

A Comparative Study of Machine Learning Models for Heart Disease Prediction Using Grid Search and Random Search for Hyperparameter Tuning

Suman Arshad¹, Syed Muhammad Junaid Zaidi¹, Muazzam Ali^{1*}, M U Hashmi¹, Abdul Manan¹, and Affan Ahmad²

¹Department of Basic Sciences, Superior University, Lahore, 54000, Pakistan.

²Department of Computer Science, Superior University, Lahore, 54000, Pakistan.

*Corresponding Author: Muazzam Ali. Email: muazzamali@superior.edu.pk

Received: April 11, 2024 Accepted: September 28, 2024

Abstract: An important global health concern is heart failure, for which early detection can greatly improve patient outcomes. Machine learning has proved to be useful in predicting the likelihood of heart disease by looking at factors like age, high blood pressure, and cholesterol. This study compares popular machine learning models, such as Random Forests, Gradient Boosting, Stacking, KNN, SVM, and Logistic Regression. We utilized a Grid Search as well as Random Search to improve the models' efficiency and perform-ability. Following model tuning, the models were determined using metrics like accuracy, recall, F1 score, and precision, AUC, Cohen's Kappa, and MCC. With grid search accuracy of 94.95% and random search accuracy of 94.54%, Random Forest produced the best results. This highlights how important it is to select the right model and adjust its parameters for the best results, and it also shows how well Random Forest predicts heart disease.

Key words: Hyperparameters Modification; Random Forests; Machine Learning; Medical Testing and Diagnostics; Heartbeat Predicting; Ensembles Methods.

1. Introduction

Heart attack is the biggest cause of death worldwide, accounting for several new cases and deaths each year. The World Health Organization, reports the 31% of all deaths globally are related to cardiovascular disorders (CVDs), with 17.9 million deaths being caused by these conditions. Heart disease is the least common type of CVD, and early recognition is essential to decreasing its consequences. Early detection permits for on time healing procedures that may improve affected person outcomes and prevent principal headaches. Heart disorder has historically been diagnosed the use of processes like cardiograms, electrocardiograms and the know-how of clinical professionals [1, 2]. These methods, while powerful, can be pricey, time-consuming, and aid-extensive, especially in regions with constrained access to healthcare. The prediction of heart sickness is converting because of gadget gaining knowledge of (ML) and artificial intelligence (AI) [3]. Healthcare vendors can use those technologies to forecast affected person results and pick out developments via examining patient records. Algorithms for machine studying are able to locate complicated relationships among multiple variables that conventional statistics ought to have ignored, like affected person demographics and medical records. For this reason, they come in in particular on hand while calculating the danger of strokes the use of large records units [4-6].

Especially within the discipline of coronary heart disease prediction, system getting to know has considerably modified healthcare over the past few many years. There are many models used, from simple methods like Logistic Regression to more complex strategies like Random Forests and Gradient Boosting. The treatment evaluation, outcome for affected person's prediction, and hazard thing identity are supported with

the aid of these fashions [7]. The simplicity and understandability of Logistic Regression make it a famous version, however greater complicated fashions, including Random Forest and Support Vector Machines (SVM), are preferred for dealing with complex information and connections among variables. Studies show that ensemble techniques, like Random Forest and Gradient Boosting, are often extra accurate than separate fashions in predicting heart disorder. Gradient Boosting came in 2nd with 87% accuracy, and Random Forest came in first with 89%, [8]. Through the mixing of predictions from a couple of models, ensemble getting to know can help increase accuracy. However, the performance of these fashions relies upon on the choice of the right hyperparameters, so making the vital changes to the hyperparameters will yield the exceptional consequences [9-11]

Grid Search is a not unusual approach for tuning hyper Parameters as it assessments each viable aggregate of the parameters in a predefined grid. While it unearths the surest hyperparameters, it's far computationally pricey, especially for big amounts of information and complex models. Selecting samples at random to more than a few hyper parameters is the extra efficient and faster choice called Random Search. When efficiency is extensively inspired by way of a limited set of parameters, it features in particular nicely. To expect heart disease, machine getting to know is being applied in multiple research [12-14]. In [15] ensemble methods to acquire favorable consequences the use of a variety of type algorithms, inclusive of Naive Bayes in addition to Decision Trees. Random Forest as well as Gradient Boosting executed higher than different algorithms. However, many research focused on the whole on accuracy, neglecting precision and keep in mind, two important healthcare metrics which can be vital to avoid overlooked diagnoses [16]. The advantages of collaborative methods in healthcare were emphasized [17-19]. XGBoost, Random Forest, and Adaboost had been some of the classifiers they in comparison. They found that ensemble methods improve predictive performance by using the strengths of each individual classifier. Comparing different hyperparameter tuning methods for machine learning models, like Grid Search as well as Random Search, hasn't been extensively researched, but. We compare the efficacy of both tuning methods for heart condition prediction using various classifiers in an effort to close this gap [20].

