POLITECNICO DI TORINO

MASTER's Degree in COMPUTER ENGINEERING

MASTER's Degree Thesis

Interpretability Guided ECG Denoising with Tsetlin Machines

Supervisors	Candidate
Prof. Alessandro ALIBERTI	Giulio MASELLI (306125)

DECEMBER 2025

Interpretability Guided ECG Denoising with Tsetlin Machines

Giulio Maselli

Abstract

Deep Learning models have recently established new benchmarks in electrocardiogram (ECG) denoising, yet their "black-box" nature and susceptibility to generative hallucinations pose significant barriers to clinical adoption. This thesis addresses the trade-off between performance and transparency by proposing a novel "Safety-by-Design" framework based on the **Regression Tsetlin Machine (RTM)**.

Unlike conventional end-to-end neural networks that reconstruct waveforms directly, the proposed system decouples noise quantification from signal restoration. A multi-head RTM architecture estimates the scalar intensity of Baseline Wander, Muscle Artifacts, and Power Line Interference using transparent propositional logic. These estimates subsequently drive a bank of adaptive digital filters through a non-linear gating mechanism, ensuring that physiological signal morphology is preserved when noise is absent.

A key contribution of this work is the utilization of intrinsic interpretability as an engineering driver. By analyzing the global feature importance of the logic clauses, we identified a dominance of spectral features over time-domain attributes, enabling a **95% reduction in the feature space** while improving predictive correlation by 4.2%. Experimental validation on the MIT-BIH Arrhythmia Database demonstrates that the system achieves a Signal-to-Noise Ratio (SNR) improvement of **+5.20 dB**, outperforming classical filters by 134% and reaching competitive performance relative to state-of-the-art Convolutional Neural Networks. Rigorous statistical analysis confirms cross-subject robustness (Coefficient of Variation $< 6\%$) and negligible bias. With an inference latency of just **0.070 ms** on a standard CPU and a minimal memory footprint, the proposed solution is uniquely suited for trusted, real-time deployment on wearable medical devices.

Table of Contents

1 Introduction **1**

1.1 Clinical context and importance of the ECG 1
1.2 The problem of noise in ECG signals 1
1.3 Denoising approaches: classical filters, Deep Learning, and black-box models 2
1.4 Motivation for interpretable and white-box models 3
1.5 Thesis objectives 3
1.6 Main contributions 4
1.7 Thesis outline 5

2 Background **6**

2.1 Signal Processing Foundations 6
2.1.1 Signal Classification and Properties 6
2.1.1.1 Deterministic versus Stochastic Signals 6
2.1.1.2 The Concept of Stationarity 7
2.1.2 Digital Representation: Sampling and Quantization 7
2.1.2.1 The Sampling Theorem 7
2.1.2.2 Quantization and Dynamic Range 8
2.1.3 Frequency and Spectral Analysis 8
2.1.3.1 The Fourier Transform 8
2.1.3.2 Power Spectral Density and Welch's Method 8
2.1.4 Noise Theory and Metrics 9
2.1.4.1 Spectral Characteristics of Noise 9
2.1.4.2 Quantitative and Qualitative Metrics 9
2.2 The Electrocardiogram (ECG) 9
2.2.1 Physiology and Morphology 10
2.2.1.1 The Cardiac Cycle and Electrical Conduction 10
2.2.1.2 Waveform Components 10
2.2.2 Artifacts and Noise Sources in Ambulatory ECG 11
2.2.2.1 Baseline Wander 11
2.2.2.2 Electromyographic Interference 12
2.2.2.3 Power Line Interference 12
2.2.3 Classical Denoising Techniques 12
2.2.3.1 Linear Digital Filtering 12

2.2.3.2 Adaptive Filtering 13
2.2.3.3 Transform Domain Decomposition 13
2.3 Machine Learning in Signal Denoising 13
2.3.1 Machine Learning Paradigms 13
2.3.1.1 Supervised Learning 13
2.3.1.2 Unsupervised Learning 14
2.3.1.3 Reinforcement Learning 14
2.3.2 Core Computational Concepts: Regression and Convolution . 14
2.3.2.1 Regression Analysis 14
2.3.2.2 The Convolution Operation 15
2.3.3 Deep Learning Architectures for ECG 15
2.3.3.1 Convolutional Neural Networks (CNNs) 15
2.3.3.2 Recurrent Neural Networks (RNNs) and LSTMs . . 15
2.3.3.3 Denoising Autoencoders 16
2.3.4 Denoising Paradigms: Reconstruction versus Estimation . . . 16
2.3.4.1 Direct Denoising 16
2.3.4.2 Noise Estimation 16
2.3.5 The Interpretability Challenge 17
2.3.5.1 Opacity and Lack of Trust 17
2.3.5.2 Hallucinations and Generative Artifacts 17
2.4 The Tsetlin Machine Framework 17
2.4.1 Historical Origins: Cybernetics and Automata Theory 18
2.4.1.1 The Linear Tactic Automaton 18
2.4.2 The Tsetlin Machine Architecture 18
2.4.2.1 Propositional Input and Literals 19
2.4.2.2 Conjunctive Clauses 19
2.4.2.3 The Team of Automata 20
2.4.3 Learning Dynamics: The Game of Feedback 20
2.4.3.1 Resource Allocation and Thresholding 20
2.4.3.2 Type I Feedback: Pattern Recognition and Refinement 20
2.4.3.3 Type II Feedback: Discrimination and Suppression . 21
2.4.4 State-of-the-Art in Biomedical Tsetlin Machines 21
2.4.5 The Regression Tsetlin Machine (RTM) 21
2.4.5.1 From Polarity to Magnitude 21
2.4.5.2 Error-Driven Feedback 22
2.4.6 A Practical Example: Learning the XOR Function 22
2.4.7 Advantages for Biomedical Signal Processing 23
2.4.7.0.1 Intrinsic Interpretability. 23
2.4.7.0.2 Computational Efficiency and Edge AI. . . 23
2.4.7.0.3 Robustness to Noise and Overfitting. . . . 23

3 Related Works **24**
3.1 Classical Signal Processing for ECG Denoising 24
3.1.1 Digital Filtering Techniques 24
3.1.2 Adaptive Filtering 25
3.1.3 Transform Domain Decomposition 25
3.2 Deep Learning Approaches: The Current State-of-the-Art 26
3.2.1 Direct Denoising: End-to-End Reconstruction 26
3.2.2 The Shift to Noise Estimation 26
3.2.3 Generative Models 27
3.2.4 The “Black-Box” Problem in Clinical AI 27
3.3 The Tsetlin Machine in Biomedical Applications 27
3.3.1 Interpretable Classification of Arrhythmias 28
3.3.2 Continuous Estimation via Regression Tsetlin Machines . . . 28
3.3.3 Efficiency and Edge Computing Potential 28
3.3.4 Intrinsic vs. Post-Hoc Interpretability 29
3.4 Gap Analysis and Thesis Positioning 29
3.4.1 Synthesis of Current Limitations 29
3.4.2 Thesis Contribution 29

4 Methodology **31**
4.1 System Architecture 31
4.1.1 Pipeline Overview 31
4.1.2 The Noise Intensity Paradigm 32
4.2 Data Preparation and Noise Mixing 34
4.2.1 Source Databases 34
4.2.2 Synthetic Mixing Protocol 34
4.2.3 Segmentation and Data Partitioning 35
4.2.4 Noise Distribution Analysis 35
4.3 Explainability-Driven Feature Engineering 36
4.3.1 Feature Domains 36
4.3.1.1 Spectral Features (Frequency Domain) 36
4.3.1.2 Bitplane Features (Amplitude Domain) 36
4.3.2 Optimization via Global Feature Importance 37
4.3.3 The "Zero-Variance" Phenomenon 37
4.3.4 Binarization Strategy: Thermometer Encoding 37
4.4 Regression Tsetlin Machine Architecture 38
4.4.1 Multi-Head Configuration 38
4.4.1.1 The EMG and BW Regressors 38
4.4.1.2 The PLI Regressor and Bimodality 38
4.4.2 Hyperparameter Specification 38
4.4.3 Training Dynamics 39
4.5 Post-Processing and Adaptive Control 39
4.5.1 Isotonic Calibration 39

4.5.2 The Gating Mechanism 40
4.5.3 Adaptive Filtering Modules 40
4.5.3.1 Baseline Wander Removal 41
4.5.3.2 Muscle Artifact Denoising 41
4.5.3.3 Power Line Interference Suppression 41
4.6 Statistical Validation Framework 41
4.6.1 Cross-Subject Robustness Analysis 41
4.6.2 Bootstrap Confidence Intervals 42
4.6.3 Agreement Analysis: Bland-Altman Method 42
4.6.4 Evaluation Metrics 42
4.6.4.0.1 Regression Metrics: 42
4.6.4.0.2 Denoising Metrics: 43

5 Experimental Results 44
5.1 Noise Intensity Estimation Performance 44
5.1.1 Regressor Accuracy on Test Set 45
5.1.2 Impact of Isotonic Calibration 45
5.2 Statistical Validation and Robustness 46
5.2.1 Cross-Subject Consistency 47
5.2.2 Reliability Analysis via Bootstrap 47
5.2.3 Agreement Analysis: Bland-Altman 48
5.3 Downstream Denoising Performance 48
5.3.1 Quantitative SNR Improvement 49
5.3.2 Qualitative Waveform Analysis 50
5.4 Interpretability and Knowledge Discovery 50
5.4.1 Global Feature Importance 51
5.4.2 Clause Analysis and Logic Rules 52
5.4.3 Feature Concentration and Sparsity 52
5.5 Ablation Studies 53
5.5.1 Feature Engineering Evolution 53
5.5.2 The "Zero-Variance" Feature Cleaning Paradox 53
5.5.3 Encoding Resolution Sensitivity 54
5.6 Computational Efficiency and Benchmarking 54
5.6.1 Runtime and Memory Footprint 54
5.6.2 Comparison with Neural Architectures 55

6 Discussion 57
6.1 The Accuracy-Transparency Trade-off 57
6.1.1 Benchmarking against Deep Learning 57
6.1.2 The Value of Trust in Clinical Settings 58
6.2 Interpretability as an Engineering Driver 58
6.2.1 The Feedback Loop: From Insight to Optimization 58
6.2.2 Overcoming the Curse of Dimensionality 58

6.3 Clinical Safety and the "Do No Harm" Principle 59
6.4 Generalization and Patient Variability 59
6.5 Operational Feasibility and Green AI 60
6.5.1 The Bitwise Advantage 60
6.5.2 Enabling Edge AI in Healthcare 60
6.6 Methodological Limitations 61
6.6.1 The Bimodality of Power Line Interference 61
6.6.2 Saturation in High-Noise Regimes 61
6.6.3 Feature Dependency and the "Zero-Variance" Paradox 61

7 Conclusion and Future Work **63**
7.1 Synthesis of Achievements 63
7.2 The Shift Towards Transparent Signal Processing 63
7.3 Future Research Directions 64
7.3.1 Algorithmic Extensions 64
7.3.2 Hardware Deployment and TinyML 64
7.3.3 Clinical Validation and Multi-Lead Integration 64

Bibliography **66**

List of Figures

2.1 Typical ECG waveform of a normal heartbeat cycle. The annotations highlight the P-wave (atrial depolarization), the QRS complex (ventricular depolarization), and the T-wave (ventricular repolarization), which are critical morphological features for diagnosis. 11
2.2 Schematic representation of the standard Tsetlin Machine architecture. Input features are converted into literals, which are processed by teams of Tsetlin Automata to form conjunctive clauses. The final output is determined by aggregating the votes from active clauses. 19

4.1 Block diagram of the proposed TMU-Optimized pipeline. The system decouples the estimation path (Intelligence Unit, top) from the filtering path (Signal Processing Unit, bottom). The Tsetlin Machine estimates noise intensities which modulate the adaptive filters via the control signal α. 33

5.1 Scatter plots of Predicted vs. Ground Truth noise intensity for EMG, BW, and PLI. The red dashed line represents perfect prediction ($y = \hat{y}$), while the blue line indicates the linear fit. Note the bimodal clustering in the PLI plot. 45
5.2 Reliability diagrams for EMG predictions before (left) and after (right) isotonic calibration. The calibrated model aligns with the ideal diagonal, confirming that the predicted intensities accurately reflect the observed noise levels. 46
5.3 Boxplots of Pearson correlation coefficients calculated per subject. The narrow spread indicates that the model generalizes well across different patients. 47
5.4 Bland-Altman plots showing the difference between predicted and ground truth intensities against the mean. The solid line represents the mean bias, while dashed lines indicate the 95% Limits of Agreement. 49
5.5 Signal-to-Noise Ratio (SNR) Improvement comparison. The proposed TMU-based method significantly outperforms classical techniques and bridges the gap towards Deep Learning performance. 50

5.6 Qualitative assessment of denoising performance. (a) Clean Ground Truth. (b) Noisy Input corrupted by EMG and BW. (c) Output of the TMU-Guided Adaptive Filter. (d) Detailed zoom on the QRS complex, demonstrating the preservation of peak amplitude and morphology. . 51
5.7 Global feature importance for EMG and BW estimation. Red bars indicate Spectral features, while Green bars indicate Bitplane features. The model autonomously identified that spectral energy is the primary discriminator for muscle noise. 52

List of Tables

3.1 Comparison of ECG Denoising Paradigms vs. Proposed Method. . . 30

4.1 Comparison of Denoising Paradigms: Waveform Estimation vs. Intensity Estimation. 33
4.2 Distribution of Noise Intensities in the Test Set ($N = 53,256$). . . . 35
4.3 Hyperparameter configuration for the TMU-Optimized architecture. 39

5.1 TMU Regressor Performance on Test Set (After Isotonic Calibration). 45
5.2 Impact of Isotonic Calibration on Model Performance. 46
5.3 Per-Subject Performance Analysis (N=48 patients). 47
5.4 Bootstrap 95% Confidence Intervals (1000 resamples). 48
5.5 Bland-Altman Agreement Statistics. 48
5.6 Comparison of Denoising Performance (SNR Improvement). 49
5.7 Ablation Study: Evolution of Feature Engineering (EMG Regressor). 53
5.8 Runtime and Memory Benchmarks for the TMU-Optimized System. 55
5.9 Comparison: Proposed TMU Framework vs. Standard CNN Architectures. 55

Chapter 1

Introduction

1.1 Clinical context and importance of the ECG

Cardiovascular diseases are among the leading causes of death worldwide, with ischemic heart disease increasing substantially global mortality. In this context, the electrocardiogram (ECG) is one of the most widely used diagnostic tools to check the electrical activity of the heart in a non-invasive, low-cost, and repeatable way.

An ECG is obtained by placing electrodes on the skin and recording, over time, the potential differences generated by the depolarization and repolarization of cardiac tissue. The resulting waveform provides information on cardiac rhythm, conduction, and morphology, and is essential for detecting arrhythmias, ischemic events, and other pathological conditions. Thanks to portable and wearable devices, ECG monitoring is no longer limited to hospital settings but is common in ambulatory and home environments.

However, the clinical usefulness of the ECG strongly depends on the quality of the signal. When the recorded ECG is corrupted by noise, interpretation becomes harder and, in the worst case, can lead to delayed or incorrect diagnoses. Improving ECG denoising is not only a technical challenge, but also has a direct impact on patient care.

1.2 The problem of noise in ECG signals

In practice, an ECG rarely coincides with an ideal, noise-free trace. During acquisition, several factors introduce disturbances that alter the waveform and may hide physiological components of interest. The most relevant sources of noise include:

- **Baseline Wander (BW):** a slow drift of the baseline, typically at very low frequencies (around 0.05–0.5 Hz), caused by respiration, changes in electrode–skin contact, and patient movement.
- **Muscular artifacts (EMG):** broadband noise generated by skeletal muscle activity, particularly relevant in ambulatory monitoring, Holter recordings, and wearable devices.

- **Power Line Interference (PLI):** narrow-band sinusoidal interference at 50 or 60 Hz induced by the power grid and nearby electrical equipment.

These components degrade visual quality and can also distort quantitative parameters such as amplitudes, intervals, and slopes. In some cases, noise may emulate pathological patterns (e.g., apparent ST-segment elevation due to baseline wander) or mask anomalies that are actually present.

For these reasons, the signal processing and biomedical engineering communities have committed considerable effort to ECG denoising algorithms that remove or attenuate noise while preserving the morphology and diagnostic content of the original signal.

1.3 Denoising approaches: classical filters, Deep Learning, and black-box models

Early ECG denoising was based on classical digital signal processing. High-pass filters are commonly used to mitigate baseline wander; notch filters suppress power line interference; and more advanced methods such as Independent Component Analysis (ICA), Empirical Mode Decomposition (EMD), and wavelet denoising have been proposed to deal with muscular artifacts.

Although effective in controlled scenarios, these methods show limitations. Performance is sensitive to design choices (e.g. cutoff frequencies, wavelet type and levels), they can struggle with non-stationary noise, and they may accidentally remove relevant ECG components together with noise, degrading clinical information.

More recently, machine learning, and especially Deep Learning, has been explored as a flexible alternative. Architectures such as denoising autoencoders, recurrent neural networks (LSTM/GRU), and generative models (GAN/VAE) learn a mapping from noisy to clean ECG directly from data and often outperform traditional filters under complex noise conditions.

A particularly relevant line of work is *noise estimation*. Instead of reconstructing the clean signal, the model estimates the noise component(s) and subtracts them from the input. Prior work has shown that a fully convolutional multi-task autoencoder can estimate the three main ECG noise sources (baseline wander, muscular artifacts, and power line interference) and achieve better denoising than a conventional autoencoder trained for direct reconstruction, when both share the same backbone.

Despite their strong results, most of these systems are *black-box*. They stand on networks with hundreds of thousands of parameters, whose internal representations are difficult to interpret. Moreover, analyzes have reported problematic behavior such as the "hallucination" of noise components that are not present (e.g., predicting BW/PLI on EMG-only signals), which can degrade the output in specific scenarios.

1.4 Motivation for interpretable and white-box models

In medical applications, interpretability is increasingly regarded as a key requirement. The main focus is not only on numerical performance, but also on understanding *why* a system produces a certain output, which features it stands on, and how robust its conclusions are to input variations.

