Prediction of Mortality in Patients with Atrial Fibrillation: Analysis of the AFRICA Registry

**Amadou Diabagate[1]*, Yazid Hambally Yacouba[2]* , Doffou Jérôme Diako[3],
Katienefowa Sekou Koulibaly[4], Awa Fofana[4]**

[1]Faculty of Mathematics and Computer Science, University Felix Houphouët-Boigny, Abidjan, Côte d'Ivoire
[2]National High School of Architecture and Urban Planning, University of Bondoukou, Bondoukou, Côte d'Ivoire
[3]Ecole Superieure Africaine des TIC, Abidjan, Côte d'Ivoire
[4]Faculty of Medicine, University Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire
Email: *amadou1.diabagate@ufhb.edu.ci, *ahmadou.diabagate@gmail.com

Abstract

Atrial fibrillation (AF) is a leading cardiac arrhythmia associated with elevated mortality risk, particularly in low-resource settings where early risk stratification remains challenging. This study investigates the potential of supervised machine learning models to predict all-cause mortality in patients with AF, using real-world clinical data from Côte d'Ivoire. Six classification algorithms, namely Decision Tree, Random Forest, XGBoost, Support Vector Machine, K-Nearest Neighbors, and Neural Network, were evaluated using key metrics such as accuracy, F1-score, precision, recall, MCC, and AUC-ROC. Among these models, XGBoost achieved the highest overall performance, demonstrating strong calibration, robust predictive capacity, and balanced handling of class imbalance. Random Forest also performed competitively, while the Decision Tree offered a viable trade-off between interpretability and efficiency, making it suitable for clinical deployment in resource-constrained environments. The integration of SHapley Additive exPlanations (SHAP) analysis further enhanced model transparency by identifying key predictors such as the EHRA score, heart failure status, blood pressure, and left atrial diameter—variables aligned with current clinical knowledge. Despite promising results, the study acknowledges limitations, including a modest sample size, single-center design, and absence of external validation. Nevertheless, these findings underscore the feasibility of applying explainable AI methods to support early identification of high-risk AF patients and inform personalized care strategies. This work contributes to the growing body of evidence supporting AI-driven clinical decision-making and highlights the need for further validation studies, integration with real-time workflows, and enhanced model interpretability to foster trust and adoption in diverse healthcare settings.

1. Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia worldwide, affecting over 33 million individuals and markedly increasing the risk of stroke, heart failure, and premature mortality [1]. As populations age and cardiovascular risk factors proliferate, AF has become a pressing public health concern, particularly in low- and middle-income countries, where health systems often struggle to provide early diagnosis and continuous monitoring [2] [3]. In sub-Saharan Africa, this challenge is compounded by underdiagnosis and fragmented data systems, hindering the development of reliable clinical tools for risk stratification [4].

While traditional risk scores such as CHA_2DS_2-VASc and HAS-BLED are widely used in clinical practice, they rely on a limited set of static variables and assume linear relationships between risk factors, thus failing to capture the complex, non-linear dynamics typical of real-world patient populations [5] [6]. As a result, their prognostic accuracy in heterogeneous settings remains modest and may contribute to suboptimal care decisions [7].

In this context, supervised machine learning (ML) models have emerged as promising alternatives for clinical risk prediction. These approaches can process high-dimensional data and uncover subtle interactions that often escape conventional statistical methods [8] [9]. Ensemble-based algorithms such as Random Forest and XGBoost have shown superior performance in a variety of cardiovascular applications, including heart failure, stroke prediction, and mortality risk assessment [10] [11].

Nonetheless, the limited interpretability of many ML algorithms has raised concerns among clinicians, particularly regarding trust and transparency in high-stakes environments. Explainable artificial intelligence (XAI) frameworks, especially SHAP (SHapley Additive exPlanations), have gained traction as they provide individualized insights into how input variables influence predictions, thereby enhancing clinical usability and alignment with domain knowledge [12] [13].

This study seeks to contribute to the growing body of literature on AI in cardiology through three key innovations. First, it compares the predictive performance of six supervised ML classifiers, including both interpretable and complex models, for all-cause mortality in AF patients. Second, it is grounded in real-world clinical data from Côte d'Ivoire, a region where the application of AI in cardiology remains scarce [14]. Third, it integrates SHAP analysis to illuminate clinically meaningful predictors and foster trust in model interpretation.

The remainder of this paper is structured as follows. Section 2 outlines the clinical aspects of atrial fibrillation (AF), introduces the AFRICA Registry, and discusses recent work on how artificial intelligence can be used to predict clinical outcomes in AF. Sections 3 and 4 describe the methodological framework and present the empirical results. Section 5 discusses the implications, strengths, and limitations of the findings, and Section 6 concludes the study with key takeaways and future perspectives.

2. Literature Review

Atrial fibrillation (AF) is one of the most common supraventricular arrhythmias. It results from chaotic atrial electrical activity that disrupts effective atrial contraction and compromises ventricular filling [15]. This irregular rhythm promotes blood stasis, particularly in the left atrial appendage, and creates favorable conditions for thrombus formation and severe embolic events [5]. Its association with all-cause mortality is well established [16], extending beyond direct complications such as stroke, systemic embolism, and acute heart failure, to include comorbidities like hypertension, diabetes mellitus, and structural cardiac abnormalities [17]-[19]. Overall, AF nearly doubles the risk of death compared with individuals without arrhythmia [20]. This excess risk is particularly pronounced among elderly patients, women, and those not receiving adequate anticoagulation [21].

From a clinical perspective, AF is categorized as paroxysmal, persistent, permanent, or de novo, depending on the pattern and duration of arrhythmic episodes [22]. These categories have direct therapeutic implications, guiding decisions on rhythm or rate control as well as anticoagulation strategies [23]. Disease progression is often characterized by a gradual decline in left ventricular function, which may culminate in heart failure, an independent predictor of mortality [24]. This bidirectional relationship, sometimes referred to as the electromechanical vicious cycle, can be particularly detrimental in healthcare systems where diagnosis is delayed or treatment is inconsistent [25].

In sub-Saharan Africa, AF remains markedly underdiagnosed and undertreated [26]. Data from the AFRICA Registry have shown that one-year mortality after diagnosis is alarmingly high, largely due to embolic complications, limited follow-up, inadequate anticoagulation, and disparities in access to cardiology services [27] [28]. Although international guidelines recommend the systematic use of thromboembolic and bleeding risk scores such as $CHADS_2$, CHA_2DS_2-VASc, and HAS-BLED [29], their application in many African settings is inconsistent. Combining these conventional tools with artificial intelligence approaches could significantly improve early identification of patients at risk of stroke or cardiovascular death [30].

The AFRICA Registry was established precisely to fill the gap in large-scale standardized AF data across the continent [27]. As one of the first multicenter registries in Africa, it collects demographic, clinical, biological, and therapeutic information, with follow-up extending to at least twelve months [31]. This design

makes it possible to track disease progression and document severe outcomes such as ischemic stroke, heart failure–related hospitalizations, and all-cause and cardiovascular mortality. Early findings from Côte d'Ivoire, Senegal, and Cameroon report one-year mortality rates approaching 48% [28], with most deaths linked to cardioembolic complications, decompensated heart failure, and inadequate anticoagulation [32]. These results underscore persistent challenges, including delayed diagnosis, limited availability of novel anticoagulants, unequal access to specialized care, and poor treatment adherence [29]. The registry also highlights notable heterogeneity in clinical practice, particularly regarding anticoagulant use, risk-score application, and the balance between rate and rhythm control [5].

Beyond its observational value, the registry provides a robust platform for the development and validation of predictive models [33]. Its comprehensive dataset supports the training of machine learning algorithms in a real-world, multicenter African context. At the same time, it contributes to capacity building by informing clinical training, supporting the development of locally adapted guidelines, and encouraging the use of digital decision-support tools in low- and middle-income countries [34]. Planned extensions include modules on social determinants of health, barriers to treatment adherence, and cost-effectiveness analyses of anticoagulation strategies [35]. In the long term, the registry aims to evolve into a continental platform for surveillance, evaluation, and improvement of arrhythmia management within broader public health strategies [30].

Despite the widespread use of conventional risk scores, their predictive accuracy for all-cause mortality remains limited. These tools are based on simple additive models and fail to capture the complex, nonlinear interactions among risk factors [36] [37]. Machine learning approaches have the potential to overcome these limitations. Algorithms such as Random Forest, Decision Tree, Support Vector Machine, XGBoost, and deep neural networks have demonstrated strong ability to model mortality risk from longitudinal clinical data [38]-[41]. They are capable of detecting subtle interactions, handling imbalanced datasets, and integrating diverse structured and unstructured variables without relying on predefined assumptions.

Several recent studies illustrate this potential. Kao and colleagues reported the excellent performance of XGBoost in predicting AF risk among elderly patients using electronic medical records [42]. Zhang and collaborators showed that ensemble models consistently outperformed conventional risk scores in predicting fatal outcomes [43]. Other investigations have identified additional mortality predictors such as anemia, inflammatory markers, socioeconomic status, and polypharmacy, which are often overlooked by traditional clinical tools [44] [45]. While deep learning models generally require large datasets, they have demonstrated remarkable accuracy when applied to electrocardiograms, imaging, and sequential clinical data, with some models exceeding 90% accuracy for in-hospital mortality prediction [46] [47].

Yet, the application of artificial intelligence in African clinical settings remains

rare, primarily due to limited access to large, high-quality datasets and underdeveloped digital infrastructure [48]. Initiatives such as the AFRICA Registry [49] are therefore invaluable, providing the foundation needed to develop predictive models adapted to regional realities. Building on this foundation, the present study evaluates the feasibility of using machine learning to predict one-year mortality in a cohort of Ivorian AF patients. Recent studies have also shown that this registry can directly support the development of predictive models tailored to African healthcare needs [50]. By focusing on mortality, a dimension often overlooked at the intersection of AF and AI, this work addresses a critical gap in knowledge and proposes concrete avenues for advancing personalized, data-driven medicine in resource-limited contexts.

