ANALYSIS AND PREDICTION OF HEPATITIS USING MACHINE LEARNING

Abstract

This study presents a systematic method using machine learning to diagnose hepatitis disease. By leveraging distinct classifiers and applying class balancing techniques, the proposed approach aims to improve accuracy and reduce human error in diagnosis. The hepatitis virus is considered one of the deadliest diseases in the world. Machine learning could help in diagnosing this condition. In order to develop a comprehensive strategy for diagnosing hepatitis, the authors evaluated the performance of the classifiers on the UCI dataset. The Classification algorithms are logistic regression model, the support vector machine, and the random forest. These methods are tested without class balancing and also with class balancing through SMOTE. The efficiency of these tools improved significantly after adopting class balancing.

Another method is to use classification algorithms. They can be compared using methods that are based on feature selection. Principal components of the classification process, namely the information gain attribute, the cfs subset eval, and the F-measure and ROC Area performance, for filter-based features selection. It was found that the Nave Bayes model had the highest accuracy when it came to analyzing the data set of hepatitis patients. The best feature selection method for determining hepatitis patients is the filter-based model.

Keywords – Machine learning, Hepatitis, Feature Selection, SVM, Diagnosis, LR.

Authors

B. Siva Kumar Reddy
Professor
Department of ECE
CMR Engineering College
Hyderabad, Telangana, India.
Bsivakumar100@gmail.com

K. Sai Divya Sri
Student
Department of ECE
CMR Engineering College
Hyderabad, Telangana, India.

K. Krishna
Student
Department of ECE
CMR Engineering College
Hyderabad, Telangana, India.

K. Sravan
Student
Department of ECE
CMR Engineering College
Hyderabad, Telangana, India.

K. Atheesh
Student
Department of ECE
CMR Engineering College
Hyderabad, Telangana, India.
208r1a0423@gmail.com

I. INTRODUCTION

The common types that are known to cause the disease are hepatitis A, B, C, D and E [1]. A disease diagnosis is a vital aspect of medicine as it involves the classification of a condition based on its symptoms [2]. For instance, hepatitis is a life-threatening condition that affects the liver. The two main forms of hepatitis are acute and chronic. The former is characterized by minimal damage to hepatocytes, while the latter usually involves more extensive damage that can lead to long-term health issues [3 – 5]. Acute hepatitis may result in mild impairment and inflammation, while chronic hepatitis may result in severe liver dysfunction and even cancer.

It is important to note that the five different types of hepatitis are responsible for numerous fatalities and outbreaks worldwide [6]. In particular, the C and B strains are known to cause cancer and liver cirrhosis. Around 130 to 150 million individuals worldwide are infected with hepatitis C. This viral infection usually leads to inflammation of the liver, which is why it is responsible for about 1.5 million deaths each year [7 – 9]. Hepatitis causes: alcohol consumption, drug events, bacterial & viral infections. Early detection aids recovery. Machine learning aids efficient diagnosis, supporting inexperienced clinicians. Inexperienced clinicians make human errors when it comes to diagnosing hepatitis. [10] These can result from a combination of factors such as inexperience, exhaustion, and lack of thought. Disease classification using supervised machine learning. Researchers train the computer with a dataset and evaluate model accuracy with test data [11]. The proposed method for hepatitis diagnosis employs machine learning techniques and is tested using a Python simulation environment. The hepatitis dataset from the UCI repository is utilized for testing [12].

By employing machine learning techniques, the project aims to enhance diagnostic accuracy, provide decision support to inexperienced clinicians, and identify hidden patterns in the data that may not be evident to human experts [13]. This reduction in human error can lead to more efficient and reliable diagnoses, ultimately improving patient outcomes and optimizing healthcare resources [14].
Here are some key motivations behind the project of analyzing and predicting hepatitis using machine learning include:

1. **Improved Diagnostic Accuracy:** Hepatitis can have various causes and manifestations, making it challenging for healthcare professionals to diagnose accurately, especially for inexperienced clinicians. By employing machine learning, the project aims to enhance the diagnostic accuracy of hepatitis, leading to better patient outcomes and timely interventions.

2. **Efficiency and Speed:** Machine learning models can process large amounts of medical data quickly, enabling rapid and efficient diagnosis of hepatitis. This can be crucial in situations where timely diagnosis can significantly impact the course of the disease and the effectiveness of treatment.

3. **Assisting Inexperienced Clinicians:** Machine learning models can act as decision support tools for less experienced clinicians, providing them with additional insights and

guidance during the diagnosis process. This support can boost their confidence and competence in handling hepatitis cases.

4. **Reduction of Human Error:** Human error is an inherent risk in medical diagnosis, and it can have serious consequences. By leveraging machine learning, the project aims to minimize the impact of human errors in hepatitis diagnosis, resulting in safer and more reliable healthcare practices.

5. **Optimization of Healthcare Resources:** Accurate and early diagnosis of hepatitis can help healthcare providers allocate resources efficiently. It can ensure that patients receive the appropriate care and treatment, leading to optimized resource utilization.

6. **Potential for Personalized Medicine:** Machine learning models can pave the way for personalized medicine approaches. By understanding the unique characteristics of each patient and their response to treatments, tailored and more effective therapeutic strategies can be developed.

7. **Scientific Advancement:** The project contributes to the scientific community by exploring the potential of machine learning in medical diagnosis. It adds to the growing body of research on applying artificial intelligence techniques to healthcare challenges.

