Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Unmasking Hidden Waves: A Proof-of-Concept Transformer-Based Electrocardiogram Signal Decomposition for Complex Arrhythmias

Michał Szafarczyk

michal.szafarczyk@student.uj.edu.pl

Jagiellonian University Medical College

Peter van Dam

University Hospital in Krakow

Article

Keywords:

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10680222/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Unmasking Hidden Waves: A Proof-of-Concept Transformer-Based Electrocardiogram Signal Decomposition for Complex Arrhythmias

Michał Szafarczyk, B.Eng.[1,2,*] and Peter M. van Dam, PhD[2,3]

[1]Jagiellonian University Medical College, Center for Digital Medicine and Robotics, Kopernika 7E, 31-034 Krakow, Poland
[2]University Hospital in Krakow, Department of Diagnostic Imaging, Functional and Virtual 3D Medical Imaging Laboratory, 2 Jakubowskiego St., 30-688 Krakow, Poland
[3]Jagiellonian University Medical College, Department of Radiology, Botaniczna St. 3, 31-503 Kraków, Poland
[*]Corresponding author. Email: michal.szafarczyk@student.uj.edu.pl

Abstract

Standard automated electrocardiogram (ECG) algorithms rely on sequence classification, frequently failing to provide morphological insight during complex tachycardias where critical physiological events, such as P waves, are buried. Here, we explore the feasibility of shifting the paradigm from proxy-based classification to explicit signal decomposition to unmask hidden cardiac electrical activity. We developed the ECG Signal Decomposition Transformer (ESD-former), a hybrid 1D U-Net architecture with a Transformer bottleneck, designed to continuously disentangle raw ECG signals into explicitly separated P, QRS, and T tracks. Trained on physiologically constrained synthetic overlapping arrhythmias and optimized using a custom ShapeAwareLoss, the model was tested for proof-of-concept real-world feasibility on continuous 5-minute resting ECGs. While baseline models failed on overlapping waveforms, the ESD-former reconstructed hidden features like retrograde P waves in synthetic examples. In an exploratory clinical validation, it achieved a Positive Percent Agreement (PPA) of 92.9% for QRS, 76.0% for T waves, and 82.7% for atrial activity against a reference software, while successfully generating 81 additional deeply buried P wave detections that were subsequently confirmed by manual expert review. This explicit signal decomposition framework provides a foundational proof-of-concept that, following rigorous clinical validation, could potentially assist electrophysiologists in the assessment of complex arrhythmias.

Keywords: electrocardiogram; deep learning; signal decomposition; complex arrhythmias; atrioventricular block; overlapping waveforms.

Introduction

The electrocardiogram (ECG) remains the fundamental diagnostic tool for identifying cardiac arrhythmias. While the exponential growth of continuous monitoring data necessitates automated analysis, standard algorithms frequently struggle with complex paroxysmal supraventricular tachycardias (SVTs), specifically atrioventricular nodal reentrant tachycardia (AVNRT) and atrioventricular reentrant tachycardia (AVRT)[1]. The primary diagnostic challenge during these extreme tachycardias lies in severe waveform superposition: retrograde P waves collide with and become entirely buried within the QRS complexes or subsequent ST-T segments.

Traditional automated diagnostic software often circumvents this problem by relying on macroscopic, surrogate features, such as regular R-R interval shortening or "pseudo-R'" waves in lead V1[2]. Similarly, contemporary deep learning architectures formulate ECG analysis predominantly as a sequence classification task, outputting discrete diagnostic labels[3]. While achieving high statistical accuracy for rhythm classification, these systems generally do not resolve the underlying physiological components, offering limited morphological insight into the hidden P waves — information that is highly valuable for electrophysiologists when evaluating mechanisms of tachycardia.

Inspired by advancements in general time-series forecasting — specifically the Autoformer architecture[4], which demonstrated that intrinsic signal decomposition enhances representation learning — we investigate a shift from proxy-based classification to explicit signal separation. We hypothesize that continuously disentangling a complex ECG signal into independent tracks can provide a more transparent, physiologically interpretable representation of extreme arrhythmias.

To test this proof-of-concept, we introduce the ECG Signal Decomposition Transformer (ESD-former), a hybrid 1D U-Net architecture integrating a Transformer bottleneck, designed to map ECG signals into separated P, QRS, and T waveforms. To address the inherent lack of absolute ground truth for superimposed events in clinical records, we developed

a synthetic generation pipeline and optimized the network using a custom ShapeAwareLoss. Through a progressive ablation study and subsequent exploratory validation on continuous 5-minute resting ECGs, we evaluate the feasibility of this framework to untangle buried components and explicitly separate hidden P waves during complex tachycardias and complete heart blocks.