3. Methodology

This section introduces the methodological approach adopted to predict mortality in patients with atrial fibrillation. The proposed workflow follows a supervised machine learning pipeline that includes data preprocessing, class balancing, model training, and predictive evaluation. Each step was implemented using standard tools and good practices to ensure both accuracy and reproducibility. **Figure 1** below provides an overview of the system's architecture, illustrating the overall workflow.
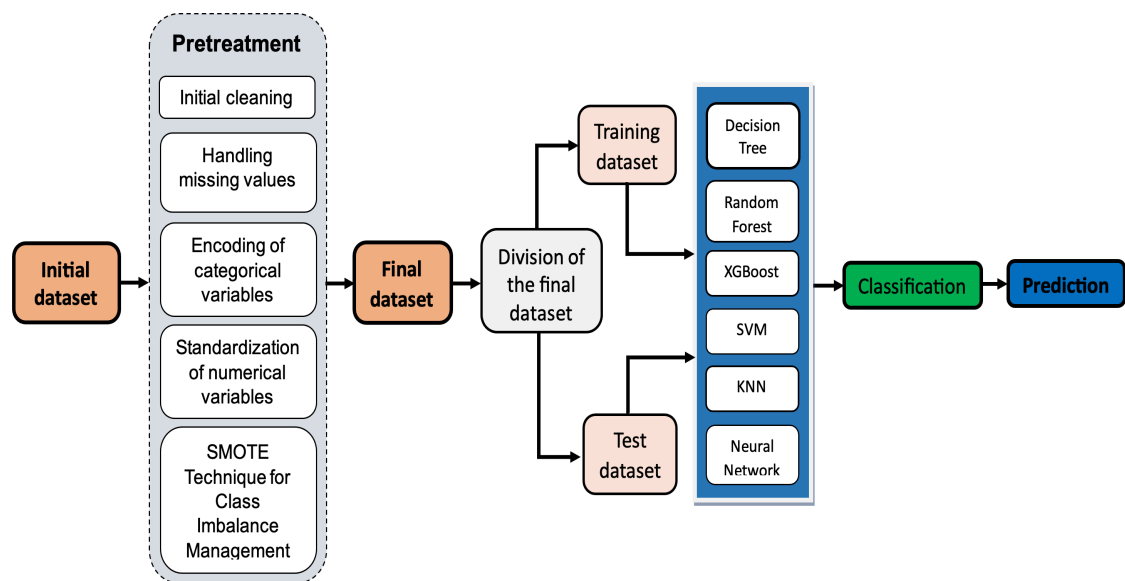

Figure 1. Overview of the machine learning pipeline used.

To operationalize this workflow, the pseudocode below details each technical stage used to build, validate, and explain the predictive model.

Algorithm 1—Machine Learning Workflow for Mortality Prediction
Input:
Structured clinical dataset of atrial fibrillation (AF) patients
Output:

Mortality prediction and interpretability insights

Step 1: Import essential libraries and packages

Step 2: Load the dataset and remove

- Irrelevant columns

- Columns with to many (>100) missing values

Step 3: Handle missing data

Step 4: Encode categorical variables and normalize numerical features

Step 5: Split dataset into training (80%) and testing (20%)

Step 6: Address class imbalance using SMOTE

- Apply oversampling only on the training set

- Set target minority-to-majority ratio (e.g., 0.75)f

Step 7: Define candidate machine learning models

- Decision Tree, Random Forest, XGBoost, SVM, KNN, Neural Network

Step 8: Perform hyperparameter tuning using RandomizedSearchCV

- Identify optimal configurations for each model

Step 9: Train and validate models using Stratified K-Fold Cross-Validation

Step 10: Evaluate performance on key metrics

- Accuracy, F1-score, Precision, Recall, AUC-ROC, MCC, Log-loss

Step 11: Interpret best-performing model using SHAP

- Identify most influential features

- Visualize variable impacts and directionality

To prevent any information leakage, all preprocessing steps were performed separately within each cross-validation fold. In each fold, missing values in the training set were first imputed, after which the Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the outcome classes. Feature scaling was then carried out using parameters derived exclusively from the training data, and these transformations were applied to the corresponding test set. This ensured that oversampling, scaling, and model training were all based solely on training data for each fold.

3.1. Clinical Dataset Overview

The dataset employed in this study originates from the AFRICA registry (Atrial Fibrillation Registry in Countries of Africa), a large-scale continental initiative designed to standardize the collection of clinical data on atrial fibrillation (AF) across Africa. For the purposes of this research, data were drawn specifically from the Ivorian cohort of the registry. This extraction covers two time periods: from January 1, 2016 to January 31, 2018, and from January 2021 to December 2023.

This dataset was compiled through a multicenter study conducted at several major healthcare institutions in Côte d'Ivoire, including the Institut de Cardiologie d'Abidjan (ICA), Centre Médical Cardio-Respiratoire des Jardins de Cocody (CMCARE), Home Medical Service du Plateau (HMS), Clinique Saintes Myriades de Marcory, Polyclinique Internationale Sainte Anne-Marie (PISAM), Hôpital Militaire d'Abidjan (HMA), CHU de Yopougon and CHU de Bouaké.

Eligible participants were adult patients (aged 18 and above) residing in Côte d'Ivoire, with a confirmed diagnosis of AF documented by either standard electrocardiogram (ECG) or Holter ECG monitoring. Only individuals with accessible clinical follow-up for a minimum of 12 months and who had provided informed consent were included. Exclusion criteria comprised pregnant women, patients not formally enrolled in the registry, and cases of situational or transient AF.

1) Collected Variables and data structure

This dataset encompasses a broad spectrum of information, including demographic, clinical, biological, and therapeutic data, alongside records of disease progression and patient outcomes. It offers a static yet representative snapshot of patient profiles, drawing from both general and structural elements found in the medical documentation.

The general descriptors provide contextual and epidemiological insights into the study population. The primary condition under investigation is atrial fibrillation (AF), a common arrhythmia characterized by rapid and disorganized atrial activity. The dataset includes distinctions between AF subtypes, paroxysmal, persistent, and permanent, each reflecting different clinical presentations and disease trajectories. Patients are categorized according to age ranges and sex (male or female), providing a foundational stratification for analysis. Key clinical variables captured in the dataset include:

Risk scores such as $CHADS_2$ and CHA_2DS_2-VASc, which estimate thromboembolic risk, and the EHRA score, which quantifies symptom burden;

Comorbid conditions, notably hypertension, diabetes mellitus, and heart failure;

AF subtypes: paroxysmal (self-terminating episodes), persistent (requiring medical intervention), Long-standing persistent (sustained despite multiple cardioversion attempts), and permanent (accepted as the baseline rhythm).

Clinical outcomes tracked over the observation period include rates of hospitalization, mortality, and major complications like ischemic stroke or heart failure exacerbation. Therapeutic strategies are also documented, spanning rhythm control (e.g., cardioversion and antiarrhythmic therapy), rate control (using chronotropic agents), and anticoagulation approaches (including both vitamin K antagonists and direct oral anticoagulants). The organizational structure of the dataset reflects a patient-level observational framework, with each unit corresponding to an individual followed for at least one year. Variables such as age, gender, residence, comorbidities, clinical scores, and outcome events were collected using two main methods:

Prospective clinical data collection during routine consultations and hospital admissions;

Digital entry through standardized forms and electronic platforms aligned with the AFRICA registry protocol.

Over the study period, 239 patients with a confirmed diagnosis of atrial fibrillation were recorded in the AFRICA Registry. All of these met the eligi-

bility criteria and were included in the analysis. No exclusions were made at any stage. In the raw dataset, the completeness of information varied across variables: some were fully documented for all patients, while others showed high proportions of missing values, particularly when the variables were not applicable to the entire cohort. These missing values were handled during preprocessing, and the final dataset used for machine learning contained complete information for all variables retained in the analysis.

Table 1 provides a structured overview of the collected variables, detailing their types, codes, and definitions.

Table 1. Overview of key variables in the dataset.

Data Type	Name	Coding	Description
Demographics	Age	Numeric (years)	Age of the patient at the time of atrial fibrillation diagnosis.
	Sex	1 = Male, 2 = Female	Biological sex of the patient.
	Location	Text (City or Region)	Geographic area where the patient received care.
Clinical	Type of Atrial Fibrillation	1 = Paroxysmal, 2 = Persistent, 3= Long-standing persistent, 4 = Permanent	Classification based on the duration and stability of the arrhythmia.
	Symptoms	1 = Palpitations, 2 = Dyspnea, 3 = Fatigue, 4 = Syncope	Main symptoms reported by patients at presentation.
	EHRA Score	1 = None, 2 = Mild, 3 = Severe, 4 = Disabling	Functional classification of symptom burden.
	CHADS$_2$ Score	Numeric (0 - 6)	Score estimating stroke risk based on comorbidities.
	CHA$_2$DS$_2$-VASc Score	Numeric (0 - 9)	Extended stroke risk score including additional clinical factors.
	Comorbidities	1 = Hypertension, 2 = Diabetes, 3 = Heart Failure, 4 = Thrombocytopenia, etc.	Coexisting medical conditions that impact clinical risk.
	HAS-BLED Score	Numeric (0 - 9)	Score estimating bleeding risk under anticoagulation therapy.
Therapeutics	Type of Treatment	1 = Flecainide, 2 = Amiodarone, 3 = Anticoagulants, etc.	Pharmacological strategies used for AF management.
	Cardioversion Method	1 = Pharmacological, 2 = Electrical	Method employed to restore sinus rhythm.
Outcomes	Complications	1 = Stroke, 2 = Thromboembolism, 3 = Bleeding, etc.	Adverse events occurring during the follow-up period.
	Death	0 = No, 1 = Yes	Indicates whether the patient died within one year of follow-up.
	Cause of Death	Text (Narrative)	Specific clinical reason for death, when applicable.
	Hospitalizations	Numeric (Count)	Total number of hospital admissions recorded during the study.
Paraclinical Tests	ECG Results	Text or Code	Electrocardiogram findings documenting AF or related abnormalities.
	Echocardiography	Numeric and/or Text	Echographic parameters including structural heart measures and thrombus status.
Biological	Laboratory Values	Numeric	Blood test results used to support diagnosis and treatment decisions.

2) Descriptive Statistics and Patient Characteristics

The descriptive statistical analysis, presented in Table 2, highlights the considerable clinical and biological heterogeneity among patients diagnosed with atrial fibrillation (AF) in this cohort.

Table 2. Summary of descriptive statistics for key clinical variables.

Variable	Mean	Variance	Standard Deviation	Median	Mode	Range	Min	Max
Left Atrium Mean Diameter	48.28	50.47	7.10	48.00	48.00	68.00	6.00	74.00
Patient Weight	67.84	180.72	13.44	68.00	68.00	120.00	10.00	130.00
Diastolic Blood Pressure	84.26	246.34	15.70	80.00	80.00	127.00	3.00	130.00
Systolic Blood Pressure	131.05	538.24	23.20	130.00	120.00	128.00	82.00	210.00
Heart Rate	83.87	552.84	23.51	80.50	80.50	170.00	6.00	176.00
Patient Age	61.74	248.89	15.78	65.00	67.00	79.00	12.00	91.00
Left Ventricular End Diastolic Volume	54.20	82.74	9.10	54.00	54.00	86.00	5.00	91.00
Patient Height	167.89	34.36	5.86	168.00	168.00	45.00	150.00	195.00
EHRA Score	2.19	0.71	0.84	2.00	2.00	3.00	1.00	4.00
CHA2DS2VAS Score	2.99	2.09	1.45	3.00	3.00	9.00	0.00	9.00

The statistical exploration of the quantitative variables reveals meaningful insights into the clinical and demographic characteristics of the patient population included in this study. The average left atrial diameter is 48.28 mm (SD = 7.10), with values ranging from 6 to 74 mm. This wide dispersion suggests heterogeneous atrial remodeling, which is a known structural hallmark in patients with atrial fibrillation.

