RESEARCH ARTICLE | MARCH 31 2025

Coagulation parameters as predictors of paroxysmal atrial fibrillation: Data modeling by logistic regression and ROC analysis

Krasimira Prodanova; Maria Negreva

AIP Conf. Proc. 3182, 110001 (2025)
https://doi.org/10.1063/5.0246073

Coagulation Parameters as Predictors of Paroxysmal Atrial Fibrillation: Data Modeling by Logistic Regression and ROC Analysis

Krasimira Prodanova[1, a)] and Maria Negreva[2, b)]

[1]*Technical University of Sofia, 8 Kl. Ohridski Blvd., 1000 Sofia, Bulgaria*
[2]*Medical University of Varna "Prof. P. Stoyanov", 55 Marin Drinov sreed, 9002 Varna, Bulgaria*

[a)]Corresponding author: kprod@tu-sofia.bg
[b)]mnegreva@abv.bg

Abstract. Our research on patients with paroxysmal atrial fibrillation (PAF) found significant deviations in fourteen coagulation indicators already occurring in the first twenty-four hours of the clinical manifestation of the disease, namely: significantly increased plasma activity of coagulation factors FII, FV, FVII, FVIII, FIX, FX, FXI, FXII and plasma activity of von Willebrand factor, as well as significantly increased plasma levels of tissue factor, FVIII, vWF, fibrinopeptide A and prothrombin fragment F1+2. The early nature of the deviations raises an important question, namely, to what extent are they a consequence of atrial fibrillation or are they closely related to the clinical manifestation of the arrhythmia and precede it. In search of their predictive value for PAF occurrence, the performed logistic regression analysis and ROC analysis showed that, of the investigated indicators, plasma FVIII activity gives the best diagnostic opportunity to identify the occurrence of the disease (AUC=0.85, Acc=0.85, Se 99 %, Sp 69%, with over 70% correctly classified cases). Our results are important for clinical practice and are a good prerequisite for further prospective clinical studies. Establishment of significant biomarkers predicting PAF appearance will enable detection of patients at increased risk of thromboembolic events, prevention optimization and reduction of the frequency of thromboembolic events associated with the rhythm disorder.

INTRODUCTION

PAF high morbidity and significant thromboembolic potential impose a need for its reliable prediction, which would lead to therapeutic approach optimization [1, 2]. In recent years, the idea that coagulation markers can be used as predictors of arrhythmia occurrence has been increasingly suggested [3, 4]. This assumption has its logical justification. It is well known that the role of coagulation goes far beyond the limits of hemostasis and that it is involved in a number of pathological processes, such as oxidative stress, inflammation, tissue remodeling, etc., which at the same time are part of AF pathophysiological mechanisms [5, 6]. In our study on the disease, we found significant deviations in fourteen key coagulation parameters already occurring in the first 24 hours from the clinical appearance of the arrhythmia (on average 8.14 ± 0.76 hours from the onset of the arrhythmia), namely: significantly increased plasma activity of coagulation factors FII , FV, FVII, FVIII, FIX, FX, FXI, FXII and plasma activity of von Willebrand factor (vWF), as well as significantly increased plasma levels of tissue factor (TF), FVIII, vWF, fibrinopeptide A (FPA) and prothrombin fragment 1+2 (F1+2) [7, 8, 9, 10, 11]. For the first time convincing laboratory evidence has been presented for the early development of hypercoagulability in PAF, already occurring in the first 24 hours of the disease. The early nature of the deviations raises an important question, namely, to what extent are they a consequence of atrial fibrillation or are they closely related to the clinical manifestation of the arrhythmia and precede it. This determined our goal to evaluate the prognostic value of the studied coagulation parameters by PAF occurrence probability.

Proceedings of the 49th International Conference "Applications of Mathematics in Engineering and Economics"
AIP Conf. Proc. 3182, 110001-1–110001-9; https://doi.org/10.1063/5.0246073

STATISTICAL ANALYSIS

The statistical software STATISTICA 13.0 and R core pacage are use to analyze the data. The continuous variables are shown as mean and standard deviation. The level of significance for testing of statistical hypothesises was p=0,05.

The predictive role of the plasma parameters (factors) of the coagulation system for PAF occurrence is analysed using univariate logistic regression analysis and Receiver Operating Characteristic (ROC) analysis.

The dependent variable *y* is binary - presence/absence of PAF. We use logistic regression to estimate the probabilities for *y* (response states) based on the *x* values of the predictor (studied coagulation factor) measured in the sample [12]. The probability *p(x)* of PAF depends on the linear function:

$$d(x) = B_0 + B_1 x,$$

where B_0 and B_1 are the parameters of the model. The estimate of the occurrence probability *p(d(x))* of PAF at a given value of the factor is calculated with:

$$p(d) = \frac{e^d}{1+e^d},$$

where *e* is Napier's constant.

The odds ratio (OR), which measures the strength of the association between the two event states is calculated using the logistic model. OR is defined as:

$$odd\ ratio = \frac{p(d)}{1+p(d)},$$

i.e. as the ratio of the chances of being in one or the other category (absence/presence of the PAF).

The end result of the logistic model is a classification of the cases to one of the two states of *y*, generating errors of type I and type II, namely:

- Type I error - positive cases wrongly classified as negative (false negative observations, false omission, False Negatives);
- Type II error - negative observations wrongly classified as positive (false positive cases, false detection, False Positives).

Their analysis is the subject of the ROC (Receiver Operating Characteristic) analysis presented below.

ROC (Receiver Operating Characteristic) analysis

In the last four decades, ROC analysis has been the main method for evaluating accuracy of a given diagnostic test, resp. ability of a certain biomarker to distinguish "sick" from "healthy" subjects [13]. In dichotomous statistical models, such as logistic regression analysis, it gives a correct assessment of their predictive ability. In ROC analysis terminology, it is assumed that the classifier has a parameter, that by varying, gives the separation into two classes. This parameter is called threshold point or cut-off value. Depending on it, different magnitudes of type I and II errors are obtained.

The ROC curve is a graph of dependencies, with the Se values (TPR, true positive rate) on the Y axis, and the fraction of false positive cases FPR=1−Sp (FPR, false positive rate) on the X axis (Figure 1) . Thus, each point on it represents a sensitivity/specificity ratio corresponding to a decision threshold. The closer the curve is to the upper left corner, the higher the predictive ability of the model. The diagonal line corresponds to the "useless" classifier (probability *p(x)*=0,5), i.e. complete indistinguishability of the two classes.

The ideal model has 100% sensitivity and specificity, which is practically impossible to obtain. The compromise is found using an optimal threshold point (optimal cut-off value), which is established according to a set criterion for its determination. The most commonly used is the Youden's index, also known as the Youden's J statistic, since it reflects the intention to maximize the correct part of the classification. The main objective of this criterion is to maximize the difference between TPR (Se) and FPR=(1–Sp) (Figure1).

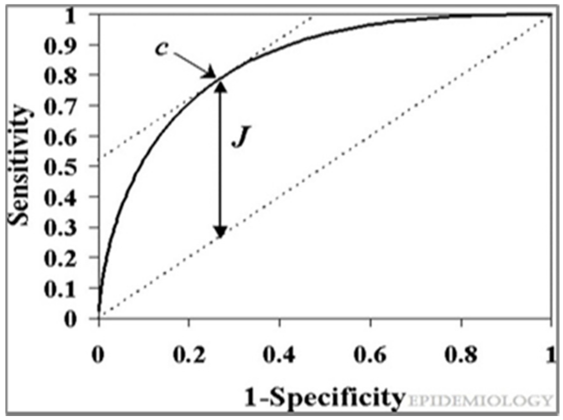

FIGURE. 1. Receiver Operating Characteristic (ROC) curve, Youden's (J) statistic and optimal cut-off value (c).

We constructed ROC-curves with their respective optimal threshold points for the studied plasma parameters of the coagulation system, determined according to Youden's index.

RESULTS AND DISCUSSION

Hemostasis indicators were examined once in peripheral venous blood in 51 PAF patients with an episode duration of ≤24 hours and 51 controls in the period from October 2010 to May 2012 in the First Clinic of Cardiology with an ICU at the St. Marina University Hospital – Varna, after approval by the Research Ethics Committee of the same hospital (№35/29.10.2010) and Medical University of Varna (No. 9/14.10.2010).

The logistic models of studied indicators, except for plasma levels of F1+2, showed a predictive value ($p<0.001$) for PAF occurrence. Thus, the null hypothesis of independence of PAF appearance from the values of these indicators was rejected. The estimated parameters of the logistic model, as well as the correspondingly obtained values for the odd ratio and percentage of cases correctly classified with the model for each of the factors are shown in Table 1.

As can be seen from the values of B_1 coefficients, with increasing plasma activity of FII, FV, FVII, FVIII, vWF, FIX, FX, FXI, FXII and plasma levels of TF, FVIII, vWF and FPA, the probability of PAF occurrence increases.

However, the created logistic models differed significantly in their correctness. The percentage of correctly classified cases varied widely from 54.37% for the FPA model to 70.87% and 72.82% for vWF levels and FVIII activity, respectively. This predetermines a great difference in their prognostic value and benefits for their clinical application.

Three of the hemostasis indicators enabled correct clasiffication of over 70% of observed cases, namely: FVIII plasma activity and vWF plasma levels and activity.

Figure 2, Figure 3 and Figure 4 show the estimated probabilities using logistic models for FVIII plasma activity, vWF plasma levels and activity, respectively.

From these models, the values of the indicator can be calculated at a probability of 0.9 (i.e. a 90% chance for the occurrence of the disease) and at 0.1.

Using the probability function in Figure 2, we report a 90% chance of PAF occurrence at FVIII activity of 240% and 10% at an indicator value of 60%.

TABLE 1. Estimated parameters of the logistic models, odd ratios and % of correctly classified cases.

Coagulation marker	B_0	B_1	OR	Correct (%)
FII act (%)	-2.854	0.022	2.49	61.17%
TF (pg/mL)	-3.243	0.015	2.70	62.14%
FV act (%)	-3.379	0.022	4.52	67.96%
FVII act (%)	-3.893	0.030	3.52	65.05%
FVIII lv (%)	-2.118	0.021	2.12	59.22%
FVIIIact (%)	-3.626	0.025	5.94	70.87%
vWF lv (%)	-1.641	0.011	8.53	72.82%
vWF act (%)	-2.928	0.020	5.94	70.87%
FIX act (%)	-2.854	0.025	2.41	60.78%
FX act (%)	-3.243	0.017	2.30	60.19%
FXI act (%)	-3.379	0.029	2.11	59.22%
FXII act (%)	-3.893	0.013	2.11	59.22%
FPA(ng/mL)	-2.118	0.611	1.43	54.37%

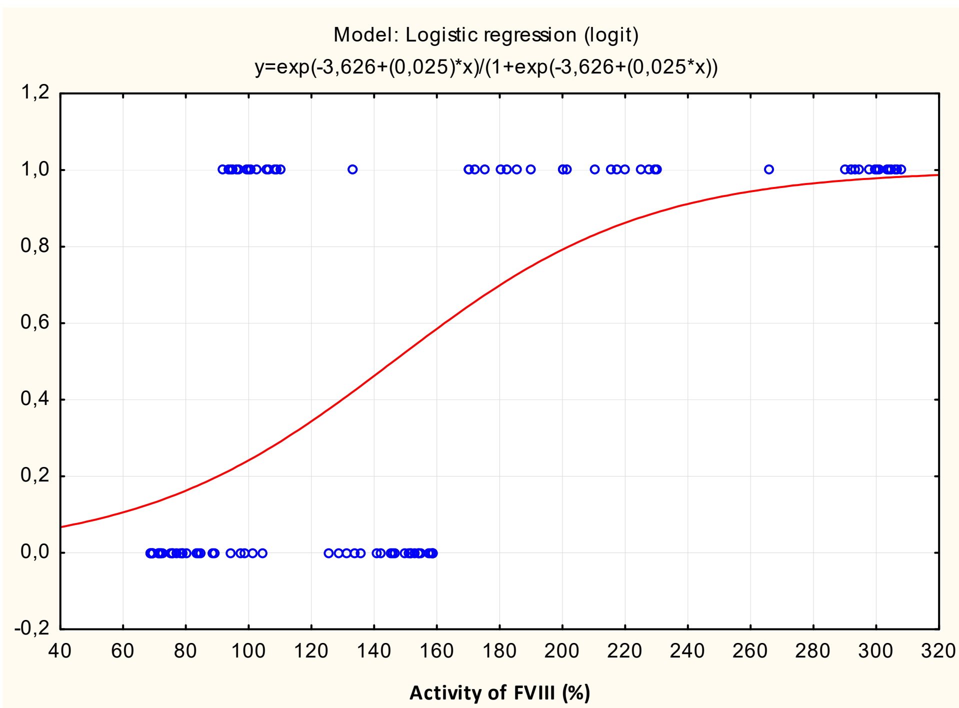

FIGURE 2. Probability of PAF occurrence in relation to FVIII plasma activity

Using the probability function in Figure 2, we report a 90% chance of PAF occurrence at FVIII activity of 240% and 10% at an indicator value of 60%.

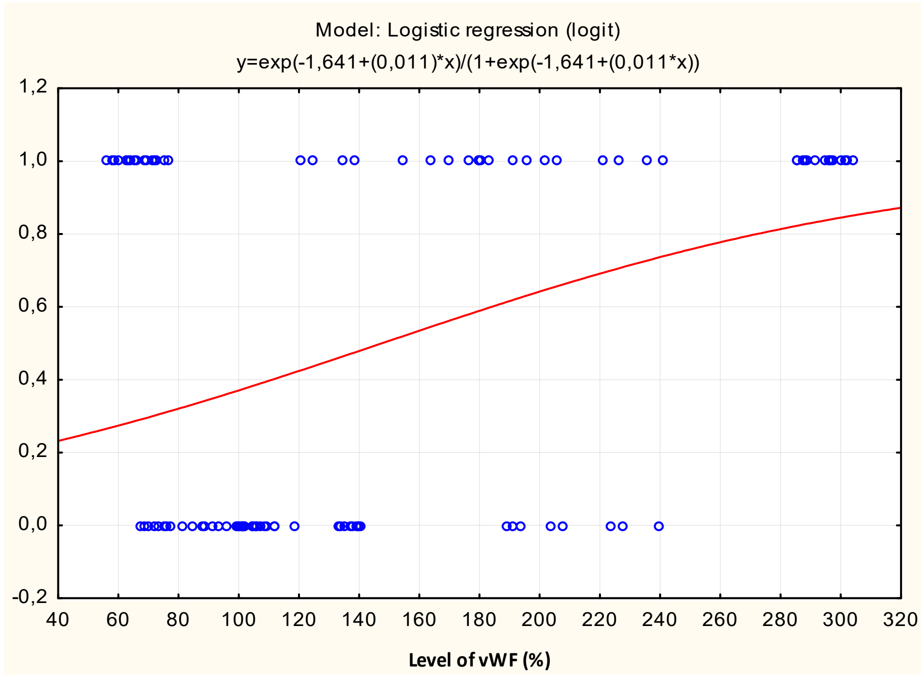

FIGURE 3. Probability of PAF occurrence in relation to vWF plasma levels

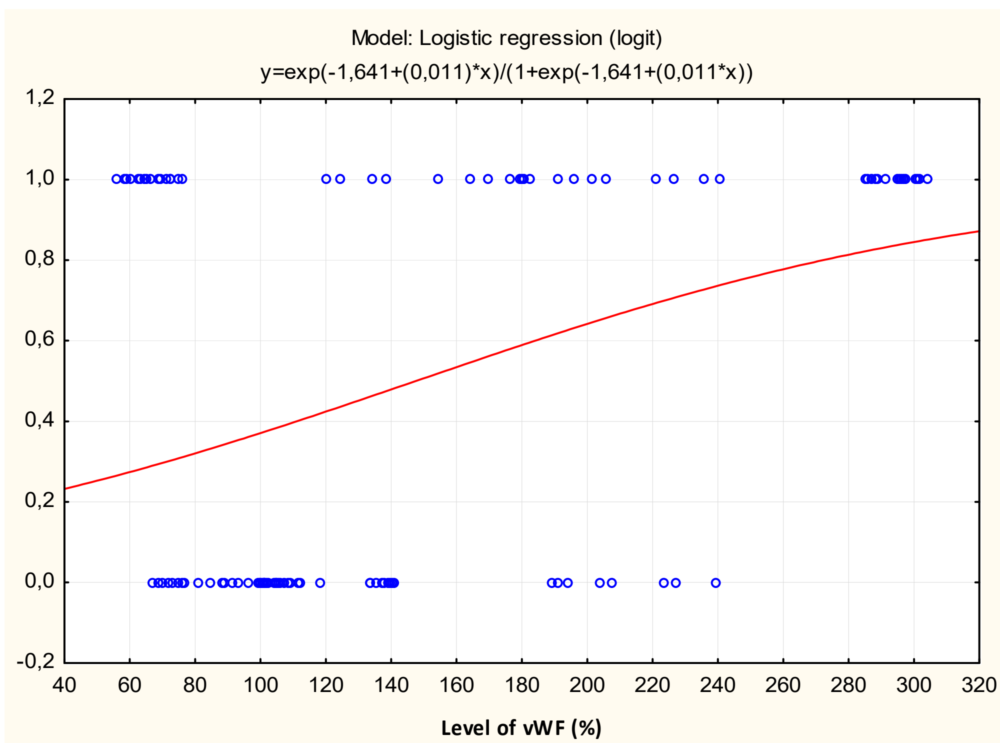

FIGURE 4. Probability of PAF occurrence in relation to vWF plasma activity

Figure 3 shows that the probability of developing the disease is 85% at vWF levels of 302% and only 25% at 58%. The indicator vWF plasma activity of 260% determines a 90% probability of PAF occurrence (Figure 4). As the values of the indicator decrease, the probability also decreases and reaches 10% at a vWF activity of 40%.

The constructed logistic models give a good idea about the ability of the investigated coagulation indicators to predict PAF in its binary manifestation absence/presence. However, they do not indicate the probability of type I and type II errors, occurring when classifying the values, i.e. probability of false negative or false positive results, respectively.

In order to evaluate the quality of the obtained logistic models, resp. accuracy of the studied hemostasis indicators as possible laboratory diagnostic markers for PAF appearance, we performed a ROC analysis of those that presented statistical significance in the logistic regression. The calculated AUC, optimal threshold point found according to the Youden's index and corresponding accuracy (Acc), sensitivity (Se) and specificity (Sp) for each indicator are presented in Table 2.

TABLE 2. Data from ROC analysis of studied coagulation indicators

Indicator	AUC	Optimal trhreshhold point	Youden 's index	Accura cy (Acc)	Sensiti vity (Se)	Specifi city (Sp)
FII act (%)	-2.854	0.022	2.49	61.17%	0.99	0.59
TF (pg/mL)	-3.243	0.015	2.70	62.14%	0.94	0.63
FV act (%)	-3.379	0.022	4.52	67.96%	0.99	0.63
FVII act (%)	-3.893	0.030	3.52	65.05%	0.99	0.59
FVIII lv (%)	-2.118	0.021	2.12	59.22%	0.98	0.37
FVIIIact (%)	-3.626	0.025	5.94	70.87%	0.99	0.69
vWF lv (%)	-1.641	0.011	8.53	72.82%	0.85	0.61
vWF act (%)	-2.928	0.020	5.94	70.87%	0.98	0.63
FIX act (%)	-2.854	0.025	2.41	60.78%	0.57	0.96
FX act (%)	-3.243	0.017	2.30	60.19%	0.99	0.55
FXI act (%)	-3.379	0.029	2.11	59.22%	0.99	0.58
FXII act (%)	-3.893	0.013	2.11	59.22%	0.87	0.57
FPA(ng/mL)	-2.118	0.611	1.43	54.37%	0.98	0.55

Comparison of ROC curves by AUC estimation showed that the most effective models for PAF prediction were the models of FVIII activity (AUC=0.85, Figure 5), FVII activity (AUC=0.81, Figure 6), TF levels (AUC=0.81 , Figure 7) and FXI activity (AUC=0.81, Figure 8), ranked in descending order, respectively (Table 2). Their model quality was determined as excellent (AUC>0.8). Indicator vWF levels and activity, despite their good ability to correctly classify cases (72.82% and 70.87% of cases, Table 1), showed lower model efficiency (AUC=0.64 and AUC=0.79, Table 2).

The four best performing models presented high predictive model accuracy, except for FXI activity, where at AUC=0.81, the accuracy was only 0.57.

Indicator FVIII activity showed an accuracy of 0.85, followed by FVII activity (Acc=0.79) and TF levels (Acc=0.79) (Table 2). They also showed a very good combination of sensitivity and specificity at the determined optimal cutoff values (Table 2). In them, the logistic regression correctly classified 70.87%, 65.05% and 62.14% of PAF cases (Table 1).

Summarizing logistic regression and ROC analysis data, FVIII activity emerges as the coagulation indicator optimally combining the ability to correctly classify cases with good efficiency and accuracy.

The optimal threshold values of all indicators were determined using the Youden's index (Table 3). Based on the previously constructed logistic regression models, we calculated PAF occurrence at these threshold values. At an optimal FVIII activity threshold value of 158.89%, the sensitivity of the indicator for detecting PAF presence is 99% with a specificity of 69% (Table 2).

Of the indicators studied, FVIII plasma activity gives the best diagnostic opportunity to identify PAF occurrence (Table 1, Table 2).

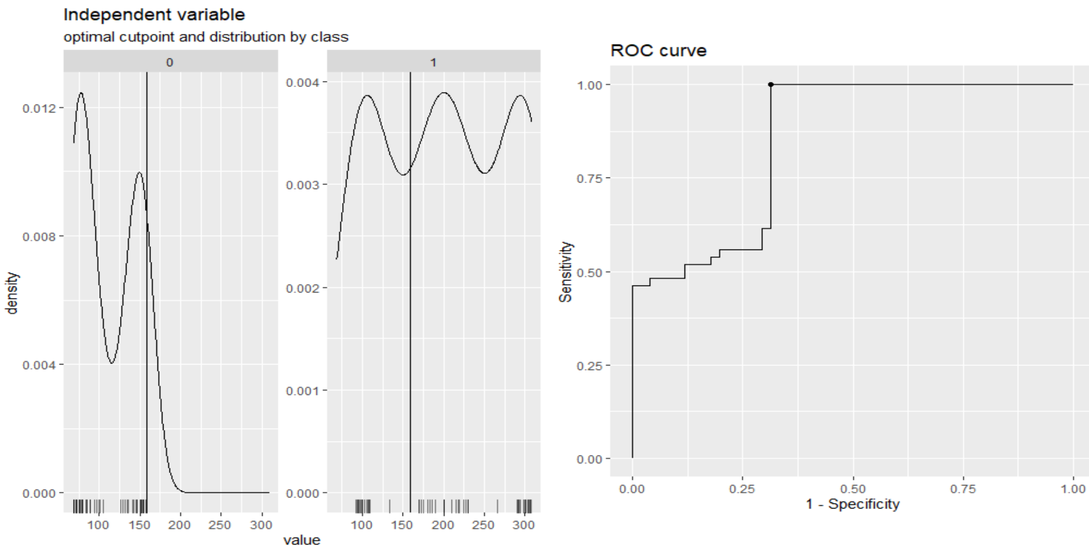

FIGURE 5. Left: Distribution of FVIII activity values in PAF classes absence (0) and presence (1) relative to the vertical line of the optimal threshold point. **Right:** ROC curve of the indicator with optimal threshold point

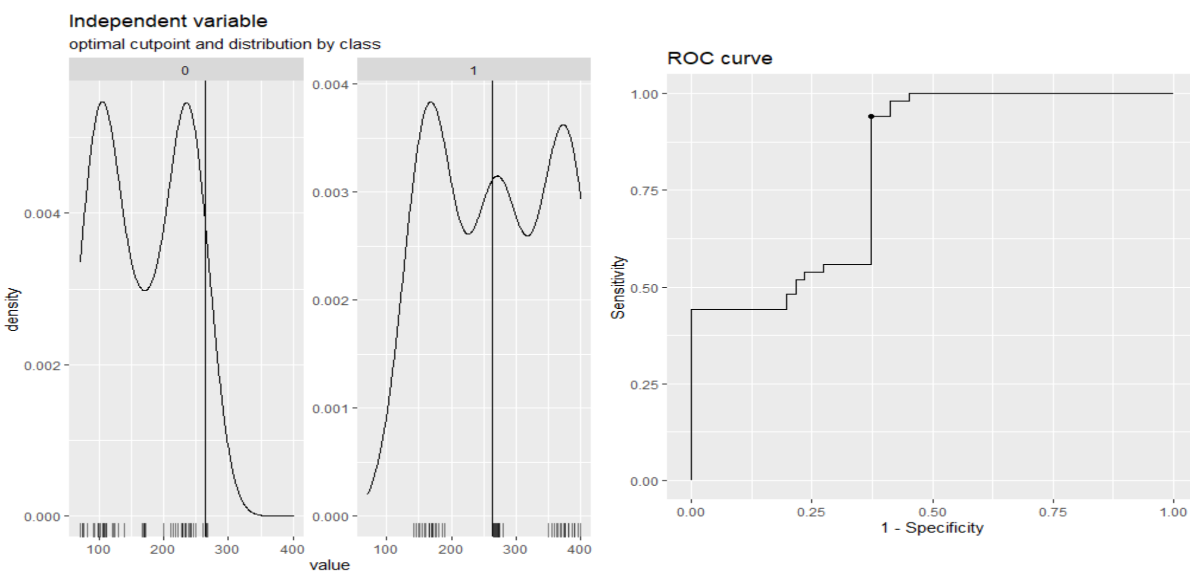

FIGURE 6. Left: Distribution of TF level values in PAF classes absence (0) and presence (1) relative to the vertical line of the optimal threshold point. **Right:** ROC curve of the indicator with optimal threshold point

