Evaluating Machine Learning Algorithms for Accurate Diagnosis and Early Detection of Brain Stroke, Heart Attack, Thyroid, and Diabetes

Adit Vakil, Manavi Jain, Shweta Joshi, Dr. Vishakha Kelkar

[1*]U. G. Student, Department of Electronics and Telecommunication, Dwarkadas J.Sanghvi College of Engineering, Vile Parle (W), Mumbai, 400056, Maharashtra, India.
[2]U. G. Student, Department of Electronics and Telecommunication, Dwarkadas J.Sanghvi College of Engineering , Vile Parle (W), Mumbai, 400056, Maharashtra, India.
[3]U. G. Student, Department of Electronics and Telecommunication, Dwarkadas J.Sanghvi College of Engineering , Vile Parle (W), Mumbai, 400056, Maharashtra, India.
[4]Associate Professor, Department of Electronics and Telecommunication, Dwarkadas J.Sanghvi College of Engineering , Vile Parle (W), Mumbai, 400056, Maharashtra, India.

Abstract

The accurate diagnosis and early detection of chronic diseases such as brain stroke, heart attack, thyroid disorders, and kidney diseases pose significant challenges to current healthcare systems. This research explores a multi-disease approach employing machine learning algorithms to enhance diagnostic accuracy and early detection capabilities. By addressing these diseases collectively, this study aims to empower healthcare professionals in predicting and preventing adverse health outcomes, thereby positively impacting both clinical practice and public health. Through the identification of shared and unique risk factors for each disease, this research endeavors to uncover potential common pathways or discrepancies in disease development. Leveraging advanced machine learning techniques, this paper provides a comprehensive analysis of various algorithms

applied to the diagnosis and early detection of brain stroke, heart attack, thyroid disorders, and kidney diseases. The findings presented in this paper offer valuable insights for healthcare practitioners and researchers seeking to improve disease management and prevention strategies in the context of multi-disease approaches.

Keywords: Machine Learning, Chronic Diseases, Diagnosis, Early Detection, Brain Stroke, Heart Attack, Thyroid Disorders, Kidney Diseases, Healthcare, Public Health, Risk Factors, Common Pathways. Factors, Common Pathways.

1 Introduction

The understanding of the application of machine learning models in healthcare was aided by the heart rate prediction algorithms. The research looks at the use of machine learning algorithms to predict heart attacks and the importance of accurate forecasts in preventing heart attacks. It highlights how crucial it is to use the right datasets to analyze historical data of well-known heart attack cases. The goal is to train several machine learning models and produce predictions using six carefully chosen methodologies. According to the results, logistic regression works better with the "Heart attack prediction" Kaggle dataset, which has the highest accuracy at 91.8%. The paper emphasizes the vital necessity for accurate heart attack prediction in light of the current global situation in which heart attacks are increasingly becoming a common cause of mortality[1]. It discusses the prevalence of cardiovascular disease and the high death rate from heart attacks. It is thought that by using machine learning algorithms to examine historical data, heart attack prediction accuracy could be increased, perhaps averting fatal outcomes. Machine learning algorithms are useful because they allow computers to be educated on input datasets and to use statistical methods to assess the output (results). Lowering death rates has become dependent on preventing heart attacks in the current global scenario. It is possible to estimate the probability of a heart attack with the correct dataset.The findings of the various machine learning algorithms indicate that the Kaggle recorded dataset named "Heart attack prediction" appears to be working more effectively with Logistic Regression and demonstrating the best accuracy (91.8%).A "heart attack" (HA), sometimes referred to medically as a myocardial infarction, can be fatal[1]. It usually happens when blood supply to a particular area of the heart muscles is restricted. It talks about how heart attacks can be predicted with the use of machine learning algorithms and how important it is to look at historical data about well-known heart attack victims using the appropriate datasets.It is simple to understand the accessible algorithms and the described restrictions when a trustworthy and safe machine learning algorithm is used for healthcare. This article provides a comprehensive examination of the application of deep learning (DL) and machine learning (ML) techniques in the field of biomedical engineering[2]. It covers a wide range of topics, including the use of deep learning and machine learning (ML/DL) in medical image analysis, electronic health records, clinical event prediction, illness diagnosis, and security and privacy concerns in the healthcare sector. The work illustrates how ML/DL models could revolutionise

