# Interpretable machine learning for cardiogram-based biometrics

Ilija Tanasković[a,b]*
Ljiljana B. Lazarević[c]
Goran Knežević[d]
Nikola Milosavljević[e]
Olga Dubljević[f]
Bojana Bjegojević[g]
Nadica Miljković[a,h]

[a]University of Belgrade – School of Electrical Engineering, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia
[b]Institute for Artificial Intelligence Research and Development of Serbia, Fruškogorska 1, 21000 Novi Sad, Serbia
[c]Institute of Psychology and LIRA lab, Faculty of Philosophy, University of Belgrade, Čika Ljubina 18-20, 11000 Belgrade, Serbia
[d]Department of Psychology and LIRA lab, Faculty of Philosophy, University of Belgrade, Čika Ljubina 18-20, 11000 Belgrade, Serbia
[e]University of Belgrade – Faculty of Special Education and Rehabilitation, Visokog Stevana 2, 11000 Belgrade, Serbia
[f]Institute for Biological Research "Siniša Stanković", National Institute of the Republic of Serbia, University of Belgrade, Bulevar Despota Stefana 142, 11000 Belgrade, Serbia
[g]Technological University Dublin, Human Factors in Safety & Sustainability Research Group, Grangegorman Lower, Dublin 7, D07 H6K8, Ireland
[h]Faculty of Electrical Engineering, University of Ljubljana, Tržaška cesta 25, 1000 Ljubljana, Slovenia

**Abstract**

This study investigates the role of electrocardiogram (ECG) and impedance cardiogram (ICG) features in biometric identification, emphasizing their discriminative capacity and robustness to emotional variability. A total of 29 features spanning four domains (temporal, amplitude, slope, and morphological) are evaluated using random forest (RF) models combined with multiple interpretability methods. Feature importance shows that both ECG- and ICG-derived features are consistently ranked among the top 10 by Gini importance, permutation importance, and SHAP values, with ECG features, particularly QRS-centric descriptors, occupying the highest positions. In parallel, ICG BCX features contribute complementary, however, with lower cross-method stability. Correlation analysis reveals substantial

*Corresponding Author: Tel: +381 11 3218 348, E-mail: ilija.tanaskovic@ivi.ac.rs (Ilija Tanasković)
Other e-mail addresses: ljiljana.lazarevic@f.bg.ac.rs (Ljiljana B. Lazarević), gknezevi@f.bg.ac.rs (Goran Knežević), nikola.milosavljevic@f.bg.ac.rs (Nikola Milosavljević), olga.dubljevic@ibiss.bg.ac.rs (Olga Dubljević), bojana.bjegojevic@TUDublin.ie (Bojana Bjegojević), nadica.miljkovic@etf.bg.ac.rs (Nadica Miljković)

multicollinearity, where the RF distributes and diminishes importance across highly correlated pairs, confirming reduced independent contributions. Statistical analysis identifies 14 features with significant differences between baseline and anger, without a clear pattern by domain. Feature selection with recursive feature elimination and genetic algorithms converges on a subset (12 features) that attains accuracy within 1% of the full set (99%), improving efficiency in storage and computation. Proposed complementary analyses indicate that the individuality is primarily encoded in the QRS-related ECG features across all four domains. Meanwhile, BCX-derived ICG features contribute mainly through amplitude, providing supportive but less stable discriminatory cues. The confirmed resilience of QRS-centric descriptors to emotional variation, where stable inter-individual differences in the QRS complex could be traced to variations in ventricular mass, conduction pathways, and thoracic geometry, may indicate their central role in future models of cardiogram-based identity.

**Keywords:** biometric identification, electrocardiogram, feature importance, feature selection, impedance cardiography, random forest

## 1. Introduction

Biometrics leverages measurable physiological and behavioral characteristics to establish and verify the individual's identity [1]. While fingerprints and iris patterns remain standard, recent advances highlight the value of biomedical signals, which are harder to steal or forge. Among these, the electrocardiogram (ECG) has emerged as a promising modality, due to its inherent liveliness detection capability [2–4]. Additionally, multimodal approaches, combining ECG with other biomedical signals, have attracted increasing interest [3]. For example, the impedance cardiogram (ICG), which captures the changes in thoracic impedance linked to the heart mechanical activity [5,6], could provide complementary information to the ECG for identification purposes [7,8].

Machine learning (ML) has further advanced the field of biometric identification, consistently achieving high accuracy with cardiogram-based systems [2,3,7]. However, the limited explainability of ML models remains a challenge, as research is

not focused on identifying discriminative contribution of specific ECG and ICG features. This gap restricts insight into the physiological mechanisms that underlie identity-specific signatures in both healthy and patients.

Furthermore, cardiogram-based biometrics face several real-world challenges, including inconsistency over time and susceptibility to emotional states [3]. As the autonomic nervous system modulates cardiac function, emotional arousal alters ECG and ICG morphology [3], with emotions, such as anger causing pronounced drops in identification accuracy, as demonstrated in our previous work [8], whereas others, such as anxiety or stress, exert a relatively small impact [9,10].

### 1.1 Research questions

Our previous studies [8,11] demonstrated that a multimodal approach utilizing fiducial-based features extracted from both ECG and ICG signals achieved the highest biometric identification accuracy, with statistically significant results. However, those efforts focused mainly on predictive performance, without examining how individual features contribute to individuality.

In this study, we therefore adopt an exploratory approach [12], with aim is to identify which properties of ECG and ICG carry individual-specific information, including heartbeat amplitude, waveform morphology, and slope, or the variability of heart intervals across subjects. Accordingly, the study is guided by the following research questions:

1. What part of the cardiogram signals contains the essential information for biometric identification?
2. How does correlation among features influence their importance in ML models, such as random forests (RF)?
3. Which features are the most affected by emotional state changes, and do particular feature types (*e.g.,* time, amplitude, slope, or morphology) exhibit consistent sensitivity to emotional variations?

## 2. Materials and methods

We use a publicly available dataset of simultaneously recorded ECG and ICG signals from 202 healthy young adults [8,13], with full details on the dataset, recording setup, and acquisition procedures provided in the Appendix. The visual representation of the proposed methodology is shown in Figure 1. The pipeline begins with preprocessing to reduce noise, followed by delineation to identify fiducial points, *i.e.*, characteristic landmarks tied to key cardiovascular events. These two steps are adopted from our previous work [8,11]. An RF is firstly trained and evaluated using all 29 features, which serves as the base model and benchmark, with identification accuracy used as the reference metric for all subsequent comparisons.

Several techniques are then applied in parallel to identify a specific subset of features that are crucial for identification. Firstly, feature importance is estimated with three methods (Gini importance, permutation feature importance, and shapley additive explanations - SHAP values), from which the top 10 features are taken and intersected to produce the first subset, providing a stable consensus ranking. Secondly, feature selection is performed using recursive feature elimination with cross-validation (RFECV) and a genetic algorithm (GA), each producing a data-driven subset without manual thresholds. The intersection of these two subsets defines the second subset, which emphasizes selection stability. Thirdly, correlation analysis is conducted using Pearson and Spearman coefficients to identify clusters of highly correlated features, which are then used to assess cluster-level contribution and to examine how multicollinearity influences importance estimation. Finally, statistical testing across emotional states identifies features that change significantly, and this inference is complemented by training the RF on three subsets: all features, the non-significant set, and the significant set, to compare predictive behavior. All evaluations are conducted on independent test set, which has not participated in the training process. The outcomes from these four paths are synthesized to localize where identity-related information resides in ECG and ICG and to characterize how emotional variability shapes those signals. To support clarity and ease of navigation, the detailed steps of the proposed pipeline are provided in the Appendix.

