Toward Low-Complexity Arrhythmia Classification for Ambient Assisted Living

Hangze Wu[1,2], Ruben Schlonsak[1,3], and Denys J.C. Matthies[1,3]

[1] Technical University of Applied Sciences Lübeck, Germany
[2] East China University of Science and Technology, Shanghai, China
[3] Fraunhofer IMTE, Lübeck, Germany

Abstract. Population aging increases the need for accessible cardiac rhythm screening outside clinical settings. We present a proof-of-concept, low-complexity pipeline for at-home electrocardiogram (ECG) rhythm screening tailored to Ambient Assisted Living (AAL). To mimic single-lead devices, we extract interpretable features from lead II, including ventricular rate, QRS duration, QT/QTc, RR-interval statistics, and simple counts derived from wavelet-based delineation. We compare a compact Fully Connected Neural Network (FCN) against a tree-based model (XGBoost). Models are developed on ChapmanECG under an 11-class label space and evaluated in matched splits; XGBoost yields higher mean accuracy than the FCN. For an external, system-level check, we apply the trained pipeline to short excerpts from the MIT-BIH Arrhythmia Database and harmonize labels via a graded protocol (Correct/Partially Correct/Incorrect) to account for taxonomy differences. Under this protocol, the pipeline attains 82.98% accuracy. While not a diagnostic study, the findings indicate that feature-based, interpretable models can provide practical, low-weight rhythm screening suitable for AAL contexts and merit further validation with clinically curated, single-lead home recordings.

Keywords: ECG · Arrhythmia Classification· Ambient Assisted Living · Feature Engineering · Low-Complexity Models.

1 Introduction

The global demographic shift toward population aging is accelerating and will strain healthcare delivery models that were optimized for episodic, clinic-centered care. Cardiovascular diseases (CVDs) remain the leading cause of mortality, accounting for an estimated 32% of all deaths worldwide in 2019, underscoring the need for continuous risk assessment and timely intervention [14]. In high-income countries such as Australia, the proportion of adults aged $\geq$ 65 years is projected to reach 22% by 2057, a transition that is expected to increase healthcare expenditures and disability-adjusted life years attributable to age-related conditions [1, 5]. Traditional, reactive care pathways are poorly suited to managing chronic, fluctuating CVD risks in older adults, who often present with multimorbidity, polypharmacy, and functional decline.

Ambient Assisted Living (AAL) and Machine Learning (ML) together offer a scalable, preventative alternative. AAL systems embed unobtrusive sensing into the home and community, e.g., wearables, contactless radar, and environmental sensors, to create longitudinal, context-rich physiological and behavioral time series.
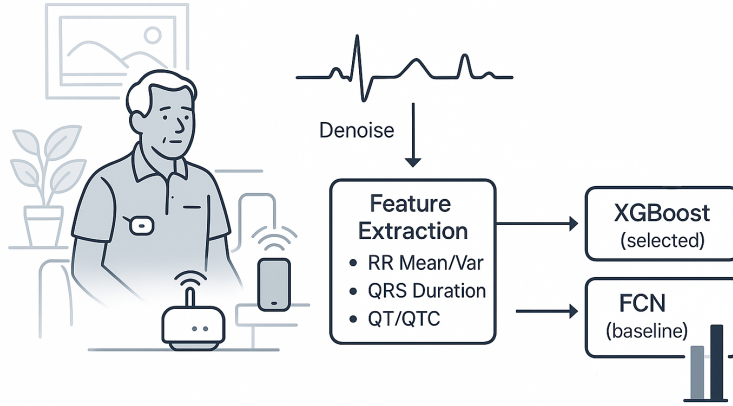
Ambient Assisted Living ECG monitoring pipeline

Fig. 1: **Overview of the AAL ECG monitoring pipeline:** single-lead ECG acquired in a home setting, denoised and processed for feature extraction (RR mean/variance, QRS duration, QT/QTc), followed by lightweight classification (XGBoost selected; Fully Connected Neural Network (FCN) as baseline).