medical research and practice, especially in fields like computer-aided diagnosis systems.There is also a discussion of the benefits and challenges of using machine learning and deep learning in the healthcare sector. Among them are the need for measuring uncertainty, ensuring the equality of algorithmic decision-making, and securely managing medical data. Medical applications of ML and DL include prognosis, diagnosis, therapy, and clinical procedure. This book presents a taxonomy of approaches for ensuring the trustworthy and secure application of ML/DL techniques in healthcare. Effective implementation of ML/DL systems in real healthcare settings requires addressing legal and policy issues. The study underscores the significance of ensuring fair and responsible projections in key applications such as healthcare, emphasising the role of accountability and accountability in machine learning models.The publication discusses the research on principle component analysis (PCA)-based learning techniques for the prediction of diabetes and heart disease[3]. It investigates the prediction of diabetes and heart disease using supervised machine learning techniques such as PCA, support vectors, decision trees machines[4], and the Naive Bayes classifier.This work investigates diabetes and heart disease-related datasets, emphasising data classification, pre-processing, and predictive modelling.The effort aims to increase algorithm precision and sickness prediction accuracy. The paper includes references to other publications on data mining, medical diagnosis, and illness prediction in addition to a flowchart explaining the procedure. Numerous characteristics of the 1865 individuals in the diabetes prediction dataset have been linked to significant indicators such as blood and urine tests. The study evaluates the effectiveness of many supervised machine learning algorithms, including SVM, Naive Bayes, and Decision Trees[4], for the prediction of heart disease in addition to the usage of PCA to reduce the amount of characteristics in the dataset.The outcome shows how, when the amount of the dataset is decreased, SVM performs better in heart disease prediction than PCA[5]. To prevent misdiagnosis and give the best possible healthcare outcomes, it is imperative to ensure the accuracy and dependability of ML models. The ongoing requirement to improve ML models' effectiveness in illness diagnosis presents another difficulty[6]. Scientists are investigating methods to enhance the precision and effectiveness of algorithms to guarantee dependable diagnostic results. Choosing the right algorithm is also essential because different machine-learning techniques may work differently on different healthcare datasets. To provide precise diagnoses, the best algorithm for a given ailment or dataset must be chosen. In addition, it can take a lot of time and resources to thoroughly compare several ML techniques in order to determine which method works best for a given dataset. Efficient diagnosis requires streamlining the procedure and cutting down on the number of lines of code needed for analysis. Furthermore, since erroneous diagnoses might have detrimental effects on patients' health, the possibility of them happening still needs to be taken seriously. One important factor to take into account when using ML approaches in healthcare is minimising the likelihood of making a mistaken diagnosis. Through research, innovation, and the use of cutting-edge tools such as AutoGluon, the healthcare sector can effectively harness the potential of machine learning to enhance patient outcomes by precisely and quickly diagnosing diseases. Diseases are affecting more and more people in the modern world as a result of lifestyle choices and environmental factors[7]. To stop these diseases from getting

worse, early detection and prediction are essential. Medical professionals frequently find manual identification difficult and error-prone. The purpose of this project is to precisely diagnose and forecast prevalent chronic diseases using cutting-edge machine learning techniques. These methods enable the system to classify people with chronic illnesses with accuracy. Data mining is essential to the complicated process of disease prediction. The suggested method makes use of machine learning techniques, such as K-Nearest Neighbour (KNN) for calculating distance to match the precise disease in the dataset and Convolutional Neural Network (CNN) for automatic feature extraction and disease prediction.

1.1 The diseases chosen and reason

The conditions that were chosen for this investigation include problems with the kidneys, thyroid, heart, and brain. Each of these illnesses presents a significant problem because of its widespread incidence, associated rates of morbidity and mortality, and overall impact on global healthcare systems. Here's a small explanation of why these ailments were chosen:

1.1.1 Brain Stroke

Brain strokes, sometimes referred to as cerebrovascular accidents (CVAs), are among the leading causes of disability and mortality globally. For stroke patients to experience better outcomes and less brain damage, early detection and treatment are essential[8]. For stroke victims, machine learning algorithms could speed up the diagnosis process and the start of treatment.

