AutoML-Driven ECG Classification of Cardiac Pathologies with Explainable AI*

Dmytro Tymoshchuk[1,*,†], Andriy Sverstiuk[2,†], Vitaliy Tymoshchuk[1,†], Lyubomyr Mosiy[1,†] and Oksana Bahrii-Zaiats[2,†]

[1] *Ternopil Ivan Puluj National Technical University, Ruska str. 56, Ternopil, 46001, Ukraine*
[2] *I. Horbachevsky Ternopil National Medical University, Maidan Voli St., 1, Ternopil, 46002, Ukraine*

Abstract

Cardiovascular diseases (CVD) remain one of the leading causes of mortality in the world, which emphasizes the need to create accurate, interpretable and effective diagnostic systems. Electrocardiography is a key non-invasive method that provides critical information about the functional state of the heart. In this work, an information technology for automated diagnosis of CVD based on the time rhythm function taking into account the extreme amplitudes of the characteristic electrocardiogram (ECG) waves (P, Q, R, S, T) was proposed. A dataset of 924 samples from the open PhysioNet databases was formed, covering four diagnostic categories (conditional norm, norm with a pacemaker, arrhythmias and morphological pathologies). Ten statistical descriptors (Mean, Median, Mode, Standard Deviation, Sample Variance, Kurtosis, Skewness, Range, Minimum, Maximum) were used to describe temporal variability. The EvalML AutoML framework was used to build the models, which automatically determined the optimal data processing. The Extra Trees Classifier algorithm turned out to be the best, achieving an average classification accuracy of about 96.5% for four classes and an AUC of more than 0.92, which confirms its generalization capabilities. To ensure the transparency of the results, the SHAP method was used, which showed that the most significant features are Skewness and Kurtosis. The integration of AutoML and Explainable AI methods provided high accuracy and reliability of diagnostics while maintaining interpretability, which makes the proposed approach promising for clinical application and analysis of other biomedical signals.

Keywords

ECG signal, machine learning, EvalML, AutoML, Explainable AI, SHAP

1. Introduction

In recent decades, cardiovascular diseases (CVD) have become one of the most serious challenges for the global healthcare system. According to the World Health Organization, they account for almost a third of all deaths in the world. Such statistics indicate the extreme relevance of research in the field of cardiology and the search for effective methods of early diagnosis, monitoring and prevention. Timely detection of pathological changes in cardiac activity allows significantly reducing the risk of complications and saving the lives of millions of patients.

One of the key tools in this area is electrocardiography — a non-invasive method that provides a safe and relatively simple registration of the electrical activity of the heart. The ECG signal contains multilayered information about the functional state of the myocardium, which is reflected in the shape, amplitude and time parameters of the characteristic P, Q, R, S, T waves, as well as the intervals and segments between them. It is the analysis of the morphology of these components that allows us to detect a wide range of pathologies — from rhythm and conduction disorders to signs of ischemic disease or structural changes in the heart muscle. Due to this, the ECG remains

* *ProfIT AI'25: 5th International Workshop of IT-professionals on Artificial Intelligence, October 15–17, 2025, Liverpool, UK*
1* Corresponding author.
† These authors contributed equally.
dmytro.tymoshchuk@gmail.com (D. Tymoshchuk); sverstyuk@tdmu.edu.ua (A. Sverstiuk); Tymoshchuk@tntu.edu.ua (V. Tymoshchuk); lmosiy@gmail.com (L. Mosiy); bagrijzayats@tdmu.edu.ua (O. Bahrii-Zaiats)
0000-0003-0246-2236 (D. Tymoshchuk); 0000-0001-8644-0776 (A. Sverstiuk); 0009-0007-2858-9434 (V. Tymoshchuk); 0009-0000-9778-331X (L. Mosiy); 0000-0002-5533-3561 (O. Bahrii-Zaiats)

an indispensable method in modern clinical practice, serving as the basis for both routine medical examinations and specialized scientific research.

Despite significant progress in digital processing of ECG signal, there remains a gap between the complexity of the temporal organization of the heart rhythm and the methods for its analysis. Existing approaches often ignore interwave temporal variability or consider it in isolation from amplitude characteristics. This limits the ability to detect complex arrhythmias, early signs of ischemia, or conduction disorders, which may manifest themselves precisely in changes in the temporal structure of individual ECG signal components.

Therefore, there is a need to use approaches that can overcome the limitations of classical methods. In this context, modern machine learning (ML) methods are of particular importance, as they are a universal tool for working with large data sets, providing the ability to find hidden dependencies where traditional approaches are powerless. ML is actively used in the financial sector to predict risks and optimize investment strategies [1,2], in transport to develop intelligent control systems and road safety [3,4], in medicine to create intelligent decision-making support systems and predict the course of diseases [5,6], in materials science to model the structural properties and durability of materials [7,8], in energy to assess consumption and improve resource efficiency [9,10], as well as in cybersecurity to detect anomalies in network traffic and prevent attacks [11,12]. Due to their ability to adapt to different types of data and conditions, ML methods have become the basis for creating systems that provide accurate predictions, high speed of analysis, and scalability in various application contexts.

Recent advances in ML for ECG signal analysis demonstrate significant progress in automated diagnosis of cardiovascular diseases. Based on a comprehensive review of publications in the Scopus database, the following analysis presents advances in ML methods for the analysis of heart rate variability and cardiac signals.

Classical ML algorithms remain powerful tools for ECG signal classification. A hybrid approach that combines Dual-Tree Complex Wavelet Transform (DTCWT) with ML classifiers to detect six classes of cardiac arrhythmias was proposed [13]. Such methodology, tested on the full 48-record MIT-BIH Arrhythmia Database, demonstrates the effectiveness of combining advanced feature extraction methods with traditional classifiers. A comprehensive approach to ECG classification using wavelet scattering to extract low-variability features was developed [14]. Using the MRMR (Minimum Redundancy and Maximum Relevance) algorithm for feature selection, it was achieved an outstanding accuracy of 99.84% using a cubic SVM model among twenty tested ML models. Optimized ensemble methods for ECG anomaly detection were investigated, and the optimized XGBoost model using Bayesian hyperparameter optimization achieved 100% accuracy, significantly outperforming modified gradient boosting with an accuracy of 96.58% and SVM with 91.69% [15].

