Machine Learning-Driven Algorithm for Improved Detection of Brief Cardiac Arrhythmias

Lucia Vavassori[1,2], Valentina DA Corino[2,3], Raphael Schneider[1], Javier Saiz-Vivo[1]

[1] Medtronic Bakken Research Center, Maastricht, The Netherlands
[2] Department of Electronics, Information and Bioengineering, Polytechnic of Milan, Milan, Italy
[3] CardioTech Lab, IRCCS Centro Cardiologico Monzino, Milan, Italy

Abstract

With the integration of implantable cardiac monitors such as Medtronic's LINQ II™ in clinical practice, continuous monitoring has improved arrhythmias detection. However, current devices only detect atrial fibrillation (AF) episodes lasting $\geq$ 2 minutes, potentially missing short but clinically relevant arrhythmias such as brief AF and non-sustained ventricular tachycardia (NSVT), particularly in hypertrophic cardiomyopathy patients. A Random Forest classifier was trained on rhythm-labeled ECG PhysioNet data from Long Term AF and VTaC databases, and tested on MIT-BIH Arrhythmia Database. Different ECG segment durations (2-min, 1-min, 30-s, and 10-s) were evaluated to assess detection performance. Extracted features included RR variability, QRS morphology, wavelet, high-order statistics, and Hermite coefficients. Feature selection combined correlation filtering and Least Absolute Shrinkage and Selection Operator regularization. With 2-min segments, AF detection achieved F1-score of 0.70 and specificity of 0.95, while NSVT detection was lower (F1 = 0.44, specificity = 0.70). Shorter segments improved NSVT detection (specificity = 0.82 at 10-s), but reduced AF specificity (0.87). Overall, shorter windows enhance NSVT detection, while AF benefits from longer ones. The selected features are computationally efficient and suitable for LINQ II™.

1. Introduction

According to the European Society of Cardiology guidelines, AF is defined as an irregular rhythm without distinct P waves, lasting at least 30 seconds. Individuals with AF are at risk of stroke, heart failure and increased mortality [1]. Implantable Cardiac Monitors (ICMs), such as the Medtronic LINQ II™, are widely used for continuous ECG monitoring in outpatient settings. LINQ II™ continuously analyzes heart rhythm based on RR-interval irregularity and detects AF episodes using a classification algorithm that operates over fixed 2-minute windows [2]. This duration balance sensitivity, specificity, and computational efficiency. Since the device is designed to operate for more than four years, the algorithm must be accurate enough to detect true AF episodes while sufficiently selective to avoid excessive false positives. In ICMs, each detected episode triggers its storage and remote transmission, which consumes limited on-device memory and electrical power for local processing and wireless telemetry. As a result, only arrhythmic events lasting at least 2 minutes are currently stored [3]. However, this limitation may lead to the omission of shorter but clinically relevant AF episodes that could support earlier diagnosis. In addition, current ICM algorithms are not optimized to detect other short arrhythmias, such as NSVT, or to precisely capture their onset and offset.

NSVT is defined as a sequence of more than three (typically five) consecutive ventricular beats at rates exceeding 100 beats per minute, lasting less than 30 seconds [4]. NSVT is a well-established marker of increased risk for sustained ventricular arrhythmias and sudden cardiac death, especially in hypertrophic cardiomyopathy patients, where it plays a key role in decisions regarding implantable cardioverter-defibrillator (ICD) therapy. Although the 2020 AHA/ACC guidelines recommend 48-hour Holter monitoring for symptom evaluation, they also advise periodic ambulatory ECG monitoring every 1 to 2 years for routine surveillance. Notably, several studies have shown that more frequent and prolonged monitoring significantly improves the detection rates of both AF and NSVT [5–7]. Detecting NSVT is particularly challenging due to its brief duration and its similarity to other arrhythmias such as AF or supraventricular tachycardia, which may exhibit wide QRS complexes and elevated ventricular rates. [8, 9]. To the best of our knowledge, this is the first study to propose a unified machine learning (ML) pipeline for the detection of both short AF episodes and transient NSVTs. The aim of this study is to develop a lightweight ML model suitable for integration into ICMs, enabling en-

hanced diagnostic capabilities without compromising performance and memory.

