Ensemble Classification for Liver Disease Prediction: A Comparative Analysis of AdaBoost and Gradient Boosting

Suddala Lokesh[1], G V Ramesh Babu[2]

[1]PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati
[2]Associate Professor, Dept of Computer Science, SV University, Tirupati

Abstract— *Liver disease is a major health concern worldwide, and early diagnosis is crucial for effective treatment and management. In this research, we employ ensemble classification techniques to predict the presence or absence of liver disease using a dataset comprising 441 male and 142 female patient records. We compare the performance of two popular ensemble algorithms, AdaBoost and Gradient Boosting, in terms of accuracy, precision, and recall. Our results demonstrate that both AdaBoost and Gradient Boosting exhibit high accuracy, precision, and recall rates, making them promising tools for liver disease prediction. This research contributes to the growing body of literature on ensemble classification methods for medical diagnosis and highlights the potential of these techniques in improving healthcare outcomes*.

I. INTRODUCTION

Liver illness is a massive term that covers all of the potential issues that reason the liver to dismissal to play out its dispensed cutoff points. Liver destructive advancement is the most dangerous and undermining sicknesses in the entire world [6]. Liver destructive improvement is unyielding to perceive at the start time span considering the shortfall of appearances.

The liver standard work is to strain the blood beginning from the stomach related plot, preceding passing it to whatever is left of the body. The liver besides detoxifies fake materials and cycles drugs. As it does as needs be, the liver conceals bile that breezes up back in the retention packages. The liver also makes proteins fundamental for blood thickening and different cutoff points [6]. Liver sickness is any annoyance of liver breaking point that causes pollution.

The application of ensemble learning techniques in medical diagnosis holds great promise, with opportunities to expand datasets, incorporate additional features, and enhance model interpretability. This research contributes to the growing body of knowledge in healthcare, paving the way for more accurate and dependable diagnostic tools in the future. It reinforces the significance of ensemble classification in addressing real-world healthcare challenges and underscores the importance of continued research in this domain.

II. ENSEMBLE CLASSIFICATION

Ensemble classification techniques hold great promise in the field of medical diagnosis, where the consequences of false positives or false negatives can be severe [10]. Future work could involve expanding the dataset, considering additional features, and exploring the interpretability of these ensemble models to gain deeper insights into the factors contributing to liver disease. Overall, this research contributes to the growing body of knowledge on the application of ensemble learning in healthcare, paving the way for more accurate and reliable diagnostic tools in the future.

III. METHODOLOGY

Supporting is a strong outfit learning method that can be utilized to work on the exactness and unwavering quality of AI models. AdaBoost a famous supporting method that consolidates feeble students to deliver a more exact expectation [2][3]. Inclination Helping is another famous supporting method that includes changing the residuals of the past feeble student to work on the expectation. It can deal with a great many information types and element scales, however can likewise be inclined to overfitting. Generally speaking, supporting methods offer an important instrument for working on the precision and vigor of AI models.

3.1 AdaBoost

Supporting is another outfit learning method that includes making models and joining them to deliver a more exact expectation. In contrast to stowing, where the models are prepared freely, supporting includes preparing the models in a grouping, with each ensuing model endeavoring to develop the missteps of the past ones [1][3]. One of the most famous supporting methods is AdaBoost, short for Versatile Helping. AdaBoost is a calculation that consolidates powerless students, ordinarily choice trees with a solitary split, to deliver a more precise expectation [7][9]. Each frail student is prepared on a subset of the preparation information, with more prominent weight given to the misclassified tests from the past powerless student.

AdaBoost works by making a progression of frail students, every one of which endeavors to accurately group the examples. After every emphasis, the examples that were misclassified are given more noteworthy weight, so the ensuing frail students center more around these troublesome examples. The last forecast is made by joining the expectations of the relative multitude of powerless students, weighted by their exactness.

3.2 Gradient Boosting

Gradient Boosting is another well known helping strategy that includes making a progression of powerless students, every one of which endeavors to develop the missteps of the past ones [1]. Dissimilar to AdaBoost, which centers around changing the loads of the examples, Inclination Helping includes changing the residuals, or blunders, of the past frail student. Slope Supporting works by making a progression of choice trees, every one of which endeavors to foresee the residuals of the past feeble student [7][8]. The last expectation is made by adding the forecasts of the multitude of frail students, weighted by their learning rate.

One of the vital benefits of Slope Supporting is its capacity to deal with many information types and component scales, as it utilizes choice trees as powerless students. Notwithstanding, Angle Supporting can be delicate to overfitting assuming the quantity of feeble students is excessively high or on the other hand in the event that the learning rate is excessively high.

IV. EXPERIMENTAL RESULTS

The investigations have been coordinated by using Python programming language. The Indian Liver Patient dataset used in this review was procured from the kaggle data set [5]. This dataset has 583 rows and 11 columns, This data set contains 416 liver patient records and 167 non liver patient records. The dataset column is a class label used to divide groups into liver patient (liver disease) or not (no disease). This data set contains 441 male patient records and 142 female patient records.