One significant area of unresolved research need is the lack of comparative studies between different hyperparameter tuning techniques (e.g., Grid Search and Random Search) over machine learning models utilized in heart disease prediction. Most studies focus on improving specific algorithms without taking the possible effects of different tuning strategies into account. Moreover, while accuracy is generally disclosed, other important metrics precision, recall, F1 score, AUC, including Kappa and MCC are often overlooked. For a comprehensive evaluation of the model's performance, these metrics are crucial, especially in the healthcare sector where error rates and incorrect outcomes can significantly impact patient care. With a focus on tuning hyper parameters using Grid Search as well as Random Search, we compare various machine learning classification techniques for predicting the likelihood of heart disease in order to bridge this research gap. We evaluate SVM, Random Forest, Gradient Boosting, Stacking, K-Nearest Neighbors (KNN), and Logistic Regression classifiers. We examine those fashions with more than a few performance metrics in an effort to supply a complete evaluation of their predictive strength.

2. Methodology

2.1. Dataset

The "heart_statlog_cleveland_hungary_final.Csv" dataset become utilized on this examine. To predict heart disease, this popular dataset is extensively used. The affected person's age, gender, blood strain, cholesterol, and blood pressure are only a few of the traits that would be coronary heart sickness danger elements. It is indicated by using the target variable whether or not heart disease is gift or not. With its aggregate of non-stop and express facts, the dataset is like minded with an extensive variety of device getting to know models. It additionally incorporates facts from multiple sources to offer a number of patient traits that beautify the generalizability of the predictive fashions. The dataset contains a number of features that are relevant to the prediction of heart disease. Maximum heart rate and blood pressure are continuous and vary with time, but age and cholesterol are constants. Categorical variables include exercise-induced angina, sex,

fasting blood sugar levels, resting ECG readings, and the ST segment's slope. ST depression is once another variable that is constant. It is crucial to choose a model that can manage this combination and correctly predict heart disease because its characteristics include both continuous and categorical data.

2.2. Preprocessing the Data

Data preprocessing is necessary when applying machine learning algorithms in order to maximize model performance. We focused on growing and standardized continuous features in this work while keeping categorical variables unaffected.

2.2.1. Establishing a Standard

For standardizing continuous variables like age, cholesterol levels, and resting blood pressure, the learning StandardScaler was employed. The standard deviation and mean of each feature are set to one and zero, respectively, as result of standardization. This is crucial for models that are sensitive to input scale, like support vector machines (SVM) and K-Nearest Neighbors (KNN). Without standardization, smaller features could be hidden by larger numerical range features, which would distort the outcomes.

2.2.2. Classification of Variables

Quantitative features were utilized to encode categorical features such as sex, resting ECG, and fasting blood sugar levels. Scaling is unnecessary for categorical data, so these variables (male/female, normal/abnormal, etc.) were left in their original representations without scaling.

2.3. Tuning the Hyperparameters

Tuning Hyperparameters is fundamental to maximize the performance of machine learning model. Hyperparameters are crucial to set before training begins and have a good sized impact at the version's overall performance, in comparison to version parameters which might be discovered during education. For hyperparameter tuning, we employed Grid Search and Random Search.

2.3.1. Grid Search

For every classifier, Grid Search iteratively looks through a predetermined set of hyperparameters. For each hyperparameter such as the number of neighbors for KNN or the normalization strength for Logistic Regression we set ranges and then analyzed each combination to determine which worked best. It can, however, be computationally costly, particularly for big datasets or intricate models.

2.3.2. Random Search

Conversely, Random Search selects random hyperparameter combinations from pre-established ranges. Because it concentrates on most promising regions rather than checking every combination, this method is typically more effective than Grid Search. Even though it might not identify the ideal hyperparameters, it frequently yields positive outcomes with fewer assessments.

For every classifier in our study, we used both approaches, and we compared the results.

2.4. Used Classifiers

Because of their distinct advantages and capacity to work with the structure of the heart disease dataset, we have chosen the following machine learning models:

2.4.1. Logistic Regression

As a model for binary classification, logistic regression is straightforward and efficient. Utilizing a logistic function fitted to the data, it forecasts the likelihood of a given event. It is particularly helpful in medical applications such as heart disease prediction when there is a linear relationship between features and the target. This makes it both predictive and simple to interpret.

2.4.2. Support Vector Machine (SVM)

Using SVM, a hyperplane is made to divide data into different classes. The radial basis function and other kernel functions are used to model intricate, non-linear relationships between features and the target variable. Because SVM performs well in high-dimensional spaces, it is a good fit for datasets with a large number of characteristics, such as the heart disease dataset. Additionally, it functions well in datasets with distinct class boundaries, which are typical in the medical domain.

2.4.3. Random Forest

Several choice trees are generated the usage of the Random Forest ensemble learning technique, which then aggregates the predictions to growth precision and decrease overfitting. To encourage diversity, it creates arbitrary subsets of samples and features for every tree. Because this method can handle different feature types and variable interactions, it works well with the heart disease dataset. It also captures potential non-linear relationships between outcomes of heart disease and risk factors.