ML methods can transform these multimodal streams into digital biomarkers and individualized risk models, enabling early detection of decompensation (e.g., evolving arrhythmia, heart-failure exacerbation), anomaly detection against personal baselines, and prediction of adverse events to trigger just-in-time interventions. Edge deployment further reduces latency and preserves privacy by processing locally, while periodic, secure aggregation supports population-level model refinement [10, 7]. By shifting from episodic to continuous, personalized monitoring and decision support, integrated AAL-ML pipelines have the potential to improve outcomes for older adults living with or at risk for CVDs, while alleviating pressure on overstretched health systems [14, 1, 5]. The main contributions of this work are:

- A compact feature set and reproducible extraction workflow for single-record ECG analysis.

- A controlled comparison of a Fully Connected Neural Network (FCN) versus an eXtreme Gradient Boosting (XGBoost) model under matched train/test splits.
- An integrated MATLAB application demonstrating the end-to-end use of the system from signal input to classification.

2   Related Work

A review of the literature reveals two primary areas of research that are pertinent to this study: the application of Ambient Assisted Living (AAL) technologies for elderly care and the use of deep learning for cardiovascular diagnosis.

Ambient Assisted Living for Cardiovascular Health aims to improve the quality of life for elderly individuals, allowing them to live independently and safely in their own homes for longer. Using advanced Ambient Intelligence technologies, AAL systems create smart environments that support daily activities. These systems integrate sensors, accelerometers, glucometers, and ECG to monitor health, thereby extending the period older adults can live autonomously. In addition, home-monitoring can help reduce the inconvenience of visiting a doctor [12]. Triantafyllidis et al. [13] proposed a framework for integrating sensors into Ambient Assisted Living (AAL) systems to enable remote monitoring of patients with chronic conditions, particularly elderly individuals. The architecture relies on wearable sensors that continuously capture physiological parameters and transmit the data via Bluetooth to a smartphone. The smartphone processes the incoming signals and forwards the information to a medical center, where it is reviewed by professional caregivers within the AAL infrastructure. Such a system facilitates reactive services that support patient self-management, while also enabling proactive interventions initiated by the system to enhance patient safety.

Deep Learning in Cardiovascular Diagnosis has emerged as a transformative field over the past decade, with electrocardiogram (ECG) analysis representing one of the most successful applications of artificial intelligence in cardiovascular medicine. Landmark studies trained large CNNs on single-lead or 12-lead recordings and reported cardiologist-level performance for rhythm detection; subsequent benchmarking on PTB-XL confirmed strong results for ResNet/Inception-style architectures [6, 11]. However, these gains often require considerable compute, memory, and energy budgets, and models can be brittle under domain shift, which are issues that matter in AAL/edge contexts. Parallel to end-to-end deep learning, feature-based pipelines remain competitive when curated clinical features (e.g., rates, intervals, morphology cues) are available. Gradient-boosted trees, notably XGBoost [3], have shown strong ECG rhythm classification on tabular feature sets and hybrid fusion settings, offering favorable complexity and interpretability compared to deep networks [8]. Several recent studies in cardiology and biomedical signal processing report XGBoost (or ensembles with

XGBoost meta-learners) as effective baselines that are easier to deploy on constrained hardware.

3   Methods

This section details the end-to-end pipeline used for ECG rhythm classification in an Ambient Assisted Living (AAL) context: signal preprocessing and wave delineation, feature computation, learning algorithms, a controlled train/test protocol, and a minimal MATLAB application for system integration and usability. The approach emphasizes a compact, clinically motivated feature set and lightweight learners suitable for resource-constrained settings.

3.1   Dataset

The ChapmanECG dataset of Zheng et al. [15] was used for model development. It comprises 12-lead ECG recordings from 10,646 patients with accompanying rhythm and cardiac-condition labels, as well as a per-patient spreadsheet of basic features. Each lead contains 5,000 samples over a 10-second interval (sampling rate 500 Hz) with an amplitude scale of 1,000 mV. For external evaluation, the MIT-BIH Arrhythmia Database [9, 4] was employed. It consists of 48 half-hour two-channel ambulatory ECG recordings from 47 subjects, collected between 1975 and 1979 at the BIH Arrhythmia Laboratory. The signals were digitized at 360 samples/s per channel with 11-bit resolution.

