Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 2 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

A COMPARISON OF MACHINE LEARNING CLASSIFIERS AND UNCERTAINTY QUANTIFICATION TECHNIQUES FOR PREDICTING HEART DISEASE

Batool Sharif
Department of Basic Sciences, Superior University Lahore, Pakistan

M U Hashmi
Department of Computer Sciences, Superior University Lahore, Pakistan

Muazzam Ali
Department of Basic Sciences, Superior University Lahore, Pakistan

Muhammad Azam
Department of Computer Sciences, Superior University Lahore, Pakistan

Abdul Manan
Department of Basic Sciences, Superior University Lahore, Pakistan

Nayab Imtiaz
Department of Basic Sciences, Superior University Lahore, Pakistan

****Corresponding author: Muazzam Ali (muazzamali@superior.edu.pk)***
DOI: https://doi.org/10.71146/kjmr261

Article Info

Abstract

In all over the world, Heart disease is the leading disease which causes sudden death, so for early detection of this disease and treatment need a reliable and accurate prediction models. This study highlights comprehensive approaches of machine learning classifiers such as Logistics Regression, Naïve Bayes, Random Forest Tree, K-Nearest Neighbor, Decision Tree and Support Vector Machine are applied on a dataset of indication and different symptoms of this disease. The uncertainty quantification techniques to increase the accuracy such as dropout uncertainty, ensemble and quantile regression are used. Also F1 score, accuracy, precision, recall,, support, confusion matrix are computed and the Receiver Operating Characteristic and area under the curve to evaluate the overall performance of the classifiers are discussed. The highest accuracy of testing is 99.02% with random forest classifier, when integrated with quantile regression it gives same output. Moreover, this study highlights the value of uncertainty quantification to improve the accuracy of many classifiers to predict the heart disease. This study highlights the value of uncertainty quantification in enhancing version reliability and shows how well the Random Forest classifier predicts coronary heart disease. The results provide comprehensive understandings of how to apply machine learning models in academic contexts to improve patient experiences and diagnostic accuracy.

Keywords:
Heart Disease Prediction, Uncertainty quantification, Random Forest Tree, ROC, Predictive Modeling.

Introduction

Heart disease is the leading cause of mortality globally and has an important effect on global health. One third of all deaths annually, or approximately 17.9 million, are due to heart disease and cardiovascular disease (Gaziano et al., 2022). Increasing older people, the process of urbanization, and the rise of risk factors such as diabetes, hypertension, and obesity are contributing to an increase in the incidence of heart disease. Heart disease has an economic effect on people, families, and health care systems due to clinical costs, skipped salaries, and long-term disability (Johnson et al., 2021). Heart damage harms public activities by causing deep and mental laziness in both the affected person and others around them. Heart disease is becoming increasingly common and has a significant impact on the lives of many people. This includes the basic requirement for accurate and trustworthy expectation models that could help in early diagnosis and treatment, improve outcomes for those affected, and reduce financial and societal costs (Belard et al., 2017).

Predicting heart disease accurately is extremely difficult due to the complexity and unpredictable nature of its symptoms and indicators. Heart disease can have a wide range of side effects, from weakness and sickness to rigidity in the chest and fatigue, making diagnosis and treatment more difficult and time-consuming. This requirement requires using precise and reliable degree models with the intention to offer correct patient-particular treatment and proof (Hendriksen et al., 2013). Research specializes in a first rate deal of risk; fake results can also cause troubles and needless clinical attention, even as false terrible outcomes can also result in overlooked opportunities for early intervention and probably existence-saving medications (Hoffmann et al., 2017). Furthermore, offering suited examples is greater complex because of the inherent weaknesses in efforts. Measuring strategies need to be used to ensure that the predictions are particular and that the levels of confidence are understood so that it will treatment this problem. Precise fashions can also provide more goal and dependable records by means of comparing hazard, so as to ultimately increase their usefulness in medical contexts (Eck et al,2016, Mirams et al., 2016). Predicting cardiac sickness can be challenging since its side effects are complex and vary widely. Numerous adverse symptoms, such as queasiness, dyspnea, drowsiness, and chest pain, may also be associated with heart disease. Treatment delays may also result from testing and analysis prompted by the wide range of adverse effect manifestations. High precision and dependability are essential for prescient styles since false negatives and upsides may have major effects in logical situations (Neal et al., 2015, Walter et al., 2012, Jorden et al., 2011). While a false negative might result in lost opportunities for early intervention and have major fitness effects, a false positive can expose patients to needless strain, intrusive procedures, and therapies.

System learning, which has become a wonderful tool for sickness detection and prognosis, is revolutionizing healthcare practices. By examining large datasets that may not immediately be visible to human observers, machine learning algorithms can identify patterns and connections in patient data (Zhang et al., 2023). In the context of sickness analysis, algorithms that examine devices may employ these patterns to precisely calculate a patient's risk of developing a certain ailment. This enables scientists to respond quickly and maybe prevent unfavorable outcomes. Classifiers are a crucial component of the device learning algorithms used in infection prediction (Uddin et al., 2019). These algorithms are trained on category datasets, where each data point is associated with a certain outcome or condition (e.g., the existence or non-existence of a disease). Classifiers are able to identify patterns in the statistics and classify recently found, unlabeled information sections by examining those labeled examples (Bekker et al., 2020). Many classifiers are employed often, including Naive Bayes, Random Forest, K-Nearest Neighbors, Decision Trees, Support Vector Machines, and Logistic Regression; each has advantages and disadvantages based on the clinical situation. Measuring uncertainty techniques are crucial to increasing the precision of system mastery forecasts. By comparing the level of uncertainty in the model's predictions,

these techniques provide important details on the level of confidence in each forecast. Quantile regression, ensemble methods, and dropout uncertainty are a few common ways to measure uncertainty in sickness prediction. When uncertainty quantification is included into prediction models, especially when the consequences of false positives or false negatives are significant, healthcare professionals may also be able to make more informed decisions (Houben et al., 2019).

This paper includes uncertainty assessment procedures and examines the prediction ability of six machine learning classifiers for heart disease. The objective is to improve diagnostic accuracy and increase the prediction of these effects in real-world scientific scenarios by carefully assessing many performance metrics, such as accuracy, ROC curve, and F1 score. This finding might significantly impact patient care by enhancing outcomes for individuals at risk for heart disease and providing doctors with more dependable methods for early identification and treatment. The application of predictive modeling in clinical practice and healthcare has benefited from this endeavor. The goal is to arm doctors with more advanced tools to anticipate cardiac illness by thoroughly comparing device learning classifiers, including methods for measuring uncertainty. There are significant implications for scientific exercise since reliable and specialized prediction models may also help with early analysis and management, with the goal of eventually causing greater patient outcomes and carrier satisfaction. Moreover, this work contributes to the larger field of predictive modeling by expanding the understanding of the effectiveness of various algorithms and uncertainty quantification methods. This study opens up new avenues for healthcare innovation and research by providing a foundation for future investigations into the optimization of prediction models for a variety of clinical conditions.

Literature Review

Many studies have been conducted in recent years to anticipate cardiac illness utilizing system analysis techniques to create accurate and dependable models. Verma and Gupta examined the software of statistics mining and gadget learning approaches for heart disease prediction in one of the first studies on this topic. Their research tested a way to improve the accuracy of coronary heart sickness prediction models by using device mastering techniques such as logistic regression, random forests, and support vector machines (Verma et al,2021). Building on these first results, different studies have investigated the use of state-of-the-art gadget studying strategies, which include switch mastering and deep getting to know, for the prediction of coronary heart disorder. Determined the extent to which deep gaining knowledge of fashions, such recurrent and convolutional neural networks, can recognize complex styles from huge datasets and enhance the precision of coronary heart ailment prediction models. (Patel et al., 2020) confirmed multiplied efficacy and efficiency when transfer getting to know strategies were carried out to hire pre-expert fashions and adapt them for obligations that anticipate coronary heart disorder.

Researchers have also labored with hybrid models and ensemble methods to decorate the forecasting talents of fashions for cardiac problems. Several device studying algorithms, together with Random Forest, Support Vector Machines, and Gradient Boosting, are applied to the setting of coronary heart sickness prediction so one can demonstrate how ensemble techniques may want to enhance the accuracy of such designs. Similar to this, studies on using a lively learning device for the prediction of cardiovascular infection confirmed the efficacy of hybrid models of coronary heart disease that safeguarded present day tool mastering strategies and deep gaining knowledge of strategies (Hassan et al., 2023). Additionally, studies has examined gadget learning software program for chance detail identity and affected man or woman category, critical areas of coronary heart ailment prediction. For instance, tool gaining knowledge of may be used to decorate affected person effects by way of utilizing algorithms to discover patients that pose an excessive chance primarily based on their scientific traits and demographics. Similarly, Singh et al.'s artwork proven how machine getting to know may be used to increase patient elegance accuracy by growing a prediction version for intricate

character stratification (Motwani et al., 2017, Olsen et al., 2020). A top notch deal of research has been finished these days on the aggregate of device studying with uncertainty quantification strategies, especially in the discipline of infection prediction. By imparting an extra thorough expertise of the uncertainty related to prediction styles, this technique seeks to decorate choice-making in healthcare contexts (Costa Bal et al., 2019). For instance, certain strategies for very well examining uncertainty quantification have pressured the want to include uncertainty estimates into tool mastery models on the way to enhance their reliability and safety.