Conventional Deep Learning models offer limited transparency, especially when operating directly on raw waveforms. When a black-box network fails to denoise an ECG correctly, it is difficult to identify the root cause and to control how the model interacts with morphology. Removing parts of the QRS complex, P wave, or T wave together with noise is a critical clinical risk.

White-box models such as *Tsetlin Machines* provide a viable alternative. They operate on interpretable binary features and learn sets of propositional if–then rules implemented by collections of Tsetlin automata. Decisions can be traced back to human-readable clauses, offering much greater transparency than typical deep architectures.

This thesis lies at the intersection of these directions. On one hand, it adopts the noise estimation paradigm, where denoising is driven by explicit estimates of noise. On the other hand, it replaces the convolutional network with a white-box framework based on Tsetlin Machines, focusing not on reconstructing noise waveforms but on predicting an *interpretable intensity* for each noise type.

1.5 Thesis objectives

The overall objective is to develop an *interpretable* ECG denoising system based on Tsetlin Machines and the *noise intensity estimation* paradigm. The goal is to combine:

- competitive quantitative performance on intensity estimation and denoising;
- reduced computational complexity;
- transparency of the decision process;
- the ability to drive an adaptive filtering pipeline.

The specific objectives are:

1. **Define a framework for ECG noise intensity estimation.**
   Reformulate the task from predicting the noise waveform to estimating, for each time window, a continuous intensity in $[0, 1]$ for baseline wander, muscular artifacts, and power line interference.
2. **Design a Regression Tsetlin Machine (RTM) for intensity estimation.**
   Develop RTM models for each noise type, operating on feature sets derived from the ECG (e.g. bitplane-based features and high-frequency spectral features), and study their accuracy and interpretability.

3. **Integrate intensity estimation into an adaptive filtering pipeline.**
   Use the predicted intensities to control digital filters (notch, high-pass, wavelet-based denoising), modulating their strength based on the estimated noise level to improve SNR while avoiding over-filtering and hallucination.

4. **Exploit interpretability to guide model optimization.**
   Utilizes intrinsic explainability (feature importance, clause inspection, sparsity) to refine the feature set and hyperparameters, treating interpretability as a driver for optimization rather than a post-hoc explanation.

5. **Evaluate overall performance.**
   Assess the pipeline on intensity estimation (correlation, mean absolute error, root mean squared error) and denoising (signal-to-noise ratio and its improvement) and compare against neural-network-based approaches.

1.6 Main contributions

The main contributions of this thesis are:

- **Tsetlin Machines for ECG noise intensity estimation.**
  This is one of the first applications of regression Tsetlin Machines to quantitative noise intensity estimation on biomedical signals. The system employs three independent RTM models for baseline wander, muscular artifacts, and power line interference.

- **Intensity-based noise estimation instead of waveform prediction.**
  The framework predicts normalized intensity values per window rather than full noise waveforms. This simplifies the regression target, reduces sensitivity to local waveform errors, and enables direct control of downstream filters.

- **Compact, interpretable hybrid feature set.**
  Through an explainability-guided process, we derive a hybrid set combining bitplane-based amplitude statistics and high-frequency spectral descriptors. It outperforms the individual groups while reducing the number of features by roughly an order of magnitude compared to the initial baseline.

- **Interpretability as an optimization lever.**
  Feature-importance analysis reveals a highly sparse structure, where a small subset of features accounts for most predictive power. We use this insight to progressively simplify the model (from baseline to hybrid and optimized variants), improving average correlation while reducing complexity.

- **Adaptive filtering that avoids noise hallucination.**
  Intensity estimates drive a gating logic that activates each filter only when the corresponding intensity is significantly above zero, mitigating the risk of subtracting non-existent noise.

- **Rigorous statistical validation.**
  We validate the clinical reliability of the system through extensive per-subject analysis, bootstrap confidence intervals, and Bland-Altman agreement analysis, demonstrating consistent performance across a diverse patient population.

- **Real-time feasibility.**
  The optimized configuration achieves millisecond-level inference on a standard CPU with a small memory footprint, suggesting suitability for real-time or near real-time ECG monitoring.

1.7 Thesis outline

The remainder of the document is organized as follows.

- **Chapter 2 – Background.**
  Fundamental concepts: ECG morphology, noise sources, signal quality metrics, and traditional denoising methods.

- **Chapter 3 – Related Works.**
  Deep learning for ECG denoising, noise estimation vs. direct reconstruction, and interpretable ML in biosignals.

- **Chapter 4 – Methodology.**
  Problem formulation, feature engineering, RTM models, training and calibration, and the intensity-guided denoising pipeline.

- **Chapter 5 – Experimental Results.**
  Results on intensity estimation, end-to-end denoising, explainability analyzes, and error analysis.

- **Chapter 6 – Discussion.**
  Comparison with CNN-based noise estimation, strengths and limitations, and clinical implications.

- **Chapter 7 – Conclusions and Future Work.**
  Summary of contributions and directions for future work.

Chapter 2

Background

2.1 Signal Processing Foundations

The analysis of biomedical data relies fundamentally on the rigorous application of Digital Signal Processing theory [1, 2]. A signal, in its most abstract definition, is a physical quantity that varies as a function of one or more independent variables. It conveys information about the behavior or attributes of a phenomenon. In the context of cardiovascular monitoring, the primary signal of interest is the Electrocardiogram [3]. This signal is a manifestation of the electrophysiological currents generated by the heart muscle over time. While the underlying biological processes are continuous, modern analysis pipelines operate exclusively within the digital domain. This necessitates a robust understanding of the transformation from analog to digital representations and the statistical characterization of stochastic processes.

2.1.1 Signal Classification and Properties

Signals can be categorized based on the predictability of their behavior and the nature of their domain. This distinction is critical for selecting appropriate denoising strategies and understanding the limitations of classical filtering approaches.

2.1.1.1 Deterministic versus Stochastic Signals

A deterministic signal is one whose future values can be precisely calculated from a mathematical model or a known rule. For example, the 50 Hz or 60 Hz sinusoidal hum generated by power lines is largely deterministic, as its frequency and waveform are stable and predictable. In contrast, biomedical signals such as the ECG are inherently stochastic [4]. They are realizations of random processes that cannot be described by an explicit time function. Instead, they must be characterized by statistical properties. These include their probability density function, mean (first moment), and variance (second moment). This stochastic nature implies that an ECG signal cannot be predicted with absolute certainty. However, its statistical behavior often exhibits patterns that machine learning models can learn to recognize.

2.1.1.2 The Concept of Stationarity

A stochastic process is defined as Strict-Sense Stationary if its statistical properties are invariant to a shift in the time origin. In simpler terms, the fundamental nature of the signal generator does not change over time. However, physiological signals are rarely stationary [3]. Heart rate varies due to autonomic nervous system regulation, known as Heart Rate Variability. The amplitude of the signal may fluctuate with respiration, and the characteristics of noise artifacts change abruptly with patient movement. This non-stationarity poses the primary challenge for classical linear filtering techniques. These techniques often assume static noise profiles, such as a fixed cutoff frequency. This limitation motivates the need for adaptive, data-driven approaches that can dynamically adjust their parameters in response to changing signal statistics.

2.1.2 Digital Representation: Sampling and Quantization

The transition from a continuous physical phenomenon to a digital sequence involves two distinct discretization steps. These are sampling, which is the discretization of time, and quantization, which is the discretization of amplitude [1].

2.1.2.1 The Sampling Theorem

Sampling is the process of converting a continuous-time signal into a discrete-time sequence by measuring its instantaneous value at periodic intervals. The sampling frequency is defined as the reciprocal of this interval. The relationship between the continuous signal and its discrete counterpart is governed by the Nyquist-Shannon Sampling Theorem [5, 6]. This fundamental theorem states that a band-limited continuous signal with a maximum frequency component can be perfectly reconstructed from its samples if and only if the sampling frequency is at least twice the maximum frequency. The threshold frequency is known as the Nyquist frequency. If the analog signal contains energy at frequencies higher than this threshold, a phenomenon known as aliasing occurs. During the sampling process, these high-frequency components fold back into the lower frequency spectrum. They are indistinguishable from the genuine low-frequency signal components.

In the context of ECG analysis, aliasing is catastrophic. High-frequency noise, such as electromyographic artifacts, can alias into the frequency band of the QRS complex or the P-wave if not properly filtered before sampling. This mimics pathological markers or obscures diagnostic features. For diagnostic ECGs, standard guidelines typically recommend sampling rates of 500 Hz or higher to accurately capture rapid depolarization of the ventricles [4]. However, ambulatory Holter monitors often trade temporal resolution for storage capacity, operating at lower rates.

2.1.2.2 Quantization and Dynamic Range

While sampling discretizes the independent variable, quantization discretizes the dependent variable. The continuous voltage range of the analog signal is mapped onto a finite set of discrete levels. This is determined by the bit-depth of the Analog-to-Digital Converter. An ADC with a specific bit resolution divides the input voltage range into discrete levels. Unlike sampling, which can be theoretically lossless given bandwidth constraints, quantization is an inherently lossy process [2]. The difference between the actual analog value and the quantized digital value is known as the quantization error. This is often modeled as additive white noise uniformly distributed across the quantization step.

In ECG monitoring, sufficient bit-depth is crucial. The signal amplitude can range from a few millivolts for the R-wave peak to mere microvolts for P-wave or late potentials. A low-resolution quantization results in a staircase effect that destroys subtle morphological details essential for diagnosis. Modern acquisition systems typically use 12-bit or 16-bit converters to ensure that the quantization noise floor remains well below the physiological noise floor.

2.1.3 Frequency and Spectral Analysis

Analyzing signals in the time domain, or observing the waveform shape, is intuitive but often insufficient for separating signal from noise. This is especially true when they overlap over time. The frequency domain offers a complementary perspective, where signal components are distinguished by their oscillation rate.

2.1.3.1 The Fourier Transform

The mathematical tool used to decompose a signal into its constituent frequencies is the Fourier Transform. For discrete signals of finite length, the Discrete Fourier Transform (DFT) transforms the sequence into a sequence of complex numbers [7]. The magnitude spectrum reveals the energy distribution of the signal across different frequency bins. This representation is critical for understanding the nature of noise. For instance, Baseline Wander is concentrated at very low frequencies, while Power Line Interference appears as a sharp peak at 50 Hz or 60 Hz.

2.1.3.2 Power Spectral Density and Welch's Method

For stochastic signals like ECG mixed with random noise, a simple DFT of the entire recording is often too noisy to be useful for robust feature extraction. The spectrum of a random signal is itself a random variable with high variance. To obtain a stable estimate of the frequency content, the Power Spectral Density (PSD) is utilized. This describes how the power of the signal is distributed over frequency.

Specifically, Welch's Method is preferred for PSD estimation in this context [8]. Welch's method improves on the standard periodogram by dividing the signal into overlapping segments. It applies a window function, such as the Hamming window

[9], to each segment to reduce spectral leakage. It then computes the DFT for each segment and averages the results. This averaging process significantly reduces the variance of the spectral estimate, providing a smooth and reliable profile of the noise characteristics. This spectral characterization forms the basis of the feature engineering phase. Features derived from specific spectral bands are often the most discriminative inputs for estimating the intensity of muscle artifacts.

2.1.4 Noise Theory and Metrics

In the signal processing domain, noise is formally defined as any unwanted disturbance that obscures or interferes with the signal of interest [10]. While mathematically noise is often modeled as an additive term, the physical reality involves complex interactions between the signal source and the interference.

2.1.4.1 Spectral Characteristics of Noise

Noise is often characterized by its color, an analogy to optics referring to the shape of its Power Spectral Density. White Noise has a constant power spectral density across all frequencies. Its samples are uncorrelated over time. Thermal noise in electronic circuits is classically modeled as White Gaussian Noise. Colored Noise has a non-flat PSD. For example, Pink noise has a power density proportional to the inverse of the frequency. Many biological systems and low-frequency electronic drifts exhibit these characteristics. Narrowband Noise is concentrated at a specific frequency range. The most common example is sinusoidal interference from power lines.

2.1.4.2 Quantitative and Qualitative Metrics

To quantify the severity of noise and the efficacy of removal techniques, specific metrics are required. The Signal-to-Noise Ratio (SNR) is the standard measure. It is defined as the ratio of signal power to noise power, typically expressed in decibels. A higher SNR indicates a cleaner signal. For denoising tasks, the key performance indicator is the SNR Improvement. This is defined as the difference between the output SNR and the input SNR. Additionally, the Root Mean Square Error (RMSE) provides a direct measure of the average deviation between the reconstructed signal and the ground truth in the signal units.

However, in medical applications, mathematical closeness is not always equivalent to diagnostic quality. A filtering algorithm might achieve a low RMSE by aggressively smoothing the signal. If this smoothing obliterates the small Q-wave of a myocardial infarction, the clinical utility is lost. Therefore, quantitative metrics must always be interpreted alongside qualitative assessments of morphological preservation.

2.2 The Electrocardiogram (ECG)

The Electrocardiogram is the graphical representation of the electrical potential generated by the heart as it propagates to the skin surface. Since its invention by

Willem Einthoven in the early 20th century, the ECG has remained the cornerstone of cardiovascular diagnosis. It provides a low-cost, not invasive, and immediate window into the electromechanical function of the heart [4]. This section explores the physiological genesis of the signal, its morphological characteristics, and the various sources of interference that complicate its automated analysis in ambulatory settings.

2.2.1 Physiology and Morphology

To correct interpret the ECG signal, it is necessary to understand the underlying physiology that generates it. The heart operates as an electromechanical pump composed of four chambers. Its pumping action is coordinated by a specialized electrical conduction system that ensures synchronized contraction.

2.2.1.1 The Cardiac Cycle and Electrical Conduction

The cardiac cycle initiates at the Sinoatrial Node, located in the right atrium. The Sinoatrial node acts as the natural pacemaker of the heart. It spontaneously generates an electrical impulse through the movement of ions across cell membranes [3]. This wave of depolarization spreads through the atrial muscle fibers. It causes them to contract and pump blood into the ventricles. The electrical signal then converges at the Atrioventricular Node. This node acts as a critical gatekeeper. It delays the signal to ensure that the atria have fully emptied before the ventricles contract. From the Atrioventricular node, the impulse travels rapidly down the Bundle of His and spreads through the Purkinje fibers. This causes a synchronized and massive depolarization of the ventricular muscle mass.

2.2.1.2 Waveform Components

This sequence of electrical events projects onto the body surface as a characteristic waveform comprising waves, segments, and intervals. The fidelity of these morphological features is paramount for clinical diagnosis.

The P-wave is a small, rounded deflection representing atrial depolarization. Its presence, shape, and timing relative to the QRS complex are vital for diagnosing arrhythmias, such as atrial fibrillation or atrioventricular blocks.

The QRS Complex represents the rapid depolarization of the large ventricular mass. It is the most prominent feature of the ECG and serves as the primary fiducial point for heart rate calculation. It typically consists of the Q-wave, the R-wave, and the S-wave. The width of the QRS complex indicates the speed of conduction. A widened QRS suggests a bundle branch block or a ventricular origin of the beat. The amplitude is also significant as high voltage can indicate hypertrophy of the heart muscle.

The T-wave follows the QRS complex and represents ventricular repolarization. This is the recovery phase of the heart muscle where the ions return to their resting state.

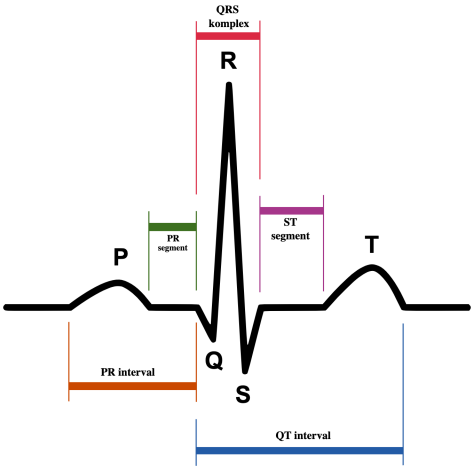


Figure 2.1: Typical ECG waveform of a normal heartbeat cycle. The annotations highlight the P-wave (atrial depolarization), the QRS complex (ventricular depolarization), and the T-wave (ventricular repolarization), which are critical morphological features for diagnosis.

The ST-Segment is the isoelectric line between the end of the QRS complex and the start of the T-wave. Elevation or depression of this segment relative to the baseline is the hallmark sign of myocardial ischemia and acute myocardial infarction.

The diagnostic power of the ECG lies in the precise timing and shape of these components. For instance, a shift in the ST-segment of just 0.1 mV is clinically significant. This sensitivity highlights why noise removal is a delicate task. A denoising algorithm that suppresses noise but also suppresses a pathological elevation could lead to a missed diagnosis.

2.2.2 Artifacts and Noise Sources in Ambulatory ECG

In a controlled clinical setting, obtaining a clean ECG is relatively straightforward. However, modern cardiology is moving towards ambulatory monitoring using Holter monitors or wearable patches to detect intermittent arrhythmias [11]. In these real-world scenarios, the ECG is susceptible to various noise sources. Three dominant artifacts are particularly relevant to this research [10].

2.2.2.1 Baseline Wander

Baseline Wander is a low-frequency oscillation of the signal's isoelectric line. It typically operates in the frequency range of 0.05 Hz to 0.5 Hz. The primary cause is respiration. As the patient breathes, the chest expands and contracts. This alters the impedance between the electrodes and the heart and physically moves the electrodes relative to the cardiac vector. Other causes include perspiration and slow body movements. While Baseline Wander does not typically obscure the high-frequency QRS complex, it poses a serious threat to the diagnosis of ischemia. Since ischemia is detected by measuring the absolute amplitude of the ST-segment relative to the

baseline, a wandering baseline can mimic or mask pathological elevation. Furthermore, large amplitude drifts can saturate the dynamic range of the recording equipment, leading to signal loss.

