# Robust ECG Signal Decomposition Enables Interpretable Distinction Between Atrial Flutter and Atrial Fibrillation

Marc Goettling[1], Alexander Hammer[1], Nadine K. Wäßnig[1,2], Hagen Malberg[1], Martin Schmidt[1]

[1] TU Dresden, Institute of Biomedical Engineering, Dresden, Germany
[2] Kardiologen am Fetscherplatz, Dresden, Germany

## Abstract

*Atrial fibrillation (AFIB) and atrial flutter (AFL) require different clinical management, but their diagnostic distinction from short 12-lead electrocardiograms (ECGs) can pose a challenge. We propose an interpretable pipeline that combines a synthetic-data-trained 1D U-Net for atrial signal extraction with an autocorrelation-based distinction feature. A total of 50,000 synthetic 10-second ECGs were generated using cubic Hermite splines with diverse atrial and ventricular morphologies, that were further augmented by multiple noise profiles. The U-Net was trained on synthetic data to decompose each signal into atrial, ventricular and noise components, achieving Pearson correlations of 0.980 (atrial) and 0.998 (ventricular) on validation data. The trained model was then applied to 5,203 AFIB and 726 AFL 12-lead ECG recordings from the CinC challenge 2021 datasets in a leave-one-dataset-out cross-validation. The optimized distinction feature achieved a cross-dataset accuracy of 93.6%. Dataset-specific performance increased consistently over all datasets by up to +10.3% compared to the literature. Around 90% sensitivity, the method achieved 89.0% specificity compared to only 66.7% for a standard method. Cross-dataset and dataset-specific performance optimization showed similar findings, demonstrating generalizability and, thus, clinical usability in further applications.*

## 1. Introduction

Atrial fibrillation (AFIB) and atrial flutter (AFL) are distinct atrial arrhythmias that overlap substantially in rate control, rhythm control, and ablation planning, but still require disease-specific treatment decisions. Accurate differentiation is therefore clinically important. However, in emergency department settings, AFL has been reported to be misdiagnosed as AFIB in up to 44% of cases [1].

A common clinical distinction between the two arrhythmias is based on the ventricular response. AFIB is typically associated with an absolute arrhythmia without organized P waves, whereas AFL is often expected to produce a more regular ventricular pattern because of fixed atrioventricular conduction. However, ventricular rate and rhythm do not reliably distinguish AFIB from AFL [2]. In particular, AFL with variable atrioventricular conduction can produce irregular RR intervals that closely resemble AFIB on a short 10 s ECG, so ventricular rhythmicity alone is not a robust basis for discrimination [2].

A key difference lies in atrial activity. AFL is driven by organized macro-reentrant circuits that generate repetitive atrial waves in at least one lead [3]. AFIB, by contrast, is characterized by chaotic, low-amplitude fibrillatory activity without stable periodicity [1]. This suggests that atrial activity may provide a more robust basis for discrimination than ventricular rhythm alone.

However, previous studies on automated AFIB/AFL distinction provide only limited evidence for robust generalization of any feature set. Some relied on small datasets [4], whereas others were affected by data leakage [5]. In addition, a recent deep learning study reported reduced discrimination performance in cross-dataset evaluation, suggesting reliance on dataset-specific patterns rather than stable physiological markers [6]. Because AFIB and AFL differ in the temporal organization of atrial activity, we hypothesize that quantifying atrial regularity enables more reliable, cross-dataset robust distinction. Since real surface ECG recordings lack ground truth decomposition into atrial and ventricular signal components, we employ synthetic ECGs (sECG) to model these separate signals and enable supervised training. To test our hypothesis, we trained a 1D U-Net on sECG decomposition and used it to extract atrial activity from real 12-lead ECG recordings, compute an autocorrelation (AC) based regularity measure from the extracted signal and evaluate its discriminative performance using leave-one-dataset-out cross-validation (LODO) across the CinC Challenge 2021 [7] datasets.

