POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Heartbeat by Heartbeat - Lightweight Variational Learning for Resource-Constrained Ischemia Detection

Laurea Magistrale in Computer Science and Engineering - Ingegneria Informatica

Author: Matteo Corti

Advisor: Prof. Giuseppe Pozzi

Academic year: 2024-2025

1. Introduction

Myocardial ischemia, a condition arising from inadequate perfusion of the cardiac muscle, remains a leading cause of morbidity and mortality worldwide. Because ischemic episodes are frequently transient and unpredictable, continuous electrocardiographic (ECG) monitoring is essential for timely clinical intervention [5]. Contemporary cloud-based diagnostic systems have demonstrated considerable efficacy in automated detection; however, their reliance on persistent network connectivity precludes deployment in resource-constrained settings—remote regions, disaster-affected areas, and low-resource healthcare environments where reliable internet access remains unavailable to a substantial fraction of the global population [2].

This thesis addresses the resulting gap through the development and systematic evaluation of a computational framework for automated ischemia detection from single-lead ECG recordings, expressly engineered for execution on mobile and embedded platforms subject to severe computational and memory constraints. The central research question is whether machine learning-based cardiac ischemia detection can be deployed on such devices while preserving clinically acceptable diagnostic performance in fully offline operation.

Three principal contributions are presented. First, a *beat-centric processing paradigm* in which individual cardiac cycles—rather than fixed-duration temporal windows—constitute the fundamental unit of analysis, yielding a 16-fold reduction in per-beat memory requirements relative to standard 10-second window approaches. Second, a *Variational Autoencoder (VAE)* with a lightweight convolutional encoder totalling 56,096 trainable parameters, which learns compact 16-dimensional latent representations that preserve morphological features of diagnostic relevance—notably ST-segment deviation and T-wave abnormalities. Third, a *hierarchical temporal aggregation pipeline* that transforms beat-level probabilistic predictions into clinically interpretable episode-level detections through hysteresis-based filtering aligned with clinical annotation granularity.

Experimental evaluation on the European ST-T Database (EDB) [5] yields an episode-level sensitivity of 80% with 100% precision, while post-training INT8 quantisation compresses the model to 309 KB without measurable accuracy degradation.

2. Proposed Methodology

2.1. Beat-Centric Processing

The proposed system departs from conventional fixed-duration window approaches (typically 10 seconds) and operates at the granularity of individual cardiac cycles. This design choice rests on two complementary arguments.

From a physiological standpoint, ischemia manifests at the single-beat level through characteristic alterations in ST-segment morphology, T-wave amplitude and polarity, and QT interval duration; these markers are intrinsic to individual cardiac cycles rather than emergent properties requiring multi-beat context for identification. From a computational standpoint, processing isolated beats of 320 samples (after resampling) reduces per-beat encoder inference memory to approximately 5.6 KB, compared with roughly 90 KB for standard 10-second windows at 250 Hz, enabling deployment on platforms with as little as 512 KB of available RAM.

Beat segmentation employs a modified Pan-Tompkins R-peak detector [4] with adaptive thresholding, validated against EDB annotations with sensitivity, positive predictive value, and F1 score each exceeding 99%. Beats are extracted via symmetric 1,28 s windows centred on R-peaks, resampled to a canonical length of 320 samples, and normalised per-beat to zero mean and unit variance. A wavelet-based fiducial point delineation stage identifies P-wave, QRS complex, and T-wave boundaries; a quality assessment framework then assigns each beat a quality level (EXCELLENT, GOOD, or ACCEPTABLE) that determines eligibility for downstream inference.

