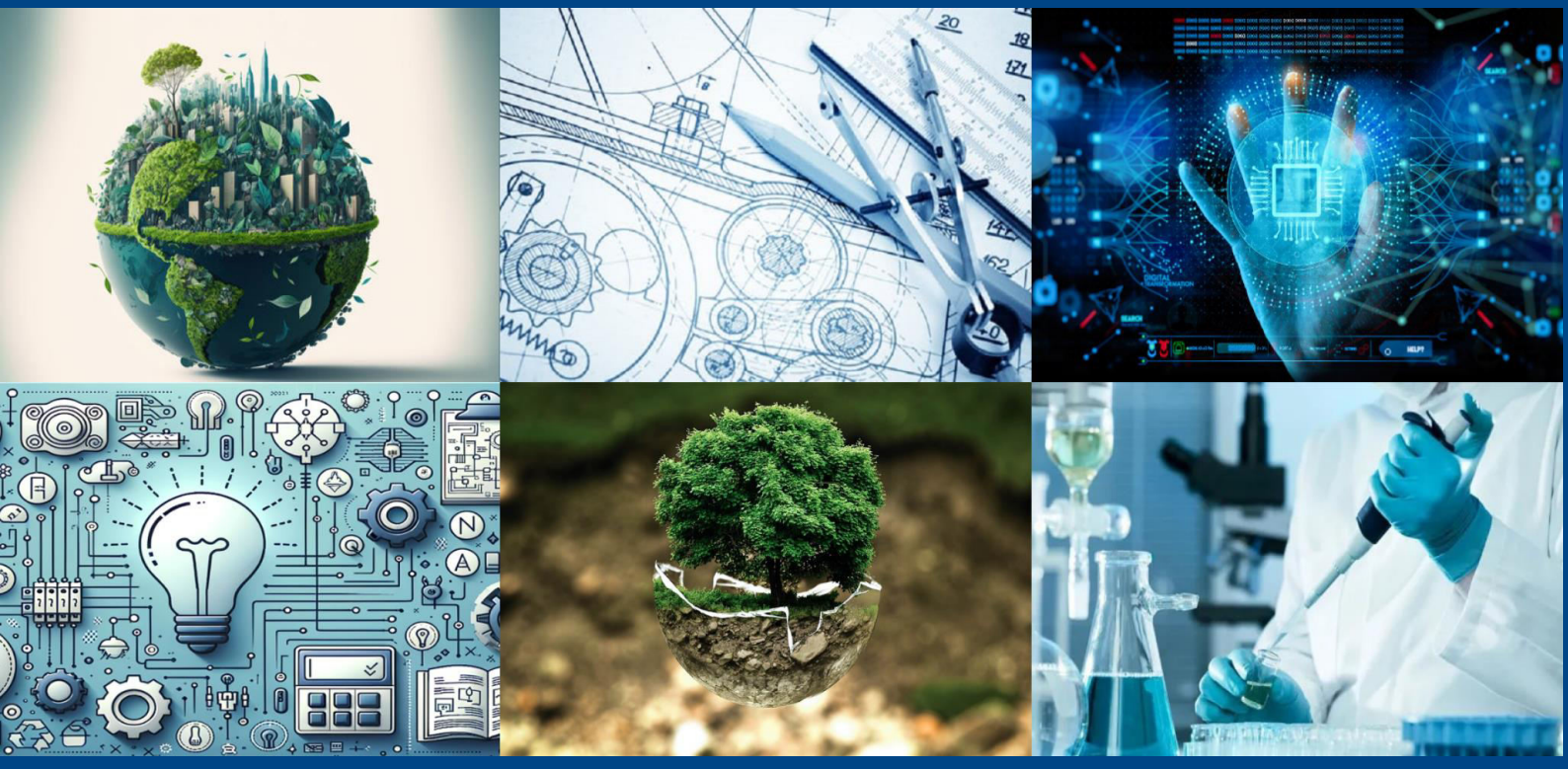
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 8, Issue 3, March 2025**

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803313

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Earlier Detection of Liver Diseases using KNN Algorithm

DR.D.J. Samatha Naidu, K. Venkata Subbaiah

Professor, Dept. of MCA, APGCCS, New Boyanapalli, Rajampet, AP, India

PG Student, Dept. of MCA, APGCCS, New Boyanapalli, Rajampet, AP, India

ABSTRACT: In a human body function of the liver is important. Many persons are suffering from liver disease, but they don't know it. The identification of liver diseases in the early stage helps a patient get better treatment. If it is not diagnosed earlier, it may lead to various health issues. To solve these issues, physicians need to examine whether the patient has been affected by liver disease or not, based on the multiple parameters. In this paper, we classify the patients who have liver disease or not by using different machine learning algorithms by comparing the performance factors and predicting the better result. The liver dataset is retrieved from the Kaggle dataset.

