

### **Research Article**

## Machine Learning Modeling of Disease Treatment Default: A Comparative Analysis of Classification Models

# Michael Owusu-Adjei , James Ben Hayfron-Acquah , Frimpong Twum , and Gaddafi Abdul-Salaam

Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Correspondence should be addressed to Michael Owusu-Adjei; mowusuadjei@st.knust.edu.gh

Received 14 November 2022; Revised 21 December 2022; Accepted 23 December 2022; Published 3 January 2023

Academic Editor: Chandrabose Selvaraj

Copyright © 2023 Michael Owusu-Adjei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Generally, treatment default of diseases by patients is regarded as the biggest threat to favourable disease treatment outcomes. It is seen as the reason for the resurgence of infectious diseases including tuberculosis in some developing countries. Sadly, its occurrence in chronic disease management is associated with high morbidity and mortality rates. Many reasons have been adduced for this phenomenon. Exploration of treatment default using biographic and behavioral metrics collected from patients and healthcare providers remains a challenge. The focus on contextual nonbiomedical measurements using a supervised machine learning modeling technique is aimed at creating an understanding of the reasons why treatment default occurs, including identifying important contextual parameters that contribute to treatment default. The predicted accuracy scores of four supervised machine learning algorithms, namely, gradient boosting, logistic regression, random forest, and support vector machine were 0.87, 0.90, 0.81, and 0.77, respectively. Additionally, performance indicators such as the positive predicted value score for the four models ranged between 98.72%-98.87%, and the negative predicted values of gradient boosting, logistic regression, random forest, and support vector machine were 50%, 75%, 22.22%, and 50%, respectively. Logistic regression appears to have the highest negative-predicted value score of 75%, with the smallest error margin of 25% and the highest accuracy score of 0.90, and the random forest had the lowest negative predicted value score of 22.22%, registering the highest error margin of 77.78%. By performing a chi-square correlation statistic test of variable independence, this study suggests that age, presence of comorbidities, concern for long queuing/waiting time at treatment facilities, availability of qualified clinicians, and the patient's nutritional state whether on a controlled diet or not are likely to affect their adherence to disease treatment and could result in an increased risk of default.

#### 1. Introduction

Statistical data compiled [1] from all the regional and district hospitals except the two leading teaching hospitals (Korle-Bu Teaching Hospital and Komfo Anokye Teaching Hospital) in Ghana show a depressing outlook on the incidence of hypertension and other cardiovascular diseases that has resulted in increase patient mortality and morbidity rates with devastating consequence. This has been described as a consequence of patient nonadherence/default or noncompliance to prescribed treatment appointments. Hypertension and associated cardiovascular diseases were listed as the leading cause of deaths according to the report in [1] among the population in 2017 and ranked third as the biggest cause of patient admission in hospitals countrywide. These statistical descriptions are also confirmed in [2], which attributes 13 percent of total global deaths to the incidence of hypertension and estimates them to be the largest risk factor for deaths globally.

Cardiovascular diseases including hypertension are chronic diseases that require long-term treatment management. Recent reports healthcare personnel [1, 2] confirming increasing mortality rates among suffering patients require a critical look at treatment adherence together with default risks by considering both biographic and contextual behavioral variables among patients and healthcare givers. The importance of behavioral analysis in healthcare is underscored by the authors of [3], whose work affirms that patient behaviors are important considerations for disease treatment including the determination of disease causality. Further studies to determine the impact of behavior on patients stipulate that performance behavior of patients is aimed at preventing disease occurrence, detecting onset of diseases, and improving disease treatment outcomes, but admittedly, the anticipated outcomes are also influenced by the behavior of healthcare professionals [4]. Their definition of behavior by health care professionals emphasizes on patient needs. behavior that takes into account the needs of patients [4]. Additionally, studies regarding healthcare provider behavior and its impact on patients, healthcare working environments, colleague workers, etc., are emphasized by the author of [5], with a tacit admission that there are negative outcomes of disrespectful behavior, some of which cause recipients to experience fear and feel isolated, including many others. The impact of negative consequences of these behavioral traits by healthcare providers on patient treatment outcomes shows that disease treatment default by patients can be established in context. This includes the consideration of contextual variables about both parties (healthcare providers and patients). Applying predictive algorithms with the capability of showing hidden patterns of information can assist in this endeavor.

