

1 An interpretable tensor model for identifying atrial fibrillation in
2 sinus-rhythm ECG: block-term tensor regression on PTB-XL,
3 benchmarked against deep and foundation models

4 Dries Cornelissen¹ Axel Faes^{2,*}

5 ¹ Data Science Institute (DSI), Hasselt University, Hasselt, Belgium.

6 ² Biomedical Data Sciences, Data Science Institute (DSI) and Biomedical Research Institute (BIOMED),
7 Hasselt University, Hasselt, Belgium.

8 * Corresponding author: Axel Faes, axel.faes@uhasselt.be

9 **Manuscript facts.** Word count from the title page through the figure legends, references excluded:
10 4031. Abstract word count: 197. Number of figures: 2. Number of tables: 3.

11 **Author contributions.** Dries Cornelissen: Data curation, Investigation. Axel Faes: Conceptu-
12 alization, Methodology, Software, Formal analysis, Validation, Visualization, Supervision, Writing
13 (original draft), Writing (review and editing). Both authors approved the final manuscript.

Abstract

Atrial fibrillation may leave a substrate that an electrocardiogram in sinus rhythm already carries, even when the trace looks normal. Reading it usually requires a deep network trained on private cohorts that does not expose what it uses. We test whether an intrinsically interpretable tensor model can identify atrial fibrillation in sinus rhythm on public data. On PTB-XL we tensorize the multilead median beat into a samples-by-time-by-lead array and fit block-term tensor regression (BTTR), a supervised model whose parameters are readable spatio-temporal weight maps. We benchmark BTTR against clinical features, a convolutional network, an unsupervised CP surrogate, and frozen embeddings from HuBERT-ECG, a foundation model pretrained on 9.1 million recordings. BTTR matches or exceeds every baseline: all paired patient-level bootstrap differences in AUROC are non-negative. The substrate maps are faithful to the decision by construction and stable across resampling: the per-lead weight profile reproduces at Pearson $r = 0.88$. R-peak registration is necessary for the tensor model and immaterial for a convolutional network under the same protocol. Interpretability therefore comes at no accuracy cost. To our knowledge this is the first intrinsically interpretable tensor model for this task and the first such study using only public data.

Keywords: Atrial Fibrillation; Electrocardiography; Machine Learning; Interpretable machine learning; Block-term tensor regression

1 Introduction

Atrial fibrillation is the most common sustained arrhythmia and a leading cause of ischemic stroke. Much of its burden falls on patients who are in sinus rhythm at the moment of recording, because the arrhythmia is paroxysmal and intermittent. A standard screening electrocardiogram captured between episodes can therefore look entirely normal while the underlying atrial substrate is already present. Reading that substrate from a single sinus-rhythm trace would let clinicians flag patients for monitoring before an event, which is the practical motivation behind a line of work that began with the demonstration that a neural network can identify atrial fibrillation from a sinus-rhythm electrocardiogram [2].

That result has been replicated and extended across cohorts and device classes [3, 8, 11, 18, 21]. The models are accurate, and the clinical signal is real. Two features recur across the literature and limit what the field can learn from it. First, the strongest results sit on private institutional data, which makes head-to-head comparison between methods difficult and external scrutiny harder still. Second, the models are black boxes. A clinician is told that a normal-looking trace carries risk but not which part of the beat, or which leads, the model is reacting to. For a finding that contradicts the visible reading of the electrocardiogram, the absence of an account is a real obstacle to trust and to physiological interpretation.

The usual response is to attach a post-hoc explanation: saliency maps, attention overlays, or surrogate attributions fitted after the fact [12–14, 17]. These methods estimate what a fixed black box might have used. They do not change what it actually used, and different attribution methods can disagree on the same model. The explanation is a separate object from the decision, and its fidelity is itself in question.

Tensor models offer a different starting point. Arranging the multilead beat as a samples-by-time-by-lead array turns the electrocardiogram into a structured signal on which a supervised low-rank model can be fit directly, and the fitted factors are the decision. Each per-block time-by-lead weight map is a parameter of the model, so it reads as a spatio-temporal signature rather than an approximation of one. Block-term tensor regression (BTTR) is one such estimator [7]. It was introduced for a different problem; here we transfer it to a new task and ask whether an intrinsically

interpretable tensor model can compete with the black-box approach on its own ground, using only public data.

We study identification of atrial fibrillation in sinus rhythm on PTB-XL [20], a public clinical electrocardiogram database. The signal pipeline builds an R-peak-aligned median beat per lead, stacks it into the input tensor, and an alignment ablation isolates what that registration step contributes to each model family. We evaluate BTTR against clinical features, a convolutional network in the style of [2], a capacity-matched unsupervised CP surrogate, and a foundation model [4]. The benchmark is like-for-like: identical cohorts, identical patient-grouped folds, and a paired patient-level bootstrap for every method comparison.

The contributions are as follows.

