International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue.1, Jan 2023

COMPARATIVE ANALYSIS OF LIVER DISEASES BY USING MACHINE LEARNING TECHNIQUES

Mr T. Kanakaiah[1],A. Swetha Reddy[2],M. Sneha Reddy[3],A. Meghana[4],B.Sandhya[5],Daniel Firdous[6]

[1]Assistant professor, Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

[2,3,4,5,6]UG Students,Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

ABSTRACT

In the human body, the liver plays a crucial role in various physiological functions. Unfortunately, many individuals suffer from liver diseases without being aware of it. Early detection of liver diseases is paramount as it enables prompt and effective treatment, preventing the development of serious health complications. Failure to diagnose liver diseases at an early stage can lead to a multitude of health issues. To address these challenges, healthcare professionals must assess whether a patient is afflicted with liver disease by analyzing multiple parameters. This paper focuses on the classification of patients based on the presence or absence of liver disease using various machine learning algorithms. By comparing performance metrics and predicting optimal outcomes, we aim to assist physicians in accurately diagnosing liver diseases. The liver dataset utilized in this study is sourced from the Kaggle dataset, providing a comprehensive foundation for our analysis and evaluation.

I.INTRODUCTION

Liver diseases represent a significant public health challenge globally, impacting millions of individuals and straining healthcare systems worldwide. Timely and accurate diagnosis of liver diseases is critical for effective treatment and management, as these conditions can lead to severe complications if left untreated. With the advent of machine learning (ML) techniques, there is growing interest in leveraging data-driven approaches to enhance the diagnosis and prognosis of liver diseases.

The project "Comparative Analysis of Liver Diseases by Using Machine Learning Techniques" aims to explore and evaluate various ML algorithms for the analysis of liver diseases. By harnessing the power of ML, this project seeks to improve our understanding of

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue.1, Jan 2023

liver diseases, facilitate early detection, and optimize treatment strategies.

Through this project, we endeavor to conduct a comparative analysis of different ML techniques, including supervised learning, unsupervised learning, and deep learning, to classify and predict liver diseases accurately. By evaluating the performance of these algorithms, we aim to identify the most effective models for diagnosing different types of liver diseases and predicting disease progression.

Furthermore, this project aims to contribute to the existing body of knowledge by providing insights into the comparative performance of ML techniques in liver disease analysis. By elucidating the strengths and limitations of different ML algorithms, we aspire to develop more accurate and efficient diagnostic tools for liver diseases, ultimately improving patient outcomes and healthcare delivery.

Overall, the "Comparative Analysis of Liver Diseases by Using Machine Learning Techniques" project represents a proactive effort to leverage ML technologies to address the challenges associated with liver disease diagnosis and management, ultimately leading to improved patient care and public health outcomes.

II.LITERATURE REVIEW

1. Machine learning techniques have emerged as valuable tools for automating the diagnosis of various diseases, offering the potential to significantly impact medical practices. In the context of liver diseases, which have seen a notable increase in prevalence, early detection is imperative for effective management and improved patient outcomes. In their paper, "Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients," Maria Alex Kuzhippallil, Carolyn Joseph, and A. Kannan delve into the realm of machine learning applications in liver disease diagnosis. Their study highlights the importance of early disease identification and proposes a novel approach to classification using an extended XGBoost classifier with genetic algorithm optimization. Through a comprehensive review of existing literature, the authors explore the efficacy of various classification models and visualization techniques in predicting liver diseases with feature selection. Additionally, they address the

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue.1, Jan 2023

challenge of outlier detection and elimination, crucial for ensuring the reliability of machine learning models. Performance evaluation metrics, including accuracy, precision, recall, f-measure, and time complexity, are employed to assess the proposed classifier's effectiveness. Overall, their study contributes valuable insights into the application of machine learning techniques for liver disease diagnosis, paving the way for improved diagnostic accuracy and efficiency in clinical settings.