The software of uncertainty quantification strategies, inclusive of dropout uncertainty, ensemble methods, and quantile regression, in infection prediction has been the concern of several studies (Kabir et al,2018). It became evaluated how exceptional uncertainty quantification techniques, consisting of Bayesian Neural Networks, may be used to device evaluation for bio signal packages. Similar to this, the need of expressing doubt in clinical device research introduced interest to the need of tool study models to be able to specifically identify uncertainty by using affirming "I do no longer know" and maybe refraining from creating a diagnosis or prediction (Crisan et al., 2023). The impact of uncertainty quantification on the accuracy and dependability of prediction models for different illnesses has been thoroughly studied by the researcher (Kwon et al,2020). It has been demonstrated how similar tactics may also increase the precision and dependability of predictive models using the software for uncertainty quantification in the prediction of heart infections. Similarly, by assessing the effectiveness of multiple device learning algorithms in heart disease prediction, such as Random Forest, Support Vector Machines, and Gradient Boosting,(Aly et al.2022) paintings confirmed how ensemble techniques might improve the predictive version accuracy.

It is generally recognized that integrating uncertainty quantification techniques into device learning has great potential to improve the accuracy and reliability of prediction models for many diseases. However, further research is needed to address the challenges and limitations associated with these approaches, such as the models' interpretability, scalability, and generalizability.

Methodology

Data Collection

The dataset used for this observation was sourced from the well-known website Kaggle, which houses datasets given by individuals and businesses for use in research and analytical contests. This particular dataset is extremely relevant to the study's goals as it consists of thorough patient records that have been hand-selected especially to aid in the prediction of cardiac sickness. Each of the 1,000 afflicted individual data in the collection contains thirteen fantastic health indicators and markers of a heart condition. These comprise angina, blood pressure, low-density lipoprotein (LDL) cholesterol, heart rate variability, and kind of chest discomfort, among others. The dataset was divided into seven hundred examples (or 70%) for training the device to master fashions, and three hundred instances (or 30%) were set aside for validation and inspection. This ratio was adopted for dataset partitioning. During the preliminary investigation, it was observed that there was a 5% occurrence of missing data across many variables. This meant that preprocessing was necessary to ensure the stability of the developed prediction models. In order to ensure that the dataset was anonymized and that permission was obtained for its usage in research settings, ethical clearance was verified with the Kaggle information issuer before to using the dataset for this study. Any identifiers that may be used to identify specific victims have been either encrypted or removed in order to protect the privacy of the affected individuals. To safeguard from any misuse of the facts, all records processing methods were in compliance with international facts protection standards. Regulations specific to utilizing scientific records have been strictly followed, upholding the integrity of the study and its moral standards.

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Data Preprocessing

The dataset was cleaned as the initial step in the training procedure to ensure the consistency and accuracy of the observation. Many imputation strategies were employed to address the five percent cost of missing information throughout certain attributes. Lacking values for continuous variables like blood pressure and low-density lipoprotein cholesterol were imputed using the mean of the associated feature in order to maintain the statistical integrity of the dataset. The type of chest pain and the imputation method for expressing variables led to the hiring of the mode. After the facts were cleared, were standardized such that all of the numerical values fell within the same range. This became significant since the medical indicators had unique scales (heart rate in beats per minute, cholesterol in milligrams per deciliter, etc.). Min-max scaling was used to modify the data so that each feature had a maximum of 1 and a minimum of 0. For distance-primarily based classifiers such as K-Nearest Neighbors, which are explored, this normalization approach may be quite helpful, as all attributes contribute similarly to the gap calculations.

The processes of feature engineering and selection had proved essential to improving the device's performance in learning models. The function selection approach begins with a correlation evaluation in order to identify and eliminate qualities that are highly associated, hence reducing duplication and the potential for multicollinearity. Features were eliminated if their correlation coefficient with any other feature was higher than 0.85. Feature engineering was used not only to choose the most informative features but also to generate new variables that can enhance model performance. For example, a new feature named "Risk Score" was created by aggregating clinically significant variables, such age, cholesterol, and systolic blood pressure, and weighting them according to a preliminary analysis's determined importance. The goal of this manufactured feature was to create a single predictive variable that would incorporate several danger indicators. Creating dummy variables out of category variables to make them easier to employ in machine learning algorithms that handle numerical input was another facet of feature engineering. Binary representations of categorical features, such as the kind of chest pain and the presence of angina, were created using techniques like one-hot encoding. In order to improve the accuracy and effectiveness of the heart disease prediction models, feature selection and engineering were used to refine the input data and customize it to better fit the analytical methodologies utilized later in the study.

Model Selection

A variety of machine learning classifiers are used in this work, each with unique features and approaches that are appropriate for heart disease prediction. Among the classifiers were;

(a) Logistics Regression

A statistical technique for simulating the likelihood of a binary outcome is called logistic regression. It models a binary dependent variable based on one or more predictor variables using a logistic function. In medical situations, logistic regression is used because it yields probabilities that may be readily understood as risk

factors. This is especially helpful in illness prediction, where knowing the probability of occurrence is essential. Furthermore, its simplicity makes it simple to convey to medical specialists. Mathematically Logistic Regression is given as

$$P\left(Y = \frac{1}{X}\right) = \frac{1}{1 - e^{-(\alpha_0 X_1 + \alpha_1 X_2 + \cdots \ldots + \alpha_i X_i)}}$$

Where $P\left(Y = \frac{1}{X}\right)$ represents the probability that the output Y equals 1 given the input featuresX,

$\alpha_0, \alpha_1, \ldots\ldots \alpha_i$, are the coefficients or weights associated with the input features and x_i denote the input features.

(b) Naïve Bayes

A family of straightforward probabilistic classifiers known as "naive Bayes classifiers" is based on using the Bayes theorem under the strong (naive) independence assumptions between the features. Because Naive Bayes works well with high-dimensional datasets which are frequently seen in medical diagnostics where a variety of symptoms and test results are taken into account—it is used in this situation. It offers quick calculation and respectable accuracy, making it ideal for preliminary screenings. It works well with a large data.

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

where $P(C|X)$ is posterior probability of class (C, Target) given predictor $(X, \text{Attributes})$, $P(C)$ is the prior probability of class $P(X|C)$ is the likelihood which is the probability of predictor-given class and $P(X)$ is the prior probability of predictor.

(c) Random Forest

During training, Random Forest creates a large number of decision trees and outputs the mean prediction (regression) or mode of the classes (classification) of each individual tree. Random Forest is an ensemble learning technique for classification and regression. Because Random Forest manages over fitting in huge datasets more effectively than many other classifiers, it is the algorithm of choice. Its robustness for medical predictions—where several factors may interact in unforeseen ways comes from its ability to manage intricate feature interactions and excellent accuracy. A Random Forest Classifier is composed of a collection of classification trees h;

$$\{h(x, T, \varphi_k), \quad k = 1, 2, \ldots, K\}$$

Where φ_k represents identically and independently distributed random vectors, and each tree casts a unit vote for the most likely class at input x

(d) K-Nearest Neighbors

A non-parametric technique for regression and classification is K-Nearest Neighbors. In both scenarios, the object is assigned to the class that has the highest frequency among its k nearest neighbors. The input is made up of the k closest training instances in the feature space, and the output is decided by a majority vote of its

neighbors. KNN is used because it is easy to use and efficient, especially since the logical assumption in medical diagnostics is that comparable instances would have similar results. Additionally, it is quite intuitive, which helps practitioners grasp the forecasts. Mathematically it is represented by

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

(e) Decision Tree

A tree-like model of decisions and their potential results, such as chance event outcomes, resource costs, and utility, is used by the Decision Tree decision assistance tool. A clear and simple-to-understand paradigm is what Decision Trees offer, and this is crucial in therapeutic situations. This may readily handle qualitative (categorical) aspects as part of the diagnosis and assist in determining important threshold values for medical testing. The nodes of decision is selected on the high information gain of attribute which is calculated using the following formula

$$\text{Entropy} = -\sum_{i=1}^{n} p_i \log_2 p_i$$

(f) Support Vector Machine

Encouragement For two-group classification issues, Vector Machine is a supervised machine learning model that makes use of classification techniques. An SVM model can classify fresh text after being given sets of labeled training data for every category. Because of its efficiency in high-dimensional spaces and its adaptability to non-linear decision limits thanks to the kernel method, support vector machines (SVM) are widely utilized. Because of this, it is very well suited for intricate and subtle medical prediction tasks for which linear bonds are insufficient. Support Vector Machines were introduced by Vladimir Vapnik and his colleagues.

