Volume 9, Issue 7, July – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1960

Liver Failure and Cirrhosis Prediction- Using Methods for Machine Learning

[1]S. Selvakumar, [2]S Sanjay
[1]Assistant Professor, [2]Student,
Department of Computer Science, SRM Arts and Science College,
Kattankulathur - 603 203, Chengalpattu District, Tamilnadu, India

Abstract:- Liver disease continues to be a major global health concern, accounting for a considerable portion of global mortality. It results in a variety of symptoms such aberrant nerve function, blood in the cough or vomit, renal and liver failure, jaundice, and liver encephalopathy. It is caused by a myriad of variables that influence the liver, including obesity, untreated hepatitis infection, and alcohol misuse. In order to effectively treat liver infections, early detection is essential, and sensor-based medical technology is frequently used in modern medical procedures to identify illnesses. But diagnosing a condition can be expensive and difficult. Thus, the purpose of this paper is to compare the effectiveness of different machine learning algorithms in order to judge how well they function and have what potential to categorize liver diseases.

Keywords:- *Random Forest, Decision Tree, Support Vector Machine (SVM), Logistic Regression, and Categorization.*

I. INTRODUCTION

The liver is the biggest inside organ in the human body and is a vital organ for carrying out numerous substantial errands. such as digestion system, squander departure, and absorption. Working like a organ, it secretes bile to encourage assimilation. In expansion, the liver detoxifies substances, metabolizes drugs, and channels blood some time recently it is dispersed all through the body. The liver is pivotal for by and large wellbeing since it moreover makes crucial proteins for blood clotting and other organic capacities.

Liver infection is a predominant condition around the world, and the liver is helpless to different clutters. One common clutter is greasy liver, where huge fat vacuoles collect in liver cells, frequently related with over the top liquor utilization, but it can happen in non-drinkers as well.

Hepatitis ordinarily caused by viral contaminations transmitted through sullied substances Another liver clutter stems from coordinate contact with tainted body fluids.

Cirrhosis of the liver is a extreme shape of liver malady characterized by noteworthy cell misfortune, driving to the liver contracting and getting to be harsh and weathered. In spite of the fact that the liver endeavors to recover, the misfortune of liver cells surpasses the rate of substitution. Liver cancer is too a concerning condition influencing the liver.

The objective of this consider is to foresee and liver ailment utilizing information mining classification calculations in arrange to accomplish made strides execution exactness. To confirm and affirm the precision of the forecasts, different machine learning strategies are utilized.

II. LITERATURE REVIEW

- The paper titled "The Diagnosis of Chronic Liver Disease Using Machine Learning Techniques" by authors Golmei Shaheamlung and Harshpreet Kaur, published on March 26, 2021, addresses the increasing global prevalence of liver disease in the twenty-first century. With the mortality rate reaching 3.5%, liver disease has become a significant cause of death, emphasizing the need for early detection and prompt treatment to improve patient outcomes.

To accomplish precise determination, the consider utilizes a verifiable liver malady database and utilizes progressed machine learning procedures. By dissecting this comprehensive dataset, the analysts point to reveal covered up designs and fundamental data related to liver malady. The extreme objective is to give a dependable and effective strategy for diagnosing liver infections, which frequently display complex challenges and require exact answers for compelling quiet administration.

- In their 2018 paper titled "Survey on the Prediction and Analysis of Liver Diseases Using Data Mining Techniques," authors Shambel Kefelegn and Pooja Kamat explore liver disease, one of the most prevalent conditions worldwide.

The liver, being a basic organ and organ, plays noteworthy parts in distinctive physiological shapes, such as bile era and emanation, protein union, blood clotting, and absorption framework of fats, sugars, and press. Liver brokenness can lead to distinctive liver-related sicknesses and prosperity issues.

III. METHODOLOGY

The strategy of this ponder points to address different viewpoints related to liver malady classification. It centers on include determination and demonstrate execution utilizing distinctive classification calculations. The inquire about utilizes The Indian Liver Quiet Dataset (ILPD) starts from the UCI Machine Learning Store, and it serves as a important

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1960

dataset for analysts centering on liver malady expectation. This dataset contains different quiet traits, counting age, sexual orientation, liver work test comes about (such as bilirubin, ALT, AST), protein levels, and a parallel name demonstrating the nearness or nonappearance of liver malady. By utilizing the ILPD, analysts can create and assess information mining and classification calculations to precisely anticipate the probability of liver infection in patients based on these highlights. Its accessibility has encouraged noteworthy progressions in the field of liver infection forecast and empowered productive inquire about endeavors inside the logical community., comprising of 583 cases and ten natural variables. The dataset is isolated into yes (416 cases) and no (167 cases) classes, speaking to liver conditions.

