



## Minirocket Kullanarak Güçlendirilmiş ve Verimli Atriyal Fibrilasyon Tespiti

### Robust and Efficient Atrial Fibrillation Detection from Intracardiac Electrograms Using Minirocket : A Comparative Study with Machine Learning Algorithms

Celal Alagoz<sup>1</sup>

<sup>1</sup>Kırıkkale, TÜRKİYE

Başvuru/Received: 25/12/2023

Kabul / Accepted: 16/01/2024

Çevrimiçi Basım / Published Online: 31/01/2024

Son Versiyon/Final Version: 31/01/2024

#### Öz

Atriyal Fibrilasyon (AF) tespiti, intrakardiyak Elektrogram (EGM) sinyallerinin kritik bir yönüdür ve kardiyovasküler sağlık izlemesinin önemli bir parçasını oluşturur. Bu çalışma, güçlendirilmiş ve verimli AF tespiti için bir zaman serisi sınıflandırma (TSC) algoritması olan Minirocket'ın uygulanmasını keşfeder. Rodrigo et al. (2022) tarafından elde edilen veri setinin bir alt kümesi üzerinde karşılaştırmalı bir analiz gerçekleştirilir. Çalışma, Minirocket'ın kısa EGM dizileri ve değişen eğitim büyüklükleri karşısındaki direncini araştırır; bu, giyilebilir ve implante edilebilir cihazlar gibi gerçek dünya uygulamaları için önemlidir. Ampirik çalışma süresi analizi, Minirocket'ın geleneksel makine öğrenimi algoritmalarına kıyasla verimliliğini değerlendirir. Sonuçlar, özellikle kısa sinyaller ve değişen eğitim büyüklükleri senaryolarında Minirocket'ın dikkate değer performansını sergiler, böylece geleceğin kardiyovasküler izleme teknolojilerinde AF tespiti için umut vadeden bir aday olarak öne çıkar. Bu araştırma, AF tespit algoritmalarının verimliliğinin artırılması ve dinamik klinik senaryolara adapte edilmesi konusunda katkıda bulunur.

#### Anahtar Kelimeler

*"Atriyal fibrilasyon, İntrakardiyak elektrogram analizi, Makine öğrenimi, Zaman serisi sınıflandırması"*

#### Abstract

Atrial fibrillation detection from intracardiac electrogram signals is a critical aspect of cardiovascular health monitoring. This study explores the application of Minirocket, a time series classification algorithm, for robust and efficient atrial fibrillation detection. Comparing it with standard machine learning algorithms, including Extreme Gradient Boosting and Random Forest, using electrograms from 44 patients, Minirocket surpasses both traditional and deep learning approaches with F1 scores around 0.98 for 4 seconds signal duration. The algorithm demonstrates robustness in data-limited scenarios. Additionally canonical feature ablation is implemented and it emphasizes the nuanced importance of temporal features, with distribution-related features proving detrimental. Minirocket's efficiency, adaptability to varying signal sizes, and competitive accuracy position it as a promising tool for real-time AF detection, addressing key challenges in emerging cardiovascular monitoring technologies. This research contributes to the optimization of AF detection algorithms for increased efficiency and adaptability to dynamic clinical scenarios.

#### Key Words

*"Atrial fibrillation, Intracardiac electrograms analysis, Machine learning, Time series classification"*

## 1. Introduction

Atrial Fibrillation (AF), recognized by rapid and irregular heartbeats, is a cardiac arrhythmia posing not only a substantial lifetime risk factor but also significant health risks, including stroke and heart failure [Weng et al. (2018)]. It falls within the broader category of supraventricular tachycardia (SVT) and can coexist with other SVT types, such as Atrial Tachycardia (AT) and Atrial Flutter (AFL) [Haissaguerre et al. (1994), Katritsis et al. (1996), Palma et al. (2000), Jang et al. (2010)]. The timely and accurate identification of AF is imperative for effective management, preventing complications, and holds crucial clinical implications, influencing the utilization of pacemakers, defibrillators, and the interpretation of remote-monitoring tracings from implantable devices.

As data analysis techniques continue to evolve, extensive studies focus on cardiac electrophysiological signals to enhance data interpretation and expedite diagnostic decision-making. Electrocardiograms (ECG) are commonly used for detecting cardiac disorders due to their accessibility and low cost. Recent studies, exemplified by Smith et al. (2019) and Hannun et al. (2019), showcase the efficacy of deep learning methods in AF detection, surpassing traditional ECG interpretation by physicians. Despite the widespread availability of ECGs, their data collection from the body surface provides coarser data when measuring tissue-level electrical activities.

In contrast, EGM analysis, obtained through direct contact with cardiac tissue, offers a refined understanding of fibrillatory activities. This type of analysis facilitates improved mapping and navigation for therapeutic actions such as ablation. However, due to the demanding and invasive nature of EGM collection, simulated EGMs are often used for regularity and organization analysis. Earlier studies employed conventional ML and signal processing techniques, while recent endeavors embraced deep learning for EGM analysis. Nevertheless, in the realm of AF treatment, the finer-level EGM analysis benefits significantly from a coarser-level interpretation, especially since AF may coexist with other SVT types and exhibit diverse characteristics in various atrial regions.

While few studies have analyzed intracardiac EGMs for AF detection, including the work by Isa et al. (2007) and Rodrigo et al. (2022), the latter's use of deep learning techniques faced intensive computational demands. This current study builds upon Rodrigo et al. (2022) and seeks to explore the effectiveness and efficiency of state-of-the-art time series classification (TSC) algorithms, with a specific focus on Minirocket. The motivation arises from the necessity for efficient algorithms capable of real-time AF detection in wearable and implanted devices. This is particularly crucial considering the varying characteristics of AF episodes and different atrial regions.

### 1.1. Motivation

The motivation behind this study arises primarily from the intricate nature of AF, which may coexist with other SVT types and exhibit diverse characteristics. The increasing prevalence of wearable and implanted devices, coupled with ambulatory monitoring technologies, underscores the need for efficient algorithms capable of processing shorter EGM sequences and adjusting to varying training sizes to enable real-time AF detection. Secondly, while deep learning models are recognized for their high performance accuracy, they are resource intensive often requiring substantial amounts of data and are prone to longer training times, which may not be readily available in many scenarios. Conversely, standard Machine Learning (ML) algorithms, while efficient, struggle with the non-stationary, non-local, and non-monotonic properties inherent in time series data. This study specifically focuses on Minirocket due to its tailored design for time series data, minimal computational complexity, seamless implementation without the need for extensive pre-processing steps or feature extraction, minimal parameter tuning requirements, and superior adaptability to diverse signal sizes. By testing Minirocket alongside conventional ML algorithms, which is a novel inclusion in the literature, this study aims to demonstrate its superior accuracy and runtime efficiency compared to the current state-of-the-art methods. The overarching goal is to comprehensively assess the performance, efficiency, and robustness of Minirocket in the context of AF detection.

### 1.2. Objectives

**Algorithmic Comparison:** Evaluate the performance of Minirocket in AF detection and compare it with conventional ML algorithms.

**Robustness Analysis:** Investigate the robustness of Minirocket in the face of shorter EGM sequences and varying training sizes, reflecting real-world scenarios and the demands of streaming data.

**Efficiency Considerations:** Conduct an empirical runtime analysis to evaluate the efficiency of Minirocket concerning other conventional ML algorithms.

### 1.3. Contribution

This study contributes to the evolving field of AF detection by providing insights into the effectiveness of Minirocket as a TSC algorithm. The focus on robustness in the context of short-duration signals and varying training sizes addresses practical challenges in the deployment of AF detection algorithms, particularly in the context of emerging technologies.

### 1.4. Structure of the Paper

The remainder of this paper is organized as follows: Section 2 provides a comprehensive overview of the methodology, including data collection, preprocessing, and classification algorithms. Section 3 presents the results of the experiments, highlighting the performance, robustness, and efficiency of Minirocket. Section 4 offers a detailed discussion of the findings and their implications. Finally, Section 5 concludes the paper with a summary of key insights and suggestions for future research.