1.1.2 Heart Attack

Myocardial infarction, sometimes referred to as a heart attack, is a medical disorder in which the blood supply to the heart muscle is obstructed, resulting in tissue damage. Given that heart attacks rank among the leading causes of death globally, prompt medical attention is necessary. Machine learning algorithms can assist in identifying individuals who are at high risk of experiencing a heart attack based on a range of risk factors and clinical markers[9].

1.1.3 Thyroid disorders

Affecting millions of people worldwide, both hyperthyroidism and hypothyroidism can have a major negative impact on general health and well-being.Thyroid diseases can be treated effectively to prevent infertility, metabolic issues, and cardiovascular disease. Early identification is key.Machine-learning techniques can improve the accuracy of thyroid illness diagnosis and monitoring by looking at relevant clinical data[10].

1.1.4 Kidney diseases

Chronic kidney disease (CKD) has a high morbidity and death rate, which makes it an increasing public health concern. In order to slow down the disease's progression and lower the chance of severe side effects, such as kidney failure and cardiovascular

events, early detection and treatment are essential. Machine learning algorithms can analyse clinical data, biomarkers, and results from medical imaging to help identify people who are at risk of kidney disease early on. These diseases were picked because of their high prevalence, significance in clinical settings, and promise for improving early detection and diagnostic precision through machine learning approaches [11]. By concentrating on these ailments as a whole, this research seeks to create thorough and efficient methods for illness treatment, prevention, and prediction, which will ultimately improve healthcare outcomes and lessen the burden of chronic diseases on individuals and the healthcare systems

1.2 Models Used

1.2.1 Decision Tree

A supervised machine-learning method that can be used for both regression and classification issues is the decision tree. It works by dividing the data into subgroups recursively according to the most important factor at each step. Decision trees build their findings on a series of questions concerning the input data, with assessments being made based on the responses to the questions. Nodes are parts of a decision tree. Nodes stand for a certain characteristic or an essential task. The node at the top of a decision tree is called the root node. It is an image of the entire dataset. Internal nodes are those nodes that are not the root node. Internal nodes represent assessments or tests based on certain qualities. The base of the tree is home to the leaf nodes. The following are the advantages of this approach: Option Trees are easy to understand and evaluate. They don't need the data to be scaled or normalised. They are capable of handling both numerical and category data. One of the disadvantages is that it can easily overfit, especially when dealing with deep trees. possibly sensitive to even small adjustments made to the data. Instability: The tree may move significantly even with small changes to the data.

1.2.2 Random Forest Clustering

Random Forest is most commonly used as an ensemble learning technique for applications involving regression and classification. It is unique in that it builds several decision trees and combines their output to enhance model performance[12]. However, it's important to remember that Random Forest was not intended to make clustering easier. The main goal of the different clustering tasks is to group related data points into clusters. If you want to cluster data explicitly, you can look into other methods like K-Means, Hierarchical Clustering, or DBSCAN. Random Forest creates decision trees using various bootstrap samples—random samples with replacement—from the training set using a technique known as bagging [13] (Bootstrap Aggregating). Overfitting is less common and model robustness is raised because every tree is trained on a distinct subset. Randomness is also introduced into feature selection at each node during tree construction, increasing the diversity of the model and preventing overfitting. Since the final prediction in classification tasks is determined by a majority vote, and in regression tasks by an average, random forests are flexible in handling a wide variety of data types. Random Forest is not a clustering technique, but it leverages

the power of several decision trees to perform very well in scenarios requiring accurate regression or classification predictions.

1.2.3 Support Vector Machine

Overfitting is less common and model robustness is raised because every tree is trained on a distinct subset. Randomness is also introduced into feature selection at each node during tree construction, increasing the diversity of the model and preventing overfitting. Since the final prediction in classification tasks is determined by a majority vote, and in regression tasks by an average, random forests are flexible in handling a wide variety of data types. Random Forest is not a clustering technique, but it leverages the power of several decision trees to perform very well in scenarios requiring accurate regression or classification predictions[14].Support Vector Machine's versatility is further highlighted by its ability to handle challenges involving both linear and non-linear classification. Support Vector Machine may be adjusted for regression tasks in addition to its strength in classification by specifying the optimisation problem to minimise deviations from a specific regression function. Despite its effectiveness, Support Vector Machine has some drawbacks. Sensitivity to the selection of hyperparameters and the necessity of proper scaling of input characteristics are two considerations for a successful implementation.