2.2. VAE Architecture

The core of the classification system is a VAE trained to learn compact, discriminative latent representations of individual cardiac beats. The encoder $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ maps a 320-sample beat waveform $\boldsymbol{x}$ to the parameters of a diagonal Gaussian in a d-dimensional latent space: mean $\boldsymbol{\mu}$ and log-variance $\log \boldsymbol{\sigma}^2$. Sampling employs the reparameterisation trick [3]:

$$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \tag{1}$$

The CNN encoder comprises two convolutional blocks with aggressive stride-4 configurations, calibrated to the temporal scales of ECG waveform components. The first layer (kernel size 15, 32 output channels) targets rapid deflections characteristic of the QRS complex, while the second (kernel size 7, 64 output channels) captures the slower morphological variations of the ST segment and T wave. Batch normalisation and GELU activations follow each convolution. Two parallel fully connected layers project the 1280-dimensional flattened feature vector to $\boldsymbol{\mu}$ and $\log \boldsymbol{\sigma}^2$. The complete encoder totals 56,096 parameters and approximately 168,000 multiply-accumulate operations (MACs) per beat (Table 1).

Table 1: CNN encoder architecture. B: batch size.

Layer	Params	MACs	Output
Conv1D (1→32, k=15, s=4)	512	40,960	(B,32,80)
Conv1D (32→64, k=7, s=4)	14,400	86,400	(B,64,20)
FC_{μ} (1280→16)	20,496	20,480	(B,16)
$\text{FC}_{\log \sigma^2}$ (1280→16)	20,496	20,480	(B,16)
Total	**56,096**	**∼168K**	

A fully connected (FC) encoder with 126,304 parameters serves as a baseline architecture and enables subsequent ensemble combination. Classification operates on the deterministic latent mean $\boldsymbol{\mu}$ (not on samples $\boldsymbol{z}$) via a three-layer head (16→64→32→2) with batch normalisation, LeakyReLU activations, and dropout ($p = 0.4$), adding 3,426 parameters. The decoder, employed only during training to enforce reconstruction-based regularisation, is discarded at inference time; the deployed model (encoder + classifier) thus comprises 59,522 parameters.

The latent dimensionality $d = 16$ was selected through systematic ablation (Section 3.3): lower dimensionalities yielded degraded reconstruction fidelity for subtle ST-segment features, whereas 32 dimensions achieved the highest classification performance but at the cost of a larger parameter count. The 16-dimensional space provides a 20× compression ratio (320→16) while preserving the morphological information most relevant to ischemia discrimination.

2.3. Multi-Objective Training

The training objective combines six loss components addressing complementary goals: reconstruction fidelity ($\mathcal{L}_{\text{recon}}$, mean squared error),

KL divergence regularisation ($\mathcal{L}_{\text{KL}}$) with linear β-warmup to prevent posterior collapse, focal classification loss ($\mathcal{L}_{\text{focal}}$, $\gamma = 2.0$, class weights $\alpha_0 = 0.25$, $\alpha_1 = 0.75$), a waveform smoothness regulariser ($\mathcal{L}_{\text{smooth}}$), and unsupervised and supervised contrastive losses ($\mathcal{L}_{\text{con}}$, $\mathcal{L}_{\text{sup\_con}}$). Dynamic balancing between the reconstruction and classification terms is realised through the GradNorm algorithm [1], which adjusts the respective loss weights at each iteration so as to equalise the gradient norms flowing to the shared encoder. A free-bits mechanism ($\lambda_{\text{FB}} = 0.25$ nats per dimension) ensures that each latent dimension encodes a minimum amount of information, mitigating the risk of unused latent capacity.
Optimisation employs AdamW (lr $= 10^{-3}$, weight decay 10^{-4}) with ReduceLROnPlateau scheduling. Online data augmentation—polarity flip, Gaussian noise injection, amplitude scaling, and time warping ($\sigma_{\text{warp}} = 0.028$)—is applied stochastically during training to improve robustness to recording variations. Early stopping with a patience of 15 epochs on validation loss terminates training; convergence is typically reached within 40–60 epochs on an NVIDIA T4 GPU.