## 2. Methods

Figure 1 summarizes the proposed workflow. sECGs with known atrial and ventricular components are noise-

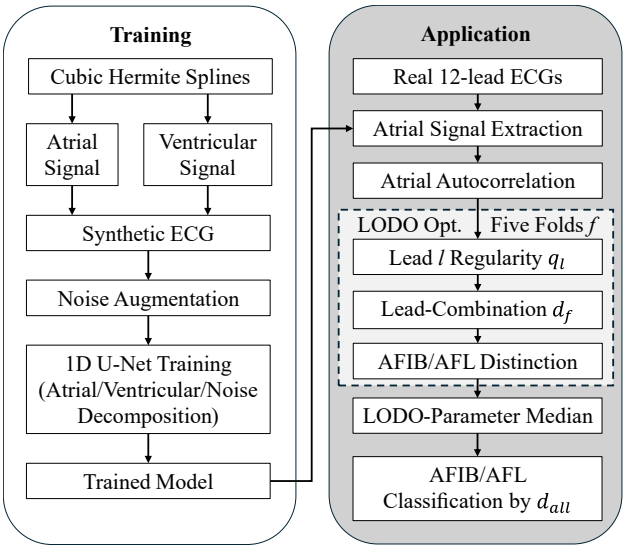


Figure 1. Overview of the proposed workflow. AFIB, Atrial fibrillation; AFL, Atrial flutter; ECG, Electrocardiogram; LODO Opt., Leave-one-dataset-out optimization.

augmented and used to train a 1D U-Net for signal decomposition. The trained model is then applied to real 12-lead ECGs to extract atrial activity, from which an AC-based regularity feature is computed per lead. In the LODO setting, the regularity feature window and lead-combination weights are optimized across folds, and their median values are used for the final AFIB/AFL classification score.

## 2.1. Synthetic Dataset Generation

Using Cubic Hermite splines [8], we generated 50,000 10-second single-lead sECGs at 500 Hz by summing atrial and ventricular signals. The ventricular signal consisted of repeated QRS-T beats with diverse spline-defined morphologies and the atrial component was modeled separately for fibrillation and flutter using waveforms with varying amplitude, frequency, and frequency modulation. Beat number and timing were sampled from predefined heart rate and heart rate variability ranges, except in a subgroup of AFL signals, where beat timing was coupled to flutter wave phase to mimic different atrioventricular block patterns.

To model realistic recording conditions, we generated nine noisy versions of each sECG, with three cases each for low, medium and high noise amplitude. Noise types included baseline wander, white noise, pink noise, quantization artifacts, muscle activity spikes, high-frequency bursts, and lead disconnect segments.

## 2.2. Application Datasets

For AFIB versus AFL classification, we used the CinC Challenge 2021 datasets [7], excluding Ningbo and PTB due to AFIB/AFL label ambiguity and insufficient recordings. The final dataset comprised 5,203 AFIB (median age 73 [Q1-Q3: 65-81] years; $57.6\%/42.4\%$ male/female) and 726 AFL (median age 71 [61-79] years; $59.0\%/41.0\%$ male/female) 12-lead ECGs, each 10 s long at 500 Hz.

## 2.3. Decomposition Model

This proof-of-concept study did not include systematic hyperparameter tuning of the 1D U-Net. Instead, we selected a moderate-complexity architecture based on signal-domain considerations. Five double convolution layers with stride size $s = 4$ in the second convolution of each block and kernel size $k = 13$ yield a receptive field of $r = 4{,}105$. This resides close to the input length $n = 5{,}000$, enabling modeling of low-frequency components such as baseline wander across a 10 s ECG. The network uses $m = 16$ initial feature maps, doubled at each level, resulting in about $2.7 \times 10^7$ trainable parameters.

## 2.4. Model Training and Evaluation

The sECGs were split into 80% training and 20% validation sets. The U-Net was trained for 200 epochs using an equally weighted loss combining Pearson correlation for the predicted atrial and ventricular signals with atrial and ventricular SNR terms based on the residual after subtracting both predictions from the input. The model with the lowest validation loss was retained.

As a reference value, we also implemented standard beat template subtraction (TS). For each lead, R peaks were detected [9] and windows around the detected beats were aligned. A median ECG beat template was then constructed and subtracted at each R peak position. The residual signal was used as reference estimate of atrial activity.