The mean patient weight is 67.84 kg, with a relatively high variability (SD = 13.44), extending from 10 to 130 kg. This dispersion highlights the need for careful adjustment in any clinical model, as body weight may influence drug dosing and cardiovascular risk stratification. Similarly, patient height is relatively homogeneous (mean = 167.89 cm; SD = 5.86), indicating a more consistent distribution within the cohort.

Regarding blood pressure, the diastolic pressure has a mean of 84.26 mmHg and a SD of 15.70, while the systolic pressure exhibits a wider spread (mean = 131.05 mmHg; SD = 23.20). The extreme values observed (e.g., diastolic BP as low as 3 mmHg and systolic BP up to 210 mmHg) suggest the presence of both hypotensive and severely hypertensive individuals, underlining the clinical diversity of the cohort.

The heart rate displays substantial variability (mean = 83.87 bpm; SD = 23.51), with extreme values ranging from 6 to 176 bpm. This may reflect a mixture of patients in different rhythm control states (e.g., bradycardic or tachyarrhythmic), which is common in atrial fibrillation populations.

The average patient age is 61.74 years (SD = 15.78), with a wide age range (12 to 91 years). Notably, the median age is slightly higher (65 years), indicating a mild right skew in age distribution, consistent with the fact that atrial fibrillation becomes more prevalent in older adults.

With respect to cardiac function, the left ventricular end-diastolic volume (LVEDV) averages 54.20 mL (SD = 9.10), within expected physiological ranges, though extreme low values (as low as 5 mL) may warrant further clinical interpretation or data validation.

In terms of risk stratification scores, the EHRA score, which reflects symptom severity, averages 2.19 (SD = 0.84), suggesting that most patients experience moderate symptoms. The CHA_2DS_2-VASc score, a pivotal metric for stroke risk estimation, has a mean value of 2.99 (SD = 1.45), with some patients exhibiting a score as high as 9, indicating a high-risk subgroup.

Collectively, these statistics portray a clinically diverse population with substantial variability across several key parameters. This heterogeneity reinforces the need for robust predictive modeling approaches that account for both physiological extremes and central tendencies within the dataset.

The categorical variables provide a nuanced understanding of the clinical reality faced by patients with atrial fibrillation in this cohort. Figure 2 shows that heart failure affected nearly half of the study population (45.6%), underscoring its substantial burden in this clinical context. This high prevalence calls for strengthened diagnostic and therapeutic strategies to address this comorbidity within integrated care pathways. Hospitalization was reported in 67.4% of cases, reflecting the high level of clinical instability and the frequency of acute episodes requiring inpatient care. A closer look at the type of atrial fibrillation reveals that 65.3% of patients presented with the permanent form, far exceeding the proportions of paroxysmal (12.1%), persistent (15.1%), and long-standing persistent AF (7.5%). This predominance suggests a population with advanced arrhythmic profiles, for whom rhythm control is often no longer feasible and management relies largely on rate control and symptom palliation.
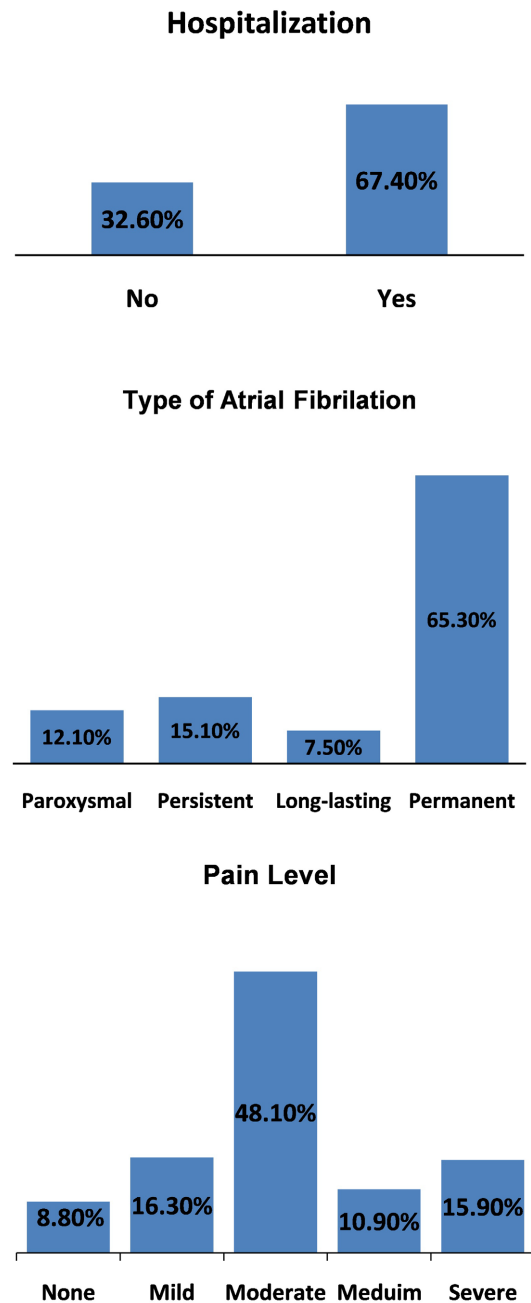
Heart Failure

Hospitalization

Type of Atrial Fibrilation
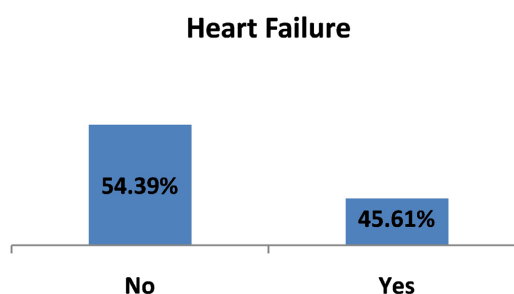
Pain Level

Figure 2. Graphical representation of selected qualitative variables.

Symptom burden, as captured by self-reported pain levels, further illustrates the vulnerability of this population. Nearly half of the patients (48.1%) reported moderate pain, while 15.9% experienced severe discomfort. Only 8.8% declared an absence of pain, indicating that the majority of individuals live with a considerable degree of physical suffering, which may compound the clinical risk and affect adherence to treatment.

These distributions confirm that the study population is characterized by significant clinical deterioration, both in terms of disease progression and quality of

life. Such a context strongly supports the use of predictive models to anticipate adverse outcomes and guide timely, personalized interventions.

3) Correlation Matrix Analysis of Key Predictive Features

Figure 3 presents the heatmap of correlations among the 15 most important variables identified by the Random Forest model reveals several clinically meaningful associations that support the robustness of the feature selection process. The correlation analysis among provides valuable insights into the underlying relationships within the patient population. As expected, a moderate positive correlation is observed between systolic and diastolic blood pressure (r = 0.58), reflecting the physiological coupling of these hemodynamic parameters. Similarly, the CHA_2DS_2-VASc score demonstrates a positive association with patient age (r = 0.42), which is consistent with its scoring criteria that explicitly integrate age as a contributing factor to thromboembolic risk.
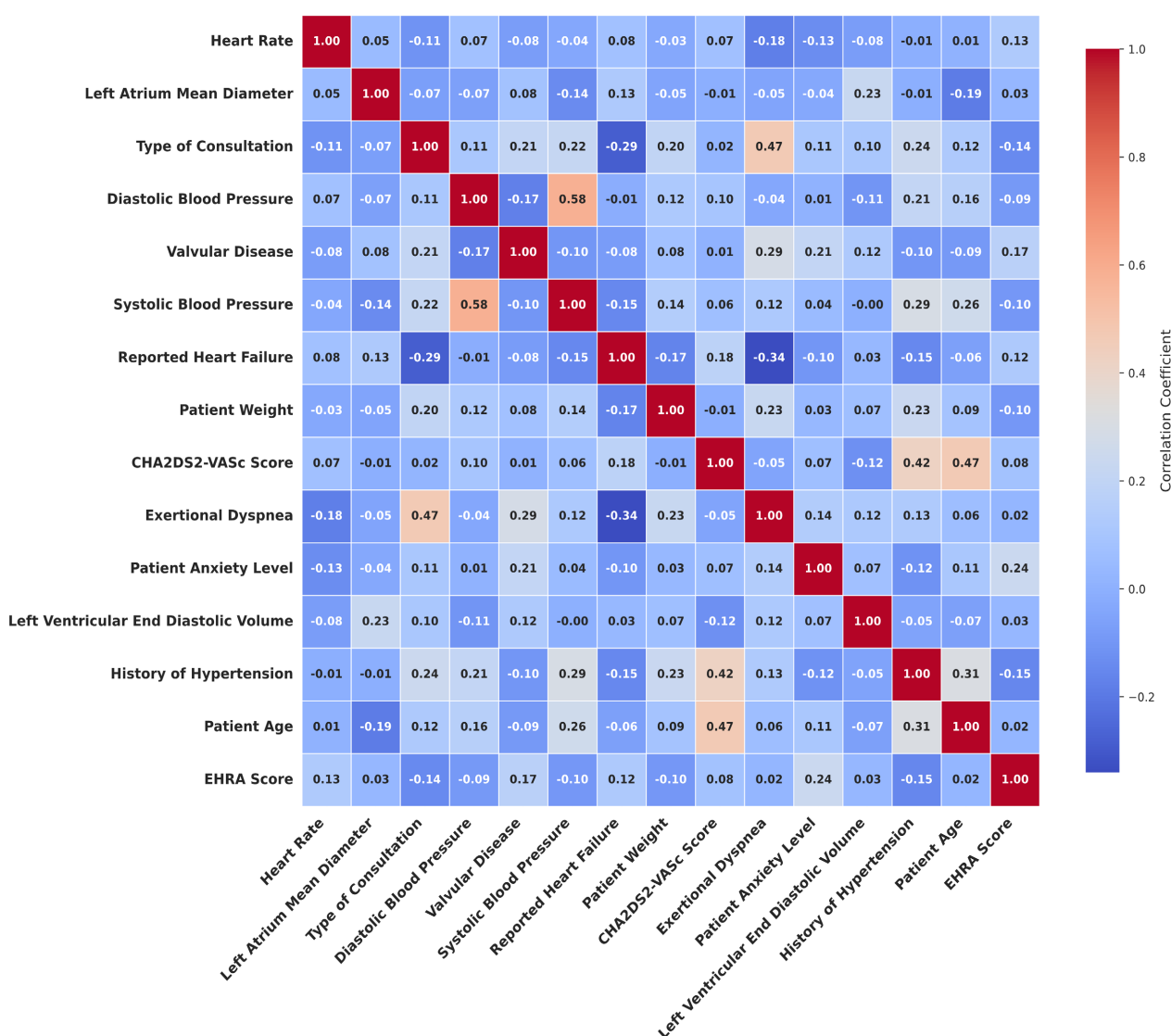

Figure 3. Heatmap of the correlation between the top 15 predictive features.

Another notable finding is the moderate correlation between exertional dyspnea and type of consultation (r = 0.47), suggesting that patients presenting with more severe symptoms are more likely to require urgent or specialized clinical encounters. This relationship may also reflect a bias toward more intensive evaluation among symptomatic individuals.

The left atrial diameter appears weakly associated with the left ventricular end diastolic volume (r = 0.23), indicating that atrial enlargement may partly reflect ventricular structural changes, albeit with modest intensity. Interestingly, patient weight shows very weak or no meaningful correlation with most variables, highlighting its relative independence in this clinical context.

Anxiety levels and pain-related indicators, such as the EHRA score, do not exhibit strong correlations with traditional clinical metrics, suggesting that symptom perception may follow a distinct psychological or subjective trajectory, decoupled from purely physiological markers.

Overall, the matrix reveals predominantly low correlation coefficients, which supports the use of multivariate modeling approaches such as machine learning. The weak linear dependencies between variables reduce the risk of multicollinearity, thereby enhancing the robustness of predictive models. These findings justify the integration of diverse and partially independent features into mortality prediction frameworks for patients with atrial fibrillation.

3.2. Data Preprocessing and Ethical Compliance

Prior to model training, a thorough data preprocessing phase was conducted to ensure quality and consistency. Initial cleaning involved the removal of columns with excessive missing values or low predictive relevance. For the retained variables, appropriate imputation strategies were applied to handle missing data: categorical variables were filled based on the most frequent values, while numerical variables were imputed using the median, mode, or distribution-based constants (e.g., 0 or 1), depending on the shape and clinical relevance of the distribution. All categorical minority class corresponds to high-risk conditions.

The final dataset, free of missing values and composed entirely of normalized numerical features, was used for subsequent training and evaluation. All procedures were conducted in compliance with the ethical principles outlined in the Declaration of Helsinki, and the dataset was fully anonymized prior to analysis to safeguard patient confidentiality.

3.3. Machine Learning Methods

This study incorporated six supervised machine learning algorithms, chosen to reflect a diverse range of modeling strategies. This approach ensures a comprehensive evaluation of predictive performance, model robustness, interpretability, and computational efficiency. It aligns with recent literature highlighting the importance of adapting model selection to the structure and complexity of the clinical problem [51] [52].

Random Forests, as introduced by Breiman [40], operate by constructing multiple decision trees using resampled subsets of the training data. The final prediction is obtained through a majority vote across all trees, enhancing the model's stability and resistance to overfitting. The method also provides an estimate of generalization error using Out-Of-Bag (OOB) samples and quantifies the relative importance of input features, offering valuable insights into the drivers of prediction [53].

Multilayer perceptrons, a widely used architecture in neural networks, consist of interconnected layers of neurons that apply nonlinear transformations to their inputs. This structure enables the model to capture intricate patterns and interactions among variables. Neural networks have demonstrated strong performance in medical classification tasks, particularly when relationships between predictors and outcomes are highly nonlinear [54] [55].

Decision trees classify observations by sequentially partitioning the feature space according to thresholds that maximize class separation. Each terminal node, or leaf, represents a predicted class or outcome based on the distribution of training instances. This method has been extensively used in clinical decision support because of its transparency and ease of interpretation [56] [57].

Support Vector Machines (SVM) aim to find the optimal hyperplane that separates data points belonging to different classes with the largest possible margin. When data are not linearly separable, SVM can map the original feature space into higher dimensions using kernel functions. This flexibility makes it particularly effective in domains with complex boundaries between classes [42].

XGBoost, an optimized form of gradient boosting developed by Chen and Guestrin [41], builds additive models using decision trees that minimize a regularized loss function. It incorporates advanced techniques such as shrinkage, subsampling, and second-order optimization to improve convergence and generalization. XGBoost is recognized for its superior performance on structured datasets and is frequently used in competitive modeling environments [58].

K-Nearest Neighbors (KNN) classifies new observations by identifying the most common class among a set of nearby examples in the feature space. The distance metric, typically Euclidean, determines proximity. Although simple, KNN can perform well in cases where class boundaries are locally consistent and the dataset is not highly dimensional.

As a point of reference, we added a very simple baseline model using Scikit-learn's DummyClassifier, configured with the most_frequent strategy. This model, while clearly not suitable for any practical application, simply predicts the most common class in the training data. Its purpose is not performance, but perspective, it helps illustrate what minimal predictive ability looks like, making it easier to appreciate the real contributions of the more sophisticated algorithms tested in this study.

To ensure fair comparison across models, we applied the conventional probability threshold of 0.50 to distinguish between survival and mortality. This choice

was made to maintain consistency among the six machine learning methods as well as the baseline DummyClassifier. We recognize, however, that in clinical contexts alternative thresholds, such as those adjusted to balance sensitivity and specificity, could provide additional relevance.

3.4. Hyperparameter Optimization and Model Selection

In this study, multiple machine learning algorithms were investigated to predict mortality using a dataset composed of 239 observations and 66 clinical variables. To optimize the predictive accuracy of each model, a comprehensive hyperparameter tuning phase was conducted using the RandomizedSearchCV method. This technique enables an efficient exploration of the hyperparameter space by sampling configurations probabilistically, thereby reducing computational cost compared to the exhaustive GridSearchCV.

Table 3 provides a concise overview, highlighting each method along with the hyperparameter values that achieved the best performance.

Table 3. Summary of machine learning models and their optimized hyperparameter settings.

Model	Optimized Hyperparameters and Values
Decision Tree	criterion = "entropy", max_depth = None, min_samples_leaf = 1, min_samples_split = 5
Random Forest	bootstrap = True, max_depth = None, max_features = "log2", min_samples_leaf = 2, min_samples_split = 9, n_estimators = 129
XGBoost	colsample_bytree = 0.9631, gamma = 0.3255, learning_rate = 0.2845, max_depth = 3, n_estimators = 427, subsample = 0.7247
SVM	C = 3.7554, gamma = "scale", kernel = "rbf"
KNN	metric = "manhattan", n_neighbors = 6, weights = "distance"
Neural Network	activation = "relu", alpha = 0.0061, early_stopping = True, hidden_layer_sizes = (100,), learning_rate = "constant", n_iter_no_change = 10, solver = "adam"

Hyperparameter tuning was performed using cross-validation to ensure that the selected configurations maximize the models' generalization capacity. This step is critical for ensuring the reproducibility of the results and establishing a robust foundation for the comparative analysis of model performance.

To rigorously assess model performance, we implemented a Stratified K-Fold cross-validation strategy. This method partitions the dataset into K subsets (folds) while preserving the class distribution across each fold, a particularly important consideration in imbalanced classification tasks. During each iteration, one fold is used for testing, while the remaining $K–1$ folds serve as the training set. This process is repeated K times, and the results are averaged to provide a more robust estimate of model performance.

In our case, we set K to 10, a commonly recommended choice in the literature [59] [60], which offers a good trade-off between bias and variance. This evaluation

protocol plays a central role in ensuring fairness and consistency across models when comparing their predictive capabilities.

3.5. Performances Evaluation

In classification tasks, evaluating the effectiveness of predictive models relies on four fundamental outcomes: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These metrics form the basis for several performance indicators that help determine the robustness and clinical relevance of the model. Below, we outline the mathematical definitions of some of the most widely used evaluation metrics applied in this study.

Accuracy, or overall correctness, reflects the proportion of correct predictions—both positive and negative—over the total number of instances. It provides a general measure of how often the model makes correct predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Recall, also referred to as sensitivity or the true positive rate, assesses the model's ability to correctly identify all actual positive cases. It is particularly relevant in clinical contexts where missing positive instances carries serious implications:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

Precision quantifies the proportion of correctly predicted positive cases among all cases predicted as positive. It reflects the reliability of positive predictions made by the model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

F1-score represents the harmonic mean between Precision and Recall. It offers a balanced metric that is especially useful in datasets with class imbalance, as it accounts for both false positives and false negatives.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{4}$$

Together, these metrics provide a comprehensive assessment of model performance, allowing for informed comparisons and selection of the most clinically effective algorithm.

4. Results

A total of six machine-learning algorithms were evaluated to predict the onset of mortality in patients with atrial fibrillation. Performance was assessed via stratified 10-fold cross–validation to ensure robust estimates.

4.1. Model Performance Comparison

Table 4 summarizes the cross-validated results for each algorithm across key met-

rics: accuracy, f1-score, precision, recall, specificity, AUC_ROC, MCC and logloss.

Table 4. Cross-Validation performance of machine learning models across multiple metrics.

Model	Accuracy	f1_score	Precision	Recall	Specificity	AUC_ROC	MCC	LogLoss
Dummy Classifier	62.76	22.62	26.00	20.00	78.72	49.36	−1.4	13.43
Decision Tree	67.38	43.29	41.11	44.97	71.43	61.41	19.42	1298.41
Random Forest	74.06	22.06	72.22	15.63	82.86	73.04	19.83	53.43
XGBOOST	75.32	43.90	57.77	35.42	80.00	70.76	30.66	85.09
SVM Classifier	72.81	0.00	0.00	0.00	85.71	53.67	0.00	58.63
KNN	68.18	28.67	35.73	24.76	80.00	56.19	9.99	267.91
Neural Network	72.81	12.16	6.78	50.00	80.00	52.46	0.00	1813.16

The Dummy Classifier, used as a baseline, showed the weakest overall performance. Its accuracy was 62.76 percent, with an F1-score of only 22.62 percent. Precision reached 26.00 percent and recall 20.00 percent, while specificity was relatively high at 78.72 percent. However, the AUC-ROC of 49.36 percent and the negative MCC of −1.40 percent confirm that this model performs essentially at the level of random guessing. The log-loss of 13.43 further indicates poor probability calibration. This baseline is important because it allows the improvements achieved by more sophisticated models to be clearly appreciated.

The comparative evaluation of six classification models reveals distinct differences in their ability to predict mortality in patients with atrial fibrillation. XGBoost offered the most balanced performance, with the highest accuracy (75.32%), a strong F1-score (43.90%), and the best MCC (30.66%). Its AUC-ROC of 70.76% reflects a solid discriminative capacity, although the recall (35.42%) suggests some mortality cases remained undetected. Random Forest also performed well in terms of accuracy (74.06%) and precision (72.22%), highlighting its ability to limit false positives. However, its recall (15.63%) was notably low, reducing its usefulness for early identification of at-risk patients. Still, its low log-loss (53.43) and strong AUC (73.04%) suggest reliable probability estimates. The Decision Tree showed more modest results, with an acceptable recall (44.97%) and specificity (71.43%), but its high log-loss (1298.41) and moderate MCC (19.42) point to limited predictive stability. KNN and Neural Network models yielded less consistent results. KNN had low recall (24.76%) and weak discrimination (AUC = 56.19%), while the Neural Network displayed high recall (50.00%) but very poor precision (6.78%) and extremely high log-loss (1813.16), suggesting unstable output. The SVM classifier failed to identify any death cases, with a recall and F1-score of zero, despite a high specificity (85.71%), making it unsuitable for this clinical application.

Overall, the inclusion of the Dummy Classifier confirms that all other machine

learning models provide tangible improvements over random prediction. In summary, XGBoost and Random Forest stood out as the most promising candidates for mortality prediction in this cohort. XGBoost demonstrated the best overall balance between sensitivity, discrimination, and calibration, while Random Forest achieved high precision and strong AUC-ROC, despite a lower recall. These findings provide a valuable starting point for building predictive tools to support clinical decision-making in patients with atrial fibrillation.

These comparisons are further illustrated in **Figure 4**, which provides a visual overview of the results by metric.
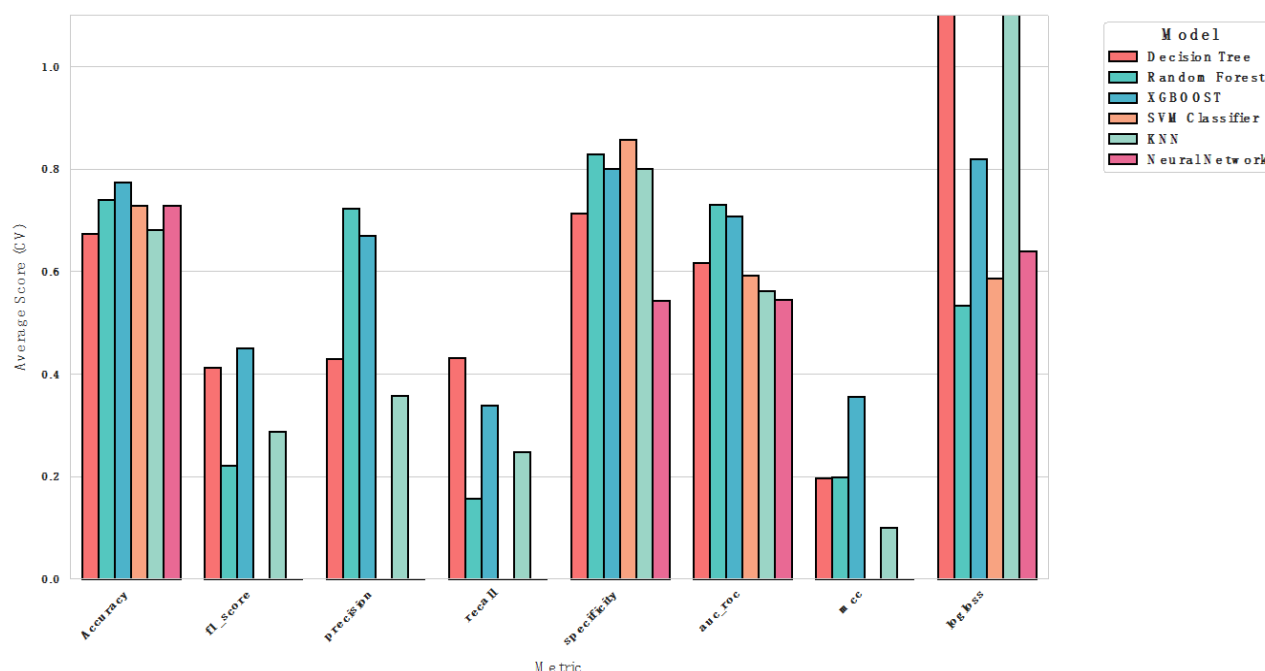

Figure 4. Comparative Barplot of Model Performance by Metric Using Cross-Validation.

4.2. Confusion-Matrix Analysis

Figure 5 presents the confusion matrices of the machine learning models obtained through cross-validation. These visualizations provide a detailed view of each model's behavior in predicting mortality among patients with atrial fibrillation. The Decision Tree and XGBoost classifiers exhibit the most balanced confusion matrices, correctly identifying 49 out of 53 mortality cases (true positives), while misclassifying only 3 to 4 non-death cases as positive. These results translate into strong recall and low false-positive rates, reinforcing the models' clinical relevance in detecting at-risk patients without excessively flagging false alarms. The Random Forest model, while achieving a slightly higher overall accuracy, reveals a tendency toward conservative prediction. It identifies only 40 deaths correctly and misclassifies 13 actual deaths as non-deaths, suggesting a high precision but limited sensitivity, a trade-off that may not be ideal in mortality prediction where false negatives carry significant consequences.
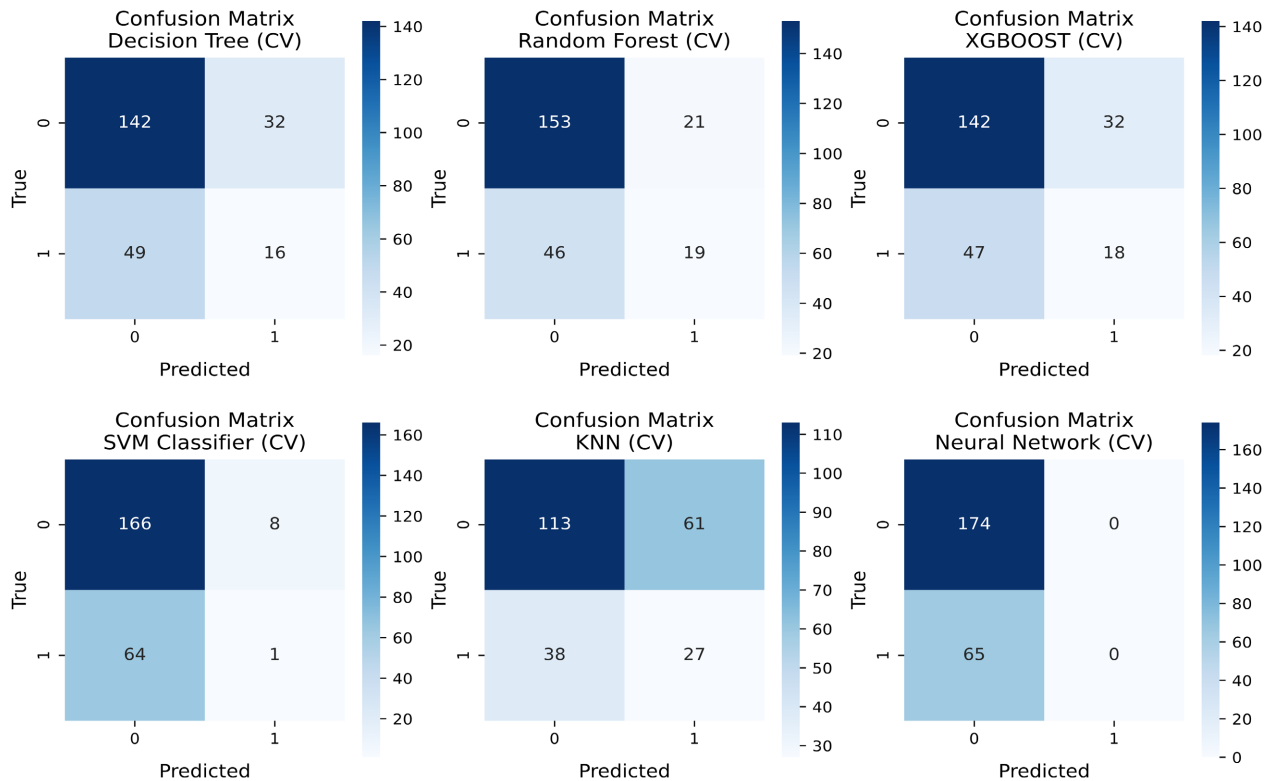
Figure 5. Confusion matrices of the machine learning models (Cross-Validation).

In contrast, the SVM classifier underperforms considerably, capturing only 8 of the 53 death cases while misclassifying 6 non-deaths. This results in a very low recall and renders the model inadequate for clinical applications where sensitivity is crucial. The K-Nearest Neighbors (KNN) model offers a slightly better balance, with 22 correct identifications of death cases, but also suffers from a relatively high number of false positives (43). Its tendency to misclassify many surviving patients could lead to unnecessary clinical alerts. Lastly, the Neural Network demonstrates the most unbalanced performance: although it identifies 17 of the 53 deaths, it misclassifies a large number of survivors (56), suggesting poor discrimination and overfitting to certain patterns in the data.

Overall, XGBoost emerges as the most robust and clinically viable model, offering an optimal compromise between true positive identification and false positive minimization, followed by the Decision Tree. These findings are consistent with prior quantitative performance metrics and support their integration in predictive workflows.

4.3. Feature Importance

The feature importance analysis provides valuable insights into the variables that most strongly influenced the model's ability to predict mortality in patients with atrial fibrillation.

Figure 6 highlights the top 15 features ranked by the Random Forest model and visually underscores their relative importance in predicting mortality among pa-
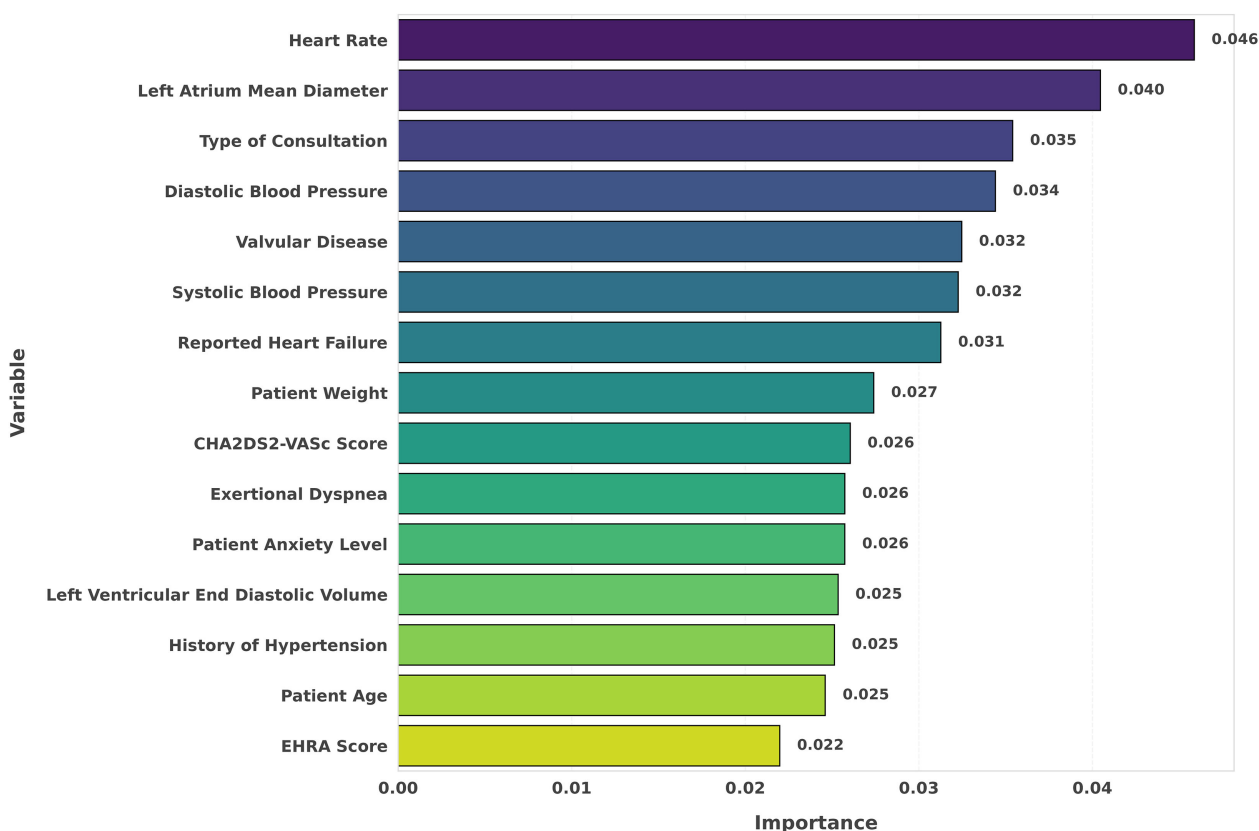
tients with atrial fibrillation.

Figure 6. Top 15 most important features according to random forest.

At the top of the ranking, the EHRA score emerged as the most influential predictor, underscoring the clinical significance of symptom severity in determining patient outcomes. This score, which reflects the impact of atrial fibrillation on daily activities, appears to encapsulate a multifaceted view of patient status, making it a powerful mortality indicator. Closely following are diastolic blood pressure and left atrial mean diameter, both of which are known to reflect underlying hemodynamic and structural cardiac stress. Their high predictive weight reinforces existing evidence linking elevated blood pressure and atrial remodeling to adverse outcomes. Heart rate and reported heart failure also occupy prominent positions, further confirming the central role of cardiac function and rhythm control in mortality risk. Notably, patient age and systolic blood pressure, while traditionally emphasized in cardiovascular risk stratification, appear slightly less discriminative than some symptom-based or structural markers in this context, though they still contribute meaningfully.

The model also attributes notable importance to patient weight and anxiety level, suggesting that broader physiological and psychosocial parameters may play non-negligible roles in prognosis. Interestingly, CHA$_2$DS$_2$-VASc score, a standard tool for stroke risk in atrial fibrillation, appears with moderate influence, potentially reflecting its partial overlap with other, more directly predictive features in

the dataset. Lower-ranked yet still relevant features include left ventricular end-diastolic volume, type of consultation, and valvular disease, which may signal underlying cardiovascular conditions without providing direct mortality risk signals. At the bottom of the ranking, exertional dyspnea and history of hypertension demonstrated minimal influence in this model, suggesting that, while clinically relevant, they offer limited incremental value beyond other more dominant predictors.

Overall, the distribution of feature importance highlights a complex interplay between symptom burden, cardiac structure and function, and patient demographics, supporting the multifactorial nature of mortality risk in atrial fibrillation.

4.4. ROC-Curve Analysis

The ROC (Receiver Operating Characteristic) curves provide a comparative visualization of each model's ability to distinguish between deceased and surviving patients. Among the six classifiers evaluated, Random Forest achieved the highest Area Under the Curve (AUC = 0.71), indicating a relatively good trade-off between sensitivity and specificity. This suggests that the Random Forest model demonstrated the most robust discriminative power in identifying high-risk patients within the studied cohort. Decision Tree and XGBOOST models followed closely, with AUC scores of 0.66 and 0.65, respectively. While slightly lower than Random Forest, these values still reflect moderate classification performance, reinforcing their utility in mortality prediction when interpretability or speed is prioritized.

In contrast, K-Nearest Neighbors (KNN) yielded an identical AUC score of 0.65 but with a less stable curve, reflecting more variability in threshold behavior. The Support Vector Machine (SVM) and Neural Network models both performed poorly, with AUC scores of 0.50, equivalent to random guessing, indicating a lack of predictive power in this specific clinical context. This underperformance may be attributed to limited sample size, suboptimal hyperparameters, or the inability of these models to capture the underlying structure of the data given its imbalanced nature.

Overall, the ROC analysis confirms that tree-based ensemble methods, particularly Random Forest, are better suited for this classification task. The clear separation of their curves from the diagonal line of random prediction validates their practical applicability for risk stratification in atrial fibrillation–related mortality.

Figure 7 illustrates the ROC curves of the six machine learning models, offering a comparative view of their discriminative performance in predicting mortality among patients with atrial fibrillation.

4.5. Training Time and Computational Complexity

Training time is a key consideration when selecting models for deployment, particularly in real-time or resource-constrained environments. Table 5 summarizes

the training durations for each evaluated method, highlighting the trade-offs between computational efficiency and algorithmic complexity.
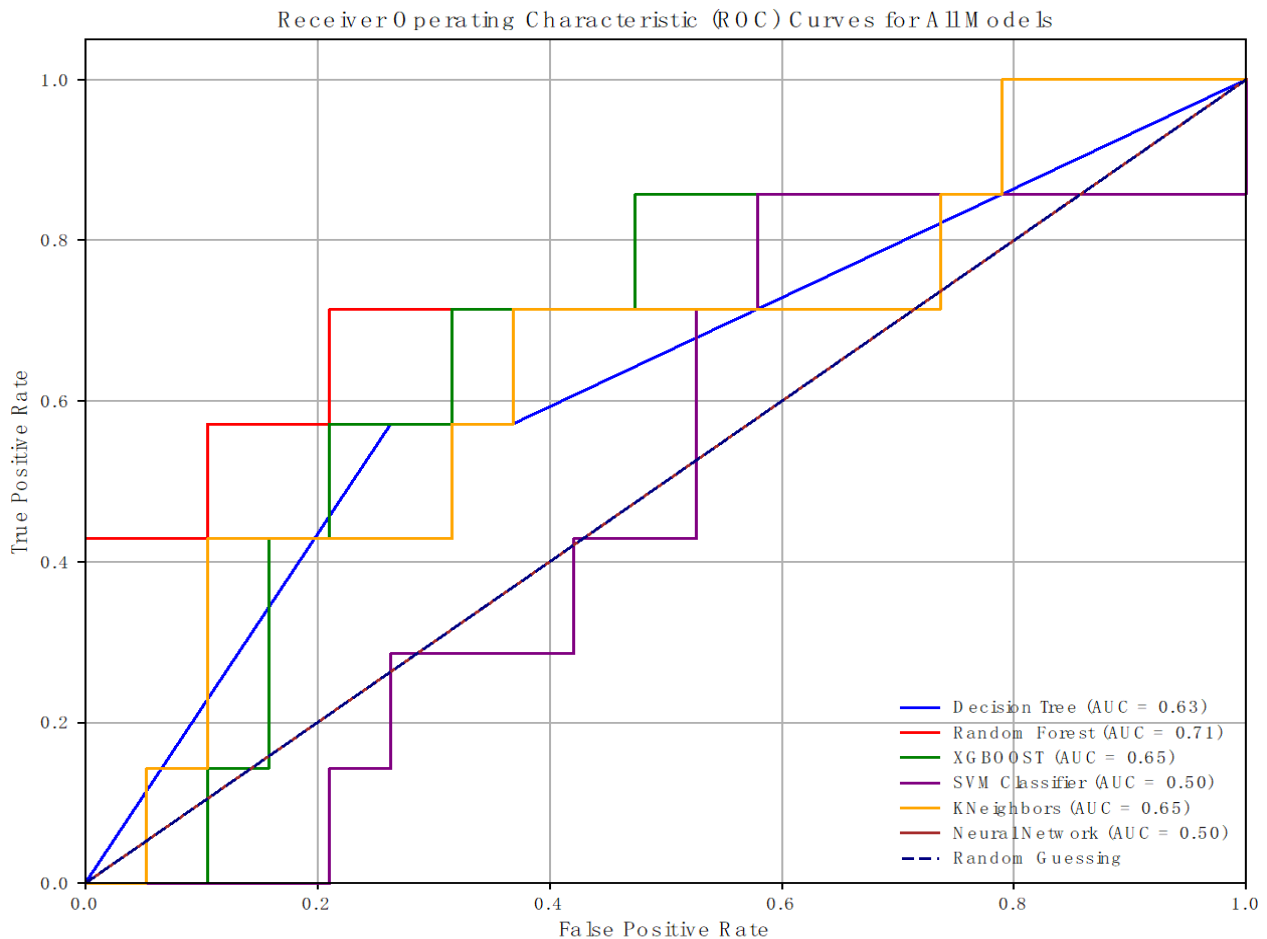

Figure 7. ROC curves of the machine learning models for mortality prediction.

Table 5. Training time comparison of machine learning models.

Model	Decision Tree	Random Forest	XGBoost	SVM Classifier	KNN	Neural Network
Training Time	0.27 s	9.76 s	10.29 s	1.69 s	0.46 s	5.94 s

The training times varied significantly across the evaluated models. Simpler algorithms, such as the Decision Tree and K-Nearest Neighbors, completed training rapidly in 0.27 and 0.46 seconds, respectively, reflecting their straightforward design and low computational demands.

In contrast, ensemble methods like XGBoost (10.29 s) and Random Forest (9.76 s) required more time, owing to their iterative and tree-based structures. The Neural Network also showed a longer training time (5.94 s), consistent with the resource-intensive nature of deep learning. The SVM classifier occupied a middle ground at 1.69 seconds.

These differences highlight the balance between model complexity and computational efficiency. While simpler models may suit real-time or constrained environments, more advanced models offer greater predictive potential at the cost of longer training durations.

4.6. SHAP-Based Interpretability Analysis of the Mortality Prediction Model

The SHAP summary plot provides a nuanced view of how each variable influences the prediction of mortality among patients with atrial fibrillation. Features are ranked by their overall impact on model output, while the color gradient indicates the actual feature value, from low (blue) to high (red).

Figure 8 presents the SHAP summary plot, which illustrates the contribution of each feature to the mortality prediction model and highlights how individual variables influence the classification outcomes.

Figure 8. SHAP Summary Plot Showing Feature Contributions to the Mortality Prediction Model.

Among the most influential predictors, reported heart failure and EHRA score stand out prominently. For these variables, high values (in red) are associated with a strong positive SHAP impact, pushing the model toward a prediction of death. This pattern underscores the clinical intuition that severe symptoms or underlying cardiac dysfunction significantly elevate mortality risk.

Heart rate also shows a similar trend, where higher rates are linked to a positive contribution to death prediction. In contrast, valvular disease appears to behave in a more complex fashion: both low and high values span a wider SHAP range, suggesting its impact might vary based on interactions with other features.

Left atrium mean diameter, systolic and diastolic blood pressure, and patient weight exhibit more balanced effects, although higher atrial dimensions and systolic pressures tend to shift predictions toward the positive class. Interestingly, patient anxiety level, CHA2DS2-VASc score, and consultation type also contribute meaningfully, indicating that both clinical and behavioral factors are taken into account by the model.

On the lower end of the ranking, variables such as exertional dyspnea and history of hypertension demonstrate more modest influence, with SHAP values clustering near zero, suggesting limited standalone predictive power.

Overall, this analysis confirms the multifactorial nature of mortality risk in atrial fibrillation, with a few dominant clinical features, such as heart failure status and symptom severity (EHRA), playing pivotal roles in model predictions.

4.7. Review of Model Findings and Recommendations

The results of this study confirm the potential of supervised learning for predicting mortality in patients with atrial fibrillation. Among the evaluated models, XGBoost and Random Forest demonstrated the most compelling performance across key metrics such as accuracy, AUC-ROC, and MCC. XGBoost stood out for its balanced trade-off between sensitivity and precision, along with strong calibration and overall robustness. Random Forest, while achieving high precision and the best AUC, showed slightly lower recall, making it more suitable for conservative risk stratification strategies.

SHAP-based interpretation further validated these models by highlighting the clinical relevance of predictors such as heart failure, EHRA score, and heart rate. These variables align with established medical indicators, strengthening the interpretability and trustworthiness of the predictions.

In summary, XGBoost appears best suited for clinical deployment, with Random Forest offering a precise alternative when minimizing false positives is critical. Both models, supported by SHAP explainability, provide effective and interpretable tools for mortality risk prediction in atrial fibrillation.

This section synthesizes the main findings, underscores the contributions of the proposed methods, and lays the groundwork for future work in AI-assisted clinical risk assessment.

5. Discussion

This study confirms the relevance of supervised machine learning in predicting mortality among patients with atrial fibrillation, using clinical data collected from multiple healthcare centers in Côte d'Ivoire. The comparative analysis of six algorithms revealed clear differences in terms of accuracy, interpretability, and clinical

applicability. This section discusses the key findings, the strengths of the evaluated models, the reasonably identified limitations, and potential directions for future research.

5.1. Model Performance and Interpretability

Among the evaluated approaches, XGBoost and Random Forest emerged as the most effective models, particularly with respect to AUC-ROC, MCC, and the balance between sensitivity and specificity. XGBoost demonstrated highly consistent results, positioning it as a strong candidate for clinical risk assessment. These findings are consistent with previous research in cardiology, where gradient boosting often outperforms simpler algorithms [51] [61].

Interpretability was enhanced through the use of SHAP values, which highlighted major predictors such as the EHRA score, the presence of heart failure, and certain echocardiographic parameters, especially left atrial size. These results are in line with well-established cardiological risk factors. Such transparency is essential for fostering clinician trust and supporting the adoption of decision-support tools in high-stakes medical settings [8].

5.2. Added Value of Machine Learning Compared to Traditional Tools

Traditional clinical scores such as CHA_2DS_2-VASc and HAS-BLED, though widely used, are based on additive point systems involving a limited number of variables. In contrast, machine learning models are capable of capturing complex, non-linear interactions across clinical, biological, behavioral, and structural data [51]. In our study, rarely considered variables, such as anxiety level or type of consultation, emerged as secondary yet meaningful predictors.

Furthermore, the use of SHAP for analyzing feature contributions provided insights not only into the predictors of mortality but also into how these variables influence outcomes. This level of explanation is essential to encourage ethical and informed deployment of AI tools in clinical settings [8].

5.3. Study Limitations

Some limitations must be acknowledged, though they do not undermine the validity of the conclusions. The sample size, while multicenter in nature, remains relatively modest (n = 239), which may limit statistical power and the detection of subtle associations in specific subgroups.

The imbalance between the deceased and non-deceased classes is a classic challenge in this type of study [61]. Although this issue was partially addressed through stratified cross-validation and the use of MCC, it must be considered when interpreting the results.

In addition, although most models produced reasonably well-calibrated probability estimates, the decision tree and the neural network showed unusually high log-loss values even when accuracy and recall were acceptable. This behavior re-

flects limitations inherent to these algorithms: decision trees tend to generate extreme, unsmoothed probabilities, while neural networks may yield unstable outputs due to overfitting. These results do not compromise the overall predictive performance but rather point to the value of exploring probability calibration techniques in future research.

Finally, while the models demonstrated promising performance, their robustness should be further evaluated through testing on larger datasets and operational scenarios. Such efforts would help assess their adaptability to clinical variability and their long-term reliability.

It is important to note, however, that these limitations do not challenge the observed trends but rather support the need for further work to strengthen and extend the current findings.

5.4. Future Research Directions

Future research should focus on expanding the dataset and integrating the models into real clinical environments. A more refined use of the longitudinal data already available could lead to dynamic predictions better suited to patient follow-up.

Additionally, embedding these tools into hospital information systems would enhance their practical value, especially for the early identification of at-risk patients. Continued efforts to develop explainable AI methods (e.g., SHAP, counterfactuals) will also be essential to ensure transparency, regulatory compliance, and clinician acceptance [8].

6. Conclusions

This study demonstrates the potential of supervised machine learning models in predicting mortality among patients with atrial fibrillation, using real-world clinical data from a sub-Saharan African cohort. Among the six classifiers evaluated, XGBoost and Random Forest emerged as the most promising candidates, offering a valuable trade-off between predictive accuracy, robustness, and clinical interpretability. XGBoost achieved the highest overall performance across several key metrics, while Random Forest provided stable precision and a strong AUC, confirming its reliability. Although the Decision Tree showed lower performance, it stood out for its simplicity, rapid training time, and ease of interpretation, features particularly relevant in low-resource settings.

The integration of SHAP-based explainability significantly enhanced the transparency and clinical relevance of the models, identifying EHRA score, heart failure status, blood pressure, and left atrial diameter as major contributors to mortality risk. These findings resonate with existing medical knowledge, reinforcing the credibility of the models and offering opportunities to strengthen clinical decision-making processes.

Despite these encouraging outcomes, certain limitations must be acknowledged. The relatively modest sample size, single-center origin of the dataset, and absence

of external validation constrain the generalizability of the results. Additionally, model performance on certain metrics, such as recall or log-loss, revealed areas that warrant further optimization.

Future research should prioritize external validation across diverse populations, explore dynamic modeling strategies such as survival analysis, and investigate integration with real-time clinical workflows. It will also be crucial to assess the ethical and regulatory implications of using AI in risk prediction, particularly in vulnerable or underrepresented populations.

In summary, this work illustrates the feasibility and utility of applying interpretable AI methods to improve risk stratification in atrial fibrillation. With appropriate safeguards and validation, these tools may contribute meaningfully to more personalized, proactive, and equitable patient care.

Conflicts of Interest

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

References

[1] Chugh, S.S., Havmoeller, R., Narayanan, K., Singh, D., Rienstra, M., Benjamin, E.J., *et al.* (2014) Worldwide Epidemiology of Atrial Fibrillation: A Global Burden of Disease. *Circulation*, **129**, 837-847.
https://doi.org/10.1161/circulationaha.113.005119

[2] Elliott, A.D., Middeldorp, M.E., Van Gelder, I.C., Albert, C.M. and Sanders, P. (2023) Epidemiology and Modifiable Risk Factors for Atrial Fibrillation. *Nature Reviews Cardiology*, **20**, 404-417. https://doi.org/10.1038/s41569-022-00820-8

[3] Mensah, G.A., Roth, G.A. and Fuster, V. (2019) The Global Burden of Cardiovascular Diseases and Risk Factors. *Journal of the American College of Cardiology*, **74**, 2529-2532. https://doi.org/10.1016/j.jacc.2019.10.009

[4] Balla, S., Nkum, B.C. and Burch, V. (2019) The Current State of Cardiovascular Disease in Sub-Saharan Africa. *Cardiovascular Journal of Africa*, **30**, 314-321.

[5] Lip, G.Y.H., Nieuwlaat, R., Pisters, R., Lane, D.A. and Crijns, H.J.G.M. (2010) Refining Clinical Risk Stratification for Predicting Stroke and Thromboembolism in Atrial Fibrillation Using a Novel Risk Factor-Based Approach: The Euro Heart Survey on Atrial Fibrillation. *Chest*, **137**, 263-272.
https://doi.org/10.1378/chest.09-1584

[6] Pisters, R., Lane, D.A., Nieuwlaat, R., de Vos, C.B., Crijns, H.J.G.M. and Lip, G.Y.H. (2010) A Novel User-Friendly Score (HAS-BLED) to Assess 1-Year Risk of Major Bleeding in Patients with Atrial Fibrillation. *Chest*, **138**, 1093-1100.
https://doi.org/10.1378/chest.10-0134

[7] Zheng, X., Wang, F., Zhang, J., Cui, X., Jiang, F., Chen, N., *et al.* (2022) Using Machine Learning to Predict Atrial Fibrillation Diagnosed after Ischemic Stroke. *International Journal of Cardiology*, **347**, 21-27.
https://doi.org/10.1016/j.ijcard.2021.11.005

[8] Ribeiro, M.T., Singh, S. and Guestrin, C. (2016) "Why Should I Trust You?". *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 1135-1144.
https://doi.org/10.1145/2939672.2939778

[9] Lundberg, S.M. and Lee, S.I. (2017) A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems* (*NeurIPS*), Long Beach, 4-9 December 2017, 4766-4777.

[10] Luo, Y., Szolovits, P., Dighe, A.S. and Baron, J.M. (2016) Using Machine Learning to Predict Laboratory Test Results. *American Journal of Clinical Pathology*, **145**, 778-788. https://doi.org/10.1093/ajcp/aqw064

[11] Hung, M., Hon, E.S., Lauren, E., Xu, J., Judd, G. and Su, W. (2020) Machine Learning Approach to Predict Risk of 90-Day Hospital Readmissions in Patients with Atrial Fibrillation: Implications for Quality Improvement in Healthcare. *Health Services Research and Managerial Epidemiology*, **7**, Article 2333392820961887. https://doi.org/10.1177/2333392820961887

[12] Carvalho, D.V., Pereira, E.M. and Cardoso, J.S. (2019) Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics*, **8**, Article 832. https://doi.org/10.3390/electronics8080832

[13] Rajkomar, A., Hardt, M., Howell, M.D., Corrado, G. and Chin, M.H. (2018) Ensuring Fairness in Machine Learning to Advance Health Equity. *Annals of Internal Medicine*, **169**, 866-872. https://doi.org/10.7326/m18-1990

[14] Mwita, J.C., Ocampo, C., Molefe-Baikai, O.J., Goepamang, M., Botsile, E. and Tshikuka, J.G. (2019) Characteristics and 12-Month Outcome of Patients with Atrial Fibrillation at a Tertiary Hospital in Botswana. *Cardiovascular Journal of Africa*, **30**, 168-173. https://doi.org/10.5830/cvja-2019-013

[15] January, C.T., Wann, L.S., Calkins, H., Chen, L.Y., Cigarroa, J.E., Cleveland, J.C., *et al.* (2019) 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients with Atrial Fibrillation. *Journal of the American College of Cardiology*, **74**, 104-132. https://doi.org/10.1016/j.jacc.2019.01.011

[16] Benjamin, E.J., Wolf, P.A., D'Agostino, R.B., Silbershatz, H., Kannel, W.B. and Levy, D. (1998) Impact of Atrial Fibrillation on the Risk of Death. *Circulation*, **98**, 946-952. https://doi.org/10.1161/01.cir.98.10.946

[17] Chugh, S.S., Roth, G.A., Gillum, R.F. and Mensah, G.A. (2014) Global Burden of Atrial Fibrillation in Developed and Developing Nations. *Global Heart*, **9**, 113-119. https://doi.org/10.1016/j.gheart.2014.01.004

[18] Hindricks, G., Potpara, T., Dagres, N., *et al.* (2021) 2020 ESC Guidelines for the Diagnosis and Management of Atrial Fibrillation. *European Heart Journal*, **42**, 373-498.