Prediction of Liver Disorders using Machine Learning Algorithms: A Comparative Study,Md. Fazle Rabbi; S. M. Mahedy Hasan; Arifa Islam Champa; Md. AsifZaman; Md. Kamrul Hasan,Liver, a crucial interior organ of the human body whose principal tasks are to eliminate generated waste produced by our organism, digest food, and preserve vitamins and energy materials. The liver disorder can cause various fatal diseases, including liver cancer. Early diagnosis, and treating the patients are compulsory to reduce the risk of those lethal diseases. As the diagnosis of liver disease is expensive and sophisticated, numerous researches have been performed using Machine Learning (ML) methods for classifying liver disorder cases. In this paper, we have compared four different ML algorithms such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Extra Trees (ET) for classifying Indian Liver Patient Dataset (ILPD). Pearson Correlation Coefficient based feature selection (PCC-FS) is applied to eliminate irrelevant features from the dataset. Also, a boosting algorithm (AdaBoost) is utilized to enhance the predictive performance of those algorithms. The comparative analysis is evaluated in terms of accuracy, ROC, F-1 score, precision, and recall. After comparing experimental results, we have found that boosting on ET provides the highest accuracy of 92.19%.

III.EXISTING SYSTEM

Various studies have explored the use of machine learning techniques in diagnosing and predicting liver diseases, each employing different approaches and algorithms to improve accuracy and effectiveness. For instance, researchers have investigated the implementation of decision tree techniques, such as LMT, J48, Hoeffding Tree, Decision Stump, and Random tree, to predict the onset of

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue.1, Jan 2023

liver diseases. Notably, some studies have found that certain algorithms, like Decision Stump, demonstrate higher accuracy rates in disease prediction.

Other researchers have focused on utilizing data mining techniques, particularly Medical Data Mining (MDM), to diagnose liver diseases, emphasizing early-stage prediction and complexity assessment. Additionally, studies have examined the prediction of specific liver diseases, including cirrhosis, hepatitis, and liver cancer, using classification techniques like Naive Bayes and Decision Trees.

Moreover, researchers have explored the use of ensemble learning techniques and sampling methods to enhance accuracy and stability in liver disease prediction models. Techniques such as Stability Selection and Random Forest have been combined to improve performance, while confusion matrices have been used to assess accuracy and effectiveness.

Furthermore, researchers have proposed intelligent medical decision support systems, employing various classification techniques to diagnose liver disorders effectively. These systems analyze data from different scenarios to compare the effectiveness and correction rates of different algorithms.

Disadvantages

- An existing methodology doesn't implement an effective Machine Learning Classifiers.
- The system not implemented Decision Trees which leads complexity in testing and training datasets.

IV.PROPOSED SYSTEM

To analyze liver disease, data from the Indian Liver Patient dataset on Kaggle is utilized. Patients are categorized based on disease status, denoted as either 1 or 2. The dataset contains various attributes, detailed in Table I. During data preprocessing, the gender attribute is converted to an integer value (0 for male, 1 for female). The proposed system's workflow includes data collection, handling categorical values, splitting the dataset for training and testing, performing feature selection, and applying machine learning techniques. Finally, the predicted results are compared to determine the accuracy of the model.

Advantages

- It is a supervised Machine learning technique applied for

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue.1, Jan 2023

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.

- This Classifier technique is effective when only a small amount of training data is required to derive approximation parameters. With highly scalable model creation, it can tackle a wide range of challenging real-world problems.
- This is a pattern recognition system that involves the training datasets for finding the k closest relatives in new conditions. When using k-NN for classification, we must calculate the location data within the nearest neighbor's category.

V.MODULES

- **Data Collection Module:**

The data collection module is responsible for retrieving relevant datasets containing information on Indian liver disease patients. This process involves sourcing datasets from reliable sources such as Kaggle or healthcare databases. The collected data should encompass various attributes such as patient demographics, medical history, laboratory test results, and disease diagnoses.

- **Data Preprocessing Module:**

The data preprocessing module plays a crucial role in preparing the collected datasets for analysis. This involves cleaning the data to remove any inconsistencies, missing values, or outliers that may adversely affect the performance of machine learning algorithms. Additionally, data preprocessing encompasses tasks such as feature scaling, normalization, and encoding categorical variables to ensure uniformity and compatibility with the chosen machine learning models.

- **Model Building Module:**

The model building module involves the development and training of machine learning models for predicting liver disease in Indian patients. This process begins by selecting appropriate algorithms, such as decision trees, support vector machines, logistic regression, or neural networks, based on

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue.1, Jan 2023

the nature of the dataset and the problem at hand. The selected algorithms are then trained on the preprocessed data to learn patterns and relationships between input features and target variables.

- **Applying the Algorithms Module:**

Once trained, the machine learning models are applied to unseen data to make predictions about the presence or absence of liver disease in Indian patients. This involves feeding new patient data into the trained models and obtaining predictions based on learned patterns. The performance of the algorithms is evaluated using various metrics such as accuracy, precision, recall, and F-measure to assess their effectiveness in accurately diagnosing liver disease.