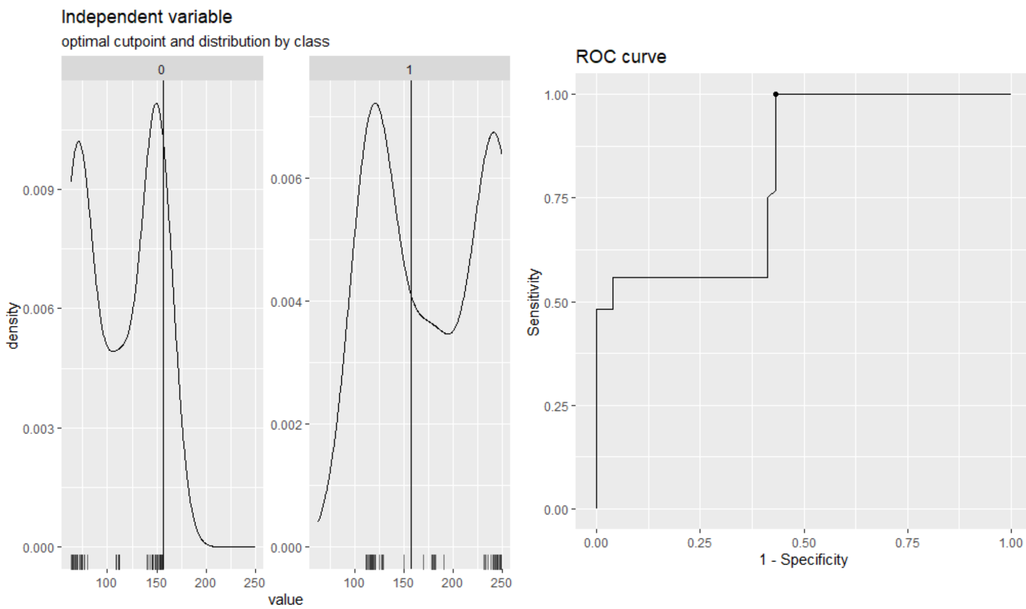

FIGURE 7. Left: Distribution of FXI activity values in PAF classes absence (0) and presence (1) relative to the vertical line of the optimal threshold point. **Right:** ROC curve of the indicator with optimal threshold point.

CONCLUSIONS

A number of possibilities for disease prediction have been sought over the years, both through demographic and clinical indicators, and through creation of assessment scales such as STAF, LADS and iPAB, which, however, do not present clinical benefits [14, 15, 16, 17].

The stratifying scales CHARGE-AF score, HAVOC score, ATLAS score, etc. also do not find an established place in the diagnostic process [18, 19]. That is why our results are important for clinical practice and a good premise for further prospective clinical studies. Establishing significant biomarkers predicting PAF appearance will enable detection of patients at increased risk of thromboembolic events, optimizing prevention and reducing frequency of PAF associated thromboembolic events. At the same time, it could provide clarity on optimal duration and timing of heart rate monitoring in patients with suspected atrial fibrillation.

ACKNOWLEDGEMENTS

The authors would like to tank the Research and Development Sector at the Technical University of Sofia for the financial support.

REFERENCES

1. A. Arauz at al, Neurologia **37**, 362-370 (2019).
2. E. J. Benjamin at al, Circulation **139**, e56–e528 (2019).
3. D. Ellis at al, Medicine **97**, e13830 (2018).
4. W. Alonso at al, Int J Cardiol **155**, 217-222 (2016).
5. R. B. Schnabel at al, Circulation **121**, 200–207 (2010).
6. H. M. Spronk at al, Eur Heart J **38**, 38-50 (2017).

7. M. Negreva at al, Minerva. Cardiol. Angiol. **69**, 269–276 (2021).
8. M. Negreva at al, Cardiol. Res. **11**, 22–32 (2020).
9. M. Negreva at al, J. Atr. Fibrillation **13**, 2297 (2020).
10. M. Negreva at al, Arch. Med. Sci. Atheroscler. Dis. **5**, e140–e147 (2020)..
11. M. Negreva at al, Medicine (Baltimore) **95**, e5184 (2016).
12. D. W. Hosmer at al, *Applied Logistic Regression*, John Wiley & Sons, (2013).
13. K. Hajian-Tilaki, Caspian Journal of Internal Medicine **4**, 627–635 (2013).
14. E. J. Benjamin at al, , JAMA **271**, 840–844 (1994).
15. A. D. Krahn at al, Am. J. Med. **98**, 476–484 (1995).
16. P. A. Wolf at al, Am. Heart. J. **131**, 790–795 (1996).
17. X. Chen at al, Clin. Cardiol. **41**, 1507–1512 (2018).
18. C. Kwong at al, Cardiology **138** , 133–140 (2018).
19. J. Mesquita at al, Europace **20**, f428–f435 (2018).