LIVER DISEASE PREDICTION FOR AGE FACTORS

[1]VANGETI BABU RAO, [2]G.V.RAMANA

[1] PG SCHOLAR, VAHINI INSTITUTE OF SCIENCE & TECHNOLOGY

[2] ASSOCIATE PROFESSOR IN DEPT OF CSE IN SREE VAHINI INSTITUTE OF SCIENCE & TECHNOLOGY, TIRUVURU, KRISHNA DIST, ANDHRA PRADESH, INDIA.

ABSTRACT

Chronic liver disease is a serious threat to public health that claims many lives worldwide. Numerous liver-damaging factors, including alcohol misuse, obesity, and undiagnosed hepatitis infections, contribute to this illness. Numerous symptoms, including aberrant nerve activity, redness from a cough or vomiting, anaemia, kidney failure, jaundice, or liver encephalopathy, can be caused by this condition. It is difficult and costly to diagnose chronic liver disease. As a result, this study aims to evaluate the predictive power of several machine learning algorithms for chronic liver disease in order to reduce the high costs of detection.

In this work, six distinct algorithms were employed: Naïve Bayes, Decision Tree, Random Forest, K Nearest Neighbours, Support Vector Machine, and Logistic Regression. A number of parameters, such as the f-1 score, specificity, accuracy, precision, and recall, were used to evaluate the effectiveness of various categorization techniques. Our findings showed that the respective accuracy percentages for Logistic Regression, Random Forest, Decision Tree, Support Vector Machine, K Nearest Neighbours, and Naïve Bayes were 75%, 74%, 69%, 64%, 62%, and 53%. According to the investigation's findings, logistic regression was the most accurate technique.

Furthermore, we looked into several approaches to analyse this data in order to visualise the main objective of our work, which was to forecast liver disease using clinical data.

KeyWords: Algorithmn, Detection, Logistic Regression, Accuracy, Precesion, Naïve Bayes.

INTRODUCTION

The liver, one of our bodies' major organs, is essential for both food digestion and the removal of toxic compounds from the body. Viral infections and excessive alcohol consumption can affect or even endanger a person's life because of their vulnerability to the liver. There are many different types of liver diseases, such as liver cancer, cirrhosis, hepatitis, and liver tumours. The two most common causes of death among these are cirrhosis and liver disorders [1]. Thus, liver disease is one of the primary worldwide health issues. Every year, diseases associated to the liver claim the lives of around 2 million people worldwide [2].While a million people died from cirrhosis in 2010, many patients in the Global Burden of Disease (GBD) study published in BMC Medicine [3] fought liver cancer. Data mining has significantly changed the field of medical science, especially with regard to diagnosing and prognosticating liver disease [4-6]. Deep learning can be used to improve disease detection and forecast, encourage more objectivity in decision-making, and raise interest in the biological field [16]. Machine learning techniques can help diagnose diseases more quickly and with less expensive tests. Enhancing prediction accuracy and lowering diagnostic costs are the two main objectives of this research for the healthcare sector. LR, KNN, DT, SVM, NB, and RF were among the classification algorithms that were used as a result. Precision, recall, efficacy, and F1 score were among the dimensions on which each approach was evaluated. Furthermore, the data were compared using the Receiver Operating Characteristic (ROC). The next sections cover the details of the dataset, the data preparation techniques, and the methodology used in Chapter 2.A serious global health concern is liver sickness, which encompasses a wide range of conditions that profoundly affect the composition and operations of the liver. Innovative methods for identifying risk factors and developing predictive models are desperately needed because of the disease's broad prevalence and variety of causes. As a ground-breaking effort in the field of preventive medicine, the "Liver Disease Prediction for Age Parameters" project makes use of machine learning skills to determine an individual's risk of liver disease based on age-related factors as well as a number of other health indicators.

It is impossible to exaggerate the significance of liver disease in the context of world health. Serious diseases that contribute to morbidity and death globally, such as cirrhosis, hepatitis, and fatty liver disease, represent significant risks to public health. Due to the intricate nature of liver illnesses, focused strategies for early identification and intervention must be developed, along with a nuanced understanding of the components that contribute to the condition.

Predictive Models' Imperative:

Identifying risk variables and creating predictive models are essential components of a proactive approach to healthcare. Not only may early detection of liver disease provide timely medical interventions, but it also provides individuals with the necessary knowledge to make informed lifestyle choices that lower their risk. Predictive models are essential in this situation because they provide individuals and healthcare providers with actionable knowledge.

The Cutting-Edge Initiative:

The "Age-Related Factors in Liver Disease Prediction" study represents a revolutionary advance in medical innovation. This study uses algorithms to analyse big information and identify minute trends that traditional diagnostic approaches can overlook. Data mining is a subset of artificial intelligence.