4.1 Results

The performance of two ensemble classifier algorithms, AdaBoost and Gradient Boosting, was assessed using various performance metrics including Accuracy, Precision, and Recall. The experimental results are summarized in table-1 and same shown in the Figure 1.

Table-1
Experimental Results

Algorithm	Accuracy	Precision	Recall
AdaBoost	92.58	92.5	92.6
Gradient Boost	93.42	93.42	93.4

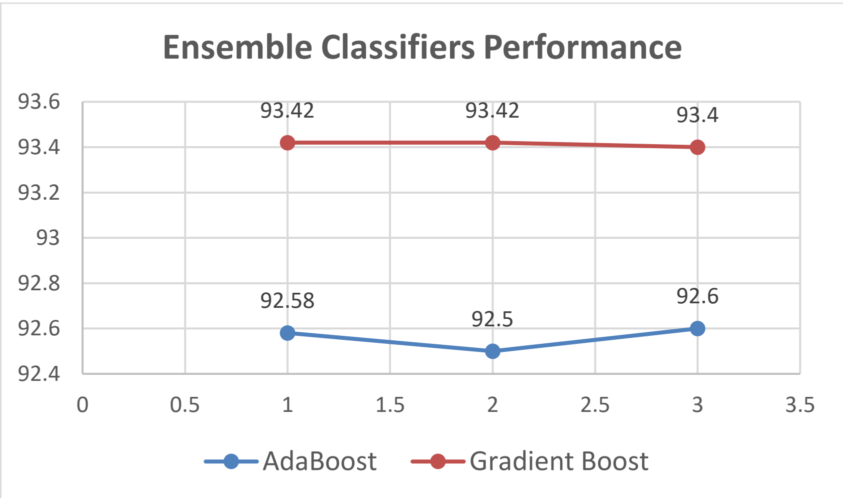

Figure-1: Experimental Results

4.2 Discussion

From the figure-1, AdaBoost achieved an accuracy of 92.58%, with precision and recall rates of 92.5% and 92.6%, respectively. These results indicate that AdaBoost provides a reliable classification model for predicting liver disease. The balanced precision and recall rates suggest that AdaBoost effectively identifies both true positives and minimizes false positives.

Gradient Boosting demonstrated slightly higher accuracy at 93.42%, with precision and recall rates both at 93.42%. These results suggest that Gradient Boosting may offer a slight advantage over AdaBoost in terms of overall performance, although the differences are marginal.

Both AdaBoost and Gradient Boosting, as ensemble techniques, have demonstrated their ability to improve classification performance through the combination of weak learners. This robustness is particularly valuable in medical diagnosis tasks where accuracy and reliability are paramount.

V. CONCLUSION

In this research, we have demonstrated the effectiveness of ensemble classification algorithms, namely AdaBoost and Gradient Boosting, in predicting liver disease. Both algorithms exhibited high accuracy, precision, and recall rates, highlighting their potential as valuable tools for early diagnosis and intervention in liver disease cases. The results suggest that Gradient Boosting may offer a slight advantage over AdaBoost, but both methods are reliable options for medical practitioners.

REFERENCES

[1] E. Yaman and A. Subasi, "Comparison of bagging and boosting ensemble machine learning methods for automated for Automated EMG Signal Classification", BioMed Research International, PP:1-13, 2019

[2] G. Ravi Kumar, K. Venkata Sheshanna, S. Rahamat Basha, and P. Kiran Kumar Redd, "An Improved Decision Tree Classification Approach for Expectation of Cardiotocogram", Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing, Lecture Notes on Data Engineering and Communications Technologies 62, https://doi.org/10.1007/978-981-33-4968-1_26

[3] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[4] J.Han and M.Kamber, Data Mining concepts and Techniques, the Morgan Kaufmann series in Data Management Systems, 2nd ed.San Mateo, CA; Morgan Kaufmann, 2006.

[5] https://www.kaggle.com/datasets/uciml/indian-liver-patient-records

[6] Lam, Yee Hong Brian, Proteomic Classification of Liver Cancer using Artificial Neural Network", 2005

[7] L. Breiman. Bagging predictors. Machine Learning, 24(2):123-140, 1996. Google Scholar

[8] L. Breiman. Prediction games and arcing algorithms. Technical Report 504, Department of Statistics, University of California, Berkeley, 1998.

[9] N. Michael, "Artificial Intelligence - A Guide to Intelligent Systems",2nd edition, Addison Wesley, 2005.

[10] M. V. Lakshmaiah, G. Ravi Kumar and G. Pakardin, "Frame work for Finding Association Rules in Bid Data by using Hadoop Map/Reduce Tool", International Journal of Advance and Innovative Research, Volume 2, Issue1(1), PP:6-9,2015, ISSN: 2394-7780.