1. We present, to our knowledge, the first intrinsically interpretable tensor model for identifying atrial fibrillation in sinus rhythm, with substrate maps that are stable across resampling and faithful by construction, evaluated entirely on the public PTB-XL database [20].
2. We show that beat alignment is necessary for tensor methods but not for the convolutional network under the same protocol, locating alignment as a model-dependent requirement and complementing [15].
3. On a like-for-like benchmark, BTTR matches or exceeds clinical features, a convolutional network, an unsupervised CP surrogate, and a foundation model, with all paired patient-level differences in AUROC non-negative.
4. We show that this interpretability comes at no accuracy cost, since the glass-box model is not outperformed by any black-box baseline, the 9.1-million-recording foundation model included.

2 Related work

Atrial fibrillation in sinus rhythm. The task of inferring atrial substrate from a sinus-rhythm electrocardiogram was established with a convolutional network trained on a large institutional cohort [2], and the finding has held up across replications, populations, and recording hardware [3, 8, 11, 18, 21]. These studies share an architecture-and-data recipe: deep supervised learning on private data, with the model treated as an end-to-end predictor. Performance is the reported quantity, and the internal representation is not.

Interpretability. For clinical electrocardiogram models the dominant interpretability strategy is post-hoc. Saliency, gradient, and attention-based attributions are computed for a trained network and read as evidence of what it used [12–14, 17]. Such attributions are estimates layered onto a fixed model, and their stability and fidelity are active concerns [9]. Intrinsic interpretability takes the alternative position that the model’s parameters should be the explanation. Our work sits on that side: the substrate map is a fitted factor, not an attribution.

Tensor and block-term models on the electrocardiogram. Tensor decompositions have a long history on the electrocardiogram, almost entirely in an unsupervised, source-separation role: extracting atrial activity, separating maternal and fetal signals, and related blind decompositions [5, 6, 16]. Supervised block-term modeling, where the decomposition is fit to a label, has been developed on the electroencephalogram [19], and the block-term tensor regression estimator we use was introduced in that broader context [7]. We do not claim to be first to apply tensors

to the electrocardiogram, which would be inaccurate; our contribution is to bring a supervised, intrinsically interpretable block-term estimator to the specific task of identifying atrial fibrillation in sinus rhythm, and to benchmark it openly.

Alignment and foundation models. The need to align beats before tensorization has been examined for ECG learning [15], and we revisit it here under a controlled comparison across model families. Foundation models pretrained on large unlabeled electrocardiogram corpora have recently appeared; HuBERT-ECG [4] provides reusable embeddings and serves here as the strongest available black-box baseline, again an existing estimator transferred to our task rather than a method we propose.

3 Methods

3.1 Data and cohorts

We use PTB-XL [20], a public clinical 12-lead electrocardiogram database with patient identifiers and diagnostic labels. The task is identification of atrial fibrillation in sinus rhythm, and we are explicit that this is not future prediction: PTB-XL carries no longitudinal follow-up. A sinus-rhythm record is labeled positive if the same patient has a concurrent atrial-fibrillation recording within 31 days. The atrial-fibrillation recording may either precede or follow the sinus-rhythm trace, and the labeling does not order the two. We therefore describe the setting as concurrent-association identification rather than detection of a pre-arrhythmic substrate, and we keep that distinction throughout.

The primary cohort, AFIB, comprises 42 positive patients and 126 controls. A secondary cohort, AFIB-or-flutter, adds atrial flutter and comprises 55 positive and 165 control patients. Both cohorts are patient-grouped with one record per patient, and controls are matched at roughly three to one. Grouping at the patient level prevents records from the same patient appearing in both training and test partitions.

3.2 Preprocessing and tensorization

Each record is denoised with a wavelet filter chosen to preserve the low-frequency atrial band, so that P-wave and PR morphology survive the cleaning step. R-peaks are detected, and a median beat is formed over the window from -250 to $+400$ ms around the R-peak. Stacking median beats yields a samples-by-time-by-lead tensor that is the input to the tensor models. For the alignment ablation we also build an unaligned tensorization that uses a fixed full window without R-peak registration, so that the only difference between conditions is whether beats are aligned.

3.3 Models

BTTR. Block-term tensor regression [7] is a supervised, inductive, deflation-based estimator. The binary label is mapped to a ± 1 target and fit by least squares, giving a linear decision score rather than a logistic one. The model proceeds in K iterations. At each iteration a block is extracted, and both the data-tensor residual and the target residual are deflated before the next block is fit, so each block accounts for structure the previous blocks left behind. Automatic component estimation selects each block’s multilinear rank within the training fold, searching a grid defined by signal-to-noise and energy-ratio criteria. With the rank-1 setting enforced, each block is rank-1 in the time and lead modes, so the model is a supervised, label-fit sum of rank-1 terms. That structure

is what makes each block’s time-by-lead weight map directly readable as a substrate signature. The block count is fixed a priori at $K = 8$ and is never tuned on test folds.