Results

Model Generalization to Unseen Morphologies

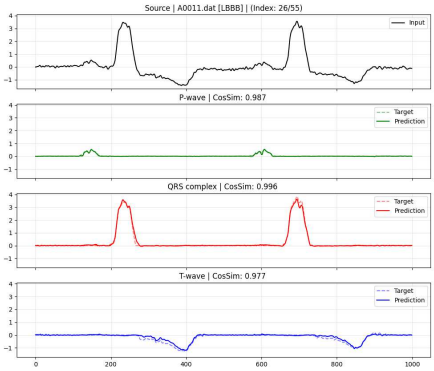
Table 1 presents the comparative evaluation of Models 1, 2, and 3 across the dataset, measured by the cosine similarity (CosSim) for P, QRS, and T waves. The results indicate the limitations of models trained strictly on distinct, non-overlapping waveforms when tasked with zero-shot generalization to complex arrhythmias.

Model 1 (trained only on normal rhythms) and Model 2 (trained on non-overlapping arrhythmias) failed to decompose cases characterized by severe wave superposition. For instance, in predicting the P wave for AVNRT and AVRT, Model 1 achieved CosSim scores of -0.106 and 0.008, respectively, while Model 2 achieved 0.026 and 0.015. These near-zero or negative values suggest that the models lacked the physiological prior to search for retrograde or buried P waves within the ST-segment or T wave.

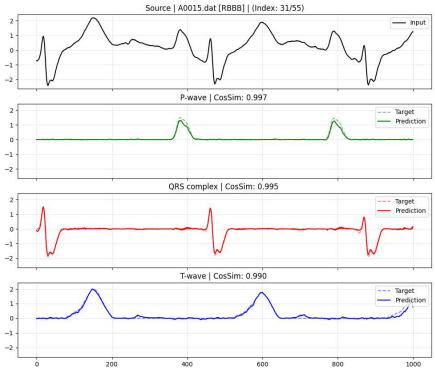
Conversely, Model 3, which was exposed to the broader dataset including synthetic overlapping cases, learned the complex signal decomposition, showing improved performance on extreme 1AVB, 3AVB, AVRT, and AVNRT cases (Fig. 1 and Fig. 2). It should be noted that the performance gain in Model 3 is attributable to a combination of synthetic overlap augmentation and exposure to additional arrhythmia classes.

Table 1: **Comparative Evaluation of Models 1, 2, and 3.** Values represent the Cosine Similarity (CosSim) for P, QRS, and T wave predictions across different arrhythmia classes. *Abbreviations:* 1AVB, first-degree atrioventricular block; 2AVB, second-degree atrioventricular block; 3AVB, third-degree atrioventricular block; AFIB, atrial fibrillation; AVNRT, atrioventricular nodal reentrant tachycardia; AVRT, atrioventricular reentrant tachycardia; LBBB, left bundle branch block; N/A, not applicable; PVC, premature ventricular contraction; RBBB, right bundle branch block; WPW, Wolff-Parkinson-White syndrome.

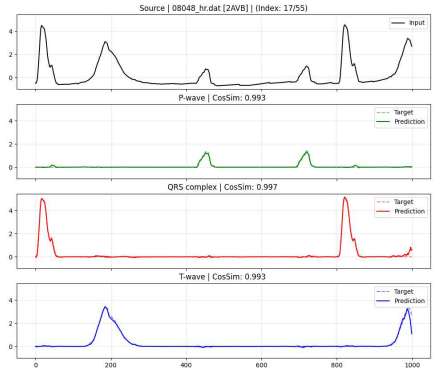
Class	**Model 1 (Normal Only)**			**Model 2 (Non-Overlapping)**			**Model 3 (Full)**		
	P	QRS	T	P	QRS	T	P	QRS	T
1AVB	0.878	0.992	0.839	0.884	0.995	0.889	0.883	0.995	0.900
2AVB	0.897	0.983	0.985	0.688	0.997	0.983	0.627	0.997	0.993
3AVB	0.792	0.981	0.842	0.819	0.988	0.882	0.897	0.988	0.893
AFIB	N/A	0.992	0.684	N/A	0.995	0.944	N/A	0.994	0.887
AVNRT	-0.106	0.990	0.860	0.026	0.991	0.935	0.855	0.997	0.988
AVRT	0.008	0.982	0.615	0.015	0.986	0.675	0.814	0.988	0.901
LBBB	0.786	0.942	0.745	0.942	0.992	0.876	0.870	0.993	0.791
NORMAL	0.836	0.986	0.875	0.817	0.991	0.884	0.802	0.984	0.899
PVC	0.710	0.972	0.791	0.782	0.976	0.674	0.840	0.987	0.645
RBBB	0.970	0.967	0.960	0.976	0.995	0.982	0.985	0.970	0.980
WPW	0.660	0.956	0.790	0.948	0.994	0.966	0.864	0.992	0.973