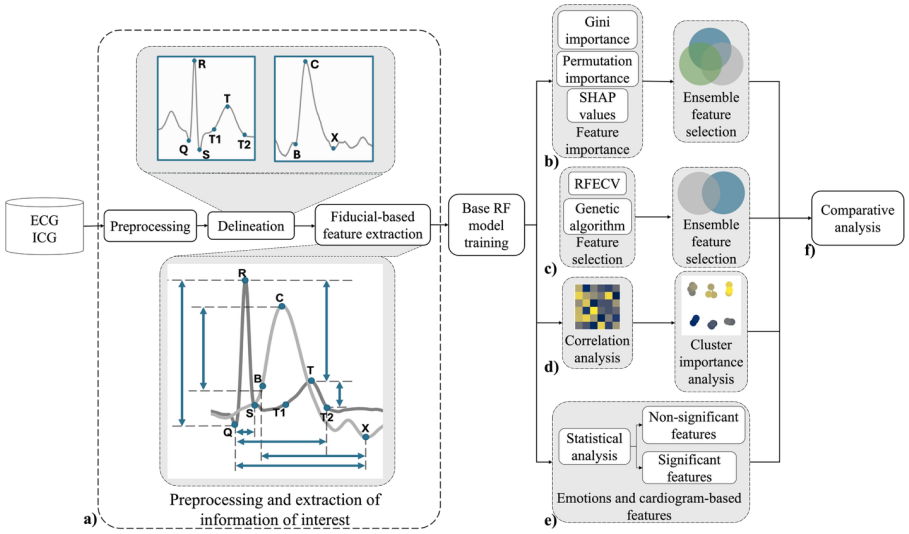


*Figure 1: Block diagram of the proposed methodology: a) preprocessing and extraction of information of interest; b) feature importance; c) feature selection; d) correlation and cluster-importance analysis; e) emotions and cardiogram-based features; f) comparative analysis. electrocardiogram - ECG, impedance cardiogram - ICG, random forest - RF, shapley additive explanations - SHAP, recursive feature elimination with cross-validation - RFECV.*

## 3. Results

Feature cross-correlations are visualized as heatmaps, with significant associations ($p < 0.05$) marked by an asterisk. As Pearson and Spearman matrices show nearly identical patterns, only Pearson coefficients ($r$) are presented in Figure 2, while the Spearman matrix ($\rho$) is provided in Figure S1 of the Supplementary Materials [14]. Highly correlated pairs ($|r|$ or $|\rho| > 0.7$) are summarized in Table 1. All reported correlations are significant ($p < 0.001$) and every pair identified by Spearman is also captured by Pearson correlation coefficient. Therefore, subsequent analysis relies on Pearson correlations. Using a graph search over the adjacency list of these pairs, seven clusters of highly correlated features are identified for further analysis.

### 3.1 Feature importance findings

The RF model achieves 99.17% accuracy, 99.16% F1 score, 99.33% precision, and 99.17% recall on the test set using all features from Table A1, described in the Appendix. These results are referred to as the base model performance in subsequent

comparisons. Detailed feature importances are provided in Figures S2-S4 in the Supplementary Materials [14].

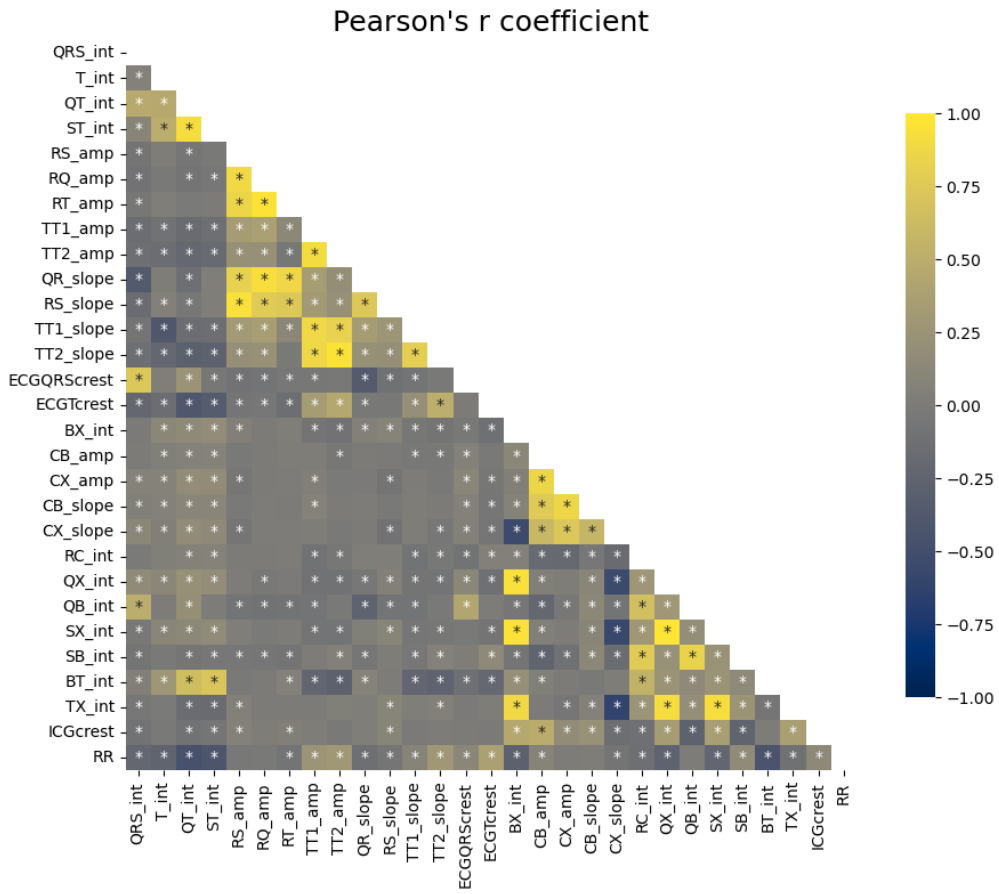


*Figure 2: Heatmap of Pearson correlation coefficients between extracted features, with asterisks (*) indicating statistical significance.*

Using an ensemble feature-selection strategy, the intersection of the three top 10 lists yields a consensus subset of 8 features, visualized as a Venn diagram in Figure 3, consisting of ECGQRScrest, RS_amp, RS_slope, RT_amp, TT2_amp, QRS_int, RQ_amp, and QR_slope. Retraining the RF model with this subset results in 96.92% accuracy, 96.91% F1 score, 96.46% precision, and 96.92% recall.

### 3.2 Correlation and cluster analysis outcomes

The effect of multicollinearity on feature importance is summarized in Table 2 which reports the seven identified feature clusters and the corresponding accuracy changes, reflecting their collective contribution to classification, visualized in Figure 4. For each feature (Variable 1), a permutation is applied to quantify the relative change in the importance of its correlated pair (Variable 2), while the resulting changes in importance across all other features are summarized using descriptive statistics.

*Table 1: Pairs of correlated features, denoted as Feature 1 and Feature 2, with the corresponding Pearson and Spearman correlation coefficients; the Method column indicates whether the high correlation ($|r|$ or $|\rho| > 0.7$) is identified by Pearson only, Spearman only, or by both. All correlations are statistically significant with $p < 0.001$.*

| Feature 1 | Feature 2 | Pearson | Spearman | Method |
|---|---|---|---|---|
| QX_int | SX_int | 0.981 | 0.971 | Both |
| RQ_amp | RT_amp | 0.958 | 0.949 | Both |
| TT2_amp | TT2_slope | 0.954 | 0.964 | Both |
| BX_int | SX_int | 0.953 | 0.954 | Both |
| RS_amp | RS_slope | 0.943 | 0.942 | Both |
| BX_int | QX_int | 0.939 | 0.937 | Both |
| SX_int | TX_int | 0.926 | 0.902 | Both |
| RQ_amp | QR_slope | 0.922 | 0.918 | Both |
| QT_int | ST_int | 0.918 | 0.846 | Both |
| QX_int | TX_int | 0.907 | 0.873 | Both |
| TT1_amp | TT2_amp | 0.898 | 0.903 | Both |
| RS_amp | RQ_amp | 0.891 | 0.866 | Both |
| TT1_amp | TT1_slope | 0.881 | 0.874 | Both |
| BX_int | TX_int | 0.876 | 0.857 | Both |
| TT1_amp | TT2_slope | 0.874 | 0.888 | Both |
| RT_amp | QR_slope | 0.871 | 0.858 | Both |
| CX_amp | CB_slope | 0.865 | 0.903 | Both |
| RS_amp | RT_amp | 0.857 | 0.833 | Both |
| CB_amp | CX_amp | 0.852 | 0.835 | Both |

| Feature 1 | Feature 2 | Pearson | Spearman | Method |
|---|---|---|---|---|
| QB_int | SB_int | 0.830 | 0.710 | Both |
| TT2_amp | TT1_slope | 0.821 | 0.830 | Both |
| RS_amp | QR_slope | 0.816 | 0.793 | Both |
| TT1_slope | TT2_slope | 0.792 | 0.799 | Both |
| RC_int | SB_int | 0.770 | 0.271 | Pearson only |
| RQ_amp | RS_slope | 0.770 | 0.748 | Both |
| CB_amp | CB_slope | 0.764 | 0.817 | Both |
| QR_slope | RS_slope | 0.739 | 0.721 | Both |
| RT_amp | RS_slope | 0.735 | 0.714 | Both |
| QRS_int | ECGQRScrest | 0.734 | 0.391 | Pearson only |
| CX_amp | CX_slope | 0.731 | 0.676 | Pearson only |
| ST_int | BT_int | 0.712 | 0.649 | Pearson only |

### 3.3 Feature selection performance

RFECV selected a subset of 15 features, while the GA approach selected 17 features. Both methods achieved very high performance, with RFECV reaching 99.00% accuracy and F1-score, 99.18% precision, and 99.00% recall, and GA achieving 98.58% accuracy, 98.56% F1-score, 99.81% precision, and 98.58% recall. Despite differences in the specific subsets, there was substantial overlap, with 12 features consistently selected by both methods. The visualization of the overlapping and unique features is shown in Figure 5. The RF model is retrained using the 12 intersected features, achieving 98.17% accuracy, 98.15% F1-score, 98.45% precision, and 98.15% recall, which is comparable to the base model, despite reducing the feature set by more than a half.

### 3.4 Effect of emotions on cardiogram-based features

Statistical analysis reveals that 14 features (QRS_int, T_int, QT_int, ST_int, TT1_amp, TT1_slope, BX_int, CB_amp, CB_slope, RC_int, QX_int, BT_int, ICGcrest, and RR) differ significantly between the baseline and anger segments. The remaining 15 features, such as RS_amp, RQ_amp, RT_amp, TT2_amp, QR_slope, RS_slope, TT2_slope, ECGQRScrest, ECGTcrest, CX_amp, CX_slope, QB_int,

SX_int, SB_int, and TX_int, do not exhibit significant changes across emotional states. Detailed results of statistical analysis are presented in Table S1 in the Supplementary Materials [14]. The effect sizes are negligible for all features except T_int, which demonstrated a small effect. The highly correlated pairs, which show opposite significance between segments, as shown in Table 3.

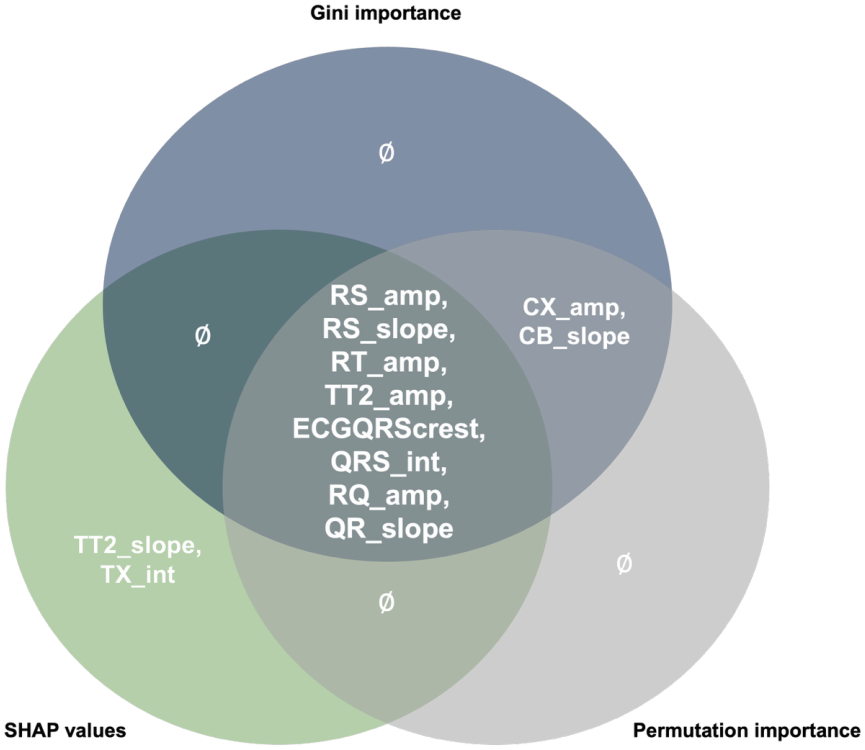


*Figure 3: Venn diagram showing the overlap among the top 10 features identified by Gini, permutation importance, and SHAP methods.*

The predictive evaluation of features, based on sensitivity to emotions, is summarized in Table 4. RF models are trained using three feature subsets: all features, only emotion-insensitive (non-significant) features, and only emotion-sensitive (significant) features. Utilizing all features, within-segment performance remains high (97–99%), while cross-segment generalization declines (88–89% baseline → anger; 91–92% anger → baseline). Using only emotion-insensitive features preserves strong within-segment accuracy (97–98%) but does not improve generalization (86–

89%). In contrast, emotion-sensitive features alone yield substantially lower performance (within-segment 89–90%; cross-segment 66–72%).

## 4. Discussion

The RF classifier achieves strong performance, with all evaluation metrics reaching approximately 99%, surpassing the results reported in our previous studies [8,11]. As the focus of this work is not ML optimization, but rather the analysis of feature relationships and their contributions to classification, this improvement is likely driven by several methodological differences. The number of samples per subject is increased to 18 in the present study, compared with 8 or 12 in earlier work, and the feature set is expanded from 17 to 29 features, inspired with Patro et al. (2022) [15]. In addition, only baseline and anger segments are used for feature extraction, with neutral segments excluded.

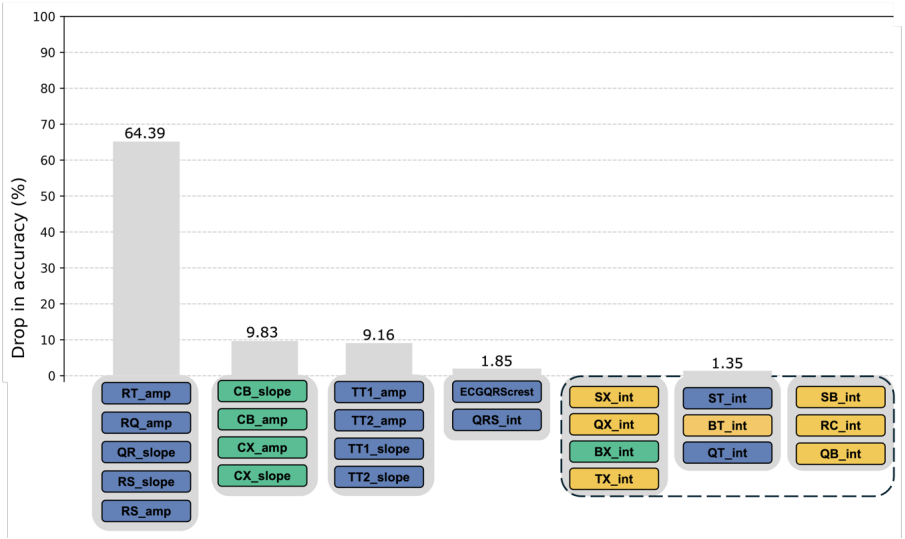


*Figure 4: Clusters of highly correlated features and the associated drop in identification accuracy (%) when each cluster is shuffled. Feature tiles are color-coded by signal of origin: dark blue = ECG only, green = ICG only, yellow = ECG\ICG. The final dashed group shows the cumulative effect of the remaining low-impact clusters.*

### 4.1 Interpretation of correlation and clustering patterns

The correlation analyses reveal substantial interdependencies among variables. Twenty-seven feature pairs exhibit a strong Spearman correlation coefficient, indicating robust monotonic associations insensitive to outliers. Pearson correlations confirm these findings with additional four pairs exceeding the 0.7 threshold in absolute value, suggesting that linear relationships predominate, but are complemented by monotonic non-linear patterns. Overall, these results emphasize the redundancy inherent in ECG- and ICG-derived features, reflecting the physiological interconnectedness of cardiac signals. Even with strong multicollinearity, RF performance does not deteriorate, consistent with its known robustness to correlated inputs [16].

Based on the results presented in Table 2, almost all correlated features (except for RC_int) showed an increase in their importance scores, and the rate of increase does not align with the strength of the correlation. This pattern is consistent with the known tendency of Gini importance to distribute importance across highly correlated variables [17], since trees in the ensemble select among them inconsistently. Cluster-level shuffling further indicates that QRS-related features (RQ_amp, RT_amp, QR_slope, RS_amp, RS_slope) exhibit the highest drop, as shown in Table 2 and Figure 3, aligning with prior studies highlighting the central role of QRS complex in decision-making [18,19], while ST_int, QT_int, and BT_int exhibit the lowest decrease in accuracy (only 0.17%), despite their physiological importance.

### 4.2 Insights from feature importance analysis

ECG-derived features, primarily those associated with the QRS complex, appear at the top of the rankings across all importance rankings (Figures S2-S4 in the Supplementary Materials [14]). Repolarization-related descriptors such as RT_amp and TT2_amp also appear among the top features, indicating that T-wave morphology contributes additional individuality, though with lower stability than QRS-based metrics. This supports evidence that ECG alone provides strong discrimination [2–4,9,20], while the present results further indicate that the QRS complex plays a central role in the decision-making process. ICG-derived descriptors, particularly CX_amp

and CB_slope, also recur across methods, indicating that the BCX complex of the ICG provides meaningful but less stable information compared with ECG features, supporting prior findings on its biometric potential [7,8,11]. This lower stability requires further investigation, as it may be influenced by highly correlated ICG features that exhibit opposite sensitivity to emotional variation, as shown in Table 3. Retraining the RF model using common subset yields performance only ~2% lower than the full model with ~72% reduction in feature space, raising a practical question of balancing feature set reduction with the potential gains in storage efficiency and computational cost.

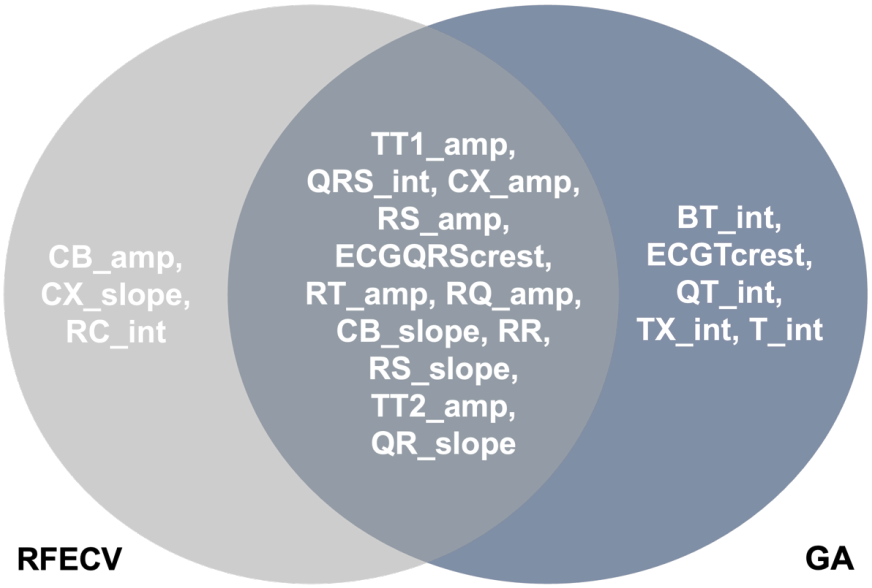


*Figure 5: Venn diagram illustrating the overlap of features selected by recursive feature elimination with cross-validation (RFECV) and genetic algorithm (GA).*

### 4.3 Evaluation of feature selection strategies

Feature selection using RFECV results in a reduced subset of 15 features while maintaining the evaluation metrics (~99%) compared to the base model. On the other hand, the GA approach selects 17 features and achieves similar performance. Two methods converge on 12 common features, shown in Figure 5, and this intersection fully overlaps with the 8 features consistently identified across all importance metrics.

This further reinforces the central role of QRS-derived features in driving identification performance, as well as complementary information within T wave. At the same time, the retention of CX_amp and CB_slope among the 12 shared features indicates that BCX features also contribute complementary information, although their absence from the SHAP-derived subset suggests that their discriminative value is less consistent across methods. These 12 common features constitute only 42% of the original feature set yet achieve performance within ~1% of the base model, underscoring that dimensionality reduction can retain nearly all discriminative power while substantially lowering computational, storage, and energy costs.

*Table 2: Impact of multicollinearity on feature importance and identification accuracy.*

| Variable 1 | Variable 2 | Correlation coefficient | Relative change (%) | Relative change in importance score of remaining variables (%) | | | | | Decrease in accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | min | 25% | median | 75% | max | |
| RQ_amp | RT_amp | 0.96 | 16.71 | -5.81 | 2.46 | 5.66 | 7.04 | 16.62 | 64.39 |
| | QR_slope | 0.92 | 11.98 | | | | | | |
| | RS_amp | 0.89 | 10.92 | | | | | | |
| | RS_slope | 0.77 | 5.53 | | | | | | |
| QX_int | SX_int | 0.98 | 13.47 | -4.32 | -0.89 | 1.18 | 4.34 | 9.35 | 0.59 |
| | BX_int | 0.94 | 7.12 | | | | | | |
| | TX_int | 0.91 | 5.76 | | | | | | |
| TT2_slope | TT2_amp | 0.95 | 11.49 | -9.29 | 0.61 | 2.50 | 4.73 | 6.61 | 9.16 |
| | TT1_amp | 0.87 | 14.26 | | | | | | |
| | TT1_slope | 0.79 | 9.83 | | | | | | |
| CX_amp | CB_slope | 0.86 | 18.65 | -2.29 | 0.32 | 2.82 | 5.60 | 13.64 | 9.83 |
| | CB_amp | 0.85 | 17.68 | | | | | | |
| | CX_slope | 0.73 | 20.41 | | | | | | |
| ST_int | QT_int | 0.92 | 3.67 | -3.79 | 0.09 | 1.79 | 3.30 | 5.46 | 0.17 |
| | BT_int | 0.71 | 2.59 | | | | | | |
| SB_int | QB_int | 0.83 | 7.36 | -6.28 | -0.90 | 2.25 | 3.01 | 8.12 | 0.59 |
| | RC_int | 0.77 | -0.34 | | | | | | |
| QRS_int | ECGQRScrest | 0.73 | 6.79 | -4.47 | 0.35 | 3.62 | 6.99 | 17.00 | 1.85 |

Inter-individual variability in the selected features plausibly arises from stable anatomical, electrophysiological, and hemodynamic differences. QRS-centric ECG features primarily capture ventricular depolarization amplitude, synchrony, and conduction velocity. These measures depend on myocardial mass and geometry, Purkinje network distribution, thoracic anatomy (heart position, chest wall thickness, lung volume), and tissue conductivity, which differ among individuals and yield

unique waveform morphology [21,22]. Repolarization-related measures, which describe T-wave size and shape, reflect differences in ventricular recovery dynamics, while RR captures intrinsic sinus rate and autonomic balance, which indirectly influence morphology [5]. BCX descriptors index ventricular ejection [5,6,23], as their amplitude and slopes depend on contractility, ejection velocity, aortic compliance, valve motion, and thoracic impedance, all of which vary across individuals. Overall, QRS-based features dominate because they arise from relatively fixed structural and conduction properties, while BCX features provide subject-specific hemodynamic information, and T-wave measures contribute individuality through repolarization gradients.

*Table 3: Correlation and significance of feature pairs. The table shows the correlation coefficient between Variable 1 and Variable 2, as well as whether each variable is statistically significant.*

| Variable 1 | Variable 2 | Correlation | Variable 1 Significant | Variable 2 Significant |
|---|---|---|---|---|
| QX_int | SX_int | 0.98 | True | False |
| BX_int | SX_int | 0.95 | True | False |
| QX_int | TX_int | 0.91 | True | False |
| TT1_amp | TT2_amp | 0.90 | True | False |
| BX_int | TX_int | 0.88 | True | False |
| TT1_amp | TT2_slope | 0.87 | True | False |
| CX_amp | CB_slope | 0.87 | False | True |
| CB_amp | CX_amp | 0.85 | True | False |
| TT2_amp | TT1_slope | 0.82 | False | True |
| TT1_slope | TT2_slope | 0.79 | True | False |
| RC_int | SB_int | 0.77 | True | False |
| QRS_int | ECGQRScrest | 0.73 | True | False |

### 4.4 Implications of emotion-sensitive cardiogram features

Our previous study [8] shows that emotional mismatch between enrollment and evaluation can reduce identification accuracy by up to 13%. In the present analysis, 14 features appear as emotionally-sensitive, without a clear pattern regarding domain or signal of origin. Notably, the QRS-derived features (RQ_amp, RS_amp, QR_slope, RS_slope, ECGQRScrest), that are most influential for decision-making, show no

significant differences between emotional states, indicating that the core identity-relevant descriptors remain stable.

*Table 4:Random Forest performance across emotional segments using all features, non-significant features, and significant features. Cross-validation (CV) refers to training and evaluation within the same segment, while generalization (Gen) refers to training on one segment and testing on the other*

| Training/Evaluation | Setting | All features | | Non-significant | | Significant | |
|---|---|---|---|---|---|---|---|
| | | Acc (%) | F1 (%) | Acc (%) | F1 (%) | Acc (%) | F1 (%) |
| Baseline → Baseline | CV | 98.62 | 98.53 | 97.74 | 97.60 | 89.99 | 89.12 |
| Baseline →Anger | Gen | 89.44 | 88.16 | 87.73 | 86.31 | 72.11 | 69.43 |
| Anger → Anger | CV | 97.63 | 97.36 | 97.96 | 97.78 | 90.54 | 89.83 |
| Anger → Baseline | Gen | 92.13 | 91.22 | 89.00 | 87.54 | 69.58 | 66.82 |

Training on only emotion-insensitive features yields a further reduction in cross-emotion performance, indicating that removing emotion-sensitive features does not improve robustness. In contrast, models trained exclusively on emotion-sensitive features show larger degradation in performance, demonstrating that emotion-driven variability provides little subject-discriminative value. This suggests that emotionally-sensitive features carry limited information on individuality, which is expected given that identity-specific descriptors must remain stable across varying emotional states for a biometric system to function reliably. Although Bazett's correction is applied, several intervals still differ significantly, suggesting that additional normalization strategies may be useful. Interestingly, when using all or only emotion-insensitive features, cross-segment generalization is consistently better from anger to baseline than other way around which is rather unexpected given that baseline is typically considered as segment with no influence.

Beyond the stable QRS descriptors, T-wave and ICG BCX features are split between significant and non-significant sets, as shown in Table 3, which could cause their lower stability and smaller overall contribution to identification. Overall, the results indicate that not all features provide equivalent identity-specific information, and optimal feature choice depends on the intended goal: enhancing biometric performance or mitigating emotional effects.