2.2.2.2 Electromyographic Interference

Electromyographic Interference, also known as muscle artifact, originates from the electrical activity of skeletal muscles. Every time the patient moves, tenses, or shivers, the muscles generate electrical potentials that are picked up by the surface electrodes. This type of noise is arguably the most challenging to remove because of its spectral characteristics. Unlike the rhythmic cardiac signal, Electromyographic noise is stochastic and broad-band. Its energy is distributed from DC to several hundred Hertz. Crucially, there is a significant spectral overlap between this noise and the diagnostic bandwidth of the ECG, particularly the QRS complex. This means that simple low-pass filtering often fails. Removing high-frequency muscle noise inevitably blurs the sharp edges of the QRS complex. This reduces its amplitude and extends its duration, potentially mimicking conduction abnormalities.

2.2.2.3 Power Line Interference

Power Line Interference is an environmental noise caused by capacitive coupling with AC power lines and electrical equipment. It appears as a continuous sinusoidal component at the main frequency. This frequency is 50 Hz in Europe and Asia, and 60 Hz in the Americas. While visually disruptive, Power Line Interference is stationary and narrow-band. However, its amplitude can be massive, completely swamping the P and T waves. Furthermore, non-linearities in the acquisition chain can generate harmonics at multiples of the fundamental frequency. These harmonics can also be within the diagnostic ECG bandwidth.

2.2.3 Classical Denoising Techniques

Before the advent of machine learning, ECG denoising relied on deterministic algorithms rooted in signal processing theory. These approaches remain the industry standard for many embedded devices due to their low computational footprint.

2.2.3.1 Linear Digital Filtering

The most straightforward approach involves linear frequency-selective filters. High-pass filters are used to remove Baseline Wander, typically with a cutoff frequency around 0.5 Hz. Low-pass filters are used to attenuate high-frequency muscle noise. Notch filters are specifically designed to surgically remove the 50 Hz or 60 Hz Power Line Interference. While computationally efficient, fixed-coefficient filters lack in adaptability. They cannot distinguish between noise and signal components that overlap in frequency. A filter aggressive enough to clean muscle noise will often degrade sharp physiological waveforms.

2.2.3.2 Adaptive Filtering

To address non-stationarity, Adaptive Filtering techniques were introduced [12, 13]. Algorithms such as Least Mean Squares or Recursive Least Squares [14] can adjust filter coefficients in real-time to track changing noise environments. However, adaptive filters typically require a reference channel. This is a separate signal that correlates with noise but not with the ECG. In standard single-lead wearable devices, such a reference is rarely available, limiting the applicability of this method.

2.2.3.3 Transform Domain Decomposition

More sophisticated methods utilize transform domains. The Discrete Wavelet Transform (DWT) decomposes the signal into different scales or frequency bands [15]. Denoising is performed by thresholding the wavelet coefficients, a technique pioneered by Donoho [16]. This involves keeping the large coefficients, which are assumed to vary with the QRS complex, and setting to zero the small ones, which are assumed to correspond to noise. While being powerful, wavelet denoising requires expert selection of the mother wavelet and complex thresholding rules. It can also introduce localized artifacts near signal discontinuities. Similarly, Empirical Mode Decomposition (EMD) adaptively breaks the signal into Intrinsic Mode Functions [17]. While this avoids the selection of a fixed basis function, it is computationally expensive and prone to mode-mixing artifacts.

2.3 Machine Learning in Signal Denoising

The limitations of classical filtering techniques, particularly their inability to adapt to non-stationary noise environments without external reference channels, have driven the exploration of data-driven approaches. Machine Learning represents a paradigm shift from designing explicit rules for noise removal to learning these rules from data [18]. This section provides a comprehensive overview of learning paradigms, details the core mathematical operations utilized in modern signal processing models, and critically analyzes the dominant deep learning architectures and their associated interpretability challenges.

2.3.1 Machine Learning Paradigms

Machine Learning algorithms can be categorized based on the nature of the signal available to the learning system and the type of feedback provided during the training process.

2.3.1.1 Supervised Learning

Supervised learning is the most prevalent approach in signal processing tasks where ground truth data is available. In this paradigm, the model is provided with a dataset consisting of input-output pairs. The goal is to learn a mapping function

that predicts the output of new unseen inputs. In the context of ECG denoising, this typically involves training a model using pairs of noisy signals (input) and clean signals (target). The algorithm iteratively adjusts its internal parameters to minimize the discrepancy, or loss, between its predictions and the provided targets.

2.3.1.2 Unsupervised Learning

Unsupervised learning deals with unlabeled data. The system attempts to infer the inherent structure or distribution of the input data without explicit guidance on what the output should look like. Common applications include clustering and dimensionality reduction. In biomedical signal processing, unsupervised techniques like Principal Component Analysis (PCA) [19] or Independent Component Analysis (ICA) [20] have historically been used to separate signal sources (e.g., separating fetal ECG from maternal ECG) based on statistical independence rather than labeled examples.

2.3.1.3 Reinforcement Learning

Reinforcement Learning (RL) differs fundamentally from the previous two categories. Instead of being told what to predict, an agent interacts with an environment and learns to make a sequence of decisions to maximize a cumulative reward signal. The agent explores the environment, takes actions, and receives feedback in the form of rewards or penalties. While less common for direct end-to-end signal filtering, RL concepts are foundational to the operation of Tsetlin Machines [21]. As will be discussed in the following section, the internal automata of a Tsetlin Machine operate as independent reinforcement learning agents, adjusting their states based on rewards and penalties derived from the prediction accuracy.

2.3.2 Core Computational Concepts: Regression and Convolution

Two mathematical concepts are central to understanding how modern machine learning models process physiological time-series: regression (the task) and convolution (the operation).

2.3.2.1 Regression Analysis

Machine learning tasks are generally divided into classification and regression. While classification assigns inputs to discrete categories (e.g., "Arrhythmia" vs. "Normal") [22], regression involves predicting a continuous quantity. Formally, regression seeks to estimate the relationship between a dependent variable (the target) and one or more independent variables (the predictors). In signal denoising, the problem is inherently a regression task. If the goal is to reconstruct the clean signal, the model must predict the continuous voltage amplitude of the ECG at each time step. Alternatively, in the noise estimation approach proposed in this work, the model performs regression to predict the continuous intensity of the noise artifacts (a value

ranging from 0 to 1). The performance of a regression model is typically evaluated using metrics such as Mean Squared Error (MSE) or Mean Absolute Error (MAE), which quantify the distance between the predicted continuous values and the actual values.

2.3.2.2 The Convolution Operation

Convolution is the mathematical backbone of most modern deep learning architectures for signal and image processing. In the context of 1-D signals like ECG, convolution involves sliding a fixed size filter (or kernel) over the input sequence. At each step, the dot product between the filter weights and the local segment of the input signal is computed [18]. This operation possesses two properties that make it highly effective for biomedical signals:

1. **Local Connectivity:** The network learns to analyze local patterns (such as the sharp rise of an R-peak or the high-frequency oscillation of muscle noise) rather than looking at the entire signal at once.
2. **Parameter Sharing:** The same filter is applied on the entire time series. This creates translation invariance, meaning a specific feature (like a QRS complex) can be recognized regardless of where it appears in the cardiac cycle.

However, convolution relies heavily on floating-point matrix multiplications, which contributes to the high computational cost and energy consumption of Convolutional Neural Networks (CNNs).

2.3.3 Deep Learning Architectures for ECG

Deep Learning models, which utilize neural networks with many layers to learn hierarchical representations of data, have become the standard for high-performance signal restoration [18].

2.3.3.1 Convolutional Neural Networks (CNNs)

While originally designed for computer vision, CNNs have been successfully adapted for one-dimensional time series. In these architectures, multiple layers of convolutional filters are stacked. Lower layers may learn to detect simple features like edges or spikes, while deeper layers combine these features to recognize complex morphological structures like the P-QRS-T complex. The translation-invariant nature of CNNs makes them particularly robust for analyzing continuous monitoring data where heartbeats occur at irregular intervals. Kiranyaz et al. demonstrated the efficacy of 1-D CNNs for patient-specific ECG classification [23], and Acharya et al. extended this to deep CNNs for arrhythmia detection [24].

2.3.3.2 Recurrent Neural Networks (RNNs) and LSTMs

Recurrent Neural Networks are designed to process sequential data by maintaining an internal memory state. Advanced variants, such as Long Short-Term Memory

(LSTM) networks [25] or Gated Recurrent Units (GRU) [26], address the limitations of standard RNNs in learning long-term dependencies. These architectures are capable of exploiting the temporal context of the signal, using the history of previous cardiac cycles to inform the reconstruction of the current segment. However, the sequential nature of RNNs often results in slower training and inference times compared to parallelizable convolutional models.

2.3.3.3 Denoising Autoencoders

The Denoising Autoencoder (DAE) is a specific architectural framework used for signal restoration [27]. It consists of an encoder, which compresses the noisy input into a lower-dimensional latent representation, and a decoder, which attempts to reconstruct the clean signal from this latent code. The bottleneck forces the model to learn robust features that capture the essential structure of the physiological signal while discarding the stochastic noise, which is less compressible. Chiang et al. applied Fully Convolutional Denoising Autoencoders to ECG signals, achieving significant noise reduction [28].

2.3.4 Denoising Paradigms: Reconstruction versus Estimation

Within the Deep Learning landscape, two distinct strategies for signal restoration have emerged based on the target of the regression.

2.3.4.1 Direct Denoising

The most intuitive approach is Direct Denoising, or End-to-End Reconstruction. Here, the neural network is trained to map the noisy input signal directly to the clean output signal [28]. The model is tasked with reconstructing the complex morphology of the ECG. This approach is effective, but places a heavy burden on the model to preserve minute diagnostic details. Errors in the model can lead to signal distortion, such as smoothing of high-frequency pathological markers.

2.3.4.2 Noise Estimation

Recent advancements have shifted focus towards Noise Estimation, or Residual Learning. Instead of predicting the clean signal, the model is trained to estimate the additive noise component. The clean signal is then recovered by subtracting the estimated noise from the noisy input. This approach relies on the observation that noise artifacts, such as sinusoidal Power Line Interference or low-frequency Baseline Wander, often possess a simpler or more distinct morphology than the ECG signal itself. Empirical studies, such as the multitask noise estimation framework proposed in [29], have suggested that this paradigm yields superior Signal-to-Noise Ratio improvements.

2.3.5 The Interpretability Challenge

Despite the impressive numerical performance of Deep Learning models, their adoption in clinical practice faces a significant barrier known as the Black-Box Problem.

2.3.5.1 Opacity and Lack of Trust

Deep neural networks consist of millions of floating-point parameters and non-linear activation functions. The decision-making process is distributed across this vast web of connections, making it mathematically opaque. When a CNN filters an ECG, it is virtually impossible for a clinician to trace the logic behind the specific alteration of a waveform segment. There is no explicit rule stating why a particular voltage deflection was suppressed. This lack of transparency erodes trust, as medical professionals cannot verify whether the model's actions are based on physiological principles or spurious statistical correlations.

2.3.5.2 Hallucinations and Generative Artifacts

A critical risk associated with generative Deep Learning models is the phenomenon of hallucination. Because these models learn the statistical distribution of the training data, they may reconstruct plausible features that are not actually present in the input. In high-noise scenarios, a model might "invent" a P-wave to conform to the learned pattern of a normal heartbeat, potentially masking an arrhythmia. In contrast, it might suppress a pathological Q-wave, interpreting it as noise. In healthcare, where decisions can have life-altering consequences, the inability to distinguish between a restored signal and a generated artifact is a profound liability. This necessitates the exploration of alternative machine learning frameworks that offer both high performance and intrinsic interpretability.

2.4 The Tsetlin Machine Framework

To address the dichotomy between the high accuracy of Deep Learning and the critical need for transparency in medical applications, this thesis adopts the Tsetlin Machine (TM). Introduced by Ole-Christoffer Granmo in 2018 [21], the TM represents a fundamental departure from the arithmetic-based paradigm of neural networks. Instead of relying on matrix multiplications, differentiation, and backpropagation, the TM learns patterns using propositional logic and bitwise operations. This unique architecture inherently produces human-readable rules, offering a “White-Box” alternative that maintains competitive accuracy while addressing the opacity issues intrinsic to connectionist models.

The following subsections provide a comprehensive analysis of the framework, tracing its origins from cybernetic theory to the specific regression architecture used for ECG noise estimation.

2.4.1 Historical Origins: Cybernetics and Automata Theory

While the Tsetlin Machine is a contemporary invention, its intellectual lineage traces back to the golden age of Soviet cybernetics in the 1960s. The central figure in this domain was **Michael Lvovitch Tsetlin**, a physicist and mathematician who investigated how simple biological organisms could adapt to complex environments without having a prior model of the world [30].

Tsetlin formalized the concept of the **Learning Automaton** (LA), a decision-making unit that interacts with a stochastic environment through a trial-and-error process. He proposed the "Linear Tactic Automaton," a finite state machine designed to solve the multi-armed bandit problem. The objective of such an automaton is to select the action that maximizes the probability of receiving a reward from an environment whose statistical properties are unknown.

2.4.1.1 The Linear Tactic Automaton

The Linear Tactic Automaton is defined by a finite memory of $2N$ states. The states are divided into two sets corresponding to two possible actions, typically Action 1 and Action 2.

- States 1 to N are associated with Action 1.
- States $N+1$ to $2N$ are associated with Action 2.

The interaction follows a reinforcement learning loop. At any given time step, the automaton selects an action based on its current state. The environment responds with a feedback signal: either a *Reward* (success, penalty=0) or a *Penalty* (failure, penalty=1). In Tsetlin's design, a Reward reinforces the current action by moving the state "deeper" into the memory (away from the center), representing increased confidence. Conversely, a Penalty degrades confidence by moving the state towards the central boundary (N or $N+1$). If the automaton receives sufficient penalties while in a boundary state, it transitions to the opposing set of states, effectively switching its action.

Tsetlin proved mathematically that, despite its extreme simplicity—relying only on integer states and simple increments/decrements—this automaton is ϵ-optimal. This means that as the memory depth N approaches infinity, the probability of the automaton choosing the optimal action approaches unity, even in stochastic environments. This theoretical guaranty provides the robust foundation upon which the modern Tsetlin Machine is built.

2.4.2 The Tsetlin Machine Architecture

Granmo's breakthrough was to organize thousands of these simple automata into a collective intelligence capable of learning complex logic patterns. The architecture of a standard Tsetlin Machine is composed of three layers: the input layer (literals), the pattern recognition layer (conjunctive clauses), and the summation layer (voting) [21].

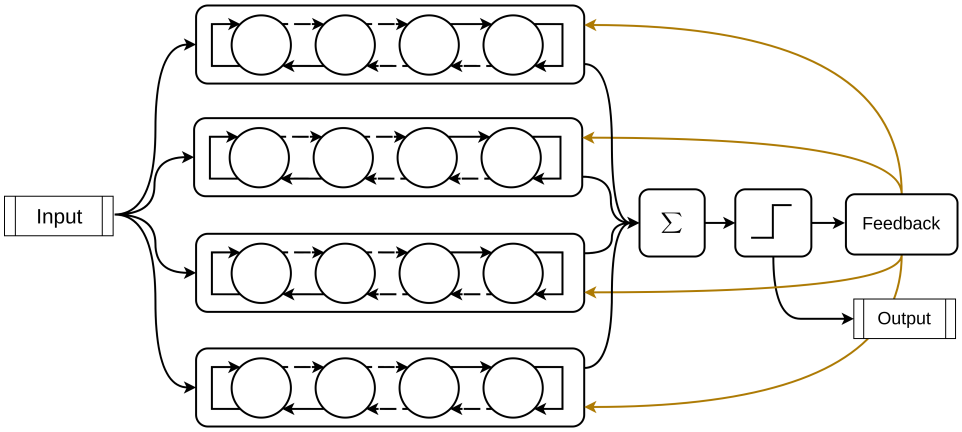


Figure 2.2: Schematic representation of the standard Tsetlin Machine architecture. Input features are converted into literals, which are processed by teams of Tsetlin Automata to form conjunctive clauses. The final output is determined by aggregating the votes from active clauses.

2.4.2.1 Propositional Input and Literals

The TM operates natively on Boolean inputs. To process continuous data, such as ECG signal features, a binarization step is required (discussed in Chapter 4). Given a binary feature vector $\mathbf{X} = [x_1, \ldots, x_o]$ of size o, the TM generates a set of **Literals** L. To maximize expressivity and allow the model to detect both the presence and absence of features, the literal set includes both the original features and their logical negations:

$$L = \{x_1, \ldots, x_o, \neg x_1, \ldots, \neg x_o\} \tag{2.1}$$

Thus, for a feature vector of dimension o, the total number of literals available to the system is $2o$.

2.4.2.2 Conjunctive Clauses

The core learning unit of the TM is the **Conjunctive Clause**. A clause C_j is a logical AND rule formed by selecting a specific subset of the available literals. Formally, if I_j denotes the set of literals included in clause j, the clause output is:

$$C_j(\mathbf{X}) = \bigwedge_{l_k \in I_j} l_k \tag{2.2}$$

where l_k represents the k-th literal present in the subset I_j. A clause outputs 1 (True) if and only if all its included literals evaluate to 1. If the set I_j is empty, the clause outputs 1 by definition during learning, ensuring that all clauses participate in the initial exploration of the solution space. The composition of the set I_j is dynamic and is determined by the Tsetlin Automata.

2.4.2.3 The Team of Automata

For every clause C_j and for every literal l_k, there is a dedicated Tsetlin Automaton $TA_{j,k}$. This automaton decides whether to **Include** or **Exclude** the literal k from clause j.

- **Exclude:** If the automaton state is between 1 and N.
- **Include:** If the automaton state is between $N + 1$ and $2N$.

This structure means that a Tsetlin Machine with m clauses and o input features utilizes $m \times 2o$ independent automata. These automata evolve independently based on feedback, yet they self-organize to form coherent patterns.

2.4.3 Learning Dynamics: The Game of Feedback

The learning process in a Tsetlin Machine is not driven by gradient descent or backpropagation. Instead, it is formulated as a game played by the automata against the environment. The goal of the game is to maximize the number of correct classifications. The mechanism relies on two distinct types of feedback signals, Type I and Type II, which are dispensed probabilistically based on the state of the clause and the correctness of the prediction [21].