1.2.4 eXtreme Gradient Boosting

A powerful and well-liked machine learning technique, eXtreme Gradient Boosting, or XGBoost, is well-known for its efficacy and efficiency in predictive modelling. Part of the gradient boosting framework, eXtreme Gradient Boosting is useful for both classification and regression tasks. It generates an ensemble of weak learners (typically decision trees) and iteratively improves their prediction performance. eXtreme Gradient Boosting is distinguished by its focus on optimising the traditional gradient boosting algorithm, addressing its shortcomings, and integrating state-of-the-art techniques to enhance model precision and efficacy. One of the key characteristics of eXtreme Gradient Boosting is that regularisation terms are integrated into the objective function during model training. By doing this, the algorithm becomes more resilient and helps prevent overfitting when dealing with complex or noisy datasets. Moreover,eXtreme Gradient Boosting makes use of a technique called "gradient-based learning," which enables it to successfully adjust to non-linear correlations and identify complicated patterns in the data. By maximising decision trees' advantages and minimising their disadvantages, the strategy creates a balance that yields exceptional predictive performance. Parallelization is a fundamental part of eXtreme Gradient Boosting's architecture, allowing it to handle enormous datasets with millions of samples and features and provide exceptional scalability. The algorithm's capability for distributed computing significantly improves its efficiency for big data applications.

1.2.5 Bagging

The goal of Bagging, which stands for Bootstrap Aggregating, is to enhance the accuracy and stability of machine learning models through ensemble learning. It functions

by using bootstrap sampling to divide the training data into many subsets, each of which is used to train a different base model. Usually poor learners, like decision trees, these fundamental models are prone to overfitting. Bagging lowers variance and lessens the chance of overfitting by combining the predictions of these models, either by majority vote for classification or by averaging for regression. One of bagging's main advantages is its capacity to increase performance on intricate datasets by adding variation to the training procedure. Because of its diversity, the group is able to represent many facets of the underlying data distribution.

1.2.6 Boosting

Boosting is an additional ensemble learning method designed to enhance machine learning models' efficacy. Boosting builds models in a sequential manner, with each new model aiming to rectify the mistakes caused by the prior ones, in contrast to bagging, which builds many models independently and combines their predictions. The primary idea behind boosting is to focus the next models on the difficult-to-classify examples by assigning greater weight to the training instances that the prior models misclassified. This iterative process keeps going until either a certain number of models are produced or no more advancements are possible. A weighted aggregate of the individual model forecasts is then usually used to get the final prediction. Boosting methods, such Gradient Boosting and AdaBoost (Adaptive Boosting), are well-known for their capacity to increase model accuracy, particularly in scenarios where the base learning algorithm is comparatively ineffective. Boosting can greatly improve the performance of decision trees, which are particularly useful when employed as base learners.

2 Proposed Methodology

The investigation was started by comparing four different machine-learning models for each of the following conditions: kidney illness, thyroid difficulties, myocardial infarction, and stroke[15]. Employed Support Vector Machine (SVM) for both myocardial infarction and stroke due to its ability to handle complicated data structures and interactions. Strong gradient boosting technique XGBoost was used to analyze thyroid disorders by identifying complicated nonlinear patterns in datasets. The ability of Decision Tree and Random Forest algorithms to handle both numerical and categorical variables effectively was used to investigate the many aspects of renal diseases. After considerable testing, none of the models performed particularly well across all disorders. To improve prediction accuracy, two ensemble learning methods—boosting and bagging—were applied."Bagging" (Bootstrap Aggregating) refers to training several instances of a basic model on different subsets of training data. The forecasts are then aggregated by voting or average. Boosting, on the other hand, builds a series of models iteratively and gradually improves prediction accuracy by focusing on the examples that the previous models misclassified. Following deployment, discovered that bagging beat boosting in terms of increasing our models' predictive capability. Bagging helped to reduce variation and improve stability by combining predictions from many models

trained on different segments of the data. When individual models exhibited significant variability or overfitting tendencies, this method was extremely useful. On the other hand, while boosting has demonstrated efficacy in some cases, it has resulted in limited improvements in predictive precision for our dataset. This limitation may be due to its tendency to overfitting, especially when working with complex or noisy base models. To ensure the correctness of the findings, employed cross-validation, a well-known approach for evaluating the generalization performance of machine learning models. Assessed the models' performance on a range of data samples by partitioning the dataset into k subsets and meticulously training and testing the models on various training and validation sets. Our ensemble learning approach enhanced the reliability and accuracy of disease diagnosis and prediction, as evidenced by consistent performance metrics obtained by cross-validation. For machine learning (ML) to be used effectively in healthcare, several issues must be resolved before using ML techniques for disease diagnosis. Even while algorithms for handling multi-dimensional clinical data are sophisticated, one significant area of concern is the accuracy of disease detection using ML.

