



OPEN

## Association of Premature Ventricular Contraction (PVC) with hematological parameters: a data mining approach

Nafiseh Hosseini<sup>1,2,4,10</sup>, Sara Saffar Soflaei<sup>2,3,10</sup>, Pooria Salehi-Sangani<sup>4</sup>, Mahdiyeh Yaghooti-Khorasani<sup>5</sup>, Bahram Shahri<sup>6</sup>, Helia Rezaefard<sup>2,3</sup>, Habibollah Esmaily<sup>7,8</sup>, Gordon A. Ferns<sup>9</sup>, Mohsen Moohebati<sup>2,3</sup>✉ & Majid Ghayour-Mobarhan<sup>2,3</sup>✉

Premature ventricular contraction (PVC) is characterized by early repolarization of the myocardium originating from Purkinje fibers. PVC may occur in individuals who are otherwise healthy. However, it may be associated with some pathological conditions. In this research the association between hematological factors and PVC was studied. In this study, 9,035 participants were enrolled in the Mashhad stroke and heart atherosclerotic disorder (MASHAD) cohort study. The association of hematological factors with PVC was evaluated using different machine learning (ML) algorithms, including logistic regression (LR), C5.0, and boosting decision tree (DT). The dataset was divided into training and test, and each model's performance was appraised on the test dataset. All data analyses used SPSS version 26 and SPSS Modeler 10. The results show that the Boosting DT was the most effective algorithm. Boosting DT had an accuracy of 98.13% and 96.92% for males and females respectively. According to the models, RDW and PLT were the most significant hematological factors for both males and females. WBC, PDW, and HCT for males and RBC, MCV, and MXD for females were also important. Some hematological factors associated with PVC were found using ML models. Further studies are needed to confirm these results in other populations, considering the novelty of the exploration of the relationship between hematological parameters and PVC.

**Keywords** Premature ventricular contraction, PVC, Complete blood count, CBC, Data Mining, Hematological parameters

Premature ventricular contraction (PVC) is characterized by early repolarization of the myocardium that originates from the Purkinje fiber rather than the sinoatrial node. Previous studies have shown that PVC may be related to heart diseases<sup>1</sup>. However, PVC may be triggered by non-cardiac causes, such as excessive alcohol consumption<sup>2</sup>, hyperthyroidism<sup>3</sup>, and sarcoidosis<sup>4</sup>. The results of previous studies about the clinical importance of PVCs have been controversial; nevertheless, the clinical benefits of PVC suppression are necessary in patients with cardiac diseases<sup>5-7</sup>. PVC is seen among 1-4% of the general population<sup>8,9</sup> and can lead to symptoms such as palpitations, chest pain, heart failure, and syncope<sup>10</sup>.

The complete blood count (CBC) is commonly requested in clinical practice. The characteristics of CBC components (hematological factors) include white blood cell (WBC), red blood cell (RBC), hemoglobin (Hb), mean corpuscular volume (MCV), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular

<sup>1</sup>Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. <sup>2</sup>International UNESCO Center for Health-Related Basic Sciences and Human Nutrition, Mashhad University of Medical Sciences, Mashhad, Iran. <sup>3</sup>Metabolic Syndrome Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. <sup>4</sup>Student Research Committee, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. <sup>5</sup>Radiation Oncology Research Center, Cancer Research Institute, Tehran University of Medical Sciences, Tehran, Iran. <sup>6</sup>Department of Cardiology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. <sup>7</sup>Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran. <sup>8</sup>Social Determinants of Health Research center, Mashhad University of Medical Sciences, Mashhad, Iran. <sup>9</sup>Division of Medical Education, Brighton and Sussex Medical School, Brighton, UK. <sup>10</sup>Nafiseh Hosseini and Sara Saffar Soflaei have equally contributed as first authors. ✉email: Ghayourmobarhan@yahoo.com; GhayourM@mums.ac.ir; MouhebatiM@mums.ac.ir

hemoglobin (MCH), red cell distribution width (RDW), platelet distribution width (PDW), mean platelet volume (MPV), mixed cell (MXD), and platelet count. Previous studies have shown that CBC components are associated with certain cardiac diseases<sup>11</sup>, such as coronary heart disease<sup>12</sup>, left ventricular hypertrophy<sup>13</sup>, and endothelial dysfunction<sup>14,15</sup>. In addition, it has been reported that PVC is associated with ventricular dysfunction<sup>16,17</sup>. Also, it has been shown that CBC components may be prognostic in patients with ventricular dysfunction<sup>18,19</sup> and predict ventricular dysfunction<sup>18–21</sup>.

Given these findings, it is proposed that some hematological factors may also be associated with PVC. Two meta-analyses showed that hematological factors could predict the occurrence and recurrence of atrial fibrillation<sup>22,23</sup>. Moreover, hematocrit was found to be associated with ventricular arrhythmias in patients with chronic renal failure on dialysis<sup>24</sup>. Bindra et al. investigated hemoglobin levels in patients with acute coronary syndrome (ACS)<sup>25</sup>. Their findings indicated that individuals with reduced Hb concentrations were predisposed to ventricular arrhythmias. Similarly, another study indicated that ventricular arrhythmia is significantly more prevalent among patients with higher Hb levels<sup>26</sup>. Recently, Arienzo et al. proposed a framework for detection PVC based on explainable AI (XAI). They used GRADient-weighted Class Activation Mapping algorithm and achieved high accuracy (96.21%)<sup>27</sup>. Also, Shen et al. proposed a model to detect PVC based on the network dynamic features. In this method each heartbeat was divided into several segments, and their statistical properties were calculated to construct sequences. Subsequently, network topology-related features were extracted from the constructed sequences and input into a decision tree-based Gentleboost. F1-measure of their method was about 97%<sup>28</sup>. However, limited research has been undertaken on the relationship between arrhythmia and CBC components, and no large-scale study has explored the potential association between hematological factors and PVC. Thus, the purpose of this research is to investigate the potential association between hematological factors and PVC using a data mining method.

## Materials and methods

### Participants and blood sampling

The 9,035 subjects were recruited in the MASHAD study. A stratified cluster random sampling technique was used to select participants aged between 35 and 65 years old. Participants who chose not to remain in the study and those who did not return for testing were excluded<sup>29</sup>. Also, individuals which did not have complete CBC component or unreadable (ECG) were excluded for this study. Participants were divided into two datasets based on their gender. After an overnight period of fasting, blood samples of 20 milliliters were collected from each participant involved in the study. The collection was performed through antecubital vein venipuncture following standardized protocols, with subjects in a seated position in an EDTA-coated tube. Hematological parameters were assessed by using the Sysmex auto analyzer system KX-21 N in whole blood samples. Complete blood count (CBC) components, including white blood cell (WBC), red blood cell (RBC), hemoglobin (Hb), hematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), Corpuscular hemoglobin concentration (MCHC), platelet (PLT), lymphocytes count (LYM), mixed cell count (MXD), Neutrophil count (NEUT), red cell distribution width (RDW), platelet distribution width (PDW), and mean platelet volume (MPV) were measured. The research protocol was approved by the School of Medicine, Mashhad University of Medical Sciences, Biomedical Research Ethics Committee (IR.MUMS.MEDICAL.REC.1399.783).