If there are given n training examples $\{x_i, y_i\}, i = 1, \cdots, n$, where each example has m inputs ($xi \in R^m$), and a class label with one of two values ($y_i \in \{-1, 1\}$). Now, all hyperplanes in R^m are parameterized by a vector w, and a constant b, expressed in the equation $w \cdot x + b = 0$

Since each of these classifiers has a special set of advantages, using them all together can help create a strong prediction model for the medical industry. When used in tandem, it allow the utilization of numerous information capabilities and theoretical frameworks to decorate prediction precision and dependability.

Uncertainty Quantification Techniques

Three main methods for measuring uncertainty were applied to all classifiers in the study in order to improve the predicted patterns and provide actionable insights with measurable confidence: Decision trees, Random Forest, Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machines, and Logistic Regression. Below are the definition and explanation of each approach, along with the manner it was used with each classifier to extract positive facts.

(a) Quantile Regression Method

Beyond mean projections, a statistical technique called quantile regression provides a more comprehensive perspective on ability outcomes by predicting the quantiles of the underlying variable's distribution conditional on impartial components. It provides information about the risk factors and variability in the predictions by enabling the version to anticipate not just the mean result but also the likelihood of seeing occurrences in many quantiles. The use of quantile regression yielded distinct probability outcomes for every classifier. In order to comprehend the hazard classification of afflicted individuals in coronary heart disease forecasts, it is critical to assess the range of potential medical consequences and the associated opportunities. It was especially helpful in defining the self-assurance intervals around the life of the contamination probability in SVM and logistic regression. The following formula defines the quantile regression,

$$Y_i = X_i\beta_r + E_i$$

Where the vector of unknown parameters associated with the rth quantile is denoted by β_r.

(b) Ensemble Methods

Ensemble approaches combine different machine learning approaches into a single predictive model to achieve certain goals such as increasing bias, improving predictions, or decreasing variation. By synthesizing many styles' results, these techniques successfully reduce the effects of noise and capacity over fitting of any single variation, leading to more accurate and dependable forecasts. To increase stability and accuracy, the Bagging classifier, which makes use of the bootstrap aggregating principle are used. This method involves training a few prediction models using excellent subsets of the original data, which may be selected at random with an alternative. By adding together the effects of each model's autonomous operation, the final consensus forecast is generated. This technique currently reduces variance to improve prediction robustness and prevent overbearing. Certain decision trees and other high-variance patterns are among the most basic bagging systems. By merging several choice timbers to form a Bagging ensemble, which performs additional tasks to a Random Forest, this may obtain a significant reduction in prediction variance without seeing a significant increase in bias. Furthermore, to get a consensus forecast, the used ensemble methodologies similar to those of various classifiers, such as Naive Bayes, providing a thorough understanding of prediction stability and dependability of the relevant fashions. It will make sure that predictions are not only accurate and dependable but also make use of the combined information provided by various expert viewpoints, demonstrating a sophisticated understanding of the available information, by combining predictions from exclusive configurations and models using ensemble techniques like bagging.

(c) Dropout Uncertainty

To deal with dropout uncertainty in tool mastery algorithms, the Monte Carlo dropout is used. To avoid overfitting and provide an empirical representation of the prediction uncertainty of the particular version, this method involves a random cutoff of neurons throughout the training process. This method, which was first created for neural networks, has been modified to operate with KNN and SVM in addition to other classifiers by putting the delay on positive capabilities or quantities of the training set. Monte Carlo dropout will raise forecast confidence across a wide range of scenarios by improving version dependability, increasing accuracy, and providing insights into how resistant trends are to statistical change.

Performance Evaluation Metrics

A wide range of measures became applied to examine and punctiliously determine the effectiveness of the classifiers used to expect coronary heart disease. Every indicator offers distinct perspectives on the precision and dependability of the prediction models, guaranteeing a comprehensive evaluation of their efficacy. Each metric used is explained in depth below, along with information on how it was determined and utilized.

(a) Accuracy

This metric calculates the percentage of genuine results—true positives and true negatives—out of all the instances that were looked at. It provides a straightforward indicator of how well a classifier predicts both classes (disease present or absent) with accuracy. The ratio of accurately anticipated observations to total observations is used to compute it.

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

(b) Recall

The frequency with which a machine learning model properly detects positive examples (true positives) out of all the real positive samples in the dataset is known as recall.

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

(c) Precision

The ratio of accurately predicted positive observations to all expected positives is known as precision. It is significant because it illustrates the model's capacity to provide precise positive detections with few false positives in the medical domain.

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

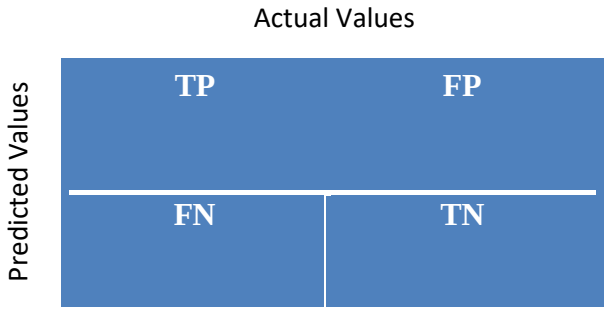
(d) F1 Score

The harmonic mean of recall and accuracy is the F1 Score. It is employed as it moves a compromise between don't forget and precision, supplying a solitary accuracy metric that debts for each fake positives and false negatives. It's quite helpful in cases when there may be an unequal distribution of classes.

$$\text{F1 score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

(e) Confusion Matrix

When describing how nicely a class model performs on a hard and fast of information for which the real values are known, a confusion matrix is usually a desk. It describes the general overall performance of the model and outputs a matrix.



(f) The ROC Curve

The Receiver Operating Characteristic Curve, or ROC Curve, is a graphical representation that indicates how a binary classifier system can be diagnostically assessed when its discrimination threshold is modified. Plotting True Positive Rate (TPR), also referred to as keep in mind, versus False Positive Rate (FPR) is used.

(g) Area under the ROC Curve

The two-dimensional area under the entire ROC curve, from (0,0) to (1,1), is measured as the area underneath the ROC curve, or AUC. It gives a mean performance throughout all feasible class stages. A model with an AUC of 0, predicts the consequences are a 100% wrong, whereas a model with an AUC of 1.0 predicts the results are a hundred% proper. All metrics were calculated the use of the testing records set, with the model schooling segment being the exception. Scikit-Learn and other Python modules offer capabilities for without problems computing those metrics.

Results and Discussion

Logistics Regression

The computed results provide an understanding of how uncertainty quantification techniques used to logistic regression and how impact the prediction accuracy with relation to heart disease.

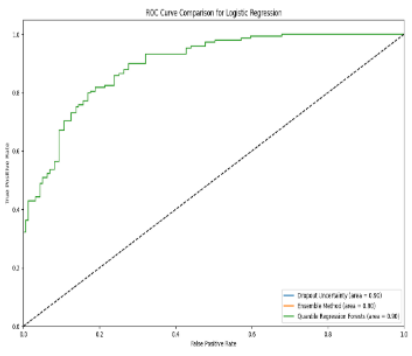
Classifier	Training Accuracy (%)	Test Accuracy (%)	Precision	Recall	F1 Score	Confusion Matrix	AUC
Logistics Regression	87.30	80.51	0.76	0.86	0.81	$\begin{bmatrix} 119 & 40 \\ 20 & 129 \end{bmatrix}$	0.90
Logistic Regression with Dropout Uncertainty	87.90	81.39	0.77	0.87	0.82	$\begin{bmatrix} 119 & 40 \\ 20 & 129 \end{bmatrix}$	0.90
Logistic Regression with Ensemble Method	86.05	77.92	0.73	0.83	0.78	$\begin{bmatrix} 115 & 44 \\ 24 & 125 \end{bmatrix}$	0.89