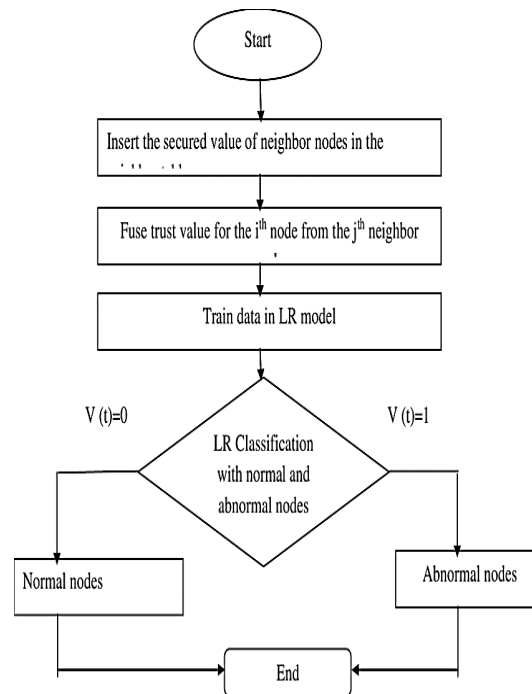


Figure 1. LR Flow Diagram

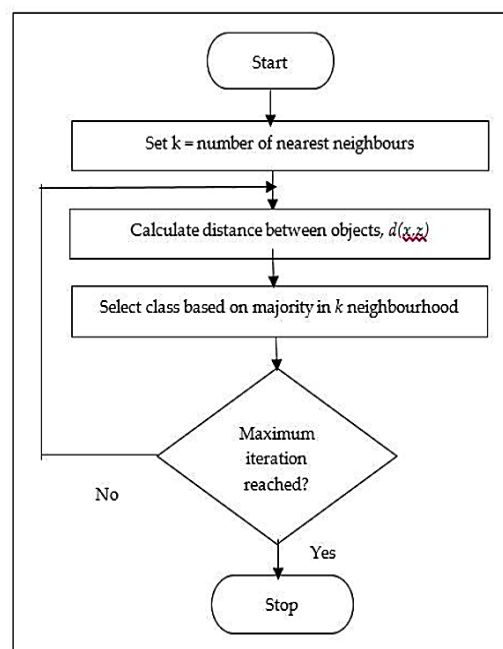


Figure 2. Distance Calculation Flow Diagram

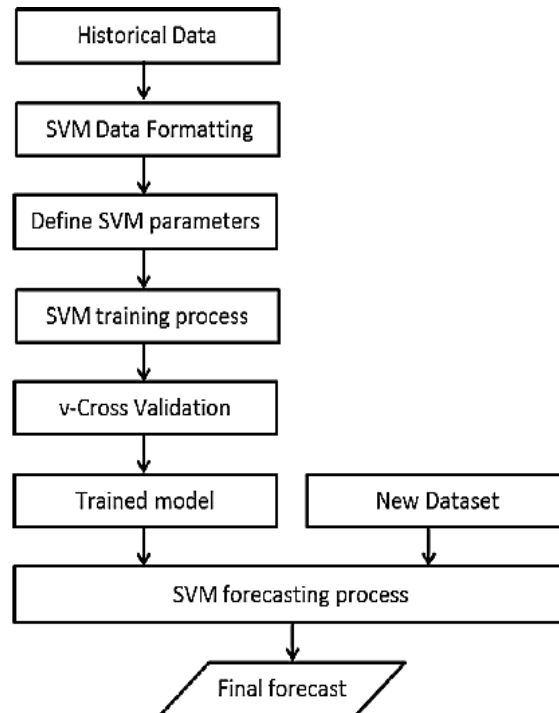


Figure 3. SVM Prediction Flow chart

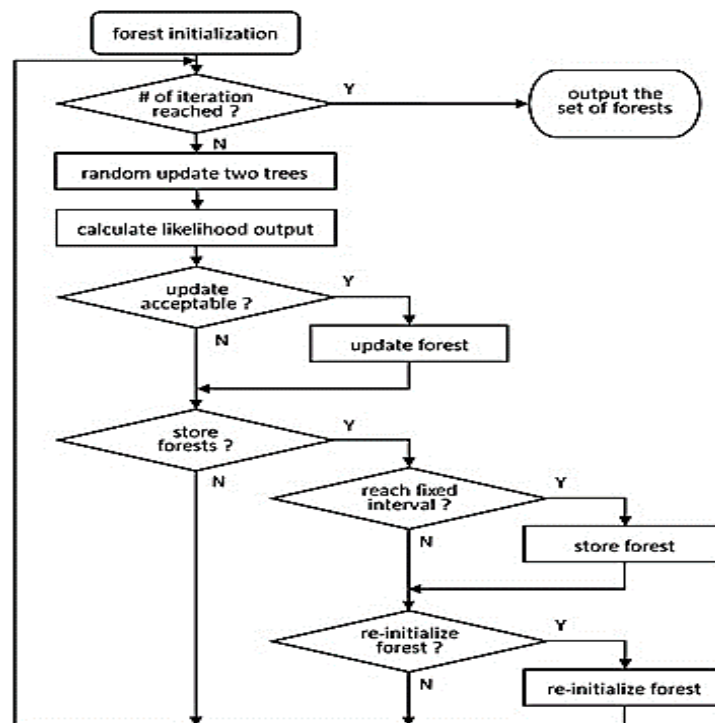


Figure 4. Random Forest Flow

2.4.4. Gradient Boosting

A technique called gradient boosting builds models in an ensemble, one after the other, fixing the mistakes of the one before it. The cardiac disease dataset is a good fit for this iterative method since it improves overall performance and handles large, feature-rich datasets. It is mainly powerful in clinical packages where facts disparities are common.

