Chapter 7

Book Chapter

Liver Disease Prediction Using Ensemble Technique

Pallavi M.O [*1], Mrityunjay K Gavimath [†2], and Elizabath Thomas [‡3]

[1]Assistant Professor, Dept. of MCA, Acharya Institute of Technology, Bangalore
[2]Assistant Professor, Dept. of Mathematics, Shri. S. R. Kanthi Arts, Commerce and Science College, Mudhol
[3]Student, Dept. of MCA Acharya Institute of Technology, Bangalore

Abstract

According to the World Health Organization (WHO), chronic illnesses account for over 59% of global mortality, with liver diseases being a leading cause of death worldwide. Due to the liver's ability to function even when partially damaged, liver issues often go undiagnosed until advanced stages. This paper presents a framework using clinical data and machine learning algorithms to predict liver disease. An ensemble approach processes data from liver patients and healthy donors through Gradient Boosting and AdaBoost classifiers. The model aims to identify high-risk individuals, enabling early detection and treatment, and highlights future integration with the broader healthcare industry.

Keywords: Liver Disease. Machine Learning Algorithm. Chronic Illnesses. Ensemble Technique.

*Email: pallavi250@acharya.ac.in Corresponding Author
†Email: mkgavimathphd@gmail.com
‡Email: elizabath285@gmail.com

1 Introduction

Artificial Intelligence is a system of computer calculations that acquires knowledge from a model through incremental improvement without explicitly requiring coding from a developer. Artificial Intelligence (AI) can be defined as a human-made reasoning system which integrates data with quantifiable devices to predict a future state and produce significant insights . Artificial intelligence combines computation with detection, cognition and action (Patil, Patel, & Lawand, 2023). The cutting edge operates hand in hand with a machine's ability to specifically benefit from the information (i.e., guidance) in order to deliver precise results. Information mining and Bayesian predictive modeling are closely related to artificial intelligence. The machine receives data in the form of info and then executes a computation to schedule responses. Usually AI projects involve making a proposal. For instance , all proposals for films and television shows for Netflix subscribers are based on the client's verified information. Tech companies are using individual problem-solving to enhance their consumer experience by tailoring proposals. AI is also used for other tasks, such as automating tasks, maintaining vision, enhancing portfolios, and detecting extortion.In a number of patient care settings as well as intelligent health systems, artificial intelligence can support healthcare professionals (Gautam & Mittal, 2022; Kumar et al., 2023).In the field of automated illness diagnosis, machine learning has a lot of promise(Al Kuwaiti et al., 2023; Ghazal et al., 2022).

As of late, Data Frameworks and key apparatuses are being consolidated as extra means to help the course of finding of sicknesses in clinical exploration. The liver, a fundamental organ is urgent in catalyst enactment, bile creation, and digestion of fats and capacity of nutrients, glycogen and minerals. Liver illnesses are challenging to analyze and thus are frequently ignored because of the absence of appropriate side effects at the underlying stages. One of the most widely recognized side effects of most liver illnesses is hyper bilirubinemia which is difficult to recognize in early assurance. Anyway, this isn't exactly sure and the view of protein level is expected to recognize and insist the proximity of liver sickness. Different AI procedures have been utilized in the forecast of liver illnesses. In this examination, we propose the use of Choice Tree, Irregular Timberland Calculation and Backing Vector Machine methods in the expectation of liver illness by Paired Arrangement of the dataset into two given classes of patient encountering liver affliction or not. The dataset contains data about quiet credits like Complete Bilirubin, Alanine Aminotransferase, Direct Bilirubin, Aspartate Aminotransferase, Age, Orientation, Egg whites, All out Proteins, Antacid Phosphatase, Egg whites and Globulin Proportion and the Outcome.

Editors: S. Pandikumar, Manish Kumar Thakur, Pallavi M O
DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5

The forecast from the previously mentioned calculations are thought about on the boundaries of Precision and different blunder computations to decide the most appropriate calculation. Liver conditions generated an abundance of data, such as metabolomics analyses, electronic health records, and reports with patient clinical information and problems. Nevertheless, if these data are to produce models concerning physiological systems of pathogenesis, they need be dissected and integrated. We adopt artificial intelligence (AI) as a classifier for vast data sets in the liver in order to predict and restore disclosure. Liver illnesses typically affect adults in the 40–60 age range, with men being the most commonly affected (Gupta et al., 2022). A 23-credit dataset containing documentation of 7000 patients—5295 of whom were male and the remaining patients were female—was generated. The proposed approach for the prediction of liver infections incorporates the use of information mining techniques such as Support Vector Machine (SVM), Helped C5.0, and Credulous Bayes (NB). These classifier strategies' exhibits are evaluated with particularity, responsiveness, and accuracy.