The revolutionary impact of deep learning on cardiac signal analysis is evident through numerous innovative architectures. The study [16] presented an advanced approach to ECG classification by integrating adaptive segmentation of heart beats with relative heart rate information in a deep learning network. This methodology achieved a sensitivity of 99.81% for normal beats, 99.08% for premature ventricular beats, and 97.83% for premature atrial beats. A hybrid model combining a convolutional neural network (CNN) with a Vision Transformer (ViT) was proposed for analyzing 12-lead ECG recordings [17]. The model achieved an average accuracy of 74% for five-class and 80% for four-class classification on the PTB-XL dataset, demonstrating the potential of transformer architectures in cardiac diagnostics. A CNN model for automatic diagnosis of multiple heart diseases from phonocardiographic signals was developed [18]. Using data augmentation techniques to improve robustness in noisy environments, the model achieved 98.60% accuracy on the test set, highlighting the effectiveness of deep learning for non-invasive diagnosis. The integration of deep learning paradigms has fundamentally transformed the capabilities of cardiac signal analysis. The application of WaveGRU-Net for non-contact ECG reconstruction using millimeter-wave technology with multiple inputs and outputs was pioneered [19]. This innovative approach successfully distinguishes respiratory and cardiac components in the time-frequency domain while maintaining robust semantic representation capabilities. A significant architectural advance has been made with the development of a frequency-guided hierarchical shifted window (FG-HSWIN) transformer incorporating inter-frequency attention mechanisms [20]. This architecture demonstrates exceptional classification performance, achieving 98.72% accuracy on the MIT-BIH arrhythmia dataset through frequency-stratified window attention mechanisms. The integration of frequency-aware positional encoding (FAPE) and lightweight multiscale feature fusion (LMFF) represents a significant methodological contribution to the field. The neuro-fuzzy paradigm exemplifies the trend towards hybrid architectures, where a multimodal feature fusion framework combining transformer-based processing with neuro-fuzzy systems achieves 98.46% accuracy and 99.1% F1-score, demonstrating the effectiveness of integrated computational approaches [21]. A sophisticated three-phase framework incorporating change point detection via autoencoder architectures, facilitating real-time processing of sequential data for cardiac anomaly detection, has been presented [22].

Combining different cardiac signal modalities improves diagnostic accuracy. A cardiovascular disease prediction system integrating ECG and phonocardiogram using Hidden Semi-Markov Model was developed [23]. The proposed HSMM achieved sensitivity of 0.952, specificity of 0.92, F-score of 0.94, accuracy of 0.91, and AUC of 0.96. A multimodal approach to emotion recognition by combining ECG signals with facial features was presented [24]. Using LightGBM, Bagged Decision Trees, Linear SVM, and Gaussian Naive Bayes, the combined model achieved an accuracy of 93.80%, demonstrating the effectiveness of multimodal analysis. Researchers in [25] demonstrated the effectiveness of hybrid approaches by integrating a dual tree complex wavelet transform with traditional ML classifiers. Their methodology successfully distinguishes six different classes of ECG beats, highlighting the continued relevance of traditional signal processing methods combined with modern classification algorithms. The ResNet-34 architecture proposed in [26], improved through transfer learning mechanisms, further confirms the effectiveness of deep residual networks in analyzing cardiac signals in the time-frequency domain. The integration of multiple sensory modalities improves diagnostic complexity. Scientists [27] developed a device with the support of fog computing that integrates phonocardiography, ballistocardiography and seismocardiography to estimate systolic blood pressure, achieving a mean absolute error of 3.5 mm Hg. This multimodal approach uses additional information from mechanical and electrical cardiac signals. The MDD2DG-IRA methodology presented by the researchers in [28] is an example of sophisticated multi-channel processing via dynamic graph convolution, achieving 99.94% classification accuracy for myocardial infarction localization. This approach demonstrates the potential of graph-based representations to capture inter-channel dependencies in multi-lead ECG analysis.

Computational efficiency remains paramount for real-time and embedded applications. This challenge was addressed by analyzing the reconstructed phase space with optimized delay state networks, achieving 99.3% accuracy while reducing hardware requirements by an order of magnitude [29]. An ensemble compression technique combining CEEMD with LSTM autoencoder architectures achieves compression ratios of 38.26 with minimal signal distortion (PRD=0.37) [30]. The development of AutoML greatly simplifies the process of developing models for medical diagnostics. The prospects of automated ML in biomedical signal processing, namely automation for practical implementation in clinical practice, are relevant [31]. A study demonstrated improved ECG-based stress classification using optimization techniques [32]. By applying genetic algorithms, artificial bee colony and improved particle swarm optimization to tune Multi-kernel SVM, the authors achieved an average accuracy of 98.93%, precision of 96.83%, completeness of 96.83% and specificity of 96.72%.

The aim of our study is to develop and experimentally test information technology for automated diagnosis of cardiovascular pathologies based on the time function of the rhythm taking into account the extreme amplitude values of the characteristic waves of the ECG signal, using AutoML and Explainable AI methods to increase the accuracy and reliability of the diagnosis of cardiovascular diseases.

2. Methods and Models

2.1. Feature Extraction and Dataset Construction

To construct datasets suitable for ML algorithms, the ECG signals were processed using a time rhythm function taking into account extreme values of the amplitude of the characteristic waves of the ECG signal, which include extreme values of the amplitude of the characteristic waves (P, Q, R, S, and T) [33].