2.4.3.1 Resource Allocation and Thresholding

To ensure that the clauses distribute themselves to cover different sub-patterns in the data, the TM employs a resource allocation mechanism controlled by a summation threshold T. The votes of the clauses are summed to form a total output score. During training, the feedback probability is scaled based on how close this score is to the threshold T. If the system is already confident in a prediction (the vote sum exceeds T), the probability of updating the automata drops to zero. This prevents the "over-hiring" of clauses for easy patterns and forces idle clauses to learn rare or difficult patterns, effectively regularizing the model.

2.4.3.2 Type I Feedback: Pattern Recognition and Refinement

Type I Feedback is triggered when a clause outputs 0 but should have outputted 1 (a False Negative scenario), or when it correctly outputs 1 but needs reinforcement. This feedback has two effects:

1. **Boost True Literals:** For literals that are 1 in the input, the corresponding automata receive a reward (Include). This reinforces the recognition of the pattern present in the input.
2. **Combat Overfitting:** For literals that are 0 in the input, the automata receive a penalty (Exclude). This removes literals that would contradict the pattern. This action is stochastic and controlled by a user-defined parameter s (specificity). A higher s makes the clauses more specific and less prone to overfitting.

2.4.3.3 Type II Feedback: Discrimination and Suppression

Type II Feedback is triggered when a clause outputs 1 but should have outputted 0 (a False Positive scenario). The goal is to silence the erroneous clause. The feedback forces the automata to **Include** literals that are 0 in the current input. By including a False literal in an AND-rule, the clause output is forced to 0, effectively suppressing the false activation. This mechanism ensures that the clauses learn to discriminate against patterns that belong to other classes.

2.4.4 State-of-the-Art in Biomedical Tsetlin Machines

Recent research has begun to explore the application of Tsetlin Machines to biomedical signals, demonstrating their potential as a viable alternative to deep learning. A pivotal study by Zhang et al. (2023) applied the Tsetlin Machine to the identification of Premature Ventricular Contractions (PVC) in ECG signals [22].

Their work provided a direct comparison between a TM-based classifier and state-of-the-art Convolutional Neural Networks (CNNs). The results showed that the TM achieved a classification accuracy (98.6%) comparable to that of deep learning models (98.9%). However, the crucial distinction lay in **interpretability**. Zhang et al. highlighted a fundamental limitation of CNNs: techniques like "Attention Maps" (e.g., Grad-CAM) can only indicate *where* the model is looking, but not *what* specific pattern it has found. In contrast, the TM produced explicit logical clauses. For instance, the model learned rules that directly correlated specific pixel patterns in the 2D ECG representation to the presence of a PVC. This allowed cardiologists to verify the diagnostic criteria used by the algorithm, confirming that the model was focusing on the QRS width and morphology, consistent with medical knowledge. This thesis builds upon this precedent, extending the application of Tsetlin Machines from *classification* (detecting the presence of an arrhythmia) to *regression* (quantifying the intensity of noise), leveraging the same benefits of transparency and trust.

2.4.5 The Regression Tsetlin Machine (RTM)

While the original TM was a classifier, the objective of estimating noise intensity requires a continuous output. To address this, the **Regression Tsetlin Machine (RTM)** is employed, as formalized by Abeyrathna et al. [31].

2.4.5.1 From Polarity to Magnitude

In the classification TM, clauses are assigned polarities to vote for different classes. In the RTM, this concept is replaced by **magnitude accumulation**. The system assumes a linear relationship between the number of active clauses and the target value. The continuous output $\hat{y}$ is computed as the normalized sum of all clause outputs:

$$\hat{y} = \frac{1}{T} \sum_{j=1}^{m} C_j(\mathbf{X}) \tag{2.3}$$

Here, the threshold T acts as a scaling factor, defining the maximum resolution of the regressor. An RTM can be viewed as a "digital" potentiometer: the more clauses are active (matching the input pattern), the higher the output value.

2.4.5.2 Error-Driven Feedback

The learning dynamics in the RTM are driven by the magnitude of the error between the predicted scalar $\hat{y}$ and the ground truth y. Unlike the simple correct/incorrect feedback of the classifier, the RTM modulates the probability of updating the automata based on the error size [31].

- **Underestimation ($\hat{y} < y$):** The model needs more active clauses. Type I Feedback is applied to inactive clauses to encourage them to recognize the current input. The probability of update is proportional to the error magnitude $|y - \hat{y}|$.
- **Overestimation ($\hat{y} > y$):** The model has too many active clauses. Type II Feedback is applied to active clauses to suppress them. Again, the aggressiveness of the update scales with the error.

This feedback loop creates a highly stable regression mechanism that gradually converges to the target intensity distribution.

2.4.6 A Practical Example: Learning the XOR Function

To demystify the operation of the TM and illustrate its non-linear capabilities, consider the classic “Noisy XOR” problem. This is a benchmark problem that simple linear models (like Logistic Regression) cannot solve because the classes are not linearly separable. Consider two binary inputs x_1, x_2 and a target output y such that $y = 1$ if $x_1 \neq x_2$ (Exclusive OR). The valid patterns for $y = 1$ are $(1, 0)$ and $(0, 1)$.

A Tsetlin Machine with just two clauses can solve this perfectly by discovering the underlying logic:

- **Clause 1 Formulation:** Through Type I feedback on examples where the input is $(1, 0)$, the automata for literals x_1 and $\neg x_2$ receive rewards and move to the "Include" state. The resulting rule converges to: $C_1 = x_1 \wedge \neg x_2$.
- **Clause 2 Formulation:** Similarly, training on $(0, 1)$ examples guides the second team of automata to learn the complementary pattern. The resulting rule converges to: $C_2 = \neg x_1 \wedge x_2$.

The final output of the machine is the logical sum (OR) of these clauses:

$$y = (x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_2) \tag{2.4}$$

Crucially, this logic is not "hidden" in a matrix of floating-point weights; it is explicit. A human can read the clauses and verify that the machine has correctly understood the logical definition of the XOR gate.

2.4.7 Advantages for Biomedical Signal Processing

The selection of the Tsetlin Machine for this research is driven by the specific convergence of three critical advantages that address the limitations of current Deep Learning approaches in the medical domain:

2.4.7.0.1 Intrinsic Interpretability. As demonstrated by Zhang et al. [22], the TM produces logic rules. In the context of ECG analysis, this means the model can "explain" its decisions. For instance, it might learn a rule stating: *"IF High Frequency Power is High AND R-peak amplitude is Low THEN Muscle Noise is High"*. This transparency allows clinicians to validate the model's reasoning, building trust in the automated system.

2.4.7.0.2 Computational Efficiency and Edge AI. The inference phase of the TM relies solely on bitwise operations (AND, NOT, XOR, Population Count) and integer addition. It requires no floating-point arithmetic or matrix multiplication [32]. This drastically reduces the computational complexity and energy consumption of the model. Consequently, TMs are uniquely suited for deployment on battery-powered medical wearables and low-cost microcontrollers, enabling advanced signal processing at the edge without the need for heavy hardware accelerators.

2.4.7.0.3 Robustness to Noise and Overfitting. The discrete nature of the automata states provides a buffer against overfitting. Unlike neural networks, which can memorize noise in the infinite precision of their weights, the TM seeks robust logic structures. This reduces the risk of "hallucinating" artifacts where none exist—a critical safety feature for automated denoising systems, preventing the distortion of physiological morphologies in clean signals.

Chapter 3

Related Works

This chapter situates the proposed research within the broader landscape of biomedical signal processing and interpretable Machine Learning. The literature review is organized into four main streams. Section 3.1 discusses traditional digital signal processing techniques, establishing the baseline for performance and computational efficiency. Section 3.2 explores the rise of Deep Learning (DL) for ECG restoration, distinguishing between direct waveform reconstruction and the emerging paradigm of noise estimation. Section 3.3 reviews the specific applications of the Tsetlin Machine framework in the biomedical domain, highlighting its potential for transparent classification and regression. Finally, Section 3.4 provides a critical gap analysis, defining the specific research questions addressed by this thesis.

3.1 Classical Signal Processing for ECG Denoising

Before the advent of data-driven approaches, ECG denoising was primarily addressed through deterministic algorithms rooted in signal theory. These approaches remain the industry standard for many embedded medical devices due to their determinism and low computational footprint.

3.1.1 Digital Filtering Techniques

The most straightforward approach involves linear frequency-selective filters, which assume that the signal of interest and the noise occupy distinct frequency bands.

- **Baseline Wander Removal:** High-pass filters are commonly employed to remove low-frequency drifts caused by respiration and patient movement. Finite Impulse Response (FIR) filters are often preferred for their linear phase response, which prevents phase distortion of the ECG waveform, although Infinite Impulse Response (IIR) filters are used when computational efficiency is paramount. The typical cut-off frequency is set between 0.5 Hz and 0.67 Hz to preserve the ST-segment information [3].
- **Power Line Interference Suppression:** To eliminate the 50/60 Hz mains hum, Notch filters are the standard solution. While effective for stationary

interference, fixed notch filters can introduce "ringing" artifacts (Gibbs phenomenon) around the sharp QRS complex if the bandwidth is too narrow, or attenuate physiological information if it is too wide.

- **Muscle Noise Reduction:** Low-pass filters are used to attenuate high-frequency Electromyographic (EMG) artifacts. However, this presents a significant challenge: the spectral content of EMG noise significantly overlaps with the high-frequency components of the QRS complex (typically up to 100-150 Hz). Consequently, aggressive low-pass filtering often results in the smoothing of the R-peak amplitude and widening of the QRS duration, potentially leading to diagnostic errors.

3.1.2 Adaptive Filtering

To address the non-stationarity of biomedical signals, Adaptive Filtering techniques were introduced. Algorithms such as Least Mean Squares (LMS) and Recursive Least Squares (RLS) iteratively adjust the filter coefficients to minimize the error between the noisy signal and a reference. While adaptive filters offer superior performance in dynamic environments compared to fixed filters, they suffer from a critical limitation: the requirement for a **reference channel**. This is a secondary signal that correlates with the noise but not with the ECG (e.g., a dedicated sensor for respiration or mains voltage). In standard single-lead ambulatory monitoring (e.g., Holter monitors or patches), such a pure noise reference is rarely available, limiting the practical applicability of this method [12].

3.1.3 Transform Domain Decomposition

A significant leap in denoising performance was achieved with transform domain methods, most notably the **Discrete Wavelet Transform (DWT)**. DWT decomposes the signal into multiple resolution levels, allowing for the separation of signal and noise based on their time-frequency characteristics. Denoising is typically performed via *Wavelet Shrinkage* (thresholding), where coefficients below a certain value are set to zero (Hard Thresholding) or attenuated (Soft Thresholding), under the assumption that noise is represented by small, high-frequency coefficients. Similarly, **Empirical Mode Decomposition (EMD)** adaptively decomposes the signal into Intrinsic Mode Functions (IMFs) without a predefined basis. Despite their efficacy, these methods introduce complexity in hyperparameter tuning. The choice of the "mother wavelet" (e.g., Daubechies, Symlet), the decomposition level, and the thresholding strategy requires expert domain knowledge and manual tuning, making them difficult to generalize across diverse patient populations without introducing artifacts [16].

3.2 Deep Learning Approaches: The Current State-of-the-Art

The availability of large-scale biomedical datasets, such as the MIT-BIH Arrhythmia Database [11], has fueled the widespread adoption of Deep Learning (DL) for ECG denoising. Unlike classical methods, neural networks can learn complex, non-linear mappings from noisy to clean signals directly from data, without relying on rigid mathematical assumptions about the noise spectrum.

3.2.1 Direct Denoising: End-to-End Reconstruction

The dominant paradigm in early DL applications was "Direct Denoising" or End-to-End Reconstruction. In this framework, a neural network (typically a Denoising Autoencoder or a U-Net) receives the noisy signal x as input and is trained to output the clean signal $\hat{s}$ directly.

- **Convolutional Autoencoders (CDAE):** Chiang et al. demonstrated that fully convolutional networks could effectively remove noise by learning a compressed latent representation of the ECG morphology [28]. These models leverage the translation invariance of convolutions to detect features like QRS complexes regardless of their temporal position.
- **Recurrent Neural Networks (RNNs):** Architectures such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU) have been employed to capture long-term temporal dependencies in the cardiac cycle.

While these models often achieve high Signal-to-Noise Ratio (SNR) scores, they place a heavy burden on the network: it must learn to reconstruct the precise morphology of the P-QRS-T complex while simultaneously suppressing noise. This often leads to **signal distortion**, such as the smoothing of high-frequency pathological markers (e.g., the Q-wave notch) or the alteration of R-peak amplitudes.

3.2.2 The Shift to Noise Estimation

To mitigate the risk of signal distortion, recent research has shifted focus towards the **Noise Estimation** paradigm (also known as Residual Learning). Instead of predicting the clean signal, the model is trained to estimate the additive noise component $\hat{n}$. The clean signal is then recovered by subtraction:

$$\hat{s} = x - \hat{n} \tag{3.1}$$

This approach relies on the observation that noise artifacts (such as sinusoidal PLI or low-frequency Baseline Wander) often possess a morphologically simpler structure than the ECG signal itself.

A significant contribution in this domain is the work by Gancitano (2024), who proposed a **Deep Multitask Noise Estimation** framework [29]. By utilizing a

multi-head Convolutional Neural Network (CNN), this approach estimates separate waveforms for Baseline Wander, Muscle Artifacts, and Power Line Interference. Gancitano demonstrated that separating the estimation into distinct auxiliary tasks yields superior results compared to direct denoising (+6.26 dB vs +2.22 dB in SNR improvement). The multitasking approach allows the encoder to learn shared features while specific decoders specialize in the unique spectral characteristics of each noise type.

3.2.3 Generative Models

Beyond deterministic regression, Generative Adversarial Networks (GANs) have been applied to ECG denoising. In this setup, a generator creates clean signals while a discriminator attempts to distinguish between real clean ECGs and denoised ones. While GANs produce visually realistic waveforms with sharp details, they suffer from a high risk of **hallucination**: the generator may invent plausible-looking features (e.g., a normal P-wave) that conceal actual pathologies (e.g., Atrial Fibrillation), rendering them dangerous for diagnostic automation.

3.2.4 The "Black-Box" Problem in Clinical AI

Despite the numerical superiority of DL models, their adoption in clinical practice faces a critical barrier: **Opacity**. Deep neural networks operate as "Black Boxes." A CNN with millions of floating-point parameters offers no insight into its decision-making process. When the model filters a signal, the clinician cannot verify *why* a specific segment was altered. Furthermore, noise estimation models based on CNNs can suffer from a specific type of failure mode: predicting and subtracting noise patterns even when the input signal is clean (ghosting). Without an interpretability layer to audit the model's internal state, it is impossible to distinguish between a correct noise removal and an erroneous artifact introduction. This lack of transparency necessitates the exploration of alternative frameworks that can offer comparable performance while ensuring the **trustworthiness** and **verifiability** required by medical regulations.

3.3 The Tsetlin Machine in Biomedical Applications

While DL currently dominates the landscape of biomedical signal processing, the Tsetlin Machine (TM) has recently emerged as a compelling alternative, particularly in high-stakes domains where transparency is mandatory. By replacing floating-point arithmetic with propositional logic, Tsetlin Machines offer a unique combination of high accuracy, low computational cost, and intrinsic interpretability. This section reviews key contributions that validate the use of TMs for physiological signal analysis.

3.3.1 Interpretable Classification of Arrhythmias

The efficacy of Tsetlin Machines in ECG analysis was rigorously demonstrated by Zhang et al. (2023) in their work on Premature Ventricular Contraction (PVC) identification [22]. In a comparative study using the MIT-BIH Arrhythmia Database, the authors trained a Convolutional Tsetlin Machine (CTM) to classify heartbeats. The results indicated that the TM achieved a classification accuracy (98.6%) comparable to state-of-the-art Deep Neural Networks (98.9%). However, the critical contribution of this work lay in the interpretability analysis. While CNNs relied on Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize vague "regions of interest," the TM provided explicit logic clauses. Specifically, the model learned rules that directly correlated the spatial distribution of pixels in the binary ECG image (e.g., width of the QRS complex) to the pathology. This allowed clinical experts to verify that the algorithm was making decisions based on established cardiological criteria, rather than spurious correlations.

3.3.2 Continuous Estimation via Regression Tsetlin Machines

While the standard TM is a classifier, signal denoising and quality assessment often require continuous outputs. To address this, Abeyrathna et al. introduced the Regression Tsetlin Machine (RTM) [31]. By mapping the summation of active clauses to a continuous output space, the RTM enables the modeling of non-linear functions using logic rules. This architecture has been successfully applied to various time-series forecasting tasks and, notably, to predicting Dengue outbreaks based on continuous environmental variables. In the context of 1-D signal processing, RTMs have demonstrated the ability to learn complex temporal dynamics while maintaining the "White-Box" property: the magnitude of the prediction can be decomposed into the contributions of specific active rules, providing a granular explanation of the output value [33].

3.3.3 Efficiency and Edge Computing Potential

Beyond accuracy and interpretability, Tsetlin Machines offer significant advantages in computational efficiency, a critical factor for wearable medical devices (e.g., Holter monitors). Research by Lei et al. on Audio Keyword Spotting demonstrated that TMs can achieve inference speeds and energy efficiency orders of magnitude superior to Neural Networks when deployed on low-power microcontrollers [32]. Since the inference phase relies exclusively on bitwise operations (AND, NOT, XOR, Population Count), TMs avoid the heavy floating-point calculations required by convolutions. This characteristic is particularly relevant for the development of "Edge AI" denoising systems capable of running directly on battery-powered sensors without cloud connectivity.

3.3.4 Intrinsic vs. Post-Hoc Interpretability

A recurring theme in the literature comparing TMs and DNNs is the distinction between intrinsic and post-hoc interpretability. As argued by Rudin [34], "Stop explaining black boxes" is a necessary paradigm shift for high-stakes AI. Post-hoc methods (like LIME or SHAP) approximate the decision boundary of a complex model, often introducing new errors or oversimplifications. In contrast, Tsetlin Machines are intrinsically interpretable: the learned clauses *are* the model. There is no approximation layer between the decision logic and the human observer. This property is essential for regulatory compliance in medical software (e.g., EU AI Act), where the ability to audit the decision-making process is a legal requirement.

3.4 Gap Analysis and Thesis Positioning

The review of the state-of-the-art reveals a clear dichotomy in the field of ECG denoising, characterized by a trade-off between performance and transparency. Table 3.1 summarizes the capabilities and limitations of the existing paradigms discussed in this chapter.

3.4.1 Synthesis of Current Limitations

1. **Classical Methods (Filters/Wavelets):** While computationally efficient and theoretically grounded, these methods lack the adaptability required to handle non-stationary and spectral-overlapping noise sources (such as EMG) without degrading the ECG morphology. They require manual tuning and cannot "learn" from data.

2. **Deep Learning (Direct & Noise Estimation):** Neural networks, particularly CNNs trained for noise estimation, have established new benchmarks for signal restoration quality. However, they suffer from the "Black-Box" problem. The lack of interpretability and the risk of generative hallucinations (inventing or suppressing features) present significant regulatory and safety hurdles for clinical deployment.

3. **Tsetlin Machines (Current State):** The TM framework has proven its capability in biomedical classification tasks (e.g., arrhythmia detection) and general regression problems. However, its application to *signal enhancement* and *continuous noise quantification* remains largely unexplored. To date, there is no established literature utilizing Regression Tsetlin Machines to drive adaptive filtering pipelines for ECG.

3.4.2 Thesis Contribution

This thesis aims to bridge the identified gap by proposing a novel hybrid architecture. We adopt the effective **Noise Estimation** paradigm (learning to identify noise rather

Table 3.1: Comparison of ECG Denoising Paradigms vs. Proposed Method.

Paradigm	**Adaptability**	**Interpretability**	**Safety (Hallucination Risk)**
Classical Filters	Low	High (Analytic)	High (Deterministic)
Deep Learning (CNNs)	High	Low (Black-Box)	Low (Generative Artifacts)
Proposed (TMU-RTM)	**High**	**High (Logic Rules)**	**High (Gated Control)**

than the clean signal) but replace the opaque CNN backbone with an interpretable **Regression Tsetlin Machine**.

Unlike end-to-end Deep Learning approaches that blindly reconstruct the waveform, our proposed system estimates the **intensity** of specific noise components (Baseline Wander, EMG, PLI) using transparent logical rules. These intensity estimates are then used to modulate standard digital filters. This "Grey-Box" approach combines the data-driven adaptability of Machine Learning with the safety and determinism of classical signal processing. Specifically, this work contributes to the field by:

- Demonstrating the first application of Regression Tsetlin Machines for ECG noise intensity estimation.
- Developing an explainability-driven feature engineering pipeline that reduces computational complexity by 95% compared to brute-force approaches.
- Validating a "Safety-by-Design" denoising controller that prevents signal distortion in clean segments through logic-based gating.

Chapter 4

Methodology

This chapter presents the architectural and methodological framework developed in this thesis for interpretable ECG denoising. The proposed system, named *TMU-Optimized*, represents a paradigm shift from conventional black-box restoration to a logic-based, white-box estimation of noise intensity.

Section 4.1 provides a high-level overview of the multi-stage pipeline. Section 4.2 details the construction of the dataset and the synthetic mixing protocol used to simulate realistic clinical artifacts. Section 4.3 describes the novel feature engineering process, mathematically defining the extraction of bitplane and spectral features and detailing the explainability-driven optimization that led to the final feature set. Section 4.4 formalizes the multi-head Regression Tsetlin Machine (RTM) architecture, including hyperparameter tuning and the handling of specific noise distributions like Power Line Interference. Section 4.5 introduces the crucial stages of isotonic calibration and adaptive filtering control. Finally, Section 4.6 outlines the rigorous statistical protocols adopted to validate the system's robustness and clinical reliability.

4.1 System Architecture

The architecture proposed in this study, called *TMU-Optimized*, implements a modular pipeline for ECG signal restoration. The fundamental design philosophy decouples the problem of *noise quantification* from the problem of *noise removal*. Unlike end-to-end deep learning architectures that perform a direct mapping from noisy to clean waveforms using millions of parameters, the proposed system operates as a transparent supervisory controller. It utilizes a logic-based regression model to estimate the intensity of specific artifacts and subsequently modulates standard digital filters based on these estimates.

4.1.1 Pipeline Overview

The processing flow is structured into five sequential stages, transforming raw physiological signals into clean diagnostic data through a series of interpretable transformations. The pipeline operates on single-lead ECG signals sampled at 360 Hz.

1. **Signal Segmentation:** The continuous ECG stream is divided into overlapping windows. A window size of $N = 1024$ samples (approximately 2.84 seconds) is used to ensure local stationarity while providing sufficient spectral resolution for low-frequency analysis.

2. **Feature Extraction and Encoding:** Instead of feeding raw samples to the model, a compact set of 57 hybrid features is extracted. These features combine spectral power densities and morphological bitplane statistics. The continuous feature values are then discretized into Boolean literals using thermometer encoding, making them compatible with the propositional logic of the Tsetlin Machine.

3. **Intensity Regression (Multi-Head TMU):** The core inference engine consists of three independent Regression Tsetlin Machines (RTMs). Each "head" is specialized to predict the intensity scalar $\hat{y} \in [0, 1]$ for a specific noise type: Baseline Wander (BW), Muscle Artifacts (EMG), and Power Line Interference (PLI).

4. **Isotonic Calibration:** The raw outputs from the Tsetlin Machines are refined using a non-parametric isotonic regression model. This stage corrects non-linear biases and ensures that the predicted intensities map linearly to the probability of noise presence.

5. **Adaptive Filtering:** The calibrated intensities act as control signals for a bank of digital filters. A gating logic enables the filters only when the estimated noise exceeds a safety threshold, modulating the filter strength (e.g. cut-off frequency or attenuation level) proportionally to the detected intensity.

Figure 4.1 provides a schematic representation of this data flow, highlighting the transition from the time domain to the logic domain and back to the signal domain.

4.1.2 The Noise Intensity Paradigm

The most significant architectural divergence from the state-of-the-art lies in the definition of the learning target. Contemporary Deep Learning approaches, such as the Multitask CNN proposed by Gancitano [29], typically frame the problem as **Waveform Estimation**. In that paradigm, the model outputs a high-dimensional vector (N samples) representing the noise signal itself, which is then subtracted from the input.

In contrast, this work adopts a **Noise Intensity Estimation** paradigm. The model predicts a single scalar value representing the global level of contamination within the window. Table 4.1 contrasts these two approaches. The intensity-based approach offers three critical advantages for clinical safety:

- **Elimination of Hallucinations:** Generative waveform estimators can introduce artifacts (hallucinations) by predicting noise patterns where none exist,

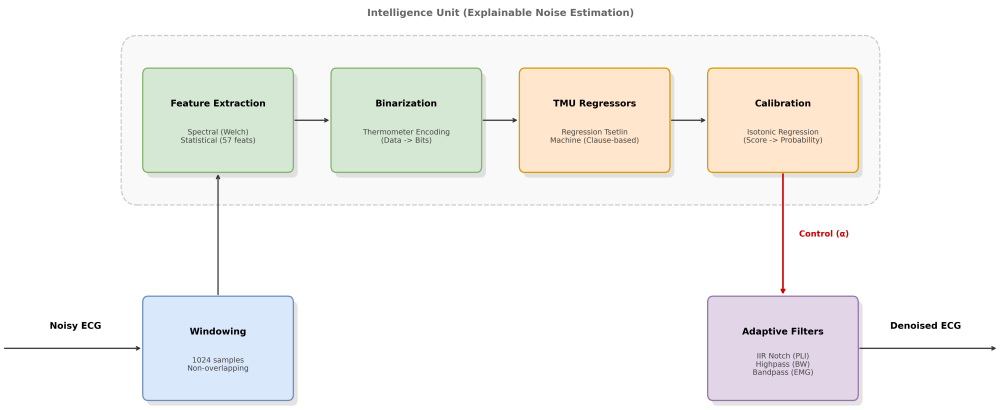


Figure 4.1: Block diagram of the proposed TMU-Optimized pipeline. The system decouples the estimation path (Intelligence Unit, top) from the filtering path (Signal Processing Unit, bottom). The Tsetlin Machine estimates noise intensities which modulate the adaptive filters via the control signal α.

which are then subtracted from the signal, creating distortion. By predicting a scalar intensity, the proposed system implements a "gate": if the predicted intensity is near zero, the filter is bypassed entirely, preserving the original signal bit-perfectly.

- **Simplified Learning Task:** Mapping a complex input to a single scalar (Regression) is computationally less demanding than mapping it to a high-dimensional vector (Generation). This allows for the use of lighter models like the Tsetlin Machine.
- **Deterministic Filtering:** The actual removal of noise is performed by standard DSP filters (e.g., Butterworth, Wavelet Shrinkage) whose frequency responses and phase characteristics are mathematically deterministic and well-understood by biomedical engineers, unlike the opaque non-linear transformations of a neural network.

Table 4.1: Comparison of Denoising Paradigms: Waveform Estimation vs. Intensity Estimation.

Aspect	**Neural CNN (State-of-the-Art)**	**TMU-Optimized (Proposed)**
Target Output	Noise Waveform ($\mathbb{R}^{1024}$)	Noise Intensity ($\mathbb{R}^1 \in [0, 1]$)
Denoising Method	Direct Subtraction	Adaptive Filtering
Hallucination Risk	High (Generative artifacts)	None (Gated logic)
Model Complexity	~121,000 params (Float32)	9,000 Clauses (Binary)
Interpretability	Black-Box	White-Box (Logic Rules)

4.2 Data Preparation and Noise Mixing

The development of a supervised regression model requires a dataset with precise ground truth labels. In the context of clinical ECGs, obtaining a perfect separation between the clean physiological signal and the noise artifacts is impossible. To overcome this limitation, this study adopts a data synthesis approach based on the additive mixing of clean signals and calibrated noise sources. This section details the source databases, the mixing protocol, and the partitioning strategy employed to ensure rigorous evaluation.

4.2.1 Source Databases

Two standard repositories from PhysioNet were utilized to construct the dataset:

- **Clean ECG Source:** The **MIT-BIH Arrhythmia Database** [35] serves as the source of physiological signals. It contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. This dataset provides a wide variety of pathological morphologies (including ventricular arrhythmias and conduction abnormalities), ensuring that the denoising model learns to distinguish noise from genuine cardiac anomalies.
- **Noise Source:** The **MIT-BIH Noise Stress Test Database (NSTDB)** [36] provides the artifact templates. It contains calibrated recordings of three specific noise types: Baseline Wander (BW), caused by patient respiration; Muscle Artifact (EMG), caused by muscle contraction; and Electrode Motion artifacts. For Power Line Interference (PLI), synthetic sinusoidal noise at 50 Hz (with harmonics) was generated to simulate mains hum.

4.2.2 Synthetic Mixing Protocol

To simulate realistic ambulatory conditions, a dynamic mixing process was implemented. For each window of clean ECG, noise is added according to a target Signal-to-Noise Ratio (SNR). Let $x_{clean}[n]$ be a segment of the clean signal and $\eta_k[n]$ be a noise source of type $k \in \{\text{BW, EMG, PLI}\}$. The contaminated signal $x_{noisy}[n]$ is generated as:

$$x_{noisy}[n] = x_{clean}[n] + \sum_k \alpha_k \cdot \frac{\eta_k[n]}{||\eta_k||_2} \cdot ||x_{clean}||_2 \tag{4.1}$$

The scaling factor α_k is determined stochastically to cover a wide range of signal qualities, spanning from -6 dB (severe contamination, where noise power exceeds signal power) to +24 dB (clean signal). The **ground truth targe**t y_k for the regression task is defined not as the SNR, which is unbound, but as the **fractional**

noise power. This maps the noise intensity to a bounded interval $[0, 1]$, which is ideal for the Tsetlin Machine regression:

$$y_k = \frac{P_{\eta_k}}{P_{total}} = \frac{\sum(\alpha_k \eta_k[n])^2}{\sum(x_{noisy}[n])^2} \tag{4.2}$$

4.2.3 Segmentation and Data Partitioning

The continuous signals are segmented into overlapping windows. A window size of $N = 1024$ samples was selected, corresponding to approximately 2.84 seconds at 360 Hz. This duration is sufficient to capture multiple QRS complexes (for spectral resolution) while maintaining local stationarity. An overlap of 512 samples (50%) is used to increase the volume of training data.

A crucial methodological choice in this work is the **Subject-Wise Splitting**. Randomly shuffling segments from all patients into training and test sets would lead to data leakage, as the model could memorize the specific ECG morphology of a patient seen during training. To prevent this, the 48 records were partitioned at the patient level:

- **Training Set (70%):** Segments from 34 patients ($N = 258,672$ windows).
- **Validation Set (15%):** Segments from 7 patients ($N = 53,256$ windows). Used for hyperparameter tuning and calibration fitting.
- **Test Set (15%):** Segments from 7 patients ($N = 53,256$ windows). Used exclusively for the final performance reporting.

4.2.4 Noise Distribution Analysis

The resulting dataset exhibits a class imbalance that reflects real-world monitoring conditions, where severe artifacts are intermittent. Table 4.2 summarizes the distribution of noise intensities in the test set. Notably, the EMG noise is heavily skewed towards high intensities (58.0% of samples have $y_{EMG} \geq 0.5$), while Power Line Interference (PLI) exhibits a bimodal distribution: it is either negligible (< 0.001) or significant, with very few samples in the intermediate range. This distribution characteristic poses a specific challenge for regression models, particularly for the PLI head, as discussed in the results chapter.

Table 4.2: Distribution of Noise Intensities in the Test Set ($N = 53,256$).

Noise Type	**Zero** (< 0.001)	**Low** ($0.001 - 0.1$)	**Medium** ($0.1 - 0.5$)	**High** (≥ 0.5)
EMG	1.6%	25.9%	14.4%	**58.0%**
BW	16.4%	35.5%	13.2%	34.9%
PLI	43.6%	27.3%	0.0%	29.1%

Note: The PLI distribution is notably bimodal, lacking intermediate values.

4.3 Explainability-Driven Feature Engineering

A critical challenge in applying Tsetlin Machines to raw signal processing is the conversion of continuous time-series data into a Boolean format suitable for propositional logic. While the standard approach often involves generating massive amounts of binary descriptors (e.g., raw bitplanes of every sample), this leads to computational inefficiency and model bloating. In this work, feature engineering was not treated as a static preprocessing step but as an iterative optimization process guided by the model's intrinsic interpretability.

4.3.1 Feature Domains

Two complementary domains were explored to capture the distinct characteristics of ECG noise artifacts: the frequency domain for stochastic noise (EMG) and interference (PLI), and the amplitude domain for morphological distortions.

4.3.1.1 Spectral Features (Frequency Domain)

Given the spectral separation between physiological ECG components (mostly $<$ 40 Hz) and muscle noise (broadband high-frequency), spectral power is a highly discriminative feature. The Power Spectral Density (PSD) is estimated using Welch's method with a Hamming window of 256 samples and 50% overlap. From the PSD, **12 spectral features** are extracted, representing the normalized power in specific bands:

- **PLI Band ($45-55$ Hz):** Targets the mains hum.
- **Low EMG ($20-40$ Hz):** Overlaps with QRS but captures lower-frequency muscle artifacts.
- **Mid EMG ($40-80$ Hz) & High EMG ($80-150$ Hz):** Capture the dominant energy of skeletal muscle contraction, largely devoid of cardiac signal.

For each band, three statistics are computed: Mean Power, Standard Deviation of Power, and Maximum Power.

4.3.1.2 Bitplane Features (Amplitude Domain)

Bitplane slicing decomposes the integer representation of the signal amplitude into binary levels. For an N-bit signal x, the k-th bitplane is the binary sequence formed by the k-th bit of each sample. Rather than using the raw bit sequences (which would result in thousands of inputs), we compute statistical summaries (Mean, Variance, Skewness, Kurtosis, Signal Energy) for **9 selected bitplanes**. These features capture the "roughness" or texture of the signal, which correlates with the presence of high-frequency noise superimposed on the smoother ECG waveform. This yields **45 bitplane features**.

4.3.2 Optimization via Global Feature Importance

The final feature set consisting of 57 features (12 Spectral + 45 Bitplane) is the result of a rigorous pruning process. An initial baseline model was trained using a brute-force approach with over 1200 features. By leveraging the Tsetlin Machine's global feature importance analysis (based on literal usage frequency in active clauses), a distinct hierarchy emerged:

- **Dominance of Spectral Features:** The high-frequency spectral bands were found to be 15-21 times more important than bitplane features for EMG estimation.
- **Redundancy:** The vast majority of raw amplitude features contributed negligible information to the regression task.

This insight allowed for a **95% reduction in the feature space** (from > 1200 to 57), which paradoxically resulted in a $+4.2\%$ improvement in correlation performance. This demonstrates that removing irrelevant literals simplifies the solution space, allowing the automata to converge on more robust logic rules.

4.3.3 The "Zero-Variance" Phenomenon

A specific ablation study was conducted to test the removal of 6 features that exhibited zero variance across the entire training set. Surprisingly, removing these ostensibly useless features resulted in a performance degradation on the test set (-4% for EMG, -19% for BW). This counter-intuitive finding suggests that "constant" features may serve as necessary anchors for the logical clauses (e.g., serving as a default 'TRUE' condition in an AND-chain). Consequently, the final model retains all 57 features, prioritizing robustness over minimal theoretical compression.

4.3.4 Binarization Strategy: Thermometer Encoding

To translate these continuous features into Boolean literals, **Thermometer Encoding** is employed. This technique preserves the ordinal relationship of magnitude. For a normalized feature value $x \in [0, 1]$ and a resolution B (number of bins), the feature is converted into a bit vector $\mathbf{b} = [b_1, \ldots, b_B]$ where:

$$b_k = \begin{cases} 1 & \text{if } x \geq \frac{k}{B} \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

Through hyperparameter tuning, a resolution of $B = 60$ **bins** per feature (resolution $r = 0.6$) was found to be optimal. This allows the model to learn fine-grained threshold rules such as *"IF Feature* $X > 0.75$ *THEN..."*. The total input size for the Tsetlin Machine is thus:

$$\text{Input Bits} = 57 \text{ features} \times 60 \text{ bins} = 3420 \text{ bits}.$$

4.4 Regression Tsetlin Machine Architecture

The core predictive engine of the proposed pipeline is the Regression Tsetlin Machine (RTM). Unlike standard classification TMs which vote for a class label, the RTM is designed to produce a continuous scalar output representing the intensity of the noise. To handle the distinct characteristics of different artifact types, a decoupled multi-head architecture was implemented.