### Statistical analysis and learning models

Quantitative and qualitative variables were delineated through the reporting of mean  $\pm$  standard deviation (SD) and frequency (%) measures, respectively. Participants were randomly assigned to test and training datasets with proportions of 20% and 80%, respectively. Moreover, because the data was unbalanced a mixed sampling approach was utilized which uses under-sampling and over-sampling methods that make synthetic minority class samples and balance the data<sup>30</sup>. To detect the relationship between hematological factors and PVC three machine learning (ML) models, including Logistic regression (LR), C5.0, and boosting decision tree (DT) were used. Area under the curve (AUC), accuracy, precision, F1-measure, and specificity were utilized to evaluate the performance of each model on the test dataset.

All analyses were made using SPSS version 26 and SPSS Modeler 10. Significant cut-off values were considered at 0.05.

#### Logistic regression

Logistic regression (LR) is a model that is applied to data with binary outcomes and multiple variables<sup>31</sup>. The formula below represents the mathematical utilization of the LR model, in which  $Y_i$ ,  $X$ , and  $\beta$  denote the response variable (zero or one), vectors of covariates, and regression coefficient.

#### Decision tree model and boosting

In the domain of big data analytics, typified by complex underlying frameworks and interrelationships, ML algorithms have emerged as effective methodologies for uncovering latent associations among variables and consequent outcomes<sup>32</sup>. Classification is a primary application of ML algorithms, aimed at discerning novel and intricate connections between predictors and outcomes, demonstrating notable efficiency in this regard<sup>33</sup>. DT is one of the main classification algorithms that simply, classify the data based on the different values of variables. It comprises different elements: 1) A root node, which classifies the data into different main subsets. 2) The internal nodes, symbolize a feasible vertex within the hierarchical arrangement, linking to the root node at the apex and the leaf nodes at the end. 3) The leaf nodes, which represent the terminus of a tree's hierarchical structure, illustrate the conclusive outcomes of the process of partitioning records into distinct target groups<sup>34</sup>. Choosing the factors and split data are done by the Information gain parameter which is based on the reduction of entropy of factors.

$$IG(T, a) = H(T) - H(T|a).$$

where  $H(T|a)$  is the conditional entropy of  $T$  given the value of attribute  $a$ .

Boosting operates through the sequential construction of multiple models. Initially, a base model is developed conventionally. Then, subsequent models are constructed, emphasizing instances misclassified by the preceding model. This iterative process continues, with each successive model addressing the errors of its predecessor. Ultimately, cases are classified by applying the collective ensemble of models, employing a weighted voting mechanism to incorporate individual predictions into a unified outcome. Although boosting can markedly enhance the accuracy of a C5.0 DT model, it necessitates extended training durations<sup>35</sup>.

## Results

### Baseline characteristics

Of the participants who were recruited into this study 3,615, and 5,420 participants were male and female, respectively. A total of 42 males and 66 females were found to be experiencing PVC. The baseline characteristics of males and females are shown in Table 1. As expected, baseline characteristics of this study showed that WBC, RBC, Hb, HCT, and MXD are slightly higher among males in comparison to females. Since these baseline differences might affect the models, they have been constructed separately for each gender group.

To investigate the feasible relationship between PVC and hematological factors, three machine learning methods have been utilized. These methods are the LR, DT (C5.0), and boosting DT. For this purpose, the dataset was randomly split into training (80%) and test (20%) datasets. Models were trained on the training dataset and subsequently evaluated on the test dataset. The evaluation was repeated 10 times, and the average scores were presented.

### LR model

According to the multivariate LR model, the variables that were significantly associated with PVC ( $P$  value  $< 0.001$ ) among males and females were indicated in Table 2. The WBC, MCHC, RDW, and Hb revealed the highest OR among variables in the LR model for males. However, in the LR model of female participants, the Hb and LYM were the most remarkable variables. To evaluate the models' performances, LR models were tested on the test dataset. The LR model for males performed accuracy, specificity, and sensitivity of 62.91%, 28.47%, and 93.09% respectively (Table 3). LR model of female subjects indicated accuracy, specificity, and sensitivity of 61.5%, 17.04%, and 93.61% respectively (Table 4). Moreover, according to the ROC curve, LR models presented the lowest Area under the curve (AUC) in comparison to the other models (Fig. 1).

### C5.0 DT model and boosting

Various hematological factors were evaluated using the DT algorithm, leading to their classification into 10 layers. The most important variable within the first layer often denotes the most significant contributor to data categorization, with subsequent variables arranged in descending order of importance. Also, boosting has been used to improve the accuracy of results. This model proceeds by focusing on correcting errors in each step. The results of C5.0 were better than the LR model for both males and females (Tables 3 and 4). Based on the final tree among males, WBC, PLT, RDW, PDW, and HCT had the greatest importance in PVC presence (Fig. 2 and

Variables*	Male		Female	
	With PVC (42)	Without PVC (3573)	With PVC (66)	Without PVC (5354)
WBC ( $\times 10^3/\mu\text{l}$ )	$6.72 \pm 1.79$	$6.21 \pm 1.66$	$5.87 \pm 1.52$	$5.97 \pm 1.46$
RBC ( $\times 10^6/\mu\text{l}$ )	$5.20 \pm 0.42$	$5.14 \pm 0.46$	$4.73 \pm 0.38$	$4.66 \pm 0.40$
Hb (mg/dl)	$15.04 \pm 1.31$	$14.79 \pm 1.32$	$13.28 \pm 1.37$	$13.01 \pm 1.69$
HCT (%)	$44.25 \pm 3.32$	$43.94 \pm 3.17$	$39.85 \pm 3.27$	$39.40 \pm 3.34$
MCV (fl.)	$84.09 \pm 9.63$	$85.51 \pm 5.61$	$84.39 \pm 5.29$	$84.58 \pm 6.27$
MCH (pg)	$29.01 \pm 2.24$	$28.89 \pm 2.59$	$28.10 \pm 2.48$	$27.99 \pm 2.67$
MCHC (g/dl)	$33.971 \pm 1.23$	$33.64 \pm 1.86$	$33.25 \pm 1.43$	$32.97 \pm 1.51$
PLT ( $\times 10^3/\mu\text{l}$ )	$208.67 \pm 42.51$	$211.97 \pm 53.05$	$225.81 \pm 55.28$	$241.26 \pm 63.65$
LYM ( $\times 10^3/\mu\text{l}$ )	$2.307 \pm 0.82$	$2.24 \pm 2.01$	$2.22 \pm 0.74$	$2.17 \pm 1.03$
MXD ( $\times 10^3/\mu\text{l}$ )	$0.71 \pm 0.36$	$0.66 \pm 0.338$	$0.47 \pm 0.20$	$0.58 \pm 0.40$
NEUT ( $\times 10^3/\mu\text{l}$ )	$3.71 \pm 1.2$	$3.36 \pm 1.33$	$3.24 \pm 1.07$	$3.32 \pm 3.18$
RDW (%)	$14.45 \pm 3.64$	$13.71 \pm 2.66$	$13.71 \pm 1.28$	$13.92 \pm 2.21$
PDW (%)	$13.25 \pm 1.83$	$12.69 \pm 1.95$	$13.02 \pm 2.08$	$12.83 \pm 2.04$
MPV (fl.)	$10.23 \pm 0.89$	$10.09 \pm 3.04$	$10.20 \pm 0.98$	$10.09 \pm 1.05$

**Table 1.** Baseline characteristics of male and female. PVC, premature ventricular contraction; WBC, white blood cell; RBC, red blood cell; Hb, hemoglobin; HCT, hematocrit; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; PLT, platelet; LYM, lymphocyte; MXD, mixed; NEUT, neutrophil; RDW, Red blood cell distribution width; PDW, platelet distribution width; MPV, mean platelet volume. \* Expressed as mean  $\pm$  standard deviation.

Variables	OR (CI 95%)	P-value
Male		
WBC	1.42 (1.31,1.54 )	>0.0001
Hb	1.14 (1.04,1.24)	>0.003
MCV	1.02 (1,1.05)	>0.019
MCHC	1.52 (1.36,1.7)	>0.0001
LYM	0.723 (0.63,0.84)	>0.0001
RDW	1.27 (1.18,1.37)	>0.0001
PDW	1.14 (1.09,1.2)	>0.0001
Female		
Hb	34.48 (14.63,81.22)	>0.0001
HCT	0.34 (0.26,0.45)	>0.0001
MCH	0.87(0.82,0.91)	>0.0001
MCHC	0.34 (0.24,0.47)	>0.0001
PLT	0.996 (0.995,0.998)	>0.0001
LYM	1.26 (1.1,1.41)	>0.0001
MXD	0.13 (0.09,0.19)	>0.0001

**Table 2.** Variables associated with PVC in males and females. PVC, premature ventricular contraction; OR, odds ratio; WBC, white blood cell; RBC, red blood cell; Hb, hemoglobin; HCT, hematocrit; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; PLT, platelet; LYM, lymphocyte; MXD, mixed; RDW, Red blood cell distribution width; PDW, platelet distribution width.

Method	Accuracy (%)	Specificity (%)	Sensitivity (Recall) (%)	Precision (%)	F1-measure (%)
LR	62.91	28.47	93.09	73.32	82.03
DT(C5.0)	93.82	86.39	97.26	93.92	95.56
DT (Boosting)	<b>98.13</b>	<b>94.41</b>	<b>100</b>	<b>97.26</b>	<b>98.61</b>

**Table 3.** Evaluation parameters for male. LR, logistic regression; DT, decision tree.

Method	Accuracy (%)	Specificity (%)	Sensitivity (Recall) (%)	Precision (%)	F1-measure (%)
LR	61.5	17.04	93.61	71.05	80.78
DT (C5.0)	94.66	89.34	97.2	95.3	96.1
DT (Boosting)	<b>96.92</b>	<b>91.34</b>	<b>99.46</b>	<b>96.19</b>	<b>97.8</b>

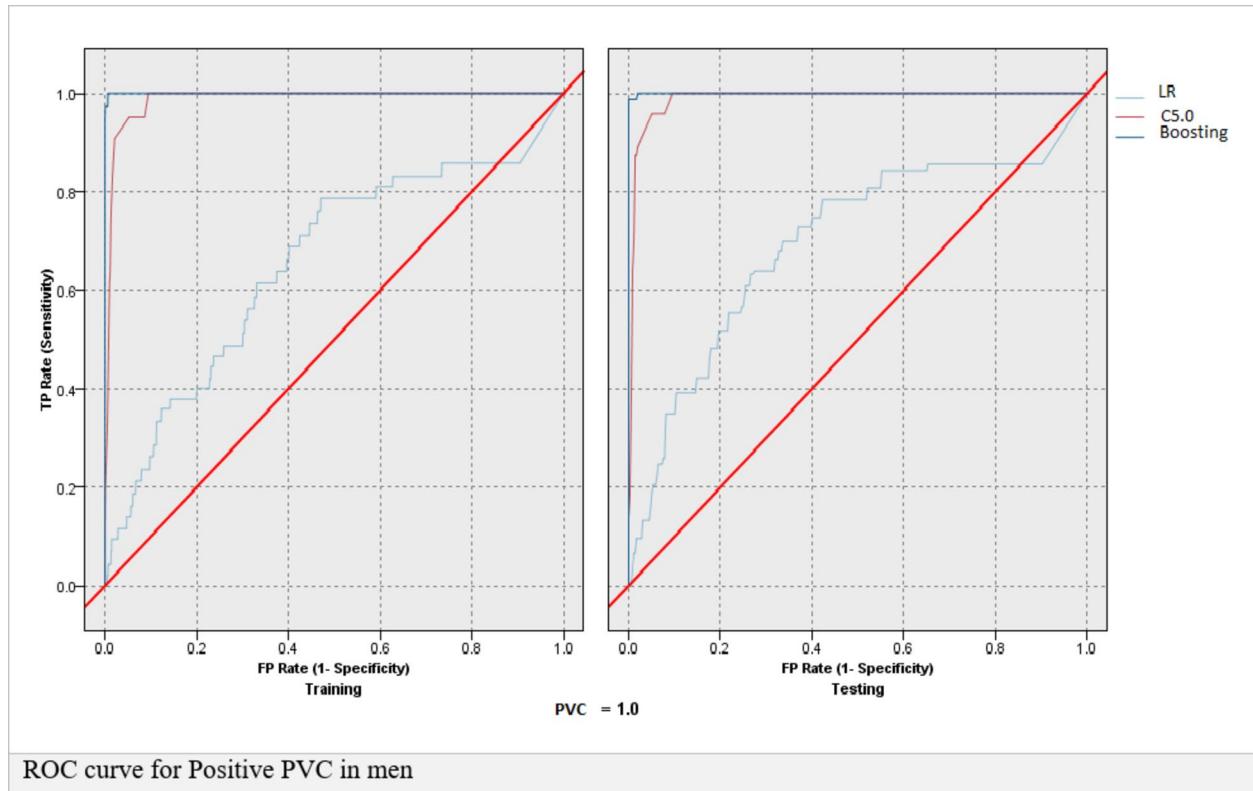
**Table 4.** Evaluation parameters for females. LR, logistic regression; DT, decision tree.

Supplement 1). Also, among females, the DT model showed that PVC presence mostly was impacted by RBC, PLT, RDW, MCV, and MXD (Fig. 3 and Supplement 2). According to the confusion matrices, the boosting model for males demonstrated the accuracy, specificity, and sensitivity of 98.13%, 94.41%, and 100% respectively (Table 3). Also, this model, performed an accuracy of 96.92%, a specificity of 91.34%, and a sensitivity of 99.46% for females (Table 4). The results of DT (C5.0) with boosting were the best performance in comparison to other models for both males and females. Additionally, the AUC of the ROC curve for boosting DT models was higher in comparison to LR and C5.0 models (Fig. 1). Also, some of the rules and upper layers of the DT model are shown in Figs. 2 and 3. The complete final Trees were shown in Supplement 1 and 2.

## Discussion

It has been shown previously that PVC may be associated with cardiac dysfunction, and it has been reported that frequent PVC could be a result of congestive heart failure<sup>36</sup>. Also, previous studies have brought up that PVC could potentially be associated with a higher risk of sudden cardiac death<sup>37–39</sup>. Hence, it is essential to identify factors that may correlate with PVC. Accordingly, this study employs three ML algorithms to explore the potential relationship between hematological factors, readily accessible in clinical settings, and the occurrence of PVC among a cohort comprising 9,035 participants.

Comparison among the models indicated that the boosting DT algorithm exhibited the most favorable performance. Conversely, the LR algorithm displayed lower AUC, sensitivity, specificity, and accuracy. The C5.0 model revealed high results but using boosting improved the final results. Based on the final DT the most



**Fig. 1.** ROC curve of models.

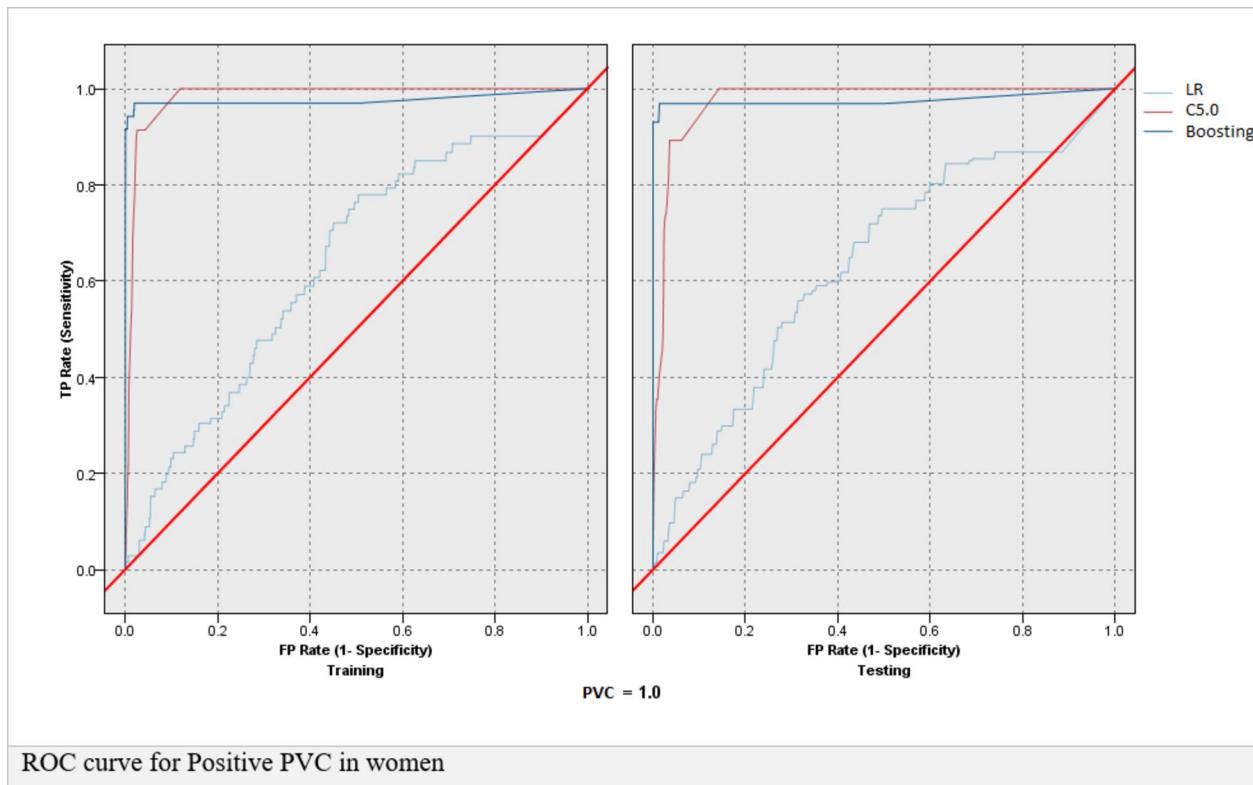
important hematology factors associated with PVC were WBC, PLT, RDW, PDW, and HCT for males and RBC, PLT, RDW, MCV, and MXD for females.

One of the hematological factors that was associated with the presence of PVC was RDW. However, there is no evident direct relationship between RDW and PVC, but according to previous studies some medical conditions might cause the co-occurrence of higher RDW and PVC. It has been shown that PVC could be associated with heart diseases such as coronary artery disease (CAD)<sup>16</sup>, structural heart disease<sup>40</sup>, myocarditis<sup>41</sup>, and arrhythmias<sup>42</sup>. In addition, it has been shown that RDW could be associated with an increased risk of different heart diseases. For instance, Tonelli et al. found that elevated levels of RDW are related to the risk of death and CVD events among patients with a history of myocardial infarction<sup>43</sup>, which might trigger PVC. Moreover, another study revealed that higher RDW is also related to a higher probability of heart failure which is one of the underlying causes of PVC<sup>44,45</sup>. Moreover, one of the major causes of higher RDW is iron deficiency, which is found to be associated with altered electrophysiological activity<sup>46</sup>.

One of the other hematological factors found to be associated with PVC was platelet count (PLT). Higher PLT may be associated with medical conditions that potentially contribute to PVC. It has been found that myocardial inflammation can lead to PVC<sup>47</sup>. In addition, this is evident that PLT is elevated during inflammation<sup>48</sup>. Moreover, beyond these findings, previous studies have demonstrated a direct relationship between PLT elevation and CVDs, which are the main risk factors for frequent PVC, as was mentioned earlier<sup>49–52</sup>. Moreover, it has been demonstrated that PLT is directly associated with cardiac arrhythmia through PLT product release which can potentially impact the cardiac cells' activities<sup>53</sup>.

According to the models, higher WBC was also associated with a higher risk of PVC. Previous studies have shown that higher WBC and neutrophil are associated with hypertrophy among patients with hypertensive blood pressure<sup>54</sup>. This statement could be one of the plausible explanations for this observation since ventricular hypertrophy is related to PVC occurrence<sup>55,56</sup>. Moreover, it has been revealed that elevated WBC is remarkably related to a higher incidence of coronary heart disease, and CVD occurrence which are tangibly related to PVC<sup>16,57–59</sup>. Moreover, studies have shown that higher WBC is an independent risk factor for coronary heart disease, which is consistent with the observation in this study<sup>12,60</sup>.

This study revealed a significant association between MCV and PVC. MCV, commonly regarded as a surrogate marker for anemia, exhibits a noteworthy correlation with PVC. Anemia has been implicated in the pathogenesis of ventricular hypertrophy, suggesting a potential long-term impact on cardiovascular health<sup>61</sup>. In concordance with this research finding, a previous study has reported that the B6 deficiency, which is caused by the lower values of MCV<sup>62</sup>, is significantly related to cardiac changes among rats, including PVC<sup>63</sup>. In parallel with the previously mentioned parameters of CBC, MCV demonstrates a relation with CVD, a connection that extends to PVC<sup>64–66</sup>.



**Figure 1.** (continued)

In addition to previous hematological factors, PDW may be associated with heart failure and cardiac events, such as PVC. A previous study has shown that PDW could be correlated with adverse prognostic events among heart failure patients<sup>67</sup>. Given this finding, previous studies have shown that the PVC could be associated with heart failure<sup>6,68</sup>. Moreover, HCT is one of the other factors that were found to be associated with PVC. Similarly, previous studies have shown that heart failure and left ventricular dysfunction are also associated with lower HCT<sup>69</sup>. It has been found that coronary heart disease (CHD), is one of the underlying causes of PVC occurrence<sup>70</sup>, was related to abnormal HCT levels<sup>71</sup>. Moreover, other similar scenarios could be plausible regarding the relationship between PVC and hematological factors, such as RBC, and MXD. It has been revealed that the PVCs could be related to high blood pressure. Regarding this finding, the RBC count potentially can influence the viscosity and other hemodynamic factors, which finally might lead to high blood pressure<sup>72</sup>. Similarly, higher MXD is associated with higher cardiovascular risk<sup>73,74</sup>.

This study has several strengths. It was conducted using a large sample dataset, which improves the reliability of the models. Secondly, this study includes several hematological factors, providing a nuanced and comprehensive understanding of the feasible relationship between hematological factors and PVC. Thirdly, the performance evaluation of the DT (C5.) with boosting model in this study indicated robust favorable AUC, accuracy, specificity, precision, and F1-measure values, which promise possible clinical applications of the models.

In conclusion, this study investigated whether it was possible to predict PVC using routinely measured hematological factors through the utilization of advanced machine learning algorithms. The most important hematological factors associated with PVC in both males and females were RDW and PLT. For males WBC, PDW, and HCT, and females, RBC, MCV, and MXD were also important. The results are consistent with prior research findings that underscored their physiological relevance. However, more research is warranted to apply and validate these findings in other population groups, as this is the first study to evaluate the association between hematological factors and PVC as far as we know.

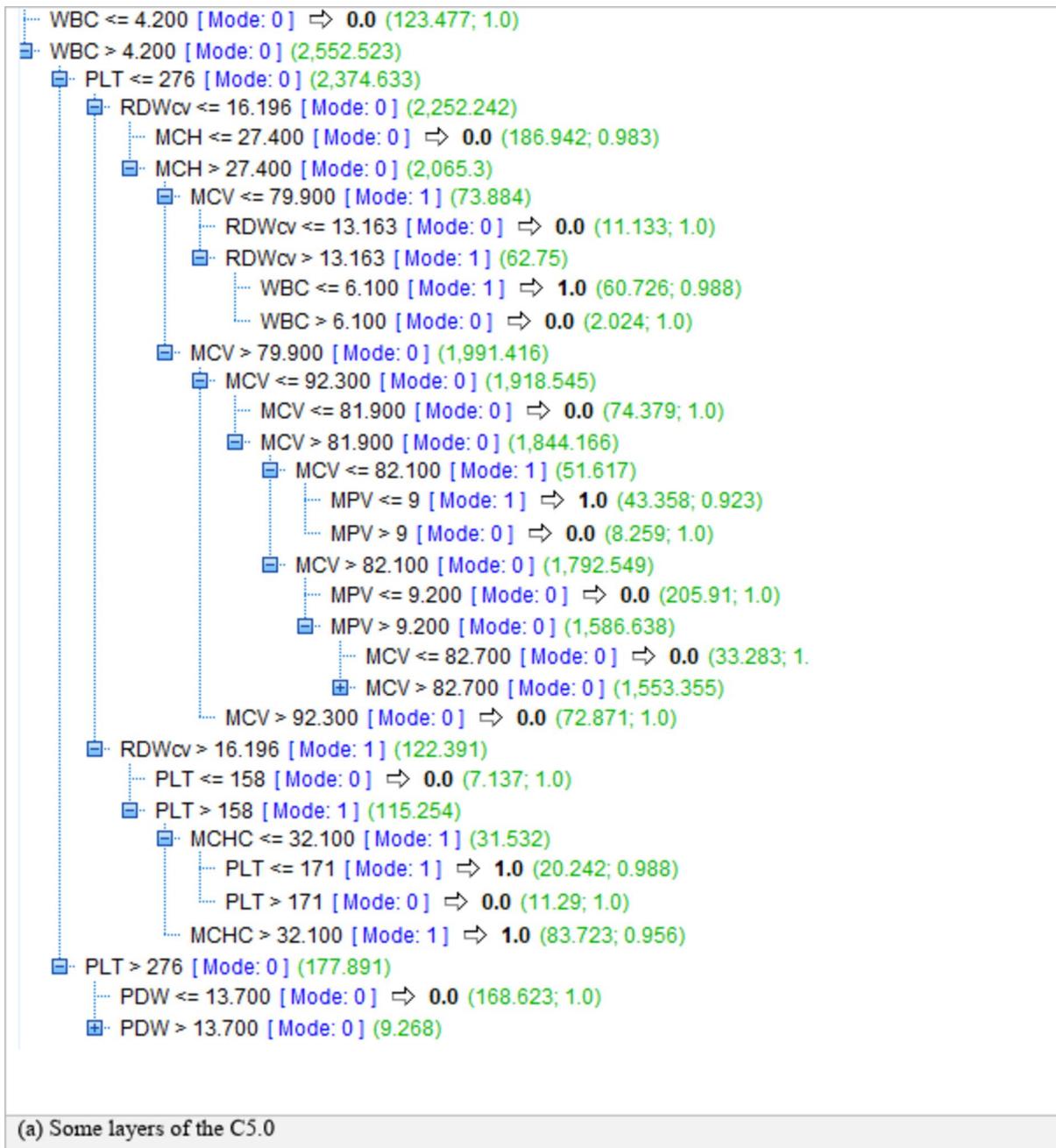
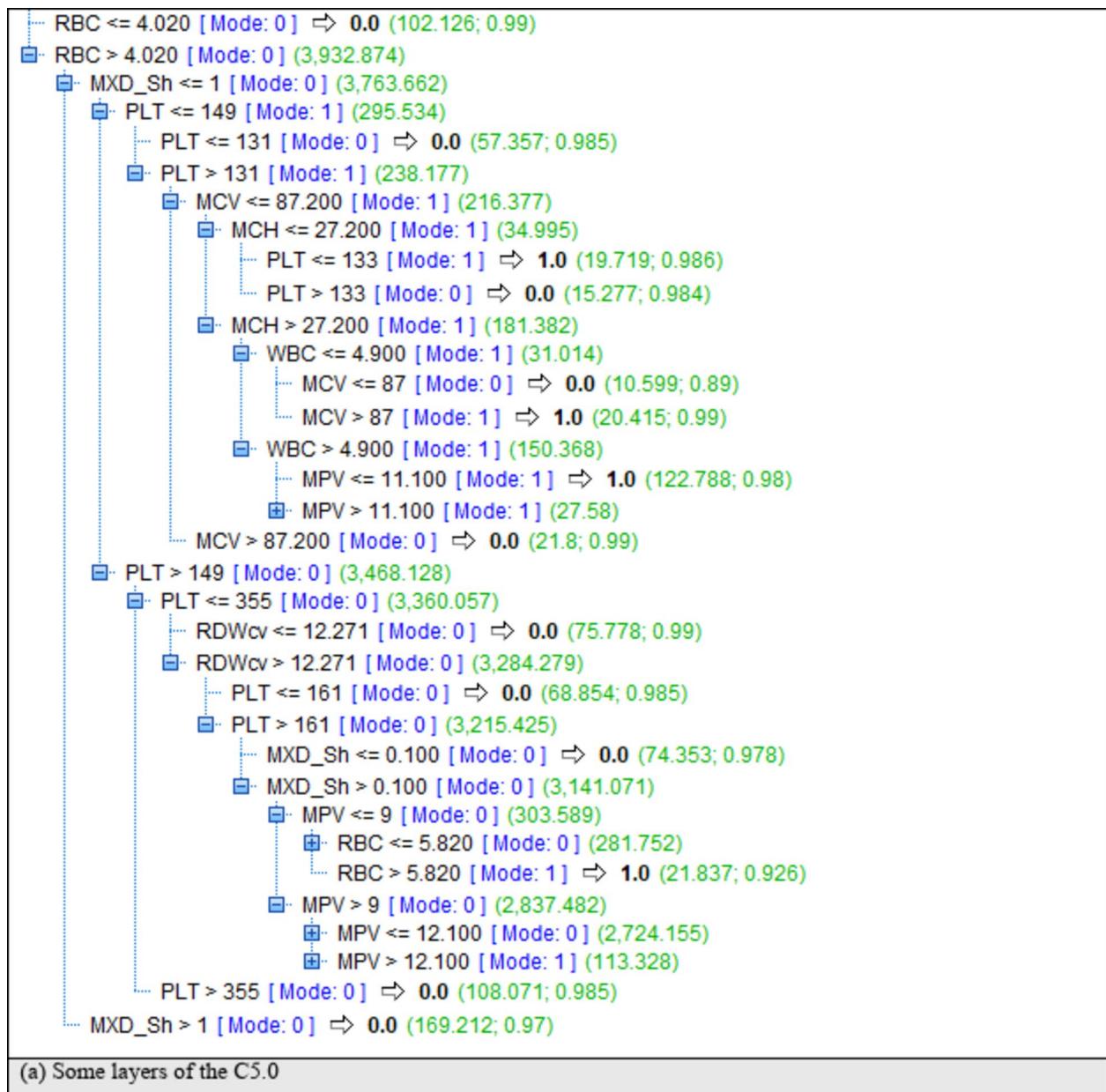


Fig. 2. Upper layers of DT and rules for positive and negative PVC for males.