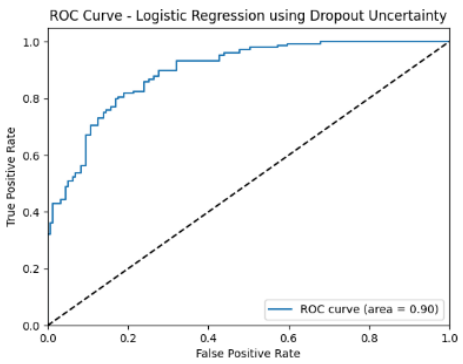
Logistic Regression with Quantile Regression Forests	88.21	81.51	0.78	0.88	0.82	$\begin{bmatrix} 119 & 40 \\ 20 & 129 \end{bmatrix}$	0.90

Table 1: Performance Comparison of Logistic Regression Models with Various Techniques of Uncertainty Quantification

The Logistic Regression method creates a baseline by balancing accuracy with other performance metrics. The high recall illustrates how important it is to efficiently identify genuine positive instances in order to decrease missed diagnoses in medical diagnostics. Accuracy and recall are somewhat enhanced by the introduction of dropout uncertainty. This technique replicates the impact of training several models by including randomization into the training process. The greater capacity of the model to prevent over fitting is probably the reason for the gain in accuracy and recall, which suggests that the model is more broadly applicable to handle unknown data. The ensemble approach performs worse than the baseline. This drop in performance might mean that the ensemble configurations (boosting, bagging, stacking, etc.) were not optimized or did not work effectively in conjunction with the Logistic Regression model. Less dependability in catching all positive instances is shown by the decline in recall and accuracy, particularly in medical diagnostics where the cost of false negatives is considerable. The greatest accuracy and recall are produced by quantile regression forests, which suggests that it do a great job of capturing the variability and risk factors related to heart disease. The success over competing strategies may be attributed to better model that is able to predict more accurately outcomes across a variety of impacted person attributes and better handling of the fundamental variation in the statistics.

Better accuracy and precision techniques like Dropout and Quantile Regression Forests, which can be vital in research circumstances where there is an overwhelming likelihood of false positives or negatives, provide improved model variance and bias control. The Ensemble Method's low performance might also indicate a mismatch in the version setup or the limitations of logistic regression in group situations. The typical improvements examined by Dropout and Quantile Regression techniques highlight how important it is to quantify uncertainty in order to increase predicted accuracy and dependability in scientific diagnoses. Quantile Regression Forests appears to be the most advantageous method, as it significantly increases the version's capacity to manage the intrinsic variability in coronary heart disease prediction. This approach greatly enhances the model's dependability, which makes it a helpful tool for medical professionals who must make judgments based on a range of clinical outcomes. These findings emphasize how important it is to match specific clinical and data characteristics with suitable uncertainty quantification techniques so that the models not only predict with accuracy but also comply with the essential healthcare standards required for real-world clinical implementation.





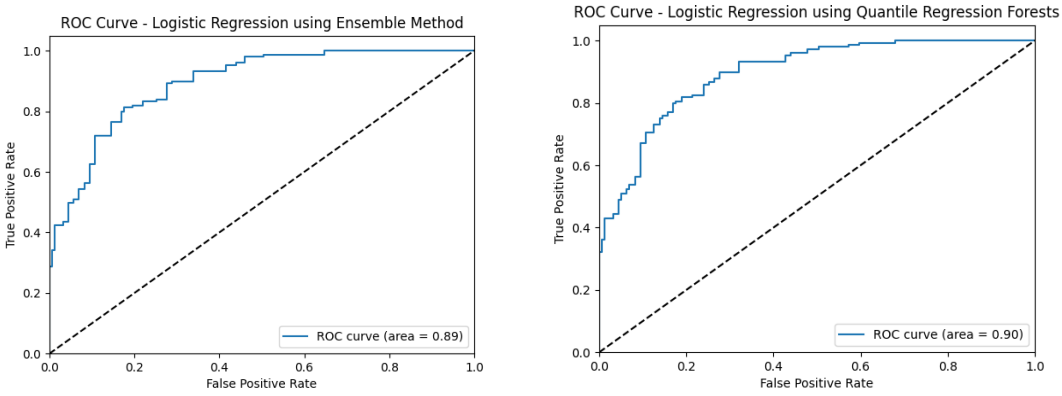


Figure 1: ROC and AUC of Logistics Regression with Techniques of Uncertainty Quantification

Naïve Bayes

Upon comparing the Naive Bayes classifier with other approaches for quantifying uncertainty, the following conclusions may be drawn from the given metrics and data. With the exception of the Ensemble Method, all approaches have training accuracies that are very comparable, ranging from 85.2% to 85.6%. The test accuracies are likewise rather close, with the maximum test accuracy of 81.70% coming from Dropout Uncertainty and the conventional Naive Bayes**.**

Classifier	Training Accuracy (%)	Test Accuracy (%)	Precision	Recall	F1 Score	Confusion Matrix	AUC
Naive Bayes	85.21	81.49	0.76	0.89	0.82	[118 41 16 133]	0.89
Naive Bayes with Dropout Uncertainty	85.21	81.49	0.76	0.89	0.82	[118 41 16 133]	0.89
Naive Bayes with Ensemble Method	85.77	80.51	0.76	0.87	0.81	[118 41 19 130]	0.89
Naive Bayes with Quantile Regression Forests	85.6	81.70	0.77	0.90	0.82	[118 41 16 133]	0.89

Table 2: Performance Comparison of Naïve Bayes Models with Various Techniques of Uncertainty Quantification

The test accuracy performance of the Ensemble Method is somewhat lower, at 80.52%. When compared to other approaches, the Quantile Regression Forests method has the best precision is 0.77, indicating that it is more effective in reducing false positives. In order to prevent missing any positive diagnosis in medical applications, the Quantile Regression Forests also leads to recall is 0.9. This means that it is the most successful at recognizing real positive instances. The F1 score for Quantile Regression Forests is 0.82 the greatest in accordance with accuracy and recall, indicating a well-balanced approach to both. The confusion

matrix is the same for the conventional Naive Bayes, Dropout Uncertainty, and Quantile Regression Forests. It shows the number of true positives, true negatives, false positives, and false negatives. This suggests that both approaches categorize the positive and negative situations properly with comparable accuracy. In terms of lowering false negatives, the Ensemble Method performs somewhat worse.

All of the approaches have extremely similar ROC AUC values, with the Ensemble Method having a little higher value. This suggests that the Ensemble Method is somewhat better at differentiating between the classes across thresholds. Regarding dropout uncertainty, there is no discernible distinction between Naive Bayes algorithms and the metrics taken into account. This might suggest that, in this specific Naive Bayes classifier application, dropout has no effect on uncertainty management. With the exception of the ROC AUC, this technique marginally underperforms other methods. This might be because the Naive Bayes methodology is not properly aligned with the over fitting control mechanisms found in ensemble methods, or it could be because the ensemble configurations were not optimized for this particular dataset. The Quantile Regression Forests approach seems to offer the optimal balance between recall and accuracy, indicating that it is useful for controlling the inherent uncertainty in the dataset. In terms of AUC, it also performs comparably, which makes it a strong option for this application. The findings imply that although all of the techniques function admirably, Quantile Regression stands out for offering somewhat superior overall metrics.