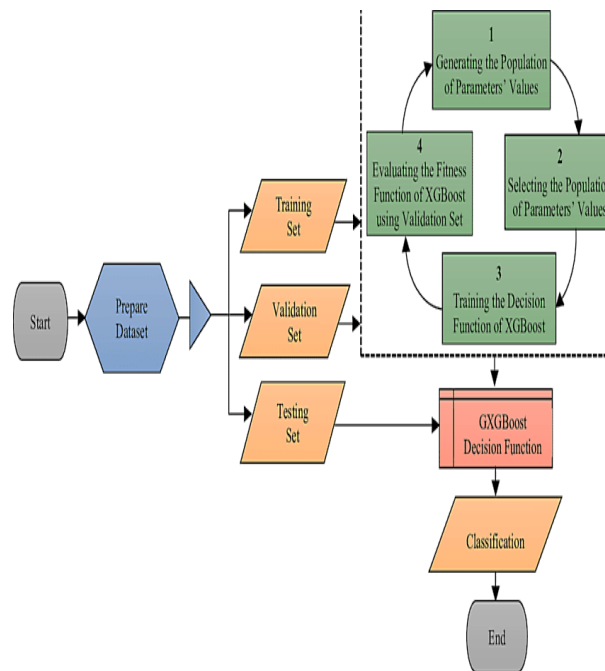


Figure 5. Gradient Boosting

2.4.5. Stacking

By combining every classifier's predictions, a final classifier is skilled thru the use of stacking, a meta-studying approach. This study used a lot of base fashions, like Logistic Regression, KNN, SVM, Random Forest, and Gradient Boosting. Stacking is a beneficial method as it brings the blessings of every model collectively to enhance normal performance. To stability the limitations of Logistic Regression in dealing with non-linear relationships, recall the non-linear talents of Random Forest or SVM. It is that this approach that makes stacking a powerful approach for heart disorder prediction.

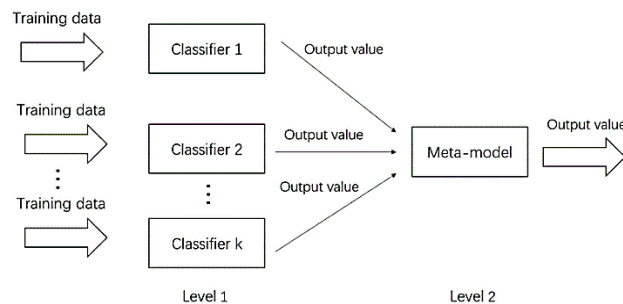


Figure 6. Stacking Flow

2.5. Computed Evaluation Metrics

For the purpose of imparting a radical analysis of the algorithms' performance, we used a number of carefully chosen measures. This is mainly important in clinical packages due to the fact each false positive incorrectly identifying a circumstance and false negatives failing to discover a situation could have negative consequences on affected person care. We are able to understand not only the general accuracy of the models but also their capability to account for specific sorts of mistakes due to the metrics we selected. By thinking about those metrics, we can make sure that the models are truthful and suitable for use in essential healthcare environments when making choices. With the help of this comprehensive evaluation and model assumptions, healthcare providers can make well-knowledgeable selections.

2.5.1. Accuracy

Accuracy is the proportion of successfully labeled instances in a dataset. While beneficial, it can be misleading in datasets that are not balanced, where a small percentage of patients have heart disease and a large proportion do. In these circumstances, a model may show high accuracy, but it might still be not able to identify patients with heart disease an important first step in receiving the right care.

$$\text{Accuracy} = \frac{\text{Total Number of Predictions}}{\text{Total Number of Predictions}}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Where:

- TP (True Positives): The variety of instances efficiently anticipated as positive.
- TN (True Negatives): The wide variety of instances effectively expected as negative.
- FP (False Positives): The quantity of times incorrectly predicted as positive.
- FN (False Negatives): The variety of instances incorrectly expected as negative.

In easier phrases, accuracy measures the proportion of correct predictions made by way of the model out of all predictions.

2.5.2. Precision

The accuracy of positive predictions as a percentage is measured by precision. High precision is crucial in medical applications because it prevents false positives, which guarantees that patients are not misdiagnosed with heart disease.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Precision} = \frac{TP}{FP+TP}$$

Where:

- TP (True Positives): The wide variety of instances efficaciously predicted as positive.
- FP (False Positives): The wide variety of times incorrectly predicted as positive.

Precision measures the accuracy of the positive predictions made via the model. A better precision shows that there are fewer fake positives most of the times expected as high quality.

2.5.3. Recall (Sensitivity)

The recall metric quantifies the model's ability to recognize real positive cases. High recall is crucial in medical diagnostics to guarantee accurate diagnosis of heart disease in patients and reduce the possibility of missed cases.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

Where:

- TP (True Positives): The number of instances correctly predicted as positive.
- FN (False Negatives): The number of instances that are actually positive but were incorrectly predicted as negative.

Recall measures the model's ability to correctly identify all relevant positive cases. A higher recall indicates that the model is good at capturing true positives and is less likely to miss positive instances.

2.5.4. F1 Score

Precision and recall are harmonic means, and this yields the F1 Score. Because it strikes a balance among remember and precision, it gives an extra accurate measure of model performance, making it valuable for datasets which are unbalanced.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{F1 Score} = 2 \times \frac{TP}{2TP+FP+FN}$$

Where:

- Recall is calculated as: $\frac{TP}{TP+FN}$

Put extra sincerely, the F1 rating is available in on hand whilst you want to decide the high-quality possible stability among recollect and precision, especially while coping with imbalanced datasets. A version with a better F1 rating performs higher.