2. Methods

2.1. Databases

To develop the ML model, three publicly available ECG datasets from PhysioNet were selected for their rhythm-level annotations. The *Long Term AF Dataset* (LTAF) includes recordings from 24 patients with paroxysmal AF and 20 with sustained AF, with a balanced gender distribution. Normal sinus rhythm (N) segments were also extracted from the same patients during periods without AF. [10, 11]. The *VTaC: A Benchmark Dataset of Ventricular Tachycardia Alarms* (VTaC) contains intensive care unit recordings from 2,260 patients and a total of 5,037 alarms. Among these, 1,441 episodes, originating from 777 unique patients, were manually validated by clinicians as true NSVT and included in this study. [8, 11]. The *MIT-BIH Arrhtyhmia Database* (MIT-BIH AD) includes 48 annotated Holter recordings from 47 subjects, aged 23 to 89 years [11, 12]. For this study, only segments labeled as N, AF, and NSVT were considered (Table 1), in line with HCM guidelines, which identify these two arrhythmias as the most clinically relevant for risk stratification and ICD therapy. To ensure generalizability of the model and prevent patient overlap, LTAF and VTaC were used for training, while MIT-BIH AD for testing.

Table 1. Overview of selected PhysioNet ECG databases.

Database	Number of Patients	Sampling Rate [Hz]	Recording Duration	Rhythm Considered
LTAF	44	128	25 h	N, AF
VTaC	777	250	6 min	NSVT
MIT-BIH AD	47	360	30 min	N, AF, NSVT

2.2. ECG Preprocessing

From each database, four derived datasets were generated by segmenting the original ECG recordings into fixed-length windows of 2 minutes, 1 minute, 30 seconds, and 10 seconds. The 2-minute window corresponds to the clinical threshold used by the LINQ II™ device, while shorter windows were explored to identify the minimum window length that maintains diagnostic performance while reducing computational demands. In LTAF and MIT-BIH AD, segmentation was aligned with rhythm annotations, retaining only windows that contained either AF or a combination of sinus rhythm and NSVT. In VTaC, where NSVT onset is standardized at minute 5, windows were centered around this point.

A unified preprocessing pipeline was applied to all recordings, following the methodology described in [8]. Briefly, the filtering stage included a high-pass filter at 1 Hz to remove baseline wander, a second-order low-pass Butterworth filter at 30 Hz to attenuate high-frequency noise, and a Notch filter at 60 Hz to eliminate powerline interference. After filtering, all signals were downsampled to 128 Hz, matching both the lowest native sampling rate among the selected databases and the acquisition frequency of LINQ II™.

2.3. Feature Extraction

A total of 211 hand-crafted features were extracted and grouped into three categories: heart rate (HR)-related, QRS complex-related, and signal morphology features. R-peak detection was performed using Pan-Tompkins algorithm, while fiducial points *(P, Q, R, S)* were identified using the method described in [13, 14]. In AF recordings, where the P-wave is absent, the algorithm accounted for baseline oscillations characteristic of fibrillatory activity. All features were then aggregated using statistical descriptors, yielding a single value per feature per segment.

- **HR-related features:** For each heartbeat, the RR interval (RR_i) was computed together with its normalized value (RR_i/RR_{avg}). Additional features included the ratio between successive intervals (RR_{i+1}/RR_i), the coefficient of variation, the root mean square of successive differences (RMSSD), the percentage of successive RR intervals differing by more than 20 ms (pNN20), and Poincaré descriptors.
- **QRS complex–related features:** Temporal and amplitude-based features were extracted for each QRS complex. These included the total QRS width (QRS_w), widths at 50% (QRS_{w2}) and 25% (QRS_{w4}) of the R-peak amplitude, and intervals between fiducial points, such as the Q-S and P-QRS onset intervals. Amplitude descriptors included absolute ($P_{peak}, Q_{peak}, R_{peak}, S_{peak}$) and differential values at each fiducial point (PQ_a, QR_a, RS_a).
- **Signal morphology features:** Hermite Basis Function descriptors were computed from 500 ms windows centered on each R-peak using orthogonal Hermite polynomials of degrees 3, 4, and 5. High-Order Statistical features, including skewness and kurtosis, were extracted at five equally spaced lags within the same window. Finally, time–frequency features were derived from a three-level Discrete Wavelet Transform using Daubechies wavelets (db1), applied to each segment to extract time–frequency descriptors.

2.4. ML Model Development

To reduce redundancy and improve model generalizability, a three-step feature selection strategy was implemented in a 5-fold cross validation setting. Within each

fold, highly correlated features (Pearson correlation > 0.9) were removed. Next, Least Absolute Shrinkage and Selection Operator (LASSO) regression was used to enforce sparsity and eliminate less informative features. Finally, only the features common to all five folds were considered to train the final model, and mutual information scores were computed to rank them by relevance.

Multiple ML classifiers were evaluated: k-Nearest neighbors, random forest (RF), support vector machine, AdaBoost, NuSVC, and logistic regression. To address class imbalance in the training set, the majority class was downsampled, resulting in 1,441 ECG segments per rhythm class. For testing, the imbalanced dataset was used to better reflect real-world rhythm distributions. To ensure comparability across ECG window lengths, the number of test segments was limited to match the smallest available count, which occurred in the 2-minute dataset (411 N, 59 AF, and 62 NSVT).

The F1-score was used as the primary evaluation metric due to class imbalance, with AF and NSVT being less frequent than N. Sensitivity and specificity were also considered, as ICMs must detect true arrhythmias reliably, while minimizing false positives to preserve memory and battery life.

3. Results

3.1. Model optimization

The RF classifier achieved the highest F1-score among all models evaluated and was therefore selected for further optimization. To fine-tune its performance, different configurations were tested using grid-search settings and by varying the size of the input feature vector. The optimal number of features varied with segment length: 18 for 2-minute and 1-minute windows, and 14 for both 30- and 10-second segments. For consistency in comparison across window durations, the final model configuration used the top 14 features ranked by Mutual Information. Among these, features related to RR intervals, QRS morphology, and skewness consistently ranked highest in relevance and also exhibited the lowest computational cost.

3.2. RF Performance Across ECG Window Durations

The performance of the RF classifier varied across segment lengths. With 2-minute windows, NSVT detection achieved high sensitivity (0.92) but a low F1-score of 0.44, indicating a high rate of false positives. This was likely due to the short duration of NSVT events, which were embedded within longer windows dominated by normal sinus rhythm. As features were aggregated over the entire segment, NSVT-specific patterns were diluted, leading to frequent misclassifications. In contrast, AF detection achieved robust results, with an F1-score of 0.70 and specificity of 0.95. Reducing the window to 1 minute preserved strong AF performance (sensitivity = 0.68, specificity = 0.94), while NSVT sensitivity remained high (0.84), although specificity was still limited (0.75), again indicating many false positives. With 30-second segments, the model maintained good AF classification (sensitivity = 0.63, specificity = 0.96), and NSVT detection improved significantly, achieving sensitivity of 0.74 and specificity of 0.98.

At 10-second windows, the classifier achieved its best NSVT detection performance, with sensitivity of 0.95 and specificity of 0.82. AF detection remained acceptable, though sensitivity (0.64) and specificity (0.87) showed a modest decline (Figure 1). These trends are further supported by *Table 2*, which shows a reduction in false positive (FP) for NSVT, along with relatively stable true positive (TP) counts and a slight increase in FP for AF as the window length decreases.

Regarding class-weighted performance across AF and NSVT, specificity remained consistently high across all window lengths, whereas both sensitivity and F1-score improved as the segment length decreased (Figure 2).

Table 2. TP, FP, true negative (TN) and false negative (FN) for AF and NSVT across different ECG durations.

		2 min	1 min	30 sec	10 sec
AF	TP	44	40	31	38
	FP	23	30	16	54
	TN	450	443	457	419
	FN	15	19	28	21
NSVT	TP	57	52	61	59
	FP	140	117	124	85
	TN	330	353	346	385
	FN	5	10	1	3

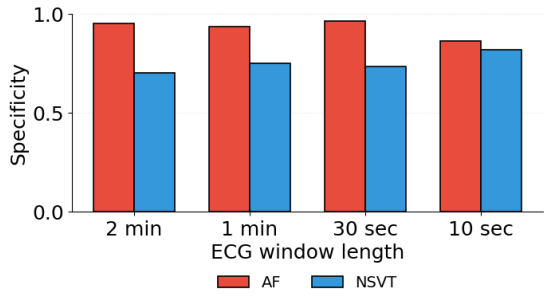

Figure 1. AF and NSVT specificity across ECG durations.

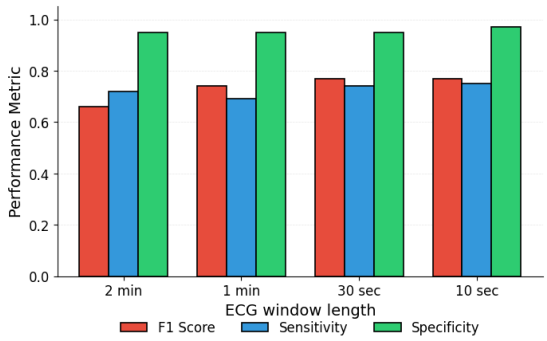

Figure 2. Class-weighted performance across ECGs.

4. Discussion

This study evaluated a RF classifier for the joint detection of NSVT and brief AF episodes. Results show a clear window-length effect: shorter segments improved NSVT detection by isolating events, while longer ones better captured AF through RR variability. This supports the 2-minute window used in LINQ II™ designed to balance performance and resource use. Most prior work focused on AF or VT separately. *Zabihi et al.,* trained an RF with 150 hand-crafted features on 30-second recordings, achieving an AF F1-score of 0.83 [15]. *Liaqat et al.,* compared ML models on MIT-BIH AF Database, reporting AF F1-scores of 0.86 [16]. These works, however did not address VT. For NSVT, *Sau et al.,* proposed a RF model to distinguish NSVT from VT, with sensitivity of 0.79 and specificity of 0.75, comparable to our results [17].

In our study, feature selection highlighted beat-to-beat RR variability as the main determinant for AF detection in longer segments, whereas QRS morphology and short-run RR descriptors were the most informative for NSVT in shorter windows. Notably, no frequency-domain features were selected at any window length, supporting an ICM-friendly design based on computationally efficient, time-domain metrics.

5. Conclusion

This study presented a simple RF model capable of detecting both AF and NSVT, with performance varying by window length. Shorter segments improved NSVT detection, while longer windows favored AF. Future work will focus on testing and fine-tuning the model on real-world ICM data to support clinical integration.

Acknowledgments

This work was supported by the INSIDE-HEART consortium. The INSIDE-HEART project has received funding from the European Union's Horizon Europe programme under the Marie Skłodowska-Curie Grant Agreement No. 101119941 (HORIZON-MSCA-2022-DN-01).

References

[1] Fredriksson T, et al. Brief episodes of rapid irregular atrial activity are a risk marker for af: a prospective cohort study. BMC Cardiovascular Disorders April 2020;20(1):167.

[2] Cho H, et al. Accuracy of ILR for detecting atrial tachyarrhythmias after AF catheter ablation. International Journal of Arrhythmia 2020;(1):6.

[3] Wechselberger S, et al. Continuous monitoring after AF ablation: the LINQ AF study. EP Europace 2018;20.

[4] Demosthenes G, et al. Nsvt. Journal of the American College of Cardiology 2012;60(20).

[5] Steve R, et al. 2020 aha/acc guideline for the diagnosis and treatment of patients with HCM. Circulation 2020;142(25).

[6] Rowin E, et al. Extended ambulatory ECG monitoring enhances identification of higher-risk VT in patients with hcm. Heart Rhythm 2024;21(6).

[7] Du D, et al. Arrhythmia monitoring for risk stratification in HCM. CJC Open 2022;4(4).

[8] Lehman, et al. VTaC: A benchmark dataset of VT alarms from ICU monitors. 2024; .

[9] Littmann, et al. Initial evaluation and management of wide-complex tachycardia: A simplified and practical approach. The American Journal of Emergency Medicine 2019;37(7):1340–1345.

[10] Petrutiu, et al. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal AF in humans. Europace 2007;9(11):803–809.

[11] Goldberger, et al. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23).

[12] Moody, et al. The impact of the MIT-BIH AD. IEEE Engineering in Medicine and Biology Magazine 2001;20(3).

[13] Cogollo S, et al. Investigating feature selection and RF for inter-patient heartbeat classification 2020;.

[14] Mondéjar-Guerra V, et al. Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers 2019;47:41–48.

[15] Zabihi M, et al. Detection of AF in ECG hand-held devices using a RF classifier. In Computing in Cardiology Conference. Rennes, France: IEEE, 2017; .

[16] Sidrah L, et al. Detection of af using a ML approach. Sensors 2021;21(4):1–17.

[17] Sau A, et al. Ml-derived cycle length variability metrics predict spontaneously terminating VT in ICD recipients. European Heart Journal – Digital Health 2024;5(1):50–59.

Address for correspondence:

Lucia Vavassori
Bakken Reserch Center, Medtronic
Endepolsdomein 5, Maastricht, 6229 GW, The Netherlands
E-mail: lucia.vavassori@medtronic.com