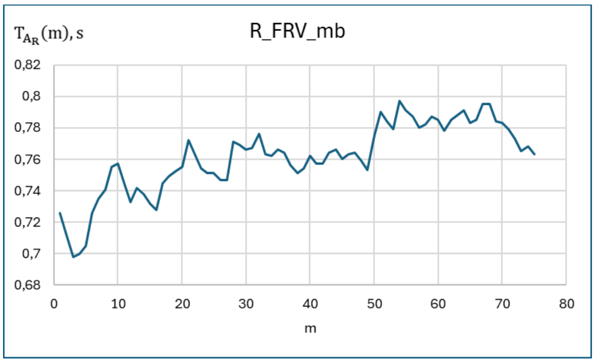
The discrete mathematical model of the time rhythm function taking into account the extreme values of the amplitude of the characteristic waves of the ECG signal [33] is represented by the function $T_{A_k}(m)$, which takes into account the extreme values of the amplitude of the characteristic waves of the ECG signal (P, Q, R, S and T):

$$T_{A_k}(m)=t_{A_k}(m)-t_{A_k}(m-1), k \in \{P, Q, R, S, T\}, m \in Z \tag{1}$$

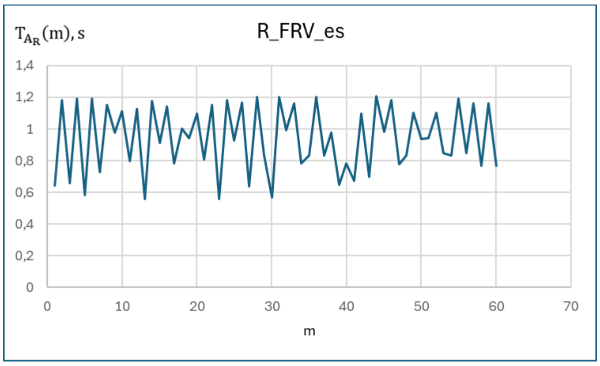
where $t_{A_k}(m)$ – the moment of reaching the peak of the k-type wave in the m-th cardiac cycle (c), $t_{A_k}(m-1)$ – time of peak k-wave in the previous cardiac cycle (m-1) (c), $T_{A_k}(m)$ – the value of the time rhythm function taking into account the extreme values of the amplitude of the characteristic waves of the ECG signal, reflecting the time interval between the peaks of the k-type waves in the current m and the previous cardiac cycle (m-1), $k \in \{P, Q, R, S, T\}$ – type of characteristic wave, $m \in Z$ – cycle number.

The data source was publicly available databases from the PhysioNet repository. The primary dataset [34] contained 12-lead electrocardiographic recordings from 45 152 subjects, digitized at a sampling rate of 500 Hz. This comprehensive collection covers a variety of cardiac arrhythmias and cardiovascular pathologies, with expert-verified annotations to ensure diagnostic accuracy. Additionally, the study included data [35] obtained from extended electrocardiographic monitoring of 15 individuals diagnosed with progressive congestive heart failure, including 11 men (age range: 22–71 years) and 4 women (age range: 54–63 years). The third dataset [36] provided over 100 two-lead electrocardiographic recordings, each covering 15-minute intervals. These recordings contained detailed morphological annotations delineating the temporal boundaries of cardiac waveform components — specifically the onset, peak, and termination of P waves, QRS complexes, T waves, and, when applicable, U waves — for a representative sample of 30 to 50 cardiac cycles per recording.

For each ECG signal recording, the values of the time rhythm function were determined taking into account the extreme values of the amplitude of the characteristic ECG signal waves. Examples of the time rhythm function taking into account the extreme values of the R wave amplitude are shown in Figure 1.



(a)



(b)

Figure 1: Examples of the time rhythm function taking into account extreme values of the R wave amplitude: incomplete blockade of the right bundle branch block (mb) (a), extrasystole (es) (b).

For quantitative description of the function $T_{A_k}(m)$, a statistical processing method is used, which allows calculating the following statistical parameters:

- Mean (arithmetic mean) is a measure of the central tendency of a distribution of temporal variability.
- Median is a robust measure of central tendency that is robust to outliers.
- Mode is the most frequently occurring value of temporal variability.
- Standard Deviation is a measure of dispersion about the mean.
- Sample Variance is the square of the standard deviation, which reflects variability.
- Kurtosis is a measure of the skewness of the distribution.
- Skewness is a measure of the asymmetry of the distribution.
- Range is the difference between the maximum and minimum values.
- Minimum is the smallest value of the time variability.
- Maximum is the largest value of the time variability.

The compiled dataset contained 924 samples. The following conventions were introduced into the dataset: Mean - Mean_t, Median - Med_t, Mode - Mo_t, Standard Deviation-StD_t, Sample Variance-SV_t, Kurtosis-Kur_t, Skewness-Sk_t, Range-Ra_t, Minimum-Min_t, Maximum-Max_t.

The classification was carried out according to four diagnostic categories. The first class represented patients without detected pathologies (conditional norm). The second class included individuals with normal cardiac function, but with installed pacemakers. The third class combined various forms of cardiac rhythm disorders. The fourth class covered pathologies associated with structural changes in the heart muscle and conduction system, manifested as stable changes in the morphology of ECG complexes. The last category also included the consequences of a previous myocardial infarction with the formation of scar tissue, violations of intraventricular conduction of varying degrees, an increase in the mass of the myocardium of individual heart departments and various forms of cardiomyopathies.

The development environment for building and testing ML models was Python, which provided the necessary tools for data preprocessing, sample partitioning, and training ML algorithms. To train and test the model, the original dataset was divided into training and test sets. The size of the test set was 20% of the total data, while the remaining 80% was used to train the model. To ensure reproducibility of the results, the initial value of the random number generator was fixed (random_state = 22). Since the problem is multi-class, the stratify=y parameter was used, which guaranteed the preservation of the initial class ratio in both the training and test sets. This approach avoided biases in the class distribution and ensured the representativeness of the model evaluation.

2.2. Machine Learning Methods

Automated machine learning (AutoML) is a concept aimed at eliminating the need for manual execution of routine tasks that accompany model building. Traditionally, the process of creating a model includes the stages of data preprocessing, feature selection, algorithm selection, hyperparameter optimization and validation of the obtained results. Performing these tasks requires significant experience and time, which limits the widespread use of ML. AutoML offers an approach that automatically combines data processing, model selection and optimization methods, reducing the influence of the human factor and ensuring the stability of the obtained results. Thanks to the use of optimization algorithms, the search for parameters is more efficient than in classical brute force, and integrated evaluation mechanisms guarantee the objectivity and reproducibility of the models.

One of the modern implementations of AutoML is the EvalML library [37]. It is available in Python and provides a full cycle of automation, starting from data preprocessing and ending with obtaining a ready-made pipeline for practical application. EvalML automatically generates a sequence of operations, which includes missing value imputation, coding of categorical variables, scaling and selection of the optimal classification or regression algorithm. Using the AutoMLSearch module, the model and parameter space is explored, and the results are presented in the form of a ranked list ordered by the selected quality metric. The system allows to extend the pipeline with custom models and transformers, which makes it suitable for both scientific research and industrial applications. Thanks to compatibility with the Woodwork library and support for exporting pipelines, EvalML can be easily integrated into a production environment. EvalML embodies the principles of AutoML in a practical tool that allows you to quickly obtain reproducible and competitive models. This makes it particularly useful in tasks that require testing a large number of combinations of algorithms and data preprocessing procedures.

2.3. Model Evaluation and Interpretation

The evaluation of the effectiveness of classification models is carried out on the basis of indicators calculated from the confusion matrix. It reflects the ratio between the correct and incorrect predictions of the model for individual samples and consists of four main components. True Positives (TP) corresponds to the number of samples of the positive class correctly classified as positive. True Negatives (TN) characterizes the number of samples of the negative class correctly classified as negative. False Positives (FP) means the number of samples of the negative class that the model mistakenly classified as positive. False Negatives (FN) reflects the number of samples of the positive class that were incorrectly classified as negative. Based on these values, the metrics Accuracy, Recall, Specificity, Precision, F1-score and G-Mean are calculated (Table 1) [38].

Table 1
Performance metrics of classification models

Metric	Formula	Interpretation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Proportion of correctly classified samples among all observations.
Recall	$\frac{TP}{TP+FN}$	Ability of the model to correctly identify positive samples.
Specificity	$\frac{TN}{TN+FP}$	Ability of the model to correctly identify negative samples.
Precision	$\frac{TP}{TP+FP}$	Proportion of correctly classified positive samples among all predicted positives.
F1-score	$2\times\frac{Precision\times Recall}{Precision+Recall}$	Harmonic mean of Precision and Recall, reflecting their balance.
G-Mean	$\sqrt{Recall\times Specificity}$	Geometric mean of Recall and Specificity, used to assess classification balance.

Taken together, these metrics provide a holistic view of the performance of a classification model, allowing to evaluate not only the overall accuracy, but also the recognition efficiency of each class and the balance of the classification.

Interpreting ML results is an important part of modern research, as it allows not only to assess the quality of the prediction, but also to understand the contribution of each feature to the model's decision-making. One of the most common approaches is the SHapley Additive exPlanations (SHAP) method, based on the concept of Shapley values from cooperative game theory [39]. In the classical formulation, Shapley values describe the contribution of each player to the overall win of the coalition. In the context of ML, the "players" are the features, the "coalition" is their set, and the "win" is the model's prediction. The idea is to fairly distribute the predicted outcome among all features depending on their contribution. The Shapley value for feature i is calculated as the average marginal contribution of this feature to all possible subsets of features.

SHAP has a number of desirable properties: local accuracy (the explanation corresponds to a specific prediction), additivity (the contributions of features are summed to the prediction), fairness (features with the same influence have the same values), which makes the method universal for interpreting different types of models. Due to this, SHAP provides not only a global interpretation, when the average influence of features in the entire sample is analyzed, but also a local one, which allows explaining an individual decision for a specific sample. This opens up the possibility of simultaneously evaluating generalized patterns and controlling the behavior of the model on individual cases.

3. Results and Discussion

The work of the AutoML algorithm, implemented in the EvalML library, made it possible to automatically select the optimal data processing pipeline and classification algorithm for the task. The AutoMLSearch function was used for training. Accuracy multiclass was chosen as the target function, which served as the main optimization criterion. The maximum number of model search iterations was set at 25, which allowed, on the one hand, to save computational resources, and on the other hand, to ensure sufficient coverage of the space of possible models. The parameter n_jobs=-1 ensured the use of all available processor cores, increasing the efficiency of calculations. The initial value of the random number generator (random_seed=22) guaranteed the reproducibility of the experiment.

In the AutoML search process, K-fold stratified cross-validation was applied to the training subsample. The number of folds corresponded to the default parameters of the framework (5 stratified splits). At each fold, all stages of data preprocessing and the selected classifier were trained on the training part, after which their effectiveness was checked on the validation part. For each pipeline, the target metric (accuracy multiclass) was calculated and averaged over the results of all folds, forming the mean_cv_score indicator. It was this average CV score that determined the model's place in the ranking (Figure 2).