2.5.5. AUC (Area under the Curve)

The charge of proper positives against the price of false positives at numerous thresholds is displayed by the place in the curve of the ROC, or AUC. A larger AUC indicates improved performance across a range of thresholds, which is helpful when assessing classifiers on datasets with imbalances.

2.5.6. Cohen's Kappa

When adjusting for chance, Cohen's Kappa gauges how closely predicted and actual labels match. Strong agreement is indicated by a Kappa value close to 1, while little to no agreement is indicated by values close to 0.

$$\kappa = \frac{Po - Pe}{1 - Pe}$$

Where:

- Po= Observed settlement (the proportion of instances where the raters agree).
- Pe = Expected settlement with the aid of threat (the percentage of instances in which the raters might agree through random risk).
- $\kappa = 1$: Perfect settlement.
- $\kappa = zero$: No settlement past risk.
- $\kappa < 0$: Agreement is much less than danger.

Values usually variety from -1 to 1, with higher values indicating better settlement among the raters.

3. Results and Discussion

We will evaluate how well various machine learning classifiers predict heart disease. We will identify the advantages and disadvantages of each model by looking at important metrics like accuracy, precision, recall, and F1 score. Tables 1, 2, and 3 provide a clear comparison of the classifiers' performances by displaying the specific results for each one using the Grid Search and Random Search techniques.

Table 1. Performance Evaluation Metrics of Different Classifiers

Model	Accuracy	Precision	Recall	F1 Score	AUC	Kappa	MCC
Logistic Regression	0.861345	0.871212	0.877863	0.874525	0.859492	0.7196	0.719626
KNN	0.886555	0.871429	0.931298	0.900369	0.881537	0.768998	0.771288
SVM	0.890756	0.867133	0.946565	0.905109	0.884497	0.776977	0.781126
Random Forest	0.945378	0.933824	0.969466	0.951311	0.942677	0.88916	0.889969
Gradient Boosting	0.915966	0.917293	0.931298	0.924242	0.914247	0.829915	0.830035
Stacking	0.945378	0.933824	0.969466	0.951311	0.942677	0.88916	0.889969

Model	Best Params	Accuracy	Precision	Recall	F1 Score	AUC	Kappa	MCC
Logistic Regression	{'C': 0.1, 'penalty': 'l2'}	0.861345	0.871212	0.877863	0.874525	0.859492	0.7196	0.719626
KNN	{'neighbors': 7, 'weights': 'distance'}	0.92437	0.912409	0.954198	0.932836	0.921024	0.846397	0.847508
SVM	{'C': 10, 'kernel': 'rbf'}	0.89916	0.92126	0.89313	0.906977	0.899836	0.796943	0.7974
Random Forest	{'max_depth': None, 'min_samples_split': 2, 'estimators': 300}	0.94958	0.947368	0.961832	0.954545	0.948206	0.897949	0.898079
Gradient Boosting	{'learning rate': 0.2, 'max_depth': 4, 'n_estimators': 100}	0.936975	0.946154	0.938931	0.942529	0.936755	0.872764	0.872795

Figure 7. Table of the Performance Evaluation Results of Classifiers with Grid search

Table 2. Performance Evaluation Metrics of Different Classifiers with Random Search

Model	Best Params	Accuracy	Precision	Recall	F1 Score	AUC	Kappa	MCC
Gradient Boosting	{'learning_rate': 0.1563987883622810, 'max_depth': 7, 'n_estimators': 120}	0.936975	0.932836	0.954198	0.943396	0.935043	0.872327	0.872611
Random Forest	{'max_depth': 48, 'min_samples_split': 5, 'n_estimators': 448}	0.945378	0.940299	0.961832	0.950943	0.943533	0.88935	0.88964
SVM	{'C': 5.90836121681994, 'kernel': 'rbf'}	0.903361	0.897059	0.931298	0.913858	0.900228	0.803898	0.804629
KNN	{'neighbor_s': 7, 'weights': 'distance'}	0.92437	0.912409	0.954198	0.932836	0.921024	0.846397	0.847508
Logistic Regression	{'C': 0.59083612168199, 'penalty': 'l2'}	0.861345	0.871212	0.877863	0.874525	0.859492	0.7196	0.719626

3.1. Comparative Discussion on Computed Evaluation Metrics

3.1.1. Accuracy

The percentage of accurate predictions among all instances is known as accuracy. After adjusting the hyperparameters, the accuracy of the majority of models increased. For instance, Random Forest's capacity to integrate several decision trees and capture intricate feature interactions allowed it to attain the highest accuracy rates, 94.95% with Grid Search and 94.54% with Random Search. By combining weaker classifiers, ensemble techniques like Gradient Boosting and Stacking also increased accuracy. KNN and logistic regression, however, performed less accurately. Though KNN's performance varied because of how sensitive it was to feature scaling and neighbor selection (k), Logistic Regression had trouble handling irregular patterns in the data.