Predictive algorithms have become powerful instruments for businesses (big and small) as a competitive tool. In mortgage financial decisions, including assessing payment behavior of homeowners [6], predictive analytics have been employed to study and understand the behavior of homeowners. The other use of predictive algorithms is providing insights into future human behavior based on present or available information [7]. In the healthcare industry, predictive algorithms have been used in the area of classification problems such as predicting the patient's nonattendance to a scheduled appointment clinic [8]. Further use of predictive algorithms is associated with early detection and diagnoses of diseases for preventive medical care due to increased medical treatment costs together with increasing morbidity and mortality rates [9]. Advances in technology, coupled with increased production of data within healthcare systems, have heightened research interests in healthcare applications for knowledge discovery and insights into patterns of change [10]. Predictive machine learning algorithms have therefore become useful in providing user-centered explanations of the factors that lead to an increased risk of adverse outcomes in healthcare settings [11].

There have been several predictive studies of disease diagnosis and detection on hypertension and other cardiovascular diseases predominantly using contextual variables such as biomedical and biographic metrics. Behavior is mentioned among the five broad determinants of health [12]. However, studies requiring the use of behavior together with either biographic or biomedical metrics for disease treatment in this regard have been limited. This research therefore explores contextual biographic and behavioral traits of both patients and healthcare providers towards disease treatment default using empirical data from a real context with reliance on nonbiomedical variable metrics.

#### 2. Related Work

A comparative analysis of different machine learning modeling techniques within the healthcare delivery system has been examined in various studies and related research works. This section discusses prediction accuracy scores of machine learning techniques used within the context of healthcare practice to highlight the significance of machine learning types in classification-based problem domains with biomedical measurements. A study [13] to predic biabetes disease with biomedical and biographic variables such as age, education, systolic and diastolic blood pressure, body mass index, direct and cholesterol using learning algorithms namely; Naive Bayes, Decision Tree, Adaboost and Random Forest showed auc score of 0.95 for Logistic regression and Random forest. This study therefore concludes that the combination of logistic regression with the random forest showed superior performance and could be used in predicting diabetic patients.

Similar predictive studies [14] with the disease prediction framework using machine learning techniques for diabetes healthcare with a combination of biographic and biomedical data such as age, blood pressure, glucose, and insulin body mass index with algorithms such as K-nearest neighbor (KNN), support vector machine (SVM), logistic regression, and random forest concluded that logistic regression had superior performance with a receiver operating characteristic curve (ROC) score of 86%. Similarly, an analysis of decision trees for diabetes prediction using both biomedical and biographic variables such as gender, plasma, insulin, glucose, body mass index, and blood pressure with decision tree algorithms such as leaf area density (LAD), naïve Bayes, and genetic j48 also showed predicted accuracy scores of 88%, 92%, and 95.8%, respectively [15].

Further studies on detailed analysis of kidney and heart disease prediction with machine learning using two distinct datasets consisting of both biomedical and biographic variables such as age, blood pressure, specific gravity, sugar, albumin, red blood cells, puss cells, pus cell clumps, bacteria, blood glucose level, blood urea, sex, education, current smoking status, cigarettes smoked per day, BP medicines, and prevalent smoking also showed accuracy scores of 100% for chronic kidney disease and 85% for heart disease. Machine learning algorithms such as logistic regression, Knearest neighbor, naïve Bayes, support vector machine, and random forest [16] also produced accuracy scores of 100% for chronic kidney disease and 85% for heart disease based on a support value of 750 for all the algorithms.

Further studies for heart disease classification predicted by the authors of [17], which aim to explore feature selections with a combination of chi-squared feature selection methods and the Bayes net algorithm, achieved an accuracy score of 85.00% using both biomedical and biographic variables such as age, sex, measuring values of fasting blood sugar, resting blood pressure, serum cholesterol, resting electrocardiographic reports, maximum heart rates, and number of major vessels colored by fluoroscopy.

A personalized modeling and prediction approach with internet-connected smart devices for data generation [18] used weighted voting logistic regression and random forest machine learning techniques for type 2 diabetes prediction and produced a model accuracy score of 0.884 with both biomedical and biographic metrics such as age, gender, body mass index, cholesterol level, marital status, employment status, and income level.