2.4. Hierarchical Temporal Aggregation

Individual beat-level classifications are inherently noisy: isolated false positives or false negatives are unavoidable at this granularity. A multi-stage aggregation pipeline addresses this limitation by exploiting the pathophysiological property that ischemia persists across many consecutive beats. In the first stage, a second-stage neural network integrates the CNN and FC encoder outputs, the Isoelectric Energy Elevation Factor (IEEF)—a morphological measure of ST-segment deviation from the isoelectric reference—and reconstruction quality metrics over sliding windows of 115 beats, producing refined per-beat ischemia probabilities. In the second stage, hysteresis-based filtering, implemented as a four-state machine (NORMAL, ONSET, ISCHEMIC, OFFSET) with upper threshold $\tau_{\text{high}} = 0.8$ and lower threshold $\tau_{\text{low}} = 0.6$, suppresses rapid oscillation at decision boundaries. Predictions are then consolidated into non-overlapping 30-second intervals matching the EDB clinical annotation granularity. The streaming implementation uses circular buffers requiring a small amount of RAM, enabling indefinite continuous monitoring without state accumulation on memory-constrained platforms.

3. Experimental Evaluation

3.1. Dataset and Experimental Protocol

Experiments were conducted on the European ST-T Database [5], a benchmark corpus comprising 90 two-hour ambulatory ECG recordings from 79 subjects. From a study population of 32 patients, beat extraction yielded 317,630 beats (80.7% normal, 19.3% ischemic; class imbalance ratio 4.18:1). To prevent information leakage of patient-specific features, data were partitioned at the patient level: 17 patients for training, 13 for validation, and 2 for testing (135,335 test beats). Leave-One-Patient-Out (LOPO) cross-validation over 18 patients was additionally performed to provide supplementary evidence of cross-patient generalisability. Precision-Recall AUC (PR-AUC) was adopted as the primary model selection metric, given the substantial class imbalance, since it is directly sensitive to false positive rates—a critical concern in clinical monitoring.

3.2. Classification and Detection Results

At the beat level, the CNN encoder achieved an AUC-ROC of 0.911 [95% CI: 0.904, 0.918] and a balanced accuracy of 88.1% on the held-out test set; the FC encoder reached a balanced accuracy of 86.4% with an AUC-ROC of 0.875. The second-stage temporal integration model, operating over 115-beat windows, improved balanced accuracy to 92.13% (ROC-AUC = 0.962, PR-AUC = 0.927).
Table 2 traces the evolution of precision and recall across the pipeline stages, illustrating the progressive transformation from a high-recall, moderate-precision regime at the beat level to a precision-oriented regime after temporal aggregation and episode consolidation.
At the sample level (evaluated at the native 250 Hz resolution), the final pipeline output achieved 97.04% accuracy, 92.71% balanced accuracy, 98.94% specificity, 86.48% sensitivity, and an F1 score of 0.899. At the episode level, the system detected 4 of 5 ground-truth is-

