Journal of Information Systems Engineering and Management
2025, 10(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Computer-Aided Diagnosis for Cardiovascular and Hepatic Disorders using Boosted Ensemble Deep Learning

Dolley Srivastava[1], Himanshu Pandey[2], Hitendra Singh[3], Rakesh Kumar Yadav[4], Ambuj Agarwal[5], Richa Sharma[6]

Maharishi University of Information Technology, Lucknow , India[1,3,4,6]

dolley89@gmail.com, hitendra.singh@muit.in, rkymuit@gmail.com,richa.sharma.ar@gmail.com

Faculty of Engineering &Technology, University of Lucknow, India[2]

hpandey010@gmail.com

Department of Computer Science And Engineering, Sharda School of Engineering & Technology, Sharda University, Greater Noida, India[5]

ambuj4u@gmail.com

ARTICLE INFO

Received: 05 Oct 2024

Revised: 30 Nov 2024

Accepted: 12 Dec 2024

ABSTRACT

In recent years, machine learning has gained traction as a potential tool for improving the accuracy and timeliness of illness diagnoses. The use of machine learning for the diagnosis of cardiovascular and renal disorders is critically examined in this research. To enhance patient outcomes, it is essential to diagnose cardiovascular and hepatic illnesses early and accurately. The interpretation of complicated clinical data and the identification of detailed patterns indicative of these disorders, however, may be difficult for standard diagnostic approaches. This study thoroughly tests three cutting-edge boosting algorithms: XGBoost (Extreme Gradient Boosting), CatBoost, and AdaBoost. Enabling the capture of complex nonlinear interactions and management of varied data sources, these ensemble approaches repeatedly integrate and optimize numerous weak learners. After conducting thorough experiments on massive clinical datasets, results show that the XGBoost algorithm is the best for certain types of diseases. Intelligent diagnostic tools that can reliably identify cardiovascular and hepatic diseases early on are within reach, according to the results of this study. This will lead to better patient care and disease management methods.

Keywords: Boosting Algorithms, CatBoost, XGBoost, AdaBoost, Cardiovascular Disorders, Hepatic Disorders, Machine Learning.

I. INTRODUCTION

An expanding subfield of AI, pattern recognition is ostensibly concerned with the mechanization of learning as it pertains to the automated finding of data regularities via the use of computer programs. An essential assurance of pattern recognition algorithms is that they will discover any data structure that permits accurate categorization. Since the null hypothesis may be accepted regardless of the data, pattern recognition has a clear advantage over traditional statistical analysis.

Like sketching a tree, the decision-making process for illness diagnosis involves considering several alternatives, chances, and decisions at each branch point. Determining diagnoses based on probability alone becomes increasingly challenging as illness complexity increases, and it is challenging to express decision-making processes as simple if-then-else rules. Here, R is a rule, algorithm, or decision something; O is an observed dataset; and T is a finite set of diagnoses; the alternative to a rule-based system is a mapping R: O -> T for diagnostic entities. This method allows for a more comprehensive evaluation of a variety of observations, making it more adaptable and less limiting in its pursuit of a diagnosis.

Both in India and across the globe, a sizable portion of the population is impacted by cardiac and hepatic problems, which constitute a significant group of diseases when considering the impact on health and death

rates[1]. The goal of better illness treatment must include early and precise diagnosis. Diagnostic imaging examinations such as X-ray, CT, or MRI, in addition to a thorough clinical examination and evaluation of symptoms, are considered the gold standard for these types of disorders[2]. Diagnostic uncertainty and inconclusiveness persist in many situations when using regular clinical and imaging data, despite the high cost and huge number of tests involved. Patients with these illnesses are often treated symptomatically or empirically without settling on a definitive diagnosis since many of these diagnostic tests are not available or affordable in rural and remote places.

BACKGROUND

In light of the present state of illness diagnosis and the possibilities presented by machine learning, we are interested in comparing the efficacy of various machine learning approaches for the diagnosis of chronic diseases, cardiovascular disease, and renal disease. By looking for patterns in the symptoms reported by patients with each condition, this research will help in the creation of diagnostic tools. The machine learning algorithms have been refined for automated diagnosis and this study's findings are used to personalize treatment plans for each patient.