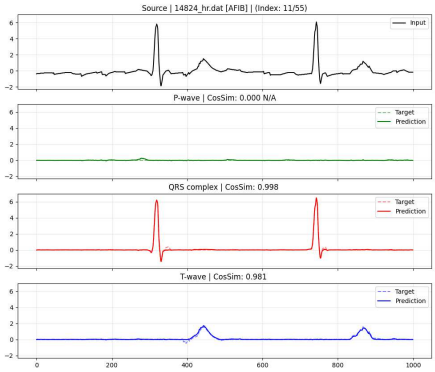
(a) LBBB Example



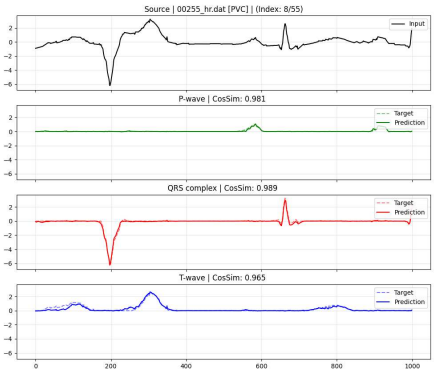
(b) RBBB Example



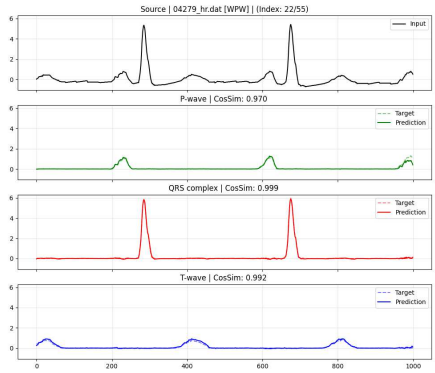
(c) 2AVB Example



(d) AFIB Example

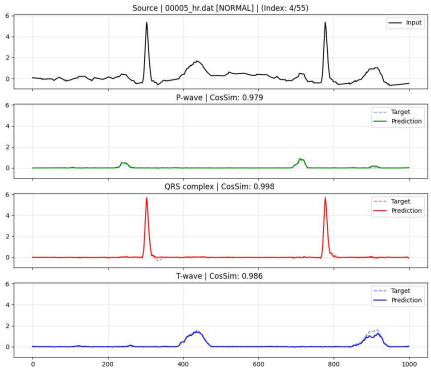


(e) PVC Example

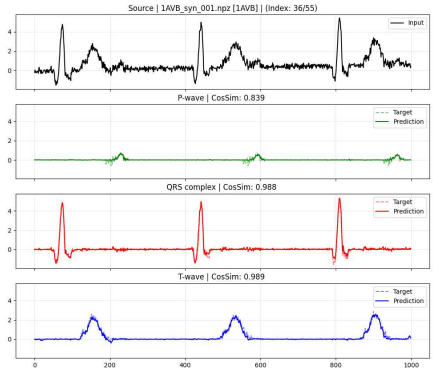


(f) WPW Example

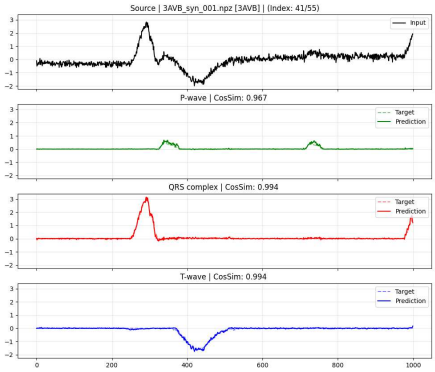
Figure 1: **Qualitative performance of the ESD-former (Model 3) on non-overlapping cases.** The raw regression outputs for P, QRS, and T components are shown below the original ECG sequence. *Abbreviations:* 2AVB, second-degree atrioventricular block; AFIB, atrial fibrillation; LBBB, left bundle branch block; PVC, premature ventricular contraction; RBBB, right bundle branch block; WPW, Wolff-Parkinson-White syndrome.



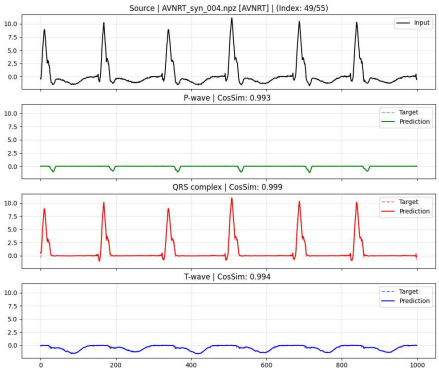
(a) Normal ECG Example



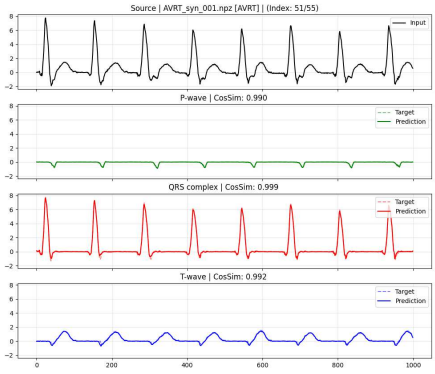
(b) 1AVB Example



(c) 3AVB Example



(d) AVRT Example



(e) AVNRT Example

Figure 2: **Qualitative performance of the ESD-former (Model 3) on overlapping cases.** Results for normal ECG and synthetic data. The raw regression outputs demonstrate continuous predictions during severe waveform superposition. *Abbreviations:* 1AVB, first-degree atrioventricular block; 3AVB, third-degree atrioventricular block; AVNRT, atrioventricular nodal reentrant tachycardia; AVRT, atrioventricular reentrant tachycardia; ECG, electrocardiogram.