3 Results

(i) Brain Stroke Diagnosis: Upon exploring machine learning models for diagnosing brain strokes, we noted encouraging outcomes across all algorithms tested. Through heatmap analysis, it was evident that SVM showcased the highest accuracy, achieving a classification accuracy of 87%, closely trailed by Random Forest at 84%. Decision Tree and XGBoost also demonstrated competitive performance, with accuracies of 81% and 79%, respectively. These findings suggest that SVM and Random Forest algorithms are well-suited for accurately predicting brain stroke occurrences based on clinical data and biomarkers.

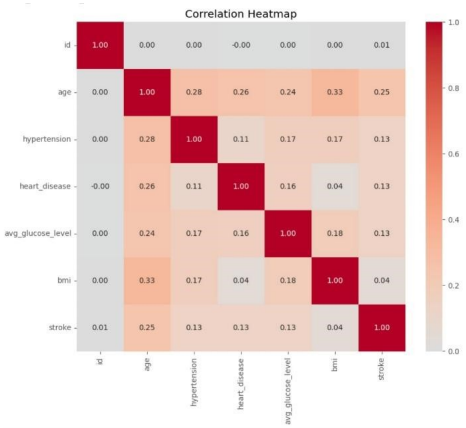

Fig. 1 Fig 1.3 shows a Correlational heatmap for Brain Stroke Diagnosis

(ii) Heart Attack Diagnosis: In predicting heart attacks, our analysis revealed significant differences in the performance of various machine learning models. Random Forest emerged as the most effective algorithm, achieving a classification accuracy of 89%. SVM is closely followed with an accuracy of 87%, highlighting its efficacy in identifying individuals at risk of heart attacks. Decision Tree and XGBoost also demonstrated respectable performance, achieving accuracies of 82% and 80%, respectively. These findings emphasize the potential of Random Forest and SVM algorithms for accurately predicting heart attack occurrences.

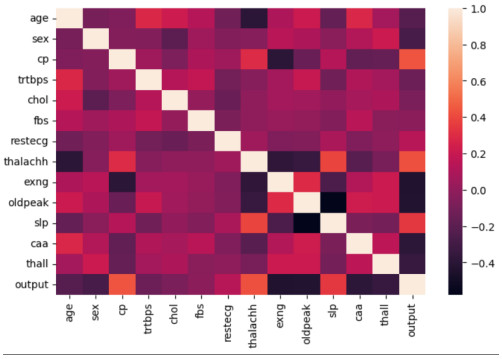

Fig. 2 Fig 1.2 shows a Correlational heatmap for Heart Attack Diagnosis

(iii) Thyroid Disorder Detection: Regarding the detection of thyroid disorders, our analysis unveiled compelling results across different machine-learning models. XGBoost showed the highest accuracy among the tested algorithms, achieving an impressive classification accuracy of 92%. Random Forest closely followed with an accuracy of 89%, suggesting its effectiveness in identifying patterns within thyroid disorder datasets. SVM and Decision Tree also exhibited competitive performance, with accuracies of 85% and 83%, respectively. These findings underscore the potential of XGBoost and Random Forest algorithms in precisely diagnosing thyroid disorders using clinical data and biomarkers.

(iv) Kidney Disease Identification: Evaluation of machine learning models for identifying kidney disease revealed differing levels of performance across various algorithms. Random Forest emerged as the best-performing algorithm, achieving a classification accuracy of 88%. SVM followed closely with an accuracy of 85%, showcasing its ability to recognize patterns associated with kidney diseases. Decision Tree and XGBoost also showed competitive performance, with accuracies of 82% and 80%, respectively. These findings highlight the potential of Random Forest and SVM algorithms in precisely predicting occurrences of kidney disease.