#### Terminology

AF	Atrial Fibrillation
AT	Atrial Tachycardia
AFL	Atrial Flutter
SVT	Supraventricular Tachycardia
EKG	Electrogram
ML	Machine Learning
TSC	Time Series Classification
Rocket	Random Convolutional Kernel Transform
Minirocket	Minimally Random Convolutional Kernel Transform
XGB	eXtreme Gradient Boosting
RF	Random Forest
LDA	Linear Discriminant Analysis
GNB	Gaussian Naïve Bayes
catch22	Canonical Time Series Characteristics
PPV	Positive Predicted Value

## 2. Materials and Methods

### 2.1. Patient Population

The dataset utilized in this investigation was generously provided by Rodrigo et al. (2022). The patient cohort was meticulously curated from the COMPARE registry (NCT02997254), comprising individuals afflicted with AF who were prospectively enrolled during ablation procedures for symptomatic AF that remained refractory to at least one anti-arrhythmic medication. In this registry, each patient underwent the recording of intracardiac EGMs via multipolar 64-pole basket catheters. A thorough examination of the registry was conducted by a panel of three cardiac electrophysiologists, who meticulously categorized each EGM tracing as either AF or atrial tachycardia AT.

In this study, a systematic selection of patients from the registry was conducted to create a balanced dataset of intracardiac recordings, encompassing both AF (N = 22) and AT (N = 22). It is important to highlight that all participants in this study provided written informed consent, adhering to protocols approved by the Human Research Protection Program. Patient demographics is summarized in Table 1.

**Table 1.** The demographic information of the patients participating in the study

	Patients (all)	Patients (AF)	Patients (AT)
Patient count	44	22	22
Age (in years)	59.2±11.5	59.7±12	58.7±11.1
Female	14(31%)	7(31%)	7(31%)
Weight	94.9±17.9	97.6±17.8	92.1±17.8

### 2.2. EGM Reading and Recording

The dataset utilized in this study was a subset of the dataset provided by Rodrigo et al. (2022). Rodrigo et al. included EGMs from 86 patients (25 Female, 65±11 years), while this study focused on EGMs from 44 patients (14 Female, 59±11 years). The entire dataset comprised N=29,340 EGM signals, whereas this study analyzed N=2,817 EGM signals. EGM collection and pre-processing step is summarized as follows. The process of EGM collection and pre-processing is summarized as follows.

The electrophysiology study commenced following the discontinuation of antiarrhythmic medications for a duration equivalent to 5 half-lives. Utilizing a 64-pole basket catheter (Abbott, Menlo Park, CA; electrode size 2 mm, inter-electrode spacing 5 mm along spline), both right and left atria were comprehensively mapped. Experienced operators skillfully manipulated the catheters to optimize contact, as previously described by Honarbakhsh et al. (2017). Unipolar EGMs spanning 60 seconds were exported from the electrophysiological recorder (Prucka, GE Marquette, Milwaukee, WI; Bard Electrophysiology, Billerica, MA) and filtered within the range of 0.05–500 Hz.

For analysis, unipolar EGMs with a duration of 4000 ms, equivalent to approximately 20 cycles of AF or AT, were employed. This duration aligns with the common practice of analyzing EGM sequences in the frequency domain, with extended durations offering negligible improvement in rhythm identification. Original EGMs had a sample frequency of 1 kHz (Bard) or 977 Hz (Prucka) and were uniformly resampled for comparative analyses between datasets. To reduce dimensionality and considering that the physiological content of AF and AT EGMs is below 200 Hz [Rodrigo et al. (2021)], EGMs were downsampled to 400 Hz using a 200 Hz anti-aliasing filter. Consequently, the findings are applicable to systems with sampling frequencies greater than 400 Hz. Ventricular artifacts were mitigated by subtracting the mean QRS complex, identified in three orthogonal ECG leads through a voltage threshold, and averaged across 1 minute [Alhousseini et al. (2020)].

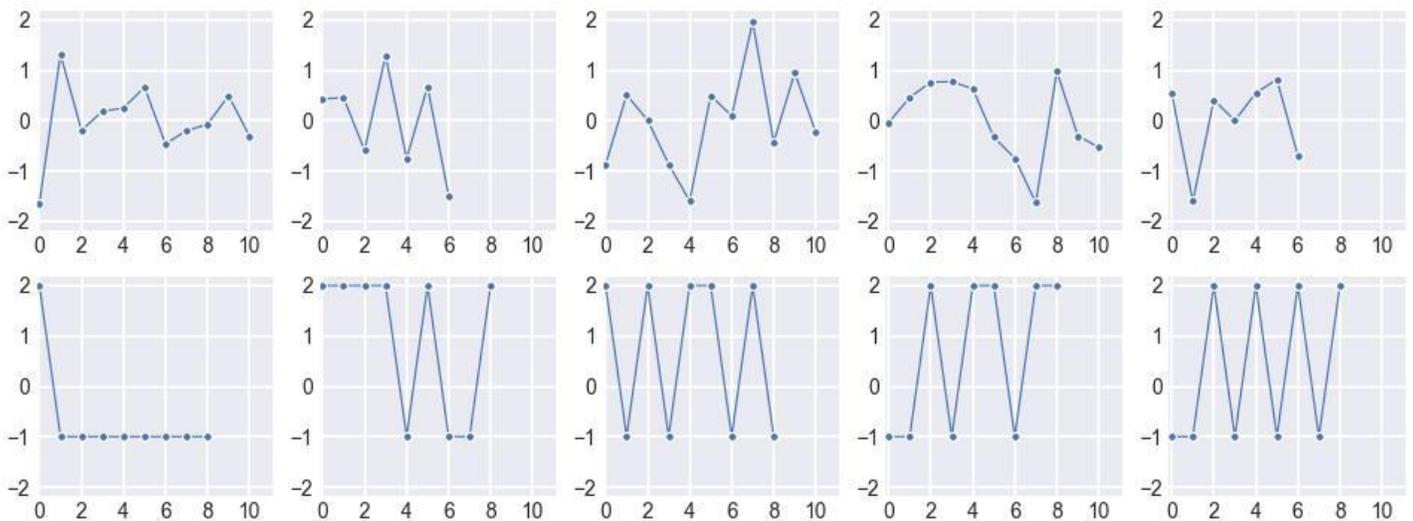
**2.3. Rocket and Minirocket**

Classifiers specialized in handling time series data are pivotal for tasks involving time-dependent information, including trend prediction, anomaly detection, and the categorization of sequential data. Prominent time series classifiers in the current landscape encompass The Hierarchical Vote Collective of Transformation-based Ensembles 2.0 (HIVE-COTEv2), introduced by Middlehurst et al. (2021), and the Time Series Combination of Heterogeneous and Integrated Embedding Forest (TS-CHIEF), developed by Shifaz et al. (2020). These algorithms demonstrated exceptional performance using 112 time series datasets from Dau et al. (2019). However, their extensive training times render them impractical for scenarios where efficiency is a critical concern.

In contrast, Dempster et al. (2020a) introduced Rocket, a significantly faster algorithm that yields comparable results to the aforementioned classifiers. Rocket can be conceptualized as an adaptation of the convolution technique employed in Convolutional Neural Networks (CNNs) for time series data. Key distinctions between Rocket and CNN include the absence of learned kernel weights in Rocket, which employs a vast number of randomly generated kernels. Implementation details are provided in Dempster et al. (2020). The transformation process can be described as follows:

$$X_i * \omega = \left( \sum_{j=0}^{l_{kernel}-1} X_{i+(j*d)} \omega_j \right) + b \tag{1}$$

Where  $X_i$  is  $i$ th instance of time series dataset,  $\omega$  is randomly generated kernel set -there are by default 10K of them-,  $l_{kernel}$  is kernel length,  $d$  is diation, and  $b$  is bias. Following the transformation, twice the number of features per time series instance is generated. Subsequently, a linear classifier, Ridge classifier, is employed for classification. In scenarios involving extensive datasets, Logistic Regression (LR) is employed, utilizing stochastic gradient descent for training. This choice is motivated by its capability to converge efficiently, potentially achieving convergence within a single iteration or even without fully utilizing all available training data. This strategy aligns with the recommendations of LeCun et al. (2015) and Bottou et al. (2018). In this study, Ridge classifier consistently outperformed LR in terms of accuracy and speed, making LR an excluded option.