### 4.5 Limitations of the study

Here, we recognize the limitations of the proposed study:

1. We plan to incorporate P-wave descriptors using delineators optimized for low-amplitude and abnormal morphologies [24], given the well-known challenges of reliable P-wave detection.
2. The 2-minute segments are too short for valid spectral HRV estimation. Guidelines recommend ≥ 1–2 minutes, preferably 5 minutes, of stationary data for LF/HF analysis [25]. Accordingly, frequency-domain HRV metrics are not extracted.
3. The dataset shows an unbalanced sex distribution reflecting the psychology-student population [26]. This imbalance is not corrected, as sex is not used for stratification, and potential sex-related differences in identification accuracy are not examined.
4. Future work explores alternative normalization strategies (*e.g.*, Israel et al., 2005[9]) to evaluate their effect on model performance.
5. The neutral condition is not used in the current analysis, although it may serve future investigations on cognitive-workload effects on cardiographic signals, as suggested by Pale et al. (2021) [27].

### 4.6 Contributions of the study

Taking into consideration all of the data provided in this study, we now revisit the initial research questions and present a summary of the findings in relation to each of them:

1. The most reliable information comes from features derived from the QRS complex. Across three feature importance methods, QRS features (ECGQRScrest, RQ_amp, RS_amp, QR_slope, RS_slope, QRS_int) rank consistently high. These patterns align with physiology: QRS amplitude and slope reflect stable structural and conduction properties, whereas BCX features capture inter-individual differences in ejection dynamics, aortic compliance, and valve timing.

2. The RF model inherently adjusts importance scores for correlated features, confirming their reduced individual contributions. The study shows that almost all highly correlated features (except RC_int) experienced an increase in importance scores upon shuffling, but the rate of increase does not always align with their correlation coefficients.
3. QRS-based features (except QRS_int) remain stable during anger, consistent with their negligible effect sizes. Predictive analysis mirrors this: using only emotion-insensitive features does not improve cross-emotion generalization (accuracy drops ~11%). Using only statistically significant features further reduces cross-emotion accuracy by up to ~23%, indicating that emotion-driven variance does not support reliable identification. Together, these findings show that QRS-derived features retain high identity specificity across emotional states, whereas emotion-sensitive features degrade generalization.

## 5. Conclusion

This study demonstrates that the most reliable identity signal resides in QRS-centric ECG features across temporal, amplitude, slope, and morphology, while ICG BCX features add complementary but less stable cues. Random forest analyses, correlation structure, and feature selection converge on a compact, high-performing subset, confirming that much of the discriminative content can be retained with fewer features. Emotional variability continues to challenge cross-state generalization, and removing emotion-sensitive variables alone does not yield systematic improvements. Multicollinearity analyses indicate that information is distributed across correlated groups rather than isolated in single predictors, guiding principled reduction strategies. Physiologically, the dominance of QRS amplitude and slope is consistent with stable inter-individual differences in ventricular mass, conduction pathways, and thoracic geometry. Taken together, these findings clarify where identity-related information is localized in cardiograms, centered on QRS features with supportive BCX contributions, and they help connect observable waveforms to stable inter-individual physiological variation.

**Authorship contribution statement**

**Ilija Tanasković:** Methodology, Software, Formal analysis, Visualization, Writing – Original Draft **Ljiljana B. Lazarević:** Conceptualization, Methodology, Data Curation, Writing - Review & Editing **Goran Knežević:** Conceptualization, Resources, Investigation, Writing - Review & Editing **Nikola Milosavljević:** Data Curation, Methodology, Investigation, Visualization, Writing - Review & Editing **Olga Dubljević:** Data Curation, Methodology, Investigation, Writing - Review & Editing **Bojana Bjegojević:** Data Curation, Methodology, Investigation, Writing - Review & Editing **Nadica Miljković:** Conceptualization, Methodology, Validation, Formal analysis, Writing - Review & Editing.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The Funder did not participate in any aspect of the study design, collection, analysis, and interpretation of data; in the manuscript preparation; or in the decision to submit the manuscript.

**Declaration of generative AI and AI-assisted technologies in the manuscript preparation process**

During the preparation of this work, the first author used GPT-5 (ChatGPT) to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

**Ethics statement**

All participants signed Informed Consents in accordance with the Helsinki Declaration and the Institutional Review Board from the Department of Psychology at the University of Belgrade approved the study (No. 2018-19) on December 5, 2018.

**Funding**

Nadica Miljković acknowledges the support from the Ministry of Science, Technological Development and Innovation of the Republic of Serbia [Grant No. 451-03-137/2025-03/200103]; Ljiljana B. Lazarević and Goran Knežević acknowledge the support from the Ministry of Science, Technological Development, and Innovation of the Republic of Serbia [Grant No. 451-03-66/2025-03/200163]; Nikola Milosavljević acknowledges the support from the Ministry of Science, Technological Development and Innovation of the Republic of Serbia [Grant No. 451-03-137/2025-03/200096]. The funder was not involved in the manuscript preparation and the decision to submit the manuscript. Also, the funder was not related to the study design, data collection, data analysis, or results interpretation.

## Appendix

To maintain readability, the full methodological description underlying the analyses, presented in Figure 1 is provided here in expanded form. This Appendix presents each step of the proposed pipeline in detail, including preprocessing, delineation, feature extraction, model training, feature-importance estimation, feature-selection procedures, correlation analysis, and statistical testing across emotional states. The structure follows the order of the corresponding subsections referenced in the Results and Discussion, enabling direct cross-navigation between the body of text and the methodological elaborations provided below, which include the dataset description, preprocessing and information extraction, decision-making using machine learning, feature selection procedures, and the analysis of emotion-related effects on cardiogram-based features.

### A1. Dataset description

Our publicly available dataset is utilized for the study, comprising ECG and ICG signals recorded from 202 healthy psychology students (37 males, 163 females, and 2 who preferred not to disclose their sex/gender; mean age = 20.18 ± 1.88 years, with 9 NA for age) [13]. ECG and ICG signals are acquired simultaneously at a sampling frequency of 2000 Hz. ECG signals are recorded using Lead II electrodes placed

according to the Einthoven's triangle, and ICG signals are acquired with four surface electrodes positioned on the subjects' backs. [8,11,28]

The recording protocol consists of two distinct parts. The initial group of 59 recordings comprises two segments, each approximately two minutes in duration: relaxation (baseline) and anger induction. The second group (143 recordings) includes three segments: baseline, neutral, and anger. A neutral phase is incorporated to control for confounding variables, such as mental engagement. Both neutral and anger-inducing fantasy materials are pre-tested in a pilot study [28]. For this study, only the baseline and anger segments are utilized. During the baseline phase, subjects were instructed to sit calmly and relax, and to optionally closing their eyes. During the anger-induction phase, participants listened to an audio-guided scenario designed to evoke anger. Detailed descriptions of the recording protocols are available in our previous publications [8,28].

## A2. Preprocessing and extraction of information of interest

This segment of analysis, which includes preprocessing and delineation steps, is adopted from our previous studies [8,11], and the scripts are available online [29]. Preprocessing and feature extraction are implemented using R programming language v4.3.0 [30], within the R Studio integrated development environment (IDE) (Posit Software, PBC, formerly RStudio, PBC).

### *A2.1 Digital filtering and delineation*

ECG and ICG signals are filtered using the $4^{th}$ order Butterworth bandpass filter to reduce both high-frequency and low-frequency noise. The frequency bandwidth is set to 1 Hz – 40 Hz for ECG recordings and 0.5 Hz – 40 Hz for ICG recordings [31,32]. All filters are applied in a bidirectional manner to compensate for zero-phase distortion. For R peak detection, we employ the modified Pan-Tompkins algorithm [33]. The position of the peaks is double-checked automatically by verifying the location as a maximum within ± 30 ms of the detected peak, which roughly corresponds to the minimum QRS duration in healthy adults [22]. The C point is detected as the maximum value of the ICG signal occurring between two consecutive

R peaks in the ECG [6,23]. To further reduce noise, ensemble averaging is conducted, which involves averaging 10 consecutive ECG and ICG beats (templates) [34]. This approach effectively minimizes noise, particularly noise with a mean value of zero (such as white noise), although it comes at the cost of reducing variability within the averaging cohort [5]. For this process, ECG and ICG templates are created with a duration of 750 ms, comprising 250 ms before the R and C peaks and 500 ms following the peaks for ECG and ICG, respectively [5,8]. A total of 9 averages is extracted per segment (baseline and anger), resulting in 18 samples per individual. The choice of nine averages is made heuristically, as it represents the maximum number of non-overlapping beats (9 averages × 10 beats) that could be extracted from each segment while ensuring equal representation across all subjects. Besides averaging the ensemble, the RR interval and RC interval are averaged on the selected cohort, which are used as features shown in Table A1. On averaged ECG signals, we also detect the positions of Q and S points, along with T1, T, and T2 points, which represent the start, peak, and the end of the T wave, respectively [9]. Additionally, from the ICG signals, besides the C point, the B and the X points are identified, corresponding to the opening and closing of the aortic valve [5,6]. These delineation steps are adopted from our previous studies [8,11].