Clinical features. Seven interval and morphology features covering RR variability, P-wave, and QRS characteristics feed a logistic regression. This baseline encodes the standard electrophysiological reading.

CP surrogate. An unsupervised rank-8 PARAFAC decomposition is fit, test records are projected onto the fixed training factors, and a logistic regression is fit on the resulting coordinates. The rank matches the BTTR block count so that capacity is held constant; the contrast between this surrogate and BTTR isolates the effect of supervision in the decomposition itself.

Convolutional network. A small, regularized one-dimensional convolutional network in the style of [2] is trained with class weighting and a three-seed ensemble to stabilize its scores. It stands in for the deep-learning approach at a scale appropriate to the cohort size.

Foundation model. HuBERT-ECG [4], pretrained on 9.1 million electrocardiograms, supplies frozen 512-dimensional embeddings on which we fit a logistic linear probe. This is an existing estimator used as the strongest black-box baseline, with the foundation model itself held fixed.

3.4 Evaluation

The primary benchmark uses patient-grouped 5-fold cross-validation. We report AUROC with patient-level bootstrap 95% confidence intervals, resampling patients over the pooled out-of-fold scores. Comparisons between methods use a paired patient-level bootstrap of the AUROC difference, which is the correct object for the question of whether one method exceeds another; overlap of two marginal confidence intervals is not, since it ignores the correlation between methods evaluated on the same patients.

Two analyses use a separate protocol. The alignment ablation and the interpretability-stability analysis refit the model over ten random control draws and are reported as mean $\pm$ standard deviation across those resamples. These resampling figures are not directly comparable to the cross-validation point estimates, and we state that explicitly wherever both appear.

Finally, to calibrate task difficulty, we include a detection-context cohort of overt sinus rhythm versus overt atrial fibrillation, 150 records per class, evaluated on a single split. This is illustrative and supports no accuracy claim of its own.

4 Results

4.1 Identifying atrial fibrillation in sinus rhythm

Table 1 reports per-method AUROC with confidence intervals for both cohorts, Fig. 1 shows the same results as a forest plot, and Table 2 gives the paired differences that drive the comparison.

On the primary cohort (42 positive, 126 control) BTTR reaches an AUROC of 0.620 [0.509, 0.727], with HuBERT-ECG at 0.616 [0.514, 0.723], the convolutional network at 0.587 [0.475, 0.693], clinical features at 0.584 [0.484, 0.684], and the CP surrogate at 0.581 [0.474, 0.686]. On the secondary cohort (55 positive, 165 control) BTTR reaches 0.700 [0.621, 0.776], the CP surrogate 0.693 [0.608, 0.774], clinical features 0.673 [0.596, 0.748], the convolutional network 0.644 [0.552, 0.731], and HuBERT-ECG 0.641 [0.557, 0.723].

Table 1 Identification of atrial fibrillation in sinus rhythm. AUROC with 95% patient-level bootstrap confidence intervals, all methods under identical patient-grouped 5-fold cross-validation. HuBERT-ECG is a self-supervised ECG foundation model (pretrained on 9.1 million recordings) used as a frozen linear probe; CNN is a compact Attia-style one-dimensional network; CP is the unsupervised tensor surrogate.

Method	AFIB (primary, 42 pos) AUROC [95% CI]	AFIB or flutter (secondary, 55 pos) AUROC [95% CI]
BTTR (ours)	0.620 [0.509, 0.727]	0.700 [0.621, 0.776]
HuBERT-ECG (foundation model)	0.616 [0.514, 0.723]	0.641 [0.557, 0.723]
CNN (Attia-style)	0.587 [0.475, 0.693]	0.644 [0.552, 0.731]
clinical features	0.584 [0.484, 0.684]	0.673 [0.596, 0.748]
CP surrogate (unsupervised)	0.581 [0.474, 0.686]	0.693 [0.608, 0.774]