The field of medical diagnostics stands to benefit greatly from machine learning (ML). Its predictive power and capacity to understand patterns in data might be a game-changer in medical diagnosis[3]. The use of machine learning for diagnosis has advanced to the point where several examples of success exist now.

Healthcare practitioners are facing a critical shortage of time and diagnostic resources due to the rising number of patients requiring medical attention. People who suffer from heart failure often have more than one co morbid illness and must make difficult choices about which treatment is best for them. There are a lot of other illnesses that seem similar[4].

Due to its potential beneficial effects on healthcare spending and quality, the subject of illness diagnosis has garnered a lot of interest during the last few decades. Medical expenses and illness progression may both be mitigated with prompt and precise diagnosis. More than 35% of the global population deals with some kind of chronic illness, says the World Health Organization (WHO). In addition, they said that diabetes and cardiovascular disease had surpassed all others as the leading killers in the last five years. The prevalence of infectious diseases is another issue that has grown in less developed nations. Ineffective disease control measures have been used in those nations due to a lack of reliable diagnoses and monitoring. Infectious diseases, in contrast to chronic diseases, are those that arise from a microbe or other pathogen. If caught early on, this condition is more amenable to treatment. Opposite to this are chronic illnesses. In spite of a 5.5-year rise in worldwide life expectancy between 2000 and 2015, research found that the number of years people lived with impairment due to chronic illness rose by as much as 52% over the same time period.

BOOSTING ALGORITHMS

Over the last ten years, boosting algorithms have exploded in popularity. Improving performance above random guessing is all that is needed for the generic boosting framework to operate[5], [6]. A committee classifier significantly lowers the training and testing error rates by merging these weak learners using a weighted majority vote. In this study, we investigate three different boosting algorithms:

Adaptive Boosting of weights (the name of the technique) and fitting of weak learners (sometimes called "base procedures") constitute a sequential additive process that, when combined, yields a single ensemble learner, according to AdaBoost, a boosting algorithm[7]. A linear combination of the predictions from the ensemble of weak learners is fed into the final learner so it can create predictions. While Freund and Schapire first presented the approach in the context of binary classification[8], it has since been modified to address regression issues. Because of its resistance to overfitting and its effectiveness in providing correct ensembles, AdaBoost was dubbed the "best off-the-shelf classifier in the world".

Chen and Guestrin presented XGBoost, short for "eXtreme Gradient Boosting," a tree boosting technique that is very scalable[9]. Quickly rising to prominence since its release, XGBoost is now among the machine learning community's most popular and effective boosting algorithms[10]. In particular, we highlight the following XGBoost features:

1. XGBoost uses a loss function + regularization formalism for tree boosting.

2. It combines regression and classification problems into one unified approach. XGBoost computes scores

or weights based on the chosen loss function and regularization parameters.

3. Instead of a first-order approach, which is used in gradient boosting implementations, it uses Newton approximations to solve the loss + regularization optimization problem.

4. Depending on data volumes, computing infrastructure, and parallelization capabilities, one can choose from multiple algorithms to grow trees.

5. A subsampling approach that includes both feature space and training samples to avoid overfitting, along with regularization.

A robust open-source gradient boosting technique for machine learning, CatBoost was created by Yandex[11], [12]. In particular, it excels at handling data with several categorical attributes, although it can manage numerical, categorical, and text data as well. The speed, precision, and good handling of missing information and outliers are some of CatBoost's well-known characteristics.