Further predictive studies on the presence or absence of heart disease using techniques such as the backpropagation multilayer perceptron [19] showed an improved model performance accuracy score of 96.30% with features such as age, sex, fasting blood sugar, and resting blood sugar. Using several machine learning techniques for supervised and unsupervised learning in diabetes research predictions [10], it was concluded that the use of support vector machines proved to be the most successful while using clinical datasets with features extracted from demographics, diagnoses, disease comorbidities and symptoms, medications, laboratory measurements, and other procedures. Furthermore, a predictive analysis of diabetic complications [20] using naive Bayes tree, C4.5 decision tree-based classification, and k-means clustering techniques with features such as age, gender, body mass index, family history of diabetes, blood pressure, duration of onset, and blood glucose level showed an overall model accuracy of 68%. A similar study to identify correlated variables such as demographic, clinical, and healthcare resource utilization variables for the diagnosis of diabetic peripheral neuropathy using the random forest algorithm/technique achieved an ROC model performance of 0.824 and a model accuracy of 89.6% with a 95% confidence interval [21]. A similar study with variables such as the number of prior convictions, age, type of index offence, diversity of criminal history, and substance abuse to predict general criminal recidivism in mentally disordered offenders using the random forest technique also produced an AUC score of 0.90 [22]. In diagnosing heart disease for diabetic patients using variables such as age, sex, blood pressure, and blood sugar in predicting the chances of a diabetic patient developing heart disease, naïve Bayes and support vector machines showed significant prediction accuracy [23]. Alternatively, a study [24] to identify the optimal model that predicts HBsAg seroclearance of patients suffering from chronic hepatitis B with selected variables such as age, gender, family history, body mass index, and drinking history used four machine learning algorithms, namely, extreme gradient boosting (XGBoost), random forest (RF), decision tree (DT), and logistic regression (LR) and identified XGBoost to show superior model performance in identifying predictive variable importance with an AUC score of 89.1%.

In the studies discussed above and many others, references are made of predicted accuracy scores (AUC) of varying percentages from various machine learning algorithms using biomedical and biographic variables. The use of biographic and behavioral metrics is noted in [22] for the prediction of criminal recidivism with a sampled population of 365. The other distinct feature of the observed related works is the use of sourced datasets from studies conducted elsewhere. The uniqueness of this study is the use of empirical data collected from a real-world context by assessing entities involved in disease treatment management for the prediction of treatment default.

#### 3. Research Contribution

One of the accomplishments in this research work is the use of biographic and behavioral metrics of both patients and healthcare providers to predict disease treatment default of patients suffering from hypertension with and without comorbidities.

#### 3.1. Research Hypothesis

3.1.1. Null Hypothesis  $H_0$ . No relationship exists between disease treatment default and patient gender.

3.1.2. Alternative Hypothesis  $H_1$ . The relationship exists between disease treatment default and patient gender.

A comparative analysis of four machine learning models, namely, logistic regression, gradient boosting classifier, support vector machine (SVM), and random forest classifier is examined to determine the best predicted values such as the true positive rate (TPR), false positive rate (FPR), positive predicted value (PPV), and negative predicted value (NPV) of each algorithm for performance evaluation.

This is accomplished in the listed steps as follows:

- Step 1. Variable independence using the chi-square correlation statistic is determined to establish the gender relationship between the output class.
- Step 2. Optimal threshold performance is determined with the application of threshold optimization using the area under the receiver operating characteristic curve.
- Step 3. The predicted accuracy scores of the true positive rate (TPR), false positive rate (FPR), positive predicted value (PPV), and negative predicted value (NPV) have been demonstrated using the four models.
- Step 4. The predicted scores such as the true positive rate (TPR), false positive rate (FPR), positive predicted value (PPV), and negative predicted value (NPV) to estimate how well these models would perform in subsequent predictions have been comparatively evaluated.

Comparative performance evaluation determines which machine learning algorithm performs better given the required dataset format.

#### 4. Materials and Methods

The electronic medical record dataset of patients was obtained through an institutional request for permission as per the request reference number DCS/S.1/VOL.1 from Kwahu Government Hospital (a district healthcare facility in the eastern region of Ghana). A total of 5,333 patients were identified as suffering from hypertension, and some of them were also suffering from both hypertension and other diseases (comorbidities). As a measure to protect the privacy and confidentiality of patients involved by the institution, individual patient names and location were omitted from the records. The features used were selected from clinical notes, and biographic data were obtained.