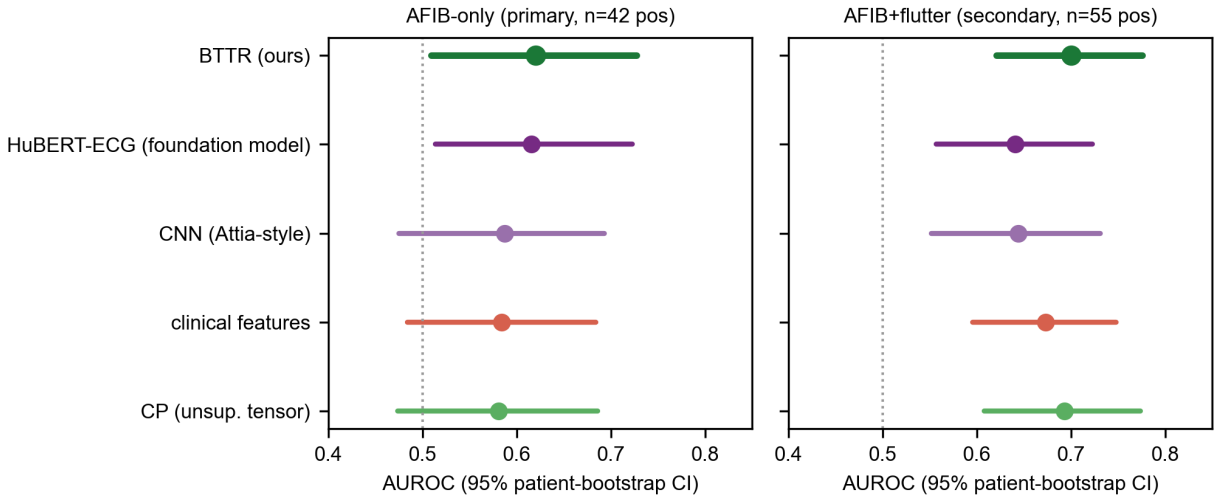


Fig. 1 AUROC with 95% patient-level bootstrap confidence intervals for all methods on both cohorts. BTTR matches or exceeds every baseline, including the HuBERT-ECG foundation model; the formal comparison is the paired difference in Table 2.

The paired test is what settles the comparison, and it is favorable across the board. Every paired BTTR-minus-baseline difference in AUROC is non-negative. On the primary cohort the difference is +0.004 against HuBERT-ECG, +0.034 against the convolutional network, +0.036 against clinical features, and +0.040 [−0.011, +0.094] against the CP surrogate. On the secondary cohort it is +0.059 against HuBERT-ECG, with BTTR higher in 90% of resamples, +0.055 against the convolutional network in 91% of resamples, +0.027 against clinical features, and +0.010 against the CP surrogate. BTTR matches or exceeds every baseline and is not outperformed by any of them, the foundation model included. Interpretability here costs nothing in accuracy. The benefit of supervision within the decomposition is clearest on the primary cohort, where BTTR exceeds the capacity-matched unsupervised CP surrogate by 0.040 in 94% of resamples; on the broader secondary cohort the surrogate nearly matches it, so the supervised advantage is largest where positive examples are scarcest.

The block count is not a tuned optimum. Sweeping K on the primary cohort gives AUROC

Table 2 Paired BTTR-minus-baseline AUROC differences (patient-level bootstrap). All differences are non-negative on both cohorts; the value in parentheses is the fraction of resamples in which BTTR exceeds the baseline.

Baseline	Primary Δ (fraction)	Secondary Δ (fraction)
HuBERT-ECG (foundation model)	+0.004 (0.52)	+0.059 (0.90)
CNN (Attia-style)	+0.034 (0.73)	+0.055 (0.91)
clinical features	+0.036 (0.72)	+0.027 (0.71)
CP surrogate	+0.040 (0.94)	+0.010 (0.72)

Table 3 Alignment ablation under a matched protocol. BTTR is reported over ten resamples (mean $\pm$ SD); the convolutional network is reported with its patient-bootstrap 95% confidence interval. Beat alignment is necessary for the tensor model and immaterial for the convolutional network.

Model	Tensorization	Primary	Secondary
BTTR	aligned median beat	0.666 ± 0.030	0.715 ± 0.012
BTTR	unaligned window	0.462 ± 0.032	0.434 ± 0.029
CNN	aligned	0.587 [0.475, 0.693]	0.644 [0.552, 0.731]
CNN	unaligned	0.586 [0.473, 0.694]	0.635 [0.543, 0.725]

0.627, 0.641, 0.666, 0.681, and 0.689 for $K = 4, 6, 8, 12, 16$, so the model is stable across this range. We fixed $K = 8$ in advance as a parsimonious block count that keeps the set of weight maps small and readable, accepting that the larger end of the range yields marginally higher AUROC.

4.2 Alignment is necessary for tensor methods but not for the convolutional network

Table 3 reports the alignment ablation. For BTTR, beat alignment makes the difference between a working model and chance. Aligned, BTTR reaches 0.666 ± 0.030 on the primary cohort and 0.715 ± 0.012 on the secondary cohort. Unaligned, it collapses to 0.462 ± 0.032 and 0.434 ± 0.029 , which is at chance. The low-rank tensor structure depends on beats being registered, and without registration there is nothing coherent for the decomposition to fit.

The convolutional network behaves differently under the same protocol. It is essentially insensitive to alignment: 0.587 [0.475, 0.693] aligned against 0.586 [0.473, 0.694] unaligned on the primary cohort, and 0.644 [0.552, 0.731] against 0.635 [0.543, 0.725] on the secondary cohort. The convolution learns its own shift tolerance, so registration adds little. Alignment is therefore a requirement of the tensor family rather than of supervised ECG learning in general, which complements the alignment analysis of [15] by separating the two model classes.