VI.CONCLUSION

In conclusion, the project "Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients" has provided valuable insights into the application of machine learning algorithms in the diagnosis of liver diseases among Indian patients. Through a comprehensive analysis of various classification models and visualization techniques, as well as the proposal of an extended XGBoost classifier with genetic algorithm optimization, this study has contributed to enhancing the accuracy and efficiency of liver disease diagnosis. The evaluation of performance metrics such as accuracy, precision, recall, f-measure, and time complexity has demonstrated the effectiveness of the proposed approach in predicting liver disease with improved accuracy and reduced computational overhead. Overall, the findings of this project underscore the potential of machine learning techniques in revolutionizing the diagnosis and management of liver diseases in clinical practice.

VII.FUTURE SCOPE

While this project has made significant strides in leveraging machine learning for liver disease diagnosis among Indian patients, several avenues for future research and development exist. One potential area for exploration is the integration of additional data sources, such as genetic information or imaging data, to further enhance the predictive capabilities of machine learning models. Moreover, continued advancements in machine learning techniques, such as deep learning and reinforcement

ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue.1, Jan 2023

learning, offer promising opportunities for improving the accuracy and robustness of liver disease diagnosis models. Additionally, expanding the scope of the project to include the prediction of disease progression and treatment response could provide valuable insights for personalized medicine approaches. Furthermore, the deployment of these machine learning models in real-world clinical settings and the evaluation of their impact on patient outcomes and healthcare delivery warrant further investigation. Overall, the future scope of this project encompasses a wide range of possibilities for advancing the field of liver disease diagnosis and improving patient care.

VIII.REFERENCES

1.R. Saravanan and Pothula Sujatha, "A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification", Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018).

2.Pushpendra Kumar and Ramjeevan Singh Thakur, "Diagnosis of Liver Disorder Using Fuzzy Adaptive and Neighbor Weighted K-NN Method for LFT Imbalanced Data", IEEE 6th International Conference on smart structures and systems ICSSS 2019.

3. Sina Bahramirad, Aida Mustapha and Maryam Eshranghi, "Classification of Liver Disease Diagnosis: A Comparative Study", published in 2013 Second International Conference on Informatics & Applications (ICIA).

4.D. Dua and C. Graff, UCI Machine Learning Repository, Irvine, CA:University of California, School of Information and Computer Science, 2019, [online] Available: http://archive.ics.uci.edu/ml.

5.Insha Arshad, Chiranjit Dutta, Tanupriya Choudhary and Abha Thakral, "Liver Disease detection due to excessive alcoholism using Data Mining Techniques", International Conference on Advances in Computing and Communication Engineering (ICACCE-2018).

6.Panigrahi Srikanth and Dharmaiah Deverapalli, "A Critical Study of Classification Algorithms Using Diabetes Diagnosis", IEEE 6th International Conference on Advanced Computing, 2016.

7.K. Thirunavukkarasu, Ajay S. Singh, Md Irfan and Abhishek Chowdhury, "Prediction of Liver Disease using

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 16, Issue.1, Jan 2023

Classification Algorithms", 4th International Conference on Computing Communication and Automation (ICCCA), 2018.

8.B.V. Baiju, Disease Influence Measure Based Diabetic Prediction with Medical Data Set Using Data Mining.

9. Noria Bidi and Zakaria Elberrichi, "Feature Selection for Text Classification Using Genetic Algorithms", 8th International Conference on Modelling Identification and Control (ICMIC-2016).

10.Akansha Ahlawat and Bharti Suri, Improving Classification in Data mining using Hybrid algorithm, IEEE, 2016.

11.Haobin Shi and Meng Xu, "A Data Classification Method Using Genetic Algorithm and K-means Algorithm with Optimizing Initial Cluster Center", IEEE International Conference on Computr and Communication Engineering Technology (CCET), 2018.

12. N. Ramkumar, S. Prakash, S. Ashok Kumar and K Sangeetha, "Prediction of liver cancer using Conditional probability Bayes theorem", 2017 International Conference on Computer Communication and Informatics (ICCCI).

13. Ke Yan, Xiaoming You, Xiaobo Ji, Guangqiag Yin and Fan Yang, "A Hybrid Outlier Detection Method for Health Care Big Data", 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud) Social Computing and Networking (SocialCom) Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom).