# Machine Learning Predicting Atrial Fibrillation as an Adverse Event in the Warfarin and Aspirin in Reduced Cardiac Ejection Fraction (WARCEF) Trial

Gue, Ying; Correa, Elon; Thompson, John L. P.; Homma, Shunichi; Qian, Min; Lip, Gregory Y. H.

*Published in:*
The American Journal of Medicine

*DOI (link to publication from Publisher):*
10.1016/j.amjmed.2023.07.019



*Publication date:*
2023

*Document Version*
Publisher's PDF, also known as Version of record

CLINICAL RESEARCH STUDY

THE AMERICAN JOURNAL of MEDICINE®

# Machine Learning Predicting Atrial Fibrillation as an Adverse Event in the Warfarin and Aspirin in Reduced Cardiac Ejection Fraction (WARCEF) Trial

**Ying Gue, MB,[a,b] Elon Correa, PhD,[a,c] John L.P. Thompson, PhD,[d] Shunichi Homma, MD,[d] Min Qian, PhD,[d] Gregory Y.H. Lip, MD[a,b,e]**

*[a]Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, United Kingdom; [b]The Department of Cardiovascular and Metabolic Medicine, University of Liverpool, United Kingdom; [c]School of Computer Science and Mathematics, Liverpool John Moores University, United Kingdom; [d]Columbia University Medical Center, New York, NY; [e]Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark.*

## ABSTRACT

**BACKGROUND:** Atrial fibrillation and heart failure commonly coexist due to shared pathophysiological mechanisms. Prompt identification of patients with heart failure at risk of developing atrial fibrillation would allow clinicians the opportunity to implement appropriate monitoring strategy and timely treatment, reducing the impact of atrial fibrillation on patients' health.

**METHODS:** Four machine learning models combined with logistic regression and cluster analysis were applied post hoc to patient-level data from the Warfarin and Aspirin in Patients with Heart Failure and Sinus Rhythm (WARCEF) trial to identify factors that predict development of atrial fibrillation in patients with heart failure.

**RESULTS:** Logistic regression showed that White divorced patients have a 1.75-fold higher risk of atrial fibrillation than White patients reporting other marital statuses. By contrast, similar analysis suggests that non-White patients who live alone have a 2.58-fold higher risk than those not living alone. Machine learning analysis also identified "marital status" and "live alone" as relevant predictors of atrial fibrillation. Apart from previously well-recognized factors, the machine learning algorithms and cluster analysis identified 2 distinct clusters, namely White and non-White ethnicities. This should serve as a reminder of the impact of social factors on health.

**CONCLUSION:** The use of machine learning can prove useful in identifying novel cardiac risk factors. Our analysis has shown that "social factors," such as living alone, may disproportionately increase the risk of atrial fibrillation in the under-represented non-White patient group with heart failure, highlighting the need for more studies focusing on stratification of multiracial cohorts to better uncover the heterogeneity of atrial fibrillation.

 *• The American Journal of Medicine (2023) 136:1099−1108*

**KEYWORDS:** Atrial fibrillation; Heart failure; Machine learning; Racial disparity; Social factors

**Funding:** The WARCEF trial was supported by the National Institute of Neurological Diseases and Stroke of the USA National Institutes of Health (grant numbers U01-NS-043975 to SH and U01-NS-039143 to JLPT).

**Conflicts of Interest:** None.

**Authorship:** All authors had access to the data and a role in the writing of the manuscript.

Requests for reprints should be addressed to Gregory Y.H. Lip, MD, Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool Heart & Chest Hospital, William Henry Duncan Building, 6 West Derby Street, Liverpool L7 8TX, United Kingdom.

E-mail address: gregory.lip@liverpool.ac.uk


https://doi.org/10.1016/j.amjmed.2023.07.019

## INTRODUCTION

Atrial fibrillation and heart failure frequently coexist due to shared pathophysiological mechanisms and risk factors,[1,2] and the combination increases the risk of hospitalization and mortality.[2,3] Both conditions are associated with a hypercoagulable state, putting patients at an increased risk of thromboembolic complications.[4,5] However, in the presence of heart failure, without associated atrial fibrillation, the routine use of oral anticoagulation (OAC) to prevent thrombotic complications is not recommended.[6] This is because although the use of warfarin significantly reduces stroke, it is at the expense of increasing major hemorrhage without any mortality benefit.[7] Therefore, the identification of atrial fibrillation in patients with heart failure not only has implications toward prognostication but also alters the recommendation toward the use of OAC for prevention of thrombotic complications.

### CLINICAL SIGNIFICANCE

- Atrial fibrillation is common in patients with heart failure with reduced ejection fraction due to common risk factors between the 2 conditions.
- Machine learning algorithms have shown that apart from clinical risk factors, social factors such as marital status and living alone can predict development of atrial fibrillation in this patient group.
- Social factors may disproportionately increase the risk of atrial fibrillation in an under-represented non-White patient group with heart failure.
- The use of machine learning can prove useful in identifying novel cardiac risk factors.

Categorizing and integrating demographics and clinical characteristics to aid clinicians in identifying patients with heart failure at risk of developing atrial fibrillation is not an easy task. Shared pathophysiological mechanisms make several of the clinical predictors common to this population of patients, thereby confounding the discriminatory power of those predictors. Interest in the use of machine learning to help in risk prediction within the clinical environment has increased in recent years. Although it is still in its pilot stages with inherent flaws and limitations, machine learning has shown great promise in improving the prediction of diseases, including atrial fibrillation.[8] There is enormous potential in its application to existing clinical databases to improve prediction and identification of high-risk patients, leading on to appropriate monitoring strategies and subsequent alterations in their pharmacotherapy.

The Warfarin and Aspirin in Patients with Heart Failure and Sinus Rhythm (WARCEF)[9] study was a double-blind multicenter clinical trial recruiting patients with left ventricular ejection fraction of 35% or less in sinus rhythm and randomizing them into either warfarin or aspirin. The study concluded that there was no significant difference in outcomes (composite of stroke, intracerebral hemorrhage or death) between treatment with warfarin and treatment with aspirin.[9] These patients were followed up for up to 6 years with recorded adverse events, including the development of atrial fibrillation. In the current analysis the primary outcome is the incidence of new-onset atrial fibrillation as an adverse event during the follow-up period. We aim to use machine learning to identify characteristics which are predictive of atrial fibrillation in patients with heart failure and sinus rhythm.

## METHODS

### Study Design

Patient-level data from the randomized, double-blind WARCEF trial (n = 2305 patients) were used to perform the analysis. Patients with atrial fibrillation at baseline (n = 86) were excluded from the analysis. Baseline characteristics, medical history, and adjudicated adverse events were extracted and compared between patients who did and did not develop atrial fibrillation as an adverse event throughout the follow-up period. Of the 2219 patients included for analysis, 215 presented atrial fibrillation as an adjudicated adverse event during the follow-up period. Logistic regression analysis and other 4 machine learning (ML) methods were employed to identify characteristics predictive of atrial fibrillation. These methods are described in further details in the supplementary materials (Appendix,[10-19] available online).

## RESULTS

### Descriptive Statistics

A total of 2219 patients were included in the analysis. Of these, 215 patients developed atrial fibrillation during the follow-up period. Patients who developed atrial fibrillation are significantly older, taller (sex-adjusted $P$ value $< .001$), and more likely to be males and White. They are also more likely to be divorced or widowed and live alone. Baseline characteristics are as shown in Table 1.

### Logistic Regression Analysis

Logistic regression was performed including all reported characteristics, and obtained a c-index for "new-onset atrial fibrillation as an adverse event during follow-up" of 0.74 (95% confidence interval [CI], 0.71-0.78). The results are shown below in Table 2. Age, White (ethnicity), marital status (divorced or widowed), and hospitalization with heart failure were all identified as significant predictors of atrial fibrillation.

### Machine Learning Analysis Results

Table 3 shows the averaged cross-validation results for predicting atrial fibrillation as an adverse event (averaged over 100 cross-validations for each fold times 5 folds).

**Table 1** Statistical Comparisons (*P* Values) Based on Wilcoxon Rank Sum Tests for Continuous Variables and Chi-Squared Tests for Others

| | Patients Who Developed Atrial Fibrillation (n = 215) | Patients Who Did Not Develop Atrial Fibrillation (n = 2004) | *P* Value |
|---|---|---|---|
| Age (years) | 59.8 ± 11.1 | 57.3 ± 11.4 | .00637 |
| Sex, n (%) | | | |
| Males | 185 (86%) | 1591 (79%) | .02034 |
| Females | 30 (14%) | 413 (21%) | |
| Race, n (%) | | | |
| White | 188 (90%) | 1596 (84%) | .01796 |
| Non-White | 20 (10%) | 300 (16%) | |
| Height (cm) | 174.1 ± 8.3 | 171.4 ± 9.3 | .00017 |
| Weight (kg) | 88.6 ± 19.6 | 85.7 ± 19.2 | .01942 |
| Body Mass Index (kg.m2) | 29.1 ± 5.5 | 29.1 ± 5.9 | .70570 |
| Education, n (%) | | | |
| ≤ 8th grade | 29 (13%) | 376 (19%) | .25730 |
| Some high school | 51 (24%) | 508 (25%) | |
| High school grad | 66 (31%) | 525 (26%) | |
| Some college | 30 (14%) | 292 (15%) | |
| College grad | 30 (14%) | 215 (11%) | |
| Post-grad education | 9 (4%) | 88 (4%) | |
| Marital status, n (%) | | | |
| Single | 17 (8%) | 257 (13%) | .02371 |
| Married | 131 (61%) | 1284 (64%) | |
| Divorced | 37 (17%) | 248 (12%) | |
| Widowed | 30 (14%) | 215 (11%) | |
| Smoking, n (%) | | | |
| Never | 60 (28%) | 622 (31%) | .58860 |
| Ex | 113 (53%) | 1028 (51%) | |
| Current | 42 (19%) | 354 (18%) | |
| Alcohol, n (%) | | | |
| Never | 111 (52%) | 1062 (53%) | .81760 |
| Ex | 46 (21%) | 441 (22%) | |
| Current | 58 (27%) | 501 (25%) | |
| Live alone, n (%) | | | |
| No | 159 (74%) | 1566 (78%) | .16050 |
| Yes | 56 (26%) | 438 (22%) | |
| Hypertension, n (%) | | | |
| No | 210 (98%) | 1950 (97%) | .74920 |
| Yes | 5 (2%) | 54 (3%) | |
| Diabetes, n (%) | | | |
| No | 148 (69%) | 1378 (69%) | .98210 |
| Yes | 67 (31%) | 626 (31%) | |

The classification results show that, overall, AdaBoost was the model with the best predictive performance, followed by Random Forest (RF). However, the limited classification results suggest that, due to the heterogeneity of atrial fibrillation, identifying patients at risk of developing atrial fibrillation is not an easy task. The next section discusses variables and clinical predictors most likely to contribute to the identification of patients at a higher risk of developing new-onset atrial fibrillation as an adverse event.

## Variable Importance for Prediction of New-Onset Atrial Fibrillation

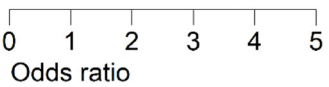
Based on the 3 decision trees classification models (ie, AdaBoost, XGBoost, and RF), variable gain contribution to each model built has been recorded. For each of these ML models, the cumulative gain of each variable has been averaged over all cross-validations. The top 15 variables (with the largest overall gain) from each model have then been selected. For 2 of the 3 models (namely, XGBoost and RF), the same top 15 variables have been identified (100% agreement). For the other model (AdaBoost), 14 variables overlapped with the other 2 models, except for "live alone." AdaBoost did not identify "live alone" as one of its top 15 predictors, instead it replaced it with the variable "ethnicity White" being the only ML mode (except for logistic regression) to identify this variable as one of the top 15 predictors of new-onset atrial fibrillation.

Figure 1 shows all 16 variables (including "White" and "live alone") identified as top 15 best predictors of new-onset atrial fibrillation (based on the WARCEF dataset

**Table 2** Logistic Regression Predicting "New-Onset Atrial Fibrillation as an Adverse Event During Follow-Up"

**Logistic regression predicting new onset AF as an adverse event**

| | Beta | OR (95% CI) | P-value |
|---|---|---|---|
| **Demographics** | | | |
| Age * | 0.017 | 1.02 (1.00 to 1.03) | 0.02467 |
| Weight | 0.038 | 1.04 (0.97 to 1.12) | 0.30214 |
| Height | -0.001 | 1.00 (0.92 to 1.08) | 0.97061 |
| BMI | -0.097 | 0.91 (0.73 to 1.12) | 0.37918 |
| Gender | 0.055 | 1.06 (0.64 to 1.77) | 0.83035 |
| White * | 0.747 | 2.11 (1.37 to 3.37) | 0.00107 |
| Live alone | 0.107 | 1.11 (0.77 to 1.59) | 0.56590 |
| Smoking | 0.068 | 1.07 (0.85 to 1.35) | 0.57182 |
| Alcohol | -0.014 | 0.99 (0.82 to 1.18) | 0.87765 |
| Education | 0.066 | 1.07 (0.96 to 1.19) | 0.22264 |
| Marital status * | 0.222 | 1.25 (1.02 to 1.53) | 0.03041 |
| **Medical history** | | | |
| DM | -0.141 | 0.87 (0.62 to 1.21) | 0.41001 |
| PVD | -0.096 | 0.91 (0.56 to 1.42) | 0.68238 |
| Defibrillator | 0.165 | 1.18 (0.77 to 1.77) | 0.43603 |
| Hypertension MH | -0.103 | 0.90 (0.66 to 1.23) | 0.51548 |
| Pacemaker | 0.112 | 1.12 (0.69 to 1.77) | 0.64093 |
| **Adverse event** | | | |
| NYHA | -0.066 | 0.94 (0.75 to 1.17) | 0.56749 |
| Embolism | -0.419 | 0.66 (0.19 to 1.73) | 0.44462 |
| Hypertension AE | -0.041 | 0.96 (0.32 to 2.32) | 0.93446 |
| Major bleeding | -0.050 | 0.95 (0.67 to 1.33) | 0.77487 |
| Stroke | 0.246 | 1.28 (0.67 to 2.27) | 0.42537 |
| MI | -0.062 | 0.94 (0.47 to 1.73) | 0.85122 |
| Heart failure * | 1.340 | 3.82 (2.82 to 5.18) | 0.00000 |




An asterisk close to the variable's name identifies a statistically significant *P* value (< .05).

AE = adverse event; BMI = body mass index; CI = confidence interval; DM = diabetes mellitus; MH = medical history; MI = myocardial infarction; NYHA = New York Heart Association; OR = odds ratio; PVD = peripheral vascular disease.

**Table 3** Prediction of Atrial Fibrillation as an Adverse Event on the WARCEF Trial

| Model | Accuracy | Sensitivity | Specificity | F1 Score | MCC | AUC | AUC (95% CI) |
|---|---|---|---|---|---|---|---|
| AdaBoost | 67.0 | 55.2 | 68.3 | 0.25 | 0.15 | 0.62 | (0.51-0.73) |
| XGBoost | 57.1 | 58.7 | 56.9 | 0.21 | 0.09 | 0.58 | (0.50-0.66) |
| Random Forest | 61.5 | 61.5 | 61.5 | 0.24 | 0.14 | 0.61 | (0.54-0.69) |
| Neural network | 56.4 | 55.5 | 56.5 | 0.18 | 0.07 | 0.56 | (0.50-0.62) |

Results are averaged over 500 independent cross-validations.

AUC = area under the receiver operating characteristic curve; CI = confidence interval; MCC = Matthews correlation coefficient; WARCEF = Warfarin and Aspirin in Patients with Heart Failure and Sinus Rhythm.

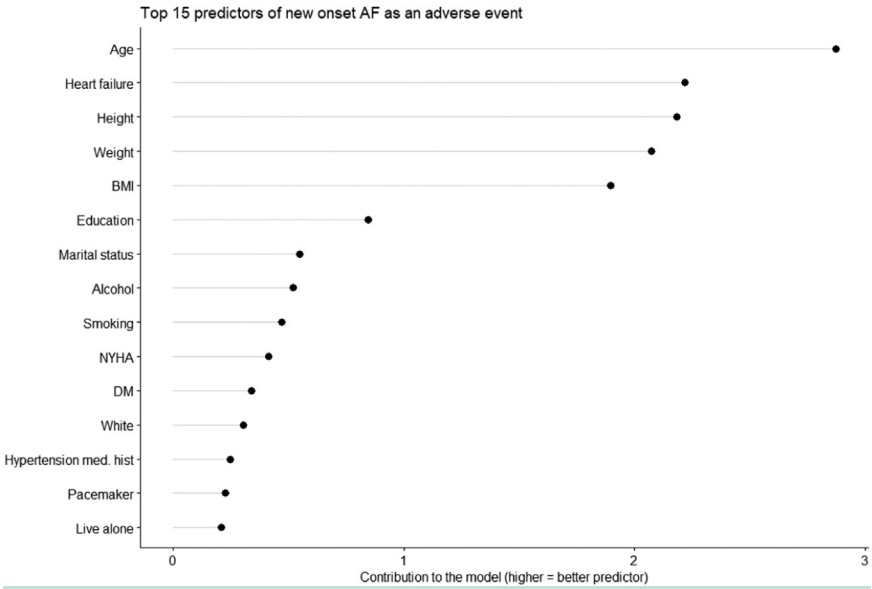



**Figure 1** Top 16 predictors of atrial fibrillation as an adverse event according to the top variables identified by 3 machine learning algorithms: XGBoost, AdaBoost, and Random Forest. AF = atrial fibrillation; BMI = body mass index; DM = diabetes mellitus; NYHA = New York Heart Association.

and AdaBoost, XGBoost, and RF results). The weighted difference between the impurity of 2 child nodes and the uncertainty of the starting node is used to calculate information gain. During the decision tree building process, the value of the information gain decides which variable should be used to split the data. The higher the information gain, the higher the contribution of the variable for splitting the data.

For marital status, comparing "single" vs "divorced" states produces a significant statistical difference, with odds ratio showing an increased risk of divorced patients presenting atrial fibrillation as adverse event of 1.5 (95% CI, 1.1-2.0) and a *P* value of .008. Studies suggest that divorce is a major life stressor that can have economic, emotional, physical, and health consequences.[20,21]

## Decision Curve Analysis

A decision curve analysis[22] for the validity of the models as an early warning system is shown in Figure 2. The models are refitted and evaluated on the whole dataset using 80%/20% split for training and testing sets, respectively, but only the 16 top predictors (see Figure 1) of atrial fibrillation as an adverse event are used to build the models. Overall, all models are valid as predictors of new-onset atrial fibrillation as an adverse event, but only when the threshold probability is between approximately 7.5% and 15%. AdaBoost is the best early warning system, followed by RF, XGBoost, and ANN.

## Clustering the Potential Predictors of Atrial Fibrillation as an Adverse Event

Cluster analysis has been proposed in previous literature to examine the phenotypic heterogeneity of new-onset atrial fibrillation. Figure 3 shows a hierarchical cluster of all variables of interest in the dataset (based on Ward's method[23]).

Figure 4 shows a biplot from a Principal Coordinates Analysis.[24] The biplot suggests that: 1) the variable ethnicity "White" splits the data into 2 clearly distinct clusters; and 2) new-onset atrial fibrillation is associated with being

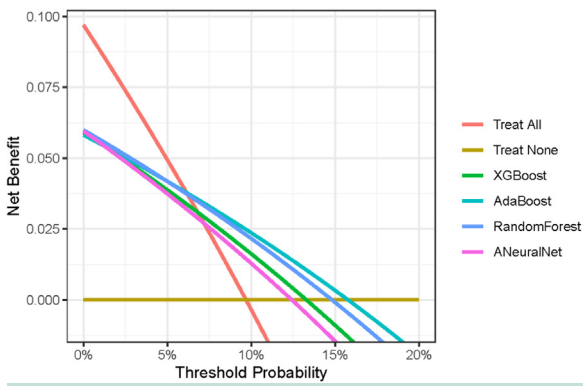



**Figure 2** Decision curve analysis for the 4 classification models evaluated in this study. Each model used the same 16 top predictors (see Figure 3) of atrial fibrillation as an adverse event as predictors.

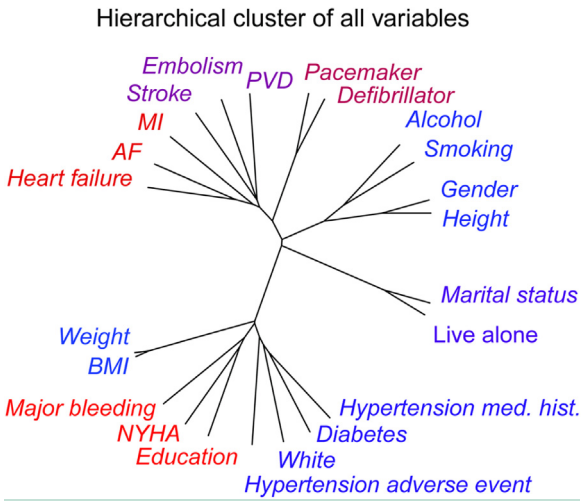



**Figure 3** Hierarchical cluster of all variables of interest in the dataset. AF = atrial fibrillation; BMI = body mass index; MI = myocardial infarction; NYHA = New York Heart Association; PVD = peripheral vascular disease.

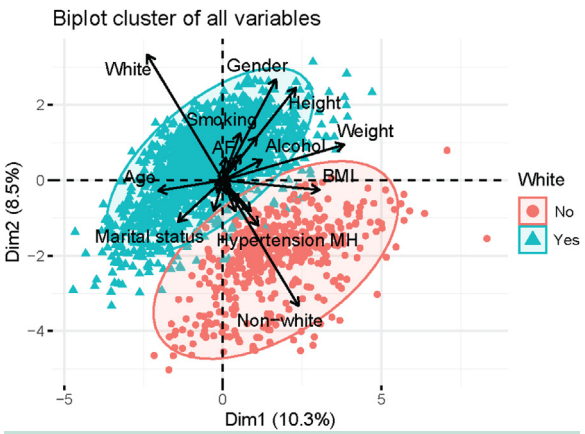



**Figure 4** Biplot cluster of all variables of interest in the dataset. White ethnicity splits the data into 2 clearly distinct clusters.

White (ethnicity), male, older, being a smoker, and marital status (being divorced or widowed).

### Forest Plots on White and Non-White Ethnicity Groups

Based on all characteristics and clinical predictors available, the cluster analysis suggests a split between the White and non-White ethnicities. Therefore, the White and non-White groups have been reanalyzed separately. Via stepwise regression, a subset of 12 common variables were identified to be used in a final regression model for both groups compared (White and non-White ethnicities). Those same 12 variables are compared below.

Table 4 shows logistic regression results obtained from the "White ethnicity (n = 1784)" data (excluding all other ethnicities) with a c-index of 0.73 (95% CI, 0.70-0.77) for new-onset atrial fibrillation as an adverse event during follow. Odds ratios for the 12 predictors of new-onset atrial fibrillation identified by stepwise regression are also shown. The results suggest that, for the "White" group, age, divorced, and heart failure are significant predictors of new-onset atrial fibrillation as adverse event.

Table 5 shows logistic regression results obtained from the "non-White (n = 435)" (other ethnicity) data (excluding White ethnicity) with a c-index of 0.92 (95% CI, 0.87-0.97) for new-onset atrial fibrillation as an adverse event during follow-up. As before, odds ratios for the 12 predictors of new-onset atrial fibrillation identified by stepwise regression are also shown. The results suggest that, for the "non-White" group, widowed, live alone, and heart failure are significant predictors of new-onset atrial fibrillation. Although the number of samples for this group is much smaller than for the "White ethnicity" group, the high c-index value for this model suggests that being widowed, living alone, and suffering from heart failure may significantly increase the risk of new-onset atrial fibrillation for the "non-White" population.

## DISCUSSION

Our analysis has highlighted interesting areas to note in atrial fibrillation prediction in patients with heart failure.

### Clustering of Risk Factors Between Ethnicities

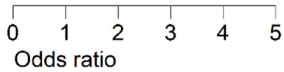
When looking at the clusters of variables (Figure 4), 2 clearly distinct clusters are visible for White and non-White ethnicities. When exploring the association of clinical characteristics with atrial fibrillation, there is a clear disproportionate impact of social circumstances between White and non-White participants, in particular marital status and living alone. The association between marital status and living alone is not the same for the White and non-White groups. In fact, on average, the White group of patients has a higher incidence of living alone than the non-White group. Nevertheless, it is patients from the non-White group living alone who have a significantly higher risk of developing atrial fibrillation when compared with their counterparts from the White group. This suggests that a complex combination of these factors: "marital status," "live alone," and "ethnicity" (and probably other factors not observed in this study) contribute to an elevated risk of developing atrial fibrillation on the non-White cohort. While White patients who are divorced also have a higher risk of developing atrial fibrillation (odds ratio [OR] 1.75; 95% CI, 1.19-2.57; *P* value = .002); non-White patients who are widowed have a 3-fold higher risk of developing atrial fibrillation (OR 3.04; 95% CI, 1.28-7.24; *P* value = .006). Moreover, non-White patients who live alone have a 2.6-fold higher risk of developing atrial fibrillation (OR 2.58; 95% CI, 1.45-4.59; *P* value = .0006).

A previous analysis of the same database[25] has shown similarly that race was a significant predictor within the univariate model but lost statistical significance

**Table 4** Forest Plot Obtained for the "White Ethnicity" Data Only (Excluding Other Ethnicities)

**White ethnicity**

| | AF | No AF | OR (95% CI) | P-value |
|---|---|---|---|---|
| Age * | 60.3+10.7 | 58.2+11.2 | 1.02 (1.01, 1.04) | 0.00142 |
| BMI | 29+5.1 | 28.8+5.4 | 0.98 (0.76, 1.24) | 0.84347 |
| Weight | 87.6+17.8 | 84.7+17.7 | 1.01 (0.94, 1.10) | 0.74032 |
| Height | 173.7+7.8 | 171.3+9.2 | 1.02 (0.94, 1.11) | 0.62855 |
| **Smoking** | | | | |
| Never | 54 | 486 | 0.92 (0.69, 1.22) | 0.71846 |
| Ex | 99 | 827 | 1.03 (0.84, 1.27) | 0.37542 |
| Current | 35 | 283 | 1.06 (0.75, 1.51) | 0.36985 |
| **Alcohol** | | | | |
| Never | 103 | 888 | 0.97 (0.79, 1.18) | 0.62966 |
| Ex | 33 | 288 | 0.97 (0.67, 1.39) | 0.57258 |
| Current | 52 | 420 | 1.07 (0.80, 1.43) | 0.32133 |
| **Education** | | | | |
| <= 8th grade | 24 | 319 | 0.59 (0.39, 0.89) | 0.99424 |
| Some high school | 47 | 426 | 0.92 (0.68, 1.24) | 0.71717 |
| High school grad | 54 | 390 | 1.25 (0.94, 1.66) | 0.06481 |
| Some college | 27 | 210 | 1.11 (0.74, 1.65) | 0.30978 |
| College grad | 27 | 180 | 1.32 (0.88, 1.98) | 0.08972 |
| Post-grad | 9 | 71 | 1.08 (0.54, 2.16) | 0.41396 |
| **Marital status** | | | | |
| Single | 12 | 157 | 0.62 (0.35, 1.12) | 0.94175 |
| Married | 121 | 1097 | 0.82 (0.68, 0.99) | 0.97995 |
| Divorced * | 31 | 162 | 1.75 (1.19, 2.57) | 0.00220 |
| Widowed | 24 | 180 | 1.15 (0.75, 1.76) | 0.25848 |
| **Live alone** | | | | |
| No | 145 | 1266 | 0.88 (0.74, 1.04) | 0.92939 |
| Yes | 43 | 330 | 1.14 (0.83, 1.56) | 0.21313 |
| **NYHA** | | | | |
| 1 | 26 | 218 | 1.01 (0.68, 1.52) | 0.47235 |
| 2 | 100 | 889 | 0.90 (0.73, 1.11) | 0.83142 |
| 3 | 62 | 467 | 1.19 (0.91, 1.55) | 0.09950 |
| 4 | 0 | 22 | | |
| **Diabetes** | | | | |
| No | 132 | 1126 | 0.98 (0.82, 1.18) | 0.57008 |
| Yes | 56 | 470 | 1.02 (0.77, 1.34) | 0.45426 |
| **Heart failure** | | | | |
| No | 93 | 1250 | 0.27 (0.22, 0.33) | 1.00000 |
| Yes * | 95 | 346 | 3.69 (2.94, 4.63) | 0.00000 |




Odds ratio for the 12 predictors of new-onset atrial fibrillation identified by stepwise regression.
AF = atrial fibrillation; BMI = body mass index; CI = confidence interval; NYHA = New York Heart Association; OR = odds ratio.

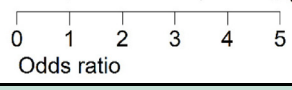
following adjustment for other covariates. This is in support of what we have found in our analysis—that the impact of race and ethnicity on the development of atrial fibrillation involves a complex interaction among various factors.

The impact of race and ethnicity has been studied extensively as a part of the social determinant in incidence of atrial fibrillation.[21] A paradoxical finding is that despite the higher prevalence of risk factors for atrial fibrillation, a lower prevalence of atrial fibrillation in non-White ethnic groups has been observed: its etiology is poorly understood due to the complex nature.[26] Apart from recognizable factors such as household income,[27] which was not recorded within the dataset, we have shown that living alone had a greater impact on ethnic minorities in terms of developing atrial

**Table 5** Forest Plot Obtained for the "Other Ethnicity (Non-White)" Data Only (Excluding White Ethnicity)

**Other ethnicity (non-white)**

| | AF | No AF | OR (95% CI) | P-value |
|---|---|---|---|---|
| Age | 55.9+13 | 54.3+11.6 | 1.02 (0.99, 1.06) | 0.22965 |
| BMI | 30.2+7.4 | 30.4+7.6 | 0.80 (0.49, 1.24) | 0.34758 |
| Weight | 95+28.7 | 89.6+23.7 | 1.08 (0.94, 1.26) | 0.31563 |
| Height | 176.9+11 | 171.6+9.6 | 0.98 (0.84, 1.15) | 0.82265 |
| **Smoking** | | | | |
| Never | 6 | 136 | 0.57 (0.25, 1.29) | 0.91012 |
| Ex | 14 | 201 | 1.11 (0.65, 1.91) | 0.35401 |
| Current | 7 | 71 | 1.66 (0.76, 3.61) | 0.10005 |
| **Alcohol** | | | | |
| Never | 8 | 174 | 0.57 (0.28, 1.15) | 0.94212 |
| Ex | 13 | 153 | 1.55 (0.88, 2.73) | 0.06531 |
| Current | 6 | 81 | 1.15 (0.50, 2.64) | 0.36791 |
| **Education** | | | | |
| <= 8th grade | 5 | 57 | 1.40 (0.56, 3.49) | 0.23556 |
| Some high school | 4 | 82 | 0.69 (0.25, 1.89) | 0.76445 |
| High school grad | 12 | 135 | 1.62 (0.90, 2.92) | 0.05514 |
| Some college | 3 | 82 | 0.50 (0.16, 1.57) | 0.88290 |
| College grad | 3 | 35 | 1.33 (0.41, 4.33) | 0.31678 |
| Post-grad | 0 | 17 | | |
| **Marital status** | | | | |
| Single | 5 | 100 | 0.70 (0.29, 1.72) | 0.78181 |
| Married | 10 | 187 | 0.70 (0.37, 1.31) | 0.86868 |
| Divorced | 6 | 86 | 1.07 (0.47, 2.45) | 0.43655 |
| Widowed * | 6 | 35 | 3.04 (1.28, 7.24) | 0.00587 |
| **Live alone** | | | | |
| No | 14 | 300 | 0.39 (0.23, 0.66) | 0.99974 |
| Yes * | 13 | 108 | 2.58 (1.45, 4.59) | 0.00062 |
| **NYHA** | | | | |
| 1 | 4 | 60 | 1.01 (0.37, 2.78) | 0.49331 |
| 2 | 17 | 244 | 1.14 (0.70, 1.87) | 0.29753 |
| 3 | 6 | 100 | 0.88 (0.39, 2.01) | 0.61949 |
| 4 | 0 | 4 | | |
| **Diabetes** | | | | |
| No | 16 | 252 | 0.90 (0.54, 1.49) | 0.65792 |
| Yes | 11 | 156 | 1.11 (0.60, 2.05) | 0.36836 |
| **Heart failure** | | | | |
| No | 7 | 270 | 0.18 (0.08, 0.38) | 1.00000 |
| Yes * | 20 | 138 | 5.59 (3.50, 8.93) | 0.00000 |




Odds ratio for the 12 predictors of new-onset atrial fibrillation identified by stepwise regression.

AF = atrial fibrillation; BMI = body mass index; CI = confidence interval; NYHA = New York Heart Association; OR = odds ratio.

fibrillation. The impact of social isolation on cardiovascular mortality has been shown previously,[28,29] and our analysis has shown that it disproportionately affects the non-White ethnic group of the WARCEF patients. As shown in Table 1, there was an under-representation of non-White ethnic group within the population recruited, therefore the conclusions drawn may be limited. However, our analysis serves not only as a reminder of the impact of social factors on health, but it also highlights the need to explore the impact of social factors on this under-represented group of patients. The study underlines the potential impact that social interventions, or the lack thereof, have on ethnic minorities in terms of developing atrial fibrillation.

## Association of Heart Failure Hospitalization and Atrial Fibrillation

The association between heart failure hospitalization and atrial fibrillation is seen in the analyses performed. This not only accentuates the common pathological background of both conditions, but also highlights the interdependency between the 2 pathologies—patients with atrial fibrillation are more likely to decompensate and be hospitalized with heart failure; similarly, patients with heart failure are more likely to have an electrocardiogram performed or monitored while in the hospital, which increases the detection of atrial fibrillation.

The OPTIMIZE-HF study identified arrhythmia as one of the most frequent (13.5%) precipitating factors for heart failure hospitalization,[30] and these arrhythmias may be transient. An interesting pilot study, CARRYING-ON HF, which implanted loop recorders in patients admitted with acute heart failure, identified 8 participants with asymptomatic paroxysmal atrial fibrillation of the 18 participants.[31] Our analysis emphasizes that the close relationship between the 2 and heart failure hospitalization, especially in patients without known atrial fibrillation, should prompt clinicians to consider looking longer and harder for atrial fibrillation, as they would benefit from the initiation of OAC.

## The Use of ML in Exploring Novel Trends/Risk Factors

The use of ML methodology in our paper has provided a different viewpoint toward identifying novel risk factors that may not be immediately apparent to clinicians. Not only will we learn more about the disease processes and latent associations, but it may also help identify other interventions that may improve clinical outcomes in patients. ML will also be able to speed up integration of large volumes of data, although the pitfall of poor data quality remains a huge concern. However, with the use of adjudicated randomized clinical trials data where the data collection process is more robust, such as WARCEF, this should be of a lesser concern.

Effective implementation of ML technologies can be a powerful tool toward early and accurate atrial fibrillation detection enabling precision in stratification of patient cohorts. The health impact of experiences of loneliness among people from Black, Asian, and Minority Ethnic groups seems to have far-reaching implications, including an apparent higher risk of developing atrial fibrillation. This needs to be further investigated and understood so that a series of recommendations for national policymakers, local authorities, and health service providers can be put forward, and people experiencing loneliness can get the help and support they need.

## Limitations

Being a post hoc analysis, there may be presence of unknown confounders that may limit the conclusions drawn from the results. Additionally, as the effective sample size used for training in each random splitting is much less than the original sample size, the results may not be representative of the whole dataset. Lastly, as there was no systematic screening for atrial fibrillation conducted in the WARCEF trial, there may be an underestimation of patients who developed atrial fibrillation.

## CONCLUSIONS

Our study has highlighted that "social factors" may disproportionately increase the risk of atrial fibrillation in the under-represented non-White patient groups with heart failure. There is a need for more studies focusing on stratification of multiracial cohorts to better uncover the heterogeneity of atrial fibrillation mechanisms across different racial groups. The integration of ML into clinical practice is mutually beneficial. It offers clinicians improved clinical workflow and diagnostic accuracy and cost-effectiveness. Clinicians offer ML the essential exposure it needs to learn complex clinical case management. This, in turn, increases critical mass and encourages buy-in from multiple disciplinary approaches.

## References


1. Santhanakrishnan R, Wang N, Larson MG, et al. Atrial fibrillation begets heart failure and vice versa. *Circulation* 2016;133(5):484–92.
2. Wang TJ, Larson MG, Levy D, et al. Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study. *Circulation* 2003;107(23):2920–5.
3. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2016;37(27):2129–200.
4. Lip GY, Gibbs CR. Does heart failure confer a hypercoagulable state? Virchow's triad revisited. *J Am Coll Cardiol* 1999;33(5):1424–6.
5. Ding WY, Gupta D, Lip GYH. Atrial fibrillation and the prothrombotic state: revisiting Virchow's triad in 2020. *Heart* 2020;106(19):1463–8.
6. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 2021;42(36):3599–726.
7. Hopper I, Skiba M, Krum H. Updated meta-analysis on antithrombotic therapy in patients with heart failure and sinus rhythm. *Eur J Heart Fail* 2013;15(1):69–78.
8. Tseng AS, Noseworthy PA. Prediction of atrial fibrillation using machine learning: a review. *Front Physiol* 2021;12:752317.
9. Homma S, Thompson JLP, Pullicino PM, et al. Warfarin and aspirin in patients with heart failure and sinus rhythm. *N Engl J Med* 2012;366(20):1859–69.
10. R Core Team. R: a language and environment for statistical computing. Available at: https://www.R-project.org/. Accessed July 14, 2023.
11. Audigier V, Husson F, Josse J. A principal components method to impute mixed data. *Adv Data Anal Classif* 2013;10:5–26.
12. Breiman L. Random Forests. *Machine Learning* 2001;45:5–32.
13. Freund Y, Schapire RE. A desicion-theoretic generalization of on-line learning and an application to boosting. In: Vitányi P, ed. *Computational Learning Theory*, Berlin, Heidelberg, Germany: Springer Berlin Heidelberg; 1995:23–37.

14. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016. https://doi.org/10.1145/2939672.2939785.
15. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85–117.
16. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurorobot* 2013;7:21.
17. Sasaki Y. The truth of the F-measure. *Teach Tutor Mater* 2007.
18. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* 1975;405(2):442–51.
19. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006;27(8):861–74.
20. Dupre ME, George LK, Liu G, Peterson ED. Association between divorce and risks for acute myocardial infarction. *Circ Cardiovasc Qual Outcomes* 2015;8(3):244–51.
21. Essien UR, Kornej J, Johnson AE, Schulson LB, Benjamin EJ, Magnani JW. Social determinants of atrial fibrillation. *Nat Rev Cardiol* 2021;18(11):763–73.
22. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26(6):565–74.
23. Ward JH. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 1963;58:236–44.
24. Mead A. Review of the development of multidimensional scaling methods. *J R Stat Soc Ser D Stat* 1992;41(1):27–39.
25. Kato TS, Di Tullio MR, Qian M, et al. Clinical and echocardiographic factors associated with new-onset atrial fibrillation in heart failure - subanalysis of the WARCEF trial. *Circ J* 2016;80(3):619–26.
26. Soliman EZ, Alonso A, Goff DC Jr. Atrial fibrillation and ethnicity: the known, the unknown and the paradox. *Future Cardiol* 2009;5(6):547–56.
27. Mou L, Norby FL, Chen LY, et al. Lifetime risk of atrial fibrillation by race and socioeconomic status: ARIC Study (Atherosclerosis Risk in Communities). *Circ Arrhythm Electrophysiol* 2018;11(7):e006350.
28. Pantell M, Rehkopf D, Jutte D, Syme SL, Balmes J, Adler N. Social isolation: a predictor of mortality comparable to traditional clinical risk factors. *Am J Public Health* 2013;103(11):2056–62.
29. Ikeda A, Iso H, Kawachi I, et al. Social support and stroke and coronary heart disease: the JPHC study cohorts II. *Stroke* 2008;39(3):768–75.
30. Fonarow GC, Abraham WT, Albert NM, et al. Factors identified as precipitating hospital admissions for heart failure and clinical outcomes: findings from OPTIMIZE-HF. *Arch Intern Med* 2008;168(8):847–54.
31. Vanoli E, Mortara A, Diotallevi P, et al. Cardiac rhythm monitoring after acute decompensation for heart failure: results from the CARRYING ON for HF pilot study. *JMIR Res Protoc* 2016;5(2):e62.

## SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.amjmed.2023.07.019.

# APPENDIX

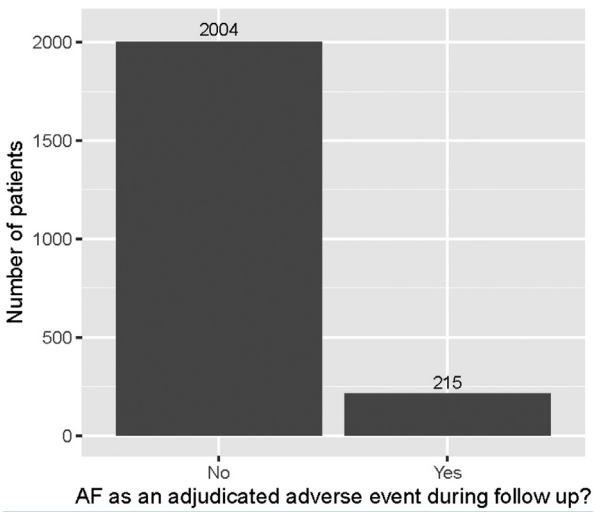
## Model Validation: 5-fold Stratified Cross-Validation

Prior to data analysis, the dataset was randomly split into 5 independent and randomly stratified folds (n = 2219, $n_1$ = 444, $n_2$ = 444, $n_3$ = 444, $n_4$ = 444, $n_5$ = 443). The class distribution inside these 5 folds is: 4 folds have exactly 444 samples, of which 43 are cases (positive atrial fibrillation as an adverse event) and 401 are controls (no atrial fibrillation as an adverse event). The last and fifth fold has exactly 443, samples of which 43 are cases (positive atrial fibrillation as an adverse event) and 400 are controls (no atrial fibrillation). At each run of the cross-validation, 4 folds are combined and used as training data and a fifth fold is set aside for validation of the models (eg, training = $n_1 + n_2 + n_3 + n_4$ = 1776 and validation = $n_5$ = 443). Each fold is used exactly once as validation set. The training data (n = 1776) are further split into stratified sets of 80% of the data for training the models and 20% of the data for testing the models. To make the results more comparable, all machine learning (ML) algorithms presented in this paper were trained, tested, and validated on the same training, testing, and validating sets. The classification results presented are always the results computed and averaged over the 5 validation folds.

As depicted in Supplementary Figure 1, the 2 classes (no atrial fibrillation as an adverse event, and atrial fibrillation as an adverse event) are highly unbalanced.

To cope with this unbalanced dataset, once the training set (4 combined folds, eg, n = 1776) is created and split into stratified sets of 80% training and 20% testing sets, random undersampling (RUS), which was the sampling technique with the best overall performance in preliminary tests, was applied to the training set only, but not to the testing set, which always keeps the original data class distribution. Therefore, when applying RUS, the models are actually trained with approximately 276 samples—of which 138 (80% of 43 cases × 4 combined folds) are cases (positive atrial fibrillation as an adverse event) and another 138 are randomly selected control samples to achieve a 50% split of the classes in the training dataset. Using only one split of the 138 control samples (which are chosen out of approximately 80% of 4 × 401 samples = 1283 samples) could bias the model building process, because approximately 1283 − 138 = 1145 samples are left out of training. Therefore, RUS is repeated 100 times for each training set (4 combined training folds), and each time the 138 control samples are independently and randomly selected to compose an RUS training dataset with the same 138 cases. For each of the 100 RUS repetition, the models are validated on the fifth hold-out validation fold.

The classification results obtained by each model are averaged over the 5-fold stratified cross-validations times 100 RUS splits = 500 independent cross-validations. Supplementary Figure 2 shows the workflow of the model training and validation.



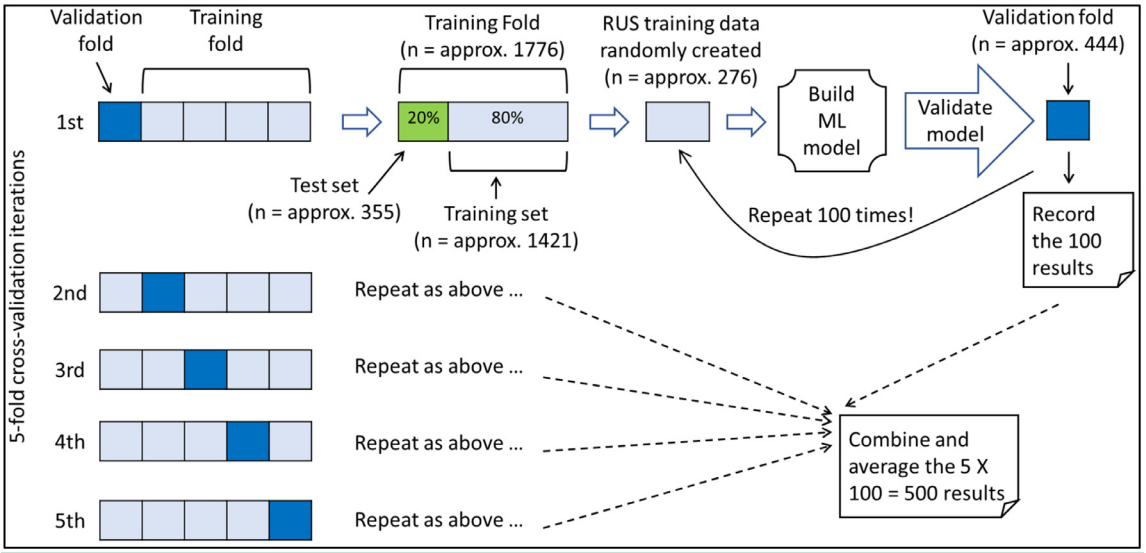

**Supplementary Figure 1** Class distribution for developing atrial fibrillation as an adverse event during the follow-up period. AF = atrial fibrillation.

## Data Pre-Processing and Imputation

Once the 5 folds were constructed, missing values were imputed using an R package[10] "*imputeFAMD*"[11] and applied independently to each training and validation fold—that is, 5 × (combined 4 training folds) and 5 × (1 validation fold). *imputeFAMD* works well with mixed data (continuous and categorical variables) and uses a principal component method "factorial analysis for mixed data" to replace missing values. To avoid data leakage, the imputation process has been independently performed on each split of the dataset; namely, on training set individually, then on test set individually, and finally on validation set individually. One limitation of this imputation process is that, even though an appropriate imputation algorithm was selected for the type of data under analysis (mixed data), no other imputation method was assessed to minimize possible randomness of the imputation process.

## Machine Learning Algorithms Predicting the Occurrence of Atrial Fibrillation as an Adverse Event

Four ML models were applied to predict the outcome of interest (incidence of new-onset atrial fibrillation as an adverse event during follow-up): Random Forest (RF),[12] Adaptive Boosting (AdaBoost),[13] eXtreme Gradient Boosting (XGBoost),[14] and Artificial Neural Network (ANN).[15] As the construction of a single decision tree model may be overly sensitive to the training data, RF uses bootstrap sampling to build a collection of random and independent decision trees, a forest; and the trees have no predetermined size. Via aggregation, each tree in the forest is used to classify new samples, and the class assignment is decided by the majority voting, the class predicted by the largest




**Supplementary Figure 2** Workflow of model training and validation under random under sampling. ML = machine learning; RUS = random undersampling.

number of trees in the forest. In RF each tree has an equal vote (weight) on the final classification. AdaBoost also builds a forest of decision trees, but each tree is composed of only 1 node and 2 leaves; these trees are termed stumps. Based on how well each stump classifies the samples, in AdaBoost, some stumps get a higher weight on the classification than others. In addition, the classification error that each stump makes influences the building of subsequent stumps. New samples are classified based on the weighted voting from all stumps. XGBoost is an advanced implementation of gradient boosting (GB)[16] algorithm, which builds fixed-size decision trees based on the errors made by previous trees and combines them into an ensemble model. XGBoost uses more complex regularization methods than GB to improve model generalization. An ANN is a learning algorithm inspired by the structure of the human brain. Using a set of nodes called neurons and a set of variable weights associated to those neurons, an ANN takes in data and trains itself to recognize patterns in the data and predict desired outputs. The neurons are the core of the network, which is usually composed of multiple layers of neurons, including input layer, output layer, and a set of hidden layers in between. Via activation functions, the neurons transmit information to the next layers and the output layer is compared with the known observed values for the samples. The weights of the neurons are adjusted to improve the outputs and minimize errors. For each of the 4 ML models used for data analysis, the following model performance metrics were computed: accuracy, sensitivity, specificity, F1-score,[17] Matthews correlation coefficient,[18] and the area under the receiver operating characteristic curve.[19] ML is a proven enabler of advancements in health care. It can unlock data insights needed to support data-driven decisions for diagnostics and treatments.

## Hyperparameter Tuning

Effective application of ML models requires appropriate model architecture design choices such as parameter selection or tuning. Parameters that define model architecture are termed hyperparameters. The process of searching for the ideal model architecture is referred to as hyperparameter tuning. This work used a grid search hyperparameter tuning method computed over the test sets. With this technique, we simply built a model for each possible combination of all hyperparameter values provided, evaluating each model on the test sets, and selecting the architecture that produced the best results. For RF, the number of decision trees that are combined to create the final prediction was tested from 200 to 2000 in steps of 200. The tree depth values were tested between 5 and 31 in steps of 2. For RF the combination of (number of trees, tree depth) that produced the best performance was (1000, 20), respectively (all other parameters were default values from R). For AdaBoost and XGBoost, the base estimator used was decision trees. For both of these algorithms, the combination of (number of trees, tree depth) that produced the best performance was also (1000, 20), respectively—values tested the same as for RF. The learning rates for AdaBoost and XGBoost were searched between 0.2 and 0.8 (in steps of 0.1), and for both algorithms the best value was 0.6 (all other parameters were default values from R). For ANNs, the number of hidden layers tested was between 1 and 10, whereas the learning rate values varied between 0.05 and 1 (in steps of 0.05). For ANNs, the combination of (hidden layers, learning rate) that produced best performance was (4, 0.15), respectively (all other parameters were default values from R).