<p>Rules for 1 - contains 20 rule(s)</p> <ul style="list-style-type: none"> <li>Rule 1 for 1.0           <ul style="list-style-type: none"> <li>if WBC &gt; 4.200</li> <li>and PLT &lt;= 276</li> <li>and RDWcv &lt;= 16.196</li> <li>and MCH &gt; 27.400</li> <li>and MCV &lt;= 79.900</li> <li>and RDWcv &gt; 13.163</li> <li>and WBC &lt;= 6.100</li> <li>then 1.000</li> </ul> </li> <li>Rule 2 for 1.0           <ul style="list-style-type: none"> <li>if WBC &gt; 4.200</li> <li>and PLT &lt;= 276</li> <li>and RDWcv &lt;= 16.196</li> <li>and MCH &gt; 27.400</li> <li>and MCV &gt; 79.900</li> <li>and MCV &lt;= 92.300</li> <li>and MCV &gt; 81.900</li> <li>and MCV &gt; 82.100</li> <li>and MPV &gt; 9.200</li> <li>and MCV &gt; 82.700</li> <li>and NEUT_Sh &lt;= 4.100</li> <li>and MCHC &gt; 33.200</li> <li>and MXD_Sh &lt;= 0.700</li> <li>and MCH &lt;= 31</li> <li>and PLT &lt;= 149</li> <li>and RBC &gt; 5.020</li> <li>and MCV &lt;= 86.100</li> <li>and WBC &lt;= 5.100</li> <li>and RDWcv &lt;= 12.437</li> <li>then 1.000</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Rule 1 for 0.0           <ul style="list-style-type: none"> <li>if WBC &lt;= 4.200</li> <li>then 0.000</li> </ul> </li> <li>Rule 2 for 0.0           <ul style="list-style-type: none"> <li>if WBC &gt; 4.200</li> <li>and PLT &lt;= 276</li> <li>and RDWcv &lt;= 16.196</li> <li>and MCH &lt;= 27.400</li> <li>then 0.000</li> </ul> </li> <li>Rule 3 for 0.0           <ul style="list-style-type: none"> <li>if WBC &gt; 4.200</li> <li>and PLT &lt;= 276</li> <li>and RDWcv &lt;= 16.196</li> <li>and MCH &gt; 27.400</li> <li>and MCV &lt;= 79.900</li> <li>and RDWcv &lt;= 13.163</li> <li>then 0.000</li> </ul> </li> <li>Rule 4 for 0.0           <ul style="list-style-type: none"> <li>if WBC &gt; 4.200</li> <li>and PLT &lt;= 276</li> <li>and RDWcv &lt;= 16.196</li> <li>and MCH &gt; 27.400</li> <li>and MCV &lt;= 79.900</li> <li>and RDWcv &gt; 13.163</li> <li>and WBC &gt; 6.100</li> <li>then 0.000</li> </ul> </li> </ul>
(b) Some rules for predicting PVC+.	(c) Some rules for predicting PVC- .

Figure 2. (continued)



**Fig. 3.** Upper layers of DT and rules for positive and negative PVC for females.

<p>Rules for 1 - contains 24 rule(s)</p> <ul style="list-style-type: none"> <li>Rule 1 for 1.0           <ul style="list-style-type: none"> <li>if RBC &gt; 4.020</li> <li>and MXD_Sh &lt;= 1</li> <li>and PLT &lt;= 149</li> <li>and PLT &gt; 131</li> <li>and MCV &lt;= 87.200</li> <li>and MCH &lt;= 27.200</li> <li>and PLT &lt;= 133</li> <li>then 1.000</li> </ul> </li> <li>Rule 2 for 1.0           <ul style="list-style-type: none"> <li>if RBC &gt; 4.020</li> <li>and MXD_Sh &lt;= 1</li> <li>and PLT &lt;= 149</li> <li>and PLT &gt; 131</li> <li>and MCV &lt;= 87.200</li> <li>and MCH &gt; 27.200</li> <li>and WBC &lt;= 4.900</li> <li>and MCV &gt; 87</li> <li>then 1.000</li> </ul> </li> </ul>	<p>Rules for 0 - contains 56 rule(s)</p> <ul style="list-style-type: none"> <li>Rule 1 for 0.0           <ul style="list-style-type: none"> <li>if RBC &lt;= 4.020</li> <li>then 0.000</li> </ul> </li> <li>Rule 2 for 0.0           <ul style="list-style-type: none"> <li>if RBC &gt; 4.020</li> <li>and MXD_Sh &lt;= 1</li> <li>and PLT &lt;= 149</li> <li>and PLT &gt; 131</li> <li>then 0.000</li> </ul> </li> <li>Rule 3 for 0.0           <ul style="list-style-type: none"> <li>if RBC &gt; 4.020</li> <li>and MXD_Sh &lt;= 1</li> <li>and PLT &lt;= 149</li> <li>and PLT &gt; 131</li> <li>and MCV &lt;= 87.200</li> <li>and MCH &lt;= 27.200</li> <li>and PLT &gt; 133</li> <li>then 0.000</li> </ul> </li> </ul>
(b) Some rules for predicting PVC+.	(c) Some rules for predicting PVC- .

Figure 3. (continued)

### Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to university data ownership policies but are available from the corresponding author upon reasonable request.