The initiative is noteworthy for emphasising age-related factors and acknowledging the dynamic interaction between ageing and the risk of liver disease development.

Function of Age-Related Elements:

Aging is an inevitable biological process that is closely linked to changes in organ function and susceptibility to a variety of health issues. Age-related changes to the liver affect its resistance to disease, despite the liver's vital role in metabolic activities. This study recognises the complex relationship between ageing and the risk of developing liver illnesses and focuses primarily on factors related to ageing.

Project Components:

The project uses a comprehensive methodology that starts with the methodical gathering and meticulous preprocessing of a large variety of health data. Uncovered patterns are uncovered by exploratory data analysis, which directs the selection of relevant features. After being carefully chosen for suitability in binary classification tasks, machine learning models are trained on age-specific datasets. To guarantee that the generated models are accurate and dependable in forecasting liver illness, they go through a thorough review process.

Impact across disciplines:

The project follows a comprehensive methodology that starts with data collection and analysis. This project encourages collaboration between scholars, data scientists, and healthcare practitioners by going beyond traditional disciplinary boundaries. Adopting an interdisciplinary approach ensures that the study addresses the practical implications of its findings in clinical settings in addition to addressing the technical challenges of predictive modelling.

Ethical Issues and Privacy:

The project's methodology takes a comprehensive approach, starting with the acquisition and application of state-of-the-art technologies in healthcare. When implementing cutting-edge technologies for healthcare, ethical issues and data protection are of utmost importance. The initiative adheres closely to stringent ethical criteria, ensuring the appropriate handling of sensitive health information and protecting the contributors' right to remain silent inside the dataset.

Future Prospects and Public Health Impact:

Scalability, regulatory compliance, and continuous improvement become critical issues as the project moves further. Its potential impact on public health can be further amplified by improving user-friendliness, integrating with healthcare systems, and doing educational outreach. Sharing the information and understanding gained from the study offers a chance to encourage community involvement and raise awareness of liver health and preventative measures.

Data Collection:

I.We used data from the UCI AI Repository for this investigation. Moreover, the original collection was made possible by the state of Andhra in northern India [7]. This data compilation includes 583 individuals with liver problems; 24.36 percent of participants are female and 75.64 percent are male. We select 10 out of the 11 unique parameters in the dataset to serve as the target class for our further research. As an illustration, Patient's age and gender are listed in Table

II. TB:TotalBilirubin

III. DB:DirectBilirubin

IV. Alkphos:AlkalinePhosphotase

V. Sgpt:AlamineAminotransferase

VI. Sgot:AsparatateAminotransferase

VII. TP:TotalProteins

VIII. ALB:Albumin

Albumin and Globulin Ratios in Relation to X.

Two sets of data were separated using a selector field and labelled by experts.

EXISTING SYSTEM

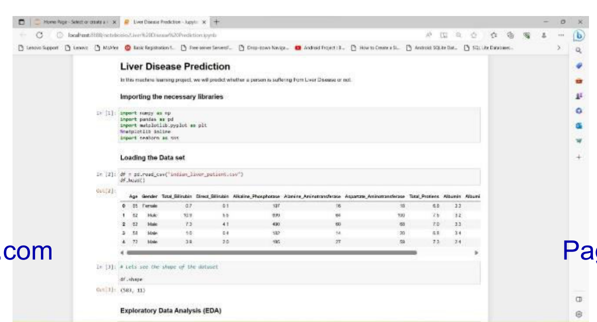
Viral infections and alcohol consumption harm the liver, placing individuals in potentially deadly circumstances. Liver disease, HIV, liver growths, liver cancer, and many more are among the various types of them. Liver diseases and cirrhosis are the two primary causes of death [1]. As such, liver disease is a major global health concern. Liver disease kills almost 2 million people worldwide every year [2]. According to the Worldwide According to the Burden of Disease (GBD) study, which was published in BMC Medicine, one million people died from liver cancer and one million people had cirrhosis in 2010.