## 2.5. Separation Feature and Optimization

For each extracted atrial signal, the AC was computed via the Wiener-Khinchin theorem. For a mean-centered signal $\tilde{x}$, the normalized AC is

$$R_k = \frac{F^{-1}\left(|F(\tilde{x})|^2\right)_k}{R_0}, \quad k = 0, 1, \ldots, L-1,$$

where $F$ denotes the discrete Fourier transform, $L$ the maximum lag index, and $R_0$ normalizes the result to $[-1, 1]$.

For each lead $l$, regularity feature $q_l$ was defined as the mean AC peak value within the lag window $[w_{\min}, w_{\max}]$.

These 12 lead-wise features were linearly combined into a single distinction feature $d = \sum_{l=1}^{12} \alpha_l \, q_l$, with per lead weights $\alpha_l$.

Generalizability was assessed by LODO cross-validation across the five application datasets, separately for U-Net and TS-derived atrial signals. In each fold $f$, one dataset was used for testing and the remainder for optimization of feature calculation parameters $w_{\min}$, $w_{\max}$ and all 12 $\alpha_l$ using a tree-structured Parzen estimator sampler [10], maximizing receiver operating characteristic curve (ROC-AUC) for AFIB/AFL discrimination by a distinction feature $d_f$. Thresholds on the AUC were then selected separately on the in-fold data and scores reported on the hold-out data. One threshold was chosen to report maximum accuracy and enable literature comparison and another at the 90% sensitivity operating point to allow estimation of clinical utility by sensitivity and specificity. An across-all datasets distinctive feature $d_{all}$ was obtained by taking the median parameter values across the five optimized parameter sets for the five different $d_f$. For a plausibility check, the median $w_{\min}$ and $w_{\max}$ values were applied to the single-channel AC of atrial signals from sECGs with predefined atrial frequencies and frequency variabilities and the Pearson correlation of the AC feature with these parameters was evaluated.

## 3. Results

On the sECG validation set, the decomposition U-Net achieved median Pearson correlations with the ground truth of 0.998 for the ventricular signal and 0.980 for the atrial signal. Figure 2 shows an example decomposition of a real AFL signal into atrial and ventricular components together with the atrial AC signal.

Across LODO folds on real data, except CPSC with no AFL signals, the median test ROC-AUC was 94.6% for the fold specific features $d_f$ computed from U-Net extracted atrial signals, compared with 89.6% for TS-based atrial signals. Table 1 summarizes classification performance at different operating points. Table 2 shows parameter distributions across the U-Net based folds, with 9 of 12 lead weights having interquartile ranges of 0.03 or less, AC window ranges of 20 ms for $w_{\min}$ and 30 ms for $w_{\max}$, and the highest lead weights assigned to leads II and V1.

Using the median parameters for $d_f$ calculation from the U-Net based folds across all real data produced the distribution for $d_{all}$ shown in Figure 3, with classification results reported in Table 1. Comparing LODO $d_f$ and $d_{all}$ distinction, per dataset accuracy changes in percentage points (pp) ranged from -1.64 pp to 3.08 pp. For the synthetic atrial signals, the AC mean peak feature showed a Pearson correlation of 0.22 ($p < 0.001$) with the base frequency and $-0.82$ ($p < 0.001$) with frequency variability.

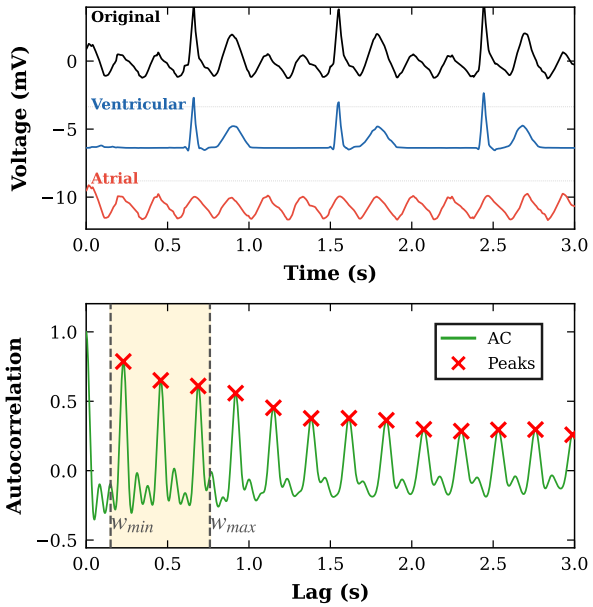


Figure 2. Real ECG lead II atrial flutter (AFL) decomposition snippet and the corresponding atrial autocorrelation (AC) with the AC feature calculation window $w$.

Table 1. Classification results per leave-one-dataset-out fold $f$ on CinC 2021 datasets by distinctive features $d_f$ and across folds by $d_{all}$ in %. $\mathrm{Acc_U}$, U-Net accuracy; $\mathrm{Acc_{Lit}}$, literature accuracy [6]; $\mathrm{Se^{90}}$/$\mathrm{Sp^{90}}$, sensitivity/specificity at 90% in-fold training sensitivity for U-Net (U) and template subtraction (TS).

| Test set $T_f$ | $\mathrm{Acc_U}$ | $\mathrm{Acc_{Lit}}$ | $\mathrm{Se_U^{90}}$ | $\mathrm{Sp_U^{90}}$ | $\mathrm{Se_{TS}^{90}}$ | $\mathrm{Sp_{TS}^{90}}$ |
|---|---|---|---|---|---|---|
| CPSC | 95.9 | 92.4 | – | 85.8 | – | 69.0 |
| CPSC-Extra | 88.4 | 78.1 | 72.2 | 87.6 | 72.2 | 68.0 |
| G12EC | 85.2 | 81.1 | 92.4 | 72.7 | 94.7 | 53.2 |
| PTB-XL | 97.8 | 92.5 | 100.0 | 91.7 | 100.0 | 63.5 |
| Chapman-Shaoxing | 87.8 | 85.2 | 85.4 | 94.1 | 81.1 | 87.4 |
| All by $d_{all}$ | 93.6 | – | 88.9 | 89.0 | 91.6 | 66.7 |

## 4. Discussion

The proposed pipeline suggests that synthetic-data-trained ECG decomposition can recover atrial activity relevant for AFIB/AFL discrimination. The U-Net showed near-perfect agreement with synthetic ground truth for ventricular and atrial signals, which translated to strong performance on real data. In LODO evaluation, $d_f$ from U-Net derived atrial signals achieved a median ROC-AUC of 94.6%, compared with 89.6% for TS-based signals. Accuracies were also consistently above literature-reported values, with a maximum gain of 10.3 pp.

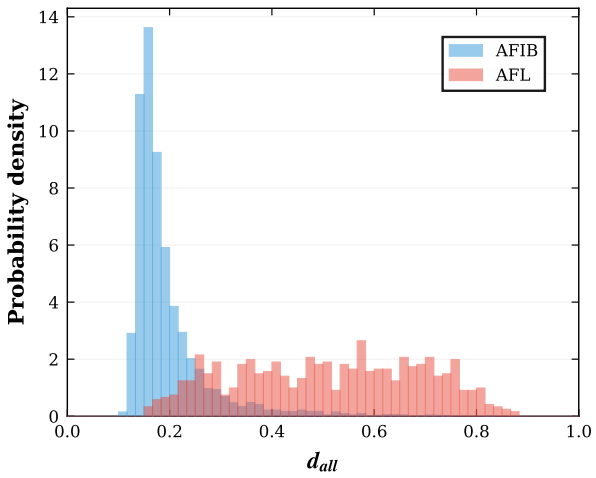


Figure 3. Probability density of the U-Net based parameter set median derived feature $d_{all}$ for all real atrial fibrillation (AFIB) and atrial flutter (AFL) recordings.