4.4.1 Multi-Head Configuration

Rather than training a single monolithic model to predict all noise types simultaneously, the system utilizes three independent RTM estimators, referred to as "heads". Each head shares the same input feature vector (the 3420 bits derived in Section 4.3) but maintains its own independent set of clauses and internal automata states.

4.4.1.1 The EMG and BW Regressors

The first two heads are specialized for Muscle Artifacts (EMG) and Baseline Wander (BW). These noise types exhibit continuous variations in intensity, making them ideal candidates for the linear regression logic of the RTM. The EMG head learns to associate high-frequency spectral power with high noise intensity, while the BW head focuses on low-frequency bitplane variations.

4.4.1.2 The PLI Regressor and Bimodality

The third head targets Power Line Interference (PLI). Analysis of the dataset revealed that PLI intensity follows a strongly bimodal distribution: the interference is either negligible or present at a significant magnitude, with very few intermediate states. Consequently, the PLI regression task functions effectively as a soft binary detector. Although the RTM architecture remains identical to the other heads, the learning dynamics adapt to this distribution, producing outputs that tend to cluster around 0 or 1.

4.4.2 Hyperparameter Specification

The configuration of the Tsetlin Machine involves three critical hyperparameters: the number of clauses (C), the voting threshold (T), and the specificity (s). To identify the optimal configuration, a grid search was conducted on the validation set. The search space explored clause counts ranging from 1,000 to 4,000, thresholds from 500 to 1,500, and specificity values from 2.0 to 10.0.

The final selected configuration, termed *TMU-Optimized*, utilizes **3,000 clauses** per head. This capacity was found sufficient to capture the diversity of noise patterns without inducing overfitting. A **Threshold** $T = 700$ was selected. In the context of regression, T acts as a scaling factor for the summation of votes; a higher T allows for a finer resolution in the output range $[0, 1]$. Finally, a **Specificity** $s = 7.0$ was

chosen. This relatively high value (compared to standard classification tasks where $s \approx 3.9$) imposes a strong penalty on false positive literal inclusions. This forces the model to learn highly precise rules, ensuring that clauses only activate when there is strong evidence of specific noise features, aligning with the "safety-by-design" philosophy of the pipeline.

Table 4.3 summarizes the final architectural parameters.

Table 4.3: Hyperparameter configuration for the TMU-Optimized architecture.

Parameter	Value	Description
Number of Heads	3	Independent models for BW, EMG, PLI
Clauses per Head	3,000	Total system capacity: 9,000 clauses
Threshold (T)	700	Voting margin and scaling factor
Specificity (s)	7.0	Pattern matching precision
Max Literals	114	Maximum literals per clause (2×57 features)

4.4.3 Training Dynamics

The models were trained using the coordinate descent-based learning scheme inherent to Tsetlin Machines. The training process was governed by the Mean Absolute Error (MAE) on the validation set. An **Early Stopping** mechanism was implemented with a patience of 3 epochs. This proved essential for computational efficiency; the EMG regressor typically converged within 4 epochs, while the BW regressor converged within 3 epochs. The total training time for the complete system on a standard CPU was approximately 5.5 minutes, highlighting the computational efficiency of the approach compared to the hours typically required for DNNs. The underlying implementation utilizes the `PyTsetlin` library, which optimizes bitwise operations for parallel execution on multi-core CPUs.

4.5 Post-Processing and Adaptive Control

The raw output of the Regression Tsetlin Machine represents a normalized vote count. While this correlates with noise intensity, strictly interpreting it as a probability or a physical magnitude can be misleading due to potential non-linearities in the feature-to-output mapping. To ensure reliable downstream control, the pipeline incorporates a calibration stage followed by a deterministic control logic that modulates the digital filters.

4.5.1 Isotonic Calibration

During the preliminary analysis of the validation set, a systematic bias was observed in the raw predictions of the RTM. Specifically, the Baseline Wander regressor exhibited a tendency to overestimate the noise intensity, characterized by a regression slope significantly greater than unity (slope ≈ 2.06). This behavior implies that a

moderate level of baseline drift could be misinterpreted as severe contamination, potentially triggering aggressive filtering that could distort the ST-segment.

To rectify this miscalibration without making strong parametric assumptions about the error distribution, **Isotonic Regression** was employed. This non-parametric technique fits a free-form line to the sequence of observations under the constraint that the fitted line must be non-decreasing. Let $\hat{y}_{raw}$ be the raw output from the Tsetlin Machine and y be the ground truth intensity. The isotonic calibrator learns a mapping function $f : [0, 1] \to [0, 1]$ by solving the following optimization problem on the validation set:

$$\min_f \sum_i (y_i - f(\hat{y}_{raw,i}))^2 \quad \text{subject to } \hat{y}_{raw,i} \leq \hat{y}_{raw,j} \implies f(\hat{y}_{raw,i}) \leq f(\hat{y}_{raw,j}) \tag{4.4}$$

The resulting calibrated output, $\hat{y}_{cal} = f(\hat{y}_{raw})$, retains the rank order of the original predictions but adjusts their magnitude to match the empirical probability of the target. As detailed in the experimental results, this step proved critical for system performance, correcting the slope of the BW regressor to near-unity and reducing the Expected Calibration Error (ECE) of the EMG regressor by over 88%.

4.5.2 The Gating Mechanism

A fundamental requirement for clinical acceptance is that the denoising system must not alter clean signals. To enforce this "do no harm" principle, a non-linear **gating mechanism** is placed between the estimator and the filters. The calibrated intensity $\hat{y}_{cal}$ acts as a control signal α for the filters. To prevent micro-adjustments due to residual regressor noise when the signal is clean, a hard threshold (gate) is applied:

$$\alpha_k = \begin{cases} 0 & \text{if } \hat{y}_{cal,k} < \tau_{gate} \\ \hat{y}_{cal,k} & \text{if } \hat{y}_{cal,k} \geq \tau_{gate} \end{cases} \tag{4.5}$$

In this implementation, the gate threshold τ_{gate} is set to 0.05. This ensures that if the estimated noise intensity is below 5%, the filter is strictly bypassed, preserving the original bit-exact morphology of the physiological signal. This mechanism effectively neutralizes the risk of "hallucinating" artifacts in clean segments, a common failure mode in end-to-end neural networks.

4.5.3 Adaptive Filtering Modules

Once the control signal α_k is determined for each noise type $k \in \{BW, EMG, PLI\}$, it modulates the parameters of standard digital filters. This hybrid approach leverages the transparency of classical Digital Signal Processing (DSP) while utilizing the Tsetlin Machine to determine the optimal operating point dynamically.

4.5.3.1 Baseline Wander Removal

Baseline wander is removed using a High-Pass Infinite Impulse Response (IIR) Butterworth filter. The control signal α_{BW} modulates the cut-off frequency f_c. When $\alpha_{BW} = 0$ (clean signal), the filter is bypassed. As the detected wander intensity increases, the cut-off frequency is shifted upwards, up to a maximum of 0.67 Hz. This allows for aggressive removal of drift during periods of high activity while preserving low-frequency components (such as the ST-segment) during stable periods.

4.5.3.2 Muscle Artifact Denoising

Electromyographic noise is broadband and overlaps with the QRS complex. To mitigate this without blurring the R-peaks, a Wavelet Denoising approach is adopted. The signal is decomposed using the `bior2.6` wavelet basis. The control signal α_{EMG} modulates the threshold λ used for wavelet shrinkage. The adaptive threshold is defined as:

$$\lambda(\alpha_{EMG}) = \sigma\sqrt{2\log N} \cdot (0.5 + 1.5 \cdot \alpha_{EMG}) \tag{4.6}$$

where σ is the estimated noise variance and N is the window length. This formulation scales the aggressiveness of the denoising linearly with the estimated EMG intensity. When muscle noise is severe ($\alpha_{EMG} \to 1$), the threshold increases, suppressing more coefficients; when noise is mild, the threshold relaxes, preserving signal detail.

4.5.3.3 Power Line Interference Suppression

For Power Line Interference, an IIR Notch filter centered at the mains frequency (50 Hz or 60 Hz) is employed. Since the PLI regressor functions primarily as a detector due to the bimodal distribution of the noise, the control logic here is binary. If $\alpha_{PLI} > \tau_{gate}$, the notch filter is engaged; otherwise, it is bypassed. This prevents the "ringing" artifacts (Gibbs phenomenon) associated with notch filters from affecting the signal when no mains hum is present.

4.6 Statistical Validation Framework

To ensure that the proposed method is robust enough for clinical application and does not merely overfit the training distribution, a comprehensive statistical validation framework was designed. This framework goes beyond standard aggregate metrics to assess patient-specific consistency, estimator reliability, and clinical agreement.

4.6.1 Cross-Subject Robustness Analysis

Standard performance metrics (such as global Mean Absolute Error) can mask poor performance on specific patient subgroups. To verify the generalization capability of the model across diverse physiological morphologies, a **Per-Subject Analysis** was implemented. The regression performance (Pearson correlation coefficient r

and MAE) is computed individually for each of the 48 subjects in the test set. The distribution of these metrics is then analyzed to calculate the **Coefficient of Variation (CV)**:

$$CV = \frac{\sigma_{\text{metrics}}}{\mu_{\text{metrics}}} \times 100\% \tag{4.7}$$

A low CV indicates that the model performs consistently regardless of the specific patient, a critical requirement for medical device certification.

4.6.2 Bootstrap Confidence Intervals

To quantify the uncertainty associated with the performance point estimates, non-parametric **Bootstrap Resampling** was employed. The test set predictions are resampled with replacement $K = 1000$ times. For each iteration, the performance metrics are recalculated. This process yields an empirical distribution for each metric, from which the **95% Confidence Intervals (CI)** are derived using the percentile method (2.5th and 97.5th percentiles). This rigorous approach confirms that reported improvements over baselines are statistically significant and not artifacts of sample selection.

4.6.3 Agreement Analysis: Bland-Altman Method

To assess the interchangeability between the Tsetlin Machine's estimated noise intensity ($\hat{y}$) and the synthetic ground truth (y), the **Bland-Altman Analysis** is utilized. This method plots the difference between the two measurements ($\hat{y} - y$) against their average ($(\hat{y} + y)/2$). Key statistics derived from this analysis include:

- **Systematic Bias:** The mean difference $\bar{d}$, indicating if the model systematically over- or under-estimates noise.
- **Limits of Agreement (LoA):** Defined as $\bar{d} \pm 1.96 \cdot SD_{diff}$, where SD_{diff} is the standard deviation of the differences.

In the context of adaptive filtering, this analysis validates whether the estimation error is within clinically acceptable tolerances to drive the filter control logic without introducing artifacts.

4.6.4 Evaluation Metrics

The performance of the system is quantified using two sets of metrics.

4.6.4.0.1 Regression Metrics:

To evaluate the noise intensity estimation:

- **Pearson Correlation (r):** Measures the linear relationship between predicted and true intensity.
- **Mean Absolute Error (MAE):** Measures the average magnitude of errors.
- **Expected Calibration Error (ECE):** Quantifies how well the predicted probabilities align with actual outcomes.

4.6.4.0.2 Denoising Metrics: To evaluate the final signal quality:

- **Signal-to-Noise Ratio Improvement (SNR_{imp}):** The difference between the output SNR and the input SNR (in dB).
- **Root Mean Square Error (RMSE):** The Euclidean distance between the denoised signal and the clean ground truth.

Chapter 5

Experimental Results

This chapter presents the empirical evaluation of the proposed interpretable denoising system. The experimental campaign was conducted on the MIT-BIH Arrhythmia Database to validate the methodology described in Chapter 4 across multiple dimensions: predictive accuracy, statistical robustness, signal restoration quality, and computational efficiency.

The presentation of the results follows a logical progression from the internal model performance to the final clinical application. Section 5.1 first assesses the precision of the Regression Tsetlin Machine in estimating noise intensity, quantitatively demonstrating the impact of isotonic calibration. Section 5.2 provides a rigorous statistical validation, examining cross-subject consistency and agreement limits to ensure the reliability of the method across a diverse patient population. Subsequently, Section 5.3 evaluates the downstream impact on ECG signal quality, comparing the proposed approach against classical filters and state-of-the-art neural networks in terms of Signal-to-Noise Ratio improvement and morphological preservation.

To demonstrate the unique advantages of the Tsetlin Machine, Section 5.4 analyzes the learned features and logic rules, providing a transparent view of the decision-making process. Finally, Section 5.5 justifies the architectural choices through ablation studies, and Section 5.6 reports the computational benchmarks, highlighting the suitability of the system for edge deployment.

5.1 Noise Intensity Estimation Performance

The first stage of the experimental evaluation assesses the capability of the Regression Tsetlin Machine (RTM) to correctly quantify the intensity of noise artifacts. This intermediate step is critical: since the denoising pipeline relies on adaptive filtering, the quality of the final signal depends linearly on the accuracy of these intensity estimates ($\hat{y} \in [0, 1]$). The performance was evaluated on the held-out test set ($N = 53,256$ segments) using Pearson correlation coefficient (r), Mean Absolute Error (MAE), and Mean Squared Error (MSE).

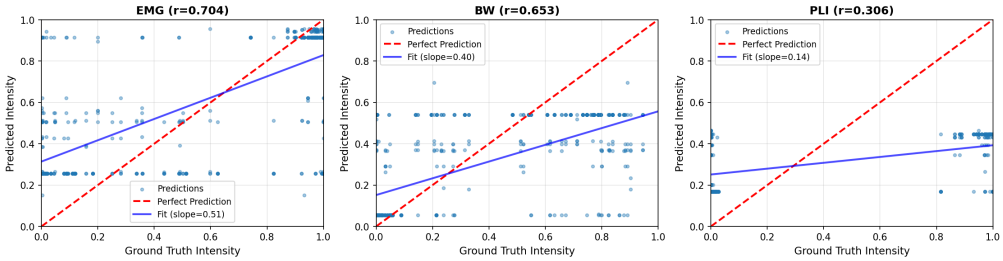


Figure 5.1: Scatter plots of Predicted vs. Ground Truth noise intensity for EMG, BW, and PLI. The red dashed line represents perfect prediction ($y = \hat{y}$), while the blue line indicates the linear fit. Note the bimodal clustering in the PLI plot.

5.1.1 Regressor Accuracy on Test Set

Table 5.1 summarizes the performance metrics for the three independent estimators: Muscle Artifacts (EMG), Baseline Wander (BW), and Power Line Interference (PLI).

Table 5.1: TMU Regressor Performance on Test Set (After Isotonic Calibration).

Regressor	Pearson r	MSE	MAE	ECE
EMG	0.7041	0.1450	0.2140	0.0368
BW	0.6529	0.1092	0.1905	-
PLI	0.3061	0.1700	0.3639	-

ECE = Expected Calibration Error (lower is better)

The EMG regressor achieved the highest correlation ($r = 0.7041$), demonstrating the model's ability to track the high-frequency stochastic nature of muscle noise using spectral features. The Baseline Wander regressor also showed strong performance ($r = 0.6529$), effectively utilizing bitplane statistics to detect low-frequency drifts.

The performance of the PLI regressor ($r = 0.3061$) requires a specific interpretation. Unlike EMG and BW, which exhibit a continuous range of intensities, the Power Line Interference in the dataset follows a strictly bimodal distribution: it is either absent or present at significant levels. Consequently, the regression task collapses into a quasi-binary detection problem. While the linear correlation metric penalizes this behavior, the model effectively functions as a binary detector, which is sufficient for triggering the notch filter in the downstream stage.

Figure 5.1 visualizes the relationship between the ground truth and the predicted intensities. Ideally, points should align along the diagonal $y = \hat{y}$.

5.1.2 Impact of Isotonic Calibration

A central finding of this study is the necessity of post-processing calibration for logic-based regression. The raw outputs from the Tsetlin Machine, being summations of clause votes, exhibited non-linear biases before calibration. As detailed in Table 5.2, the application of Isotonic Regression yielded significant improvements in linearity and reliability.

For Baseline Wander, the raw predictions systematically overestimated the noise

Table 5.2: Impact of Isotonic Calibration on Model Performance.