Exploratory Real-World Feasibility and Concordance Analysis

To evaluate the proof-of-concept feasibility of the proposed ESD-former on real-world data, the model's performance was tested on continuous 5-minute resting ECG recordings[5]. Full 5-minute recordings present clinical challenges including baseline wander, high-frequency noise, and atypical morphologies such as premature ventricular contractions (PVCs). We analyzed two independent 5-minute recordings and compared the ESD-former's segmentation outputs against predictions generated by reference software (CineECG)[6].

A concordance analysis was performed to evaluate the agreement between the ESD-former and the baseline. For ventricular depolarization (QRS complexes), the ESD-former achieved a positive percent agreement (PPA) of 92.9% relative to CineECG (694 mutually agreed detections). This indicates high agreement for fundamental beat tracking.

For atrial activity (P waves), the ESD-former achieved an 82.7% PPA relative to the reference software (591 mutually agreed detections). Furthermore, the ESD-former detected 81 additional deeply buried P wave events not flagged by CineECG. Crucially, a subsequent manual expert review of these 81 instances confirmed them as true positive hidden atrial deflections rather than false-positive artifacts, highlighting the model's capacity to successfully identify low-amplitude or partially obscured signals missed by standard algorithms.

For repolarization (T wave) detection, the PPA was 76.0% (with 568 mutually agreed detections). The reference software detected an identical number of T waves and QRS complexes (747). In contrast, the ESD-former detected 648 T waves. Qualitative review of these discrepancies suggests that the ESD-former dynamically relies on the absolute energy of the decomposed signal, which may suppress T wave predictions during episodes of high-frequency noise or highly distorted PVCs.

As illustrated in Figure 3, the qualitative analysis indicates feasibility in handling common real-world ECG artifacts. The model generally identified deep negative T waves during PVCs without misclassifying adjacent noise as false P waves.