Received: 3 July 2024; Accepted: 13 January 2025

Published online: 20 January 2025

### References

1. Lee, G. K. et al. Premature ventricular contraction-induced cardiomyopathy: a treatable condition. *Circ. Arrhythm. Electrophysiol.* **5** (1), 229–236 (2012).
2. Emilsson, K. Suspected association of ventricular arrhythmia with air pollution in a motorbike rider: a case report. *J. Med. Case Rep.* **2**, 192 (2008).
3. Farzam, K. & Richards, J. R. *Premature Ventricular Contraction*, in *StatPearls*. StatPearls Publishing Copyright © 2023, StatPearls Publishing LLC.: Treasure Island (FL). (2023).
4. Birnie, D. H. et al. HRS expert consensus statement on the diagnosis and management of arrhythmias associated with cardiac sarcoidosis. *Heart Rhythm*. **11** (7), 1305–1323 (2014).
5. Duffee, D. F., Shen, W. K. & Smith, H. C. Suppression of frequent premature ventricular contractions and improvement of left ventricular function in patients with presumed idiopathic dilated cardiomyopathy. *Mayo Clin. Proc.* **73** (5), 430–433 (1998).
6. Dukes, J. W. et al. Ventricular ectopy as a predictor of heart failure and death. *J. Am. Coll. Cardiol.* **66** (2), 101–109 (2015).
7. Huizar, J. F. et al. Outcomes of premature ventricular contraction-cardiomyopathy in the veteran Population: a secondary analysis of the CHF-STAT study. *JACC Clin. Electrophysiol.* **7** (3), 380–390 (2021).
8. Kennedy, H. L. et al. Long-term follow-up of asymptomatic healthy subjects with frequent and complex ventricular ectopy. *N. Engl. J. Med.* **312** (4), 193–197 (1985).
9. Soflaei, S. S. et al. A large population-based study on the prevalence of electrocardiographic abnormalities: a result of Mashhad stroke and heart atherosclerotic disorder cohort study. *Ann. Noninvasive Electrocardiol.* **28**(6). (2023).
10. Sheldon, S. H., Gard, J. J. & Asirvatham, S. J. Premature ventricular contractions and non-sustained ventricular Tachycardia: Association with Sudden Cardiac Death, Risk Stratification, and management strategies. *Indian Pacing Electrophysiol. J.* **10** (8), 357–371 (2010).
11. Agnello, L. et al. The value of a complete blood Count (CBC) for Sepsis diagnosis and prognosis. *Diagnostics (Basel)*. **11**(10). (2021).
12. Brown, D. W., Giles, W. H. & Croft, J. B. White blood cell count: an independent predictor of coronary heart disease mortality among a national cohort. *J. Clin. Epidemiol.* **54** (3), 316–322 (2001).
13. Amin, M. G. et al. Hematocrit and left ventricular mass: the Framingham Heart study. *J. Am. Coll. Cardiol.* **43** (7), 1276–1282 (2004).
14. Solak, Y. et al. Mean corpuscular volume is associated with endothelial dysfunction and predicts composite cardiovascular events in patients with chronic kidney disease. *Nephrology* **18** (11), 728–735 (2013).
15. Wu, T. H. et al. Gradient relationship between increased Mean Corpuscular volume and Mortality Associated with cerebral ischemic stroke and ischemic heart disease: a longitudinal study on 66,294 Taiwanese. *Sci. Rep.* **8** (1), 16517 (2018).
16. Dev Sharma, S., Ballantyne, F. & Goldstein, S. The relationship of ventricular asynergy in coronary artery disease to ventricular premature beats. *CHEST* **66** (4), 358–362 (1974).
17. Niwano, S. et al. Prognostic significance of frequent premature ventricular contractions originating from the ventricular outflow tract in patients with normal left ventricular function. *Heart* **95** (15), 1230–1237 (2009).

18. Cooper, H. A. et al. White blood cell count and mortality in patients with ischemic and nonischemic left ventricular systolic dysfunction (an analysis of the studies of Left Ventricular dysfunction [SOLVD]). *Am. J. Cardiol.* **84** (3), 252–257 (1999).
19. Li, X. et al. Predictors of early left ventricular dysfunction after mitral valve replacement for rheumatic valvular disease. *J. Card Surg.* **34** (11), 1185–1193 (2019).
20. Çakır, M. O. The Prognostic significance of Neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio for long-term survival in patients with severe left ventricular dysfunction and Implantable Cardioverter Defibrillator. *Cureus* **15** (10), e47441 (2023).
21. Ferreira, A. I. et al. Prognostic impact of red blood cell distribution width in chronic heart failure patients with left ventricular dysfunction. *J. Cardiovasc. Med. (Hagerstown)* **24** (10), 746–751 (2023).
22. Weymann, A. et al. Prediction of New-Onset and Recurrent Atrial Fibrillation by Complete Blood Count tests: a comprehensive systematic review with Meta-analysis. *Med. Sci. Monit. Basic. Res.* **23**, 179–222 (2017).
23. Weymann, A. et al. Haematological indices as predictors of atrial fibrillation following isolated coronary artery bypass grafting, valvular surgery, or combined procedures: a systematic review with meta-analysis. *Kardiol Pol.* **76** (1), 107–118 (2018).
24. Bonato, F. O. et al. Prevalence of ventricular arrhythmia and its associated factors in nondialyzed chronic kidney disease patients. *PLoS One* **8** (6), e66036 (2013).
25. Mozos, I. Mechanisms linking Red Blood Cell disorders and Cardiovascular diseases. *Biomed. Res. Int.* **2015**, 682054 (2015).
26. SAYGI, S. et al. Ventricular arrhythmia in dialysis patients: a link with higher hemoglobin levels? *Hemodial. Int.* **15** (2), 250–255 (2011).
27. Arienzo, G. et al. *CardioView: a framework for detection Premature Ventricular Contractions with eXplainable Artificial Intelligence*. in *INI-DH 2024: Workshop on Innovative Interfaces in Digital Healthcare, in conjunction with International Conference on Advanced Visual Interfaces*. (2024).
28. Shen, Y. et al. *Efficient Premature Ventricular Contraction Detection Based on Network Dynamics Features* (IEEE Transactions on Instrumentation and Measurement, 2024).
29. Ghayour-Mobarhan, M. et al. Mashhad stroke and heart atherosclerotic disorder (MASHAD) study: design, baseline characteristics and 10-year cardiovascular risk estimation. *Int. J. Public. Health.* **60**, 561–572 (2015).
30. Lusa, L. Improved shrunken centroid classifiers for high-dimensional class-imbalanced data. *BMC Bioinform.* **14**, 64 (2013).
31. Hosmer, D. W. *Applied logistic regression*, 2. (2000).
32. Saberi-Karimian, M. et al. Potential value and impact of data mining and machine learning in clinical diagnostics. *Crit. Rev. Clin. Lab. Sci.* **58** (4), 275–296 (2021).
33. Zhong, Y. *The analysis of cases based on decision tree*. in *7th IEEE international conference on software engineering and service science (ICSESS)*. 2016. IEEE. (2016).
34. Song, Y. Y. & Ying, L. Decision tree methods: applications for classification and prediction. *Shanghai Archives Psychiatry.* **27** (2), 130 (2015).
35. Pasbasi, E., Ozen, M. & Aydin, N. *Improving medical diagnosis reliability using Boosted C5.0 decision tree empowered by Particle Swarm Optimization*. Annu Int Conf IEEE Eng Med Biol Soc, 2015: pp. 7230–3. (2015).
36. Chakkro, C. S. & Gheorghiade, M. Ventricular arrhythmias in severe heart failure: incidence, significance, and effectiveness of antiarrhythmic therapy. *Am. Heart J.* **109** (3), 497–504 (1985).
37. Wang, Y. et al. Cellular mechanism of premature ventricular contraction-induced cardiomyopathy. *Heart Rhythm.* **11** (11), 2064–2072 (2014).
38. Morrow, J. P. High-fat diet, obesity and sudden cardiac death. *Acta Physiol.* **211** (1), 13–16 (2014).
39. FISHER, F. D. & TYROLER, H. A. Relationship between ventricular premature contractions on routine Electrocardiography and subsequent sudden death from Coronary Heart Disease. *Circulation* **47** (4), 712–719 (1973).
40. Saurav, A. et al. Premature ventricular contraction-Induced Cardiomyopathy. *Clin. Cardiol.* **38** (4), 251–258 (2015).
41. Lakkireddy, D. et al. Myocarditis causing premature ventricular contractions. *Circulation: Arrhythmia Electrophysiol.* **12** (12), e007520 (2019).
42. Hatanaka, K. et al. The Association between the pattern of premature ventricular contractions and Heart diseases: Assessment of Routine Electrocardiography in Health examinations. *J. Occup. Health.* **44** (5), 343–347 (2003).
43. Tonelli, M. et al. Relation between red blood cell distribution Width and Cardiovascular Event Rate in People with Coronary Disease. *Circulation* **117** (2), 163–168 (2008).
44. Makhoul, B. F. et al. Relation between changes in red cell distribution width and clinical outcomes in acute decompensated heart failure. *Int. J. Cardiol.* **167** (4), 1412–1416 (2013).
45. Y, S., Chandregowda & Cp, D. G. Relationship between red cell distribution width and clinical outcome in patients with Acute Coronary Syndrome. *J. Assoc. Physicians India.* **70** (4), 11–12 (2022).
46. Goldstein, D. et al. Experimental Iron Deficiency in rats: mechanical and electrophysiological alterations in the Cardiac muscle. *Clin. Sci.* **91** (2), 233–239 (1996).
47. Hansen, P. R. Inflammatory alterations in the myocardial microcirculation. *J. Mol. Cell. Cardiol.* **30** (12), 2555–2559 (1998).
48. van der Meijden, P. E. J. & Heemskerk, J. W. M. Platelet biology and functions: new concepts and clinical perspectives. *Nat. Reviews Cardiol.* **16** (3), 166–179 (2019).
49. Elwood, P. C. et al. Ischemic heart disease and platelet aggregation. The Caerphilly Collaborative Heart Disease Study. *Circulation* **83** (1), 38–44 (1991).
50. Thaulow, E. et al. Blood platelet count and function are related to total and cardiovascular death in apparently healthy men. *Circulation* **84** (2), 613–617 (1991).
51. Willoughby, S., Holmes, A. & Loscalzo, J. Platelets and Cardiovascular Disease. *Eur. J. Cardiovasc. Nurs.* **1** (4), 273–288 (2002).
52. McManus, D. D. & Freedman, J. E. MicroRNAs in platelet function and cardiovascular disease. *Nat. Reviews Cardiol.* **12** (12), 711–717 (2015).
53. De Jong, J. S. & Dekker, L. R. *Platelets and Cardiac Arrhythmia*. *Front. Physiol.*, 1. (2010).
54. Shi, H. et al. Association of white blood cell counts with left ventricular mass index in hypertensive patients undergoing anti-hypertensive drug therapy. *Exp. Ther. Med.* **13** (4), 1566–1571 (2017).
55. Schmieder, R. E. & Messerli, F. H. Ventricular arrhythmia and sudden cardiac death: the significance of left ventricular hypertrophy as risk factor]. *Schweiz. Med. Wochenschr.* **123** (4), 99–107 (1993).
56. Bayés-Genís, A. et al. Cardiac arrhythmias and left ventricular hypertrophy in systemic hypertension and their influences on prognosis. *Am. J. Cardiol.* **76** (13), 54D–59D (1995).
57. Abdalla, I. S. H. et al. Relation between ventricular premature complexes and sudden cardiac death in apparently healthy men. *Am. J. Cardiol.* **60** (13), 1036–1042 (1987).
58. Lee, C. D. et al. White Blood Cell Count and incidence of Coronary Heart Disease and ischemic stroke and mortality from Cardiovascular Disease in African-American and White men and women: atherosclerosis risk in communities Study. *Am. J. Epidemiol.* **154** (8), 758–764 (2001).
59. Simpson, R. J. et al. Prevalence of premature ventricular contractions in a population of African American and white men and women: the atherosclerosis risk in communities (ARIC) study. *Am. Heart J.* **143** (3), 535–540 (2002).
60. Weijenberg, M. P., Feskens, E. J. M. & Kromhout, D. *White Blood Cell Count and the Risk of Coronary Heart Disease and All-Cause Mortality in Elderly Men*. Arteriosclerosis, Thrombosis, and Vascular Biology, 16(4): pp. 499–503. (1996).