4.3 Interpretability: stable and faithful substrate maps

The block maps are model parameters, so they are faithful to the decision by construction; there is no separate attribution step whose fidelity could be questioned. Fig. 2 shows the recovered maps.

Stability is the property worth testing, and the maps are stable. Across ten resamples the per-lead weight profile reproduces at Pearson $r = 0.88 \pm 0.06$, so the recovered signature is reproducible across draws rather than an artifact of a single fortunate split. Because the maps are the fitted

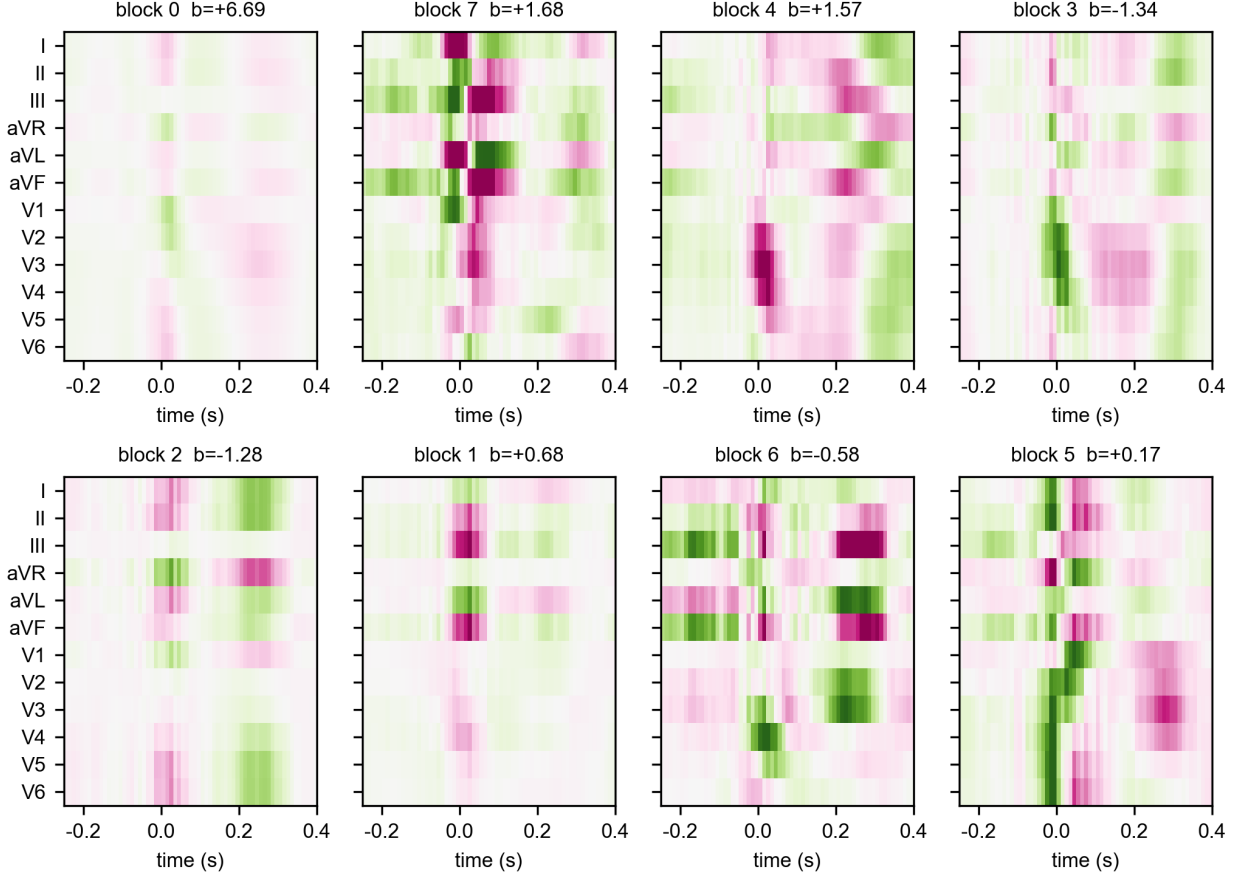


Fig. 2 BTTR block weight maps on the primary cohort with $K = 8$. Each panel is one block’s time-by-lead weight map with its scalar coefficient; color encodes the signed contribution. The maps are fitted model parameters. Across ten resamples the per-lead profile reproduces at Pearson $r = 0.88$.

model parameters, they are faithful to the decision by construction; that the decision is driven by the label, rather than by chance structure, follows from the above-chance discrimination reported above rather than from a post-hoc attribution.