The classification-based machine learning algorithms used are gradient boosting classifier, logistic regression, support vector machine, and random forest. All software developments were performed using Python version 3.10.5 and its packages for all data processing steps. The results obtained from data modeling are presented in the subsequent section.

4.1. *Results*. The sample size was 5333 patients which consisted of 4,312 females constituting 80.86% of the total sample population and 1,021 males constituting 19.14% of the total sample population. The statistical description of age bracket and distribution density among those sampled is illustrated in Figures 1 and 2, respectively.

The highest age bracket distribution from the sampled data according to Figure 1 is found among the ages between 66 and 76, and the least age bracket distribution is found between 110 and 120. Age distribution density is higher between ages 44 and 54 up to 88 and 98. Gender distribution is displayed in Figure 3.

Basic categories of the sampled data were patients diagnosed with hypertension only and those who had been diagnosed with hypertension and other cardiovascular diseases such as diabetes. Figure 4 shows the frequency distribution of patients with hypertension only (indicated by blue color) and hypertension with comorbidities (indicated by red color). The female patient population accounted for the highest number of patients with hypertension only and hypertension with comorbidities, as illustrated in Figure 4, against their male counterparts.

In Figure 3, the frequency distribution among gender indicates a higher proportion of females with fewer males.

Representation of output class distribution in Figure 5 shows that negative class patients identified as nondefaulters were far more than those identified as defaulters to treatment default.

For the new patient, the onset of treatment begins with signs and symptoms of discomfort, whereas for an existing patient undergoing treatment, it begins with a visit as a result of a scheduled appointment or realization of signs and symptoms of discomfort. Figure 6 shows the disease treatment process for both new and existing patients suffering from hypertension and other forms of cardiovascular disease. In Figure 6, the rectangles indicate processes and subprocesses, the arrows show task routes, the rhombus show decision boundaries, and the oval represents the start or end of processes. Deepened processes with texts show data collection points, but for this research work, no measuring biomedical item was collected. These points have



FIGURE 1: Age bracket distribution.

been illustrated for emphasis on collected behavioral and biographic variables used. Certain processes had been highlighted for emphasis about the behavior of health care professionals especially on delivery of service.

Figure 7 shows a flowchart diagram illustrating various stages and processes relevant to building machine learning algorithms for prediction. It includes subprocesses, decision points, and task evaluation points, as well as arrow pointers for task routes. There are processes that depend on other processes to begin, creating a dependent rule. In Figure 7, the dependent process can be seen in the data processing stage where many subprocesses are defined before the next task, which is splitting of the dataset. At certain stages of the process, decisions to determine the next course of action are made. In machine learning prediction processes, there is beginning and ending of all processes. A green oval shows beginning of processes, whereas a red oval shows ending of processes.

In order to establish variable relationships for the output variable, the chi-square correlation statistic test was performed on all the input variables. Those with statistical significance are selected and described as follows.

#### 4.1.1. Chi-Square Correlation Statistic Score

On controlled diet: p value = 4.577598097458345e - 32 Chi-square statistic value = 153.02396213387735 Dependent (reject H0) Availability of a physician: p value = 2.8205192634562076e - 55 Chi-square statistic value = 260.9682907753015 Dependent (reject H0) Concern for long waiting/queuing time: p value = 1.239782911843141e - 25 Chi-square statistic value = 122.96922267902593 Dependent (reject H0)



FIGURE 2: Age distribution density histogram.



FIGURE 3: Frequency distribution by gender.



FIGURE 4: Comorbidity distribution among sampled gender.

Hypertension with comorbidities: p value = 2.4832105751483715e - 54 Chi-square statistic value = 256.5842340588116 Dependent (reject H0) Patient's age: *p* value = 4.4969165821531835*e* - 05



FIGURE 5: Treatment output class distribution.



FIGURE 6: Flowchart diagram of the disease treatment process.

Chi-square statistic value = 259.20561138258796 Dependent (reject H0) Gender: *p* value = 0.4911070922909865 Chi-square statistic value = 3.4137909069569465

Independent (fail to reject H0)

The chi-square correlation statistic test was performed on (selected for its significance) input variables which are described above. It shows p values and correlation status of the dependent variable. Patients admitted to be on a controlled diet, physician availability to examine patients, concern for longer waiting or queuing time at healthcare facilities, patients suffering from both hypertension and hypertension with comorbidities, and the patient's age are estimated to influence the dependent variable as null hypothesis is rejected.