In this paper we are going examine how to anticipate chance of liver sickness for an individual, in light of the blood test report consequences of the client. In this paper, the gamble of liver sickness was anticipated utilizing different AI calculations. The last result was anticipated in light of the most dependable AI calculation. In light of the exact model we planned a framework which requests that an individual enter the subtleties of his/her blood test report. Then, at that point, the framework utilizes the most reliable model which is prepared to foresee, regardless of whether an individual has hazard of liver illness.

2 Methodology

- Data Collection:
  Gathering data is the first and most important stage in creating an AI model. This crucial stage has a big influence on the model's quality; the more complete and superior the data, the more effective the model will be. A few methods, such as manual intercessions, web scraping, and so on, are used to gather the data. Web scraping in the medical field can save many lives by facilitating the making of informed decisions (Lotfi et al., 2021).Figure 1 represents the architecture of the model
- Dataset:
  There are 583 distinct data points in the dataset. The dataset is divided into 11 components, which are displayed below.
  - Age: Age of the patient
  - Female: Gender of the patient (1 if Female, 0 if Male)
  - TB: Total Bilirubin

Editors: S. Pandikumar, Manish Kumar Thakur, Pallavi M O
DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5

- DB: Direct Bilirubin
- Alkphos: Alkaline Phosphotase
- Sgpt: Alamine Aminotransferase
- Sgot: Aspartate Aminotransferas
- TP: Total Protien
- ALB: Albumin
- A/R: Albumin and Globulin Ratio
- class: 1 Liver diseases and 0 no diseases

- Data Preparation:
  Combat data and set it up for planning. Eliminate redundant information, rectify mistakes, address missing characteristics, adjust data categories, calibrate, and clean up the necessary power. Data must be randomized to eradicate the impact of the particular inquiry for which we gathered the information, or perhaps prearranged. Analyze data to support the identification of pertinent correlations between variables, class imbalances (warning about predisposition!), or further exploratory investigation.
- Model Selection:
  We implemented the machine learning algorithm GradientBoostingClassifier + AdaBoostClassifier (Ensemble Technique) after achieving 92.1% precision on the test set.
- Ensemble Technique:
  Troupe tactics are methodologies that incorporate several models instead of relying just on one in an attempt to boost the accuracy of model outcomes.The robustness of the individual models is strengthened by the ensemble learning technique (Xiao, 2019). The preciseness of the results is substantially improved by the combined models. This has made such methods increasingly prevalent in artificial intelligence.
- Categories of Ensemble Methods :
  There are two broad categories of outfit strategies: equal group ways and consecutive troupe procedures. Base pupils are created in a grouping using successive outfit approaches, such as Versatile Helping (AdaBoost). The base students' growing age increases their dependence on one another. Next, by giving newly distorted students larger loads, the model's exhibition is enhanced. Base pupils are generated in an equal configuration—for example, irregular backwoods—in equal outfit operations. In order to promote autonomy among basic pupils, equal techniques take advantage of the students' equal ages. The freedom of base pupils virtually eliminates errors due to the usage of midpoints.

DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5

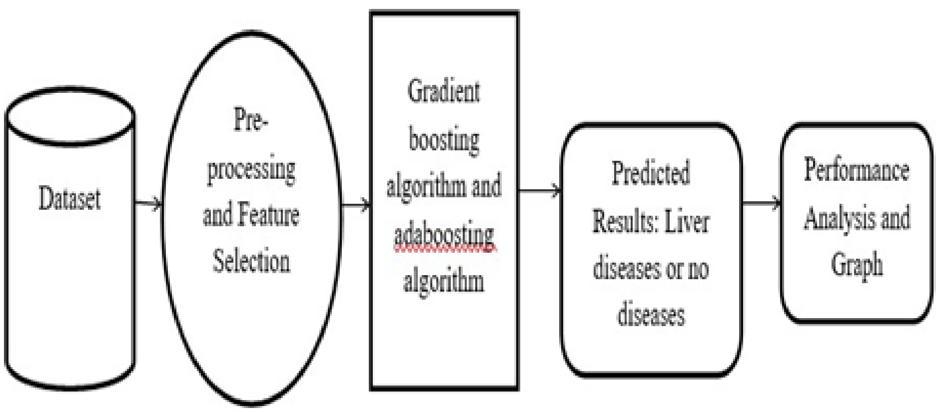

Figure 1. Architecture

DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5

Primary Types of Troupe Methods

1. Bagging

   The brief bootstrap structure, known as stowing, is mostly used in order and relapse prevention. Using choice trees, it increases the models' accuracy while typically reducing variation. A reduction in variation increases precision by eliminating overfitting, a challenge to many predictive models.There are two types of stowing: aggregation and bootstrapping. Using the substitution technique, tests are obtained from the complete population (set) in a strategy known as bootstrapping.The replacement technique for examination facilitates the randomness of the selection process. The base learning calculation is then applied to the cases to finish the approach. The final step in packing is accumulation, which combines all possible forecast outcomes and randomizes the outcome. Expectations won't be accurate without collecting since all outcomes won't be carefully considered. As a result, the total is determined by bootstrapping procedures based on likelihood or by taking into account all of the predictive models' outputs. Because stacked students combine to form a single, solid student that is more stable than a single student, stacked students are extremely important. It also eliminates any modification, which reduces the overfitting of duplicates. The computational expense of stowing is one of its drawbacks. In this approach, when the proper stowing process is ignored, it can lead to greater predilection in models.

2. Boosting

   Assisting is a data collection technique that capitalizes on previous indicator errors to enhance future forecasts. By combining a few weak base pupils, the strategy shapes one's areas of strength, so improving model consistency. Assisting involves organizing weaker students into groups and using the knowledge that the next student in line will impart to the weaker students to create stronger role models. A variety of structures are needed to provide assistance, such as XGBoost (Outrageous Slope Supporting), Versatile Helping (AdaBoost), and angle supporting. AdaBoost treats weaker pupils as choice trees, which typically contain one split popularly referred to as choice stumps. The main choice stump for AdaBoost comprises perceptions that communicate similar burdens. Slope assistance increases the precision of the model by adding indicators to the group in a sequential manner, with the addition of indications exactly before their replacements. The effects of errors in previous indicators are mitigated by new indicators. The slope sponsor can identify problems with students' expectations and address them by using the angle of fall. XGBoost improves speed and execution by using decision trees with assisted angles. It is very dependent upon the pace at which computations are

DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5

performed and the objective model's display. Because model preparation should adhere to a grouping, inclination-supported machine execution is sluggish.

3. Stacking

   Another method of outfitting is called stacking, which is commonly referred to as stacked speculation. Combining classifiers and maximizing prediction accuracy is a prevalent usage scenario for stacking (Ledezma et al., 2010). The way this strategy operates is by allowing a preparation calculation to forecast a few more similar learning calculations. Stacking has been used successfully in characterizations, thickness evaluations, relapse, and distance learning. It can also be used to calculate the error rate needed for stowing.

- Analyse and Prediction:

  We selected just ten features from the actual dataset:
  - Age: Age of the patient
  - Female: Gender of the patient (1 if Female, 0 if Male)
  - TB: Total Bilirubin
  - DB: Direct Bilirubi
  - Alkphos: Alkaline Phosphotase
  - Sgpt: Alamine Aminotransferas
  - Sgot: Aspartate Aminotransferase
  - TP: Total Protiens
  - ALB: Albumin
  - A/R: Albumin and Globulin Ratio
  - class: 1 Liver diseases and 0 no diseases

  Our accuracy on the test set was 92.1%.

3 Result

Utilizing different techniques, we start our concentrate in this part with the information handling stage and proceed to highlight extraction, classification, and expectation examination. The traits utilized in the dataset are Age, Direct Bilirubin, Complete Bilirubin, Soluble Phosphate, Alamine Aminotransferase, Aspartate Aminotrans-ferase (see figure 2), Proteins, Egg whites, and Globulin Proportion, Egg whites Dataset (where informational index is the class name). Every histogram tells us about the recurrence dispersions for different patients in that specific characteristic. Each trait included in the informative index and its relationship are plotted (see figure 3). It can be established that there is a significant correlation between immediate and total bilirubin level.