**Figure 1.** Displayed above are exemplary kernels from Rocket (top) and Minirocket (bottom). The x-axis represents the kernel length, while the y-axis illustrates the kernel weights. A noticeable reduction in randomness occurs when transitioning from Rocket to Minirocket. This is evident in the fixed kernel lengths of Minirocket compared to the varying kernel lengths of Rocket. Additionally, the kernel weights of Rocket span real values around zero, ranging between -2 and 2, whereas the kernel weights of Minirocket are confined to -1 or 2.

To enhance the accuracy performance of Rocket, Minirocket was introduced by Dempster et al. (2020). Minirocket not only outperformed Rocket but also exhibited greater efficiency. Minirocket employs almost deterministic kernels, as opposed to the random characteristics of length, weights, bias, and dilation parameters in Rocket. The selection criterion focuses on discarding the maximum value, with only the Positive Predictive Value (PPV) chosen as the feature after convolution. Changes in convolutional kernel properties are summarized in Table 2, and example kernels are visualized in Figure 2.

**Table 2.** Comparing convolutional kernel properties of Rocket and Minirocket

	<i>Rocket</i>	<i>Minirocket</i>
length, $l_{\text{kernel}}$	{7,9,11}	9
weights, $W$	$\mathcal{N}(0,1)$	{-1,2}
bias, $b$	$\mathcal{U}(0,1)$	from convolution output
dilation, $d$	random	fixed (rel. to input length)
padding	random	length)
features	PPV+max	fixed
num. features	20K	PPV 10K

Minirocket achieves a marked acceleration in transformation through four key optimizations:

- Simultaneously calculating PPV for  $W$  and  $-W$ .
- Repurposing the convolution output to compute various features.
- Eliminating multiplications in the convolution operation.
- Efficiently computing all kernels (almost) 'at once' for each dilation.

Despite these changes, the theoretical computational complexity of Minirocket remains comparable to Rocket. However, empirical runtime observations indicate a significant speed improvement. Following transformation with Minirocket, the same number of features as the number of kernels are generated per time series instance, given that only PPV is selected as a feature. Similar to Rocket, the Ridge Classifier is employed for classification, with the default number of kernels being 10,000 for both Rocket and Minirocket. In this study, different values are tested to explore their impact on classification performance..

#### 2.4. Features Extracted from EGMs

In the analysis of time series data, particularly cardiac electrophysiological signals, the extraction of various features is essential. The Canonical Time Series Characteristics (catch22), introduced by Lubba et al. (2019), represents a collection of 22 time series features specifically tailored for time-series analysis. These features were thoughtfully curated from the extensive pool of 7000+ features present in the Highly Comparative Time Series Analysis (hctsa) toolbox [Fulcher et al. (2013), Fulcher et al. (2017)]. The hctsa toolbox is a comprehensive resource for time-series analysis, offering versatility in its application based on specific analytical goals.

In order to distill a more streamlined set of canonical features, a hierarchical clustering technique was applied to the correlation matrix of features that exhibited superiority over random chance. The clusters generated were subsequently organized based on balanced accuracy, assessed through a decision tree classifier. Within each of the 22 formed clusters, a singular feature was chosen, considering balanced accuracy outcomes, computational efficiency, and interpretability.

The catch22 features span a diverse array of characteristics, encompassing distributional properties, regularity, complexity, and more. A careful selection of features from the hctsa toolbox, along with brief descriptions, is presented in Table 3.

**Table 3.** The *catch22* feature set [adopted from Lubba et al. (2019)]

<i>hctsa</i> feature name	Description
<i>Distribution</i>	
DN_HistogramMode_5	The distribution's mode represented by a 5-bin histogram of z-scores.
DN_HistogramMode_10	The distribution's mode represented by a 10-bin histogram of z-scores.
<i>Simple temporal statistics</i>	
SB_BinaryStats_mean_longstretch1	The most extended duration of successive values surpassing the mean.
DN_OutlierInclude_p_001_mdrmd	Intervals of time between consecutive occurrences of extreme events above the mean

---

DN_OutlierInclude_n_001_mdrmd	Intervals of time between consecutive occurrences of extreme events below the mean
<i>Linear autocorrelation</i>	
CO_flecac	The point at which the autocorrelation function crosses 1/e for the first time
CO_FirstMin_ac	The initial occurrence of the minimum value in the autocorrelation function
SP_Summaries_welch_rect_area_5_1	Cumulative power within the lowest quintile of frequencies in the Fourier power spectrum
SP_Summaries_welch_rect_centroid	Centroid of the power distribution in the Fourier power spectrum
FC_LocalSimple_mean3_stderr	Average deviation from a rolling 3-sample mean forecast
<i>Nonlinear autocorrelation</i>	
CO_trev_1_num	Statistic indicating time-reversibility, $\langle (x_{t+1} - x_t)^3 \rangle t$
CO_HistogramAMI_even_2_5	Automutual information, $m = 2, \tau = 5$
IN_AutoMutualInfoStats_40_gaussian_fmml	The point of first minimum in the automutual information function
<i>Successive differences</i>	
MD_hrv_classic_pnn40	The ratio of successive differences surpassing $0.04\sigma$ [Mietus et al. (2002)]
SB_BinaryStats_diff_longstretch0	The longest duration of consecutive incremental decreases
SB_MotifThree_quantile_hh	Shannon entropy of two successive letters in equiprobable 3-letter symbolization
FC_LocalSimple_mean1_ttauresrat	Change in correlation length following iterative differencing
CO_Embed2_Dist_tau_d_expfit_meandiff	Exponential fitting applied to consecutive distances in a 2-dimensional embedding space
<i>Fluctuation Analysis</i>	
SC_FluctAnal_2_dfa_50_1_2_logi_prop_r1	Fraction of slower timescale fluctuations exhibiting scaling behavior according to detrended fluctuation analysis (50% sampling)
SC_FluctAnal_2_rsrangeft_50_1_logi_prop_r1	Fraction of slower timescale fluctuations exhibiting scaling behavior according to linearly rescaled range fits
<i>Others</i>	
SB_TransitionMatrix_3ac_sumdiagcov	Sum of the diagonal elements of the covariance matrix for the transition probabilities between symbols in a 3-letter alphabet
PD_PeriodicityWang_th0_01	Periodicity measure of Wang et al. (2007)

---

The "catch22" library has been implemented in Python, offering user-friendly functionality and facilitating seamless integration into Python workflows. This library proves to be a valuable tool for extracting relevant features from EGMs, contributing to a comprehensive analysis of cardiac electrophysiological signals.

For TSC algorithms, no feature extraction is carried out, while conventional ML algorithms involve feature extraction. Nonetheless, the classification task using ML algorithms is also conducted without feature extraction, resulting in significantly inferior performances compared to utilizing catch22 features. Consequently, feature extraction is by default incorporated when employing ML algorithms for the remainder of the paper.

## 2.5. Conventional Classifiers

In the selection of conventional ML algorithms for the classification task, considerations include ease of use and broad applicability. The selected models and their concise descriptions are provided below:

- **eXtreme Gradient Boosting (XGB):** XGB is a widely adopted open-source machine learning library designed for supervised learning tasks, particularly in structured/tabular data and regression problems. It leverages gradient boosting, constructing an ensemble of weak learners, often decision trees, sequentially. Incorporating L1 (Lasso) and L2 (Ridge) regularization terms, XGBoost mitigates overfitting by penalizing complex models. Noteworthy features include "tree pruning" for model simplification and built-in cross-validation for hyperparameter tuning. XGBoost provides a feature importance score, aiding in feature selection and model interpretability.
- **Random Forest (RF):** RF is an ensemble learning method applicable to both classification and regression tasks. Comprising multiple decision trees, each trained on a random subset of the data, it amalgamates their predictions for enhanced accuracy and generalization. Bootstrap sampling and feature randomization introduce randomness, mitigating overfitting. RF offers simplicity, flexibility, and robustness, and its predictions are determined through a majority vote (classification) or averaging (regression) across individual trees. Feature importance is measured based on contributions to impurity reduction.
- **Linear Discriminant Analysis (LDA):** LDA is a statistical technique employed for dimensionality reduction and classification. Particularly useful in supervised learning scenarios, LDA aims to maximize the separation between classes. It transforms original features into linear combinations, emphasizing class discrimination. LDA is distinct from Principal Component Analysis (PCA) as it prioritizes class separation. Assuming normal distribution and a common covariance matrix within each class, LDA computes eigenvalues and eigenvectors to establish a linear decision boundary. LDA is robust to outliers compared to some other classification methods.
- **Gaussian Naive Bayes (GNB):** The GNB classifier is a probabilistic algorithm based on Bayes' theorem, assuming Gaussian distribution of features. Despite its "naive" assumption of feature independence, GNB performs effectively, especially with high-dimensional datasets. It estimates class probabilities based on the likelihood of features given a class, assuming feature independence within each class. Parameters such as mean and variance for each feature in each class are estimated during training. GNB is particularly suitable for scenarios where the normal distribution assumption holds.

All classifiers are implemented with default parameters, and hyperparameter tuning is not performed in this study. The selected classifiers offer a blend of simplicity, efficiency, and performance across diverse classification tasks

## 2.6. Classification Metrics and Evaluation

Monte Carlo cross-validation, a technique involving the random partitioning of the dataset into training and testing sets across multiple iterations, is employed for robust cross-validation. To ensure reproducibility of training and testing sets, the shuffling seed at each iteration is set to the iteration number. In this experimental setup, various training and testing sizes are explored to comprehensively evaluate classifier performance.

Cross-Validation Process:

- Monte Carlo Cross-Validation: The dataset is randomly split into training and testing sets for multiple iterations. The shuffling seed, determined by the iteration number, ensures the reproducibility of the training and testing sets.

Evaluation Metrics:

- Accuracy: This metric quantifies the overall correctness of the classifier by measuring the ratio of correctly predicted instances to the total instances. It provides an intuitive assessment of the classifier's performance.
- F1 Score: The F1 score is a metric that balances precision and recall, offering a single numerical measure of a classifier's performance. It is particularly useful in situations where there is an imbalance between classes, providing a comprehensive evaluation of both false positives and false negatives.

The choice of these metrics aims to provide a well-rounded understanding of classifier performance, capturing both the accuracy of predictions and the ability to deal with imbalances in class distribution. The utilization of Monte Carlo cross-validation enhances the reliability and robustness of the evaluation process. For the sake of consistency in comparison, the random state seed for all classifiers is set to the same number, which is 42.

## 3. Results

This section presents a comparative analysis of the classification and runtime performance of TSC algorithms, including Rocket and Minirocket, in contrast to conventional Machine Learning ML algorithms (XGB, RF, LDA, and GNB) for the purpose of EGM signal analysis to detect AF. Additionally, an XGB-specific feature ablation study is conducted.

### 3.1. Robustly Identifying AF and AT from EGMs

To assess the robustness of identifying AF and AT from intracardiac EGM signals, an analysis is conducted with varying EGM signal durations and training sizes. Signal durations are systematically reduced from 4 seconds to 2.5, 1, 0.5, and 0.25 seconds, while training sizes are decreased from 80% to 60%, 40%, and 20%.

#### *Accuracy score performances for varying durations*

In the initial experiment, classification performance is evaluated based on accuracy, and the results for different EGM durations with a fixed training size of 80% are presented in Table 4. Notably, Minirocket and Rocket, both TSC algorithms, emerged as the top-performing classifiers. For a 4-second EGM duration, they achieved an accuracy performance of approximately 0.98. As the duration was reduced, their performance declined, yet even for a 1-second duration, they maintained a performance level of around 0.92, which can still be deemed a successful classification. In the case of 0.5 and 0.25 seconds durations, their performance was more modest but still exceeded 0.80.

**Table 4.** Accuracy performance of classifiers for different EGM durations.

	4 s	2.5 s	1 s	0.5 s	0.25 s
Minirocket	<b>0.9803±0.0044</b>	<b>0.9679±0.0076</b>	0.9232±0.0043	0.8706± 0.0067	<b>0.8066± 0.0151</b>
Rocket	0.9800±0.0041	0.9652±0.0067	<b>0.9333±0.0105</b>	<b>0.8738± 0.0147</b>	0.7915± 0.0118
XGB	0.9239±0.0070	0.8957±0.0117	0.8383±0.0169	0.7727± 0.0142	0.7071± 0.0167
RF	0.9071±0.0099	0.8755±0.0110	0.8349±0.0089	0.7707± 0.0132	0.7144± 0.0112
LDA	0.7234±0.0122	0.7129±0.0151	0.6773±0.0170	0.6628± 0.0153	0.6553± 0.0186
GNB	0.6520±0.0176	0.6560±0.0117	0.6401±0.0152	0.6216± 0.0144	0.6197± 0.0200

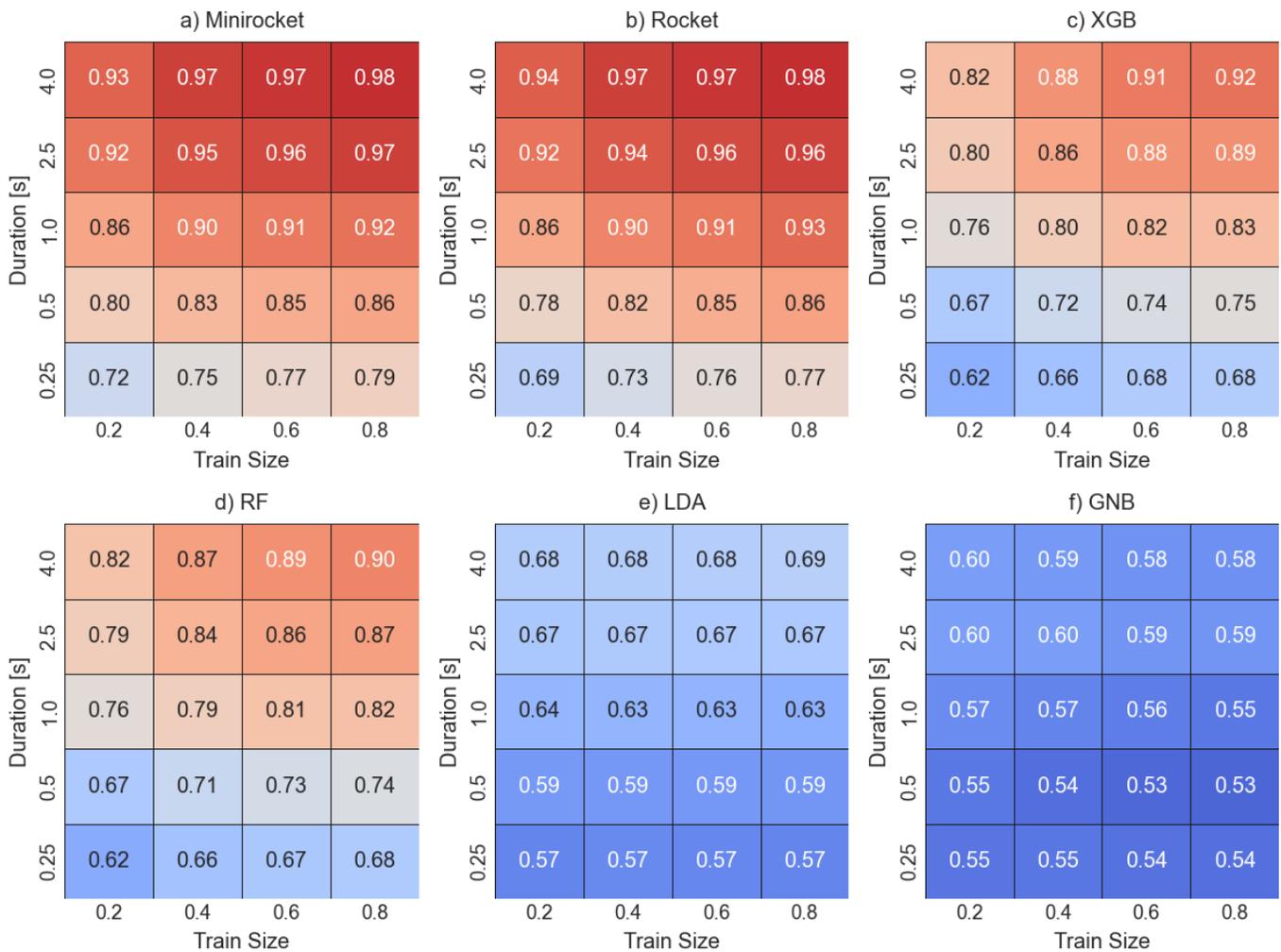
Among conventional ML algorithms, XGB and RF demonstrated relatively moderate performances, with XGB outperforming RF for longer durations. Their accuracy values were around 0.90 for 4- and 2.5-second EGM durations, decreasing to around 0.70 for a 0.25-second duration. LDA and GNB exhibited lower performances, with LDA consistently around 0.70 and GNB around 0.65.

#### *F1 score performances for varying durations and training sizes*

Figure 2 illustrates performance changes concerning different training sizes and signal durations, considering mean F1 scores as the performance measure. Minirocket consistently yielded mean F1 scores exceeding 0.95 for training sizes of 40% and above and durations of 2.5 seconds and above. Performance values remained above 0.90 for durations of 1 second or more, with training sizes of 40% and above, as well as for training sizes of 20% and durations of 2.5 seconds and above. The values hovered around 0.85 for a 0.5-second duration and training sizes of 40% and above, as well as for training sizes of 20% and durations of 1 second. For a 0.25-second duration, performance slightly declined below 0.80 but never fell below 0.70. Rocket's performance was generally close to Minirocket, with Minirocket consistently exhibiting slightly better performance.

The performance of XGB was approximately 0.90 for a training size of 40% and above and a duration of 4 seconds, and for a training size of 60% and above for a duration of 2.5 seconds. The values were around 0.85 for a training size of 40% and a duration of 2.5 seconds. Performance values remained above 0.80 for a duration of 1 second and a training size of 40% and above, as well as for a training size of 20% and a duration of 2.5 seconds and above. The values were around 0.75 for a duration of 0.5 seconds and a training size of 40% and above, and for a training size of 20% and a duration of 1 second. For a duration of 0.25 seconds, the performance declined below 0.70, but it never dropped below 0.60. RF values were, in general, slightly lower than those of XGB.

LDA and GNB once again demonstrated poorer performances compared to other classifiers. In the face of varying training sizes and durations, LDA values ranged between 0.69 and 0.57, while GNB values ranged between 0.60 and 0.55. Therefore, they displayed a higher level of robustness to changes in training size.

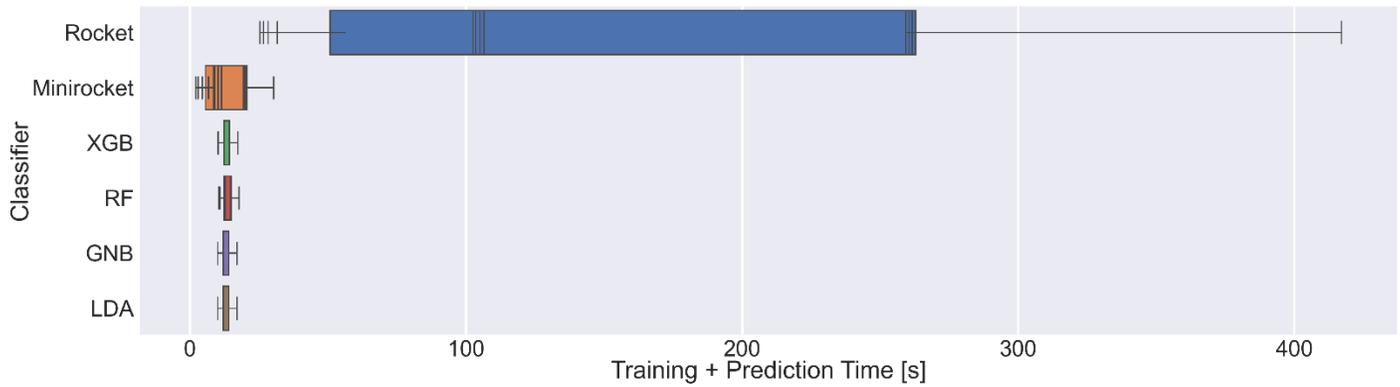


**Figure 2.** A comprehensive overview of the performance of all classifiers under different EGM signal durations and training sizes is depicted in individual heat map plots for each classifier. The assessment of performance is based on the mean of F1 scores obtained from 10 runs of Monte Carlo validation.

### 3.2. Empirical Runtime Analysis

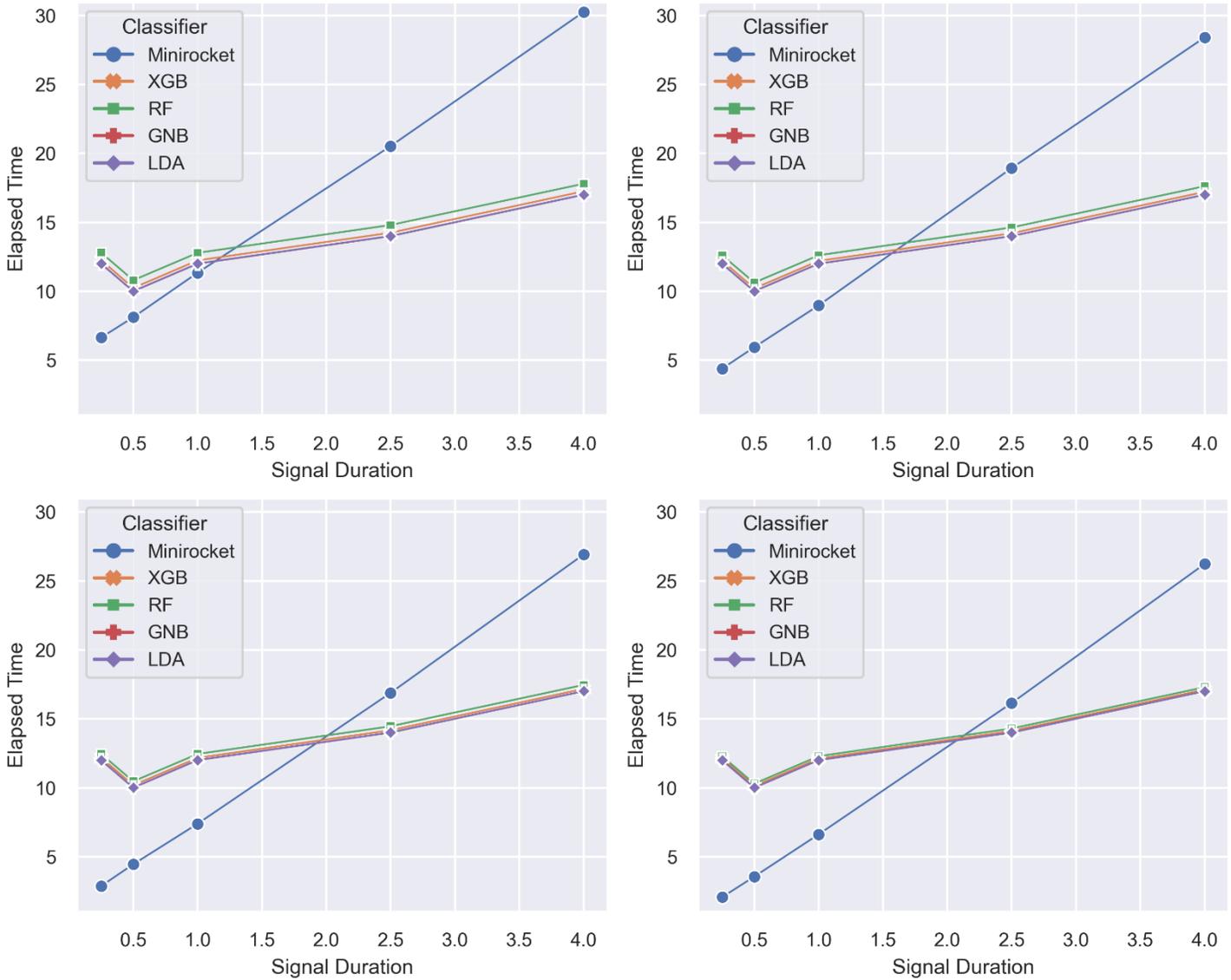
Efficiency is a crucial factor in the analysis of EGM signals, much like any medical signal. Therefore, an empirical runtime analysis is conducted among the examined classifiers. For each configuration, comprising the selection of the classifier, signal duration, and

training size, the elapsed time of each run of Monte Carlo validation is recorded. With ten runs of validation implemented, the median elapsed time among the ten is determined as the elapsed time for that specific configuration. Feature extraction time elapsed during the pre-processing phase is considered for ML algorithms. Depending on the signal duration, feature extraction time took about between 17 and 10 seconds. It is important to note that such preprocessing is not needed for TSC algorithms; hence, elapsed time for those algorithms is constituted only of training and testing stages.