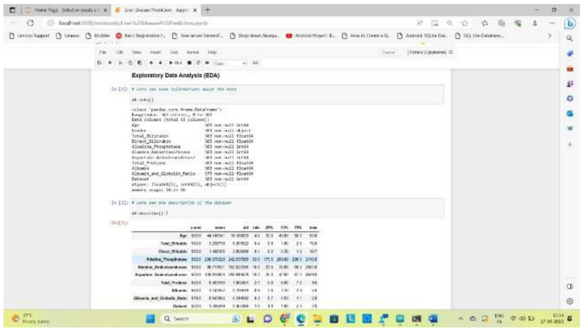
PROPOSED SYSTEM

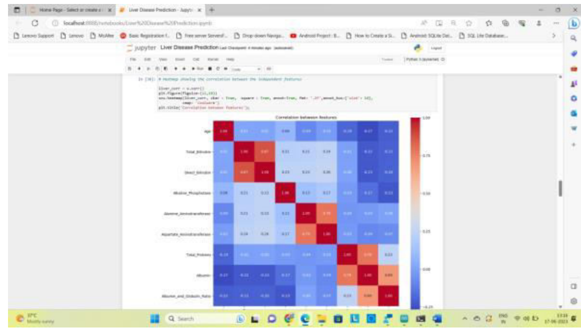
Artificial intelligence has had a major impact on the diagnosis and prognosis of liver diseases in the healthcare industry [4-6]. AI can increase the identification and prediction of sickness, which would increase the objectivity of decision-making processes [16]. Medical problems can be successfully treated with AI approaches, which lowers the cost of evaluations. The primary objectives of this project are to lower diagnosis expenses in the medical field and increase the efficiency of result prediction.

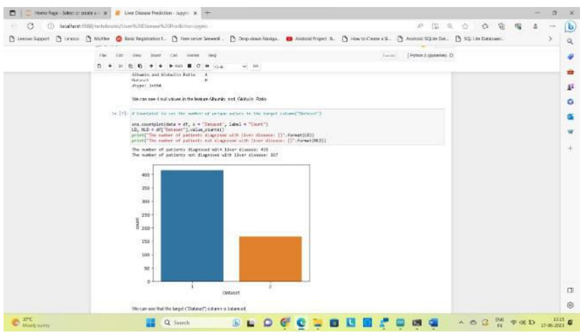
Research To determine whether or not individuals had liver disease, we employed a variety of categorization strategies. Random Forest (RF), k-Near (KNN), a decision tree (DT), the support vector machine (SVM), log-regression (LR), and naive Bayes were the six learning approaches used for the learning (NB). The efficiency of these methods was evaluated based on a number of factors, such as recall, accuracy, and precision. Moreover, theperformance was compared using the receiver operative characteristic.

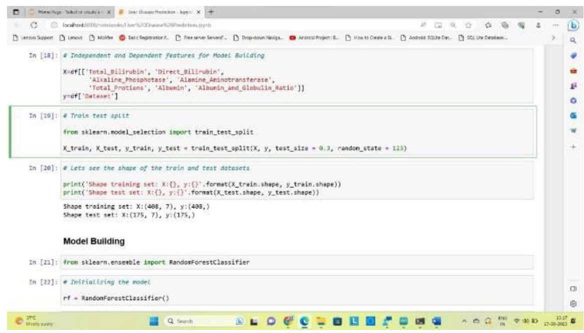
RESULT:





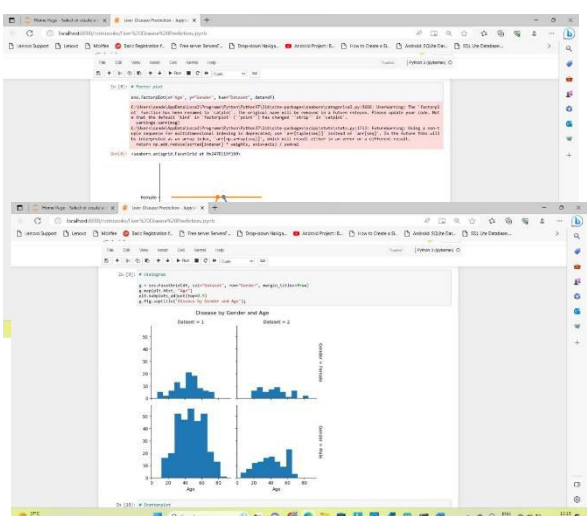


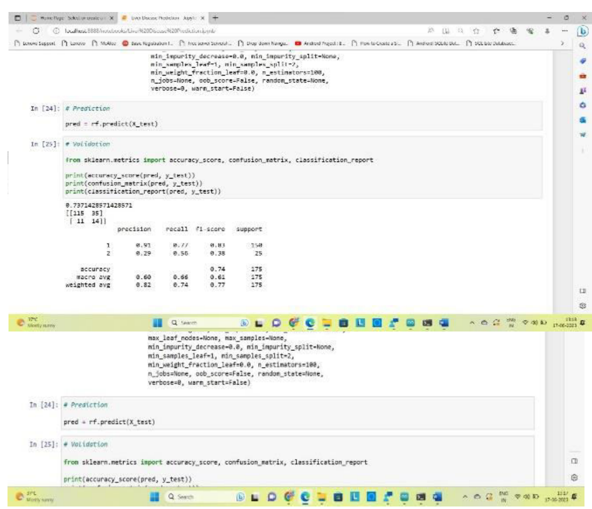




Feature Selection

Model traning





Accuracy of the model

CONCLUSION

The main goal of this study is to use six different supervised machine learning classifiers to create an effective diagnostic system for patients with chronic liver infections. Based on patient data parameters, we evaluated each classifier's performance. The results showed that, for predicting liver disease, the LR classifier produced the best results, with an accuracy of 75% based on the F1 measure, while NB had the lowest precision at 53%. Thus, for the chronic disease diagnosis and decision support system, the superior classification approach will be used. The software would be able to advise users on health issues and predict liver infections in advance. In low-income nations without adequate medical resources or trained medical personnel, this use can be especially helpful

Our investigation has opened up a number of possibilities for more research in this area. We just looked at a few popular supervised machine learning algorithms; other algorithms may be used to build a more precise model for liver disease prediction, improving performance all around. Additionally, this approach can be very helpful in medical facilities and research to prevent liver infections.

REFERENCES

[1] K. Sumeet, J.J. Larson, B. Yawn, T.M. Therneau, W.R. Kim, Underestimation of liver-relatedmortalityin theUnited States. Gastroenterology ;(2013)145:375–382, e371–372.

[2] A.A. Mokdad, A.D. Lopez, S. Shahraz, R. Lozano, A.H. Mokdad, J. Stanaway, et al, Livercirrhosis mortality in 187 countries between 1980 and 2010: a systematic analysis. BMC Med 2014;12:145.

[3] Bypass, Peter, The global burden of liver disease: a challenge for methods and for public health.BMCmedicine 12.1 (2014); 159.

[4] L.A.Auxilia, Accuracy Prediction Using Machine Learning Techniques for Indian Patient Liver Disease. 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI).IEEE(2018).

[5] Hashem, M.Esraa, S.Mai, A study of support vector machine

algorithm for liver disease diagnosis. American Journal of Intelligent Systems 4.1(2014); 9-14.

[6] P.Sajda, "Machine learning for detection and diagnosis of disease." Annu.Rev.Biomed.Eng.8 (2006); 537-565. UCI Machine Learning Repository. ILPD (Indian Liver Patient Dataset) DataSet .https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver +Patient Dataset)

[7] L.Breiman, Random Forests. Machine Learning, 45(1), (2001);5–32.https://doi.org/10.1023/A:1010933404324 Decision Trees, Retrieve from: https://dataaspirant.com/2017/01/30/how-decision-tree algorithm- works /,Last Accessed: 5 Octobor, 2019 Support vector machine, Retrieve from: http://www.statsoft.com/textbook/support- vector machines, LastAccessed:5 October,2019

[8] V.Vapnik, I.Guyon, T.H.-M, Learn, and undefined 1995. Support vector machines. statweb. stanford.edu(1995).

[9] ZhangM, ZhouZ, "ML-KNN: Alazy learning approach to multi-labellearning. "Pattern recognition 40.7 :(2007); 2038-2048.

[10] G.Guo, H.Wang, D.Bell, Y.Bi, K.Greer, KNNModel Based Approach in Classification(pp.986–996).Springer, Berlin, Heidelberg(2003).https://doi.org/10.1007/978-3-540-39964-3_62

Naive Bayes, ,Retrieve from: https://www.geeksforgeeks.org/naive-bayes-classifiers/,LastAccessed:5 Octobor,2019

[11] S. M. Mahmud, et al. "Machine Learning Based Unified Framework for Diabetes Prediction."Proceedings of the 2018 International Conference on Big Data Engineering and Technology. ACM(2018).

[12] S. Safavian, D. Landgrebe, A survey of decision tree classifier methodology. IEEE transactionsonsystems, man, andcybernetics, 21(3), (1991);660-674.

[13] A. Chervonenkis, Early history of support vector machines.

In Empirical Inference (pp. 13-20).Springer,Berlin, Heidelberg(2013).

[14] K.M.Leung, Naivebayesian classifier. Poly technic University Department of Computer Science/Finance and Risk Engineering(2007).

Guide Details:

G.V.Ramana, Associate Professor in Department of CSE in Sree Vahini Institute of Science & Technology, Qualification on M.TECH, MBA, MCSTA.

Email id: ramanagurrala999@gmail.com

Student Details:

VANGETI BABU RAO, M.Tech: Sree Vahini Institute of Science & Technology.

Email id: vangetibaburao@gmail.com