Reading where in the beat the weight concentrates gives a concrete result that a P-wave-focused analysis would not surface. The discriminative weight is distributed across the cardiac cycle. Roughly 0.17 of the total weight falls in the atrial window covering the P-wave and PR segment, with the remainder spread over the QRS (about 0.27), the ST segment (about 0.16), and the T-wave (about 0.40). The lead profile is broad, spanning the inferior leads and the mid-precordium. The signature is thus weighted toward ventricular depolarization and repolarization rather than the P-wave alone. Established sinus-rhythm atrial markers, such as P-wave indices [10], target the atrial component, which here accounts for a minority of the discriminative weight. This is a substantive use of the glass-box property: the model reports that its decision draws on distributed morphology, and a method that examined only the P-wave would have missed it.

The localization is not an effect of flutter. The primary cohort excludes flutter, and adding flutter in the secondary cohort leaves the atrial-weight fraction essentially unchanged at 0.17 against 0.17. We confine the interpretability claims to stability, faithfulness, and the measured weight

distribution, and we do not assert a specific electrophysiological mechanism behind them.

4.4 Detection context

On the overt detection cohort, sinus rhythm against atrial fibrillation at 150 records per class on a single split, BTTR reaches 0.875 [0.835, 0.913], clinical features 0.860 [0.816, 0.901], and the CP surrogate 0.881 [0.841, 0.918]. Deep-learning detection of overt atrial fibrillation runs near ceiling [17, 20]. Overt detection is a solved and visible problem; the methods cluster together here, and on a single illustrative split we make no accuracy claim. The point of the contrast is that the value lies in the harder identification task, where the substrate is not visible to a reader.

5 Discussion

The central result is a supervised tensor model that exposes a stable, faithful spatio-temporal signature alongside its prediction. Among the methods compared, BTTR was the only one to do so, and it did so at no accuracy cost: on identical public data, with identical patient-grouped folds, BTTR matched or exceeded a convolutional network and a foundation model pretrained on 9.1 million electrocardiograms, and every paired difference in AUROC was non-negative. The common trade-off, in which a glass-box model is accepted as a concession against accuracy, did not appear in this benchmark.

The map analysis adds a finding that the glass-box property made visible. The discriminative signature is distributed across the beat and weighted toward ventricular repolarization rather than confined to the P-wave. We present this as a data-driven observation rather than a mechanism. It is plausibly consistent with ventricular involvement in the atrial-arrhythmia substrate, but we state that only as plausible and not as proven; the model reports where its weight sits, not why the physiology puts it there.

The alignment result carries a methodological message. Beat alignment was necessary for the tensor model and immaterial for the convolutional network, so alignment is a property of the tensor family and not a universal preprocessing law. Practitioners choosing between model classes should treat registration as a model-dependent decision.

Scope. Four limits bound the claims. First, the task is concurrent-association identification, not future prediction. The labeling does not order the sinus-rhythm and atrial-fibrillation recordings, so we do not claim a pre-arrhythmic-substrate interpretation. Second, an absolute AUROC of 0.62 to 0.70 is a feasibility-level result on a small public cohort, in the range reported for clinical risk scores for incident atrial fibrillation such as CHARGE-AF [1]. It is appropriate to a methods and benchmarking contribution and not to a deployable screening tool, and we say so plainly. Third, the study rests on a single public cohort, and external validation on an independent database is the natural next step. Fourth, flutter enters only in the secondary cohort.

Future work. Genuine incident-AF prediction needs dated-outcome cohorts in which the sinus-rhythm recording demonstrably precedes the arrhythmia, and assembling such a cohort is the most direct extension. On the method side, interpretable extensions of the block-term estimator, such as regularized or discriminant variants, could sharpen the signature while keeping the maps readable.

6 Conclusion

We applied an intrinsically interpretable tensor model to the identification of atrial fibrillation in sinus rhythm on public data, and the four contributions hold up affirmatively. First, BTTR is, to our knowledge, the first intrinsically interpretable tensor model for this task, with substrate maps that are stable across resampling at Pearson $r = 0.88$ and faithful by construction, demonstrated entirely on PTB-XL. Second, beat alignment was necessary for the tensor model but not for the convolutional network, locating alignment as a model-dependent requirement. Third, on a like-for-like benchmark BTTR matched or exceeded clinical features, a convolutional network, a CP surrogate, and a foundation model, with all paired differences in AUROC non-negative. Fourth, this interpretability came at no accuracy cost, since the glass-box model was not outperformed by any black-box baseline. The recovered signature is distributed across the beat and weighted toward ventricular repolarization, an observation we report as the glass box makes it visible, stable and faithful, without claiming a mechanism.

Declarations

Funding. This work was supported by the Flanders AI Research Program (FAIR), an Impulse Program of the Flemish Government (Flemish Ministry of Economy, Science and Innovation).

Competing interests. The authors declare no competing interests.

Ethics approval. This study is a secondary analysis of PTB-XL, a publicly available, fully de-identified electrocardiogram database, and did not involve new human-subjects research; no additional ethics approval was required.

Consent to participate. Not applicable, because the study analyzes a public de-identified database.

Consent for publication. Not applicable, because the study analyzes a public de-identified database.

Data availability. The PTB-XL database (version 1.0.3) is publicly available [20].