8. **Accessibility and Affordability:** Once developed, machine learning models can be deployed in various healthcare settings, including resource-limited regions, making hepatitis diagnosis more accessible and affordable to a broader population.
Overall, the motivations behind this project are rooted in advancing healthcare practices, enhancing patient care, and leveraging the power of machine learning to address the challenges associated with hepatitis diagnosis.

II. RESEARCH METHODOLOGY

Developing a Machine Learning-based Analysis and Prediction System for Hepatitis would involve the following research methodology, encompassing several steps and approaches:

- **Here's an outline of a possible research methodology for the development of such a system:**

- **Model Optimization:** Refine the model by adjusting hyperparameters or employing ensemble methods to improve prediction accuracy further.
- **Interpretability:** Depending on the chosen algorithms, explore methods for interpreting the models to gain insights into which features are most influential in hepatitis prediction.
- **Ethical Considerations:** Address any ethical concerns related to data privacy and patient confidentiality. Ensure compliance with applicable regulations and guidelines.
- **Testing and Validation:** Apply the optimized models to the test dataset to assess their generalization ability and validate their predictive power.

- **Comparison with Baselines:** Compare the performance of the developed machine learning models with baseline models or existing hepatitis diagnostic methods.
- **Clinical Validation:** Collaborate with medical experts to validate the predictions made by the machine learning models against actual patient outcomes and clinical observations.
- **Discussion of Results:** Analyze and discuss the findings of the research, including the strengths, limitations, and potential applications of the developed machine learning models for hepatitis analysis and prediction.
- **Conclusion and Future Work:** Summarize the research outcomes and propose potential avenues for future research and improvements in hepatitis analysis using machine learning.

This research methodology provides a general framework for developing a Machine Learning-based Analysis and Prediction System for Hepatitis. The specific steps and approaches may vary depending on the available data, computational resources, and advancements in machine learning techniques.

III. CONCLUSION

The study demonstrates the effectiveness of machine learning in hepatitis diagnosis. By applying class balancing techniques and utilizing SVM (90% accurate), LR (93.18% accurate), KNN (85.93% accurate), and RF (92.52% accurate) classifiers, accurate and reliable predictions can be achieved. The results highlight the importance of feature selection and data preprocessing to enhance diagnostic accuracy.

REFERENCES

[1] Bayrak EA, Kirci P, Ensari T (2019) Performance analysis of machine learning algorithms and feature selection methods on hepatitis disease. Int J Multidiscip Stud Innov Technol 3(2):135–138
[2] Sachdeva, R.K., Bathla, P., Rani, P. et al. A systematic method for diagnosis of hepatitis disease using machine learning. Innovations Syst Softw Eng 19, 71–80 (2023). https://doi.org/10.1007/s11334-022-00509-8
[3] Nayeem MJ, Rana S, Alam F, Rahman MA (2021) Prediction of hepatitis disease using K-nearest neighbors, Naive Bayes, support vector machine, multi-layer perceptron and random forest. In: International conference on information and communication technology for sustainable development (ICICT4SD), pp 280–284.
[4] Nivaan GV, Emanuel AWR (2020) Analytic predictive of hepatitis using the regression logic algorithm. In: 3rd International seminar on research of information technology and intelligent systems (ISRITI), pp 106–110.
[5] BAYRAK, Ebru AYDINDAĞ, Pınar KIRCI, and Tolga Ensari. "Performance analysis of machine learning algorithms and feature selection methods on hepatitis disease." International Journal of Multidisciplinary Studies and Innovative Technologies 3.2 (2019): 135-138.
[6] Mijwil MM, Salem IE, Abttan R (2021) Utilisation of machine learning techniques in testing and training of different medical datasets. Asian J Comput Inf Syst 9:29–34
[7] P. Nancy, V. Sudha, and R. Akiladevi, "Analysis of feature Selection and Classification algorithms on Hepatitis Data", International Journal of Advanced Research in Computer Engineering & Technology, Volume 6, Issue 1, 2017.
[8] Sachdeva RK, Bathla P, Rani P, Kukreja V, Ahuja R (2022) A systematic method for breast cancer classification using RFE feature selection. In: 2nd International conference on advance computing and innovative technologies in engineering (ICACITE), pp 1673
[9] Peng J et al (2021) An explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients. J Med Syst 45(5):1–9.

[10] P. Tapkan, L. Özbakır, and A. Baykasoğlu, "Weka ile Veri Madenciliği Süreci ve Örnek Uygulama", Endüstri Mühendisliği Yazılımları ve Uygulamaları Kongresi, 2011.

[11] Rani P, Kumar R, Jain A, Chawla SK (2021) A hybrid approach for feature selection based on genetic algorithm and recursive feature elimination. Int J Inf Syst Model Des 12(2):17–38.

[12] Ksi̦ȧzek W, Gandor M, Pławiak P (2021) Comparison of various approaches to combine logistic regression with genetic algorithms in survival prediction of hepatocellular carcinoma. Comput Biol Med 134:1–13.

[13] Reddy, B. Siva Kumar, et al. "Design and Implementation of Novel FPGA Based LFSR for IOT and Smart Applications." *2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*. IEEE, 2022.

[14] Reddy, B. Siva Kumar, Raja Krishnamoorthi, and Ch Priyanka. "A Novel Compression Framework for Electro Cardiac Signals." *Journal of Physics: Conference Series*. Vol. 1964. No. 6. IOP Publishing, 2021.