CONTRIBUTIONS

This study will focus on the prediction of long-term health risks in particular. The main points of the paper are as follows:

- Building a state-of-the-art ML system that incorporates accelerated ensemble learning for precise diagnostic categorization of hepatic and cardiovascular diseases.
- Thorough assessment and comparison of three cutting-edge boosting algorithms: XGBoost (Extreme Gradient Boosting), CatBoost, and AdaBoost, within the framework of diagnosing hepatic and cardiovascular diseases.
- Capturing complex patterns and correlations within multimodal data requires the integration of multiple clinical data sources, such as patient demographics, physiological measures, laboratory findings, medical imaging data, and more.
- Results from rigorous testing on massive clinical datasets show that the XGBoost algorithm performs better than competing algorithms when it comes to diagnosing liver and cardiovascular diseases.
- In a quantitative study, the suggested strategies outperformed both conventional diagnostic procedures and standalone ML models in terms of classification accuracy.
- Setting the stage for XGBoost-based intelligent diagnostic systems to be developed for the early and accurate diagnosis of hepatic and cardiovascular diseases, with the possibility of their incorporation into clinical decision support systems.
- Aid in the widespread use of enhanced ensemble learning methods for medical diagnosis and illness categorization by providing evidence of their effectiveness in managing complicated clinical data and enhancing diagnostic precision.

What follows is an outline of the rest of the paper. In Section 1, we get a brief synopsis of the paper's goals and an overview of the global literature on cardiovascular and hepatic disorders related to bleeding and machine learning-based prediction models. The second section provides a literature overview of related works. Part 3 of the article lays out the necessary background information to tackle the issue at hand by explaining the technique that will be used in the suggested solution. Section 4 presents the outcomes that were achieved using the technique that was suggested. The article concludes with the statements made in Section 5.

LITERATURE REVIEW

Different types of liver illnesses are classified according to the causes and symptoms they cause. Infections, injuries, toxic material or medication exposure, processes, or genetic abnormalities (such hemochromatosis) may all be causes. Hepatitis, cirrhosis, and stones are all possible outcomes of the aforementioned conditions; as stones get larger, they may cause obstructions, fatty infiltration, and, very rarely, liver cancer. The buildup and accumulation of toxic substances, including iron or copper, and disruptions to essential liver activities may also result from genetic disorders[13], [14].

Lipid buildup in the liver is a hallmark of non-alcoholic fatty liver disease (NAFLD), a leading cause of liver damage. A condition known as "non-alcoholic steatohepatitis"[15] occurs when the liver cells become inflamed and damaged. Among the most devastating liver illnesses is cirrhosis. Scar tissue develops in lieu of healthy tissue as a result of this condition. As a result, the liver is crippled and will never work the same. Drunkenness, non-alcoholic fatty liver disease, chronic hepatitis C, and chronic hepatitis B are the primary factors that lead to liver cirrhosis[16].

Acute hepatitis[17] causes the liver to inflame and die quickly, whereas chronic hepatitis[18] causes the liver to inflame and die slowly over a long time. Infection with a member of the hepatitis virus family is the most common cause of hepatitis, however, any of the aforementioned may lead to the disease. Hepatitis A, B, C, D, and E are the names given to these viruses in the sequence in which they were found.

Histopathological examinations have long been the foundation upon which medical reports of patients' conditions have been built. Efficient techniques for data collecting, processing, and visualization have emerged because of advancements in information and communication technology, particularly in machine learning (ML) and artificial intelligence (AI)[19]. By integrating the results of clinical procedures with those of AI and ML models, clinicians may enhance their illness detection judgments even more. Predicting the early onset of complications in diabetes (as a classification problem[20] or a regression task for short-term glucose prediction[21]), cholesterol[22], hypertension[23], hypercholesterolemia[22], COPD, COVID-19 [24], stroke[1], CKD, lung cancer[25], sleep disorders, CVDs[26], etc. has benefited greatly from ML techniques.

MATERIAL AND METHODS

I. Dataset

Datasets used in this study include those pertaining to liver disease[27] and heart disease[28]. There are 3,197,95 records in the Heart Disease dataset, which includes 17 distinct attributes that include patients' health information. The liver dataset includes demographic information such as age, gender, height, weight, status, duration till death or last follow-up, and last follow-up. All of the data used in this study came from publicly available sources. We hoped to show that our suggested machine learning model was robust and generalizable by using these two different datasets to forecast the occurrence of liver disease and heart disease, two common health problems that have substantial social and economic consequences.

II. Methodology

Machine learning models for the categorization of hepatic and cardiovascular illnesses were developed in this work using three state-of-the-art boosting algorithms: XGBoost (Extreme Gradient Boosting), CatBoost, and AdaBoost.