[19] Wang, T.J., Larson, M.G., Levy, D., Vasan, R.S., Leip, E.P., Wolf, P.A., *et al.* (2003) Temporal Relations of Atrial Fibrillation and Congestive Heart Failure and Their Joint Influence on Mortality. *Circulation*, **107**, 2920-2925. https://doi.org/10.1161/01.cir.0000072767.89944.6e

[20] Miyasaka, Y., Barnes, M.E., Gersh, B.J., Cha, S.S., Bailey, K.R., Abhayaratna, W.P., *et al.* (2006) Secular Trends in Incidence of Atrial Fibrillation in Olmsted County, Minnesota, 1980 to 2000, and Implications on the Projections for Future Prevalence. *Circulation*, **114**, 119-125. https://doi.org/10.1161/circulationaha.105.595140

[21] Friberg, L., Rosenqvist, M. and Lip, G.Y.H. (2012) Net Clinical Benefit of Warfarin in Patients with Atrial Fibrillation: A Report from the Swedish Atrial Fibrillation Cohort Study. *Circulation*, **125**, 2298-2307. https://doi.org/10.1161/circulationaha.111.055079

[22] Kornej, J., Börschel, C.S., Benjamin, E.J. and Schnabel, R.B. (2020) Epidemiology of Atrial Fibrillation in the 21st Century. *Circulation Research*, **127**, 4-20. https://doi.org/10.1161/circresaha.120.316340

[23] Gregory, Y.H., John Camm, A., Hohnloser, S.H., *et al.* (2012) 2012 Focused Update of the ESC Guidelines for the Management of Atrial Fibrillation. *European Heart Journal*, **33**, 2719-2747.

[24] Zakeri, R., Chamberlain, A.M., Roger, V.L. and Redfield, M.M. (2013) Temporal Relationship and Prognostic Significance of Atrial Fibrillation in Heart Failure Patients with Preserved Ejection Fraction: A Community-Based Study. *Circulation*, **128**, 1085-1093. https://doi.org/10.1161/circulationaha.113.001475

[25] Anter, E., Jessup, M. and Callans, D.J. (2009) Atrial Fibrillation and Heart Failure: Treatment Considerations for a Dual Epidemic. *Circulation*, **119**, 2516-2525. https://doi.org/10.1161/circulationaha.108.821306

[26] Opie, L.H. and Mayosi, B.M. (2005) Cardiovascular Disease in Sub-Saharan Africa. *Circulation*, **112**, 3536-3540. https://doi.org/10.1161/circulationaha.105.597765

[27] Diop, K.R., Samb, C.A.B., Kane, A., *et al.* (2022) Atrial Fibrillation in Three Cardiological Reference Centers in Dakar: Senegal Data from the AFRICA Register Survey. *The Pan African Medical Journal*, **43**, 112.

[28] Coulibaly, I., Anzouan-Kacou, J.B., Konin, K.C., Kouadio, S.C. and Abouo-N'Dori, R. (2010) Atrial Fibrillation: Epidemiological Data from the Cardiology Institute in Abidjan, Côte d'Ivoire. *Medecine tropicale: Revue du Corps de sante colonial*, **70**, 371-374.

[29] Noubiap, J.J. and Nyaga, U.F. (2019) A Review of the Epidemiology of Atrial Fibrillation in Sub-Saharan Africa. *Journal of Cardiovascular Electrophysiology*, **30**, 3006-3016. https://doi.org/10.1111/jce.14222

[30] Topol, E.J. (2019) High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine*, **25**, 44-56. https://doi.org/10.1038/s41591-018-0300-7

[31] Nabyonga-Orem, J., Karamagi, H., Bazeyo, W., *et al.* (2016) Patterns of International Collaboration in Cardiovascular Research in Sub-Saharan Africa. *Cardiovascular Diagnosis and Therapy*, **6**, 436-445.

[32] Ettarh, R. (2016) Patterns of International Collaboration in Cardiovascular Research in Sub-Saharan Africa. *Cardiovascular Journal of Africa*, **27**, 194-200. https://doi.org/10.5830/cvja-2015-082

[33] Kao, Y.T., Huang, C.Y., Fang, Y.A., Liu, J.C. and Chang, T.H. (2023) Machine Learning-Based Prediction of Atrial Fibrillation Risk Using Electronic Medical Records in Older Aged Patients. *The American Journal of Cardiology*, **198**, 56-63. https://doi.org/10.1016/j.amjcard.2023.03.035

[34] Meltzer, S.N. and Weintraub, W.S. (2020) The Role of National Registries in Improving Quality of Care and Outcomes for Cardiovascular Disease. *Methodist DeBakey Cardiovascular Journal*, **16**, 205-211. https://doi.org/10.14797/mdcj-16-3-205

[35] World Health Organization (2011) Global Atlas on Cardiovascular Disease Prevention and Control. https://iris.who.int/handle/10665/44701

[36] Alonso, A. and Bengtson, L.G.S. (2014) A Rising Tide: The Global Epidemic of Atrial Fibrillation. *Circulation*, **129**, 829-830. https://doi.org/10.1161/circulationaha.113.007482

[37] Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M. and Qureshi, N. (2017) Can Machine-Learning Improve Cardiovascular Risk Prediction Using Routine Clinical Data? *PLOS ONE*, **12**, e0174944. https://doi.org/10.1371/journal.pone.0174944

[38] Obermeyer, Z. and Emanuel, E.J. (2016) Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. *New England Journal of Medicine*, **375**, 1216-1219.

https://doi.org/10.1056/nejmp1606181

[39] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32.
https://doi.org/10.1023/a:1010933404324

[40] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the* 22*nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794.

[41] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. https://doi.org/10.1023/a:1022627411411

[42] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, **521**, 436-444. https://doi.org/10.1038/nature14539

[43] Krittanawong, C., Johnson, K.W., Rosenson, R.S., Wang, Z., Aydar, M., Baber, U., *et al.* (2019) Deep Learning for Cardiovascular Medicine: A Practical Primer. *European Heart Journal*, **40**, 2058-2073. https://doi.org/10.1093/eurheartj/ehz056

[44] Zhang, Z., Ho, K.M. and Hong, Y. (2021) Machine Learning for Clinical Risk Prediction in Patients with Atrial Fibrillation: A Systematic Review. *Heart, Lung and Circulation*, **30**, 1395-1404.

[45] Attia, Z.I., Noseworthy, P.A., Lopez-Jimenez, F., Asirvatham, S.J., Deshmukh, A.J., Gersh, B.J., *et al.* (2019) An Artificial Intelligence-Enabled ECG Algorithm for the Identification of Patients with Atrial Fibrillation during Sinus Rhythm: A Retrospective Analysis of Outcome Prediction. *The Lancet*, **394**, 861-867. https://doi.org/10.1016/s0140-6736(19)31721-0

[46] Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., *et al.* (2019) Cardiologist-level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms Using a Deep Neural Network. *Nature Medicine*, **25**, 65-69. https://doi.org/10.1038/s41591-018-0268-3

[47] Raghunath, S., Pfeifer, J.M., Ulloa-Cerna, A.E., *et al.* (2021) Deep Neural Networks Can Predict New-Onset Atrial Fibrillation from the 12-Lead ECG. *Circulation*, **143**, 1287-1298.

[48] Johnson, A.E.W., Pollard, T.J., Shen, L., Lehman, L.H., Feng, M., Ghassemi, M., *et al.* (2016) MIMIC-III, a Freely Accessible Critical Care Database. *Scientific Data*, **3**, Article No. 160035. https://doi.org/10.1038/sdata.2016.35

[49] Harutyunyan, H., Khachatrian, H., Kale, D.C., Ver Steeg, G. and Galstyan, A. (2019) Multitask Learning and Benchmarking with Clinical Time Series Data. *Scientific Data*, **6**, Article No. 96. https://doi.org/10.1038/s41597-019-0103-9

[50] Kramoh, E., Coulibaly, Z., Soya, B., *et al.* (2023) Profil épidémiologique et mortalité des patients atteints de fibrillation atriale en Côte d'Ivoire: Résultats du registre AFRICA. *Les Archives des Maladies du Cœur et des Vaisseaux*, **116**, 145-152.

[51] Rajkomar, A., Dean, J. and Kohane, I. (2019) Machine Learning in Medicine. *New England Journal of Medicine*, **380**, 1347-1358. https://doi.org/10.1056/nejmra1814259

[52] Luo, Y., Dong, R., Liu, J. and Wu, B. (2024) A Machine Learning-Based Predictive Model for the In-Hospital Mortality of Critically Ill Patients with Atrial Fibrillation. *International Journal of Medical Informatics*, **191**, Article 105585. https://doi.org/10.1016/j.ijmedinf.2024.105585

[53] Breiman, L. (1996) Bagging Predictors. *Machine Learning*, **24**, 123-140. https://doi.org/10.1023/a:1018054314350

[54] Bishop, C.M. (1995) Neural Networks for Pattern Recognition. Oxford University Press.

[55] Haykin, S. (1999) Neural Networks: A Comprehensive Foundation. 2nd Edition, Prentice Hall.

[56] Breiman, L., Friedman, J., Olshen, R.A. and Stone, C.J. (1984) Classification and Regression Trees. 1st Edition, Chapman and Hall/CRC.

[57] Quinlan, J.R. (1993) C4.5: Programs for Machine Learning. Morgan Kaufmann. https://cir.nii.ac.jp/crid/1572261549005092352

[58] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. https://doi.org/10.1023/a:1022627411411

[59] Natekin, A. and Knoll, A. (2013) Gradient Boosting Machines, a Tutorial. *Frontiers in Neurorobotics*, **7**, Article 21. https://doi.org/10.3389/fnbot.2013.00021

[60] Kohavi, R. (1995) A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the* 14*th International Joint Conference on Artificial Intelligence*, Montreal, 20-25 August 1995, 1137-1143.

[61] Guo, H.X., Li, Y.J., Shang, J., *et al.* (2017) Learning from Class-Imbalanced Data: Review of Methods and Applications. *Expert Systems with Applications*, **73**, 220-239. https://doi.org/10.1016/j.eswa.2016.12.035