In line with previous studies [8,11], this study expands the analysis by using a larger number of samples per class and an extended feature set. The updated set provides broader coverage of temporal, amplitude, and morphological descriptors, with all interval-based features (except RR) normalized using Bazett's correction to reduce heart-rate–related confounding, whereas in the earlier set, only selected intervals were corrected, indicated as *int2* [8]. A key enhancement is the inclusion of slope (angular) features, such as QR_slope, RS_slope, TT1_slope, TT2_slope, CB_slope, and CX_slope, which capture the rate of change in waveform segments to complement amplitudes and temporal information. The feature space is further enriched by expanded amplitude and morphological descriptors such as RT_amp and ECGQRScrest, resulting in a more comprehensive and balanced representation of ECG and ICG dynamics.

### *A2.2 Fiducial-based feature extraction*

The fiducial-based features represent relationships between two or more fiducial points, capturing the connections between key events in the cardiovascular system. Depending on the type of relationship, these features can be broadly categorized into temporal, amplitude, slope, and morphological features [8,15,34]. All temporal features, except for RR, are corrected using Bazett's formula to compensate for potentially high heart rate. A list of the extracted features, along with their signal of origin, domain, description, and references, is provided in Table1. In this work, we extend prior feature sets by broadening temporal, amplitude, and morphological descriptors, and by introducing a new slope, also called angular, domain that links changes in amplitude with timing. We also develop additional temporal features that connect ECG and ICG fiducials, denoted as ECG\ICG features. These extensions build on and generalize ideas from Patro et al. [15] and Antić et al. [11], while maintaining compatibility with established measures in the literature.

To understand the distribution of the features listed in Table A1, we calculate the Pearson and Spearman correlation coefficients with the appropriate *p*-value. From this analysis, we identify pairs of features with high correlation, where the absolute value of the correlation coefficient exceeds 0.7. These pairs are represented as an adjacency list, which is then interpreted as a graph structure. In this representation, features are nodes and significant correlations are edges. Groups of nodes, connected in the form of islands, are treated as clusters of highly related features. These clusters represent groups of features that share similar correlation patterns, offering a descriptive view of the structure present in the dataset. It should be noted, however, that correlation does not imply causation [35], and these clusters should be interpreted as descriptive structures rather than evidence of direct underlying mechanisms.

## A3. Decision-making

All ML model development, feature importance analysis, feature selection, and statistical testing are conducted using Python 3.9 [36], employing libraries including NumPy [37], pandas [38], matplotlib [39], pickle [36], random [36], scikit-learn [40], SciPy [41], SHAP [42], statistics [36], math [36], statmodels [43], and deap [44]. The

scripts used for decision-making, feature importance analysis, and feature selection are available online [45]. In this study, we employ the RF algorithm, which demonstrated superior performance compared to other ML models in our previous research [8]. Additionally, RF inherently provides a ranked list of feature importance without requiring additional processing steps [16,40,46], making it a strong candidate for our research, as it directly supports the central aim of identifying which aspects of the ECG and ICG signals carry individual-specific information. It is important to note that hyperparameter tuning is intentionally omitted at all stages of the ML pipeline, as its computational expense is deemed excessive relative to the marginal increase in accuracy observed [8]. The focus is therefore on using ML as a benchmark to track relative changes and improve interpretability, particularly for understanding the contribution of each cardiogram to biometric identification, so the default values for hyperparameters are employed [40]. However, thorough hyperparameter tuning was performed in our previous research [8,13], in the present study, both the sample size and the feature space are substantially increased (from 8-12 up to 18 samples per individual and from 17 up to 29 extracted features), which does not guarantee that the previously optimized hyperparameters would yield the best performance under these new conditions. Moreover, prior work has shown that hyperparameter tuning in RF can indeed improve performance, but often at the cost of significant computational demand [47].

*Table A1: Description of ECG, ICG, and combined ECG\ICG features, including their signal origin, domain, description, and references. Feature names reflect the fiducial points used. The suffix indicates the domain: "int" for interval (temporal), "amp" for amplitude, "slope" for slope (angular), and "crest" without a suffix denotes morphology.*

| Feature | Signal | Domain | Description | Reference(s) |
|---|---|---|---|---|
| RR_int | ECG | temporal | Heart rate | [47,48] |
| QRS_int | ECG | temporal | Duration of depolarization of the ventricle | [15] |
| T_int | ECG | temporal | Duration of repolarization of the ventricle | [9,49] |
| QT_int | ECG | temporal | Duration of depolarization and repolarization of the ventricle | [47] |

| Feature | Signal | Domain | Description | Reference(s) |
|---|---|---|---|---|
| ST_int | ECG | temporal | Duration between the end of depolarization and the end of repolarization of the ventricle | [9,15] |
| RQ_amp | ECG | amplitude | Amplitude difference between the R and Q points in the QRS complex | [15,47–49] |
| RS_amp | ECG | amplitude | Amplitude difference between the R and S points in the QRS complex | [15,47–49] |
| RT_amp | ECG | amplitude | Amplitude difference between R peak and T peak | [47] |
| TT1_amp | ECG | amplitude | Amplitude difference between the peak and the beginning of the T wave | [34,49] |
| TT2_amp | ECG | amplitude | Amplitude difference between the peak and the endpoint of the T wave | [34,49] |
| QR_slope | ECG | slope | Slope of QR segment | [15] |
| RS_slope | ECG | slope | Slope of SR segment | [15] |
| TT1_slope | ECG | slope | Slope between the beginning and the peak of the T wave | Inspired by [15] |
| TT2_slope | ECG | slope | Slope between the peak and the endpoint of the T wave | Inspired by [15] |
| ECGQRScrest | ECG | morphological | Crest factor of the QRS complex | Inspired by [11] |
| ECGTcrest | ECG | morphological | Crest factor of the T wave | [11] |
| BX_int | ICG | temporal | Left Ventricular Ejection Time (LVET) | [11,27] |
| CB_amp | ICG | amplitude | Amplitude difference between the C peak and the B point | [27] |
| CX_amp | ICG | amplitude | Amplitude difference between the C point and the X point | [11] |
| CB_slope | ICG | slope | Slope of BC segment | Inspired by [15,27] |
| CX_slope | ICG | slope | Slope of CX segment | Inspired by [15,27] |
| ICGcrest | ICG | morphological | Crest factor of BCX complex | [11] |
| RC_int | ECG\ICG | temporal | Duration between the peak of electrical and the peak of mechanical activity | [11] |

| Feature | Signal | Domain | Description | Reference(s) |
|---|---|---|---|---|
| QX_int | ECG\ICG | temporal | Duration between the onset of depolarization and the end of the mechanical activity of the ventricles | Inspired by [11,15] |
| QB_int | ECG\ICG | temporal | Duration between the onset of depolarization and the beginning of mechanical activity of the ventricles | Inspired by [11,15] |
| SX_int | ECG\ICG | temporal | Duration between the end of depolarization and the end of mechanical activity of the ventricles | Inspired by [11,15] |
| SB_int | ECG\ICG | temporal | Duration between the end of depolarization and the beginning of the mechanical activity of the ventricles | Inspired by [11,15] |
| BT_int | ECG\ICG | temporal | Duration between the peak of repolarization and the beginning of mechanical activity of the ventricles | Inspired by [11,15] |
| TX_int | ECG\ICG | temporal | Duration between the peak of repolarization and the end of mechanical activity of the ventricles | Inspired by [11,15] |

We assume a closed-world scenario, meaning individuals not registered in the database cannot access the identification system. Consequently, the problem is designed as a multi-class (multinomial) classification task, where each individual represents a distinct class. The dataset is split into training and test sets using a 33% test set ratio with stratification. This ensures that each individual's information is proportionally represented in both the training and test datasets. RF is evaluated solely on the test set, which does not participate in the training process, using standard classification metrics, including accuracy, precision, recall, and the F1 score.

### *A3.1 Feature importance*

Feature importance in this study is evaluated using three complementary metrics: Gini importance (*i.e.*, reduction in impurity), permutation feature importance, and Shapley additive explanations (SHAP) [17,42,46]. In the first method, feature importance estimation is embedded in the training process, and the importance scores are scaled so that their sum equals 1 [40,46]. This approach is computationally efficient and

directly integrated into the RF training process, providing an immediate understanding of which features influence the model. In the second approach, the feature importance is evaluated using the permutation feature importance method with the *n_repeat* parameter set to 100 on the independent test set. Permutation-based importance is robust to features with numerous categories [50] and can be computed on an independent test set [40], ensuring a more reliable evaluation of feature relevance. On the other hand, this method is computationally more intensive, as it involves shuffling feature values multiple times and re-evaluating the model to estimate importance. As the third method, we compute SHAP values, which attribute each prediction to individual features based on cooperative game theory [42]. In the multinomial classification, SHAP produces a three-dimensional output where feature contributions are calculated for each sample and for each class, reflecting how much a given feature pushes the prediction toward or away from each possible class. To obtain a global view, these values are first averaged across classes, then converted to their absolute values and finally averaged across all samples, resulting in the mean absolute importance of each feature.