3.1.2. Precision

The percentage of true positives among predicted positives is measured by precision. High precision lowers false positives, or patients who are incorrectly diagnosed with heart disease, which makes it essential for medical diagnostics. Because Random Forest averages several decision trees to reduce variance and prevent over-predicting positives, it achieved the highest precision of any algorithm during Grid Search, coming in at 94.73%.

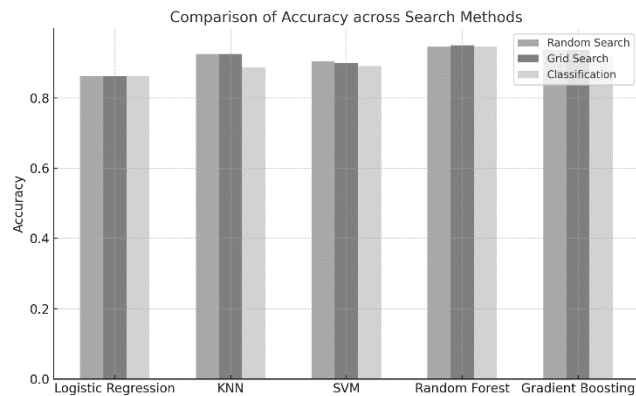


Figure 8. Accuracy across Search Methods

Reduced accuracy in SVM and Logistic Regression suggests that more negatives were mistakenly classified as positives in these models, particularly in cases of non-linear relationships. Non-linear class separations may be difficult for SVM with a linear kernel to handle, whereas KNN's sensitivity to distance and incorrect feature scaling can lead to misclassification.

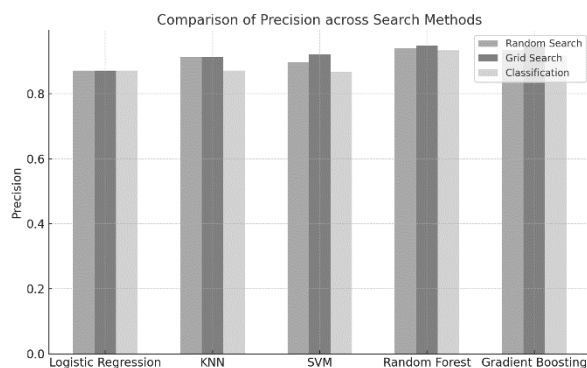


Figure 9. Precision across Search Method

3.1.3. Recall (Sensitivity)

Recall gauges a model's ability to detect genuine positive cases, which is essential for medical diagnostics to prevent patients with heart disease from being overlooked. With high recall, Random Forest and Gradient Boosting were able to identify the majority of heart disease cases. By combining several weak classifiers, these ensemble models lessen bias, aiding in the detection of subtle patterns and lowering false negatives. However, because they rely on linear or kernel-based decision boundaries, which are difficult to capture complex, non-linear relationships in the data, Logistic Regression and Support Vector Machines (SVM) had lower recall.

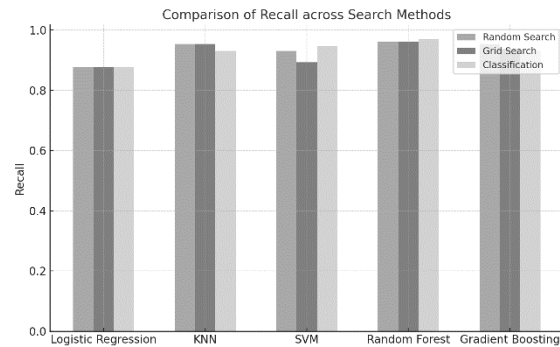


Figure 10. Recall across Search Methods

3.1.4. F1 Score

The F1 score, considered the harmonic mean of recall and precision, offers a balance between the two. The precision and recall of a model are critical in unbalanced datasets, like medical data, and a high F1 score represents this. With an F1 rating of 95.45 % in Grid Search, Random Forest became able to reveal an awesome stability among decreasing fake alarms and detecting real coronary heart sickness instances. Ensemble strategies are beneficial for complex, massive-scale datasets, as visible by the fulfillment of different models like Gradient Boosting and Stacking. KNN and logistic regression, on the other hand, produced too many false positives or missed real positives, which is why their F1 scores were lower, suggesting that they were less successful.

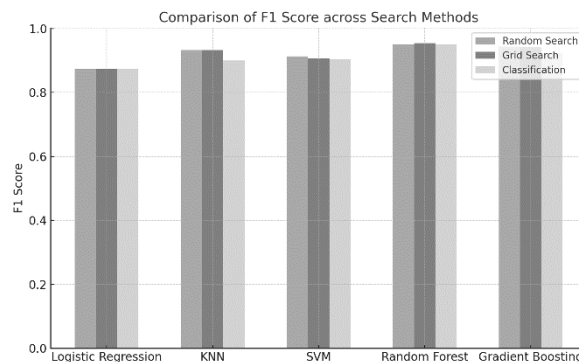


Figure 11. F1 Score across Search Methods

3.1.5. AUC (Area under the Curve)

AUC examines a model's ability to differentiate between positive and negative lessons at diverse cutoff points. An AUC that is high suggests that the model ranks predictions well, regardless of any threshold. With an AUC of 94.32%, Random Forest outperformed Gradient Boosting in Grid Search, showing good performance in identifying heart disease cases. Due to the resilience of such models used for healthcare diagnosis, physicians are able to adjust thresholds based on sensitivity specifications.