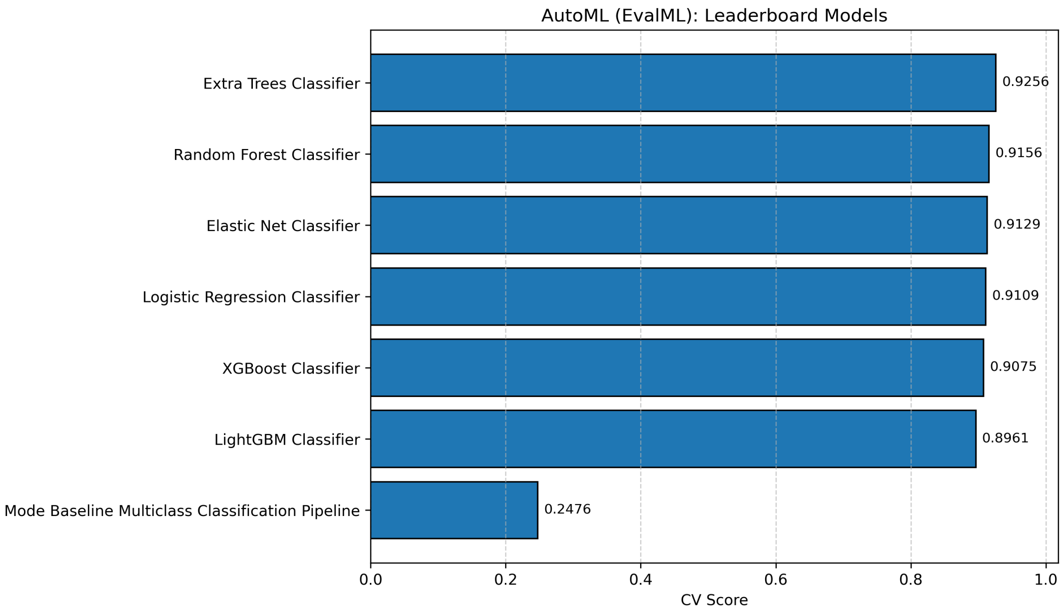
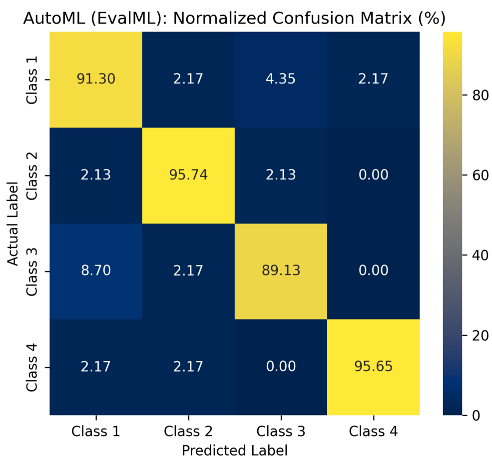


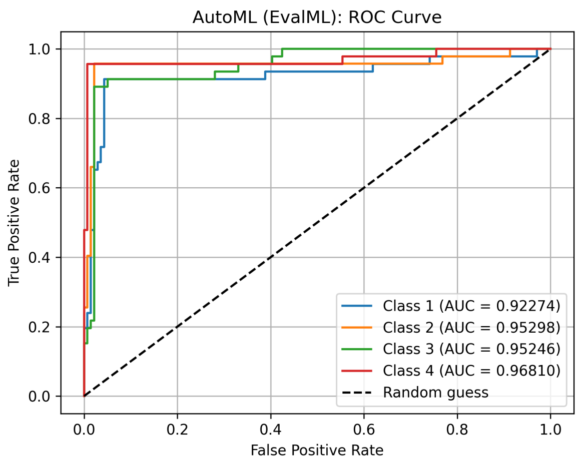
Figure 2: AutoML (EvalML) model leaderboard.

The best model was selected according to the performance values and used for further analysis. The optimal pipeline generated by EvalML consisted of four sequential steps of preprocessing and classification. The first stage used the Label Encoder, which provided encoding of categorical variables into a numerical format required for further processing. The next component was the Imputer, which performed validation and, if necessary, filled in missing values. For numerical features, the median filling strategy was used. The third step was the Select Columns Transformer, which selected a subset of the most relevant features for modeling. The final stage was the Extra Trees Classifier, which belongs to the family of ensemble methods based on random trees [40]. EvalML automatically selected the hyperparameters: the number of trees n_estimators=997, the maximum depth max_depth=10, the feature number selection strategy max_features=log2, as well as the default settings for the partitioning parameters and node weights.

Figure 3 presents the evaluation results of the Extra Trees Classifier model, including the normalized confusion matrix (%) and multi-class ROC curves.



(a)



(b)

Figure 3: Normalized confusion matrix (%) (a) and multiclass ROC curves (b) for the Extra Trees Classifier model on the test dataset.

The confusion matrix shows that the model provides a high level of correct classifications, ranging from 89.13% to 95.74%. The lowest rate is observed for the third class, where 8.70% of examples were incorrectly assigned to the first class, while for other classes the proportion of false predictions does not exceed 4.35%. On average, the classification accuracy is about 96.5%, which indicates a balanced quality of recognition between classes. The ROC curves reflect the ratio between the frequency of true positive and false positive classifications when changing the model threshold. High values of the area under the curve (AUC > 0.92) were recorded for all four classes. The curves are located significantly above the random guess line, which confirms the stable ability of the model to distinguish positive and negative examples.

In order to get a more complete picture of the model's performance, the model performance indicators were calculated. The obtained values are given in Table 2.

Table 2
Performance indicators of the Extra Trees Classifier model

Class	Accuracy	Recall	Specificity	Precision	F1-Score	G-Mean
1 (conditional norm)	0.9459	0.9130	0.9568	0.8750	0.8936	0.9346
2 (conditional norm with an implanted pacemaker)	0.9729	0.9574	0.9782	0.9375	0.9473	0.9677
3 (arrhythmia)	0.9567	0.8913	0.9784	0.9318	0.9111	0.9338
4 (morphological diseases)	0.9837	0.9565	0.9928	0.9777	0.9670	0.9744

The set of obtained values of these metrics confirmed the high efficiency of the constructed model, which not only demonstrated high overall accuracy, but also maintained the proper balance between positive class detection, correct negative class recognition, and prediction reliability.