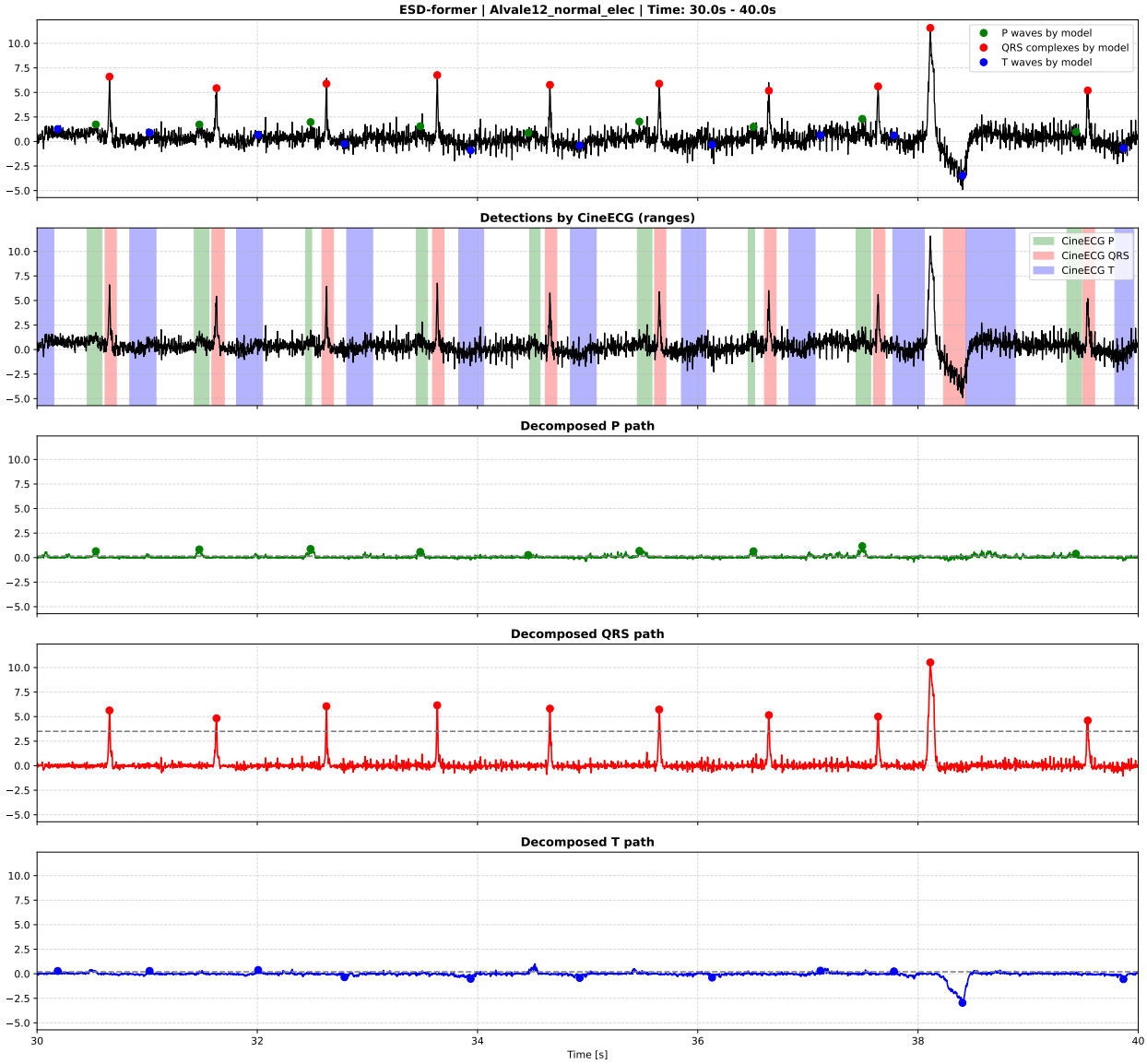


Figure 3: **Continuous 5-minute resting ECG analysis using the ESD-former.** The model's P-QRS-T peak detection behavior during morphological variations, demonstrating the handling of an inverted T wave of a premature ventricular contraction without triggering a P wave detection. *Abbreviations:* ECG, electrocardiogram; PVC, premature ventricular contraction.

Discussion

The exploratory evaluation of the ESD-former demonstrates the feasibility of moving beyond theoretical proxy-based classification toward continuous P-QRS-T signal extraction. While effective for reconstructing overlapping components in synthetic datasets, our evaluation highlighted specific edge cases, architectural limitations, and data dependencies that influence the model's inductive biases.

Morphological Confounding in Left Bundle Branch Block

While Model 3 achieved superior global performance, its P wave CosSim for left bundle branch block (LBBB) decreased from 0.942 (Model 2) to 0.870. The observed errors are consistent with the possibility that the model associates certain post-QRS negative deflections with atrial activity, an inductive bias likely developed from exposure to synthetic AVRT and AVNRT cases. Consequently, the network generated minor false-positive P wave candidate predictions within the pure, deeply inverted T waves characteristic of LBBB cases, slightly degrading the objective metric.

Architectural Limitations: The Case of Atrial Flutter

Explicit signal decomposition possesses inherent limitations when applied to continuous waveform abnormalities like atrial flutter (AFLT). Physiologically, the continuous macroreentrant flutter circuits (F waves) forming a sawtooth baseline are incompatible with discrete wave isolation. The present formulation was not designed to address continuous flutter activity. To address AFLT within a comprehensive pipeline, future approaches may need to route such signals through parallel classification models optimized specifically for macroreentrant circuits[7,8].

Impact of Data Scarcity on Second-Degree AV Block

The P wave CosSim for second-degree AV block (2AVB) plateaued at approximately 0.62 to 0.68. However, because the models were evaluated on merely 2 cases, this metric is highly susceptible to sample selection bias and essentially unstable. This should be treated as an exploratory observation highlighting the systemic bottleneck in contemporary automated electrocardiography: deep learning performance drops precipitously for underrepresented conduction disorders[9,10]. Algorithmic innovations cannot entirely compensate for the absolute absence of high-fidelity training data, necessitating the curation of dedicated datasets for future re-evaluation.

Study Limitations and Future Work

Several important limitations to this study must be acknowledged, underscoring its status as a proof-of-concept framework. First, the lack of an absolute ground truth for heavily superimposed arrhythmias in clinical records necessitated the use of synthetically generated overlapping events. While this guarantees known mathematical targets, it introduces a potential generator bias, as the model may learn to invert the synthetic construction rather than fully capture the electrophysiology of real-world superposition. Second, the clinical validation was exploratory, performed on only two 5-minute retrospective recordings. This sample size is insufficient to establish broad clinical robustness across diverse patients, hardware, or rhythm variations. Third, the reliance on Positive Percent Agreement (PPA) against a reference software limits the evaluation of over-segmentation; future large-scale studies must incorporate Positive Predictive Value (PPV) and formal false-positive analyses across larger clinical cohorts. Finally, while intra-patient consistency was enforced during synthetic generation, the rigorous patient-level separation across the aggregated public databases remains challenging, requiring future validation on a fully independent, locked external test cohort.

In conclusion, this study presents a foundational proof-of-concept for automated electrocardiogram analysis by shifting the paradigm toward explicit physiological signal decomposi-

tion. We introduced the ESD-former architecture, designed to disentangle complex arrhythmias into explicitly separated P, QRS, and T waveform tracks. The progressive ablation study demonstrated that exposure to overlapping waveforms improves the model's ability to recover those components on similarly constructed signals. While extensive external validation on real-world tachycardia cohorts is still required, this explicit signal separation framework provides an interpretable tool that may eventually support electrophysiological assessment and clinical analysis of complex arrhythmias.

Methods

Data Sources and Targeted Sampling

To explore the generalizability of the proposed ESD-former, data was aggregated from multiple publicly available, well-established electrocardiography databases. The foundational dataset used in this study was the Lobachevsky University Electrocardiography Database (LUDB)[11], which provides continuous delineations of P, QRS, and T waves. Additional records were integrated from the PhysioNet databases[12]. Specifically, selected cases were sourced from PTB-XL[13], Georgia Database, CPSC 2018, St. Petersburg INCART[1415], Ningbo, and Chapman-Shaoxing datasets[1617].