Code availability. The code to reproduce the cohorts, models, and analyses is available at <https://github.com/TheAxeC/bttr-af-ecg>.

Use of AI-assisted tools. AI-based writing-assistance tools were used to help draft and edit the manuscript text. All scientific content, analyses, numerical results, and interpretations were produced and verified by the authors, who take full responsibility for the work.

References

- [1] Alvaro Alonso, Bouwe P. Krijthe, Thor Aspelund, Katherine A. Stepas, Michael J. Pencina, Carlee B. Moser, Moritz F. Sinner, Nona Sotoodehnia, João D. Fontes, A. Cecile J. W. Janssens, Richard A. Kronmal, Jared W. Magnani, Jacqueline C. Witteman, Alanna M. Chamberlain, Steven A. Lubitz, Renate B. Schnabel, Sunil K. Agarwal, David D. McManus, Patrick T. Ellinor, Martin G. Larson, Gregory L. Burke, Lenore J. Launer, Albert Hofman, Daniel Levy, John S. Gottdiener, Stefan Kääb, David Couper, Tamara B. Harris, Elsayed Z. Soliman, Bruno H. C. Stricker, Vilmundur Gudnason, Susan R. Heckbert, and Emelia J. Benjamin. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *Journal of the American Heart Association*, 2(2):e000102, 2013. <https://doi.org/10.1161/JAHA.112.000102>.

- [2] Zachy I. Attia, Peter A. Noseworthy, Francisco Lopez-Jimenez, Samuel J. Asirvatham, Abhishek J. Deshmukh, Bernard J. Gersh, Rickey E. Carter, Xiaoxi Yao, Alejandro A. Rabinstein, Brad J. Erickson, Suraj Kapa, and Paul A. Friedman. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet*, 394(10201):861–867, 2019. [https://doi.org/10.1016/S0140-6736\(19\)31721-0](https://doi.org/10.1016/S0140-6736(19)31721-0).
- [3] Yong-Soo Baek, Sang-Chul Lee, Wonik Choi, and Dae-Hyeok Kim. A new deep learning algorithm of 12-lead electrocardiogram for identifying atrial fibrillation during sinus rhythm. *Scientific Reports*, 11(1):12818, 2021. <https://doi.org/10.1038/s41598-021-92172-5>.
- [4] Edoardo Coppola, Mattia Savardi, Mauro Massucci, Marianna Adamo, Marco Metra, and Alberto Signoroni. HuBERT-ECG as a self-supervised foundation model for broad and scalable cardiac applications. *medRxiv*, 2024. <https://doi.org/10.1101/2024.11.14.24317328>. Preprint. Pretrained on 9.1M 12-lead ECGs; model: Edoardo-Coppola/hubert-ecg-small (HuggingFace).
- [5] José Henrique de Moraes Goulart, Pedro Marinho Ramos de Oliveira, Rodrigo Cabral Farias, Vicente Zarzoso, and Pierre Comon. Alternating group lasso for block-term tensor decomposition and application to ECG source separation. *IEEE Transactions on Signal Processing*, 68: 2682–2696, 2020. <https://doi.org/10.1109/TSP.2020.2985591>.
- [6] Pedro Marinho R. de Oliveira and Vicente Zarzoso. Block term decomposition analysis in long segments of atrial fibrillation ECGs. In *XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2018)*, 2018. <https://doi.org/10.14209/sbrt.2018.182>.
- [7] Axel Faes, Flavio Camarrone, and Marc M. Van Hulle. Single finger trajectory prediction from intracranial brain activity using block-term tensor regression with fast and automatic component extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7): 8897–8908, 2024. <https://doi.org/10.1109/TNNLS.2022.3216589>.
- [8] Tove Hygrel, Fredrik Viberg, Erik Dahlberg, Peter H. Charlton, Katrin Kemp Gudmundsdottir, Jonathan Mant, Josef Lindman Hörnlund, and Emma Svennberg. An artificial intelligence-based model for prediction of atrial fibrillation from single-lead sinus rhythm electrocardiograms facilitating screening. *EP Europace*, 25(4):1332–1338, 2023. <https://doi.org/10.1093/europace/euad036>.
- [9] Jun Lei, Yuxi Zhou, Xue Tian, Qinghao Zhao, Qi Zhang, Shijia Geng, Qingbo Wu, and Shenda Hong. A deep learning method for beat-level risk analysis and interpretation of atrial fibrillation patients during sinus rhythm. *Biomedical Signal Processing and Control*, 100:107028, 2025. <https://doi.org/10.1016/j.bspc.2024.107028>.
- [10] Jared W. Magnani, Lei Zhu, Faye Lopez, Michael J. Pencina, Sunil K. Agarwal, Elsayed Z. Soliman, Emelia J. Benjamin, and Alvaro Alonso. P-wave indices and atrial fibrillation: cross-cohort assessments from the Framingham Heart Study (FHS) and Atherosclerosis Risk in Communities (ARIC) study. *American Heart Journal*, 169(1):53–61.e1, 2015. <https://doi.org/10.1016/j.ahj.2014.10.009>.
- [11] Pietro Melzi, Ruben Tolosana, Alberto Cecconi, Ancor Sanz-Garcia, Guillermo J. Ortega, Luis Jesus Jimenez-Borreguero, and Ruben Vera-Rodriguez. Analyzing artificial intelligence systems for the prediction of atrial fibrillation from sinus-rhythm ECGs including demographics