**Figure 3.** Summary of elapsed times for all configurations of the classifiers. x-axis shows the elapsed time in seconds.

Firstly, a summary of elapsed times for all configurations, constituting the selection of EGM signal duration and training sizes, of the classifiers is displayed using a whisker plot in Figure 3. The Rocket classifier stands out as the slowest algorithm among others, roughly ten times slower than all other classifiers. The rest of the classifiers had comparable speed. ML algorithms were strictly around the same line, exhibiting a tight distribution, while Minirocket showed more sensitivity to the variation in configuration, having a relatively broader distribution.



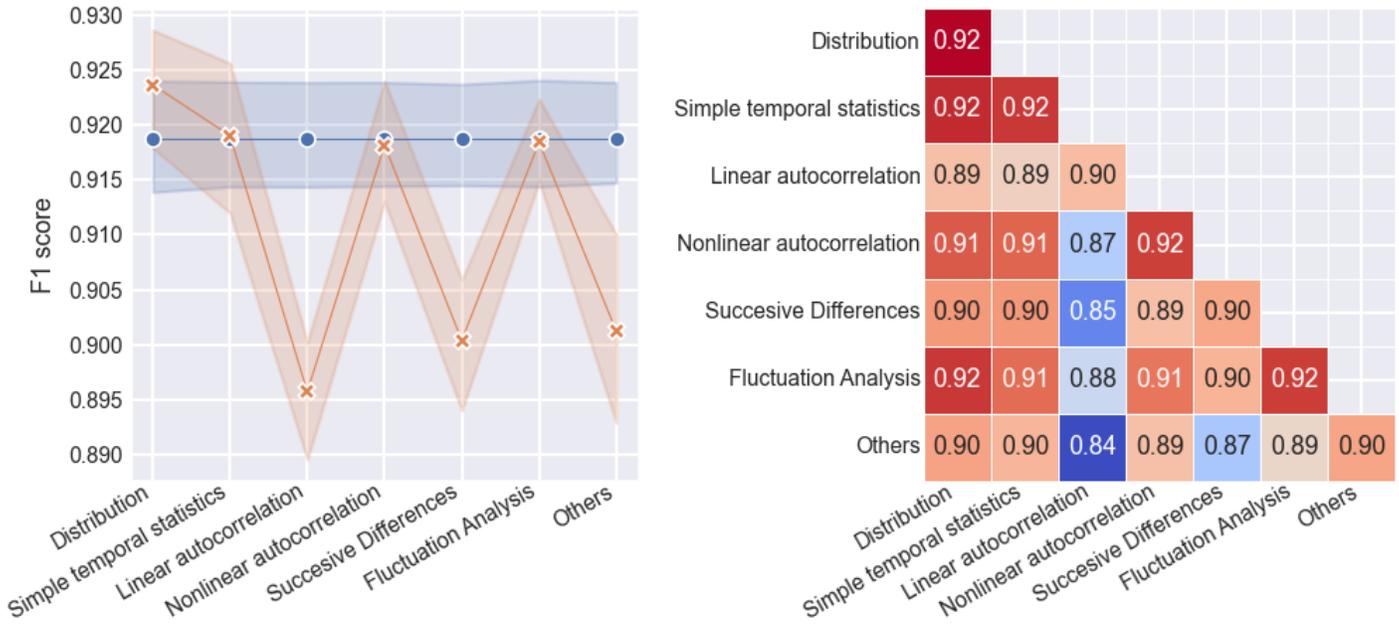
**Figure 4.** The extended elapsed times are presented, showcasing the variation based on EGM signal duration (depicted on the x-axis of each plot) and different training sizes (80% in the top left, 60% in the top right, 40% in the bottom left, and 20% in the bottom right).

Secondly, for a more detailed comparison, elapsed times varying upon EGM signal durations and training sizes are demonstrated in Figure 4. Note that Rocket was excluded from the comparison because it had outstandingly larger runtime values, which could prohibit a finer comparison. The first thing to notice is that Minirocket was slower than other ML algorithms in the case of signal durations of 2.5 seconds and 4 seconds. This difference seems to linearly increase as the signal duration increases. This is because the computational complexity of the convolution operation performed during Minirocket transformation is linear with time series length [Dempster et al. (2020b)]. However, for EGM signal durations less than 1 second, Minirocket was faster than others. As for changing the training size, elapsed time changes linearly with it. For a 1-second EGM signal duration, each time changing the training size from 80% to 60%, to 40%, and to 20%, runtime gets about 1.2 times smaller, signifying a considerable increase in speed. That rate of change in speed slightly increases as the EGM signal duration gets smaller and vice versa. For ML algorithms, for all training sizes, runtime increased almost linearly as EGM signal duration is increased. An exception to that was observed when the duration was changing from 0.25 seconds to 0.5 seconds, where a decrease in runtime is observed. This has to do with the way of working of the catch22 library with time series length. Training size did not have any effect on the runtime. This is because the runtime is virtually dependent on the feature extraction stage, and the number of training and testing sizes does not affect feature extraction because any instance, regardless of being in the training and testing set, goes through feature extraction.

### 3.3. Ablation Study for Canonical Features

An ablation study is conducted to assess the effectiveness of canonical features extracted via the catch22 algorithm. Instead of ablating every single feature among the 22, specific feature groups listed in Table 3 are ablated. The feature groups in question include distribution, simple temporal statistics, linear autocorrelation, nonlinear autocorrelation, successive differences, fluctuation

analysis, and others. The ablation study is performed exclusively for the XGB classifier, with mean F1 scores from Monte Carlo validation serving as the performance criteria. XGB was chosen due to its identification as the best-performing conventional ML algorithm among others in the experimental results of this study.



**Figure 5.** The line plot on the left illustrates the performance change during the ablation of individual feature sets. Ablated feature categories are presented on the x-axis, while on the y-axis, F1 score means are depicted using markers, and the 95% confidence interval is represented by the shaded region. On the right, a heatmap plot demonstrates the performance change during the simultaneous ablation of two sets of features. Both the x and y axes indicate the ablated feature categories. Mean F1 scores are annotated in each cell, representing the intersection of the two ablated feature categories. The diagonals showcase the mean F1 score when the respective category of features is ablated individually. The heat map provides a comprehensive view of the performance dynamics resulting from combined feature set ablations.

*Individual feature set ablations*

Initially, feature groups are systematically ablated one by one, and the change in mean F1 score is presented in Figure 5 (on the left). Ablating distribution-related characteristics, specifically the z-score of histograms with 5-bin and 10-bin, resulted in a slight performance increase from  $0.9188 \pm 0.0077$  to  $0.9236 \pm 0.0092$ . No noticeable change was observed during the ablation of simple temporal statistics, which measure extreme events above and below the mean, as well as nonlinear autocorrelation characteristics, representing automutual information and time reversibility statistics. Similarly, ablation of fluctuation analysis, which measures the fraction of slower time-scale fluctuations, showed unremarkable changes. However, a substantial performance decrease was noted during individual ablations of linear autocorrelation characteristics, including measures from the autocorrelation function and power spectrum of FFT, causing the performance to drop to  $0.8957 \pm 0.0088$ . Additionally, in the case of successive differences and others, a decrease in performance was observed. These categories include features related to difference, 3-letter symbolization statistics, and periodicity measures.