We performed a targeted, tag-based sampling procedure, actively mining these repositories specifically for underrepresented, complex minority classes to construct a curated distribution. These diverse databases served a dual purpose: they provided real-world ECG records for standard arrhythmias, and they acted as a repository of isolated physiological components utilized to synthetically generate overlapping arrhythmias. The final dataset consists of 467 unique patient records.

To evaluate feasibility in clinical scenarios, these records were supplemented with a subset of continuous, retrospective clinical ECG recordings obtained from a tertiary care hospital (University Hospital Basel, Switzerland). These data have been previously described in detail[5]. The study was carried out in accordance with the principles of the Declaration of Helsinki and all relevant guidelines and regulations. The retrospective experimental protocol was approved by the institutional review board and ethical committee associated with the University Hospital Basel. Informed consent was obtained from all subjects and/or their legal guardians for the clinical data collection. For the supplementary public datasets, formal ethical review was exempt due to their fully anonymized nature. The authors had full access to and take full responsibility for the integrity of the data.

Signal Preprocessing and Input Configuration

All raw ECG signals were preprocessed to ensure uniformity before input into the neural network architecture. The data natively sampled at 500 Hz was directly utilized without artificial resampling. The model was configured to operate on a single-lead input, specifically standardizing the extraction to Lead II, as identified dynamically via signal headers. Amplitude normalization was performed using a standard Z-score standardization: $(signal - \mu)/(\sigma + \epsilon)$, where $\epsilon = 1e - 6$ prevented division by zero. Furthermore, during the extraction of distinct physiological waveform segments, a linear baseline correction was applied based on the slope between the onset and offset points to eliminate baseline wander artifacts.

The input to the ESD-former was constrained to a fixed 1D temporal sequence of 1000 samples (equivalent to a 2.0-second window). Longer signals were accommodated via random temporal cropping during training, whereas shorter signals underwent zero-padding to meet the required dimension.

Dataset Composition and Partitioning

To accurately evaluate the model's capability, the dataset comprises two distinct categories of data: real-world clinical records and synthetically composed arrhythmias. The real-world subset includes non-overlapping or distinct-wave morphologies: normal sinus rhythm (n=199), Wolff-Parkinson-White syndrome (WPW, n=38), left bundle branch block (LBBB, n=34), premature ventricular contractions (PVC, n=31), atrial fibrillation (AF, n=22), right bundle branch block (RBBB, n=15), and second-degree atrioventricular block (2AVB, n=8).

Using the physiological components extracted from the aforementioned databases, 120 extreme cases were synthetically generated: first-degree atrioventricular block (1AVB, n=30), third-degree atrioventricular block (3AVB, n=30), atrioventricular nodal reentrant tachycardia (AVNRT, n=30), and atrioventricular reentrant tachycardia (AVRT, n=30).

Cases of atrial flutter (AFLT) were explicitly excluded from this study. Prior to model

training, the complete dataset of 467 records was partitioned into a training set (80%) and a validation set (20%). This partitioning utilized an unseeded stratified split to ensure proportional representation of all primary arrhythmia classes across both sets.

Data Annotation and Refinement

A custom software application, ECG Annotator, was developed to visually inspect, mark, delete, or fine-tune the exact onset and offset boundaries of P, QRS, and T waves across the extracted records. This process was directly applied to establish the ground truth for the real-world dataset subset, and to isolate high-fidelity physiological templates.

For highly complex arrhythmias characterized by severe waveform superposition, standard manual boundary annotation of the complete signal is physiologically unfeasible. Consequently, the ground truth for these specific classes was not manually annotated but mathematically derived by combining the previously extracted, clean physiological templates through the synthetic generation pipeline.

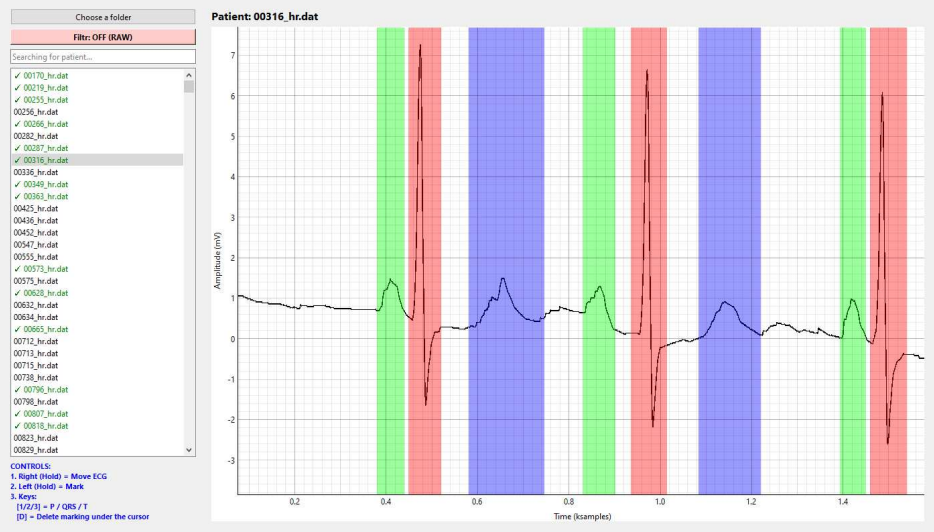


Figure 4: **User interface of the ECG Annotator software.** The tool enables operators to adjust onset and offset boundaries for ground truth masks. Components are color-coded: P waves (green), QRS complexes (red), and T waves (blue).