Pre-processing strategies are utilized to handle lost values and normalize the information. Include determination is carried out utilizing channel and wrapper strategies, counting relationship examination and arbitrary woodlands. The chosen highlights are at that point utilized to build a classification model.

The dataset is randomized and partitioned into preparing (70%) and testing (30%) sets. The preparing set comprises of 389 cases, whereas the assessment set comprises the remaining 194 examples.

Multiple classification calculations, such as Choice Trees, Irregular Timberlands, and Back Vector Machines (SVM), are utilized for examination and expectation. SVM endeavors to categorize the information by making hyperplanes, whereas Choice Trees are utilized as a flowchart-like structure to classify information based on characteristics.

To guarantee the vigor of the approach, encourage testing on a significant free test set is suggested. Also, gathering broad therapeutic information and selecting the most significant highlights stay challenging perspectives in liver infection recognizable proof utilizing machine learning procedures. Tending to these challenges may include large-scale investigate and collaboration over numerous centers, centering on information collection, division, and pre-processing.

IV. DATA COLLECTION

For this specific think about, the analysts picked to utilize the Indian Liver Persistent Dataset (ILPD) sourced from the UCI Machine Learning Store. is expecting to be a agent test of all Indians and was drawn from the Andhra Pradesh locale. It comprises of 583 illustrations and incorporates data on ten distinctive natural components. The primary course esteem in the dataset is communicated as either "yes" or "no," speaking to liver conditions. Out of the 583 cases, 416 were as "yes" (showing liver malady), and 167 were as "no" (demonstrating no liver infection) based on the particular parameters related to these natural variables. The dataset's course dispersion is basic for preparing and assessing the machine learning models for liver infection forecas

V. PREPROCESSING

During the pre-processing stage, the analysts performed information arrangement and include determination. This included taking care of lost values, managing with exceptions, and normalizing or scaling the information as required. Also, highlight choice procedures were connected to recognize and hold the most important highlights that contribute essentially to the expectation of liver malady. These steps were vital to guarantee the data's quality and to upgrade the execution and interpretability of the machine learning models utilized in the think about. organize of the ponder, a few strategies were utilized to handle and normalize lost values in the Indian Liver Quiet Dataset (ILPD). Lost values were supplanted with invalid values to guarantee a steady and total dataset. Include determination was a significant step in planning the information for classification. Both channel and wrapper strategies were utilized for this reason. In the channel strategy, relationship investigation was conducted to recognize qualities with a relationship more noteworthy than 70%. These profoundly related qualities were disposed of from the dataset as they may present excess or pointless complexity to the classification handle. Also, the wrapper strategy utilized irregular timberlands to survey the significance of different properties in the dataset. Irregular timberlands are an outfit learning method that builds numerous choice trees and totals their yields, giving profitable experiences into the importance of distinctive highlights for classification. By applying these pre-processing and highlight choice procedures, the analysts pointed to upgrade the quality of the dataset and optimize the execution of the classification calculations utilized for liver malady expectation. To make an subjective test stage, the dataset was at first haphazardly created. The dataset was isolated into preparing (which made up 70% of the dataset) and testing (which made up 30% of the dataset). The remaining 194 illustrations made up the assessment set, clearing out a add up to of 389 cases in the preparing set.to of 389 cases in the preparing set.

VI. CLASSIFICATION ALGORITHMS

Classification algorithms are essential in infection forecast and play a significant part in programmed restorative wellbeing analyze. Among them, the Bolster Vector Machine (SVM) calculation is broadly utilized and known for its tall categorization exactness. The SVM calculation points to categorize information by making hyperplanes in a high-dimensional space. In Python, the scikit-learn library is commonly utilized to actualize SVM. The pre-processed information is part into two categories: tick information and coaching information, comprising 25% and 75% of the whole dataset, individually. SVM Back Vector Machines (SVM) work by recognizing the hyperplane that maximizes the edge, speaking to the remove between the hyperplane and the closest information focuses of each lesson. This hyperplane empowers a clear division between distinctive classes, making SVM a vigorous apparatus for exact classification assignments. A bigger edge is related with lower generalization blunder, making SVM especially capable for exact classification.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1960

In the setting of liver malady expectation, SVM can be viably utilized to recognize patients with liver malady from those without, based on the chosen highlights and their comparing classifications. This renders SVM an important and commonsense information mining strategy in infection determination and forecast, giving dependable comes about and helping in restorative decision-making.