Since each of these methods captures feature importance from a different perspective, they are applied in parallel, consistent with ensemble-based strategies for feature selection [51]. As feature importance methods do not explicitly eliminate less important features, but rather provide a ranked list, we use thresholding to select only the ten highest-ranked features. The intersection of the sets is then taken, keeping only those features consistently identified among the top ten by all three methods. This strategy improves the stability of the selected features, though it comes at the expense of diversity, as unique features highlighted by a single method are excluded [51]. The resulting subset is evaluated by retraining the RF model and comparing performance against the base model that included all features. Importantly, these evaluation metrics are used solely for benchmarking the stability and informativeness of the selected feature set, and not as predictive validation in the sense of classification performance.

Assessing feature importance in datasets exhibiting multicollinearity is subject to correlation bias [17]. This arises when pairs or groups of highly correlated variables influence the importance scores assigned to individual features [17]. The analysis is

conducted using clusters of highly correlated features. The collective influence of each cluster on model accuracy is assessed by randomly shuffling all features within a cluster and monitoring the resulting decrease in classification accuracy on the test set. Moreover, we examine the relationship between the Gini importance scores of individual features within and outside their respective clusters. From each cluster, a representative feature (*i.e.*, a feature that is correlated with all features in a cluster) is selected and individually shuffled to disrupt its association with the target variable. The RF model is retrained, and Gini importance scores are recalculated. The relative changes in importance scores are analyzed for highly correlated feature pairs. In contrast, the distribution of changes across all remaining features is summarized using descriptive statistics (minimum, maximum, median, the $25^{th}$ percentile, and the $75^{th}$ percentile).

### A4. Feature selection

In contrast to feature importance methods, which provide a ranked list of features according to their estimated significance, feature selection methods explicitly eliminate less informative features from the dataset. This approach removes the need for additional thresholding strategies to determine a final subset, since the selection process itself yields a reduced feature set along with a direct evaluation of its performance, which we validate on an independent test set. In this study, feature selection is conducted using two methods: recursive feature elimination with cross-validation (RFECV) and a genetic algorithm (GA).

RFECV starts with the complete feature set and iteratively removes the least informative features, with the number of eliminated features per step determined by a predefined parameter (commonly one). At each iteration, model performance is evaluated using classification accuracy under cross-validation, and the process continues until the subset that yields the highest predictive performance is identified [40,46]. The minimum number of features is fixed at a predefined parameter of one, and the procedure continues by iteratively discarding the least informative feature until the optimal subset is determined [40,46].

In contrast, the GA follows a stochastic search strategy inspired by natural selection, where candidate feature subsets evolve toward optimal solutions through iterative refinement. The algorithm begins with an initial population of randomly generated subsets, and over multiple generations, applies crossover to exchange features between subsets and mutation to introduce random variations, thereby preserving diversity and reducing the risk of premature convergence. The fitness of each subset is assessed according to model performance, and the best-performing candidates are retained for subsequent iterations. In this study, the GA is implemented with a population size of 50 and evolved for 50 generations to balance exploration of the feature space with computational cost, while the crossover and mutation probabilities are set to 0.5 and 0.1, respectively, to compromise between preserving existing solutions and introducing sufficient variability to avoid premature convergence [52]. The parameters are chosen heuristically, as they are often tuned empirically based on the feature set and dataset properties.

As in the case of feature importance, the final set is produced by intersecting the features selected by both methods. The resulting consensus subset is then used to retrain the RF model, and its performance is compared with the results of the base model. Finally, the consensus feature set obtained through feature selection is compared with the stable subset derived from feature importance methods, allowing us to identify the most consistent discriminators across complementary approaches.

### A5. Emotions and cardiogram-based features

Our previous findings indicate that variability in emotional states challenges the model to generalize effectively [8]. In this context, generalization refers to the model ability to apply patterns learned under one emotional state to another, an ability that tends to deteriorate when affective conditions differ, thereby reducing identification performance.

To examine this effect, we conducted statistical analyses comparing features extracted from baseline and anger segments to identify those significantly affected by emotional changes. Firstly, the Shapiro-Wilk test was applied to each feature within the baseline and anger groups to assess for normal distribution. Depending on the outcome, either

a parametric test (independent t-test) or a non-parametric test (Mann-Whitney U test) was employed to evaluate group differences. To control for multiple comparisons, $p$-values were adjusted using the Bonferroni correction. Effect sizes were calculated in parallel: Cohen's $d$ for the independent t-test, and Cliff's $\delta$ for the Mann-Whitney U test, providing an estimate of practical relevance alongside statistical significance. Cliff's $\delta$ was calculated using the Python implementation available in the *cliffsDelta* repository on GitHub [53]. Features with adjusted $p$-values below the significance threshold ($p < 0.05$) were considered significantly influenced by emotional changes.

Due to multicollinearity, we also examine whether highly correlated pairs of features can yield opposite statistical outcomes, for example, one feature showing a significant baseline–anger difference while its counterpart did not. Identifying such pairs helps pinpoint potential pitfalls because it may reveal a big issue in the development of emotion-agnostic biometric systems.

To further investigate the predictive implications of these findings, we evaluated how emotional variability impacts model performance, specifically identification accuracy. Given that the dataset includes two segments corresponding to different emotional states, the RF model was first trained on one segment using all available features and validated through stratified 3-fold cross-validation. Generalization was then assessed by evaluating on samples from the other (unseen) segment. The results, reported in terms of accuracy and F1 score, served as benchmark values. The model was subsequently retrained using two additional feature sets: (i) features that did not differ significantly across emotional states, and (ii) features that did show significant emotion-related variation. Both subsets underwent the same cross-validation and generalization procedures. These complementary analyses allowed us to determine whether excluding emotion-sensitive features yields more stable identification performance, in comparison to the emotion-sensitive subset.

**References**


[1] A.K. Jain, P. Flynn, A.A. Ross, Handbook of Biometrics, Springer Science & Business Media, 2007.

[2] I. Odinaka, P.-H. Lai, A.D. Kaplan, J.A. O'Sullivan, E.J. Sirevaag, J.W. Rohrbaugh, ECG Biometric Recognition: A Comparative Analysis, IEEE

Trans.Inform.Forensic Secur. 7 (2012) 1812–1824. https://doi.org/10.1109/TIFS.2012.2215324.
[3] T.M.C. Pereira, R.C. Conceição, V. Sencadas, R. Sebastião, Biometric Recognition: A Systematic Review on Electrocardiogram Data Acquisition Methods, Sensors 23 (2023) 1507. https://doi.org/10.3390/s23031507.
[4] R. Srivastva, Y.N. Singh, A. Singh, Statistical independence of ECG for biometric authentication, Pattern Recognition 127 (2022) 108640. https://doi.org/10.1016/j.patcog.2022.108640.
[5] J.T. Cacioppo, L.G. Tassinary, G. Berntson, Handbook of psychophysiology, Cambridge university press, 2007.
[6] A. Sherwood, M.T. Allen, J. Fahrenberg, R.M. Kelsey, W.R. Lovallo, L.J. Van Doornen, Methodological guidelines for impedance cardiography, Psychophysiology 27 (1990) 1–23. https://doi.org/10.1111/j.1469-8986.1990.tb02171.x.
[7] A.S. Rathore, Z. Li, W. Zhu, Z. Jin, W. Xu, A Survey on Heart Biometrics, ACM Comput. Surv. 53 (2021) 1–38. https://doi.org/10.1145/3410158.
[8] I. Tanasković, L.B. Lazarević, G. Knežević, N. Milosavljević, O. Dubljević, B. Bjegojević, N. Miljković, CardioPRINT: Biometric identification based on the individual characteristics derived from the cardiogram, Expert Systems with Applications (2024) 126018. https://doi.org/10.1016/j.eswa.2024.126018.
[9] S.A. Israel, J.M. Irvine, A. Cheng, M.D. Wiederhold, B.K. Wiederhold, ECG to identify individuals, Pattern Recognition 38 (2005) 133–142. https://doi.org/10.1016/j.patcog.2004.05.014.
[10] R. Zhou, C. Wang, P. Zhang, X. Chen, L. Du, P. Wang, Z. Zhao, M. Du, Z. Fang, ECG-based biometric under different psychological stress states, Computer Methods and Programs in Biomedicine 202 (2021) 106005. https://doi.org/10.1016/j.cmpb.2021.106005.
[11] M. Antić, N.B. Popović, N. Milosavljević, O. Dubljević, B. Bjegojević, N. Miljković, CardioPRINT: Individual features hidden in electrocardiogram andimpedance-cardiogram, Empirical Studies in Psychology (2020) 13.
[12] G. Shmueli, To explain or to predict?, Statistical Science (2010) 289–310. https://doi.org/10.1214/10-STS330.
[13] I. Tanasković, L. Lazarević, G. Knežević, N. Milosavljević, O. Dubljević, B. Bjegojević, N. Miljković, Dataset for CardioPRINT-based Biometric Identification, (2023). https://doi.org/10.5281/ZENODO.10204955.
[14] I. Tanasković, L. Lazarević, G. Knežević, N. Milosavljević, O. Dubljević, B. Bjegojevic, N. Miljković, Supplementary Materials for “Interpretable machine learning for cardiogram-based biometrics,” (2025). https://doi.org/10.5281/zenodo.17408263.
[15] K.K. Patro, A. Jaya Prakash, M. Jayamanmadha Rao, P. Rajesh Kumar, An Efficient Optimized Feature Selection with Machine Learning Approach for ECG Biometric Recognition, IETE Journal of Research 68 (2022) 2743–2754. https://doi.org/10.1080/03772063.2020.1725663.
[16] L. Breiman, Random forests, Machine Learning 45 (2001) 5–32. https://doi.org/10.1023/A:1010933404324.