3.2   Signal Preprocessing and Wave Detection

We apply wavelet denoising (Biorthogonal 2.6, decomposition level 8) to reduce baseline wander and high-frequency noise. The signal is decomposed into approximation and detail coefficients; low-frequency drift is attenuated by zeroing the lowest-frequency approximation coefficients and high-frequency noise by zeroing detail coefficients, followed by reconstruction. R-peaks are then detected using a wavelet-based procedure, and QRS onsets/offsets as well as T-offsets are delineated to support interval computations used downstream.

From the denoised signal and delineated fiducial, we compute clinically motivated features: *Ventricular Rate*, *Atrial Rate*, *QRS Duration*, *QT Interval*, *QTc* (Bazett's correction), *QRS Count*, *Q Onset*, *Q Offset*, *T Offset*, and RR-based statistics (*mean_rr*, *var_rr*). Age and gender (numeric-mapped) complement the physiologic features. RR statistics derive from inter-R-peak intervals; QTc is computed as $QTc = QT/\sqrt{RR}$ (with RR in seconds). The operational extraction steps for each feature (e.g., counting R-peaks for QRS count, slope-change analysis for Q onset/offset, and T offset) follow the implementation described in the system.

3.3   Model training

We implemented a compact FCN in TensorFlow 2.1 with ReLU activations and a Softmax output layer. Kernel L2 regularization (λ=0.01) is applied to all dense layers. Training uses Adam (learning rate 10^{-3}) with EarlyStopping and ReduceLROnPlateau monitored on validation loss. We train up to 1,000 epochs with batch size 32 and report test accuracy on held-out data. For comparison, we used a gradient-boosted tree classifier with `multi:softmax` objective for multi-class prediction. Hyperparameters are: `learning_rate`= 0.01, `subsample`= 0.9, and `n_estimators`= 500. The number of classes is set to the number of unique rhythm labels in the training split. This configuration provides a strong tabular baseline with favorable complexity for edge deployment.

To ensure a fair comparison, we used a standardized preprocessing and splitting protocol for both model families. From the ChapmanECG feature table we form X from {*PatientAge*, *Gender*, *VentricularRate*, *AtrialRate*, *QRSDuration*, *QTInterval*, *QTCorrected*, *QRSCount*, *QOnset*, *QOffset*, *TOffset*, *mean_rr*, *var_rr*}, map *Gender* to {0,1}, and encode the target *Rhythm* with a label encoder. We adopt a 70/30 train/test split with a fixed random seed for each run and repeat the experiment for seeds 1..10 so that FCN and XGBoost are compared under matched splits with reduced variance.

4   Results

4.1   Performance on ChapmanECG

Across ten randomized, matched splits, **XGBoost** outperformed the compact **FCN** on accuracy. A paired comparison yielded $p = 1.45594 \times 10^{-11}$ (< 0.01), supporting XGBoost as the selected model. The seed-wise accuracies and summary statistics are shown in Table 1. A more detailed comparison is presented

Table 1: Minimum, average, and maximum accuracy comparison

	FCN_accuracy	XGBoost_accuracy
Average	85.43	90.16
Maximum	86.41	91.17
Minimum	84.35	89.32

in Figure 2, where the confusion matrices for both classifiers are shown using a random seed of 3. Both models showed strong performance for `SB`, `SR`, `ST`, `SVT`, acceptable performance for `AFIB`, but weak performance for `SA`; both exhibited several zero-prediction classes (`AT`, `AVNRT`, `AVRT`, `SAAWR`) owing to sample scarcity. XGBoost produced a non-zero prediction for `AF` and higher accuracy on `AFIB`, `SA`, `SVT`; FCN was slightly better on `SR`, `ST`.