*Pairwise feature set ablations*

Further ablations were conducted by removing two feature sets simultaneously, and the results are depicted in Figure 5 (on the right). The most significant performance decrease was observed when pairs of linear autocorrelation and others, as well as linear autocorrelation and successive differences, were ablated. Both individual and paired ablations consistently highlight that features related to distribution, simple temporal statistics, and fluctuation analysis characteristics might provide redundant, if not detrimental, information for the classification task. Conversely, features associated with linear autocorrelation, successive differences, and others contribute favorably to the task.

**4. Discussions**

The study assessed the robustness of various classifiers in detecting AF and AT from intracardiac EGM signals, considering variations in EGM signal duration and training sizes. Two TSC algorithms, Minirocket and Rocket, were compared with conventional ML algorithms, including XGB, RF, GNB, and LDA. The key findings are discussed below.

#### 4.1. Performance Analysis and Algorithmic Robustness

The examination of classification performance, as depicted in Table 4 and Figure 2, underscores the robustness of Minirocket and Rocket, both TSC algorithms, in identifying AF and AT from intracardiac EGM signals. Notably, Minirocket and Rocket exhibit outstanding accuracy performances, with Minirocket consistently demonstrating a slight edge in performance over Rocket. The fact that Minirocket is significantly less computationally demanding than Rocket positions it as a preferable classifier.

The study by Rodrigo et al. (2002) implemented deep learning techniques, specifically RNN, for the automatic detection of AF from a larger dataset (N=29,340 EGM signals). Their best-performing scheme achieved a mean accuracy of 0.89 and a mean F1 score of 0.96 with unipolar EGMs. In this study, despite using a smaller subsample (N=2,817 EGM signals), Minirocket outperformed the RNN-based approach with a mean accuracy of 0.98 and a mean F1 score of 0.98. It's important to acknowledge the limitation of a strict comparison due to the difference in sample sizes. However, the notable enhancement in performance suggests that Minirocket, being a faster and more efficient algorithm, could serve as a viable alternative to deep learning techniques in specific scenarios. The study's outcomes provide compelling evidence of Minirocket's potential to achieve high accuracy and F1 scores for AF detection, even with a smaller dataset.

The comparative analysis of conventional ML algorithms reveals that XGB outperforms RF, LDA, and GNB, especially for longer EGM signal durations. This superiority aligns with existing literature [Bentéjac et al. (2021)], highlighting the robustness and efficacy of XGB in various classification tasks. However, both XGB and RF exhibit a decline in accuracy as the signal duration decreases, underlining a certain sensitivity to shorter EGM sequences.

The ablation study conducted on canonical features within the XGB framework unveils the nuanced impact of specific feature groups. The removal of linear autocorrelation and other related characteristics leads to a substantial decline in performance, underscoring the pivotal role of these feature categories in the classification of AF. These categories encompass features associated with measures derived from the autocorrelation function, power spectrum of Fast Fourier Transform (FFT), and periodicity measures. This observation aligns with the findings of Rodrigo et al. (2022), where an explainability analysis of features commonly employed in clinical settings for identifying EGM signatures revealed autocorrelation, cycle length, and dominant frequency features as top contributors in distinguishing AF from atrial tachycardia (AT). Additionally, Isa et al. (2007) identified cycle length as a statistically significant feature for distinguishing AF from AT.

Conversely, the ablation of distribution-based features related to the z-score of histograms resulted in a performance increase. Therefore, it becomes apparent that these features contribute detrimentally to the classification task. This observation is logical as distributions do not account for temporal dependence, a crucial aspect in the analysis of EGMs.

#### 4.2. Robustness to Signal Duration and Training Sizes

The investigation into the robustness of Minirocket across varying EGM signal durations and training sizes provides valuable insights into its adaptability in real-world scenarios. The consistent mean F1 scores exceeding 0.95 for training sizes of 40% and above and durations of 2.5 seconds and above highlight Minirocket's stability in diverse conditions. This robustness is especially crucial in scenarios where EGM sequences may vary in length and training data availability is dynamic.

The performance of XGB, while commendable, exhibits a sensitivity to signal duration and training size changes. Notably, XGB maintains accuracy above 0.80 for durations of 1 second or more, emphasizing its suitability for scenarios with slightly shorter EGM sequences. However, the decline in performance for 0.25-second durations indicates a potential limitation in handling extremely short sequences.

#### 4.3. Empirical Runtime Analysis and Computational Efficiency

Efficiency considerations, as evaluated through empirical runtime analysis, shed light on the computational demands of the examined classifiers. The exclusion of Rocket due to significantly larger runtime values emphasizes the practical need for computational efficiency in real-time applications. Minirocket's efficiency, especially for EGM durations less than 1 second, positions it as a promising candidate for scenarios where timely processing is essential.

The linear increase in runtime for Minirocket with longer signal durations aligns with the expected computational complexity of the convolution operation in its transformation. However, the observed decrease in runtime for ML algorithms when transitioning from 0.25 seconds to 0.5 seconds signals highlights the unique considerations in the catch22 library's interaction with time series length.

#### 4.4. Practical Considerations for AF Detection

Our findings emphasize the intricate balance between classification performance and computational efficiency. Minirocket's consistent high performance and efficiency make it a compelling choice for real-time AF detection, particularly in scenarios with shorter EGM sequences. XGB, while robust, may be more suitable for scenarios with slightly longer signal durations, considering its performance trend.

The ablation study underscores the importance of specific feature groups in XGB, guiding future efforts in feature engineering for AF detection. The identified redundancy in certain feature categories suggests avenues for optimization and refinement in feature selection processes.

#### 4.5. Limitations and Future Directions

The representativeness of the dataset and the generalizability of findings to diverse populations warrant attention. Future research could address these limitations by incorporating larger and more diverse datasets, enhancing the external validity of the results.

Additionally, the study's focus on AF detection opens avenues for broader explorations into the classification of various cardiac arrhythmias. Extending the analysis to consider the interplay of different arrhythmia types could provide a more holistic understanding of the algorithms' capabilities and limitations.

#### 4.6. Clinical Implications and Integration

The robust performance of Minirocket in real-world scenarios suggests its potential integration into clinical practice for continuous AF monitoring. The computational efficiency of Minirocket aligns with the demands of wearable and implantable devices, where real-time processing is crucial for timely clinical interventions.

Furthermore, the study's emphasis on algorithmic robustness to varying training sizes contributes to the practical deployment of AF detection algorithms in dynamic healthcare settings. The adaptability of Minirocket to changes in data availability enhances its feasibility for continuous monitoring applications.

#### 4.7. Future Research Directions

Future research directions could explore the integration of Minirocket into existing clinical workflows, conducting prospective studies to validate its performance in diverse patient populations. Collaboration with clinicians and healthcare professionals is essential to bridge the gap between algorithmic advancements and practical clinical applications.

As AI continues to reshape cardiovascular care, the journey towards effective and ethical implementation unfolds. The quest for improved diagnostic accuracy, personalized treatment strategies, and seamless integration into healthcare systems persists, with the potential to revolutionize the landscape of cardiac arrhythmia management.

### 5. Conclusions

In this study, Robust Performance of TSC Algorithms: The evaluation of classification performance highlights the robustness of TSC algorithms, particularly Minirocket and Rocket, in effectively identifying AF and AT from intracardiac EGM signals. Notably, Minirocket consistently outperforms Rocket, showcasing superior accuracy performance and computational efficiency.

In comparison to a deep learning approach implemented by Rodrigo et al. (2002) for AF detection, Minirocket, despite utilizing a smaller dataset, demonstrates higher accuracy and F1 scores. This suggests that Minirocket, being faster and more efficient, could be a promising alternative to deep learning techniques, especially in scenarios with limited data availability.

Standard ML algorithms, particularly XGB, exhibit commendable performance, outperforming RF, LDA, and GNB, especially for longer EGM signal durations. However, both XGB and RF show a decline in accuracy with shorter signal durations, indicating a certain sensitivity to variations in EGM sequence length.