I. INTRODUCTION

The purpose of this study is to utilize machine learning techniques for the comparative analysis of liver diseases. By employing sophisticated algorithms, this research aims to enhance the accuracy and efficiency of diagnosing liver conditions, thereby facilitating timely intervention and treatment. The primary purpose is to develop a robust model capable of accurately distinguishing between various types of liver diseases, aiding medical professionals in making informed decisions and improving patient outcomes.Data Collection: Gathering comprehensive datasets containing clinical, genetic, and demographic information of patients diagnosed with different liver diseases.Feature Selection: Identifying relevant features from the collected data that are indicative of distinct liver conditions.Model Development: Implementing various machine learning algorithms such as decision trees, support vector machines, neural networks, and ensemble methods to develop predictive models.Performance Evaluation: Assessing the performance of developed models through metrics like accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC).Comparative Analysis: Comparing the effectiveness of different machine learning techniques in diagnosing and categorizing liver diseases.Clinical Application: Translating the developed models into practical clinical tools for assisting healthcare professionals in diagnosing liver conditions accurately.

The existing system for liver disease detection has employed various machine learning techniques such as decision trees, naive Bayes, random forests, and ensemble learning. Several studies have analyzed and compared different classifiers to determine their effectiveness in diagnosing liver diseases. Decision tree techniques like LMT, J48, Hoeffding Tree, Decision Stump, and Random Tree have been used, with Decision Stump yielding the highest accuracy. Medical data mining (MDM) has also been utilized to predict liver diseases in their early stages, helping to assess disease complexity. Classification models have been applied to distinguish between different liver conditions, including cirrhosis, hepatitis, liver cancer, and cases with no disorder, with Naive Bayes emerging as a strong performer. Additionally, feature selection methods such as Stability Selection combined with Random Forest have been employed to enhance model performance.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

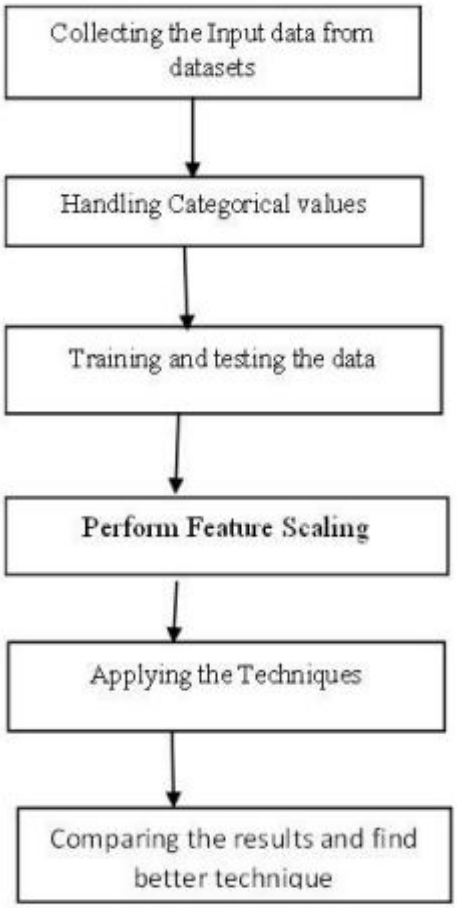


Figure 1: System Architecture

The proposed system enhances the early detection of liver diseases by utilizing advanced machine learning techniques, specifically the K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) algorithms. Unlike traditional methods, this system focuses on improving classification accuracy through optimized dataset preprocessing, automated model training, and efficient feature selection. By incorporating a user-friendly interface, it allows seamless dataset uploading, model generation, and disease prediction. Additionally, the system includes a graphical representation of accuracy comparisons, helping to evaluate the effectiveness of different algorithms. These improvements aim to provide a more reliable, accurate, and efficient approach to liver disease detection.It is a supervised Machine learning technique applied for both classification and regression kinds of problems, but it is used for classification types of problems. This model is applied to predict the categorical dependent variable with support of independent variables the output should be 0 or 1.

II. LITERATURE REVIEW

[1] Joel Jacob et al., "Diagnosis of Liver Disease Using Machine Learning Techniques", International Research Journal of Engineering and Technology (IRJET), Volume: 05, Issue: 04,2018, pp 4011-4014.