To interpret the model's performance, the SHAP method was used in the KernelExplainer variant for the multi-class classification task. The choice of this particular method is explained by the fact that the resulting pipeline can contain various preprocessing steps and classifiers, not limited to tree algorithms. KernelExplainer, being model-agnostic, ensures the correctness and stability of the obtained explanations in such conditions. To construct the explanations, a background distribution was formed based on the training subsample. Using the training data as the background prevented information leakage from the test set, preserving the purity of the evaluation. Next, a subset of test examples was formed for analysis, limited to 50 samples. This provided a balance between reducing computational complexity and preserving the statistical representativeness of the test data. The resulting subset was used to calculate local and global SHAP values, allowing us to estimate the contribution of each feature to the class probability prediction.

Figure 4 presents a global SHAP bar plot that displays the mean absolute SHAP values for all features, averaged across classes.

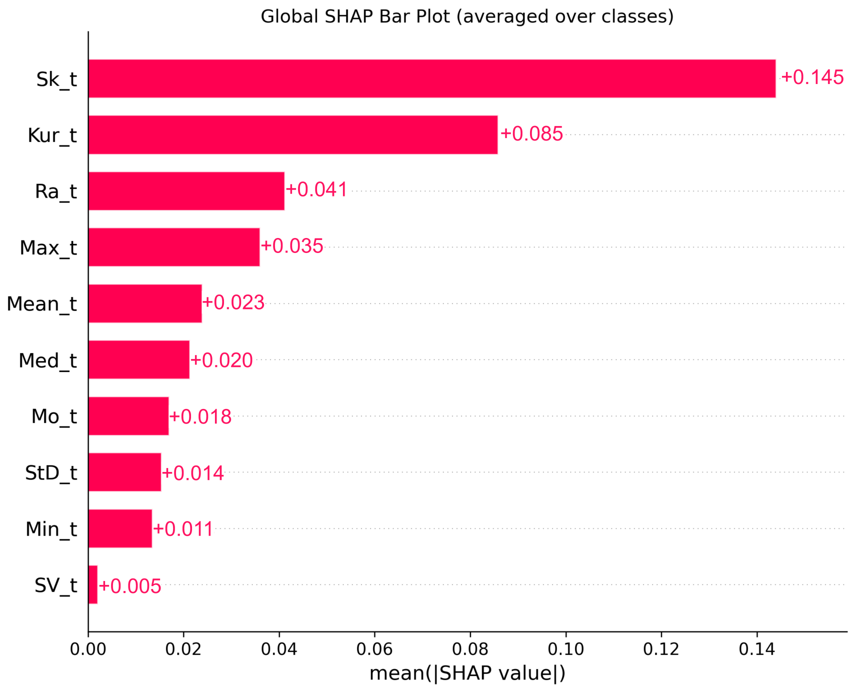


Figure 4: Global SHAP bar plot with mean absolute feature contributions.

This visualization allows us to assess the relative contribution of each feature to the decision-making process of the model. The most influential feature was the Skewness indicator (Sk_t), which indicates its greatest contribution to prediction and a decisive role in class differentiation. In second place in importance is the Kurtosis (Kur_t), which also plays a significant role in decision-making. The next most important are Range (Ra_t) and Maximum (Max_t), which demonstrate a moderate impact on the classification results. The features Mean (Mean_t), Median (Med_t), Mode (Mo_t), Standard Deviation (StD_t) and Minimum (Min_t) have close values of the average absolute impact, which indicates their additional role in the prediction process. The value of Sample Variance (SV_t) practically did not demonstrate a noticeable contribution to the model's work.

Figure 5 shows a force plot for sample #35, which reflects the local explanation of the model's prediction when assigning the sample to class 4.

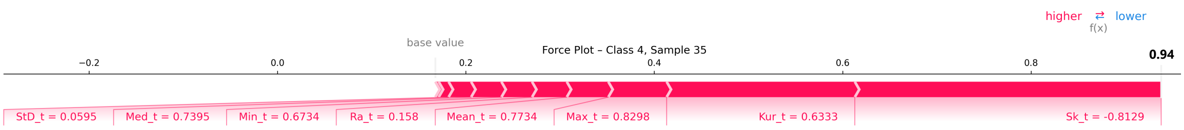


Figure 5: SHAP force plot for sample #35, class 4.

The horizontal scale shows the deviation of the forecast from the base value, which was 0.1676. The final value for class 4 is 0.94, which corresponds to the high confidence of the model in the correct classification of this sample. The visualization shows how individual features affected the forecast bias. According to the explanation given, for this sample the most significant predictors were the Skewness (Sk_t) and Kurtosis (Kur_t), which is consistent with the global results of the SHAP analysis, where these features also took leading positions. Thus, the SHAP analysis confirms that the model forecast for sample #35 is well-founded. Key features (Sk_t, Kur_t, Max_t, Mean_t) provided a confident assignment to class 4, while their role in the forecasts for other classes was minimal or even negative (Table 3).

Table 3
SHAP values for the selected sample (#35) across classes and features

Feature	Class 1	Class 2	Class 3	Class 4
Mean_t	+0.000645	-0.020046	-0.024696	+0.044083
Med_t	+0.001635	-0.007396	-0.026408	+0.032154
Mo_t	+0.001970	+0.003808	-0.016280	+0.010488
StD_t	+0.000770	-0.000953	-0.023727	+0.023896
SV_t	-0.000900	-0.003858	+0.001194	+0.003549
Kur_t	+0.000738	-0.133105	-0.067439	+0.199792
Sk_t	-0.275653	-0.021847	-0.027707	+0.325192
Ra_t	+0.001428	+0.022491	-0.060836	+0.036903
Min_t	+0.000000	-0.020272	-0.012103	+0.032504
Max_t	-0.000632	-0.009310	-0.051998	+0.061926
Base value	0.270010	0.252381	0.310000	0.167608
Predicted sum	0.000011	0.061893	0.000000	0.938095

Thus, the SHAP analysis not only confirmed the accuracy of the results obtained, but also ensured the transparency of the model's operation, which is an important factor for its scientific substantiation and practical application.