Regressor	Metric	Raw	Calibrated	Δ	Δ%
EMG	Pearson r	0.6716	0.7041	+0.0325	+4.8%
	MAE	0.3512	0.2140	-0.1372	-39.1%
	Slope	1.15	1.00	-0.15	-13.0%
	ECE	0.317	0.037	-0.280	**-88.4%**
BW	Pearson r	0.5418	0.6529	+0.1111	**+20.5%**
	MAE	0.2493	0.1905	-0.0588	-23.6%
	Slope	2.06	1.00	-1.06	-51.5%

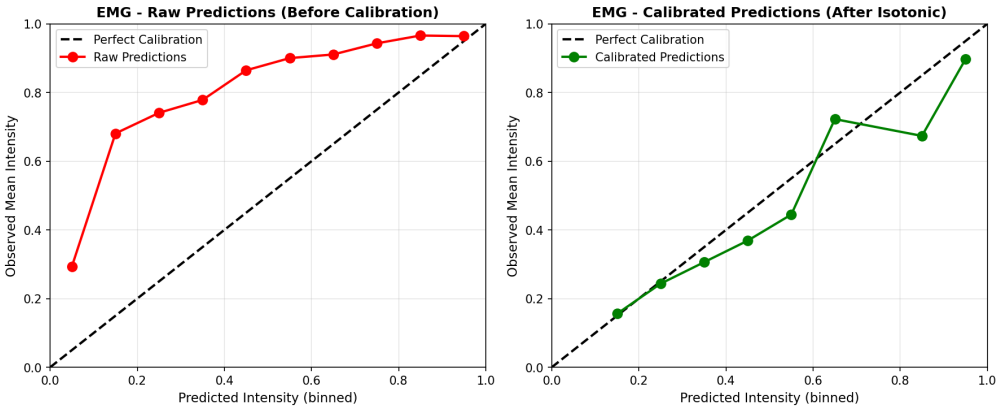


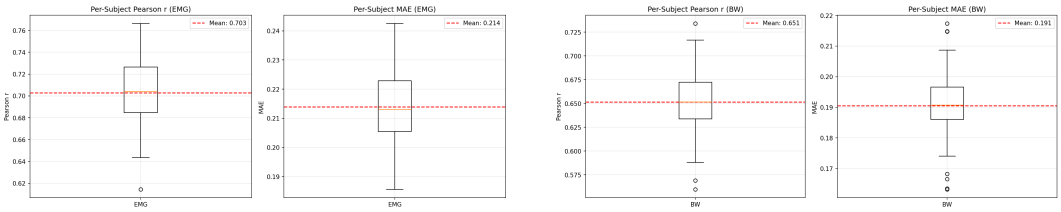
Figure 5.2: Reliability diagrams for EMG predictions before (left) and after (right) isotonic calibration. The calibrated model aligns with the ideal diagonal, confirming that the predicted intensities accurately reflect the observed noise levels.

intensity by a factor of two (Slope ≈ 2.06). Calibration corrected this slope to unity (1.00), resulting in a **20.5% improvement** in Pearson correlation. For EMG, the most significant impact was on the Expected Calibration Error (ECE), which decreased by **88.4%** (from 0.317 to 0.037). This reduction implies that the calibrated output can be interpreted as a reliable probability of noise presence, which is essential for the threshold-based gating logic used in the denoising controller.

Figure 5.2 illustrates the reliability curves. The raw predictions (left panel) deviate significantly from the diagonal, indicating miscalibration, whereas the calibrated predictions (right panel) show excellent alignment.

5.2 Statistical Validation and Robustness

In biomedical applications, reporting average performance metrics is insufficient to guarantee clinical utility. A robust model must demonstrate consistent behavior across the diverse population of patient morphologies and exhibit statistical reliability. To rigorously validate the proposed TMU-Optimized architecture, a three-tiered statistical analysis was conducted: Per-Subject consistency analysis, Bootstrap reliability estimation, and Bland-Altman agreement analysis.



(a) EMG Subject Variability **(b)** BW Subject Variability

Figure 5.3: Boxplots of Pearson correlation coefficients calculated per subject. The narrow spread indicates that the model generalizes well across different patients.

5.2.1 Cross-Subject Consistency

To assess the generalization capability of the model, performance metrics were computed individually for each of the 48 subjects included in the test set. This analysis aims to verify that the model relies on universal features of the noise artifacts rather than overfitting to patient-specific ECG characteristics.

The results, summarized in Table 5.3, indicate a high degree of consistency. The Pearson correlation coefficient for EMG estimation yielded a mean of 0.703 with a standard deviation of just 0.032. Notably, the **Coefficient of Variation (CV)** is below 6% for both EMG (4.6%) and Baseline Wander (5.5%). Figure 5.3 presents the distribution of the correlation coefficients. The compact interquartile ranges (IQR) and the absence of extreme outliers confirm that the regression performance is stable across the entire patient cohort, regardless of anatomical or physiological variations.

Table 5.3: Per-Subject Performance Analysis (N=48 patients).

Target	**Mean r**	**Std. Dev**	**Range [Min, Max]**	**CV (%)**
EMG	0.703	0.032	[0.614, 0.766]	**4.6%**
BW	0.651	0.036	[0.560, 0.734]	**5.5%**

5.2.2 Reliability Analysis via Bootstrap

To provide a probabilistic bound on the performance estimates, non-parametric bootstrap resampling was performed with $K = 1000$ iterations. This technique allows for the estimation of the sampling distribution of the metrics and the calculation of 95% Confidence Intervals (CI).

As shown in Table 5.4, the confidence intervals are extremely narrow. For EMG prediction, the 95% CI for the Pearson correlation is $[0.698, 0.710]$, with a width of only 0.011. This tight bound provides strong statistical evidence that the reported performance is intrinsic to the model architecture and not an artifact of the specific test set composition.

Table 5.4: Bootstrap 95% Confidence Intervals (1000 resamples).

Metric	Mean	Lower CI	Upper CI	Width
EMG Regressor				
Pearson r	0.7041	0.6984	0.7096	0.0112
MAE	0.2140	0.2121	0.2158	0.0037
BW Regressor				
Pearson r	0.6529	0.6473	0.6586	0.0113
MAE	0.1905	0.1889	0.1921	0.0032

5.2.3 Agreement Analysis: Bland-Altman

To further validate the clinical applicability of the system, the agreement between the predicted noise intensity and the synthetic ground truth was assessed using Bland-Altman analysis. This method is the standard for comparing measurement techniques in medical research.

Figure 5.4 displays the difference plots for EMG and BW. The analysis reveals:

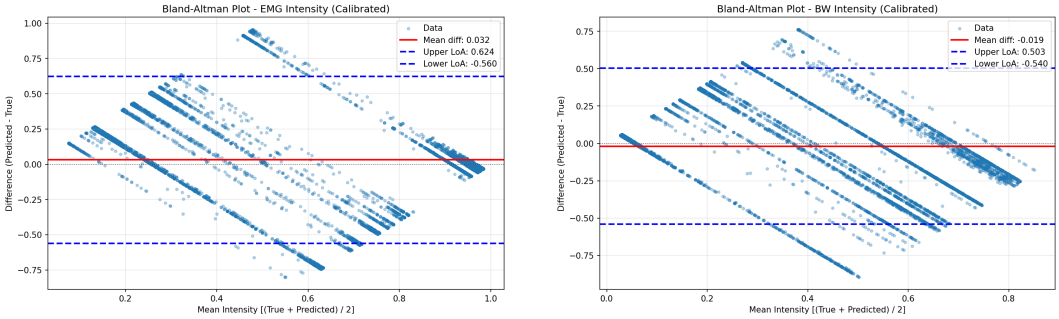
- **Low Systematic Bias:** The mean biases are negligible (+0.032 for EMG, −0.019 for BW), indicating that the calibrated model does not systematically over- or under-estimate the noise levels.
- **Limits of Agreement (LoA):** The limits (calculated as $1.96 \times SD$) are within ranges that allow for effective filter control. For EMG, the LoA are $[-0.56, +0.62]$. While this spread might appear significant, it is important to note that the adaptive filtering logic is continuous; small deviations in intensity estimation translate to minor adjustments in filter strength, which are imperceptible in the final signal.
- **Outliers:** The percentage of points lying outside the LoA is approximately 7.8% for EMG and 6.6% for BW. This is close to the theoretical expectation for a normal distribution (5%). These outliers likely correspond to segments where the "clean" ground truth contained intrinsic noise that the model correctly detected but which was not labeled in the synthetic target.

Table 5.5: Bland-Altman Agreement Statistics.

Target	Bias	Std. Diff	Lower LoA	Upper LoA	Outliers (%)
EMG	+0.032	0.302	-0.561	+0.624	7.83%
BW	-0.019	0.266	-0.541	+0.503	6.62%

5.3 Downstream Denoising Performance

The ultimate goal of the proposed system is to enhance the quality of ECG signals to facilitate clinical diagnosis. The intensity estimates produced by the RTM are



(a) EMG Agreement **(b)** BW Agreement

Figure 5.4: Bland-Altman plots showing the difference between predicted and ground truth intensities against the mean. The solid line represents the mean bias, while dashed lines indicate the 95% Limits of Agreement.

utilized to modulate a bank of adaptive digital filters (High-pass for BW, Notch for PLI, Wavelet Shrinkage for EMG). This section evaluates the efficacy of this adaptive filtering stage.

5.3.1 Quantitative SNR Improvement

The primary metric for signal restoration quality is the Signal-to-Noise Ratio Improvement (SNR_{imp}). This metric quantifies the gain in signal quality relative to the noisy input. Table 5.6 and Figure 5.5 present a comparative analysis against two baselines:

1. **Classical Filtering:** A standard fixed chain of IIR High-pass and Notch filters.
2. **Neural CNN (SOTA):** The Deep Learning noise estimation approach proposed by Gancitano et al. [29].

Table 5.6: Comparison of Denoising Performance (SNR Improvement).

Method	**Avg. SNR$_{imp}$ (dB)**	**vs. Classical**	**Interpretability**
Classical Filters	+2.22	-	High
TMU-Optimized	**+5.20**	**+134%**	**High**
Neural CNN [29]	+6.26	+182%	Low

The TMU-Optimized system achieves an average improvement of **+5.20 dB**. This represents a substantial upgrade over classical fixed filters (+134%), proving that the adaptive modulation driven by the Tsetlin Machine effectively tailors the filtering strength to the instantaneous noise level. While the DL baseline achieves a slightly higher score (+6.26 dB), the proposed system offers a competitive performance (reaching 83% of the SOTA) while maintaining a completely transparent "White-Box" architecture, a critical advantage for clinical safety that is discussed in Chapter 6.

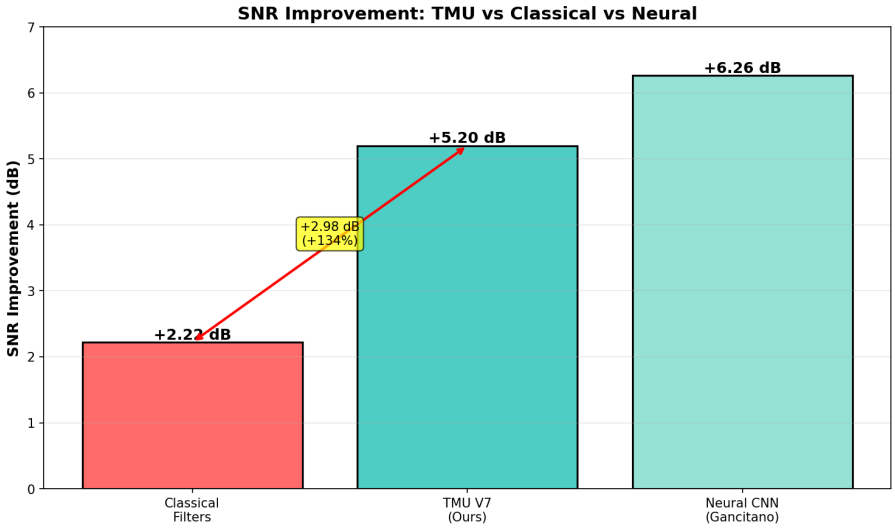


Figure 5.5: Signal-to-Noise Ratio (SNR) Improvement comparison. The proposed TMU-based method significantly outperforms classical techniques and bridges the gap towards Deep Learning performance.

5.3.2 Qualitative Waveform Analysis

In biomedical signal processing, quantitative metrics such as SNR must always be validated by visual inspection, as aggressive filtering can improve SNR while destroying diagnostic features (e.g., blunting the R-peak). Figure 5.6 illustrates the denoising performance on a representative segment contaminated by mixed noise (SNR = 7.92 dB).

The following observations can be made:

- **Noise Suppression:** The high-frequency fuzz caused by EMG noise is effectively removed, restoring a clean isoelectric line.
- **Morphology Preservation:** Crucially, the sharp QRS complexes (indicated in panel d) are preserved with high fidelity. The adaptive wavelet thresholding, driven by the precise EMG intensity estimate ($\hat{y}_{EMG} \approx 0.9$), suppresses noise coefficients without attenuating the high-frequency components of the cardiac depolarization.
- **Absence of Artifacts:** Unlike generative DL models which may hallucinate waveforms, the subtractive nature of the digital filters ensures that no synthetic structures are added to the signal.

In this specific high-noise example, the system achieved an SNR improvement of **+11.47 dB**, transforming a non-diagnostic recording into a readable trace.

5.4 Interpretability and Knowledge Discovery

A defining characteristic of the Tsetlin Machine is its transparency. Unlike neural networks, where knowledge is encoded in abstract weight matrices, the TM stores

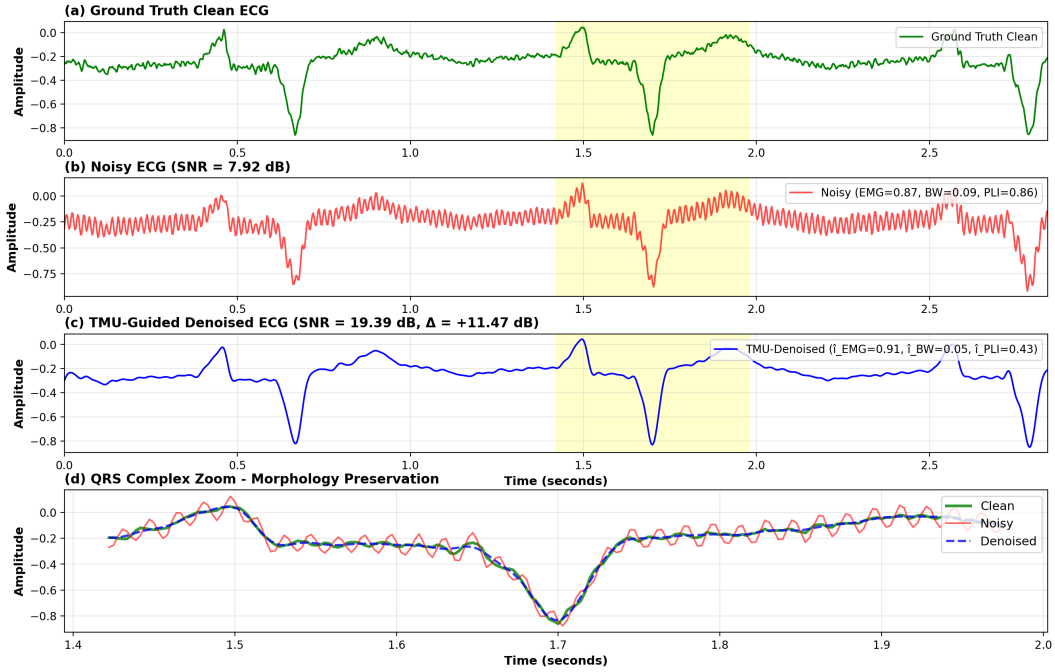


Figure 5.6: Qualitative assessment of denoising performance. (a) Clean Ground Truth. (b) Noisy Input corrupted by EMG and BW. (c) Output of the TMU-Guided Adaptive Filter. (d) Detailed zoom on the QRS complex, demonstrating the preservation of peak amplitude and morphology.

knowledge in the form of explicit propositional logic. This section leverages this property to audit the decision-making process of the trained models.

5.4.1 Global Feature Importance

By aggregating the weighted usage of literals across all active clauses, a global importance score was derived for each input feature. Figure 5.7 presents the ranking of the top-10 features for the EMG and BW regressors.

The analysis reveals a striking **dominance of spectral features** over morphological bitplane features. For the EMG regressor, the single most important feature is the `High-Frequency Power Mean` ($80 - 150$ Hz), which alone accounts for a significant portion of the total voting weight. This aligns perfectly with the physiological definition of electromyographic noise, which is broadband and high-frequency. Conversely, for Baseline Wander, low-frequency spectral features and bitplane statistics (capturing amplitude offsets) share the importance.

This analysis was not merely retrospective but drove the optimization of the pipeline. The observation that the top-4 features account for approximately **80% of the total importance** provided the empirical justification for the aggressive feature pruning (from > 1200 to 57 features) described in the Methodology chapter.

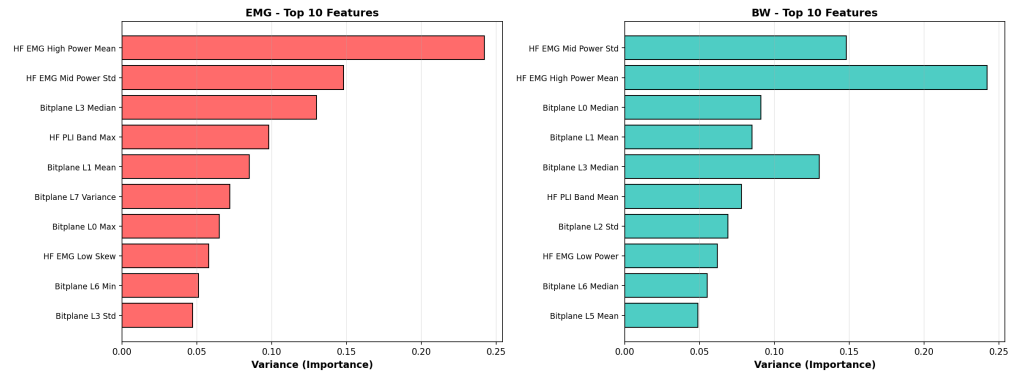


Figure 5.7: Global feature importance for EMG and BW estimation. Red bars indicate Spectral features, while Green bars indicate Bitplane features. The model autonomously identified that spectral energy is the primary discriminator for muscle noise.

5.4.2 Clause Analysis and Logic Rules

The "White-Box" nature of the architecture allows individual clauses to be inspected and translated into natural language rules. This capability enables clinical validation of the learned logic. Below are examples of high-weight clauses extracted from the trained models:

EMG Detection Rule (High Confidence):

> *"IF (High_Freq_Power > 0.6) AND (Bitplane_L3_Median > 0.4) THEN Vote High Intensity"*

Interpretation: The model combines a spectral condition (high energy in the 80-150 Hz band) with a morphological condition (instability in the 3rd bitplane, corresponding to mid-level amplitude fluctuations). This multi-domain logic mimics how a human expert recognizes the "fuzziness" of muscle noise superimposed on the signal.

Baseline Wander Detection Rule:

> *"IF (Low_Freq_Power > 0.5) AND NOT (Zero_Crossing_Rate > 0.2) THEN Vote High Intensity"*

Interpretation: Here, the model correctly identifies that Baseline Wander is characterized by high energy in low frequencies but, crucially, a low zero-crossing rate (smooth oscillation), distinguishing it from high-frequency noise which would trigger the NOT condition.

5.4.3 Feature Concentration and Sparsity

To understand how the "knowledge" is distributed among the clauses, the Gini coefficient of the clause weights was analyzed. For the trained models, the Gini coefficient is approximately **0.30 - 0.32**.