ESD-former Architecture

Designed as a multi-head regression model, the ESD-former leverages a 1D U-Net[18] topology integrated with a Transformer bottleneck[19,20]. The convolutional encoder extracts local morphological features, while the Transformer bottleneck captures global rhythm context (e.g., R-R intervals). The decoder rebuilds the temporal resolution, branching into three independent regression heads to output explicitly separated physiological tracks.

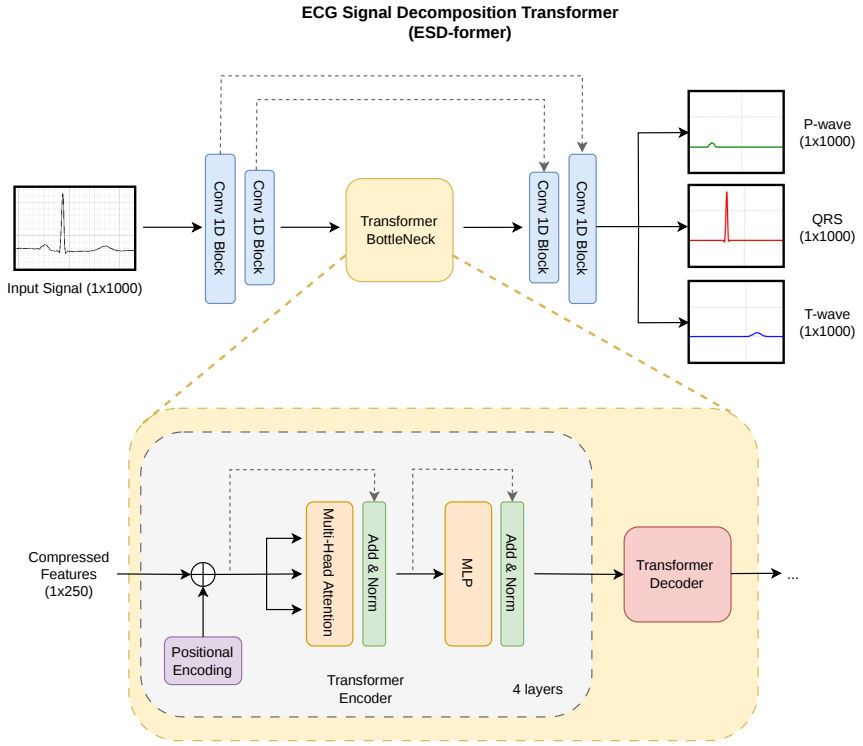


Figure 5: **Architecture of the proposed ESD-former.** The 1D input ECG signal is processed by a CNN encoder, followed by a Transformer bottleneck capturing global dependencies at a reduced temporal resolution. Decoder blocks upsample the signal, utilizing skip connections (gray dotted lines) to recover spatial precision, branching into three independent regression heads for continuous P, QRS, and T wave prediction. *Abbreviations:* CNN, convolutional neural network; ECG, electrocardiogram; MLP, multilayer perceptron.

Synthetic Overlapping Waveform Generation

We developed a synthetic generation pipeline that mathematically superimposes isolated, patient-specific templates (P, QRS, T) to simulate complex arrhythmias. By manipulating conduction intervals and rates, we generated regression targets for extreme cases. Real clinical noise was injected to simulate realistic signal-to-noise ratios. The final synthesized signal S_{total} is defined as:

$$S_{total}(t) = P(t) + QRS(t) + T(t) + Noise(t) \tag{1}$$

The comprehensive generation protocol is detailed in Supplementary Information S2.

Model Optimization and Evaluation

The network was optimized using a custom composite objective function, termed *ShapeAwareLoss*, operating independently on each output head. It effectively penalizes absolute structural dissimilarities by bypassing isoelectric baseline segments through dynamic boolean masking. The loss function combines a magnitude-based Mean Squared Error (L_{MSE}) with a structure-based Cosine Distance penalty (L_{Cos}):

$$L_{Total} = L_{MSE} + \lambda \cdot L_{Cos} \tag{2}$$

where $\lambda = 0.5$ balances the penalties. The model was trained from scratch for 100 epochs with a batch size of 32, using the AdamW optimizer[21] and a OneCycleLR scheduler[22] (max learning rate 0.001). The primary evaluation metric for decomposition quality was Cosine Similarity (CosSim).

Progressive Training and Ablation Strategy

To systematically evaluate the ESD-former's capability, we designed a progressive training strategy yielding three models. Stage 1 (Model 1) was trained exclusively on normal sinus rhythm ECGs. Stage 2 (Model 2) expanded the training dataset to include complex, yet temporally distinct non-overlapping arrhythmias. Stage 3 (Model 3) exposed the model to the full dataset, which included the synthetically generated overlapping cases.