4. Conclusion

The paper proposed and implemented an approach to diagnosing cardiovascular diseases based on the time function of the rhythm, taking into account extreme amplitude values of characteristic ECG waves. The use of the AutoML library EvalML allowed to form an optimal preprocessing and classification pipeline, where the Extra Trees Classifier algorithm showed the best results with an average accuracy of about 96.5% and high AUC indicators (>0.92) for all classes.

Interpretation of the model using SHAP confirmed the transparency of the predictions and identified the key features (Skewness, Kurtosis, Range, Maximum) that most influenced the classification results. The obtained results indicate that the combination of AutoML and Explainable AI provides high efficiency and reliability in the analysis of ECG signals, opening up prospects for the practical implementation of such systems in clinical diagnostics and their adaptation for other biomedical analytics tasks.

Acknowledgements

The authors would like to thank the ERASMUS-EDU-2025-CBHE project “Digital Transformation and Curriculum Development for Healthcare Teams” (Digi-CHange - 101233888 – GAP-101233888) for the idea, inspiration and funds that made this work possible.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to grammar and spell check, and improve the text readability. After using the tool, the authors reviewed and edited the content as needed to take full responsibility for the publication’s content.

References

[1] S. Ahmed, M. M. Alshater, A. E. Ammari, H. Hammami, Artificial intelligence and machine learning in finance: A bibliometric review, Res. Int. Bus. Financ. 61 (2022) 101646. doi:10.1016/j.ribaf.2022.101646.

[2] P.-F. Tsai, C.-H. Gao, S.-M. Yuan, Stock Selection Using Machine Learning Based on Financial Ratios, Mathematics 11.23 (2023) 4758. doi:10.3390/math11234758.

[3] X. Tao, L. Cheng, R. Zhang, W. K. Chan, H. Chao, J. Qin, Towards Green Innovation in Smart Cities: Leveraging Traffic Flow Prediction with Machine Learning Algorithms for Sustainable Transportation Systems, Sustainability 16.1 (2023) 251. doi:10.3390/su16010251.

[4] T. Yuan, W. Rocha Neto, C. E. Rothenberg, K. Obraczka, C. Barakat, T. Turletti, Machine learning for next-generation intelligent transportation systems: A survey, Trans. Emerg. Telecommun. Technol. 33.4 (2021). doi:10.1002/ett.4427.

[5] S. O. Nykytyuk, A. S. Sverstiuk, S. I. Klymnyuk, D. S. Pyvovarchuk, Y. B. Palaniza, Approach to prediction and receiver operating characteristic analysis of a regression model for assessing the severity of the course Lyme borreliosis in children, Rheumatology 61.5 (2023) 345–352. doi:10.5114/reum/173115.

[6] M. Herasymiuk, A. Sverstiuk, I. Kit, Multifactor regression model for prediction of chronic rhinosinusitis recurrence, Wiadomosci Lek. 76.5 (2023) 928–935. doi:10.36740/wlek202305106.

[7] O. Yasniy, I. Didych, D. Tymoshchuk, P. Maruschak, V. Demchyk, Prediction of structural elements lifetime of titanium alloy using neural network, Procedia Struct. Integr. 72 (2025) 181–187. doi:10.1016/j.prostr.2025.08.090.

[8] O. Yasniy, D. Tymoshchuk, I. Didych, R. Zolotyi, V. Tymoshchuk, Modeling of shape memory alloys hysteresis behavior considering the loading cycle frequency, Procedia Struct. Integr. 72 (2025) 188–194. doi:10.1016/j.prostr.2025.08.091.

[9] E. Aguilar Madrid, N. Antonio, Short-Term Electricity Load Forecasting with Machine Learning, Information 12.2 (2021) 50. doi:10.3390/info12020050.

[10] W. Pannakkong, V. T. Vinh, N. N. M. Tuyen, J. Buddhakulsomsiri, A Reinforcement Learning Approach for Ensemble Machine Learning Models in Peak Electricity Forecasting, Energies 16.13 (2023) 5099. doi:10.3390/en16135099.

[11] Tymoshchuk, D., Yasniy, O., Mytnyk, M., Zagorodna, N., Tymoshchuk, V. Detection and classification of DDoS flooding attacks by machine learning method. CEUR Workshop Proceedings, 2024, 3842, pp. 184 – 195

[12] B. Lypa, I. Horyn, N. Zagorodna, D. Tymoshchuk, T. Lechachenko, Comparison of feature extraction tools for network traffic data, CEUR Workshop Proceedings, 3896, 2024, pp. 1-11.

[13] H. K. P. Katamreddi, T. K. Battula, A hybrid approach for machine learning based beat classification of ECG using different digital differentiators and DTCWT, Comput. Biol. Med. 194 (2025) 110426. doi:10.1016/j.compbiomed.2025.110426.

[14] K.J.K., Sree Janani, K. Janani Kuralnatham; R.S., Sabeenian, R. S., Machine Learning-based ECG Classification using Wavelet Scattered Features, AIUB J. Sci. Eng. (AJSE) 23.2 (2024) 168–176. doi:10.53799/ajse.v23i2.821.

[15] A. Mohapatra, A. Dastidar, S. K. Mohapatra, M. N. Mohanty, Abnormal ECG Detection using Optimized Boosting Tree Classifier, in: 2022 OITS International Conference on Information Technology (OCIT), IEEE, 2022. doi:10.1109/ocit56763.2022.00012.

[16] J. Lim, D. Han, M. Pirayesh Shirazi Nejad, K. H. Chon, ECG classification via integration of adaptive beat segmentation and relative heart rate with deep learning networks, Comput. Biol. Med. 181 (2024) 109062. doi:10.1016/j.compbiomed.2024.109062.

[17] D. E. P. Moghaddam, M. Razavi, B. Aazhang, Cardiac Condition Classification Using 12-Lead ECG Recordings, in: 2024 58th Asilomar Conference on Signals, Systems, and Computers, IEEE, 2024, pp. 1349–1353. doi:10.1109/ieeeconf60004.2024.10942637.