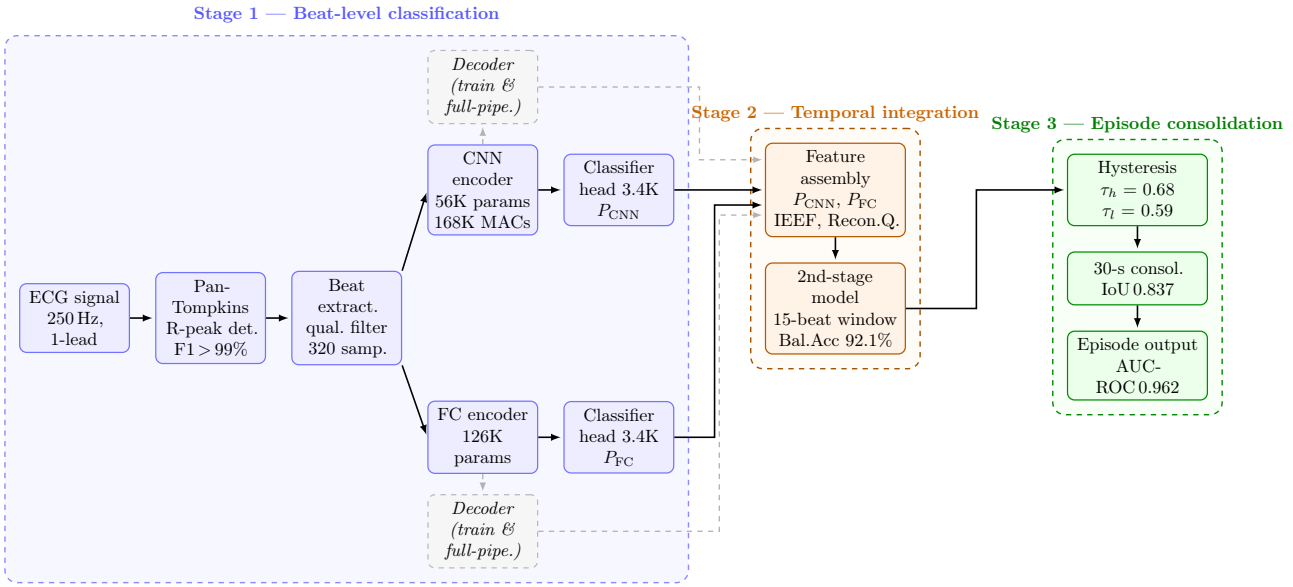

Figure 1: Three-stage ischemia detection pipeline. **Stage 1** (blue): dual-encoder beat classification with symmetric CNN and FC decoders. **Stage 2** (orange): second-stage temporal integration network. **Stage 3** (green): hysteresis thresholding and episode consolidation. Streaming state buffer estimated at <500 bytes.

Table 2: Classification metrics at each pipeline stage on the test set.

Stage	Bal. Acc.	Prec.	Recall
CNN encoder	87.80%	57.60%	87.14%
FC encoder	90.07%	63.37%	89.43%
Ensemble (baseline)	89.43%	60.93%	89.15%
2nd-stage model	92.13%	67.69%	92.17%
Hysteresis filter	87.81%	94.19%	76.48%
Final (30 s intervals)	86.79%	94.56%	74.34%

chemic episodes (sensitivity 80%, precision 100%, episode F1 = 0.889). The single missed episode exhibited attenuated ST-segment deviation that did not produce sustained probability elevation above the detection threshold. Temporal localisation of the four correctly detected episodes yielded a mean Intersection over Union (IoU) of 0.837, with a mean onset delay of +40.5 beats and a systematic duration overestimation of 15.4%—a conservative bias that ensures complete capture of ischemic periods.

LOPO cross-validation yielded a median balanced accuracy of 0.810 and median AUC-ROC of 0.919 across 18 patient folds. Inter-patient variability was substantial, consistent with the known heterogeneity of ischemic morphological presentations across individuals.

3.3. Ablation Studies and Robustness

Loss component ablation yielded two principal findings. First, replacing focal loss with standard cross-entropy under the balanced sampling strategy improved balanced accuracy by +7.2% and sensitivity by +13.5%, indicating that the down-weighting of well-classified samples by focal loss is counterproductive once the sampler has already addressed class imbalance. Second, the KL divergence term exhibited limited measurable impact on classification metrics, an effect attributable to the free-bits mechanism that attenuates the effective regularisation strength; nonetheless, the KL term contributes to latent space structure. Data augmentation improved PR-AUC by approximately 10.9% (from 0.617 to 0.684), with time warping yielding the highest individual contribution. Latent dimensionality ablation confirmed that $d = 16$ provides a favourable trade-off between compression and predictive performance, with 32 dimensions achieving the highest balanced accuracy (0.862) but at the cost of increased model complexity.

Noise robustness evaluation showed that at 15 dB SNR—representative of typical ambulatory recording conditions—balanced accuracy degraded by only 1.1% relative to the clean baseline (0.871 vs. 0.881). At 10 dB, the degradation reached 3.3%, remaining within acceptable

bounds for most monitoring scenarios.