VII. DECISION TREE

A decision tree is a visual representation taking after a flowchart, wherein each inside hub speaks to a "test" conducted on a particular characteristic (e.g., coin flip coming about in heads or tails). Each department compares to the result of the test, and each leaf hub implies a course name, speaking to the choice made after assessing all important qualities. Choice trees are effective instruments for classification errands, as they permit for an organized and natural approach to making choices based on different highlights. The classification rules are consecutively characterized from the root to the takes off. In commonsense applications with huge datasets, it has been watched that a few calculations may not abdicate ideal comes about, and they may perform superior when connected to littler subsets of the information or maybe than the whole dataset. The viability and affectability of diverse approaches can altogether change. Besides, certain methods might display varieties in exactness when managing with real-time information, highlighting the significance of cautious choice and assessment of calculations for particular utilize cases. Applying machine learning concepts to liver malady recognizable proof presents challenges, for investigate purposes, getting comprehensive restorative information is basic, and this includes gathering broad data important to the ponder of intrigued. Also, recognizing the most germane highlights from the collected information is vital, as it empowers analysts to center on the key properties that altogether affect the inquire about goals. By carefully selecting and analyzing these pertinent highlights, analysts can pick up profitable experiences and create precise models or expectations in the restorative space. for exact expectations. Overcoming these challenges may require collaborative large-scale investigate including numerous, emphasizing information collection, division, and preprocessing. To set up the vigor of the proposed approach, thorough testing on a considerable autonomous test set is fundamental. This approval prepare will help guarantee the unwavering quality and generalizability of the model's execution.

VIII. DATASET

The information is collected from UCI based on the given attributes. The Dataset utilized in the ponder comprise of 416 are emphatically tested. Based on information sorts the traits are given and ostensible information types. In our think about the dataset contains the add up to bilirubin, coordinate bilirubin, age, sexual orientation, add up to proteins, egg whites, egg whites and which is the indications of liver disease. The entire attribute and their types are given in Table 1. shown in Fig1 which is loaded in the Weka tool. The data are collected from UCI Machine Learning Repository [18] and it predicts attributes. Based on data types the attributes are given and nominal data types. In our study the dataset contains the total bilirubin, direct bilirubin, age, gender, total proteins, albumin, albumin and which is the symptoms of liver disease. The data sets consist of 538 liver and non The Dataset used in the study consist of 167 negatives tested for liver disease Class value “Yes” means having liver disease and “No” means negative.

Table 1: Attributes Dataset

Attribute Name	Possible value
Age of the patient	Numeric
Gender of the Patient	Nominal
Total Bilirubin	Numeric
Direct Bilirubin	Numeric
Alkphos Alkaline Phospotase	Numeric
Sgpt Alamine Aminotransferase	Numeric
Total proteins	Numeric
Albumin	Numeric
Albumin and Globulin Ratio	Numeric
Class	Nominal

IX. DATA ANALYSIS

A. Problem Identification

The essential point of this extend is to make and send progressed Machine Learning methods that can successfully learn comprehensive designs related with Liver Infections and construct a strong liver infection show. A key center of this conclusion is to optimize the proficiency of the calculations utilized to guarantee exact and dependable liver clutter expectations. In the domain of machine learning, inferential data sets are determined from commonly watched designs inside accumulated information or person datasets. These designs, which speak to repeating patterns, are set up through broad investigate and examination of a huge volume of pertinent information. In any case, when managing with restricted information, creating precise and self-evident inferential designs gets to be challenging, hence influencing the exactness of prescient examination. The dataset utilized in this ponder comprises of scores from five clinical trials, with each score extending from 1 to 3, where 3 shows noteworthy disintegration in organ condition.

Some use a modified version that takes into consideration the high amounts of conjugated bilirubin seen in these illnesses. The top limit for two points is 170 mol/L (10 mg/dL), whereas the maximum for a single point is 68 mol/L (4 mg/dL).