4.2   External System-Level Evaluation on MIT-BIH

For each MIT-BIH record, the first 10,000 samples were used, matching the development window. Because the MIT-BIH taxonomy differs from the 11-class scheme used during development (e.g., N≡SR, SBR≡SB), predictions were judged using three grades (Appendix: Table 2): *Correct*, *Partially Correct*, and *Incorrect*, following the source's stated criteria and using PhysioNet annotations for verification (cf. Table 6.3 of the source for MIT rhythms[2]). Counting *Partially Correct* as acceptable, the system-level accuracy is **82.98%**; this is lower than the mean development accuracy (90.16%) due to the additional feature-extraction step and cross-dataset taxonomy differences.

On ChapmanECG, XGBoost is consistently superior to FCN across matched splits (mean accuracies 90.16% vs. 85.43%). On MIT-BIH, counting *Partially Correct* as acceptable yields 82.98% system accuracy, with residual errors concentrated in classes affected by taxonomy mismatch and signal/noise-induced feature extraction errors.

5   Discussion

Our results indicate that a compact, feature-based pipeline can achieve practical performance for 11-class rhythm recognition in an AAL setting, even under cross-dataset evaluation. Despite differences between the development (ChapmanECG) and system-level evaluation (MIT-BIH)—including channel configuration, sampling rates, annotation guidelines, patient mix, and noise profiles—the system-level accuracy of 82.98% suggests that clinically motivated features (ventricular rate, QRS duration, RR-interval statistics) provide robust discrimination without relying on large deep networks. This aligns with AAL constraints such as limited device resources, the need for low latency, and strong privacy guarantees through on-device processing.

Misclassifications concentrate among clinically adjacent rhythms that share morphology and rate characteristics in short segments (for example, sinus variants and supraventricular categories). To reflect this nuance, we report a complementary *Partially Correct* metric that treats taxonomically compatible mappings as not fully incorrect, following common merges (e.g., grouping sinus variants SR/SB/ST; aligning AF with AFIB; consolidating SVT-like supraventricular rhythms). This yields a more realistic view of utility in AAL scenarios.

Cross-dataset testing highlights distribution shifts likely to occur in real deployments. While the inductive bias of hand-crafted features mitigates some shift, targeted domain adaptation remains promising: personalized baselines (e.g., subject-specific RR statistics), light-weight calibration with a small number of labeled events, and robust signal-quality indices (SQI) to exclude degraded segments. These strategies should improve transferability without materially increasing model complexity or compute cost.

Short analysis windows reduce latency and energy consumption but may miss transitions. Practical stabilizers include overlapping windows, beat-to-episode

aggregation, and simple temporal smoothing (e.g., majority voting) or light sequential models (e.g., HMM/CRF) applied to the discrete label stream. In parallel, harmonized label spaces and a specified evaluation protocol (fixed splits and seeds, consistent windowing) are as critical as model choice for reproducibility and fair comparison.

As a non-diagnostic triage tool, the system can encourage earlier clarification of suspicious rhythms, potentially reducing unnecessary visits while prioritizing cases that warrant clinician review. Clear user guidance and effective communication of uncertainty are essential to minimize false-negative risk and ensure that ambiguous outputs prompt timely follow-up. Overall, our findings support a "low-complexity first" design principle for AAL ECG screening, with a roadmap toward prospective validation, domain adaptation, and seamless integration into home-care and telecardiology workflows.

6   Limitations & Future Work

The model is trained on a single development database (ChapmanECG). Several basic rhythms are rare therein, leading to zero predictions in those classes and limiting generalization to uncommon patterns. Future work should expand training diversity to mitigate these effects.

Evaluation is performed on MIT-BIH, whose taxonomy and annotation practice differ from the 11-class scheme used for development. The study, therefore, introduces a *Partially Correct* grade to acknowledge clinically compatible predictions when labels diverge semantically or temporally. Broader cross-dataset harmonization and label mapping would further reduce such frictions.