Use of Artificial Intelligence

During the preparation of this manuscript, the authors utilized Gemini (Google) exclusively as a writing and formatting companion. All AI-assisted text was rigorously reviewed, revised, and approved by the human authors, who maintained full intellectual control and take complete responsibility for the accuracy and integrity of the final publication.

Data availability

The public ECG datasets analyzed during the current study are available in the PhysioNet repository. Direct links to the primary sources and specific versions utilized are as follows:

- Lobachevsky University Electrocardiography Database (LUDB) v1.0.1: `https://physionet.org/content/ludb/1.0.1/`
- PTB-XL, a large publicly available electrocardiography dataset v1.0.3: `https://physionet.org/content/ptb-xl/1.0.3/`
- A large scale 12-lead electrocardiogram database for arrhythmia study (Chapman-Shaoxing/Ningbo) v1.0.0: `https://physionet.org/content/ecg-arrhythmia/1.0.0/`
- St Petersburg INCART 12-lead Arrhythmia Database v1.0.0: `https://physionet.org/content/incartdb/1.0.0/`
- Additional records originating from the PhysioNet/Computing in Cardiology Challenge 2021 (including Georgia and CPSC subsets) v1.0.3: `https://physionet.org/content/challenge-2021/1.0.3/`

The clinical datasets (University Hospital Basel) supporting the findings of this study are not publicly available due to institutional privacy restrictions but may be requested from the corresponding author upon reasonable request and with permission of the providing institution.

Code availability

The custom code used for the ESD-former architecture and synthetic signal generation during the current study is available from the corresponding author on reasonable request.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions

M.S. conceptualized the model, performed data annotation, developed the code, and drafted the manuscript. P.M.v.D. supervised the research and critically reviewed the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Schläpfer, J. & Wellens, H. J. Computer-interpreted electrocardiograms: benefits and limitations. *Journal of the American College of Cardiology* **70,** 1183–1192 (2017).

2. Luz, E. J. d. S., Schwartz, W. R., Cámara-Chávez, G. & Menotti, D. ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine* **127,** 144–164 (2016).

3. Hannun, A. Y. *et al.* Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine* **25,** 65–69 (2019).

4. Wu, H., Xu, J., Wang, J. & Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)* **34,** 22419–22430 (2021).

5. Asatryan, B. *et al.* Man vs machine: Performance of manual vs automated electrocardiogram analysis for predicting the chamber of origin of idiopathic ventricular arrhythmia. *Journal of cardiovascular electrophysiology* **31,** 410–416 (2020).

6. Boonstra, M. J., Brooks, D. H., Loh, P. & van Dam, P. M. CineECG: a novel method to image the average activation sequence in the heart from the 12-lead ECG. *Computers in biology and medicine* **141,** 105128 (2022).

7. Woo, S. *et al.* ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 16133–16142 (2023).

8. Melgani, F. & Bazi, Y. Classification of electrocardiogram signals with support vector machines and particle swarm optimization. *IEEE transactions on information technology in biomedicine* **12,** 667–677 (2008).

9. Ribeiro, A. H. *et al.* Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature communications* **11,** 1–9 (2020).

10. Strodthoff, N., Wagner, P., Schaaf, T. & Roletz, W. Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE Journal of Biomedical and Health Informatics* **25,** 1519–1528 (2020).

11. Kalyakulina, A. I. *et al.* LUDB: a new open-access validation tool for electrocardiogram delineation algorithms. *IEEE Access* **8,** 186181–186190 (2020).

12. Goldberger, A. L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101,** e215–e220 (2000).

13. Wagner, P., Strodthoff, N., Bousseljot, R.-D., Samek, W. & Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset. *PhysioNet.* Version 1.0.3. `https://doi.org/10.13026/kfzx-aw45` (Nov. 2022).

14. Perez Alday, E. A. *et al.* Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020. *PhysioNet.* Version 1.0.2. `https://doi.org/10.13026/dvyd-kd57` (July 2022).

15. Perez Alday, E. A. *et al.* Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. *Physiological Measurement* **41,** 124003. `https://doi.org/10.1088/1361-6579/abc960` (Dec. 2020).

16. Zheng, J., Guo, H. & Chu, H. A large scale 12-lead electrocardiogram database for arrhythmia study. *PhysioNet.* Version 1.0.0. `https://doi.org/10.13026/wgex-er52` (Aug. 2022).

17. Zheng, J. *et al.* Optimal multi-stage arrhythmia classification approach. *Scientific reports* **10,** 2898 (2020).

18. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 234–241 (2015).

19. Vaswani, A. *et al.* Attention is all you need. In *Advances in neural information processing systems (NeurIPS)* **30** (2017).

20. Chen, J. *et al.* Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306* (2021).

21. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization (AdamW). *International Conference on Learning Representations (ICLR)* (2019).

22. Smith, L. N. & Topin, N. Super-convergence: Very fast training of neural networks using large learning rates (OneCycleLR). *Artificial intelligence and machine learning for multi-domain operations applications* **11006,** 1100612 (2019).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- UnmaskingHiddenWavesTransformerBasedECGDecompositionforComplexArrhythmiasscientificreports.pdf