The ablation study for canonical features in XGB reveals the nuanced contribution of specific feature groups. Notably, features related to linear autocorrelation, power spectrum of FFT, and periodicity measures significantly contribute to the classification of AF. This aligns with existing literature emphasizing the importance of autocorrelation measures in capturing temporal dependencies in time series data.

The ablation study highlights that features associated with distribution, particularly z-score of histograms, provide detrimental contributions to AF classification. This aligns with the understanding that distributions do not account for temporal dependence, a crucial aspect in the analysis of EGM signals.

The empirical runtime analysis emphasizes the efficiency of Minirocket, especially for shorter EGM signal durations. While Minirocket exhibits slightly slower performance than other ML algorithms for longer durations, its adaptability to varying signal sizes, low computational complexity, and competitive accuracy make it a favorable choice, particularly for real-time AF detection scenarios.

The study's findings have practical implications for the deployment of AF detection algorithms, especially in the context of emerging technologies and wearable devices. Minirocket's combination of accuracy, efficiency, and adaptability positions it as a promising candidate for real-world applications, addressing the challenges posed by varying EGM durations and training sizes.

In conclusion, this study advances our understanding of AF detection by showcasing the robustness and efficiency of Minirocket, presenting it as a valuable tool with practical implications for the evolving landscape of cardiovascular monitoring technologies.

## References

- Alhousseini, M. I., Abuzaid, F., Rogers, A. J., Zaman, J. A., Baykaner, T., Clopton, P., ... & Narayan, S. M. (2020). Machine learning to classify intracardiac electrical patterns during atrial fibrillation: machine learning of atrial fibrillation. *Circulation: Arrhythmia and Electrophysiology*, 13(8), e008160.
- Bayoumy, K., Gaber, M., Elshafeey, A., Mhaimeed, O., Dineen, E. H., Marvel, F. A., ... & Elshazly, M. B. (2021). Smart wearable devices in cardiovascular care: where we are and how to move forward. *Nature Reviews Cardiology*, 18(8), 581-599.
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54, 1937-1967.
- Bottou, L., Curtis, F. E., & Nocedal, J. (2018). Optimization methods for large-scale machine learning. *SIAM review*, 60(2), 223-311.
- Dau, H., Bagnall, A., Kamgar, K., Yeh, M., Zhu, Y., Gharghabi, S., Ratanamahatana, C., Chotirat, A., & Keogh, E. (2019). The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6), 1293-1305.
- Dempster, A., Petitjean, F., & Webb, G. I. (2020a). ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5), 1454-1495.
- Dempster, A. (2020b). A Very Fast (Almost) Deterministic Transform for Time Series Classification Angus Dempster, Daniel F. Schmidt, Geoffrey I. Webb. arXiv preprint arXiv:2012.08791.
- Fulcher, B. D., Little, M. A., & Jones, N. S. (2013). Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of the Royal Society Interface*, 10(83), 20130048.
- Fulcher, B. D., & Jones, N. S. (2017). hctsa: A computational framework for automated time-series phenotyping using massive feature extraction. *Cell systems*, 5(5), 527-531.
- Haissaguerre, M., Marcus, F. I., Fischer, B., & Clementy, J. (1994). Radiofrequency catheter ablation in unusual mechanisms of atrial fibrillation: report of three cases. *Journal of cardiovascular electrophysiology*, 5(9), 743-751.
- Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1), 65-69.
- Honarbaksh, S., Schilling, R. J., Providência, R., Dhillon, G., Sawhney, V., Martin, C. A., ... & Hunter, R. J. (2017). Panoramic atrial mapping with basket catheters: A quantitative analysis to optimize practice, patient selection, and catheter choice. *Journal of cardiovascular electrophysiology*, 28(12), 1423-1432.
- Hong, Y. J., Jeong, H., Cho, K. W., Lu, N., & Kim, D. H. (2019). Wearable and implantable devices for cardiovascular healthcare: from monitoring to therapy based on flexible and stretchable electronics. *Advanced Functional Materials*, 29(19), 1808247.

- Isa, R., Villacastín, J., Moreno, J., Pérez-Castellano, N., Salinas, J., Doblado, M., ... & Macaya, C. (2007). Differentiating between atrial flutter and atrial fibrillation using right atrial bipolar endocardial signals. *Revista Española de Cardiología (English Edition)*, 60(2), 104-109.
- Jang, J. K., Park, J. S., Kim, Y. H., Choi, J. I., Lim, H. E., Pak, H. N., & Kim, Y. H. (2010). Coexisting sustained tachyarrhythmia in patients with atrial fibrillation undergoing catheter ablation. *Korean Circulation Journal*, 40(5), 235-238.
- Katrtsis, D., Iliodromitis, E., Fragakis, N., Adamopoulos, S., & Kremastinos, D. (1996). Ablation therapy of type I atrial flutter may eradicate paroxysmal atrial fibrillation. *American Journal of Cardiology*, 78(3), 345-347.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- Lubba, C. H., Sethi, S. S., Knaute, P., Schultz, S. R., Fulcher, B. D., & Jones, N. S. (2019). catch22: CAnonical Time-series CHaracteristics: Selected through highly comparative time-series analysis. *Data Mining and Knowledge Discovery*, 33(6), 1821-1852.
- Middlehurst, M., Large, J., Flynn, M., Lines, J., Bostrom, A., & Bagnall, A. (2021). HIVE-COTE 2.0: a new meta ensemble for time series classification. *Machine Learning*, 110(11-12), 3211-3243.
- Mietus, J. E., Peng, C. K., Henry, I., Goldsmith, R. L., & Goldberger, A. L. (2002). The pNNx files: re-examining a widely used heart rate variability measure. *Heart*, 88(4), 378-380.
- Palma, E. C., Ferrick, K. J., Gross, J. N., Kim, S. G., & Fisher, J. D. (2000). Transition from atrioventricular node reentry tachycardia to atrial fibrillation begins in the pulmonary veins. *Circulation*, 102(8), 937-937.
- Rodrigo, M., Waddell, K., Magee, S., Rogers, A. J., Alhusseini, M., Hernandez-Romero, I., ... & Narayan, S. M. (2021). Non-invasive spatial mapping of frequencies in atrial fibrillation: Correlation with contact mapping. *Frontiers in Physiology*, 11, 611266.
- Rodrigo, M., Alhusseini, M. I., Rogers, A. J., Krittanawong, C., Thakur, S., Feng, R., ... & Narayan, S. M. (2022). Atrial fibrillation signatures on intracardiac electrograms identified by deep learning. *Computers in biology and medicine*, 145, 105451.
- Sana, F., Isselbacher, E. M., Singh, J. P., Heist, E. K., Pathik, B., & Armoundas, A. A. (2020). Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review. *Journal of the American College of Cardiology*, 75(13), 1582-1592.
- Shifaz, A., Pelletier, C., Petitjean, F., & Webb, G. I. (2020). TS-CHIEF: a scalable and accurate forest algorithm for time series classification. *Data Mining and Knowledge Discovery*, 34(3), 742-775.
- Siddiqi, H. K., Vinayagamoorthy, M., Gencer, B., Ng, C., Pester, J., Cook, N. R., ... & Albert, C. M. (2022). Sex differences in atrial fibrillation risk: the VITAL Rhythm Study. *JAMA cardiology*, 7(10), 1027-1035.
- Smith, S. W., Rapin, J., Li, J., Fleureau, Y., Fennell, W., Walsh, B. M., ... & Gardella, C. (2019). A deep neural network for 12-lead electrocardiogram interpretation outperforms a conventional algorithm, and its physician overread, in the diagnosis of atrial fibrillation. *IJC Heart & Vasculature*, 25, 100423.
- Wang, X., Wirth, A., & Wang, L. (2007, October). Structure-based statistical features and multivariate time series clustering. In *Seventh IEEE international conference on data mining (ICDM 2007)* (pp. 351-360). IEEE.
- Weng, L. C., Preis, S. R., Hulme, O. L., Larson, M. G., Choi, S. H., Wang, B., ... & Lubitz, S. A. (2018). Genetic predisposition, clinical risk factor burden, and lifetime risk of atrial fibrillation. *Circulation*, 137(10), 1027-1038.