Table 2: Dataset Scores for Clinical Trails

Parameter	Assign 1 point	Assign 2 points	Assign 3 points
Ascitis	Absent	Slight	Moderate
Bilirubin (mg/dL)	< 2	2-3	>3
Albumin (g/dL)	>3.5	2.8-3.5	<2.8
Prothrombin time (second over control) or INR	<4 <1.7	4-6 1.7-2.3	>6 >2.3
Encephalopathy	None	Grade 1-2 (Mild to moderate)	Grade 3-4 (Severe)

Table 3: Summary of Attributes of the Dataset

Attribute	Description	Measurement	Value range	Mean	Std	25%	50%	75%
Age (AG)	Participant's age	Years	4-90	44.107	15.981	32	45	55
Gender (GN)	Participant's gender	Categorical	0 or 1	0.775	0.483	0	1	1
Total bilirubin (TB)	Total bilirubin level in the participant's blood	mg/dl	0.4–75	3.370	6.256	0.8	1	2.7
Direct bilirubin (DB)	Direct bilirubin level in the participant's blood	mg/dl	0.1–19.7	1.528	2.870	0.2	0.3	1.3
Alkaline phosphatase (AP)	Alkaline phosphatase level in the participant's blood	U/L	63-2110	289.075	238.538	175	209	298
Alanine aminotransferase (ALA)	Alanine aminotransferase level in the participant's blood	U/L	10-2000	81.489	182.159	23	35	62
Aspartate aminotransferase (ASA)	Aspartate aminotransferase level in the participant's blood	U/L	10-4929	111.470	280.851	26	42	88
Total proteins (TP)	Total protein level in the participant's blood	g/dl	2.7–9.6	6.480	1.082	5.8	6.6	7.2
Albumin (AL)	Albumin level in the participant's blood	g/dl	0.9–5.5	3.130	0.792	2.6	3.1	3.8
Albumin and globulin ratio (AGR)	Albumin and globulin ratio in the participant's blood	g/dl	0.3–2.8	0.943	0.323	0.7	0.9	1.1
Liver disease or not (LD)	If the participant has liver disease or not	Categorical	0 or 1	0.286	0.452	0	0	1