61. Olivetti, G. et al. Myocyte cellular hypertrophy and hyperplasia contribute to ventricular wall remodeling in anemia-induced cardiac hypertrophy in rats. *Am. J. Pathol.* **141** (1), 227–239 (1992).
62. Massey, A. C. Microcytic anemia. Differential diagnosis and management of iron deficiency anemia. *Med. Clin. North. Am.* **76** (3), 549–566 (1992).
63. Mulvaney, D. A. & Seronde, J. J. Electrocardiographic changes in vitamin B61 deficient rats. *Cardiovascular. Res.* **13** (9), 506–513 (1979).
64. Peverill, R. E., Harper, R. W. & Smolich, J. J. Inverse relation of haematocrit to cardiac index in mitral stenosis and atrial fibrillation. *Int. J. Cardiol.* **71** (2), 149–155 (1999).
65. Hsieh, Y. P. et al. *Mean Corpulcular volume and mortality in patients with CKD*. *Clin. J. Am. Soc. Nephrol.* **12**(2). (2017).
66. Zhan, Y. et al. Multiplicative interaction between mean corpulcular volume and red cell distribution width with target organ damage in hypertensive patients. *J. Clin. Lab. Anal.* **31** (5), e22082 (2017).
67. Sato, Y. et al. Association between platelet distribution width and prognosis in patients with heart failure. *PLOS ONE.* **15** (12), e0244608 (2021).
68. Kim, Y. G. et al. Premature ventricular contraction increases the risk of heart failure and ventricular tachyarrhythmia. *Sci. Rep.* **11** (1), 12698 (2021).
69. Das, S. R. et al. Relation of Lower Hematocrit to Progression from Asymptomatic Left Ventricular Dysfunction to Symptomatic Heart failure (from the studies of Left Ventricular Dysfunction Prevention Trial). *Am. J. Cardiol.* **96** (6), 827–831 (2005).
70. Bet, B. B. et al. Implications of fetal premature atrial contractions: systematic review. *Ultrasound Obstet. Gynecol.* **60** (6), 721–730 (2022).
71. Ji-shen, C. *The Analysis of Association between the Hematocrit Level and Coronary Heart Disease* (Prevention and Treatment of Cardio-Cerebral-Vascular Disease, 2013).
72. Askin, L., Cetin, M. & Turkmen, S. Ambulatory blood pressure results and heart rate variability in patients with premature ventricular contractions. *Clin. Exp. Hypertens.* **40** (3), 251–256 (2018).
73. Seo, I. H. & Lee, Y. J. Usefulness of Complete Blood Count (CBC) to assess Cardiovascular and metabolic diseases in clinical settings: a Comprehensive Literature Review. *Biomedicines* **10** (11), 2697 (2022).
74. Lassale, C. et al. Elements of the complete blood count associated with cardiovascular disease incidence: findings from the EPIC-NL cohort study. *Sci. Rep.* **8** (1), 3290 (2018).

## Acknowledgements

The support provided by Mashhad University of Medical Sciences (MUMS) to conduct this study is highly acknowledged.

## Author contributions

Study concept and design: MM and MGM; Data collection: SSS, MYK, BSH, and HR; data analysis and interpretation of data: NSH, HE; drafting of the manuscript: PSS, NSH, and SSS; supervision and critical revision: GAF, MGM, and MM. All authors have approved the final draft.

## Funding

This work was supported by the Vice-chancellor of Research of Mashhad University of Medical Sciences.

## Declarations

### Competing interests

The authors declare no competing interests.

### Ethics approval and consent to participate

All experiments were performed by the declaration of Helsinki and Mashhad University of Medical Sciences ethical guidelines and regulations. The research protocol was approved by the School of Medicine, Mashhad University of Medical Sciences, Biomedical Research Ethics Committee (IR.MUMS.MEDICAL.REC.1399.783). Also, informed consent was obtained from all subjects or their legal guardians.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-86557-z>.

**Correspondence** and requests for materials should be addressed to M.M. or M.G.-M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.