A. XGBoost

It utilizes a parallel and distributed computing approach, making it suitable for handling large-scale datasets.

Step 1: Initialize the Model

Initialize the prediction for each instance with a constant value (e.g., the mean of the target variable for regression or the log-odds for classification):

$$y_i^0 = c, for\ i = 1, 2, \ldots, n (1)$$

Where:

y_i^0 is the initial prediction for instance i.

c is a constant value (e.g., mean or log-odds).

n is the number of training instances.

Step 2: Iterate for M Boosting Rounds

For m = 1 to M:

1. Calculate the Gradients and Second-Order Gradients: For each instance i, calculate the gradient g_i and the

second-order gradient h_i:

$$g_i = \frac{\partial L(y_i, y_i^{m-1})}{\partial y_i^{m-1}} (2)$$

$$h_i = \frac{\partial^{2L(y_i, y_i^{m-1})}}{\partial (y_i^{m-1})^2} (3)$$

Where:

- y_i is the true target value for instance i.
- y_i^{m-1} is the predicted value from the previous iteration.
- L is the loss function (e.g., mean squared error for regression or logistic loss for classification).

2. Fit a Decision Tree: Fit a decision tree $f_{m(x)}$ to the gradients g_i and second-order gradients h_{I} using the objective function:

$$obj^m = sum_{\{i=1\}}^{n} \left[g_i f_{m(x_i)} + \left(\frac{1}{2}\right) h_i f_m^{2(x_i)} \right] + \Omega(f_m) (4)$$

Where:

- $\Omega(f_m)$ is the regularization term to control the complexity of the tree.

3. Calculate Optimal Leaf Weights : For each leaf j of the decision tree $f_{m(x)}$, calculate the optimal weight w_j^* using:

$$w_j^* = -\frac{\left(sum_{\{i \in I_j\}} g_i\right)}{\left(sum_{\{i \in I_j\}} h_i + \lambda\right)} (5)$$

Where:

- I_j is the set of instances that belong to leaf j.
- λ is the L2 regularization parameter.

4. Update the Model: Update the predictions for each instance by adding the new decision tree with the optimal leaf weights:

$$y_i^m = y_i^{m-1} + f_{m(x_i)} (6)$$

Where:

- y_i^m is the updated prediction for instance i after iteration m.

Step 3: Output the Final Model

The final model is the sum of all the decision trees from the M boosting rounds:

$$y_i = sum_{\{m=1\}}^{M} f_{m(x_i)} (7)$$

Where:

- y_i is the final prediction for instance i.

During the training process, XGBoost also employs techniques such as column and row subsampling, and parallel and distributed computing to improve efficiency and prevent overfitting.

B. CatBoost

CatBoost employs techniques such as Ordered Target Encoding and Ordered Boosting to improve its predictive performance.

Step 1: Data Preprocessing: CatBoost performs several preprocessing steps on the input data:

1. Ordered Target Encoding: For categorical features, CatBoost replaces each category value with the corresponding average of the target variable for that category. For a categorical feature x_j and target y, the encoding is:

$$x_j^{encoded} = \text{avg}(y \mid x_j)(8)$$

2. Missing Value Handling: Missing values in numerical and categorical features are treated as separate categories during training.
3. Feature Combinations: CatBoost automatically generates new features by combining existing features, which can capture non-linear interactions.

Step 2: Initialize the Model: Initialize the prediction for each instance with a constant value (e.g., the mean of the target variable for regression or the log-odds for classification):

$$y_i^0 = c, for\ i = 1, 2, \ldots, n(9)$$

Where:

- y_i^0 is the initial prediction for instance i.
- c is a constant value (e.g., mean or log-odds).
- n is the number of training instances.

Step 3: Iterate for M Boosting Rounds

For m = 1 to M:

1. Calculate the Gradients: For each instance i, calculate the gradient g_i:

$$g_i = \frac{\partial L(y_i, y_i^{m-1})}{\partial y_i^{m-1}}(10)$$

Where:

- y_i is the true target value for instance i.
- y_i^{m-1} is the predicted value from the previous iteration.
- L is the loss function (e.g., mean squared error for regression or logistic loss for classification).