3.4. Computational Performance and Deployment

Table 3 reports inference latency across a representative spectrum of hardware platforms, from mobile processors to microcontroller-class devices. All platforms exceed the real-time requirement (approximately 1 beat per second at rest) by at least two orders of magnitude.

Table 3: Inference latency (ms/beat) for the CNN encoder.

Platform	FP32	INT8	Beats/s
Snapdragon 888	0.8	0.5	2,090
Snapdragon 680	1.9	1.0	1,045
Exynos W920	5.7	2.9	348
Cortex-M4 (168 MHz)	32	12.5	80
Cortex-M7 (480 MHz)	8	3.6	279

Post-training INT8 dynamic quantisation reduces the model from 1.0 MB (FP32) to 309 KB with no measurable classification degradation (balanced accuracy: +0.1%, AUC-ROC and PR-AUC unchanged). Memory profiling of the INT8-quantised encoder reveals a static footprint (weights stored in flash) of approximately 15 KB and a peak dynamic allocation (activations in RAM) of 4 KB, with a theoretical minimum of 3.7 KB under in-place buffer reuse. These figures confirm feasibility on microcontrollers with as little as 20 KB of RAM. BatchNorm fusion into preceding convolutional layers further eliminates runtime overhead. Model export is supported via both TorchScript and ONNX, and a Java prototype demonstrates end-to-end offline operation on Android devices.

4. Conclusions

This thesis presented a complete pipeline for ischemia detection from single-lead ECG recordings on resource-constrained embedded platforms, spanning preprocessing, beat segmentation, VAE-based feature learning, classification, and hierarchical temporal aggregation. The beat-centric paradigm addresses the tension between classification accuracy and computational efficiency: the deployed model comprises 59,522 parameters with a peak inference memory of 4 KB (INT8), while achieving episode-level detection with 80% sensitivity and 100% precision on the EDB test set. The zero false-positive rate at the episode level is of particular practical relevance, as false alarms constitute a primary barrier to clinical adoption of automated monitoring systems.

The VAE-based representation learning approach provides an effective inductive bias for morphological feature extraction under the data scarcity typical of clinical ECG applications: the 16-dimensional latent space compresses the input by a factor of 20 while preserving diagnostically relevant ST-segment and T-wave features. The hierarchical aggregation pipeline, with its streaming-compatible design (below 500 bytes of memory), converts noisy beat-level predictions into temporally localised episode detections (mean IoU = 0.837).

Several limitations warrant acknowledgement. The evaluation is restricted to the European ST-T Database and to single-lead recordings; external validation on independent, multi-centre datasets is a necessary prerequisite for clinical translation. The primary test set comprises only 2 patients, a constraint dictated by patient-level partitioning on a 32-patient study population; however, LOPO cross-validation over 18 patients provides complementary evidence of generalisability (median AUC-ROC: 0.919). Computational benchmarks on microcontroller-class hardware are based on simulation using documented processor specifications rather than direct on-device measurement, and actual streaming deployment was not demonstrated within the scope of this work.

Promising directions for future investigation include multi-lead fusion, temporal modelling of beat-to-beat ischemia evolution, uncertainty-aware prediction for risk-stratified clinical alerting, and prospective validation in clinical settings. The findings of this work provide evidence that lightweight neural-network architectures, when co-designed with their deployment targets, can bring automated cardiac monitoring to environments where computational, memory, and connectivity constraints have traditionally precluded access to such capabilities.

References

[1] Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss bal-

ancing in deep multitask networks, 2018.

[2] Gari D. Clifford, Francisco Azuaje, and Patrick McSharry. *Advanced Methods And Tools for ECG Data Analysis*. Artech House, Inc., USA, 2006.

[3] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022.

[4] Jiapu Pan and Willis J. Tompkins. A real-time qrs detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, 1985.

[5] A Taddei, G Distante, M Emdin, P Pisani, G B Moody, C Zeelenberg, and C Marchesi. The european st-t database: standard for evaluating systems for the analysis of st-t changes in ambulatory electrocardiography. *European Heart Journal*, 13(9):1164–1172, September 1992.