Evaluation uses $\approx 30\,\mathrm{s}$ excerpts (first 10,000 samples at $360\,\mathrm{Hz}$), matching development-window length. Short windows can under-represent rhythm variability and transitions. Extending or adapting windowing strategies is a direct avenue for improvement.

7   Conclusion

This proof-of-concept AAL system combines clinically motivated ECG features with a lightweight XGBoost classifier. On ChapmanECG, XGBoost surpassed a compact FCN across ten matched splits (paired t-test $p = 1.45594 \times 10^{-11}$) and was therefore integrated into the application. On MIT-BIH (first 10,000 samples per record), the whole pipeline achieved **82.98%** accuracy under an 11-class scheme, with a grading protocol to account for taxonomy and timing differences. The study also motivates a feature-based design via an explicit computational analysis of CNNs on long ECG segments. While limitations arise from class rarity, cross-dataset differences, and feature-extraction sensitivity, the results indicate that a feature-efficient, lightweight approach is viable for AAL-oriented rhythm monitoring and merits further development on broader data with strengthened signal processing.

References

1. Australian Institute of Health and Welfare: Australia's Welfare 2017. Australia's Welfare Series no. 13, AIHW, Canberra (2017), https://www.aihw.gov.au/getmedia/088848dc-906d-4a8b-aa09-79df0f943984/aihw-aus-214-aw17.pdf.aspx?inline=true
2. B.Moody, G.: Mit-bih arrhythmia database directory (1997), https://archive.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm, accessed: 16/05/2024
3. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794 (2016)
4. Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000)
5. Ha, N.T., Hendrie, D., Moorin, R.: Impact of population ageing on the costs of hospitalisations for cardiovascular disease: a population-based data linkage study. BMC Health Services Research **14**(1), 554 (2014). https://doi.org/10.1186/s12913-014-0554-9, https://doi.org/10.1186/s12913-014-0554-9
6. Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y.: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine **25**(1), 65–69 (2019). https://doi.org/10.1038/s41591-018-0268-3
7. Huang, Z., Herbozo Contreras, L.F., Leung, W.H., Yu, L., Truong, N.D., Nikpour, A., Kavehei, O.: Efficient Edge-AI models for robust ECG abnormality detection on Resource-Constrained hardware. J Cardiovasc Transl Res **17**(4), 879–892 (Mar 2024)
8. Jahangir, R., Islam, M.N., Islam, M.S., Islam, M.M.: Ecg-based heart arrhythmia classification using feature engineering and a hybrid stacked machine learning. BMC Cardiovascular Disorders **25**(1), 260 (Apr 2025). https://doi.org/10.1186/s12872-025-04678-9, https://doi.org/10.1186/s12872-025-04678-9
9. Moody, G., Mark, R.: The impact of the mit-bih arrhythmia database. IEEE Engineering in Medicine and Biology **20**(3), 45–50 (May-Jun 2001). https://doi.org/10.1109/51.932724
10. Rancea, A., Anghel, I., Cioara, T.: Edge computing in healthcare: Innovations, opportunities, and challenges. Future Internet **16**(9) (2024). https://doi.org/10.3390/fi16090329, https://www.mdpi.com/1999-5903/16/9/329
11. Strodthoff, N., Wagner, P., Schaeffter, T., Samek, W.: Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. IEEE J Biomed Health Inform **25**(5), 1519–1528 (May 2021)
12. Sun, H., De Florio, V., Gui, N., Blondia, C.: Promises and challenges of ambient assisted living systems. In: 2009 Sixth International Conference on Information Technology: New Generations. pp. 1201–1207. Ieee (2009)
13. Triantafyllidis, A.K., Koutkias, V.G., Chouvarda, I., Adami, I., Kouroubali, A., Maglaveras, N.: Framework of sensor-based monitoring for pervasive patient care. Healthcare technology letters **3**(3), 153–158 (2016)
14. WHO: Cardiovascular diseases (cvds) (2021), https://www.who.int/en/newsroom/fact-sheets/detail/cardiovascular-diseases-(cvds), accessed: 16/05/2024

15. Zheng, J.: Chapmanecg (Jun 2019). https://doi.org/10.6084/m9.figshare.c.4560497.v1, https://figshare.com/collections/ChapmanECG/4560497/1

A   Appendix

Table 2: Label harmonization between ChapmanECG (11 classes) and MIT-BIH record-level rhythms used for grading. 'PC' denotes Partially Correct.