2. Fit an Oblivious Decision Tree : Fit an oblivious decision tree $f_{m(x)}$ to the gradients g_i using the objective function:

$$obj^m = sum_{\{i=1\}}^n L\left(y_i, y_i^{m-1} + f_{m(x_i)}\right) + \Omega(f_m)(11)$$

Where:

- $\Omega(f_m)$ is the regularization term to control the complexity of the tree.

In an oblivious tree, the splits are made based on a fixed set of features at each level, rather than choosing the best feature for each node.

3. Calculate Leaf Values : For each leaf j of the oblivious tree, calculate the leaf value v_j:

$$v_j = \frac{\left(sum_{\{i \in I_j\}} g_i\right)}{\left(sum_{\{i \in I_j\}} h_i + \lambda\right)}(12)$$

Where:

- I_j is the set of instances that belong to leaf j.
- λ is the L2 regularization parameter.
- h_i is the second-order gradient, which is approximated using the Newton-Raphson method in

CatBoost.

4. Update the Model : Update the predictions for each instance by adding the new oblivious tree with the leaf values:

$$y_i^m = y_i^{m-1} + f_{m(x_i)} (13)$$

Where:

- y_i^m is the updated prediction for instance i after iteration m.

Step 4: Output the Final Model: The final model is the sum of all the oblivious trees from the M boosting rounds:

$$y_i = sum_{\{m=1\}}^{M} f_{m(x_i)} (14)$$

Where:

- y_i is the final prediction for instance i.

C. AdaBoost

It iteratively trains weak learners on reweighted versions of the data, focusing on instances that were misclassified in the previous iterations.

Step 1: Initialize the Weights: Initialize the weights for each training instance i:

$$w_i^1 = \frac{1}{n}, for\ i = 1, 2, \dots, n (15)$$

Where:

- w_i^1 is the initial weight for instance i.
- n is the number of training instances.

Step 2: Iterate for M Boosting Rounds

For m = 1 to M:

1. Train a Weak Learner- Train a weak learner (e.g., a decision tree) $f_{m(x)}$ using the weighted training data, where the weight of each instance i is w_i^m.
2. Calculate the Error Rate- Calculate the weighted error rate ε_m of the weak learner:

$$\varepsilon_m = sum_{\{i=1\}}^{n} w_i^m * \frac{I\left(y_i \neq f_{m(x_i)}\right)}{sum_{\{i=1\}}^{n} w_i^m} (16)$$

Where:

- y_i is the true target value for instance i.
- $I\left(y_i \neq f_{m(x_i)}\right)$ is an indicator function that returns 1 if $y_i$is not equal to the prediction $f_{m(x_i)}$, and 0 otherwise.

3. Calculate the Weight Update Factor -Calculate the weight update factor α_m:

$$\alpha_m = \log\left(\frac{(1 - \varepsilon_m)}{\varepsilon_m}\right) (17)$$

4. Update the Instance Weights- Update the weights for each instance i:

$$w_i^{m+1} = w_i^m * \exp\left(\alpha_m * I\left(y_i \neq f_{m(x_i)}\right)\right) (18)$$

Where:

- w_i^{m+1} is the updated weight for instance i after iteration m.

This step increases the weights of misclassified instances and decreases the weights of correctly classified instances, allowing the next weak learner to focus more on the difficult instances.

Step 3: Output the Final Model: The final model is a weighted sum of the weak learners:

$$F(x) = sum_{\{m=1\}}^{M} \alpha_m * f_{m(x)} (19)$$

Where:

- F(x) is the final prediction for instance x.
- α_m is the weight update factor for the m-th weak learner.
- $f_{m(x)}$ is the prediction of the m-th weak learner.