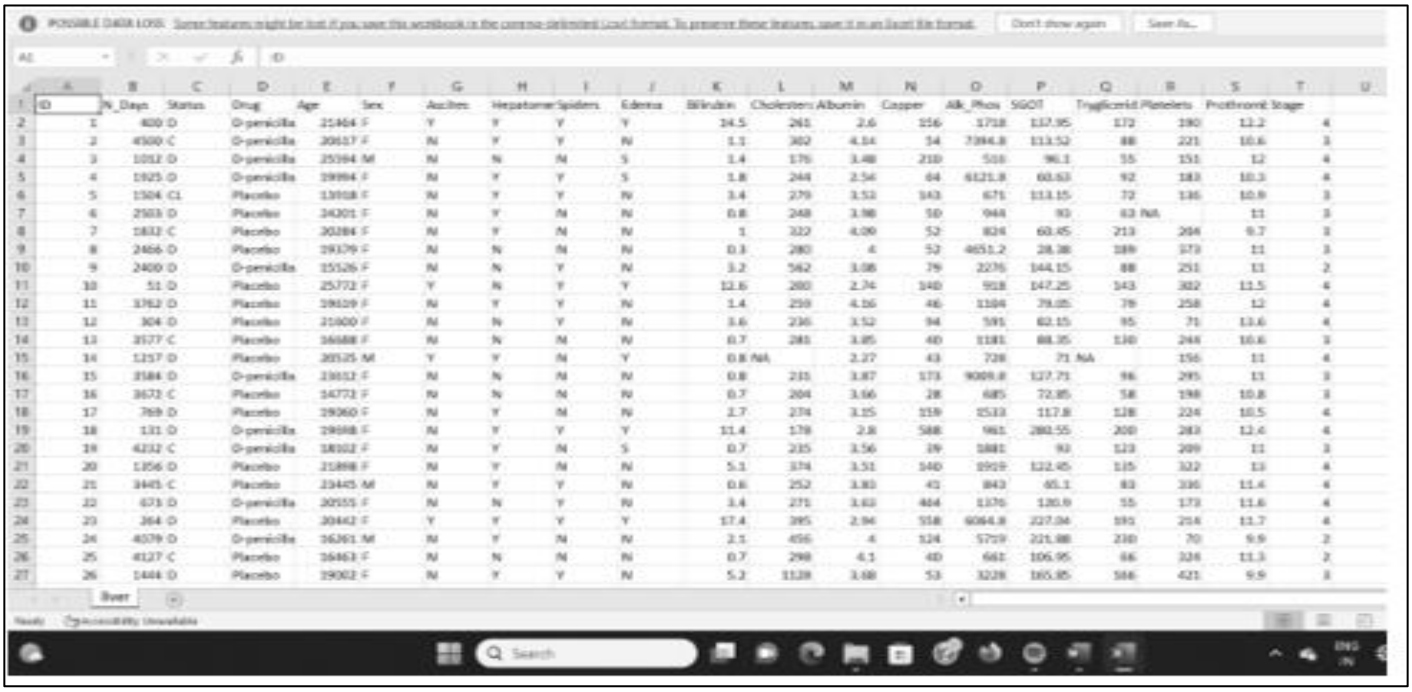

Fig 1: Dataset for Liver Disease Prediction

X. EXPERIMENTAL REPORT AND DATA ANALYSIS

The comparison of various decision tree algorithms performed on liver disease data is shown in Table 2. Summary of attributes of the dataset the highest accuracy rate. The accuracy rate of this algorithm is 70.67%. It is seen that Decision Stump is the most powerful classifier for this example. This result shows that the liver disease of a new patient is predicted successfully with an acceptable ratio 70.67%. It is also seen that J48 has the worst accuracy rate with 65.69%.

The comparison of decision tree calculation with regard to precision appeared in Fig 2. The comparison of decision tree calculation with regard to kappa measurements and runtime is appeared in Fig3 and Fig4 individually. Random Tree and Decision Tree are faster than other algorithms. MT algorithm takes a long time even though a small dataset is used.

Table 4: Comparing Between the Various Decision Tree Algorithms for the Liver Dataset

Techniques	Tree Size	ACC (%)	MAE	PRE	REC	FME	Kappa Statistics	Time
J48	65	65.69	0.3678	0.651	0.657	0.654	0.158	0.11
LMT	1	69.47	0.4116	0.632	0.695	0.628	0.065	0.88
Random Forest		69.30	0.3464	0.667	0.693	0.674	0.186	0.5
Random Tree	267	66.55	0.3382	0.662	0.666	0.663	0.183	0.01
REPTree	27	66.13	0.3800	0.630	0.691	0.629	0.067	0.03
Decision Stump	Single Level	70.67	0.4392	0.499	0.707	0.585	0.379	0.01
HoeffdingTree	1	69.75	0.4091	0.634	0.700	0.619	0.0501	0.12