- and feature visualization. *Scientific Reports*, 11(1):22786, 2021. <https://doi.org/10.1038/s41598-021-02179-1>.
- [12] Aruna Mohan, Danne Elbers, Or Zilbershot, Fatemeh Afghah, and David Vorchheimer. Deciphering heartbeat signatures: a vision transformer approach to explainable atrial fibrillation detection from ECG signals. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–6, 2024. <https://doi.org/10.1109/EMBC53108.2024.10782666>.
- [13] Sajad Mousavi, Fatemeh Afghah, and U. Rajendra Acharya. HAN-ECG: An interpretable atrial fibrillation detection model using hierarchical attention networks. *Computers in Biology and Medicine*, 127:104057, 2020. <https://doi.org/10.1016/j.compbiomed.2020.104057>.
- [14] Panteleimon Pantelidis, Samuel Ruipérez-Campillo, Julia E. Vogt, Alexios Antonopoulos, Ioannis Gialamas, George E. Zakyntinos, Michael Spartalis, Polychronis Dilaveris, Jose Millet, Theodore G. Papaioannou, Evangelos Oikonomou, and Gerasimos Siasos. ECG-XPLAIM: eXplainable locally-adaptive artificial intelligence model for arrhythmia detection from large-scale electrocardiogram data. *Frontiers in Cardiovascular Medicine*, 12:1659971, 2025. <https://doi.org/10.3389/fcvm.2025.1659971>.
- [15] Lucas Plagwitz, Lucas Bickmann, Michael Fujarski, Alexander Brenner, Warnes Gobalakrishnan, Lars Eckardt, Antonius Büscher, and Julian Varghese. The Rlign algorithm for enhanced electrocardiogram analysis through heart rate-corrected ECG alignment for explainable classification and clustering. *European Heart Journal - Digital Health*, 7(5):ztag067, 2026. <https://doi.org/10.1093/ehjdh/ztag067>.
- [16] Lucas N. Ribeiro, André L. F. de Almeida, and Vicente Zarzoso. Enhanced block term decomposition for atrial activity extraction in atrial fibrillation ECG. In *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 1–5, 2016. <https://doi.org/10.1109/SAM.2016.7569709>.
- [17] Amirsajjad Taleban, Rodney Sparapani, Patrick Noffke, Sharone Zlochiver, Qiang Lu, Michael E. Widlansky, and Jake Luo. Explainable artificial intelligence in electrocardiography: a systematic review. *Biomedical Signal Processing and Control*, 114:109325, 2026. <https://doi.org/10.1016/j.bspc.2025.109325>.
- [18] Constantine Tarabanis, Vidya Koesmahargyo, Dimitrios Tachmatzidis, Vasileios Sousonis, Constantinos Bakogiannis, Robert Ronan, Scott A. Bernstein, Chirag Barbhaiya, David S. Park, Douglas S. Holmes, Alexander Kushnir, Felix Yang, Anthony Aizer, Larry A. Chinitz, Stylianos Tzeis, Vassilios Vassilikos, and Lior Jankelson. Artificial intelligence-enabled sinus electrocardiograms for the detection of paroxysmal atrial fibrillation benchmarked against the CHARGE-AF score. *European Heart Journal - Digital Health*, 6(6):1134–1144, 2025. <https://doi.org/10.1093/ehjdh/ztaf100>.
- [19] Arne Van Den Kerchove, Hakim Si-Mohammed, François Cabestaing, and Marc M. Van Hulle. BTDA: block-term tensor discriminant analysis for brain-computer interfacing. *arXiv preprint arXiv:2511.04292*, 2025. <https://doi.org/10.48550/arXiv.2511.04292>.
- [20] Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze, Wojciech Samek, and Tobias Schaeffter. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1):154, 2020. <https://doi.org/10.1038/s41597-020-0495-6>.

393 [21] Neal Yuan, Grant Duffy, Sanket S. Dhruva, Adam Oesterle, Cara N. Pellegrini, John Theurer,
394 Marzieh Vali, Paul A. Heidenreich, Salomeh Keyhani, and David Ouyang. Deep learning of
395 electrocardiograms in sinus rhythm from US veterans to predict atrial fibrillation. *JAMA*
396 *Cardiology*, 8(12):1131–1139, 2023. <https://doi.org/10.1001/jamacardio.2023.3701>.