Chapman class	MIT-BIH rhythm(s)	Grade rule
SR (Sinus Rhythm)	N (Normal sinus rhythm)	Correct
SB (Sinus Bradycardia)	SBR (Sinus bradycardia)	Correct
ST (Sinus Tachycardia)	(no direct MIT record label)	PC vs. N/SVT-like[*]
AFIB	AFIB	Correct
AFL	AFL	Correct
SA (Sinus Arrhythmia)	N (with high RR variability)	PC
SVT	SVTA (Supraventricular tachyarr.)	Correct/PC
AT	SVTA	PC
AVNRT	SVTA	PC
AVRT	SVTA	PC
SAAWR	N or NOD (junctional), case-dependent	PC

[*] Short 10–30 s windows with elevated rate but lacking explicit MIT rhythm change are graded PC when consistent with sinus tachy features.
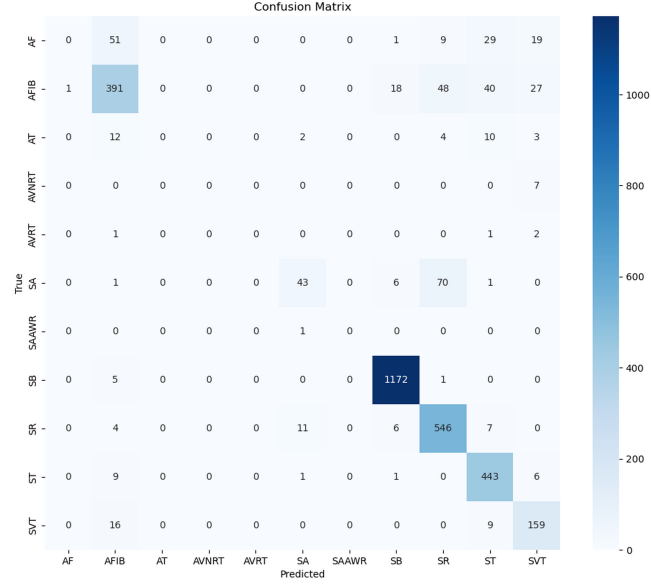
(a) Confusion matrix of the FCN model.
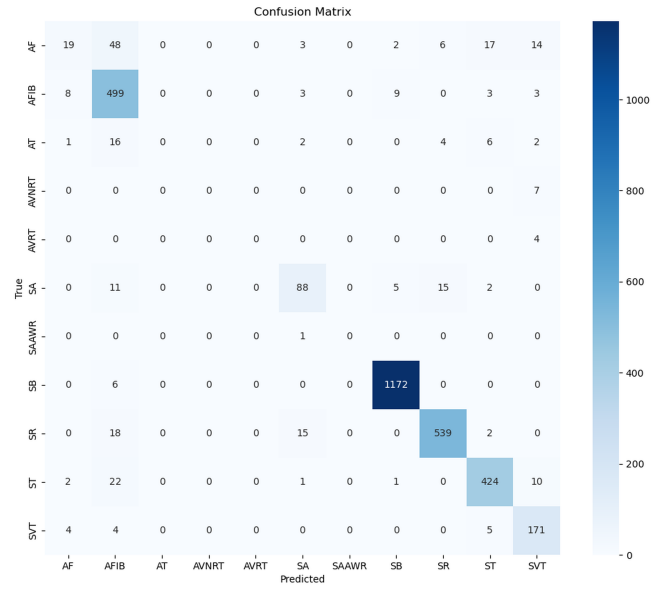
(b) Confusion matrix of the XGBoost model.

Fig. 2: Comparison of classification results for the two approaches.