[17] B. Gregorutti, B. Michel, P. Saint-Pierre, Correlation and variable importance in random forests, Statistics and Computing 27 (2017) 659–678. https://doi.org/10.1007/s11222-016-9646-1.
[18] J.R. Pinto, J.S. Cardoso, Explaining ECG biometrics: Is it all in the QRS?, in: IEEE, 2020: pp. 1–5.
[19] F.G. Silva Teodoro, S.M. Peres, C.A.M. Lima, Feature selection for biometric recognition based on electrocardiogram signals, in: 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, Anchorage, AK, 2017: pp. 2911–2920. https://doi.org/10.1109/IJCNN.2017.7966216.
[20] Y. Huang, G. Yang, K. Wang, H. Liu, Y. Yin, Robust multi-feature collective non-negative matrix factorization for ECG biometrics, Pattern Recognition 123 (2022) 108376. https://doi.org/10.1016/j.patcog.2021.108376.
[21] P.W. Macfarlane, A. Van Oosterom, O. Pahlm, P. Kligfield, M. Janse, J. Camm, Comprehensive electrocardiology, Springer Science & Business Media, 2010.
[22] A.M. Katz, Physiology of the Heart, Lippincott Williams & Wilkins, 2010.
[23] I. Karpiel, M. Richter-Laskowska, D. Feige, A. Gacek, A. Sobotnicki, An Effective Method of Detecting Characteristic Points of Impedance Cardiogram Verified in the Clinical Pilot Study, Sensors 22 (2022) 9872. https://doi.org/10.3390/s22249872.
[24] L. Maršánová, A. Němcová, R. Smíšek, M. Vítek, L. Smital, Advanced P wave detection in ECG signals during pathology: evaluation in different arrhythmia contexts, Scientific Reports 9 (2019) 19053. https://doi.org/10.1038/s41598-019-55323-3.
[25] A.J. Camm, M. Malik, J.T. Bigger, G. Breithardt, S. Cerutti, R.J. Cohen, P. Coumel, E.L. Fallen, H.L. Kennedy, R.E. Kleiger, others, Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Circulation 93 (1996) 1043–1065. https://doi.org/10.1161/01.CIR.93.5.1043.
[26] L. Olos, E.-H. Hoff, Gender Ratios in European Psychology, European Psychologist 11 (2006) 1–11. https://doi.org/10.1027/1016-9040.11.1.1.
[27] U. Pale, N. Müller, A. Arza, D. Atienza, ReBeatICG: Real-time low-complexity beat-to-beat impedance cardiogram delineation algorithm, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021: pp. 5618–5624. https://doi.org/10.1109/EMBC46164.2021.9630170.
[28] B. Bjegojević, N. Milosavljević, O. Dubljević, D. Purić, G. Knežević, In pursuit of objectivity: Physiological measures as a means of emotion induction procedure validation, Empirical Studies in Psychology (2020) 17.
[29] I. Tanasković, L. Lazarević, G. Knežević, N. Milosavljević, O. Dubljević, B. Bjegojević, N. Miljković, Luck032/CardioPRINT-biometric-identification-with-machine-learning: CardioPRINT-biometric-identification-with-machine-learning, (2023). https://doi.org/10.5281/ZENODO.10204894.
[30] R.C. Team, others, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, Http://Www. R-Project. Org/ (2016).

[31] A. Sarkar, A.L. Abbott, Z. Doerzaph, ECG biometric authentication using a dynamical model, in: IEEE, 2015: pp. 1–6. https://doi.org/10.1109/BTAS.2015.7358757.
[32] I.B. Salah, K. Ouni, Denoising of the impedance cardiographie signal (ICG) for a best detection of the characteristic points, in: IEEE, 2017: pp. 1–4. https://doi.org/10.1109/BIOSMART.2017.8095347.
[33] J. Pan, W.J. Tompkins, A Real-Time QRS Detection Algorithm, IEEE Trans. Biomed. Eng. BME-32 (1985) 230–236. https://doi.org/10.1109/TBME.1985.325532.
[34] Y. Wang, F. Agrafioti, D. Hatzinakos, K.N. Plataniotis, Analysis of human electrocardiogram for biometric recognition, EURASIP Journal on Advances in Signal Processing 2008 (2007) 1–11. https://doi.org/10.1155/2008/148658.
[35] D.E. Geer Jr, Correlation is not causation, IEEE Security & Privacy 9 (2011) 93–94. https://doi.org/10.1109/MSP.2011.26.
[36] G. Van Rossum, F.L. Drake, Python reference manual, Centrum voor Wiskunde en Informatica Amsterdam, 1995.
[37] C.R. Harris, K.J. Millman, S.J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, Array programming with NumPy, Nature 585 (2020) 357–362. https://doi.org/10.1038/s41586-020-2649-2.
[38] W. McKinney, Data structures for statistical computing in python, in: Proceedings of the 9th Python in Science Conference, Austin, TX, 2010: pp. 51–56. https://doi.org/10.25080/Majora-92bf1922-00a.
[39] J.D. Hunter, Matplotlib: A 2D graphics environment, Computing in Science & Engineering 9 (2007) 90–95. https://doi.org/10.1109/MCSE.2007.55.
[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: Machine learning in Python, The Journal of Machine Learning Research 12 (2011) 2825–2830.
[41] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature Methods 17 (2020) 261–272. https://doi.org/10.1038/s41592-019-0686-2.
[42] S.M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: Curran Associates, Inc., 2017. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.
[43] S. Seabold, J. Perktold, Statsmodels: econometric and statistical modeling with python., in: Proceedings of the 9th Python in Science Conference, 2010: pp. 92–96. https://doi.org/10.25080/Majora-92bf1922-011.
[44] F.-A. Fortin, F.-M. De Rainville, M.-A.G. Gardner, M. Parizeau, C. Gagné, DEAP: Evolutionary algorithms made easy, The Journal of Machine Learning Research 13 (2012) 2171–2175.
[45] Ilija Tanaskovic, L. Lazarevic, G. Knežević, N. Milosavljevic, O. Dubljević, B. Bjegojevic, N. Miljković, Luck032/Interpretable_machine_learning_for_cardiogram-based_biometrics: Interpretable_machine_learning_for_cardiogram-based_biometrics, (2025). https://doi.org/10.5281/ZENODO.17804969.

[46] A.C. Müller, S. Guido, Introduction to machine learning with Python: a guide for data scientists, O'Reilly Media, Inc., 2016.
[47] A. Pal, Y.N. Singh, ECG Biometric Recognition, in: D. Ghosh, D. Giri, R.N. Mohapatra, E. Savas, K. Sakurai, L.P. Singh (Eds.), Mathematics and Computing, Springer Singapore, Singapore, 2018: pp. 61–73. https://doi.org/10.1007/978-981-13-0023-3_7.
[48] S.K. Singla, A. Sharma, ECG as biometric in the automated world, International Journal of Computer Science & Communication 1 (n.d.) 12–21.
[49] R. Tan, M. Perkowski, ECG Biometric Identification Using Wavelet Analysis Coupled with Probabilistic Random Forest, in: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, Anaheim, CA, USA, 2016: pp. 182–187. https://doi.org/10.1109/ICMLA.2016.0038.
[50] A. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: a corrected feature importance measure, Bioinformatics 26 (2010) 1340–1347. https://doi.org/10.1093/bioinformatics/btq134.
[51] V. Bolón-Canedo, A. Alonso-Betanzos, Ensembles for feature selection: A review and future trends, Information Fusion 52 (2019) 1–12. https://doi.org/10.1016/j.inffus.2018.11.008.
[52] A. Lambora, K. Gupta, K. Chopra, Genetic algorithm-A literature review, in: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, 2019: pp. 380–384. https://doi.org/10.1109/COMITCon.2019.8862255.
[53] N.A. Ernst, cliffsDelta, (2025). https://github.com/neilernst/cliffsDelta (accessed November 1, 2025).