Health is Wealth. Though the medical field has grown rapidly with highly effective technologies, chronic diseases such Heart and Liver diseases are life-menacing. Various life factors such as alcohol, smoking, stress, food, lifestyle, etc. causes imbalance and add toxics to the human body leading to the occurrence of assorted diseases and disorders. The

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

medical records of the patients as a vast source of data are applied to the data mining techniques to extract the valid dataset to predict the liver disease. The classification algorithms have been widely used in the decision- making process. RNN being a text classifier of deep learning technique with the advantage of processing in multiple loops in a sequential manner to obtain best performances measured by the factor of accuracy has been proposed in this study.

2] Pragati Bhagat et al.," System for diagnosis of Liver Disease Using Machine Learning Technique", International Research Journal of Creative Research Thoughts", ISSN: 2320-2882, pp25-30.

Many people suffer from liver disease, but they don't have an idea about it. It is difficult to diagnosis of liver disease at high level. Before treatment of liver disease doctors first diagnose whether patient has liver disease or not, basis on different parameter. The system for diagnosis of liver disease using machine learning algorithms is an initiative towards better diagnosis of this disease as early as possible. Various algorithms are being studied in order to select the best algorithm which can give the best accurate results. According to the four parameters Accuracy, Precision, Sensitivity and Specificity the algorithm is being selected. After the study ANN algorithm turned out to be the best algorithm to implement and provides more accuracy than other algorithms. So, ANN is implemented using MATLAB platform and the user interface is also constructed with the help of MATLAB.

[3] Nazmun Nahar and Ferdous Ara, "Liver Disease Prediction Using Different Decision Tree Techniques", International Journal of Data Mining & Knowledge Management Process Vol.8 No.2(March 2018).

To diagnose and forecast liver disease, a variety of machine learning algorithms are widely employed in the medical industry. We analyzed several research publications in which we focused on various Data mining approaches for making use of data to support the study of high and multi-dimensional data in the health-care industry. In this regard, we have publications that are relevant to this topic in terms of methodology, algorithms, and outcomes. For selected publications, results and assessment techniques are examined, and a detailed summary of the findings is offered at the end. As a result, the purpose of this research is to use machine learning algorithms to improve the diagnosis and prediction of liver illness.

[4] A Saranya, G.Seenuvasan," A Comparative Study Of Diagnosing Liver Disorder Disease Using Classification Algorithm", International Journal Of Computer Science and Mobile Computing, Vol. 6 Issue 8mpage no 49-54(August 2017).

Liver disease is the major cause of death every year. Liver diseases is the fifth big killer in England after cancer, stroke and respiratory disease. The most common causes of liver disease worldwide are chronic hepatitis B and C, alcohol and nonalcoholic. Machine Learning has a strong potential in automated diagnosis of various diseases. With the recent upscale in various liver diseases, it is necessary to identify the liver disease at a preliminary stage. In this we propose a new classifier by extending the XG Boost classifier with genetic algorithm. This compares various classification models and visualization techniques used to predict liver disease with feature selection. Outlier detection is used to find out the extreme deviating values and they are eliminated using isolation forest. The performance is measured in terms of accuracy, precision, recall f-measure and time complexity.

[5] S. Dhamodharan, Liver Disease Prediction Using Bayesian Classification, National Conference on Advanced Computing, Application & Technologies, 2014.

In recent years in healthcare sectors, data mining became an ease of use for disease prediction. Data mining is the process of dredge up information from the massive datasets or warehouse or other repositories. It is a very challenging task to the researchers to predict the diseases from the voluminous medical databases. To overcome this issue the researchers, use data mining techniques such as classification, clustering, association rules and so on. The main objective of this research work is to predict liver diseases using classification algorithms. The algorithms used in this work are Naïve Bayes and support vector machine (SVM). These classifier algorithms are compared based on the performance factors i.e. classification accuracy and execution time. From the experimental results it is observed that the SVM is a better classifier for predict the liver diseases.

Relevance to current Research

The study on liver disease classification using machine learning techniques is highly relevant to current research trends in medical data science. With the increasing prevalence of liver diseases globally, early detection and accurate diagnosis are crucial for effective treatment. Existing studies have explored various machine learning approaches such

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

as decision trees, Naïve Bayes, logistic regression, and ensemble methods to improve diagnostic accuracy. However, many of these methods have limitations in terms of scalability, feature selection, and handling imbalanced datasets. This research builds upon previous findings by integrating advanced classification techniques, optimizing feature selection, and comparing multiple models to enhance predictive performance. By leveraging machine learning on real-world datasets, this study contributes to ongoing efforts in medical AI, supporting physicians in making more informed diagnostic decisions and ultimately improving patient outcomes.

Liver disease is a growing health concern, and early detection plays a crucial role in improving patient outcomes. Many researchers have explored machine learning techniques to help diagnose liver diseases more accurately. Previous studies have used methods like decision trees, Naïve Bayes, and logistic regression, each with its strengths and weaknesses. However, challenges like handling complex data, selecting the most relevant features, and improving accuracy still remain. This research builds on past work by comparing multiple machine learning models, refining feature selection, and optimizing prediction accuracy. By using real-world data, it aims to provide better support for doctors, making diagnosis faster and more reliable. In doing so, this study contributes to the ongoing advancements in medical technology, helping improve patient care.

III. METHODOLOGY OF PROPOSED SURVEY

Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.Provide extendibility and specialization mechanisms to extend the core concepts.Be independent of particular programming languages and development process.Provide a formal basis for understanding the modeling language.

Encourage the growth of OO tools market.Support higher level development concepts such as collaborations, frameworks, patterns and components.Integrate best practices.The use case diagram provides an overview of the interaction between the Admin and the System in the project. The Admin has several key functionalities, including registration, login, uploading datasets, pre-processing data, testing the data, viewing results, and analyzing model performance. These actions ensure the system is supplied with relevant data for processing and evaluation. On the other hand, the System handles various operations such as taking datasets, pre-processing the data, training the data, training the model, evaluating model performance, and predicting results. This diagram effectively represents the workflow of the system, illustrating the roles and responsibilities of both the Admin and the System in the machine learning-based liver disease analysis project.The class diagram represents the structure of the system by defining the relationship between the Admin and the System classes. The Admin class includes attributes such as Register and Login, which allow user authentication. Additionally, the Admin has various methods such as uploading datasets, pre-processing the data, testing the data, viewing model performance, and viewing results, indicating its role in managing data input and analysis. The System class, on the other hand, handles the core operations, including taking datasets, pre-processing the data, training the data, training the model, evaluating model performance, and predicting results. The connection between these two classes signifies that the Admin provides data and oversees the process, while the System executes the machine learning tasks to analyze liver diseases.A Sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams. The sequence diagram illustrates the step-by-step interactions between the Admin and the System in the liver disease detection process using the KNN algorithm. The process begins with the Admin registering and logging into the system, after which the system acknowledges the login. The Admin then uploads a dataset, and the System takes and preprocesses the data. Following this, the System trains the data and the machine learning model. The Admin can then test the data, and the System executes the model training. Once trained, the System evaluates the model's performance and predicts the result. Finally, the Admin views the model's performance and results. This sequence diagram effectively represents the flow of actions and interactions between the Admin and the System, ensuring smooth data processing and prediction outcomes.

IV. CONCLUSION AND FUTURE WORK

The article proposes an efficient real-time deep learning based framework to automate the process of monitoring the social distancing via object detection and tracking approaches, where each individual is identified in the real-time with

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

the 9 help of bounding boxes. The generated bounding boxes aid in identifying the clusters or groups of people satisfying the closeness property computed with the help of pairwise vectorized approach. The number of violations are confirmed by computing the number of groups formed and violation index term computed as the ratio of the number of people to the number of groups. The extensive trials were conducted with popular state-of-the-art object detection models: Faster RCNN, SSD, and YOLO v3, where YOLO v3 illustrated the efficient performance with balanced FPS and mAP score. Since this approach is highly sensitive to the spatial location of the camera, the same approach can be fine tuned to better adjust with the corresponding field of view..

REFERENCES

1. K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.
2. R. Christopher, "Port scanning techniques and the defense against them," SANS Institute, 2001.
3. M. Baykara, R. Das¸, and I. Karado ˘gan, "Bilgi g ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi," in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.
4. RashmiT V. "Predicting the System Failures Using Machine Learning Algorithms".International Journal of Advanced Scientific Innovation, vol. 1, no. 1, Dec. 2020, doi:10.5281/zenodo.4641686.
5. S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, "Surveillance detection in high bandwidth environments," in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.
6. K. Ibrahimi and M. Ouaddane, "Management of intrusion detection systems based-kdd99: Analysis with lda and pca," in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.
7. Girish L, Rao SKN (2020) "Quantifying sensitivity and performance degradation of virtual machines using machine learning.",Journal of Computational and Theoretical Nanoscience, Volume 17, Numbers 9-10, September/October 2020, pp.4055-4060(6) https://doi.org/10.1166/jctn.2020.9019
8. L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, "Detection and classification of malicious patterns in network traffic using benford's law," in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.
9. S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com