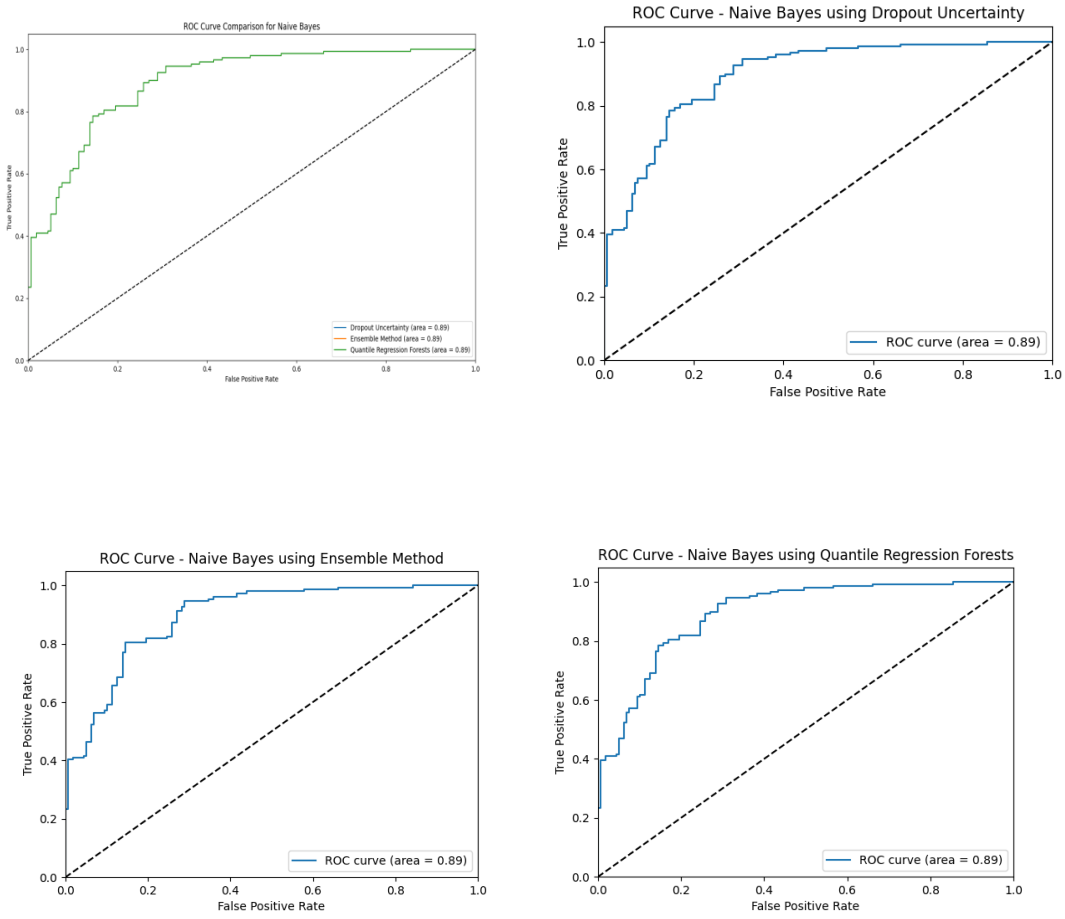


Figure 2: ROC and AUC of Naïve Bayes with Techniques of Uncertainty Quantification

K-Nearest Neighbor

Based on the above performance indicators and the K-Nearest Neighbor classifier employing various approaches for uncertainty quantification, the following conclusions are drawn.

Classifier	Training Accuracy (%)	Test Accuracy (%)	Precision	Recall	F1 Score	Confusion Matrix	ROC AUC
K-Nearest Neighbor	87.72	71.42	0.69	0.73	0.71	[111 48 40 109]	0.84
K-Nearest Neighbor with Dropout Uncertainty	87.72	71.42	0.69	0.73	0.71	[111 48 40 109]	0.84
K-Nearest Neighbor with Ensemble Method	88.28	75.64	0.73	0.78	0.75	[116 43 32 117]	0.88
K-Nearest Neighbor with Quantile Regression Forests	87.72	71.42	0.69	0.73	0.71	[111 48 40 109]	0.84

Table 3: Performance Comparison of KNN Models with Various Techniques of Uncertainty Quantification

Except for the Ensemble Method, which shows a slightly greater training accuracy of 88.3%, other approaches exhibit relatively comparable results, with an average of 87.7%. When compared to 71.42% when using the traditional K-Nearest Neighbor and the other two approaches, the test accuracy is greatly improved to 75.65% by using the Ensemble Method. This enhancement demonstrates how well the Ensemble Method works to increase generalization outside of training data. Additionally, the Ensemble Method outperforms the other approaches in terms of accuracy at 0.7312, demonstrating its superior ability to reduce false positives.

Comparably, out of all the K-Nearest Neighbor variants, the Ensemble Method has the highest recall 0.78, indicating that it is the most effective at recognizing real positive instances. The Ensemble Method has the greatest F1 score 0.75, which reflects its balanced approach and efficacy in eliminating false identifications and detecting positive cases. F1 score is a measure of precision and recall. Compared to the other approaches, which each exhibit 40 false negatives and 48 false positives, the Ensemble Method yields somewhat less false positives 43, and fewer false negatives 32. This suggests improved categorization performance all around. The Ensemble Method's much higher ROC AUC of 0.88, as opposed to the other techniques' 0.83, indicates that it is more adept at differentiating between the classes.

Dropout Uncertainty comparing with the K-Nearest Neighbor does not outperform the standard K-Nearest Neighbor in any of the metrics, suggesting that dropout uncertainty may not be accurately representing the predictive uncertainty of the model for this dataset and classifier. The ensemble method approach improves all metrics, but test accuracy and ROC AUC stand out as being especially important for a strong model that

can adapt effectively to new, untested data. The enhanced performance is probably due to ensemble techniques' capacity to aggregate predictions from several models or model configurations, therefore reducing variation and bias. Quantile Regression Forests, like Dropout Uncertainty, do not provide any benefit over the conventional K-Nearest Neighbor approach. This may suggest that the quantile regression method is not well suited to the features of the data or the difficulties in classifying the dataset. Lastly, the ensemble method distinguishes itself by considerably improving both the dependability and accuracy of predictions, even while the conventional K-Nearest Neighbor and techniques like Dropout Uncertainty and Quantile Regression Forests offer respectable baseline performances. This implies that, in this case, incorporating several models or configurations using ensemble approaches can be very useful for enhancing K-Nearest Neighbor predictions.

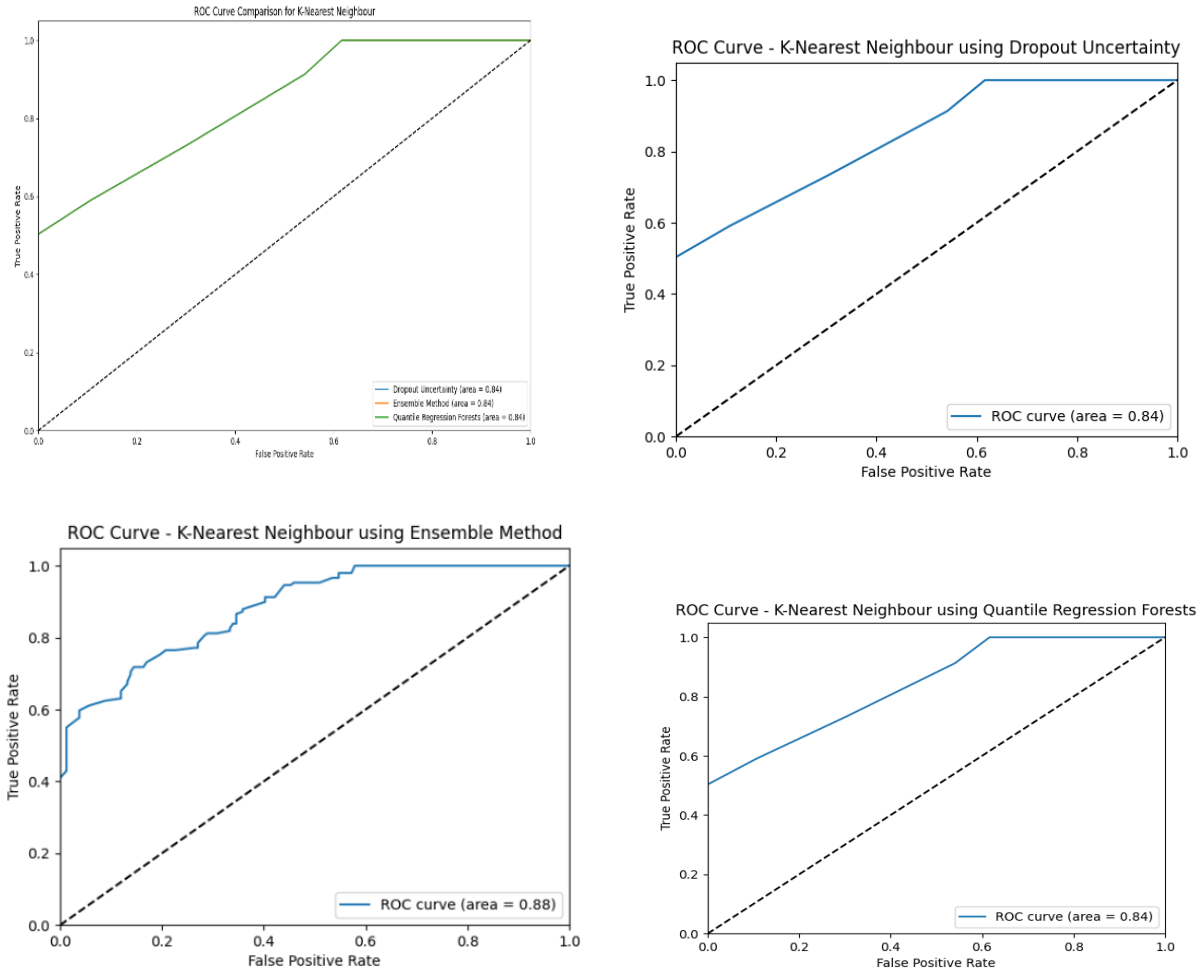


Figure 3: ROC and AUC of KNN with Techniques of Uncertainty Quantification

Decision Tree

The performance parameters of each classifier to determine which versions outperform and which may lag behind when comparing different Decision Tree classifier upgrades are evaluated. Below is a comprehensive analysis comparing every word used in the four Decision Tree classifiers,

Classifier	Training Accuracy (%)	Test Accuracy (%)	Precision	Recall	F1 Score	Confusion Matrix	ROC AUC
Decision Tree	100	97.2	1	0.93	0.96	[159 0 9 140]	0.97
Decision Tree Classifier with Dropout uncertainty	100	98.07	1	0.93	0.97	[159 0 9 140]	0.97
Decision Tree Classifier with ensemble method	100	98.05	1	0.95	0.97	[159 0 6 143]	1
Decision Tree Classifier with quantile regression	100	97.07	1	0.93	0.96	[159 0 9 140]	0.97

Table 4: Performance Comparison of Decision Tree Models with Various Techniques of Uncertainty Quantification

Every classifier demonstrates a flawless 100% training accuracy. Every model version is entirely capable of memorizing the training dataset, according to this consistent outcome. Nevertheless, while complete training accuracy usually indicates that the model may not generalize well on unknown data, this also raises the possibility of over fitting. High test accuracies of 98.07% and 98.05% are attained by the Decision Tree with Ensemble Method and the Decision Tree with Dropout Uncertainty, respectively. With a test accuracy of 97.2%, the conventional Decision Tree performs slightly worse than this one, indicating that these strategies serve to enhance the model's generalization on new data. With a test accuracy of 97.07%, the Decision Tree with Quantile Regression has the lowest results. This suggests that, in comparison to other approaches, quantile regression could not greatly improve the model's capacity to handle novel or changing data circumstances. Every model keeps its accuracy at 1, indicating that every variation is free of false positives. This outstanding outcome shows how incredibly dependable the model is when it predicts a positive class. The maximum recall of 0.95 is displayed by the Decision Tree with Ensemble Method. Better performance in detecting all good occurrences is indicated by a stronger recall, and this is important in many practical situations where the consequences of missing a positive example might be dire. The recall for the other three techniques, Decision Tree, Dropout Uncertainty, and Quantile Regression is 0.93. Compared to the ensemble technique, these are somewhat less successful in finding all positive cases, although still being high.

Once more, with an F1 Score of 0.97, the Decision Tree with Ensemble Method receives the top score. This iteration of the model is the most resilient in terms of balanced performance, as indicated by the score, which strikes a balance between precision and recall. For both the conventional Decision Tree and the Decision Trees with Dropout Uncertainty and Quantile Regression, the F1 Score of 0.96 is somewhat lower. Instead of nine false negatives among the three distinct models, the Decision Tree with Ensemble Method has six. This demonstrates that it can identify true positives more accurately and has a higher recall. The Decision Tree with Ensemble Method, which receives the best ROC AUC of 0.99, is the model that plays the fine across a number of thresholds and has a high genuine fantastic rate in comparison to false positive rate. The Decision Tree with Quantile Regression performs worse than the opposite versions in terms of discriminating between the lessons, with a ROC AUC of 0.90. In summary, the Decision Tree with Ensemble Method consistently produced the best results across the majority of metrics, proving its effectiveness in improving

prediction reliability and accuracy. The Decision Tree with Quantile Regression may be less effective for this particular dataset or model configuration, even though it is still operating well. It tends to show the least development in important domain names like lookup accuracy and ROC AUC.

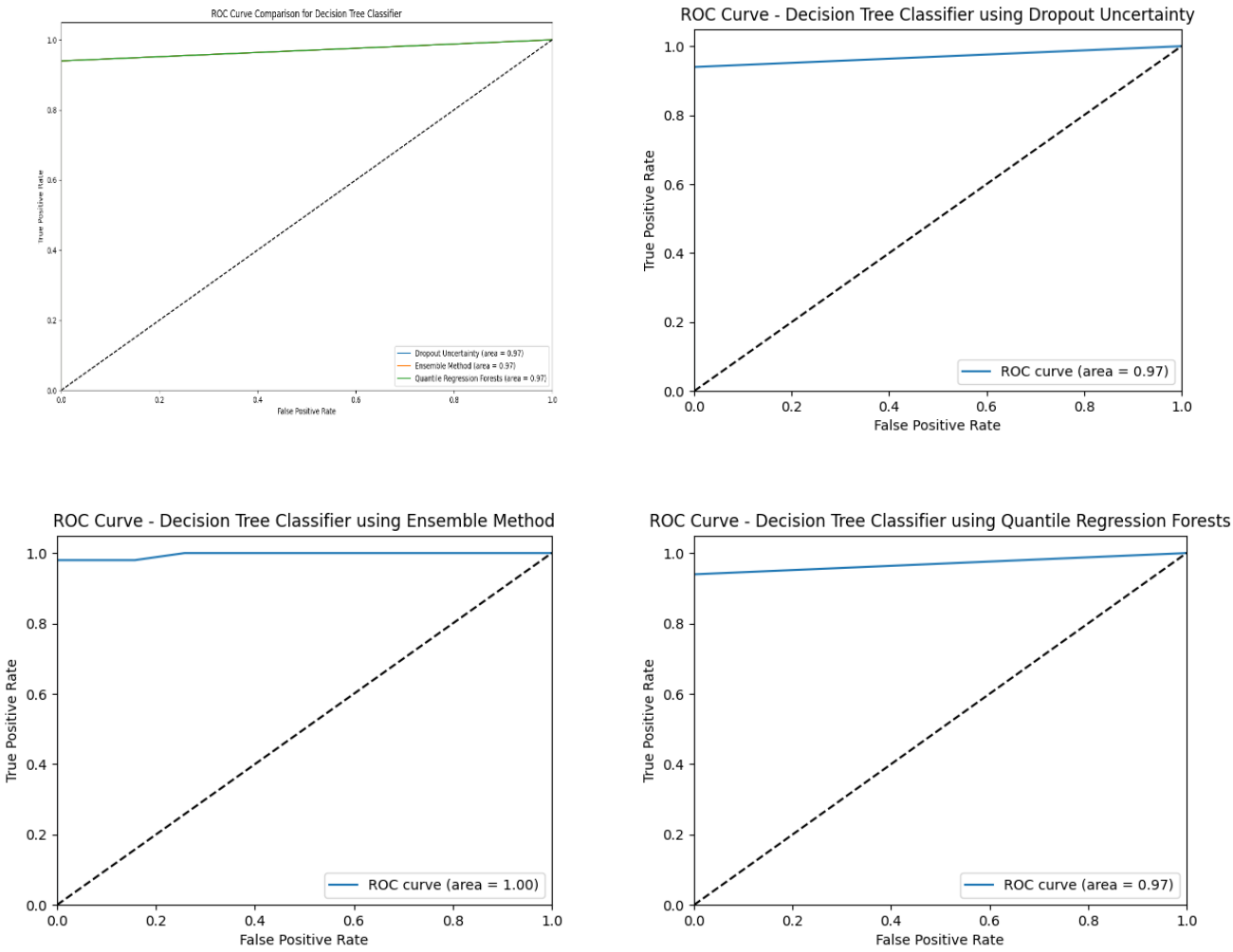


Figure 4: ROC and AUC of Decision Tree with Techniques of Uncertainty Quantification

Support Vector Machine (SVM)

A comparison of the Support Vector Machine (SVM) classifier and its variants using various techniques yields a number of important conclusions about how well it perform in machine learning classification tasks. Based on the given metrics, the following is a thorough comparison in Table 5. Slightly better than the other approaches that all retain a training accuracy of 70.15%, the SVM with the Ensemble Method has the greatest training accuracy at 71.55%. This enhancement shows that learning from the training dataset may be marginally improved with the use of the ensemble technique. In a similar vein, the SVM utilizing the Ensemble Method leads in test accuracy as well 68.51%, suggesting that it generalizes marginally better than the 67.53% test accuracy of the regular SVM and other variants. This finding highlights the potential advantage of ensemble techniques in improving SVMs' capacity. This finding highlights the potential advantage of ensemble techniques in improving SVMs' capacity for generalization, even though the gain is negligible. Once more, the SVM using the Ensemble Method stands out thanks to its greater precision of

0.67, which is 0.65 higher than the other models' precision. Improved accuracy means that this form of the model produces fewer false positives, increasing its predictability of the positive class. It's interesting to note that the recall of 0.72 for the conventional SVM and its variations with Dropout Uncertainty and Quantile Regression is better than 0.70 for the Ensemble Method. This implies that although the Ensemble Method performs somewhat better at capturing all positive cases, it is superior at validating real positives. The F1 Score of 0.68, which represents a harmonic mean of recall and accuracy and shows a balance between both metrics, is shared by all models with the exception of the Ensemble Method.