Editors: S. Pandikumar, Manish Kumar Thakur, Pallavi M O
DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5

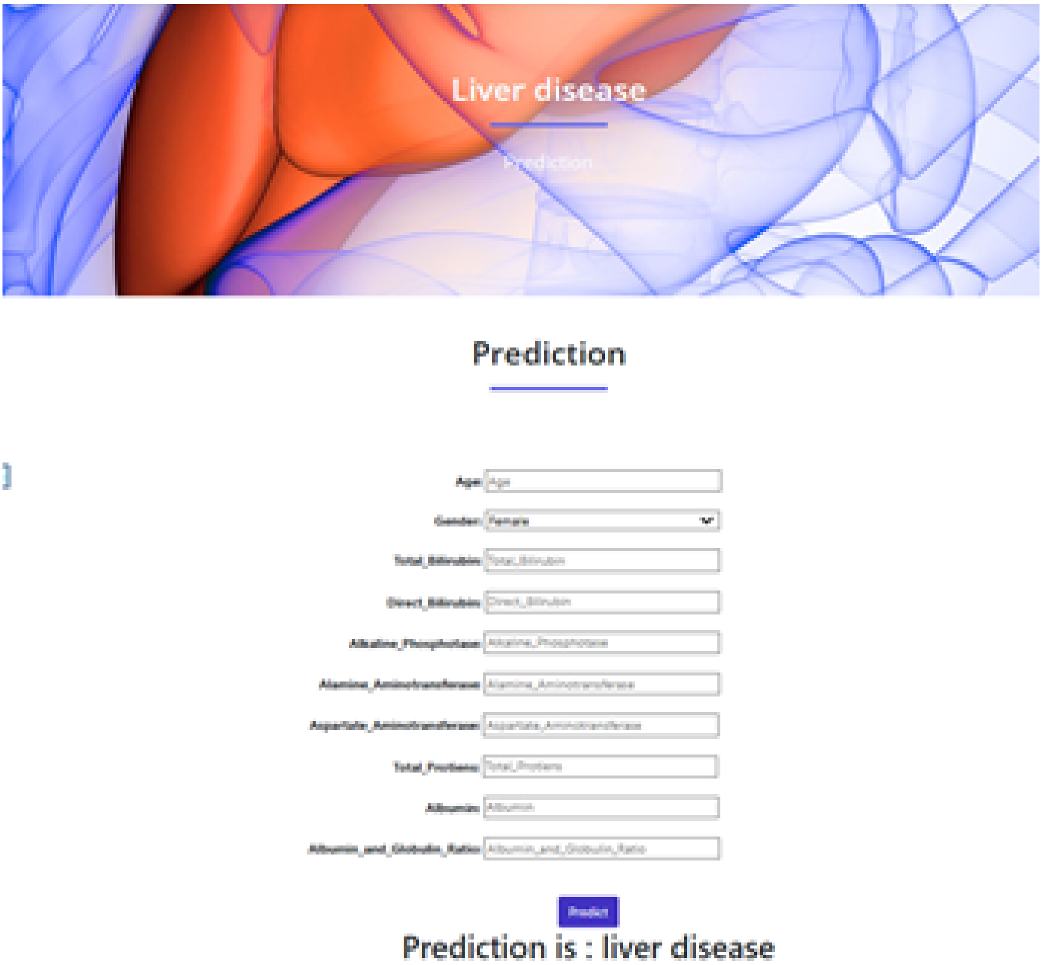

Figure 2. Prediction

DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5

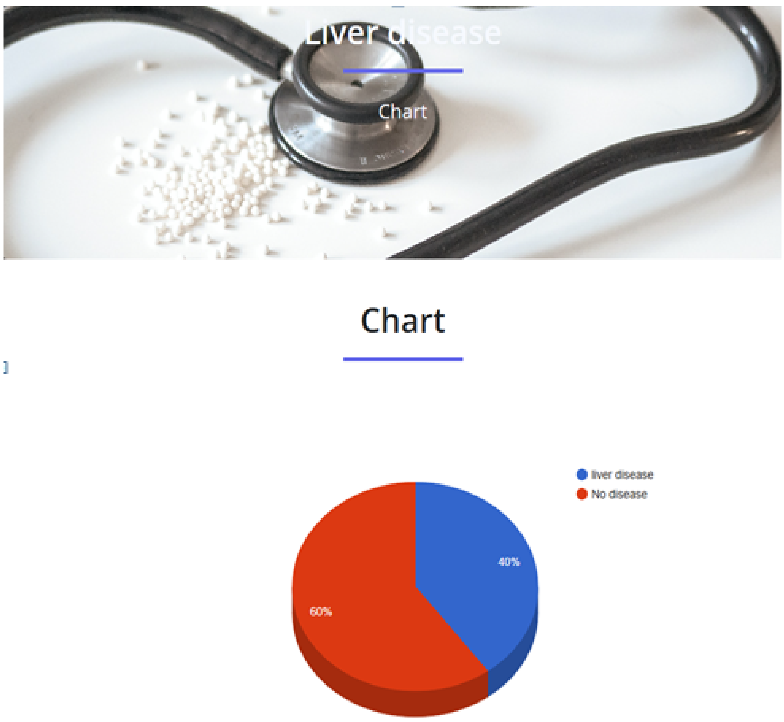

Figure 3. Analysis

Editors: S. Pandikumar, Manish Kumar Thakur, Pallavi M O
DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5

4 Conclusion

In order to develop a unique predictive model for liver disease diagnosis, the authors have developed an elaborate and organized framework that incorporates every step of the data preparation, preprocessing, model selection, and incorporation process. With a flawless training accuracy of 100% and a strong test accuracy of 92%, the model demonstrated exceptional performance, demonstrating its dependability and generalization competencies. These outcomes have important ramifications for the medical community, especially when it comes to liver disease, where prompt and accurate diagnosis is vital for successful treatment and care. The capacity of the model to generate precise projections can help medical practitioners diagnose liver illness early on, which may result in more timely therapies and improved outcomes for patients. By demonstrating how cutting-edge machine learning approaches may be adopted to address major health concerns, this research makes an essential contribution to prognostic healthcare solutions. Ultimately, the suggested model may prove to be an invaluable instrument in situations in medicine, assisting the medical field in reaching more knowledgeable and potentially life-saving conclusions.

References

Al Kuwaiti, A., Nazer, K., Al-Reedy, A., Al-Shehri, S., Al-Muhanna, A., Subbarayalu, A. V., Al Muhanna, D., & Al-Muhanna, F. A. (2023). A Review of the Role of Artificial Intelligence in Healthcare. Journal of Personalized Medicine, 13(6). https://doi.org/10.3390/jpm13060951