[18] N. Baghel, M. K. Dutta, R. Burget, Automatic diagnosis of multiple cardiac diseases from PCG signals using convolutional neural network, Comput. Biomed. 197 (2020) 105750. doi:10.1016/j.cmpb.2020.105750.

[19] D. Xu, Y. Xu, K. Xu, Z. Hu, M. Xing, F. Gini, M. S. Greco, WaveGRU-Net: Robust Non-contact ECG Reconstruction via MIMO Millimeter-Wave Radar and Multi-Scale Semantic Analysis, Signal Process. (2025) 110108. doi:10.1016/j.sigpro.2025.110108.

[20] S. Lamba, S. Kumar, M. Diwakar, A Novel Frequency-Stratified Transformer Framework with Cross-Frequency Attention for Reliable Cardiac Arrhythmia Classification, Knowledge-Based Syst. (2025) 114250. doi:10.1016/j.knosys.2025.114250.

[21] X. Lyu, S. Rani, S. Manimurugan, C. Maple, Y. Feng, A Deep Neuro-Fuzzy Method for ECG Big Data Analysis via Exploring Multimodal Feature Fusion, IEEE Trans. Fuzzy Syst. (2024) 1–13. doi:10.1109/tfuzz.2024.3416217.

[22] M. Gupta, R. Wadhvani, A. Rasool, Deep learning-based real-time diagnosis of cardiac diseases through behavioral changes in ECG signals, Biomed. Signal Process. Control 104 (2025) 107532. doi:10.1016/j.bspc.2025.107532.

[23] P. B. Patil, V. B. Reddy, A. P. S., Prediction of Cardiovascular Diseases by Integrating Electrocardiogram (ECG) and Phonocardiogram (PCG) Multi-Modal Features using Hidden Semi Morkov Model, Int. J. Recent Innov. Trends Comput. Commun. 10.10 (2022) 32–44. doi:10.17762/ijritcc.v10i10.5732.

[24] A. Dua, Emotion Detection Using Machine Learning, ECG Signals and Facial Features, in: 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), IEEE, 2024. doi:10.1109/idciot59759.2024.10467415.

[25] H. K. P. Katamreddi, T. K. Battula, A hybrid approach for machine learning based beat classification of ECG using different digital differentiators and DTCWT, Comput. Biol. Med. 194 (2025) 110426. doi:10.1016/j.compbiomed.2025.110426.

[26] S. Mavaddati, ECG arrhythmias classification based on deep learning methods and transfer learning technique, Biomed. Signal Process. Control 101 (2025) 107236. doi:10.1016/j.bspc.2024.107236.

[27] P. Salas, J.-M. Mejia-Muñoz, R. Gonzalez-Landaeta, Fog-Enabled Multimodal Chest-Worn Device for Systolic Blood Pressure Monitoring, IEEE Access (2025) 1. doi:10.1109/access.2025.3571829.

[28] X. Yang, G. Jiang, Z. Zhu, D. Wu, A. He, J. Wang, MDD2DG-IRA: Multivariate Degree Distribution to Dynamic Graph With Inter-Channel Relevance Attention Mechanism for Multi-Channel Myocardial Infarction ECG Analysis, IEEE J. Biomed. Inform. (2025) 1–12. doi:10.1109/jbhi.2025.3554309.

[29] B. B. Purkayastha, S. Barma, M. J. Saikia, A Resource-Efficient Cardiac Arrhythmia Detection Using Nonlinear Dynamics in Optimized Delay State Networks, IEEE Trans. Biomed. Eng. (2025) 1–13. doi:10.1109/tbme.2025.3605297.

[30] M. N. Mohanty, S. Baliarsingh, P. K. Panda, An Ensemble Technique for Cardiac Data Compression in Smart Healthcare System, SN Comput. Sci. 6.1 (2025). doi:10.1007/s42979-024-03605-7.

[31] L. M. Lopez-Ramos, Future Perspectives on Automated Machine Learning in Biomedical Signal Processing, in: Communications in Computer and Information Science, Springer International Publishing, Cham, 2022, pp. 159–170. doi:10.1007/978-3-031-10525-8_13.

[32] V. Malhotra, M. Sandhu, Improved ECG based Stress Prediction using Optimization and Machine Learning Techniques, ICST Trans. Scalable Inf. Syst. (2018) 169175. doi:10.4108/eai.6-4-2021.169175.

[33] L. Mosiy, A. Sverstiuk. Information technology for electrocardiographic signal analysis based on mathematical models of temporal and amplitude variability. Comput. Syst. Inf. Technol. (2) (2025) 36–44. doi:10.31891/csit-2025-2-4

[34] J. Zheng, H. Guo, H. Chu. A large scale 12-lead electrocardiogram database for arrhythmia study (Version 1.0.0), PhysioNet (2022). doi:10.13026/WGEX-ER52

[35] D.S. Baim, W.S. Colucci, E.S. Monrad, H.S. Smith, R.F. Wright, A. Lanoue, D.F. Gauthier, B.J. Ransil, W. Grossman, E. Braunwald. Survival of patients with severe congestive heart failure treated with oral milrinone. J. Am. Coll. Cardiol. 7(3) (1986) 661–670. doi:10.1016/s0735-1097(86)80478-8

[36] P. Laguna, R.G. Mark, A. Goldberg, G.B. Moody, A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG, in: Computers in Cardiology 1997, IEEE. doi:10.1109/cic.1997.648140

[37] Automated Machine Learning (AutoML) Search — EvalML 0.84.0 documentation. URL: https://evalml.alteryx.com/en/stable/user_guide/automl.html.

[38] Classification performance metrics and indices. URL: https://adriancorrendo.github.io/metrica/articles/available_metrics_classification.html.

[39] GitHub - shap/shap: A game theoretic approach to explain the output of any machine learning model. URL: https://github.com/shap/shap.

[40] ExtraTreesClassifier. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html.