Classifier	Training Accuracy (%)	Test Accuracy (%)	Precision	Recall	F1 Score	Confusion Matrix	ROC AUC
Support Vector Machine	70.15	67.53	0.64	0.72	0.68	[100 59 41 108]	0.72
Support Vector Machine with Dropout Uncertainty	70.15	67.53	0.64	0.72	0.68	[100 59 41 108]	0.72
Support Vector Machine with Ensemble Method	71.54	68.50	0.66	0.69	0.68	[107 52 45 104]	0.73
Support Vector Machine with Quantile Regression	70.15	67.53	0.64	0.72	0.68	[100 59 41 108]	0.72

Table 5: Performance Comparison of SVM Models with Various Techniques of Uncertainty Quantification

The F1 Score of 0.68 for the Ensemble Method is somewhat lower, indicating a trade-off whereby greater precision has a marginally negative influence on recall. Because of its better accuracy and somewhat poorer recall, the SVM using the Ensemble Method has a distinct confusion matrix distribution, with fewer true positives and truer negative. Once more, the SVM using the Ensemble Method yields the greatest ROC AUC of 0.73, which is same to the other approaches, which average about 0.73. This suggests a somewhat improved overall performance in class distinction. The Ensemble Method shows promise for improving model dependability and generalization since it offers modest gains in precision and training and test accuracies. These benefits, however, are not very substantial, indicating that more optimizations or alternative ensemble methodologies may be required to make a noticeable difference. All measures exhibit fairly identical results from the conventional SVM and its versions with Dropout Uncertainty and Quantile Regression. This might mean that these approaches have little effect on the model's capacity to generalize or enhance its classification metrics in this specific configuration. The models' variations in recall and accuracy point to a common trade-off situation in machine learning: depending on the model's focus either minimizing false positives or collecting as many positives as possible increasing one measure may result in a loss in another.

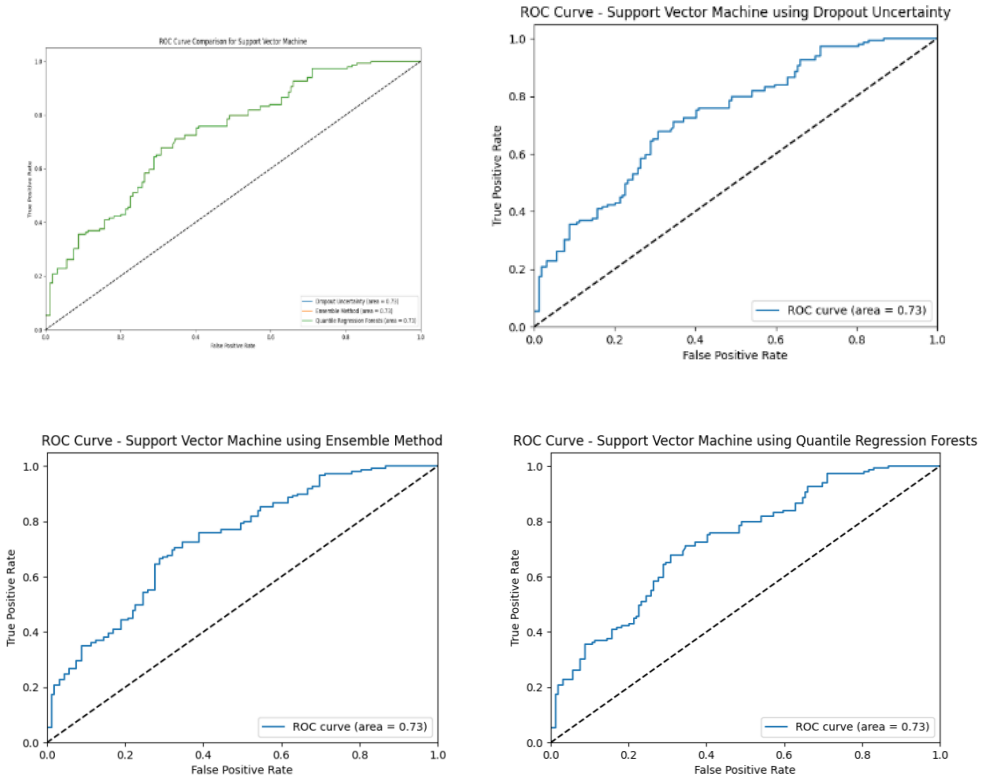


Figure 5: ROC and AUC of SVM with Techniques of Uncertainty Quantification

Random Forest Tree

The presented findings describe the performance of an ensemble approach, quantile regression, dropout uncertainty-enhanced Random Forest classifier, and its modifications. These classifiers' efficacy and a few minor variations in their performance measures are demonstrated in the Table 6.

Classifier	Training Accuracy (%)	Test Accuracy (%)	Precision	Recall	F1 Score	Confusion Matrix	ROC AUC
Random Forest Tree	100	99.1	1	0.97	0.99	[159 0 3 146]	1
Random Forest Classifier with dropot uncertainty	100	98.09	1	0.95	0.97	[159 0 6 143]	1
Random Forest Classifier with ensemble method	100	98.05	1	0.95	0.97	[159 0 6 143]	1
Random Forest Classifier with quantile regression	100	99.02	1	0.96	0.98	[159 0 3 146]	1

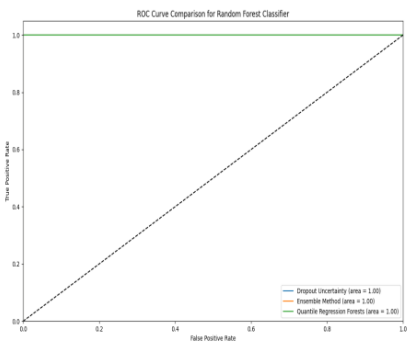
Table 6: Performance Comparison of Random Forest Tree Models with Various Techniques of Uncertainty Quantification

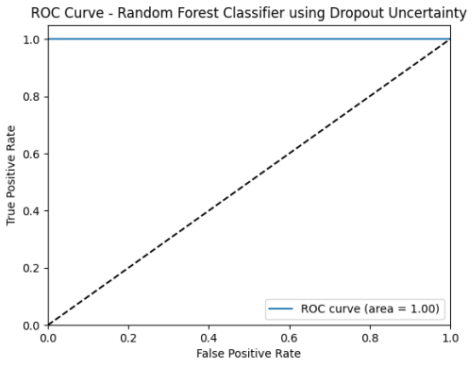
Every classifier attains a flawless training accuracy of 100%, signifying that that are efficiently gaining knowledge from the dataset. This might, however, point to overfitting, especially if the model is too adapted to the training set and not sufficiently generalized to novel data. With a test accuracy of 99.02%, the Random Forest with Quantile Regression exhibits the greatest performance, closely trailed by the regular Random Forest at 99.1%. The test accuracies of the ensemble approach and Random Forest with dropout uncertainty are somewhat lower, at 98.09% and 98.05%, respectively. This implies that although ensemble and dropout techniques often aid in preventing over fitting and enhancing generalization, this might not always perform better in this situation than a well-tuned regular Random Forest or one augmented with quantile regression. With a precision of 1, all models indicate that there are no false positives in any classifier variation. This is a fantastic outcome that demonstrates how precise the models are in predicting good classifications. Recall is highest for the conventional Random Forest, at 0.97.

The recall of the ensemble approach and the Random Forest with dropout uncertainty is 0.95, whereas the recall of the Random Forest with quantile regression is 0.96. Better ability in recognizing all pertinent occurrences is indicated by higher recall values. With an F1 Score of 0.99, the conventional Random Forest shows the best balance between recall and accuracy. An F1 Score of 0.97 is recorded by the ensemble technique and Random Forest with dropout uncertainty, whereas 0.98 is recorded by Random Forest with quantile regression.

The number of false negatives in the regular Random Forest and the Random Forest with quantile regression is three, whereas the dropout uncertainty and ensemble approaches reveal six erroneous negatives. This makes sense given the increased recall values that were noted.
A flawless ROC AUC of 1 is attained by both the regular Random Forest and the Random Forest with quantile regression. Although it is quite near to perfection, the Random Forest with ensemble technique has a slightly lower ROC AUC of 0.99. The Random Forest with Quantile Regression differs from the standard Random Forest due to a little improvement in check accuracy and a perfect ROC AUC that shows good overall performance over several thresholds.