A comparative display of the model confusion matrices obtained is shown.

Figure 8 shows a collection of the various confusion matrices obtained from the four machine learning models. Each matric has four sections indicating specific values. Description of colored section in the confused matrix named from the top left section (True positive-TP), yellow shaded section on the bottom right (True negative-TN), upper right corner (False negative-FN) and bottom left corner (False positive-FP). In Figure 8(a), gradient boosting correctly classified 1313 patients as nondefaulters and misclassified 15 nondefaulters as defaulters and 3 default patients as



FIGURE 7: A flowchart diagram of a model building process. The rectangles show processes, the arrows indicate task routes, and the rhombus show decision points and circles depict the start or end of processes (circular green, start and circular red, end).



FIGURE 8: Model confusion matrices for (a) gradient boosting, (b) logistic regression, (c) random forest, and (d) support vector machine.

nondefaulters. In Figure 8(b), logistic regression correctly classified 1315 patients as nondefaulters and misclassified 15 nondefault patients as default patients and 1 default patient

as a nondefaulter. In Figure 8(c), the random forest correctly classified 1309 nondefault patients and misclassified 16 nondefault patients as defaulters and 7 default patients as

22.22

50.0

TPR (%)

99.77 99.92

99.47

99.92

	1111	in opeenienty a	ina benontriny teot	500100.	
Model	FNR (%)	TNR (%)	FPR (%)	PPV (%)	NPV (%)
Gradient boosting	0.23	16.67	83.33	98.87	50.0
Logistic regression	0.08	16.67	83.33	98.87	75.0

11.11

5.56

TABLE 1. Specificity and sensitivity test scores

The highlighted scores tell us how much trust to put into a test result. Can a negative test result be truly trusted or not. probability of trust in a test result. Negative predicted value is the ratio of patients truly predicted as defaulters to all patients diagnosed as defaulters. It is a probability estimate for a nondefault status if described as a default patient. High percentage value means that the probability for a prediction miss is lower (minimal error). Lower percentage means a very high probability for a prediction miss.

88.89

94.44

nondefaulters. Finally, in Figure 8(c), the support vector machine correctly classified 1315 nondefault patients and misclassified 17 nondefault patients as defaulters and 1 default patient as a nondefaulter.

0.53

0.08

The predicted probability score rates in percentages are illustrated in Table 1.

Specificity and sensitivity both describe test accuracies for the predicted output. How well a model is able to identify true positives in a diagnostic test is referred to as its sensitivity, and conversely, a model's measure of true negatives is termed its specificity. The false negative rate (FNR) is the probability that a true positive result will not be true or missed.

In Table 1, the FNR, as recorded by the four models, is 0.23% for gradient boosting, 0.08% for logistic regression, 0.53% for the random forest, and 0.08% for the support vector machine. The probability of a true positive result being missed is much higher by the random forest classifier than any of the other models. Logistic regression and the support vector machine record the least probabilities of a true positive result being missed.

The true negative rate (TNR) is the probability that a true negative result will be missed by the model. The results obtained, as shown in Table 1, show the TNR rate for various models as follows: 16.67% for gradient boosting, 17.67% for logistic regression, 11.11% for the random forest classifier, and 5.56% for the support vector machine. The support vector machine is seen to have the least probability, followed by the random forest, with logistic regression and gradient boosting having the same score value.

The false positive rate (FPR) also referred to as "fall out" is described as the probability of a false alarm being raised. The FPR values for the models, as displayed in Table 1, show that the support vector machine has the highest record value of 94%, followed by the random forest (88%), with logistic regression and gradient boosting recording 83.33% each.

The positive predicted value (PPV) is the probability that a positive result is truly positive (a default patient is truly a defaulter). In Table 1, it can be seen that all the models show higher positive predicted values above 98%, that is, the probability that the prediction of a default patient will not be missed. The negative predicted value (NPV) is the probability that a nondefault patient will also not be missed. Actual prediction of patients classified as nondefaulters by the four models according to Table 1 shows that logistic regression shows a higher negative predicted value score of 75%, followed by both gradient boosting and support vector

Table	2:	Precision,	recall,	and	F1	score	weighted	macro	average
scores.									

98.79

98.72

Clas	sification repor	t	
Model	Precision	Recall	F1 score
Gradient boosting	0.98	0.99	0.98
Logistic regression	0.98	0.99	0.98
Random forest	0.98	0.98	0.98
Support vector machine	0.98	0.99	0.98

machines at 50% each, with the random forest recording the least percentage score. This means that if a predicted result of logistic regression is indicated as 75% for the NPV, then there is a 75% chance it is indeed negative. For the random forest classifier, a 22.22% NPV is indicative of only 22.22% chance of a negative result being accurate.

The true positive rate (TPR) is the probability that an actual positive will test positive. It is also known as a sensitivity test. The TPR values for all models as recorded in Table 1 are greater than 99%.

From the classification report shown in Table 2, the weighted average macro precision scores for the individual models were above 0.98, making F1 score higher above 0.9. The high weighted macroaverage F1 score is obtained for an imbalanced dataset (Figure 5) where the individual class score contribution is weighted by its size. High weighted macroaverage scores for precision, recall, and F1 score are indicative of prediction success.

The area under the receiver operating characteristic curve (roc\_auc) describes various model performances at separate threshold levels. Figure 9 shows a plot of two important parameters, the false positive rate and the true positive rate. In the ROC curve, model performance can be evaluated at different threshold points. Using the AUC technique, the aggregate model performance of the models can be calculated across all threshold points. The AUC score for logistic regression is 0.90, 0.77 for the support vector machine, 0.87 for the gradient boosting classifier, and 0.81 for the random forest, and these values indicate individual model prediction accuracy. The dotted black line shown in Figure 9 represents the baseline model, which is indicative of poor model performance.

4.2. Algorithm Performance Metric Discussion. This section is divided into two parts: the first part deals with a statistical description of the sampled data and the second part describes actual performance metrics of the models used,

Random forest

Support vector machine



FIGURE 9: Roc\_auc curve.

including results presented in various figures, graphs, and tables.

In Figure 1, a presentation of age bracket distribution of the sampled data shows that the incidence of hypertension and its associated diseases (comorbidities) is prevalent among the age bracket 66–76. The incidence of hypertension alone and hypertension with comorbidities among gender, as shown in Figure 4, indicates that the female population had the highest prevalence rate than males.

A determination of variable relationships using Chisquare correlation statistic of input variables to estimate gender, age, availability of a physician, hypertension with comorbidities, queue/waiting time and controlled diet presented under sub section Chi-square correlation Statistic score produced three scores namely; Ch-square statistic value, p value and variable correlation state. Comparison of p value with alpha value of 1 results in output decision; a rejection or failure to reject the null hypothesis.

The chi-square correlation statistic of variance showed no correlation between treatment default and gender, thereby failing to reject  $H_0$ . However, other variables such as age, presence of comorbidities, concern for long waiting/ queuing time, controlled diet, and availability of a physician showed correlation with the output variable (treatment default status), and null hypothesis was rejected.

Generally, all the models as shown in Table 1 show capacity to predict true positive values, as shown by their scoring percentages (gradient boosting, 99.77%; logistic regression, 99.92%; random forest, 99.47%; and SVM, 99.92%). The important additional performance metric in Table 1 with the highlighted text shows the negative predicted values (NPVs) of all the models. The negative predicted values in Table 1 according to individual models show that logistic regression has a higher score of 75%, followed by both gradient boosting and support vector machines scoring 50% each, with the random forest algorithm scoring the least score of 22.22%.

This was to define what negative predicted value or score mean for predictions by linking it with the score value of logistic regression. Logistic Regression's prediction of negative values will be accurate at 75% rate. This is of utmost importance for predictions in medical diagnosis, disease detection, and other problems in the healthcare domain. The accuracy of a negative or positive result being truly negative or positive is essential to avoid missing diagnosis, especially in diseases that require urgent attention. It is for this reason that a logistic regression score of 75% is considered critically important.

Additional performance evaluations with definitions of false negative rates and an explanation for the score are obtained. However, the random forest classifier shows a probability score of 0.53%, approximated as 1, which makes it very likely to miss a true positive result. As shown in Figure 5, the output class contains data imbalance; therefore, it is important to consider loss based on proportion and the use of weighted macroaverages. The weighted macro-averages for precision, recall, and *F*1 score according to the classification report shown in Table 2 are 98% for gradient boosting, 99% for logistic regression, 98% for the random forest, and 98% for the support vector machine.

The additional evaluation metric used for model performance is the roc\_auc curve, as shown in Figure 9. This curve is an aggregate measure across all threshold points for model performance. The AUC scores shown in Figure 9 for the support vector machine model are 0.77, 0.81 for the random forest, 0.90 for logistic regression, and 0.87 for gradient boosting, respectively.

From the various performance metrics, as shown in Table 1 and Figure 9, the logistic regression model is considered the obvious model of choice for this classification prediction. It has shown superior performance with an AUC score of 0.90, a negative predicted value (NPV) score of 75%, and an FNR of 0.08%.

#### 5. Discussion

The stated focus of this research was to use biographic and behavioral metrics to predict disease treatment default and to test for null hypothesis as to whether gender biases have any correlation with hypertension treatment default. The use of the chi-square correlation statistic of variable independence as shown proves no dependency of patient

TABLE 3: Contextual biographic and behavioral variables.

Biographic and behavioral metrics			
Variable	Description		
Age	Patient's age		
Gender	Gender of the patient		
Advreact	Complaints of adverse drug reaction		
Herbal	Use of herbal supplements		
Missdose	Default in medication		
Smoking	Smoking status		
Exercise	Any physical activity		
Alcohol	Use of alcohol		
Diet	Whether on controlled diet		
Staff attitude	Approval of healthcare provider attitude		
Service_stf	Service rating		
Attd_physician	Availability of a clinician		
Queue_time	Concern for waiting time		
Cmbd	Presence of comorbidities		
Defaulter_statu	Treatment status		

gender on treatment default but rejects the null hypothesis with respect to other contextual variables. Further model evaluations to determine superior model performance in Table 1 and Figure 9 also identify logistic regression as having a higher NPV, as shown in Table 1, and SVM has the least probability score for an FNR (0.08%) and the highest AUC score of 0.90, as shown in Figure 9. Based on the above statistical values, the logistic regression model used in this context is seen to be superior in performance to the other three models evaluated. This comparative analysis narrative, especially the use of NPV, proves that model accuracy prediction score evaluation linked to the problem context helps determine model performance superiority.

#### 6. Conclusion and Future Work

This research article has demonstrated how both biographic and behavioral metrics can be used to predict disease treatment default without biomedical measurements. It has also been determined (based on the collected data) that gender biases do not affect hypertension treatment default risks, but other contextual variables such as age, comorbidities, queuing time, availability of a clinician, and patients on a controlled diet can affect treatment default outcomes. Finally, the superiority of the logistic regression algorithm to predict disease treatment default in hypertensive patients has been established. Further work in this area is to investigate time complexities of the algorithms used to determine efficient and effective machine learning models among the selected variables (Table 3).

#### **Data Availability**

The dataset used for analysis can be available from the corresponding author upon request.

#### Disclosure

This publication is an extract from an academic research that is self-funded by the student.

#### **Conflicts of Interest**

The authors declare that there are no conflicts of interest regarding the publication of this paper.

#### **Authors' Contributions**

Michael Owusu-Adjei conceptualized the study. Michael Owusu-Adjei, Dr. Twum Frimpong, and Dr. Gaddafi Abdul-Salaam designed the methodology. Prof. James Ben Hayfron-Acquah supervised the study.

#### Acknowledgments

The authors continue to acknowledge the support and cooperation of management and staff of the Kwahu Government Hospital, especially the Medical Director Dr. Kobina Awotwe Wiredu and the supervisory team led by Prof. James Ben Hayfron-Acquah, Dr. Twum Frimpong, and Dr. Gaddafi Abdul-Salaam. This publication is an extract from an academic research that is self-funded by the student.

#### References

- Ghana Health Service, "The health sector in Ghana: facts and figures 2018," *Ministry Health Ghana*, pp. 1–50, 2018.
- [2] WHO, Global Health Risks, WHO, Geneva, Switzerland, 2009.
- [3] A. Tiwari, V. Dhiman, M. A. M. Iesa, H. Alsarhan, A. Mehbodniya, and M. Shabaz, "Patient behavioral analysis with smart healthcare and IoT," *Behavioural Neurology*, vol. 2021, Article ID 4028761, 9 pages, 2021.
- [4] A. M. Patey, G. Fontaine, J. J. Francis et al., "Healthcare professional behaviour: health impact, prevalence of evidence-based behaviours, correlates and interventions," *Psychology and Health*, pp. 1–29, 2022.
- [5] M. Grissinger, "Disrespectful behavior in health care: its impact, why it arises and persists, and how to address it—Part 2," *P and T*, vol. 42, pp. 74–77, 2017.
- [6] E. Siegel, Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, Revised and Updated, John Wiley & Sons, Hoboken, NJ, USA, 2015.
- [7] L. Harris, V. K. Lee, E. H. Thompson, and R. Kranton, "Exploring the generalization process from past behavior to predicting future behavior," *Journal of Behavioral Decision Making*, vol. 29, no. 4, pp. 419–436, 2016.
- [8] L. H. A. Salazar, V. R. Q. Leithardt, W. D. Parreira et al., "Application of machine learning techniques to predict a patient's no-show in the healthcare sector," *Future Internet*, vol. 14, pp. 1–21, 2022.
- [9] O. Cordón, F. Herrera, J. De La Montaña, A. M. Sánchez, and P. Villar, "A prediction system for cardiovascularity diseases using genetic fuzzy rule-based systems," *Advances in Artificial Intelligence — IBERAMIA 2002*, vol. 2527, pp. 381–391, 2002.
- [10] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 104–116, 2017.
- [11] A. J. Barda, C. M. Horvat, and H. Hochheiser, "A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare," *NCHHSTP Social Determinants of Health*, vol. 20, pp. 9–12, 2019.

- [12] NCHHSTP, "Frequently asked questions | social determinants of health," 2022, https://www.cdc.gov/nchhstp/ socialdeterminants/faq.html.
- [13] M. Maniruzzaman, M. J. Rahman, B. Ahammed, and M. M. Abedin, "Classification and prediction of diabetes disease using machine learning paradigm," *Health Information Science and Systems*, vol. 8, pp. 7–14, 2020.
- [14] R. Krishnamoorthi, S. Joshi, H. Z. Almarzouki et al., "A novel diabetes healthcare disease prediction framework using machine learning techniques," *Journal of Healthcare Engineering*, vol. 2022, Article ID 1684017, 10 pages, 2022.
- [15] K. Dwivedi, H. O. Sharan, and V. Vishwakarma, "Analysis of decision tree for diabetes prediction 4," *International Journal* of Engineering and Technical Research, vol. 9, pp. 3–6, 2019.
- [16] C. A. Salkar, "A detailed analysis on kidney and heart disease prediction using machine learning," *Journal of Computing* and Natural Science, vol. 1, pp. 9–14, 2021.
- [17] R. Spencer, F. Thabtah, N. Abdelhamid, and M. Thompson, "Exploring feature selection and classification methods for predicting heart disease," *Digital Health*, vol. 6, Article ID 205520762091477, 2020.
- [18] N. Fazakis, O. Kocsis, E. Dritsas, S. Alexiou, N. Fakotakis, and K. Moustakas, "Machine learning tools for long-term type 2 diabetes risk prediction," *IEEE Access*, vol. 9, pp. 103737– 103757, 2021.
- [19] M. Durairaj and V. Revathi, "Prediction of heart disease using Back propagation MLP algorithm," *International Journal of Scientific & Technology Research*, vol. 4, pp. 235–239, 2015.
- [20] C. Fiarni, E. M. Sipayung, and S. Maemunah, "Analysis and prediction of diabetes complication disease using data mining algorithm," *Procedia Computer Science*, vol. 161, pp. 449–457, 2019.
- [21] S. DuBrava, J. Mardekian, A. Sadosky et al., "Using random forest models to identify correlates of a diabetic peripheral neuropathy diagnosis from electronic health record data," *Pain Medicine*, vol. 18, no. 1, pp. 107–115, 2017.
- [22] M. O. Pflueger, I. Franke, M. Graf, and H. Hachtel, "Predicting general criminal recidivism in mentally disordered offenders using a random forest approach," *BMC Psychiatry*, vol. 15, pp. 62–10, 2015.
- [23] G. Parthiban, "Applying machine learning methods in diagnosing heart disease for diabetic patients," *International Journal of Applied Information Systems*, vol. 3, no. 7, pp. 25–30, 2012.
- [24] X. Tian, Y. Chong, Y. Huang et al., "Using machine learning algorithms to predict hepatitis B surface antigen seroclearance," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 6915850, 7 pages, 2019.