Applied Machine Learning Techniques for Chronic Disease Treatment Default Prediction and its Potential Benefits for Patient Outcome: A Case Series Study Approach

Michael Owusu-Adjei
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Joseph Manasseh Opong
Presbyterian University, Ghana

Apenteng Samuel
Presbyterian University, Ghana

ABSTRACT
In the context of medical diagnosis, consideration of missing diagnosis and false diagnosis of disease types is important clinical consideration for disease treatment decisions. The effect and impact of disease types, especially on others, form the basis for critical clinical decisions. Impact and consequences varies across disease types, especially for communicable and non-communicable diseases. Increasing the use of predictive techniques dueto the high use of connected internet of things devices in healthcare provides sufficient opportunity for a potential benefit assessment of the impact of predictive modeling on the treatment of diseases. Effective and efficient management of non-communicable diseases such as hypertension is hampered in part by instances of multiple forms of its occurrence in patients leading to treatment management complications. The evaluation of the effect of predictive modelingand the implications for clinical decisions to enhance the outcome of patient treatment provides important evidence-based justifications for its use in healthcare systems. The effective use of the predictive technique dependssignificantlyon the areas of its application and the consequences of error for its use in context.

Keywords
Communicable, non-communicable, diseases, prediction, effect, impact, treatment, performance, decision support systems.

1. INTRODUCTION
The effect and impact are important clinical considerations for disease treatment decisions in healthcare care. The online Cambridge dictionary defines impact as 'the force or action of one object hitting another' or 'a powerful effect that something, especially something new, has on a situation or a person'. The resulting influence of action or force is determined to be the effect of impact. Crucial to effective patient care is evidence-based clinical decisions, which include fast, intuitive, and deliberate analytical decision considerations within the context of disease type (communicable, non-communicable), severity, effect, and impact on patient care and others [1]. The clinical adoption of modeling techniques is dependent on two most important factors; clinical usefulness and trustworthiness. Therefore, the role of the model technique in clinical practice and the expected outcome must answer specific clinical questions with attendant patient benefits; otherwise, its usefulness and trust in the reliability and dependability is questioned [2][3]. Effective application of predictive techniques in medical context consists of quality and accuracy of data given, assessment of effect on clinical decisions, and impact on patient outcome taking into account disease type. The lack of performance assessment in clinical contexts is amplified in a related study [4] that examined machine learning tools in two areas; sepsis and prediction of suicide. Research on development of computerized clinical decision support systems has seen tremendous growth in recent times, but research evaluation on specific learning techniques and use and potential benefit to clinicians and patients remains a challenge. These challenges identified are mainly driven by concerns about the accountability of usage in unfavorable clinical decision outcomes, the risk of overreliance due to limitations and errors, and the problem of usability involving integration into existing clinical workflows [5][6]. The evaluation of context-based predictive modeling for potential benefit to clinical decisions on patient outcome in chronic disease treatment with comorbidity remains limited in research. This research work brings to the fore important clinical decision points for predictive modeling technique performance impact assessment on both clinical decisions and patient outcomes in chronic disease treatment default prediction.

1.1 Research Objective
The key objective of the research is to evaluate the impact of predictive performance and its effect on clinical decisions on patient outcomes in chronic diseases such as hypertension with comorbidities for potential benefits. The outcome of this research study will help address the impact and effect in the use of predictive techniques for clinical decisions that enhances patient outcome.

1.2 Importance of the Research
The burden of chronic diseases and their negative impact on health management, especially in lower and middle income countries, continue to attract the attention of both policy makers and research studies. Limited number of healthcare professionals to adequately manage the increase and its negative impact, especially in developing countries, where the destruction affects economies, increasing healthcare costs, and patient morbidity, resulting in increasing patient deaths. The use of predictive modeling techniques as a reliable alternative with trust and confidence has the potential to help address human resource challenges. Therefore, it is significant to determine the impact of predictive modeling performance and the effect on clinical decisions and patient treatment outcomes that build confidence in its application, particularly in healthcare systems.

1.3 Related Research Work
To help address potential benefits of applied predictive technique performance on clinical decisions and its effect on patient outcome, the use and impact on clinical decisions is examined in related applications in various research works,

particularly on communicable, non-communicable and other important considerations. The importance and use of predictive techniques in accelerating and improving physician tasks through effective automation processes has also enhanced and improved decision making in healthcare, as emphasized in research studies [7][8] that examined machine learning applications in healthcare, particularly public health. Furthermore, the use of predictive algorithms and its impact include a significant effect on accurate predictions that positively impacts clinical decisions and treatment outcome such as disease diagnosis and prognosis, personalized medicine, public health surveillance and outbreak detection, health behavior analysis and intervention, and utilization of healthcare resources. Some of the benefits determined in a related study show how data-driven insights and informed decision making with the use of machine learning have revolutionized healthcare and public health decision support systems. In public health delivery, the use of predictive techniques has been beneficial in spatial modeling, risk prediction, misinformation control, public health surveillance, disease prediction, pandemic / epidemic modeling and health diagnosis [9]. Other application areas for the use of predictive techniques identified include precision medicine, diagnosis and treatment recommendations, patient engagement and adherence, and administrative activities [10]. In their article on Artificial intelligence, machine learning and health systems, the potential effect of machine learning is emphasized to be the catalyst for improving healthcare systems in efficiency, effectiveness and outcome is emphasized [11]. Another scoping review of clinical decision making in the emergency department for paediatric patients [12] include among its focus models for clinical decision support. A randomized trial of 20, 563 patients admitted to a hospital was evaluated to determine the effectiveness of the computerized clinical decision support system. This provided evidence-based evaluation of actionable patient-specific recommendations at the point of care by healthcare professionals [5]. The need to improve the accuracy of the model within an acceptable time interval for early detection of diseases such as cancer, diabetes, chronic kidney disease, etc. is underscored in a research study that compared the performance of different machine learning techniques and approaches for clinical decisions [13]. Identified resource areas for predictive application decisions in healthcare were in patient care, disease treatment, resource allocation, and utilization. Results obtained indicate that the use of neural network-based deep learning techniques in the computational biology domains produced high prediction accuracy with reliability, thus making decisions based on computational biology and biomedicine dependent on such predictive modeling techniques. Similar use of machine learning techniques for the prediction of diabetes mellitus disease with feature enhancement and oversampling techniques [14] showed high prediction precision performance in two different datasets for classifiers such as random forest (RF), light gradient boost (LGB), and gradient boost (GB). The prediction accuracy scores obtained were 98.99% for LGB, 96.6% for RF and 97.64% for GB. The Conclusionss were a proposal for further improvements in prediction accuracy with advanced methodologies such as the transformer-based learning technique. Applied machine learning techniques for cervical pain assessment in patients affected by whiplash disease using techniques such as logistic regression, support vector machines, k-nearest neighbors, gradient boosting, decision trees, random forest, and neural network algorithms on 302 dataset examples [15] produced prediction precision, precision, and recall values above 90%. The predictive performance benefit and clinical impact and patient effect identified include the presence of pain in patients for clinical decisions that can improve the efficacy of treatment to help address resource utilization. A comparative study to predict measles with 1,797 suspected cases used six machine learning techniques to determine 78 positive cases of measles and 1,696 were identified as negative cases creating a class distribution imbalance. The results obtained showed superior random forest classifier performance in specificity 96%, sensitivity 88%, receiver operating characteristic curve score 92% and total prediction accuracy score 92% [16] than the other modeling techniques (generalized linear model, decision tree, nave bayes, support vector machines, artificial neural network). Applied learning technique assessment study for the detection and management of infectious diseases caused by fatal or life-threatening causative agents capable of infecting both animals and humans provided a comprehensive review of machine learning application use in pathogen detection, public health surveillance, host-parasite interaction, drug discovery, omics and vaccine discovery. The emphasis on the use of evaluation metrics (precision and recall) in classification tasks [17] as important performance measure is placed on modeling techniques such as support vector machine, random forest, and neural networks. The clinical impact is evaluated to be inclusion and emphasis on precision and recall as important evaluation metrics to identify parasitic cells as infected or not infected (positive or negative). The impact of the use of the multivariate algorithm use [18] to accurately select breast cancer without residual cancer after administering neoadjuvant therapy is identified in a related study that identified patients with complete pathologic response. Using randomized dataset examples of 457 women, partitioned into training and test set enrolled in three trials with stage 1-3 breast cancer, the false negative rate (FNR) estimate of 1.2% was achieved for logistic regression with elastic net penalty, extreme gradient boosting, support vector machines and neural network. Estimating the incidence of diabetes mellitus in a public health surveillance based on the number of reimbursements over a 2 year period [19] with four predictive techniques; Linear discriminant analysis (LDA), logistic regression (LR), flexible discriminant analysis (FDA) and decision tree for 44,659 participants showed predicted performance score of sensitivity 62%, specificity 67% and accuracy score of 67% for LDA, making it the highest performing technique of choice. Effective modeling techniques in healthcare as identified [20] in a comparative analysis of the study of modeling techniques, include the random forest classifier. The evaluation of context-based predictive modeling, especially in real-world business applications, should therefore explore specific areas in the clinical decision process for relevance to determine the potential benefits and consequences of error as a reasonable justification for its use.

2. METHODS AND MATERIALS

In this study, a case series approach was adopted using characteristics identified in 5,333 electronic health records of patients undergoing treatment for hypertension with comorbidity. The variables used were extracted from patient attendance interactions recorded by attending clinicians with an expressed permission request and approval referenced DCS/S.1/Vol.1 and dated 30 March 2022 from the main district healthcare facility in Kwahu South, Ghana. The reason for selecting this healthcare facility is informed by its long-standing management of cardiovascular diseases including hypertension, its status as a referral point and remote location serving several communities with diverse population characteristics including education, occupation, social behavior, health needs, and many others. The uniqueness of its location within a mountainous geographical location of Ghana with variable weather conditions characteristically different from other parts of Ghana makes it the ideal location for insights into such phenomena. Among the sampled patients, 4,312 of the 80.86% were women and 1,021 of the 19.14% were men. The hypertension patients, but without comorbidity, were 4,192, 78.60%, and 1,141 patients with hypertension and comorbidity were identified, which comprised 21.40%. In this study, the gender characteristics were separated into male and female, the incidence of comorbidity among patients was also separated into patients with only hypertension without comorbidity and patients with both disease conditions (hypertension with comorbidity). Selection criteria for inclusion and exclusion are defined below.

2.1 Inclusion Criteria

1. The inclusion criteria included patients diagnosed with hypertension with and without comorbidity.
2. No exclusivity rule was applied in respect of gender preferences, age, social status, occupational preferences, etc.
3. Patients who had been diagnosed from 6 months and older were considered.

2.1.1Methodology

Determining specific areas in the decision process for potential benefits in the use of predictive techniques remains the main research objective in this article. Determining the impact of model performance in these areas will provide important information for clinical decision support systems for effective healthcare care management that improves patient outcome. Ten (10) known modeling techniques namely; logistic regression, support vector machine, random forest classifier, gradient booster classifier,multilayer perceptron, extra trees

classifier, decision tree, bagging classifier, kneighbors classifier and linear discriminant analysis classifier, namely logisticregression, support vector machine, random forest classifier, gradient booster classifier, multilayer perceptron, extra trees classifier, decision tree, bagging classifier, kneighbors classifier, and linear discriminant analysis classifier are used to estimate predictive performance and how it impacts clinical decision and its effect on patient outcome. Brief but less technical descriptions of the predictive techniques used are presented in the following sub-sections.

2.1.1.1 Logistic regression

Predominantly, about 70% of problems in data science and data mining are classification problems for which classification techniques have become an essential part of machine learning applications [21]. Logistic regression use in binary classifications is evident in spam detection, disease diagnosis, and predictions such as diabetes, chronic kidney, cancer predictions, tumor classifications into malignant and benign, customer purchasing behavior predictions, etc. its use is particularly important in establishing relationships between dependent and independent variables in two-binary classifications.

2.1.1.2 Support vector machines

Its ability to handle nonlinear input space and separate data points using a hyperplane that finds optimal data points to classify new data points. The reference to a support vector machine as a discriminative classifier [22] makes it an ideal technique for intrusion detection, face detection, email

4. Patients enrolled in national health insurance and other private personal insurance including those not covered by any insurance policy were considered.
5. No consideration for the number of instances of comorbidity in a single patient.
6. Patients diagnosed and profiled as defaulters and non-defaulters were included.

2.2 Exclusion Criteria

1. Transfer patients were excluded regardless of length of stay.
2. Pregnant patients with any of these disease conditions were excluded.
3. Disabled patients with movement disabilities were excluded
4. Diagnosed patients with mental health disorders were excluded.

Ethical approval and consent: Real-world clinical data set obtained from the main district hospital in Kwahu South district of Ghana with approval notice reference number DCS/S.1/VOL.1 dated 30 March 2022.

classifications, etc. It is useful in both classification and regression problems.

2.1.1.3 Random forest classifier

The random forest classifier is a supervised learning technique used in classification, regression, and other tasks with decision trees. Decision trees are created from a randomly selected subset of training sets. The collection of votes from different decision trees determines the final prediction outcome [23].

2.1.1.4 Gradient-Boosting Classifier

Structured predictive modeling problems, such as classification and regression, find it useful on tabular data modeling [24]. It is an ensemble algorithm that minimizes error gradients with fitted boosted decision trees. It is also known as gradient tree boosting, stochastic gradient boosting, and gradient boosting machines and abbreviated as GBM.

2.1.1.5 Multilayer Perceptron

Multilayer perceptron, or MLP, is a type of neural network algorithm used mostly in complex classification tasks. Its capability lies in its ability to learn nonlinear relationships between inputs and outputs [25]. MLP has an input layer to receive inputs, a hidden layer to perform computations on inputs received, and an output layer to generate final results for display from the model.

2.1.1.6 Extra Trees Classifier

Classification and regression task algorithm that works by randomly selecting a subset of features for training using a decision tree. The tree is subsequently pruned [26] to contain features important for predicting the final outcome. It is also known as an extremely random tree.

2.1.1.7 Decision Trees

Supervised learning models are useful in both classification and regression problems. The work in a decision tree is representative of a flow chart in which each node represents a decision point that splits into two leaf nodes representing an outcome of a decision. Each of these decisions can also turn into decision nodes [27]. The decision-making process in decision trees is easy to understand and interpret. It works by splitting data into series of binary decisions for traversal.

2.1.1.8 Bagging Classifier

An ensemble algorithm that works by dividing training sets into subsets for processing using different techniques to improve performance. The combined outputs of the different techniques form the basis [28] for performance determination. Bagging is useful in the domains of classification and regression problem domains.

2.1.1.9 Kneighbors Classifier

Supervised machine learning algorithm predominantly used in both classification and regression tasks. Predictions are made by calculating the distance between the two data points (training data and test data) for the assumption of characteristics that is similar between them [29]. The characteristics learned from the training data points allow for the identification and allocation of new data points.

2.0.1.10 Linear Discriminant Analysis Classifier

Useful supervised learning technique for statistics and classification. It works by finding linear combination features in a dataset class that best discriminates or separates them. It is particularly useful in finding useful information features in classification predictions. It is widely used in fields such as pattern recognitions, bioinformatics and image processing etc [30].

3. DISCUSSION

The statistical distribution of the contributions of gender to the incidence of hypertension with comorbidity is illustrated in Table 1 which shows the occurrence of the disease among age groups and gender, including the incidence of combidity. Among the female population, the incidence of hypertension is estimated only at 3,435 and the comorbid hypertension was 877 within different age brackets. The estimated age range for only hypertension and hypertension with comorbidity is between 21 and 109 years and between 18 and 111 years. The estimated age range for the male population regarding the incidence of hypertension alone and hypertension with comorbidity was between 25-105 years and 23-100 years, respectively. These descriptions show that the incidence of hypertension and hypertension with comorbidity occurs in female populations earlier than males and this is supported by research studies [31][32] that examined gender differences in the prevalence of hypertension between various age groups and the potential association of gender-specific factors for hypertension. A comprehensive cardio metabolic risk assessment of gender found that women are at higher risk than men and that among gender-specific factors, age (younger age) at childbirth was found to be among the high-risk factors for women. This study adds that the incidence of hypertension and hypertension with comorbidity occurs in women 4-5 years earlier than in men, as shown in **Table 1**. Predictive model performance and its impact on clinical decisions and effect on patient outcome is accomplished in two different ways; **Table 2** shows macro-average model prediction precision, recall, f1 score and accuracy performance for the ten algorithms used. The importance and impact of each metric used is determined in context. As an example, it is better to misclassify a genuine email message as spam than to incorrectly misclassify a malicious spam message as genuine due to impact and effect. In medical contexts, misclassifying a compliant patient as a defaulter may have little impact and effect on clinical decisions and treatment outcome than misclassifying a default patient and this is significantly due to impact and effect on nonadherence. In Table 2 it is observed that even though LDA achieves the highest area under the curve score (auc_roc) of 0.90 or 90%, its precision is hampered by attaining a score of 0.68 or 68%. However, logistic regression in this instance achieves the highest precision score of 0.87 or 87% and an auc_roc score of 0.89 or 89%. However, macro average recall and f1 score as recorded show that random forest, extra trees, linear discriminant analysis, and gradient boosting classifiers obtained the largest macro average recall score of 0.62 or 62%, respectively, as against 0.59 or 59% recorded for logistic regression and decision tree. The use of macro-average scores in class imbalance context is to treat all classes equally regardless of variations in class distributions. Clinical significance includes determining effective and accurate predictive modeling techniques for the real-world application context where class distribution variation is a characteristic feature. Identifying various sections of the healthcare delivery process ensures effective application of predictive modeling techniques for the required impact on clinical decisions and its effect on patient outcome. Characteristic of real-world applications, collected dataset as displayed in **Fig.1** and **Fig. 2** shows output class distribution imbalance between defaulted patients and non-defaulters. Estimated patient defaulters and non-defaulters amounted to 1.52% and 98.42% of the sampled population. In **Fig. 3** important decision sections that serve as points for data collection in a typical healthcare system are provided and highlighted in red for emphasis. The highlighted sections provide useful context for predictive modeling applications. Predictive modeling contribution for impact and effect in each of these sections towards clinical decisions and outcome remains varied; therefore, predictive research in any of these or combination of sections can provide important insight into its potential benefit for clinical decisions that enhances patient outcome as each of these sections show interdependency. In its estimation of nursing processes to serve as a guide for nursing practice, subjective and objective evaluation of critical thinking and data collection is determined as an important first step to achieving patient-centered care. Subjective evaluation includes verbal statements taken from patients and healthcare providers, and objective evaluation involves measurable metrics such as vital signs [33]. The clinical importance of the utilization of the nursing process will ensure that the risk of missing out on hidden but life-threatening disease condition is reduced. Predictive technique use capable of learning patterns in clinical datasets offers important assistance to clinical decision support systems for patient management that improves treatment outcome. An outcome that includes accurate classification and prediction of disease conditions, associated risk factors, identification of hidden patterns in different segments of large clinical datasets for an effective prognosis of disease conditions. Model performance with receiver operating characteristic curve (roc_auc) for the ten algorithms is displayed in **Fig. 4** which also shows performance scores achieved by the individual techniques. Easy visualization of model performance as shown by the roc_auc score curve is displayed in **Fig. 5**

4. RESEARCH STUDY LIMITATIONS

Any assumption made in this study is based on the available collected data. Enhanced predictive performance, especially for deep neural networks, is dependent on large dataset volumes, and this research dataset volume of 5,333 is considered minimal to deep neural network contexts for effective performance evaluation hence the use of traditionally tested algorithms. The inclusion of a multilayer perceptron classifier is for comparative assessment purposes. Identified class imbalance distribution of the output class is not addressed with any optimization techniques, as its occurrence is characteristic of real-world applications.

5. APPLICATION AND SIGNIFICANCE OF THE RESEARCH

The identification of decision sections offers significant data collection points of interest in predictive modeling research for the creation of healthcare decision support systems. Prediction outcome with macro average precision, recall, and f1 score determines how much trust to put into a prediction outcome in context which may define the impact of error on the spread of diseases for communicable and non-communicable diseases. Models with low macro average precision, recall, and f1 score as recorded mean that the prediction of a default or nondefault hypertensive patient cannot be trusted in such scenarios. This could lead to a situation where the focus on default patients for necessary intervention is impaired, negatively impacting clinical decisions with attendant negative effect on patientoutcome.

6. CONCLUSION

The identification of significant decision sections that serve as important data collection points for impactful decision support in healthcare settings, the determination of the impact of the predictive technique in the use of relevant performance metrics within the context of real-worldapplications characteristic of the imbalance of the distribution of the output class, exploratory analysis showing gender disparity in the incidence of hypertension and hypertension with comorbidity between different age groups demonstrates potential benefits for applied predictive modeling techniques on the impact of clinical decisions and its effect on patient outcome. Future research will focus on the collection and effective use of deep learning techniques for the diagnosis and interpretation of X-ray images as a decision support system to address human resource challenges in the recruitment of qualified specialist radiologists in Ghana.

Table 1 Statistical distribution of the gender contribution

Gender	Age range(years) Hypertension only	Hypertension only	Age range (years) hypertension with comorbidity	Hypertension with comorbidigy
Female	21-109	3435	18-111	877
Male	25-105	757	23-100	264

Table 2 Model performance scores

Model type	macro average precision score	Macro average recall score	Macro average f1-score	auc score
Logistic regression	**0.87**	0.59	**0.65**	**0.89**
Support vector machine	**0.49**	0.50	0.50	**0.85**
Random forest	0.65	**0.62**	0.63	0.72
Gradient boosting	**0.70**	**0.62**	**0.65**	**0.84**
Multi-layer perceptron	**0.49**	0.50	0.50	**0.87**
Extra trees classifier	0.64	**0.62**	0.63	0.64
Decision tree	0.56	0.59	0.57	0.59
Bagging classifier	0.59	0.56	0.57	0.72
Kneighbors classifier	0.49	0.50	0.49	0.52
Linear discriminant analysis	**0.68**	**0.62**	0.64	**0.90**

6. ACKNOWLEDGEMENTS

The support and cooperation of both the management and staff of Kwahu Government Hospital are acknowledged.

Authors' contributions

Conceptualization: Owusu-Adjei, Michael.

Formal Analysis: Owusu-Adjei, Michael.

Methodology: Joseph Manasseh Opong

Project administration: Owusu-Adjei, Michael.

Supervision: Joseph Manasseh Opong

Writing – original draft: Owusu-Adjei, Michael.

Writing–review & editing: Apenteng Samuel

Funding: Authors declare No funding

Concent for publication: Appoved

Availability of data and Materials: Dataset used for analysis is available upon reasonable request.

Conflicts of Interest: The author(s) declare that there is no conflict of interest regarding the publication of this paper.

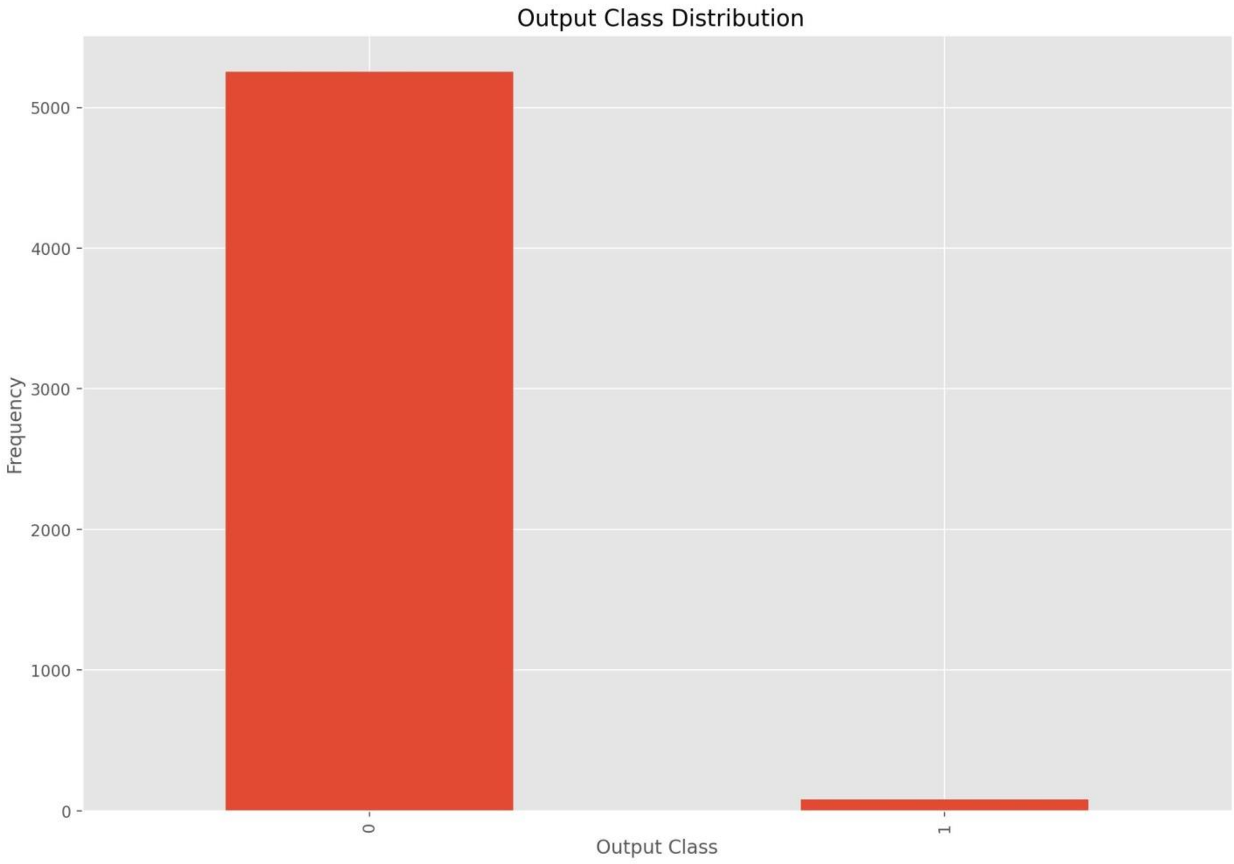

Fig 1 Class distribution frequency diagram

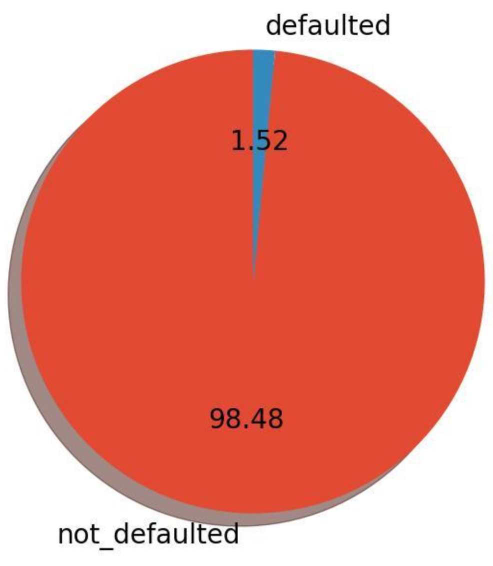

Fig 2 Pie chart display of output class distribution

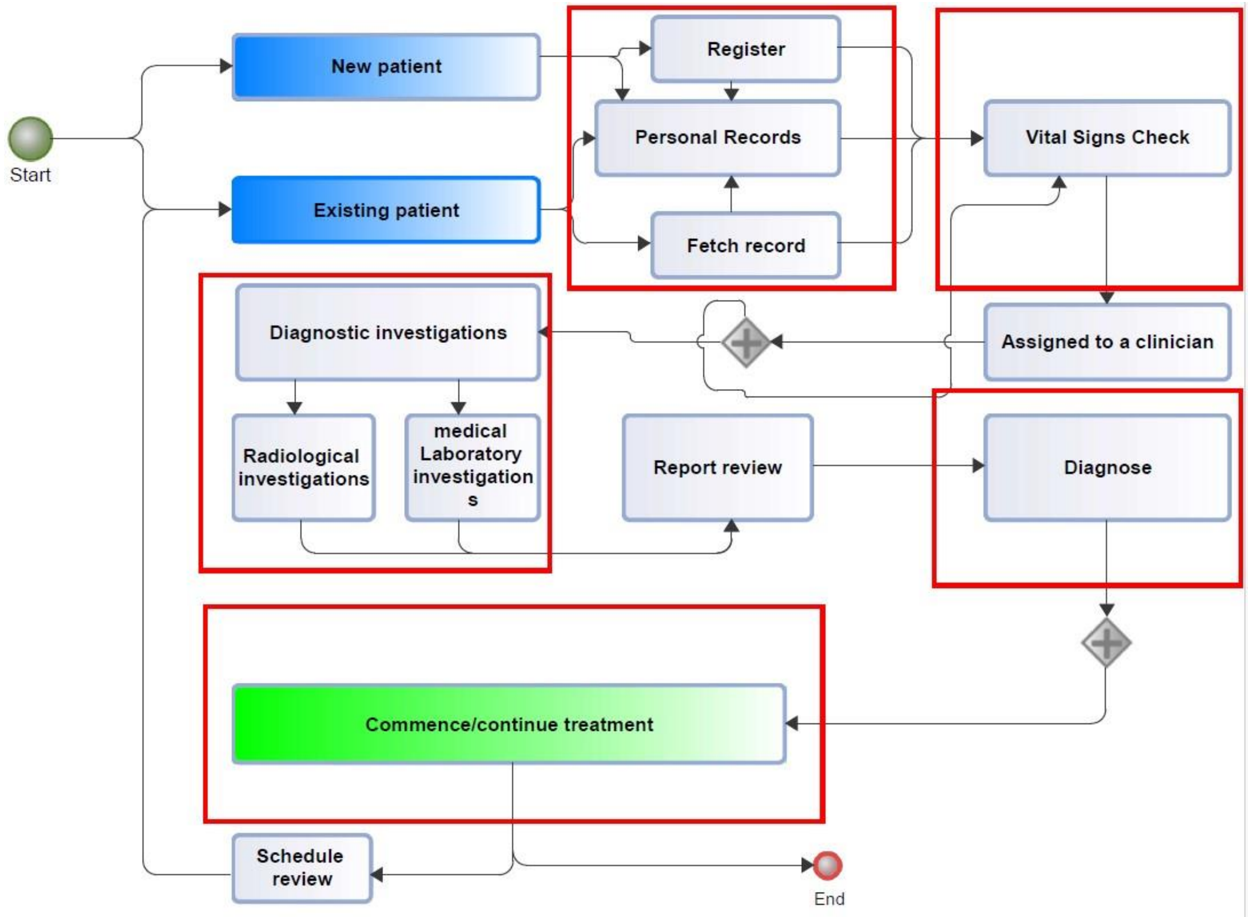

Fig 3 Simplified clinical decision sections in a typical healthcare delivery system.

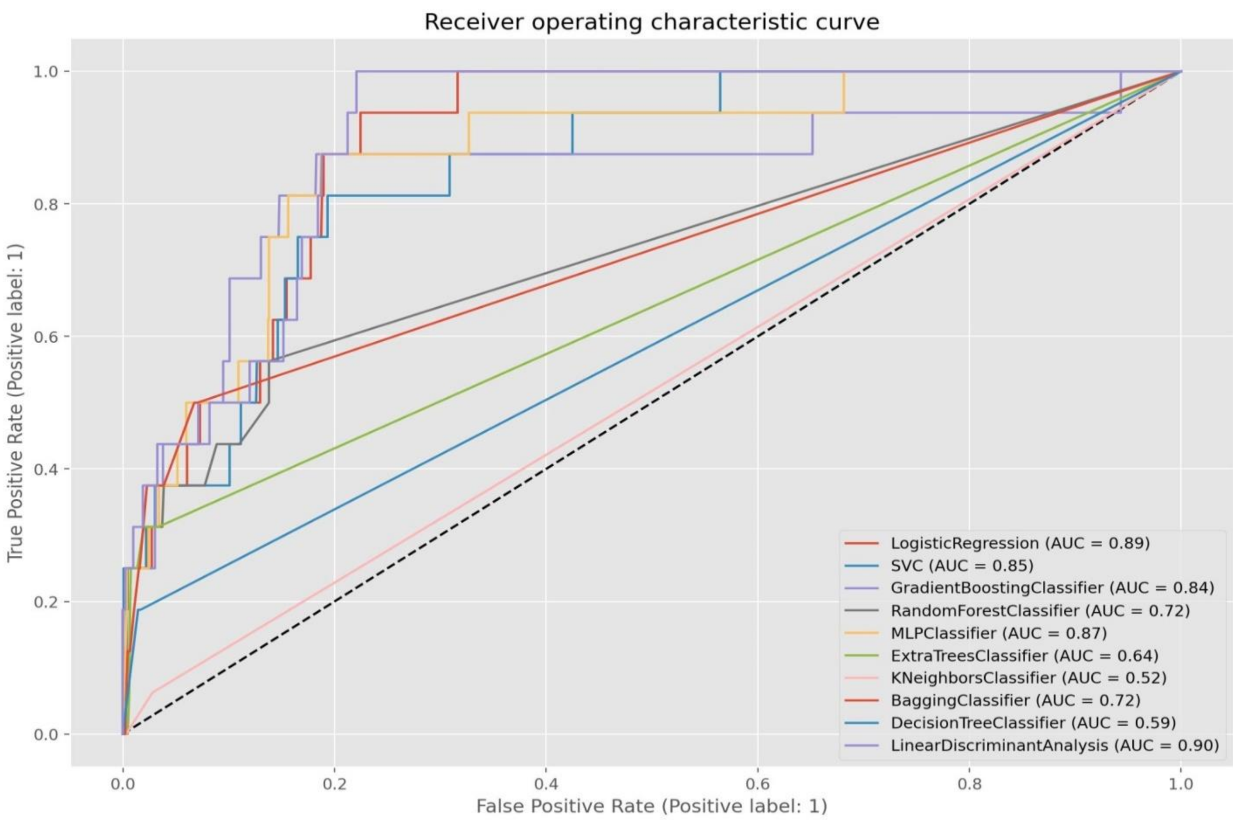

Fig 4 Receiver operating characteristic score curve

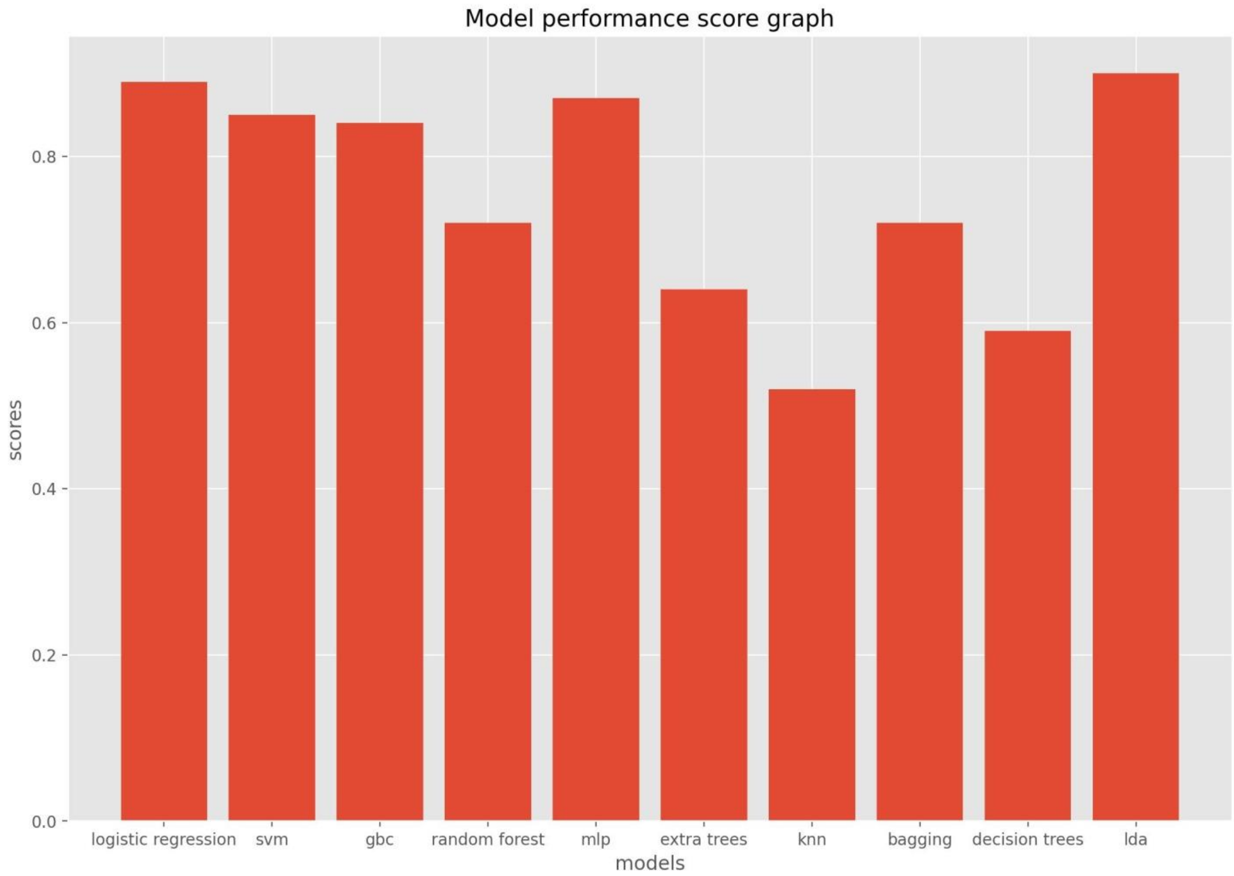

Fig 5 Model performance display

7. REFERENCES

[1] T. J. Stein, “Critical thinking in clinical practice,” *Child. Youth Serv. Rev.*, vol. 13, no. 5–6, pp. 441–443, 1991, doi: 10.1016/0190-7409(91)90032-d.

[2] S. Vollmer *et al.*, “Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness,” *BMJ*, vol. 368, pp. 1–12, 2020, doi: 10.1136/bmj.l6927.

[3] S. A. Alowais *et al.*, “Revolutionizing healthcare: the role of artificial intelligence in clinical practice,” *BMC Med. Educ.*, vol. 23, no. 1, pp. 1–15, 2023, doi: 10.1186/s12909-023-04698-z.

[4] R. Kang, E. M. Rantanen, and E. A. Youngstrom, “Machine Learning in Healthcare: Two Case Studies,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 66, no. 1, pp. 774–778, 2022, doi: 10.1177/1071181322661518.

[5] L. Moja *et al.*, “Effectiveness of a Hospital-Based Computerized Decision Support System on Clinician Recommendations and Patient Outcomes: A Randomized Clinical Trial,” *JAMA Netw. Open*, vol. 2, no. 12, pp. 1–16, 2019, doi: 10.1001/jamanetworkopen.2019.17094.

[6] A. Choudhury and C. Allen, “Artificially Intelligent? Machine Learning in Healthcare and Why It May Not Be As Advanced As You Think,” *Patient Saf.*, vol. 5, no. 2, pp. 76–83, 2023, doi: 10.33940/001c.77632.

[7] P. M. Rodrigues, J. P. Madeiro, and J. A. L. Marques, “Enhancing Health and Public Health through Machine Learning: Decision Support for Smarter Choices,” *Bioengineering*, vol. 10, no. 7, pp. 1–5, 2023, doi: 10.3390/bioengineering10070792.

[8] A. Zhang, L. Xing, J. Zou, and J. C. Wu, “Shifting machine learning for healthcare from development to deployment and from models to data,” *Nat. Biomed. Eng.*, vol. 6, no. 12, pp. 1330–1345, 2022, doi: 10.1038/s41551-022-00898-y.

[9] D. B. Olawade, O. J. Wada, A. C. David-Olawade, E. Kunonga, O. Abaire, and J. Ling, “Using artificial intelligence to improve public health: a narrative review,” *Front. Public Heal.*, vol. 11, no. October, pp. 1–9, 2023, doi: 10.3389/fpubh.2023.1196397.

[10] Z. Wen and H. Huang, “The potential for artificial intelligence in healthcare,” *J. Commer. Biotechnol.*, vol. 27, no. 4, pp. 217–224, 2022, doi: 10.5912/jcb1327.

[11] T. Panch, P. Szolovits, and R. Atun, “Artificial intelligence, machine learning and health systems,” *J. Glob. Health*, vol. 8, no. 2, pp. 1–8, 2018, doi: 10.7189/jogh.08.020303.

[12] F. Leonard, D. O’Sullivan, J. Gilligan, N. O’Shea, and M. J. Barrett, “Supporting clinical decision making in the emergency department for paediatric patients using machine learning: A scoping review protocol,” *PLoS One*, vol. 18, no. 11, p. e0294231, 2023, doi: 10.1371/journal.pone.0294231.

[13] S. Mudalige, D. Alexis, and C. Jayatilake, “Involvement of Machine Learning Tools in Healthcare,” vol. 2021, 2021.

[14] B. Shamreen Ahamed, M. S. Arya, and A. O. Nancy,

"Diabetes Mellitus Disease Prediction Using Machine Learning Classifiers and Techniques Using the Concept of Data Augmentation and Sampling," *Lect. Notes Networks Syst.*, vol. 516, pp. 401–413, 2023, doi: 10.1007/978-981-19-5221-0_40.

[15] J. de la Torre, J. Marin, S. Ilarri, and J. J. Marin, "Applying machine learning for healthcare: A case study on cervical pain assessment with motion capture," *Appl. Sci.*, vol. 10, no. 17, 2020, doi: 10.3390/app10175942.

[16] R. Gyebi *et al.*, "Prediction of measles patients using machine learning classifiers: a comparative study," *Bull. Natl. Res. Cent.*, vol. 47, no. 1, 2023, doi: 10.1186/s42269-023-01079-w.

[17] R. S. Hu, A. E. L. Hesham, and Q. Zou, "Machine Learning and Its Applications for Protozoal Pathogens and Protozoal Infectious Diseases," *Front. Cell. Infect. Microbiol.*, vol. 12, no. April, pp. 1–17, 2022, doi: 10.3389/fcimb.2022.882995.

[18] A. Pfob *et al.*, "Identification of breast cancer patients with pathologic complete response in the breast after neoadjuvant systemic treatment by an intelligent vacuum-assisted biopsy," *Eur. J. Cancer*, vol. 143, pp. 134–146, Jan. 2021, doi: 10.1016/J.EJCA.2020.11.006.

[19] R. Haneef *et al.*, "Use of artificial intelligence for public health surveillance: a case study to develop a machine Learning-algorithm to estimate the incidence of diabetes mellitus in France," *Arch. Public Heal.*, vol. 79, no. 1, pp. 1–13, 2021, doi: 10.1186/s13690-021-00687-0.

[20] M. Sharma, "Data Mining Prediction Techniques in Health Care Sector," *J. Phys. Conf. Ser.*, vol. 2267, no. 1, 2022, doi: 10.1088/1742-6596/2267/1/012157.

[21] "Python Logistic Regression Tutorial with Sklearn & Scikit | DataCamp." https://www.datacamp.com/tutorial/understanding-logistic-regression-python (accessed Dec. 29, 2023).

[22] "Scikit-learn SVM Tutorial with Python (Support Vector Machines) | DataCamp." https://www.datacamp.com/tutorial/svm-classification-scikit-learn-python (accessed Dec. 29, 2023).

[23] "Random Forest Classifier using Scikit-learn - GeeksforGeeks." https://www.geeksforgeeks.org/random-forest-classifier-using-scikit-learn/ (accessed Dec. 29, 2023).

[24] "Gradient Boosting with Scikit-Learn, XGBoost, LightGBM, and CatBoost - MachineLearningMastery.com." https://machinelearningmastery.com/gradient-boosting-with-scikit-learn-xgboost-lightgbm-and-catboost/ (accessed Dec. 29, 2023).

[25] "A Comprehensive Guide to Using Scikit-Learn's MLPClassifier - AITechTrend." https://aitechtrend.com/a-comprehensive-guide-to-using-scikit-learns-mlpclassifier/ (accessed Dec. 29, 2023).

[26] "Getting Started With The Extra Trees Algorithm In Python 2023." https://hands-on.cloud/getting-started-with-the-extra-trees-algorithm-in-python/ (accessed Dec. 29, 2023).

[27] "Decision Tree Classifier with Sklearn in Python • datagy." https://datagy.io/sklearn-decision-tree-classifier/ (accessed Dec. 29, 2023).

[28] "Bagging Classifier Python Code Example." https://vitalflux.com/bagging-classifier-python-code-example/ (accessed Dec. 29, 2023).

[29] "K-nearest Neighbors in Scikit-learn - KDnuggets." https://www.kdnuggets.com/2022/07/knearest-neighbors-scikitlearn.html (accessed Dec. 29, 2023).

[30] "Linear Discriminant Analysis Made Simple & How To Tutorial." https://spotintelligence.com/2023/11/07/linear-discriminant-analysis-lda/ (accessed Dec. 29, 2023).

[31] P. Mohanty, L. Patnaik, G. Nayak, and A. Dutta, "Gender difference in prevalence of hypertension among Indians across various age-groups: a report from multiple nationally representative samples," *BMC Public Health*, vol. 22, no. 1, pp. 1–10, 2022, doi: 10.1186/s12889-022-13949-5.

[32] P. Chhabra *et al.*, "Gender-specific factors associated with hypertension among women of childbearing age: Findings from a nationwide survey in India," *Front. Cardiovasc. Med.*, vol. 9, no. December, pp. 1–9, 2022, doi: 10.3389/fcvm.2022.999567.

[33] T. J. Toney-Butler and J. M. Thayer, "Nursing Process," *Fundam. Nurs. Made Incred. Easy! Second Ed.*, p. 4, Apr. 2023, doi: 10.5005/jp/books/14252_4.