Gautam, S., & Mittal, P. (2022). Comprehensive Analysis of Privacy Preserving Data Mining Algorithms for Future Develop Trends. International Research Journal of Computer Science, 9(10), 367–374. https://doi.org/10.26562/irjcs.2022.v0910.01

Ghazal, T. M., Rehman, A. U., Saleem, M., Ahmad, M., Ahmad, S., & Mehmood, F. (2022). Intelligent Model to Predict Early Liver Disease using Machine Learning Technique. 2022 International Conference on Business Analytics for Technology and Security, ICBATS 2022. https://doi.org/10.1109/ICBATS54253.2022.9758929

Gupta, K., Jiwani, N., Afreen, N., & Divyarani, D. (2022). Liver Disease Prediction using Machine learning Classification Techniques. Proceedings - 2022 IEEE 11th International Conference on Communication Systems and Network Technologies, CSNT 2022, 221–226. https://doi.org/10.1109/CSNT54456.2022.9787574

Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future re-

DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5

search agenda. Journal of Ambient Intelligence and Humanized Computing, 14(7), 8459–8486. https://doi.org/10.1007/s12652-021-03612-z

Ledezma, A., Aler, R., Sanchis, A., & Borrajo, D. (2010). GA-stacking: Evolutionary stacked generalization. Intelligent Data Analysis, 14(1), 89–119. https://doi.org/10.3233/IDA-2010-0410

Lotfi, C., Srinivasan, S., Ertz, M., & Latrous, I. (2021). Web Scraping Techniques and Applications: A Literature Review. Scrs Conference Proceedings on Intelligent Systems, 381–394. https://doi.org/10.52458/978-93-91842-08-6-38

Patil, N. H., Patel, S. H., & Lawand, S. (2023). Research Paper On Artificial Intelligence And It's Applications. Journal of Advanced Zoology. https://doi.org/10.53555/jaz.v44is8.3544

Xiao, J. (2019). SVM and KNN ensemble learning for traffic incident detection. Physica A: Statistical Mechanics and its Applications, 517, 29–35. https://doi.org/10.1016/j.physa.2018.10.060

DOI:10.48001/978-81-966500-2-5-7 | ISBN: 978-81-966500-2-5