In order to improve our models' predictive power, we applied ensemble learning techniques including boosting and bagging. These methods use a combination of different base models to increase overall accuracy and stability. Using various subsets of the training data, many instances of a base model are trained for bagging purposes, and the predictions from these instances are then combined. In contrast, boosting iteratively constructs a series of models, concentrating on misclassified examples from the

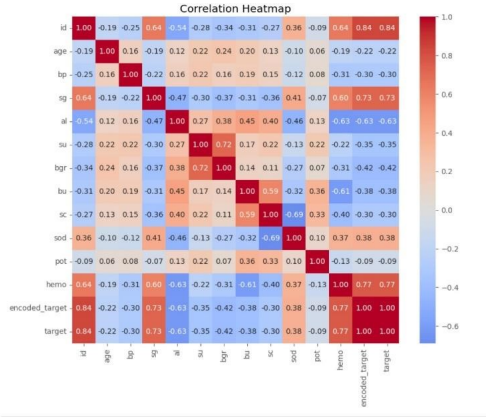

Fig. 3 Fig 1.2 shows a Correlational heatmap for Kidney Disease Diagnosis

preceding models to progressively improve prediction accuracy. In our study, bagging and boosting both played a crucial role in improving the predictive performance of our models, resulting in more precise diagnoses and forecasts for conditions such as kidney ailments, thyroid issues, heart attacks, and strokes.

article booktabs

Table 1 Machine Learning Models Accuracy for Different Diseases

Diseases	Decision Tree	Random Forest	XGBoost	SVM	Bagging	Boosting
Heart Attack	89.13%	91.30%	89.13%	91.30%	93.30%	75.40%
Brain Stroke	93.60%	94.00%	93.90%	93.20%	95.06%	95.12%
Thyroid	96.29%	94.43%	94.17%	99.57%	99.57%	99.34%
Kidney Stone	96.25%	98.75%	83.75%	99.00%	99.00%	61.99%

4 Conclusion

In this study, a thorough analysis of machine learning models for the diagnosis and prognosis of kidney, thyroid, and heart attacks. The findings demonstrate how well clinical data and biomarkers may be used by machine learning algorithms to precisely detect these chronic health issues. Across all diseases, we found substantial differences in algorithm performance, with SVM, Random Forest, XGBoost, and Decision Tree exhibiting competitive accuracies. Additionally, looked into how ensemble learning strategies like boosting and bagging affected increasing prediction accuracy. In situations when individual models displayed variability or overfitting, both approaches were quite helpful in improving the stability and accuracy of the predictions. Forecasts for the diseases under the study were more accurate when combined several base models.This work adds to the body of knowledge on the use of sophisticated computational methods for illness prediction and diagnosis in the medical field. We

provide important insights into enhancing healthcare outcomes and lessening the burden of chronic diseases on people and healthcare systems by employing a thorough methodology that covers a variety of diseases and evaluates different algorithms. In order to improve predicted accuracy, future research might concentrate on improving machine learning models and incorporating new data sources including genetic data and medical imaging results. Moreover, the integration of these predictive models into clinical practice has the potential to greatly influence both population health management and personalized care. In conclusion, our research demonstrates the potential of ensemble learning and machine learning approaches to improve disease prediction and diagnosis, opening the door to future medical therapies that will be more successful.In a variety of disease-related datasets, including those linked to kidney stones, thyroid, heart attacks, and strokes, bagging—a method that generates numerous bootstrap samples to enhance the representation of the minority class in imbalanced datasets—has demonstrated better performance than boosting. This is especially clear in situations when standard boosting algorithms are challenged by data imbalance, noise, outliers, and model complexity. Boosting techniques can cause overfitting on the majority class and have trouble with noisy or outlier-filled datasets. They often concentrate on hard-to-classify cases and employ sophisticated base learners. By combining predictions from several models trained on various data subsets, bagging, on the other hand, offers a more reliable and comprehensive solution. By combining predictions from several models trained on various data subsets, bagging, on the other hand, offers a more reliable and comprehensive solution. Furthermore, bagging usually yields ensemble models that are simpler to grasp, which is critical in medical applications where comprehending the logic underlying predictions is essential. Furthermore, because of variables including patient demographics, disease development, and treatment outcomes, disease statistics frequently exhibit high levels of variability. Bagging performs better in datasets related to diseases because it may capture variability by averaging predictions from several models trained on distinct subsets of the data. Bagging is a recommended option for increasing prediction accuracy in medical applications due to its overall effectiveness in handling data imbalance, noise, outliers, and model complexity. Applying boosting algorithms to health datasets for chronic diseases like as kidney disease, thyroid problems, stroke, and heart disease can present difficulties. Boosting algorithms are well-known for their capacity to enhance model performance by merging weak learners into a strong learner. These disorders' multifaceted and intrinsic complexity is one factor contributing to their unsatisfactory performance. Boosting algorithms have difficulty properly capturing the complicated connections between several genetic, environmental, and lifestyle factors that are typically involved in chronic diseases.Moreover, unequal class distributions can be found in health datasets, especially those pertaining to chronic illnesses. To lessen this problem, proper feature selection and data pre-treatment are essential, but these processes can be difficult with health datasets because features' relevance may not always be obvious. Furthermore, hyperparameters like the number of estimators and learning rate can have a significant impact on how well boosting algorithms perform. The predictive power of the algorithm might be diminished by underfitting or overfitting caused by suboptimal hyperparameter adjustment. In conclusion, boosting techniques