On the other hand, the AUC values of KNN and logistic regression were lower, showing that they had trouble classifying data. While the performance of KNN can vary based on the neighbors (k) chosen, the ranking power of Logistic Regression might have been limited by its focus on linear relationships.

3.1.6. Cohen's Kappa

Cohen's Kappa measures the degree of agreement between predicted categories and actual outcomes after adjusting for chance agreement. Random Forest and Gradient Boosting were the algorithms with the highest Kappa values, indicating good agreement between the observed and estimated results. This metric is vital due to the capability for an opportunity categorization in clinical programs. The excessive Kappa values of these ensemble methods display that they're more than a coincidence; they are dependable. Even though Logistic

Regression and KNN had less Kappa values, this shows that that they had hassle classifying situations effectively, particularly in instances wherein their class distributions were not uniform.

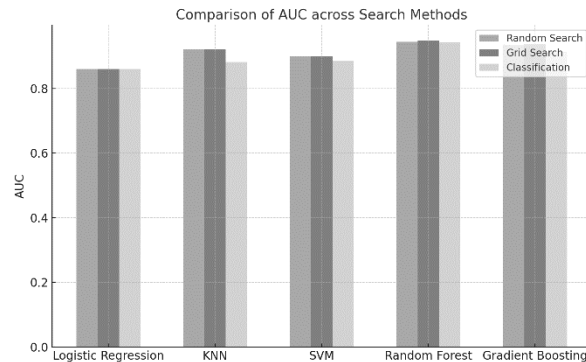


Figure 12. AUC across Search Methods

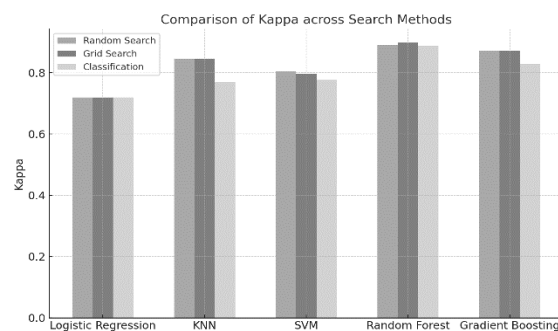


Figure 13. Kappa across Search Methods

3.1.7. Matthews Correlation Coefficient (MCC)

In the evaluation of unbalanced datasets, the Matthews Correlation Coefficient (MCC) is an Equal metric that may be useful for evaluating authentic and false positives in addition to negatives. High MCC values demonstrate that these types of models are properly-suitable for all project kinds and are capable of handle false positives as well as fake negatives. For example, Random Forest and Gradient Boosting have MCC values of zero.89 and 0.87, respectively. Their high MCC scores suggest that their estimations are accurate in spite of unequal elegance distributions. In evaluation, Logistic Regression in addition to KNN regarded to have problems with misclassifications due to their extra fundamental assumptions about the data, as proven by using their smaller MCC values.

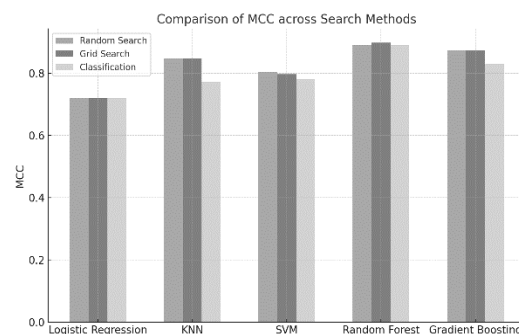


Figure 14. MCC across Search Methods

3.1.8. Discussion of Findings

More complex models, such as KNN and logistic regression, outperform ensemble methods, such as random forest and gradient boosting, according to the results. These ensemble methods can handle complex, non-linear interactions between characteristics because they combine predictions from multiple classifiers, reducing partiality and variances. With the highest scores for MCC, F1 score, accuracy, precision, recall, AUC, and Cohen's Kappa, Random Forest performed better than all other models in every metric. Its success may be defined to its ability to produce a variety of decision trees that cover various aspects of the dataset, creating a strong model for heart disease prediction.

However, the reason why Logistic Regression did not work well was because it assumed a linear connection between the target variable and features. KNN's dependence on distance metrics caused it to be sensitive to neighbor selection and feature scaling, which led to inconsistent results. SVM was flexible with kernels, but recall and precision were poor, suggesting that it failed to fully recognize all of the complex relationships in the data. Approaches for hyperparameter tuning like Grid Search and Random Search significantly enhanced model performance.

While Grid Search regarded into hyperparameters in terrific detail, Random Search yielded comparable results with a less massive computational footprint, which makes it an amazing stand-in. All matters considered, ensemble techniques like Random Forest and Gradient Boosting are strongly cautioned to be used with scientific datasets since they produce reliable and correct projections throughout a whole lot of metrics used for assessment.