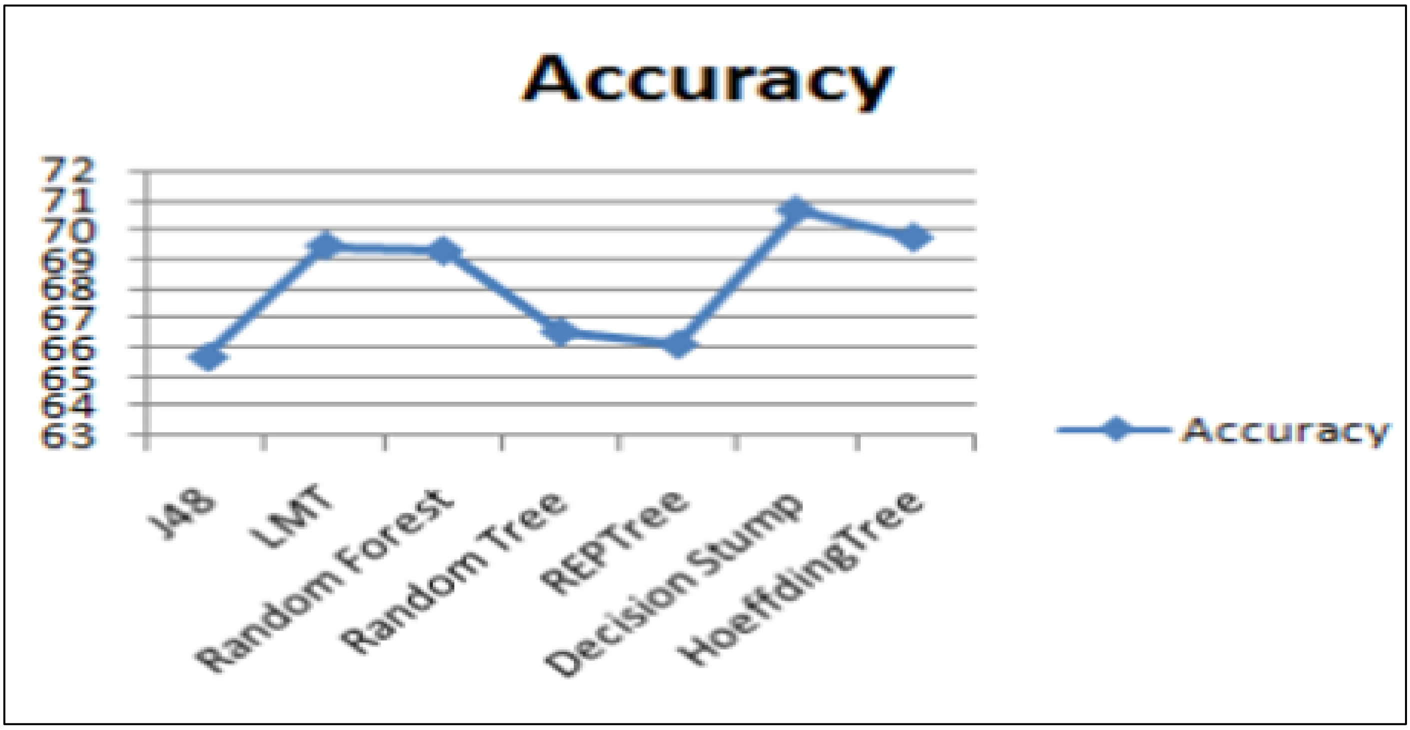

Fig 2: Comparing between the Various Decision Tree Algorithms for the Liver Dataset According to Accuracy