III. Model Training and Evaluation

For each dataset (Liver and Heart), we performed the following steps:

- Data Preprocessing: The datasets were preprocessed to handle missing values and scale numerical features. Missing values were imputed using mean imputation, and numerical features were scaled using StandardScaler from the scikit-learn library (Pedregosa et al., 2011).
- Train-Test Split: The preprocessed datasets were split into training and test sets using an 80/20 stratified split to ensure an adequate representation of both classes in each subset.
- Hyperparameter Tuning: For each algorithm, a grid search was performed to tune the hyperparameters using 5-fold cross-validation on the training set. The hyperparameters were optimized to maximize the area under the receiver operating characteristic curve (AUC-ROC) score.
- Model Training: The tuned algorithms were trained on the entire training set using the optimal hyperparameters obtained from the grid search.
- Model Evaluation: The trained models were evaluated on the held-out test set using various performance metrics, including accuracy, precision, recall, F1-score, and AUC-ROC.

To find the best method for cardiovascular and hepatic disease classification, we analyzed the three boosting algorithms' performance on the Heart dataset and the Liver dataset. The algorithms that were considered were XGBoost, CatBoost, and AdaBoost.

RESULT

Three cutting-edge boosting algorithms—XGBoost (Extreme Gradient Boosting), CatBoost, and AdaBoost—were tested in this research to see how well they classified liver and cardiovascular diseases. A large dataset including clinical data and diagnostic results for various diseases was used to test the algorithms. Tabulated in Table 1 are the key points from the performance review. Cardiovascular diseases and hepatic problems each have their own set of published outcomes.

Table 1: Performance of Boosting Algorithms on the Classification of Cardiovascular and Hepatic Disorders

Model	AUC		CA		F1		Prec		Recall	
	Cardiovascular Disorders	Hepatic Disorders	Cardiovascular Disorders	Hepatic Disorders	Cardiovascular Disorders	Hepatic Disorders	Cardiovascular Disorders	Hepatic Disorders	Cardiovascular Disorders	Hepatic Disorders
XGBoost	0.91	0.988	0.87	0.958	0.87	0.958	0.87	0.959	0.87	0.958
CatBoost	0.88	0.98	0.86	0.938	0.86	0.938	0.86	0.938	0.86	0.938
AdaBoost	0.82	0.921	0.78	0.921	0.78	0.921	0.78	0.921	0.78	0.921

Accuracy comparison for both disorders has been shown in Figure 1.

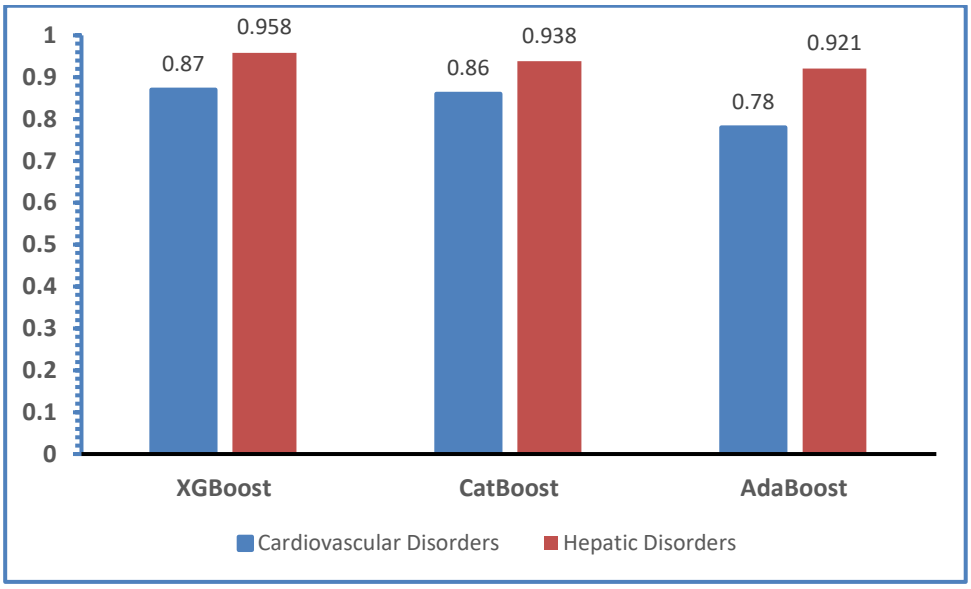

Figure 1: Accuracy Comparison

Based on these results, the following observations can be made:

- When compared to the other two algorithms, XGBoost had the best results in terms of accuracy, precision, recall, F1-score, and area under the curve (AUC) for cardiovascular ailments and hepatic illnesses, respectively.
- With higher results across the board, CatBoost outperformed AdaBoost in the classification of cardiovascular and hepatic illnesses.
- Although all three algorithms performed well, XGBoost achieved very good ratings across the board when it came to hepatic diseases.
- While XGBoost maintained its position as the best algorithm for hepatic illness classification, the algorithms' performance in cardiovascular disorder classification was worse.
- According to the area under the curve (AUC), XGBoost and CatBoost were very good at differentiating between the presence and absence of hepatic and cardiovascular diseases, although AdaBoost was somewhat less effective.

Overall, these results suggest that XGBoost is the most effective algorithm for classifying both cardiovascular and hepatic disorders, followed by CatBoost and then AdaBoost.

CONCLUSION

Three cutting-edge boosting algorithms—XGBoost (Extreme Gradient Boosting), CatBoost, and AdaBoost—were tested in this research to see how well they classified liver and cardiovascular diseases. A large dataset including clinical data and diagnostic results for various diseases was used to test the algorithms. Tabulated in Table 1 are the key points from the performance review. Cardiovascular diseases and hepatic problems each have their own set of published outcomes.

REFERENCES

[1] S. Bhardwaj, S. Jain, N. K. Trivedi, A. Kumar, and R. G. Tiwari, “Intelligent Heart Disease Prediction System Using Data Mining Modeling Techniques,” pp. 881–891, 2022, doi: 10.1007/978-981-19-0707-4_79.

[2] S. Hussain *et al.*, “Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review,” *Biomed Res Int*, vol. 2022, 2022, doi: 10.1155/2022/5164970.

[3] N. K. Trivedi, R. G. Tiwari, A. K. Agarwal, and V. Gautam, “A Detailed Investigation and Analysis of Using Machine Learning Techniques for Thyroid Diagnosis,” *2023* International *Conference on Emerging Smart Computing and Informatics, ESCI 2023*, 2023, doi: 10.1109/ESCI56872.2023.10099542.

[4] N. Ujjwal, A. Singh, A. K. Jain, and R. G. Tiwari, "Exploiting Machine Learning for Lumpy Skin Disease Occurrence Detection," *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2022*, 2022, doi: 10.1109/ICRITO56286.2022.9964656.

[5] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, "Boosting methods for multi-class imbalanced data classification: an experimental review," *J Big Data*, vol. 7, no. 1, pp. 1–47, Dec. 2020, doi: 10.1186/S40537-020-00349-Y/FIGURES/5.

[6] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *ArtifIntell Rev*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/S10462-020-09896-5/METRICS.

[7] P. Bahad and P. Saxena, "Study of AdaBoost and Gradient Boosting Algorithms for Predictive Analytics," pp. 235–244, 2020, doi: 10.1007/978-981-15-0633-8_22.

[8] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 904, pp. 23–37, 1995, doi: 10.1007/3-540-59119-2_166/COVER.

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785.

[10] R. Mitchell, A. Adinets, T. Rao, and E. Frank, "XGBoost: Scalable GPU Accelerated Learning," Jun. 2018, doi: 10.48550/arxiv.1806.11248.

[11] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," Oct. 2018, doi: 10.48550/arxiv.1810.11363.

[12] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *Adv Neural Inf Process Syst*, vol. 31, 2018, Accessed: Nov. 08, 2022. Online.. Available: https://github.com/catboost/catboost

[13] P. Carrier, M. Debette-Gratien, M. Girard, J. Jacques, P. Nubukpo, and V. Loustaud-Ratti, "Liver Illness and Psychiatric Patients," *Hepat Mon*, vol. 16, no. 12, Dec. 2016, doi: 10.5812/HEPATMON.41564.

[14] T. Horvatits, A. Drolz, M. Trauner, and V. Fuhrmann, "Liver Injury and Failure in Critical Illness," *Hepatology*, vol. 70, no. 6, pp. 2204–2215, Dec. 2019, doi: 10.1002/HEP.30824.