4. Conclusion

This observe as compared diverse system learning classifiers for coronary heart ailment prediction, such as Random Forest, Gradient Boosting, K-Nearest- Neighbors (KNN), Support- Vector Machines (SVM), and Logistic Regression. Essential parameters like accuracy, precision, remember, F1 score, AUC, Cohen's Kappa, and MCC were assessed while the models have been optimized the usage of grid seek and random search. With an accuracy of 94.95% along with brilliant achievement in different metrics, the Random Forest version proved to be the simplest, illustrating its potential to control complex facts interactions. The effectiveness of ensemble strategies changed into shown through Gradient Boosting, which completed properly even in the face of non-linear relationships, while less complicated fashions consisting of Logistic Regression as well as KNN struggled. Adjusting the hyperparameters was required, and Random Search completed extra effectively than Grid Search. Research efforts inside the future must recognition on enhancing the interpretability of excessive-appearing models like Random Forest and exploring superior strategies for dealing with imbalanced datasets. By offering facts-pushed assist with early detection of heart sickness, the models created by means of this research have the potential to decorate scientific selection help systems. By increasing the accuracy and efficacy of heart screening within healthcare systems, these models have the potential to reduce the global burden of cardiac disease.

References

1. Abdalla, M., & Wani, M. A. (2021). A Survey on Heart Disease Prediction Using Machine Learning. *International Journal of Computer Applications*, 175(3), 5-11.
2. Ahmed, E., & Ahmed, A. (2019). Heart Disease Prediction System Using Machine Learning Algorithms: A Review. *Journal of Electrical Engineering and Automation*, 1(2), 76-86.
3. Alferjani, A. A., & Fadli, R. (2020). Heart Disease Prediction Using Machine Learning Techniques: A Review. *International Journal of Computer Applications*, 975(888), 22-26.
4. Bashar, M. A., & Raza, S. A. (2021). A Novel Approach to Heart Disease Prediction Using Machine Learning Algorithms. *Journal of King Saud University - Computer and Information Sciences*.
5. Bashir, A., & Khan, M. (2018). Heart Disease Prediction Using Machine Learning Algorithms: A Review. *International Journal of Computer Applications*, 181(13), 1-7.
6. Bhatia, M., & Joshi, K. (2020). Predicting Heart Disease Using Machine Learning. *International Journal of Engineering Research and Technology*, 9(1), 34-39.
7. Chaurasia, V., & Pal, S. (2018). A Novel Approach for Heart Disease Prediction Using Hybrid Model. *International Journal of Computer Applications*, 182(21), 1-7.
8. Dhanavandan, S., & Rani, S. S. (2018). A Survey on Heart Disease Prediction System Using Data Mining Techniques. *International Journal of Advanced Research in Computer Science*, 9(5), 17-23.
9. E. H. M., & Aras, G. (2021). Heart Disease Prediction Using Machine Learning Techniques: A Review. *Proceedings of the 2021 4th International Conference on Machine Learning and Big Data Analytics*.
10. Fatima, A., & Jha, S. (2019). Heart Disease Prediction System Using Machine Learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 4(1), 104-109.
11. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189-1232.
12. Ganaie, M. A., & Wani, S. U. (2020). Heart Disease Prediction Using Machine Learning Algorithms: A Survey. *Journal of Computer and Communications*, 8(5), 16-24.
13. Ejaz, F., Tanveer, F., Shoukat, F., Fatima, N., & Ahmad, A. (2024). Effectiveness of routine physical therapy with or without home-based intensive bimanual training on clinical outcomes in cerebral palsy children: a randomised controlled trial. *Physiotherapy Quarterly*, 32(1), 78-83.
14. Ganaie, M. A., & Wani, S. U. (2021). A Comprehensive Survey on Heart Disease Prediction Using Machine Learning Techniques. *Computers in Biology and Medicine*, 138, 104907.
15. Gopikrishnan, K., & Sathya, C. (2018). A Review on Heart Disease Prediction Using Data Mining Techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 8(5), 15-21.
16. Goyal, P., & Goyal, D. (2021). Comparative Study of Machine Learning Techniques for Heart Disease Prediction. *Journal of Engineering Science and Technology Review*, 14(1), 37-44.
17. Hossain, S. S., & Rehman, M. (2020). A Review on Heart Disease Prediction Using Machine Learning Techniques. *International Journal of Computer Applications*, 975(888), 8-12.
18. Sajjad, R., Khan, M. F., Nawaz, A., Ali, M. T., & Adil, M. (2022). Systematic analysis of ovarian cancer empowered with machine and deep learning: a taxonomy and future challenges. *Journal of Computing & Biomedical Informatics*, 3(02), 64-87.
19. Kavitha, R., & Karthikeyan, M. (2019). A Review on Heart Disease Prediction Using Machine Learning Techniques. *International Journal of Recent Technology and Engineering*, 8(153), 1-5.
20. Kumar, A., & Kumar, S. (2020). Machine Learning Techniques for Heart Disease Prediction: A Survey. *Artificial Intelligence Review*, 54, 89-107.
21. Kumar, K., & Ghosh, S. (2020). A Review of Machine Learning Techniques in Heart Disease Prediction. *International Journal of Engineering Research and Applications*, 10(6), 49-53.
22. Mishra, A., & Patra, S. (2021). Heart Disease Prediction Using Machine Learning Techniques: A Review. *Journal of Physics: Conference Series*, 1886, 012014.

23. Shah, A. M., Aljubayri, M., Khan, M. F., Alqahtani, J., Sulaiman, A., & Shaikh, A. (2023). ILSM: Incorporated Lightweight Security Model for Improving QOS in WSN. *Computer Systems Science & Engineering*, 46(2).