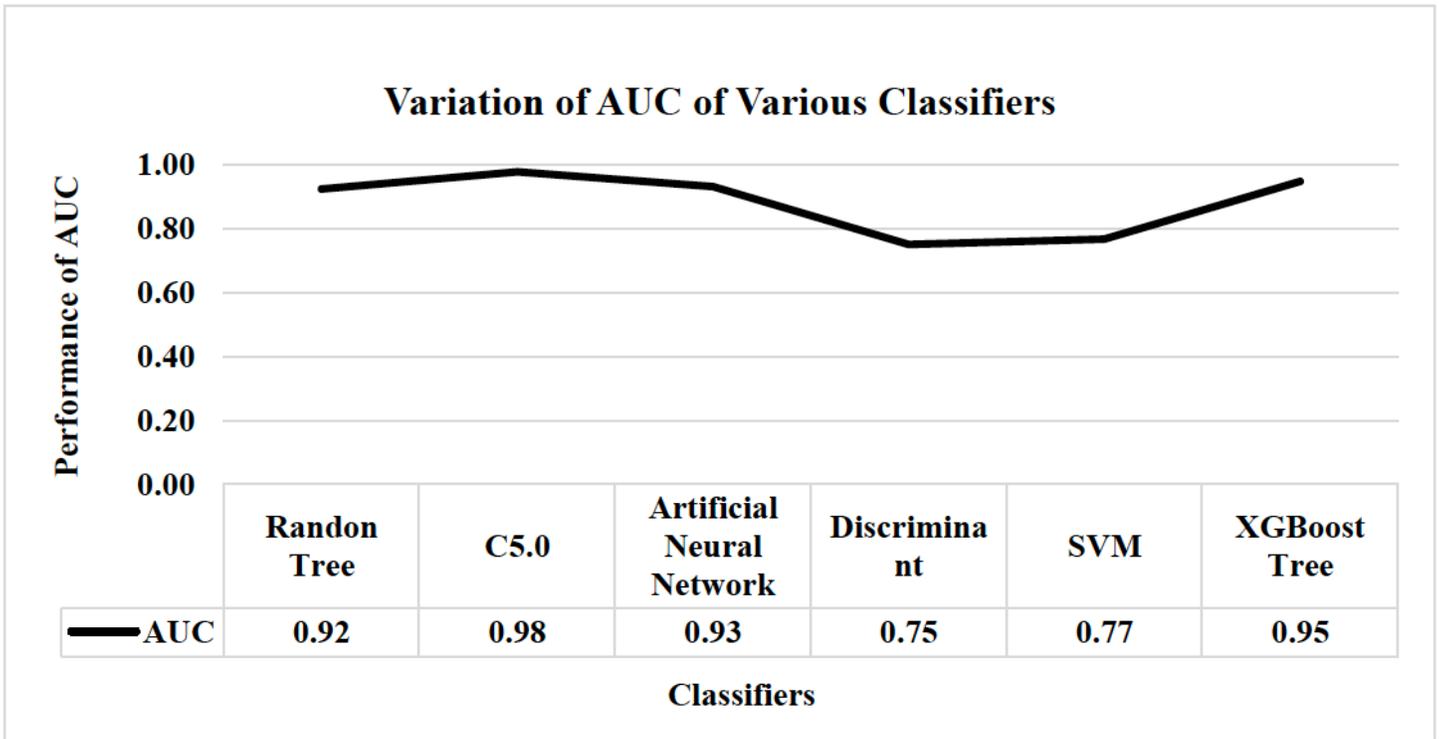


Figure 1

AUC grid.

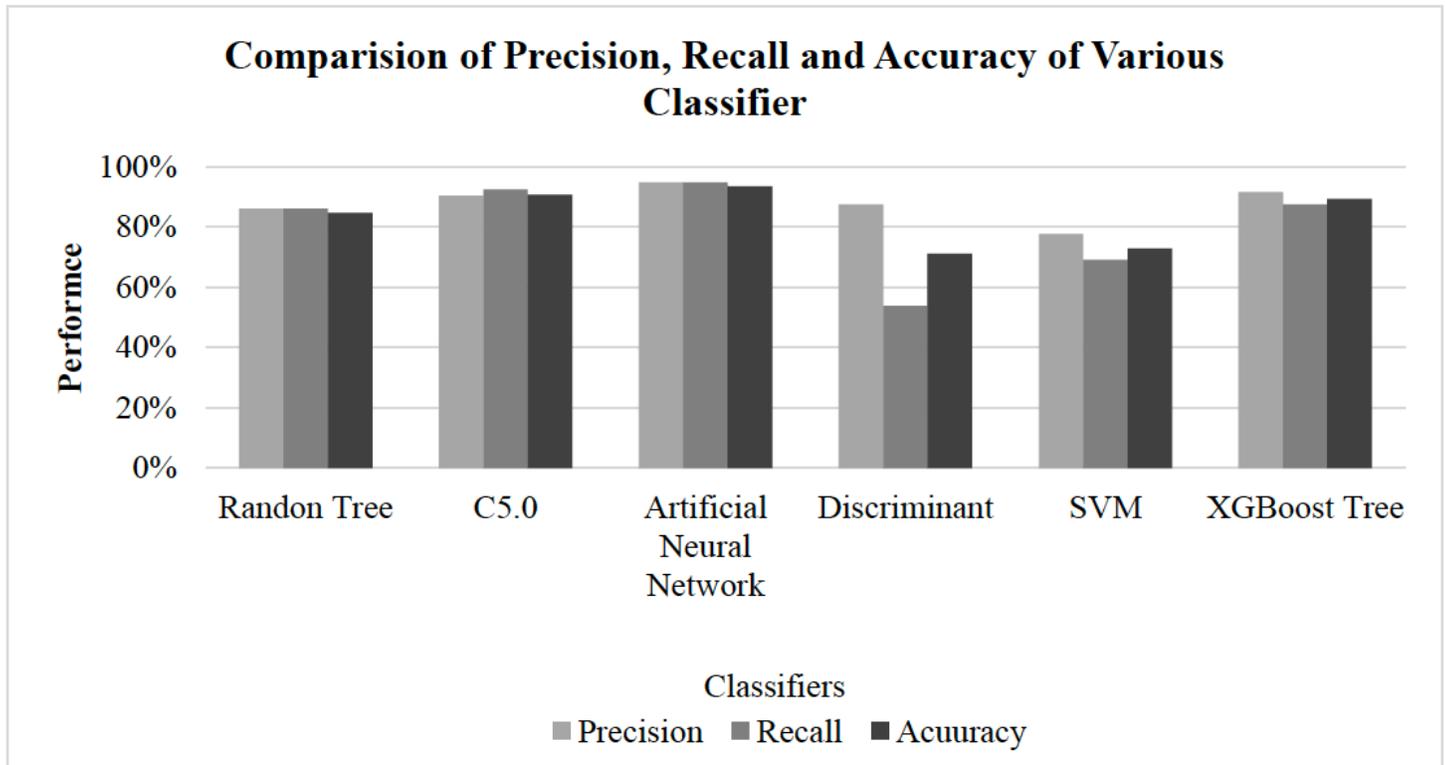


Figure 2

Precision, Recall and Accuracy Graph.