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1960

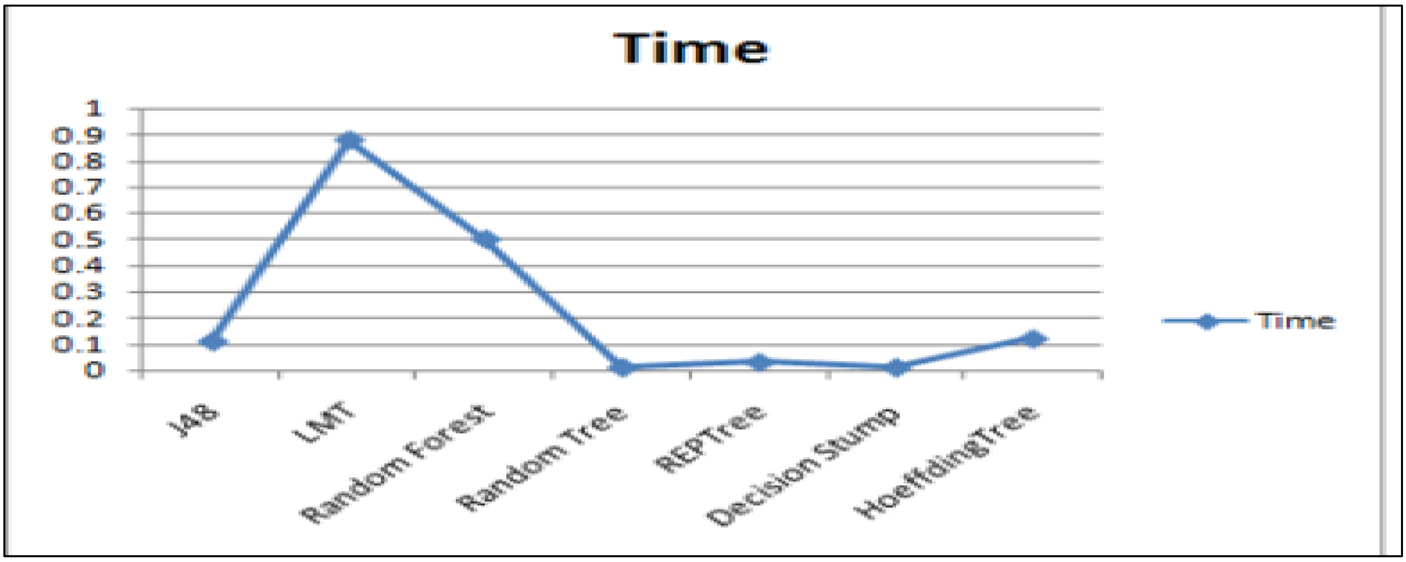

Fig 3: Comparison of the Decision Tree Algorithms According to Kappa Statistics

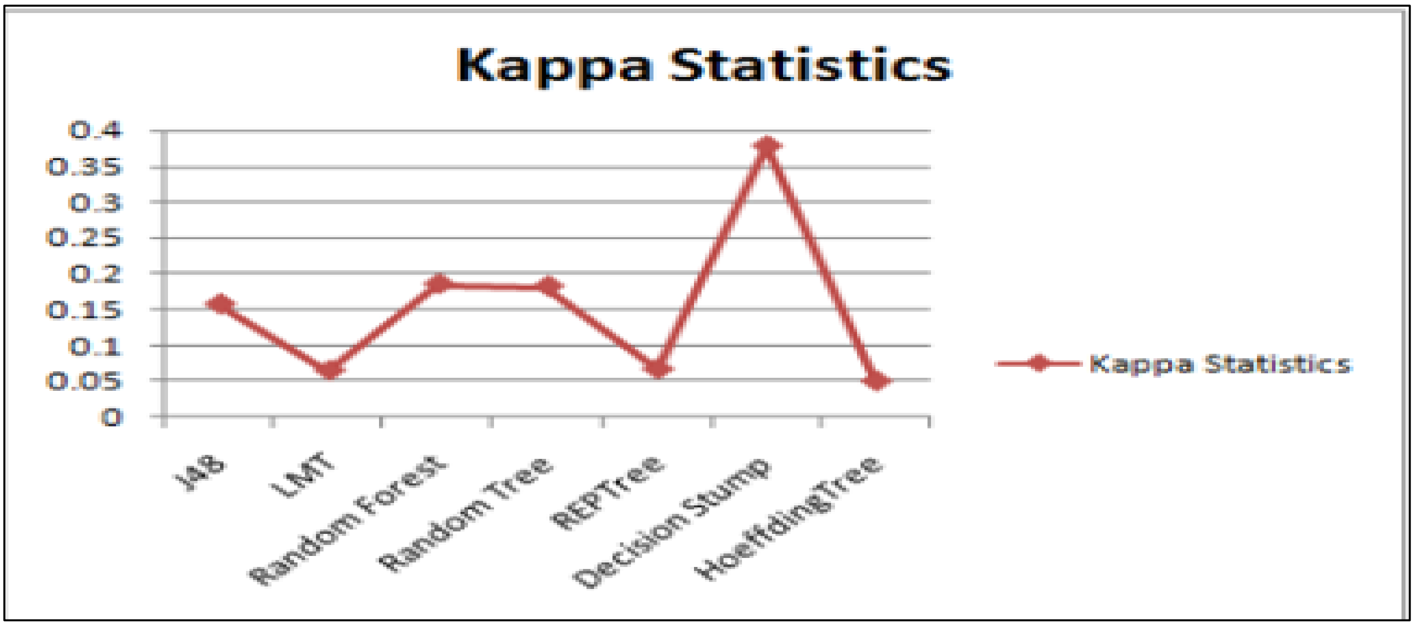

Fig 4: Comparison of the Decision Tree Algorithms According to Runtime

XI. CONCLUSION

In conclusion, the rising predominance of heart and liver conditions presents a significant challenge to open wellbeing, exacerbated by the nonstop advancement of innovation and the determined appropriation of inactive ways of life and liberal. Whereas endeavors to advance health-conscious exercises such as yoga and move classes have been made, the burden of these ailments is likely to continue in the future. Our activity the think about on tackling the capabilities of machine learning, especially the Support Vector Machine (SVM) demonstrate, to anticipate the probability of liver infections. In spite of the complexities of managing with exceptionally huge datasets, our discoveries from this test are promising. We can unquestionably conclude that the SVM demonstrate holds the potential to precisely foresee the hazard of liver illnesses.

By tackling the capabilities of machine learning calculations like SVM, society stands to advantage enormously. The capacity to figure liver infection hazard in development can lead to proactive mediations, early location, and focused on treatment techniques. Executing these measures can essentially upgrade persistent results and lighten the generally burden of liver-related wellbeing issues on healthcare frameworks. Given the expanding predominance of heart and liver conditions, it is basic to investigate and embrace imaginative innovations and data-driven approaches to viably handle these challenges. Our activity includes to the developing body of inquire about in this field and lays the establishment for future progressions in the early determination and administration of liver illnesses.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL1960

In conclusion, us ponder underscores the potential of machine learning calculations, such as SVM, as powerful devices in combating liver infections, eventually driving to moved forward wellbeing results for people and society at huge. Supported inquire about and the proceeded execution of such activities will play a urgent part in the anticipation and administration of these far reaching and concerning wellbeing conditions.

FUTURE ENHANCEMENT

Improved Performance the integration of PSO and MLP enhances the accuracy of liver disease classification, leading to more dependable predictions.