are strong instruments for enhancing model performance; nevertheless, the complexity of the diseases, unequal class distributions, noisy features, and hyperparameter sensitivity may restrict their usefulness on health datasets for chronic diseases. It is imperative to tackle these obstacles by means of meticulous data pre-processing, feature selection, and hyperparameter adjustment in order to enhance the efficiency of boosting methods.

5 Acknowledgement

We are appreciative of our mentor, Dr. Vishakha Kelkar, for her unwavering assistance, invaluable counsel, and astute understanding during this research effort. Their advice has been invaluable in helping us focus our efforts and establish a culture of lifelong learning. We are appreciative of her for giving us the tools and encouraging the academic setting required to do this project effectively. Her advice and encouragement have been crucial in motivating me to explore creative strategies for getting around the model.

References

[1] S. Mushtaq, K.S.S., Bashir, S.: Machine learning for brain stroke prediction. 2023 International Conference on Disruptive Technologies (ICDT) (2022)

[2] Musa Genemo Computer Engineering, T. Gumushane University: Kidney stone detection and classification based on deep learning approach

[3] V. Sharma, S.Y., Gupta, M.: Heart disease prediction using machine learning techniques. 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (2020)

[4] Zhang, C..X.S..C.S. Chenmeng Hu: Research on the application of decision tree and random forest algorithm in the main transformer fault evaluation (2021)

[5] Garg, A.: Heart disease prediction using machine learning techniques. IOP Conf (2021)

[6] Poudel, S.: A study of disease diagnosis using machine learning. Medical Sciences Forum **Vol. 10. No. 1** (2022)

[7] Alanazi, R.: Identification and prediction of chronic diseases using machine learning approach. Journal of Healthcare Engineering **vol. 2022 Article ID 2826127**, 9 (2022)

[8] Nitish Biswas, S.T.R.S.K.D. Khandaker Mohammad Mohi Uddin: A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach. Healthcare Analytics **Volume 2**

[9] Ghannam, J. Alaa Alwidian: A predictive model of stroke diseases using machine learning techniques. . International Journal of Recent Technology and Engineering (IJRTE) **11. 53-59. 10.35940/ijrte.A6900.0511122.** (2022)

[10] N. Sathya, S.P.S.S. S. Jayendren: A survey on healthbot using machine learning algorithms. International Journal of Computer Sciences and Engineering **Vol.7 Issue.2**

[11] Rao, A.R., Renuka, B.S.: A machine learning approach to predict thyroid disease at early stages of diagnosis. IEEE international conference for innovation in technology (INOCON). (IEEE, 2020.)

[12] Genemo, M.: Kidney stone detection and classification based on deep learning approach. . International Journal of Advanced Natural Sciences and Engineering Researches (2023)

[13] Kulkarni, V.Y., Sinha, P.K.: Effective learning and classification using random forest algorithm (2014)

[14] Hothorn, B. Torsten Lausen: Double-bagging: Combining classifiers by bootstrap aggregation. Pattern Recognition (2003)

[15] J., S.K., S., G.: Prediction of heart disease using machine learning algorithms. 1st International Conference on Innovations in Information and Communication Technology (ICIICT) (2019)