Table 2. Distribution of autocorrelation (AC) calculation parameters based on U-Net extracted atrial signals across folds. Lead weights $\alpha_l$ and AC window bounds $w_{\min}$ and $w_{\max}$ are reported as median [Q1 – Q3].

| Lead | $\alpha_l$ | Lead | $\alpha_l$ |
|---|---|---|---|
| I | 0.06 [0.02 – 0.07] | V1 | 0.20 [0.19 – 0.21] |
| II | 0.20 [0.19 – 0.20] | V2 | 0.04 [0.04 – 0.05] |
| III | 0.08 [0.08 – 0.10] | V3 | 0.01 [0.01 – 0.02] |
| aVR | 0.06 [0.01 – 0.06] | V4 | 0.06 [0.03 – 0.06] |
| aVL | 0.13 [0.13 – 0.15] | V5 | 0.04 [0.02 – 0.04] |
| aVF | 0.05 [0.04 – 0.11] | V6 | 0.07 [0.06 – 0.09] |
| $w_{\min}$ (ms) | | 150 [130 – 150] | |
| $w_{\max}$ (ms) | | 760 [730 – 760] | |

The optimized parameter sets for $d_f$ showed little variability, both in the AC window bounds and lead weights, indicating similar feature extraction across folds. Consistent with this, using the median parameters across the U-Net based folds, $d_{all}$ achieved near-excellent discrimination, with 88.9% sensitivity and 89.0% specificity across all recordings. The per-dataset accuracy difference between $d_{all}$ and the fold-specific features $d_f$ ranged only from $-1.64$ to $+3.08$ pp, further supporting generalizability across datasets.

The resulting parameter distribution was also physiologically plausible. The highest weights were assigned to leads II and V1, which are clinically relevant for assessing pathophysiological atrial activity [1, 3]. In addition, the feature distribution matched the expected physiology. Regularity values $d_{all}$ were lower in AFIB and higher in AFL, consistent with the more organized atrial activation in flutter. This interpretation was supported by the synthetic plausibility analysis, where the AC mean peak feature showed only a weak correlation with base frequency ($r = 0.22$) but a strong negative correlation with frequency variability ($r = -0.82$).

In future work, the proof-of-concept U-Net architecture should be systematically optimized, and the sECG generator should be refined toward greater physiological realism. It should also be tested whether combining the proposed AC-based regularity feature with rhythm information further improves AFIB/AFL discrimination. In addition, evaluation on further external datasets with carefully established rhythm annotations would help to better define the method's generalizability and performance limits.

Overall, the results support an interpretable and generalizing marker for AFIB from AFL distinction based on atrial signal regularity.

## References

[1] Avidan Y, et al. Recognizing atrial flutter in the emergency department: Challenges in diagnosis. Am J Emerg Med 2025;96:224–229.

[2] Krummen DE, et al. Diagnostic accuracy of irregularly irregular rr intervals in separating atrial fibrillation from atrial flutter. Am J Cardiol 2006;98(2):209–214.

[3] Bun SS, et al. Atrial flutter: More than just one of a kind. Eur Heart J 2015;36(35):2356–2363.

[4] Wang J. Automated detection of atrial fibrillation and atrial flutter in ecg signals based on convolutional and improved elman neural network. Knowledge Based Systems 2020; 193:105446.

[5] Faust O, et al. Automated arrhythmia detection based on rr intervals. Diagnostics 2021;11(8):1446.

[6] Ribeiro E, et al. Can deep learning models differentiate atrial fibrillation from atrial flutter? medRxiv 2023;2023–08.

[7] Reyna M, et al. Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021. PhysioNet 2022;.

[8] Mishra A, et al. Comparative analysis of parametric b-spline and hermite cubic spline based methods for accurate ecg signal modeling. Journal of Electrocardiology 2024; 86:153783.

[9] Makowski D, et al. Neurokit2: A python toolbox for neurophysiological signal processing. Behav Res Methods 2021; 53(4):1689–1696.

[10] Akiba T, et al. Optuna: A next-generation hyperparameter optimization framework. In Proc. KDD. 2019; 2623–2631.

Address for correspondence:

Marc Goettling
IBMT, Fetscherstraße 29, 01307 Dresden, Germany
marc.goettling@tu-dresden.de