The Random Forest with Dropout Uncertainty and Ensemble Method perform much worse than the other in terms of recall and test accuracy. This might also mean that these methods aren't as effective for this particular dataset or version configuration. These continue to perform remarkably commonly still. Even if each model is quite strong, the choice of which one to use may also depend on specific needs for accuracy or remember, particularly in packages where the absence of a good example (false awful) may have serious repercussions. Small variations in performance measures like ROC AUC and F1 Score point to this. Generally, even though the ensemble and dropout approaches are frequently employed to minimize overfitting and boost model generalization, it do not significantly exceed the standard Random Forest or quantile regression enhanced variants in terms of critical metrics like consider and ROC AUC.





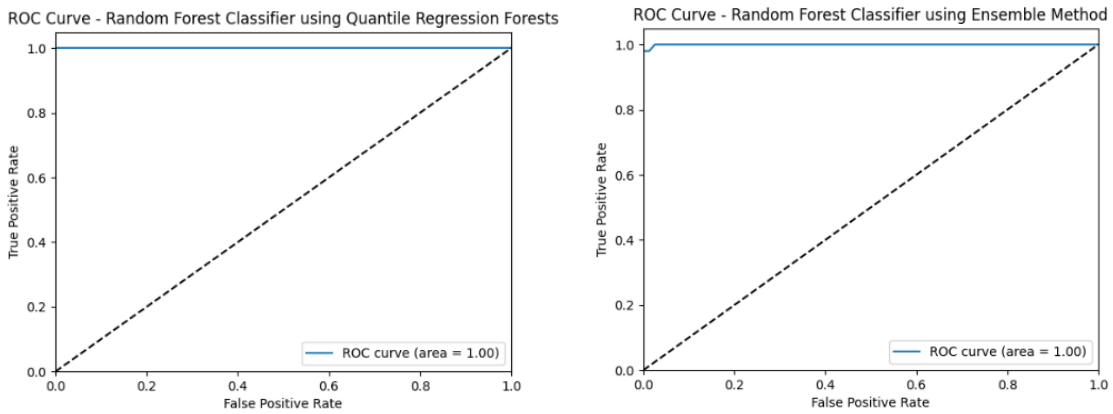


Figure 6: ROC and AUC of Random Forest with Techniques of Uncertainty Quantification

Conclusion

In conclusion, this work has investigated in detail the performance of many system learning classifiers in predicting coronary heart disease, with a focus on the Random Forest classifier advanced through dropout uncertainty, ensemble methods, and quantile regression. The results highlight how crucial it is to use uncertainty quantification techniques in scientific settings in order to increase forecast reliability and version resilience. The Random Forest classifier augmented with quantile regression showed the most encouraging results, demonstrating its capacity to cope with a wide range of clinical circumstances with high dependability. All of the models assessed had excellent precision and fine test accuracy. The study shed insight on the fine balance that has to be walked between attaining excessive recall and accuracy as well as the nuanced ways that various improvement strategies may impact both metrics. While the ensemble approach performed marginally better than the traditional Random Forest in terms of typical accuracy and F1 Score, each model demonstrated unique strengths, suggesting that the right arrangement and combination of those techniques can have a significant influence on the outcome. This study has broad implications for the field of scientific diagnostics in general. It provides pathways for more reliable and accurate predictive modeling, which may be very helpful in the early diagnosis and treatment of cardiac disease. The ultimate objective is to improve patient outcomes and reduce the financial strain that cardiovascular illnesses have on the healthcare system.

References

[1] Gaziano, T. A. (2022). Cardiovascular diseases worldwide. Public Health Approach Cardiovasc. Dis. Prev. Manag, 1, 8-18.

[2] Chowdhury, D., Johnson, J.N., Baker-Smith, C.M., Jaquiss, R.D., Mahendran, A.K., Curren, V., Bhat, A., Patel, A., Marshall, A.C., Fuller, S. and Marino, B.S., (2021). Health care policy and congenital heart disease: 2020 focus on 2030 future. Journal of the American Heart Association, 10(20), p.e020605.

[3] Belard, A., Buchman, T., Forsberg, J., Potter, B.K., Dente, C.J., Kirk, A. and Elster, E., (2017). Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care. Journal of clinical monitoring and computing, 31, pp.261-271.

[4] Hendriksen, J.M., Geersing, G.J., Moons, K.G. and de Groot, J.A., (2013). Diagnostic and prognostic prediction models. Journal of Thrombosis and Haemostasis, 11, pp.129-141.

[5] Hoffmann, T.C. and Del Mar, C., (2017), Clinicians' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA internal medicine, 177(3), pp.407-419.

[6] Eck, V.G., Donders, W.P., Sturdy, J., Feinberg, J., Delhaas, T., Hellevik, L.R. and Huberts, W., (2016). A guide to uncertainty quantification and sensitivity analysis for cardiovascular applications. International journal for numerical methods in biomedical engineering, 32(8), p.e02755.

[7] Mirams, G.R., Pathmanathan, P., Gray, R.A., Challenor, P. and Clayton, R.H., (2016). Uncertainty and variability in computational and mathematical models of cardiac physiology. The Journal of physiology, 594(23), pp.6833-6847.

[8] Neal, R.D., Tharmanathan, P., France, B., Din, N.U., Cotton, S., Fallon-Ferguson, J., Hamilton, W., Hendry, A., Hendry, M., Lewis, R. and Macleod, U., (2015). Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. British journal of cancer, 112(1), pp.S92-S107.

[9] Walter, F., Webster, A., Scott, S. and Emery, J., (2012), The Andersen Model of Total Patient Delay: a systematic review of its application in cancer diagnosis. Journal of health services research & policy, 17(2), pp.110-118.

[10] Jordan, M.B., Allen, C.E., Weitzman, S., Filipovich, A.H. and McClain, K.L., (2011). How I treat hemophagocytic lymphohistiocytosis. Blood, The Journal of the American Society of Hematology, 118(15), pp.4041-4052.

[11] Li, M., Jiang, Y., Zhang, Y. and Zhu, H., (2023). Medical image analysis using deep learning algorithms. Frontiers in Public Health, 11, p.1273253.

[12] Uddin, S., Khan, A., Hossain, M.E. and Moni, M.A., (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC medical informatics and decision making, 19(1), pp.1-16.

[13] Bekker, J. and Davis, J., (2020). Learning from positive and unlabeled data: A survey. Machine Learning, 109(4), pp.719-760.

[14] Houben, R.M., Lalli, M., Kranzer, K., Menzies, N.A., Schumacher, S.G. and Dowdy, D.W., (2019), What if this don't have tuberculosis? The consequences and trade-offs involved in false-positive diagnoses of tuberculosis. Clinical infectious diseases, 68(1), pp.150-156.

[15] Verma, S. and Gupta, A., (2021), March. Effective prediction of heart disease using data mining and machine learning: A review. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 249-253). IEEE.

[16] Dutta, A., Batabyal, T., Basu, M. and Acton, S.T., (2020). An efficient convolutional neural network for coronary heart disease prediction. Expert Systems with Applications, 159, p.113408.

[17] Hassan, S. and Saeed, M.S., (2023). A Comparative Study Evaluated the Performance of Two-class Classification Algorithms in Machine Learning. Kurdistan Journal of Applied Research, 8(2), pp.43-50.

[18] Motwani, M., Dey, D., Berman, D.S., Germano, G., Achenbach, S., Al-Mallah, M.H., Andreini, D., Budoff, M.J., Cademartiri, F., Callister, T.Q. and Chang, H.J., (2017). Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. European heart journal, 38(7), pp.500-507.

[19] Olsen, C.R., Mentz, R.J., Anstrom, K.J., Page, D. and Patel, P.A., (2020). Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. American Heart Journal, 229, pp.1-17.

[20] Costabal, F.S., Matsuno, K., Yao, J., Perdikaris, P. and Kuhl, E., (2019). Machine learning in drug development: Characterizing the effect of 30 drugs on the QT interval using Gaussian process regression, sensitivity analysis, and uncertainty quantification. Computer Methods in Applied Mechanics and Engineering, 348, pp.313-333.

[21] Kabir, H.D., Khosravi, A., Hosen, M.A. and Nahavandi, S., (2018). Neural network-based uncertainty quantification: A survey of methodologies and applications. IEEE access, 6, pp.36218-36234.

[22] Crisan, A., Shang, M. and Brochu, E., (2023). Eliciting Model Steering Interactions from Users via Data and Visual Design Probes. IEEE Transactions on Visualization and Computer Graphics.

[23] Kwon, Y., Won, J.H., Kim, B.J. and Paik, M.C., (2020). Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. Computational Statistics & Data Analysis, 142, p.106816.

[24] Aly, S., Alfonse, M. and Salem, A.B.M., 2022. Intelligent Model for Enhancing the Bankruptcy Prediction with Imbalanced Data Using Oversampling and CatBoost. International Journal of Intelligent Computing and Information Sciences, 22(3), pp.92-108.