Simplified Illness Distinguishing proof with the help of PSO-based highlight choice, deciding the sort and event of liver illnesses gets to be less complicated and more efficient.

By leveraging progressed information mining strategies and utilizing information mining ideal models and program models such as Hive and R, the prepare of information collection, preprocessing, and appraisal is streamlined, driving to made strides information quality. Upgraded Assessment Utilizing different machine learning calculations permits for time complexity and precision appraisal, empowering the estimation of distinctive characteristics based on particular prerequisites.

REFERENCES

[1]. Sebastian, Anu, and Surekha Mariam Varghese "Fuzzy logic for Child-Pugh classification of patients with cirrhosis of the liver." 2016 International Conference on Information Science (ICIS) IEEE, 2016.

[2]. Arshad, Insha, et al., "Liver disease detection due to excessive alcoholism using data mining techniques." 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE) IEEE, 2018.

[3]. Ramkumar, N., et al., "Prediction of liver cancer using Conditional probability Bayes theorem." 2017 International Conference on Computer Communication and Informatics (ICCCI) IEEE, 2017.

[4]. Hassoon, Mafazalyaqeen, et al., "Rule optimisation of boosted c5.0 classification using a genetic algorithm for liver disease prediction." 2017 International Conference on Computers and Applications (ICCA). IEEE, 2017.

[5]. Karthik, S., Priyadarshini, A., Anuradha, J., and Tripathi, B. K., Classification and Rule Extraction Using Rough Set for Diagnosis of Liver Disease and its Types, Ad.

[6]. D. Sindhuja and R. J. Priyadarsini, “A survey on classification techniques in data mining for analyzing liver disease disorder”, International Journal of Computer Science and Mobile Computing, Vol.5, no.5 (2016), pp. 483-488.

[7]. B. V. Ramana, M. R. P. Babu and N.B. Venkaeswarlu, “A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis”, International Journal of Database Management Systems (IJDMS), Vol.3, no.2, (2011) , pp. 101-114.

[8]. A.S.Aneeshkumar and C.J. Venkateswaran, “Estimating the Surveillance of Liver Disorder using Classification Algorithms”, International Journal of Computer Applications (0975 –8887) , Vol. 57, no. 6, (2012), pp. 39-42.

[9]. S.Dhamodharan, “Liver Disease Prediction Using Bayesian Classification”, 4th National Conference on Advanced Computing, Applications & Technologies, Special Issue, May 2014. International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.8, No.2, March 2018 9

[10]. P.Rajeswari and G.S. Reena, “Analysis of Liver Disorder Using data mining Algorithms”, Global Journal of Computer Science and Technology, Vol.10, no. 14 (2010), PP. 48- 52.

[11]. G. Selvara and S. Janakiraman, “A Study of Textural Analysis Methods for the Diagnosis of Liver Disease from Abdominal Computed Tomography”, International Journal of Computer Applications (0975-8887), Vol. 74, no.11 (2013), PP.7-13.

[12]. H. Sug, “ Improving the Prediction Accuracy of Liver Disorder Disease with Oversampling”, Applied Mathematics in Electrical and Computer Engineering, American-MATH 12/CEA12 proceedings of the 6th Applications and proceedings on the 2012 American Conference on Appied Mathematics (2012), PP. 331-335.

[13]. R.H.Lin, “An Intelligent model for liver disease diagnosis”, Artificial Intelligence in Medical, Vol. 47, no. 1 (2009), PP. 53-62.

[14]. B. V. Ramanaland and M.S. P. Babu, “Liver Classification Using Modified Rotation Forest”, International Journal of Engineering Research and Development ISSN: 2278-067X, Vol. 1, no. 6 (2012), PP.17-24.

[15]. H.R. Kiruba and G. T. arasu, “An Intelligent Agent based Framework for Liver Disorder Diagnosis Using Artificial Intelligence Techniques”, Journal of Theoretical and Applied Information Technology, Vol. 69 , no.1 (2014), pp. 91-100.

[16]. C.K. Ghosk, F. Islam, E. Ahmed, D.K. Ghosh, A. Haque and Q.K. Islam, “Etiological and clinical patterns of Isolated Hepatomegaly” Journal of Hepato-Gastroenterology, vol.2, no. 1, PP. 1-4.

[17]. S. S. Aksenova , ‘Machine Learning with WEKA - WEKA Explorer Tutorial for WEKA Version 3.4’, 2004.

[18]. Machine Learning Repository, Center for Machine Learning and Intelligent Systems https://archive.ics.uci.edu/ml/index.php