[15] C. Peng, A. G. Stewart, O. L. Woodman, R. H. Ritchie, and C. X. Qin, "Non-Alcoholic Steatohepatitis: A Review of Its Mechanism, Models and Medical Treatments," *Front Pharmacol*, vol. 11, p. 603926, Dec. 2020, doi: 10.3389/FPHAR.2020.603926/BIBTEX.

[16] G. Feng *et al.*, "Recompensation in cirrhosis: unravelling the evolving natural history of nonalcoholic fatty liver disease," *Nature Reviews Gastroenterology & Hepatology 2023 21:1*, vol. 21, no. 1, pp. 46–56, Oct. 2023, doi: 10.1038/s41575-023-00846-4.

[17] C. H. Liu and J. H. Kao, "Acute hepatitis C virus infection: clinical update and remaining challenges," *Clin Mol Hepatol*, vol. 29, no. 3, p. 623, Jul. 2023, doi: 10.3350/CMH.2022.0349.

[18] G. Dusheiko, K. Agarwal, and M. K. Maini, "New Approaches to Chronic Hepatitis B," *New England Journal of Medicine*, vol. 388, no. 1, pp. 55–69, Jan. 2023, doi: 10.1056/NEJMRA2211764/SUPPL_FILE/NEJMRA2211764_DISCLOSURES.PDF.

[19] R. G. Tiwari, D. S. Yadav, and A. Misra, "Performance Evaluation of Optimizers in the Classification of Marble Surface Quality Using CNN," pp. 181–191, 2023, doi: 10.1007/978-981-19-3148-2_15/COVER.

[20] N. K. Trivedi, V. Gautam, H. Sharma, A. Anand, and S. Agarwal, "Diabetes Prediction using Different Machine Learning Techniques," *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022*, pp. 2173–2177, 2022, doi: 10.1109/ICACITE53722.2022.9823640.

[21] J. Carrillo-Moreno, C. Pérez-Gandía, R. Sendra-Arranz, G. García-Sáez, M. E. Hernando, and Á. Gutiérrez, "Long short-term memory neural network for glucose prediction," *Neural Comput Appl*, vol. 33, no. 9, pp. 4191–4203, May 2021, doi: 10.1007/S00521-020-05248-0/METRICS.

[22] E. Dritsas and M. Trigka, "Machine Learning Methods for Hypercholesterolemia Long-Term Risk Prediction," *Sensors 2022, Vol. 22, Page 5365*, vol. 22, no. 14, p. 5365, Jul. 2022, doi: 10.3390/S22145365.

[23] M. H. Hung *et al.*, "Prediction of Masked Hypertension and Masked Uncontrolled Hypertension Using Machine Learning," *Front Cardiovasc Med*, vol. 8, p. 778306, Nov. 2021, doi: 10.3389/FCVM.2021.778306/BIBTEX.

[24] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for

detection of COVID-19 cases from chest X-ray images," *Scientific Reports 2020 10:1*, vol. 10, no. 1, pp. 1–12, Nov. 2020, doi: 10.1038/s41598-020-76550-z.

[25] D. Srivastava *et al.*, "Early Detection of Lung Nodules Using a Revolutionized Deep Learning Model," *Diagnostics 2023, Vol. 13, Page 3485*, vol. 13, no. 22, p. 3485, Nov. 2023, doi: 10.3390/DIAGNOSTICS13223485.

[26] J. Azmi, M. Arif, M. T. Nafis, M. A. Alam, S. Tanweer, and G. Wang, "A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data," *Med Eng Phys*, vol. 105, p. 103825, Jul. 2022, doi: 10.1016/J.MEDENGPHY.2022.103825.

[27] "Non-alcohol fatty liver disease (NAFLD)." Accessed: Mar. 04, 2024. Online.. Available: https://www.kaggle.com/datasets/utkarshx27/non-alcohol-fatty-liver-disease/data

[28] "Heart Disease prediction | Kaggle." Accessed: Jan. 16, 2023. Online.. Available: https://www.kaggle.com/datasets/mdriponmiah/heard-disease-prediction