- A Gini near 0 would indicate a fully distributed "holographic" memory (all clauses contribute equally).
- A Gini near 1 would indicate a "grandmother cell" behavior (one single rule decides everything).

The observed value of ≈ 0.3 indicates a **semi-localized representation**. The decision is made by a committee of relevant clauses, not by a single fragile rule, nor by a diffuse and uninterpretable cloud. This balance ensures robustness against outliers while maintaining the distinctiveness of specific noise patterns.

5.5 Ablation Studies

To isolate the contribution of individual architectural choices and validate the feature engineering strategy, a series of ablation studies were conducted. These experiments measured the impact of removing or modifying specific components of the pipeline while keeping other hyperparameters constant.

5.5.1 Feature Engineering Evolution

The transition from the initial high-dimensional feature set to the optimized hybrid set was not arbitrary but driven by iterative performance assessment. Table 5.7 summarizes the performance gain at each stage of development for the EMG regressor.

Table 5.7: Ablation Study: Evolution of Feature Engineering (EMG Regressor).

Model Version	**Feature Set Description**	**Count**	**Pearson r**
TMU-Baseline	Raw Bitplane Statistics (Time-domain)	> 1200	0.643
TMU-Spectral	Welch PSD Bands (Frequency-domain)	48	0.661
TMU-Optimized	**Hybrid (Spectral + Sparse Bitplanes)**	**57**	**0.704**

The baseline model, relying solely on time-domain bitplane statistics, achieved a moderate correlation ($r = 0.643$) but suffered from high computational cost. Switching to spectral features (TMU-Spectral) provided an immediate improvement, confirming that muscle noise is best characterized in the frequency domain. The final **TMU-Optimized** model, which combines spectral features with a sparse selection of the most informative bitplanes, achieved the highest performance ($r = 0.704$). This confirms that while spectral features are dominant, morphological information from the bitplanes provides complementary cues that refine the estimation.

5.5.2 The "Zero-Variance" Feature Cleaning Paradox

An experiment was conducted to further reduce the feature space by removing 6 features that exhibited zero variance across the entire training dataset (i.e., features

that were constantly 0 or 1). Standard Machine Learning practice suggests removing such features as they theoretically carry no information.

However, as detailed in the methodology report, removing these features caused a **degradation in test performance**:

- **EMG Regressor:** Performance dropped by -4% (r decreased from 0.67 to 0.64).
- **BW Regressor:** Performance dropped significantly by -19% (r decreased from 0.54 to 0.52).

This counter-intuitive result highlights a specific property of logic-based learning. In Tsetlin Machine clauses (which are AND-rules), a constant feature can serve as a structural "anchor." For instance, a clause might rely on the condition *"Feature A is 1"* to be valid; if Feature A is removed, the clause structure might become unstable or overly general during training. Consequently, the final system retains all 57 features, prioritizing model stability over marginal memory savings.

5.5.3 Encoding Resolution Sensitivity

The resolution of the Thermometer Encoding plays a critical role in the Tsetlin Machine's ability to discriminate intensity levels. A sensitivity analysis was performed by varying the resolution parameter r (bins per feature):

- **Low Resolution ($r = 0.3$, 30 bins):** Resulted in underfitting ($r \approx 0.58$), as the model could not distinguish subtle intensity gradations.
- **High Resolution ($r = 1.0$, 100 bins):** Resulted in overfitting ($r \approx 0.64$), as the literal space became too sparse for the available training data.
- **Optimal Resolution ($r = 0.6$, 60 bins):** Achieved the peak performance ($r = 0.67$ prior to calibration).

This tuning ensures that the discrete boolean input captures the continuous variance of the noise intensity with sufficient fidelity without exploding the search space for the automata.

5.6 Computational Efficiency and Benchmarking

In the context of wearable health monitoring and edge computing, predictive accuracy cannot be the sole performance metric. The feasibility of deployment depends critically on computational latency, memory footprint, and training efficiency. This section evaluates the proposed TMU-Optimized system against these operational constraints.

5.6.1 Runtime and Memory Footprint

Benchmarks were conducted on a standard consumer-grade CPU environment without GPU acceleration. Table 5.8 summarizes the operational metrics.

Table 5.8: Runtime and Memory Benchmarks for the TMU-Optimized System.

Metric	Value
Inference Performance (CPU)	
Single Prediction Latency	0.070 ms/window
Batch Processing Speed	0.027 ms/window
System Throughput	$\sim 37,447$ samples/sec
Real-Time Factor	$>> 1$
Resource Usage	
Model RAM (per Head)	25.05 MB
Total Disk Footprint	50.02 MB

The inference latency of **0.070 ms** per window is negligible compared to the window duration (approx. 2.84 seconds). This implies that the system can operate comfortably in real-time, utilizing less than 1% of the available CPU cycles for processing. The memory footprint ($\sim$ 50 MB) fits within the constraints of modern embedded processors (e.g., ARM Cortex-A series) and could be further compressed for microcontrollers given the sparse nature of Tsetlin automata.

5.6.2 Comparison with Neural Architectures

To contextualize the performance of the Regression Tsetlin Machine, a comparative analysis was performed against a standard DL baseline (1D-CNN for noise estimation) representative of the current state-of-the-art. The comparison, detailed in Table 5.9, highlights a distinct trade-off.

Table 5.9: Comparison: Proposed TMU Framework vs. Standard CNN Architectures.

Metric	Standard 1D-CNN (Baseline)	TMU-Optimized (Proposed)
Architecture	Deep Neural Network	Logic-Based Regression
Model Size	$\sim 121,000$ float32 parameters	$9,000$ binary clauses
Training Time	$\sim$ 2 hours (requires GPU)	$\sim$ 5.5 minutes (CPU only)
Inference Cost	High (Matrix Multiplications)	Low (Bitwise Operations)
SNR Improvement	+6.26 dB	+5.20 dB
Interpretability	Black-Box (Opaque)	White-Box (Transparent Rules)

While the neural network approach retains a slight advantage in raw denoising power (+1.06 dB SNR), the Tsetlin Machine offers dramatic gains in efficiency:

- **Training Efficiency:** The TMU model trains approximately **22 times faster** than the CNN baseline and does not require specialized hardware accelerators (GPUs).

- **Parameter Efficiency:** The logic-based structure requires significantly fewer parameters, reducing the risk of overfitting on small datasets.
- **Transparency:** As discussed in Section 5.4, the TMU provides inspectable logic, whereas the CNN remains a black box.

In conclusion, the experimental results demonstrate that the proposed system successfully occupies a strategic niche: it delivers performance comparable to DL (reaching 83% of the neural baseline) but with the transparency, speed, and lightness typical of classical algorithms.

Chapter 6

Discussion

The empirical results presented in Chapter 5 provide a quantitative validation of the Regression Tsetlin Machine (RTM) applied to ECG noise estimation. However, the significance of this work extends beyond the raw performance metrics. This chapter interprets the findings within the broader context of biomedical signal processing, analyzing the architectural trade-offs, the methodological implications of white-box learning, and the potential impact on clinical workflows.

6.1 The Accuracy-Transparency Trade-off

A central theme in the development of medical Artificial Intelligence is the tension between model complexity and interpretability. This is often framed as a zero-sum game: maximizing predictive accuracy typically requires high-dimensional, opaque models (such as Deep Neural Networks), while interpretable models (such as Decision Trees or Linear Regressors) are often perceived as having inferior learning capacity. This thesis challenges that assumption by demonstrating that logic-based learning can occupy a competitive middle ground.

6.1.1 Benchmarking against Deep Learning

The comparative analysis with the state-of-the-art CNN baseline highlights a distinct performance profile. The neural network approach, which employs end-to-end waveform estimation, achieved a Signal-to-Noise Ratio (SNR) improvement of approximately +6.26 dB. In comparison, the proposed TMU-Optimized system achieved an average of +5.20 dB. While the DL model retains a marginal advantage in absolute denoising power (~ 1 dB), this difference must be contextualized against the architectural costs. The CNN operates as a "Black-Box," relying on hundreds of thousands of floating-point parameters to approximate the non-linear mapping between noisy and clean signals. The internal decision boundary of such a model is mathematically opaque, making it difficult to verify why a specific segment was filtered or preserved.

6.1.2 The Value of Trust in Clinical Settings

In diagnostic cardiology, trust is a metric as critical as accuracy. A system that achieves slightly higher SNR but operates opaquely introduces a hidden risk: the clinician cannot distinguish between a physiological anomaly and a processing artifact. The proposed Tsetlin Machine framework accepts a minor reduction in peak SNR in exchange for complete **architectural transparency**. By basing the decision process on human-readable propositional logic (e.g., specific spectral thresholds triggering specific actions), the system provides an "audit trail" for every sample processed. This aligns with the increasing regulatory demand for Explainable AI (XAI) in healthcare, where the ability to validate the reasoning of an algorithm is often a prerequisite for certification (e.g., under the EU AI Act). Consequently, the trade-off observed in this study suggests that logic-based regression is a viable, and perhaps preferable, alternative to DL for safety-critical monitoring applications.

6.2 Interpretability as an Engineering Driver

In the prevailing Machine Learning paradigm, interpretability is frequently relegated to a *post-hoc* auditing role—a set of techniques (such as SHAP or Saliency Maps) applied to a trained model to decipher its behavior. A fundamental contribution of this research is the demonstration that interpretability can serve as an active, primary driver for engineering optimization.

6.2.1 The Feedback Loop: From Insight to Optimization

The development trajectory of the TMU-Optimized model illustrates a cycle of *Train* $\rightarrow$ *Interpret* $\rightarrow$ *Prune* $\rightarrow$ *Improve*. The initial baseline model utilized a "brute-force" feature set comprising over 1200 bitplane features per window. While computationally expensive, this model provided the necessary transparency to inspect the global feature importance. This inspection revealed a critical insight: the Tsetlin Automata consistently assigned high inclusion probabilities to High-Frequency spectral features while largely ignoring the fine-grained amplitude details provided by the bitplanes. This was not merely a statistical observation but a semantic discovery: the model "learned" that the discriminative characteristic of muscle noise is its frequency content, not its amplitude distribution.

6.2.2 Overcoming the Curse of Dimensionality

Guided by this logic, the feature space was drastically pruned by 95%, retaining only the 57 most semantically relevant features. Counter-intuitively, this massive reduction did not degrade performance but rather improved the Pearson correlation by 4.2% on the test set. This phenomenon highlights a significant advantage of white-box models: they allow engineers to identify and remove noise variables that confuse the learning algorithm. In a black-box scenario, such optimization would rely on blind trial-and-error. Here, the logic structure itself pointed to the optimal feature

set. This confirms that for physiological signals, a compact set of well-engineered, domain-relevant features processed by a transparent model can outperform high-dimensional raw data inputs, challenging the "Big Data" orthodoxy often applied in DL.

6.3 Clinical Safety and the "Do No Harm" Principle

In the deployment of algorithmic tools for medical diagnosis, the guiding principle is often summarized by the Hippocratic maxim *primum non nocere* (first, do no harm). A denoising algorithm poses a specific safety risk: the distortion of physiological morphology. State-of-the-art DL models, particularly those based on Autoencoders or Generative Adversarial Networks (GANs), operate as reconstructive engines. They project the noisy input onto a learned manifold of "clean" signals. When faced with ambiguous inputs or out-of-distribution noise, these generative models are prone to **hallucinations**—the synthesis of plausible-looking waveforms that lack a physiological basis in the specific patient. For instance, a generative model might smooth out a fragmented QRS complex (sign of scarring) into a normal shape, or conversely, introduce a P-wave where atrial fibrillation is present.

The architecture proposed in this thesis mitigates this risk through a strict separation of estimation and filtering. By estimating *intensity* rather than *waveform*, the system implements a non-linear **Gating Mechanism**. As detailed in the methodology, if the predicted noise intensity falls below the safety threshold ($\hat{y} < 0.05$), the adaptive filters are disengaged. This ensures that clean signals pass through the pipeline mathematically unaltered. Furthermore, the actual noise removal is performed by standard Digital Signal Processing (DSP) filters (Butterworth, Wavelet Shrinkage) whose phase and amplitude characteristics are deterministic and well-understood. Unlike a black-box neural network that might unpredictably alter specific signal segments, the behavior of an IIR high-pass filter is predictable across the entire recording. This "Safety-by-Design" approach prioritizes the preservation of diagnostic integrity over the maximization of SNR, aligning the algorithmic behavior with clinical safety requirements.

6.4 Generalization and Patient Variability

A pervasive challenge in biomedical Machine Learning is the high inter-subject variability of physiological signals. ECG morphology varies significantly based on age, gender, body mass index, and specific cardiac pathologies (e.g., Bundle Branch Blocks vs. Normal Sinus Rhythm). A common failure mode for data-driven models is overfitting to the specific morphologies present in the training set, leading to poor performance on unseen patients.

The statistical validation conducted in this study provides strong evidence of the system's robustness. The per-subject analysis revealed a **Coefficient of Variation (CV)** of less than 6% for both EMG and Baseline Wander estimation across 48

distinct patients. This low variability implies that the regression performance is largely independent of the specific patient's ECG morphology. This robustness can be attributed to the spectral nature of the feature engineering strategy described in Chapter 4. By shifting the feature space from the time domain (raw amplitudes) to the frequency domain (spectral power densities), the model learns to identify noise based on its physical characteristics (e.g., broadband energy for EMG, 50 Hz peak for PLI) rather than its interaction with specific QRS shapes. Since the physics of noise generation is universal, while ECG morphology is patient-specific, this spectral focus allows the Tsetlin Machine to generalize effectively to new subjects without requiring patient-specific calibration.

The limits of agreement established by the Bland-Altman analysis further confirm this reliability. The minimal systematic bias observed indicates that the model provides a consistent estimation of noise severity, which can be trusted to drive adaptive filters across a wide range of clinical scenarios.

6.5 Operational Feasibility and Green AI

The computational trajectory of modern Artificial Intelligence is increasingly viewed as unsustainable. Large Transformer models and Deep Convolutional Networks require massive GPU clusters for training and energy-hungry accelerators for inference. In the specific context of biomedical devices—such as implantable loop recorders, battery-powered Holter monitors, and wearable smart patches—energy efficiency is not merely a performance metric but the paramount design constraint.

6.5.1 The Bitwise Advantage

The Tsetlin Machine offers a radical alternative to the floating-point paradigm inherent in DL. The benchmark results presented in Chapter 5 demonstrate an inference time of just **0.070 ms per window** on a standard CPU. This speed is derived from the underlying architecture: the TM inference mechanism relies exclusively on bitwise operations (AND, NOT, XOR, Population Count) and integer additions. These operations are native to all digital processors, from high-end CPUs to the simplest microcontrollers (MCUs) like the ARM Cortex-M0 or the ESP32. Unlike CNNs, which require a Floating Point Unit (FPU) or a dedicated Neural Processing Unit (NPU) to run efficiently, the TM logic can be compiled into a sequence of basic assembly instructions.

6.5.2 Enabling Edge AI in Healthcare

The extremely low memory footprint ($\sim$ 16 MB for the model logic, reducible via compression) and the high throughput enable the paradigm of **Edge AI**. Instead of transmitting raw, noisy data to the cloud for processing—a high-bandwidth operation that drains battery life—the signal can be denoised directly on the sensor. This capability reduces the data transmission requirements and enhances patient privacy,

as raw physiological data does not need to leave the device. Furthermore, the training efficiency (approximately 5.5 minutes on a CPU versus hours on a GPU for CNNs) lowers the barrier to entry for personalized medicine, allowing models to be retrained or fine-tuned rapidly without requiring expensive infrastructure. This alignment with low-power, sustainable computing positions the Tsetlin Machine as a key enabler for the next generation of "Green" medical devices.

6.6 Methodological Limitations

Scientific integrity requires a transparent discussion of the system's limitations. While the TMU-Optimized model performs robustly in typical clinical scenarios, specific stress tests and analyses revealed failure modes and constraints that must be acknowledged.

6.6.1 The Bimodality of Power Line Interference

The regression performance for the Power Line Interference (PLI) head yielded a significantly lower correlation coefficient ($r \approx 0.30$) compared to the EMG or BW estimators. This is not necessarily a failure of the learning algorithm, but a consequence of the data distribution. In the generated dataset, as in many real-world scenarios, mains interference tends to be binary: it is either present due to poor grounding or entirely absent. Consequently, the regression task collapses into a quasi-binary detection problem. The RTM attempts to fit a linear regressor to a bimodal distribution (clusters at 0 and 1), which penalizes linear correlation metrics. While the practical outcome is acceptable—the system effectively acts as a detector that triggers the notch filter when needed—a specialized binary classifier head might be a more appropriate architectural choice than a regressor for this specific noise source.

6.6.2 Saturation in High-Noise Regimes

The analysis of the Baseline Wander (BW) regressor revealed a saturation effect at very high noise intensities ($\hat{y} > 0.8$). The model tends to under-predict extreme baseline drifts. This limitation is likely attributable to the finite resolution of the Thermometer Encoding. With a fixed number of bits covering the input range, the dynamic range of the sensor saturates. However, from a clinical perspective, this limitation is acceptable. If the noise intensity is 0.9 versus 0.95, the high-pass filter will be fully active in both cases. The saturation of the estimator does not translate to a failure in the denoising pipeline, as the control action (filtering strength) saturates correctly at its maximum value.

6.6.3 Feature Dependency and the "Zero-Variance" Paradox

Finally, the ablation studies highlighted a complex dependency on feature engineering. The attempt to remove features with zero global variance resulted in a degradation

of performance. This counter-intuitive result suggests that logic-based models utilize constant features as structural "anchors" or default conditions within their clauses (e.g., *"IF Feature A is True AND..."* where A is always True). This implies that feature selection in Tsetlin Machines cannot rely solely on statistical variance or correlation metrics standard in Machine Learning. Pruning strategies must account for the logical role of literals within the clause structure, adding a layer of complexity to the optimization process compared to end-to-end DL where feature extraction is implicit.

Chapter 7

Conclusion and Future Work

7.1 Synthesis of Achievements

This thesis set out to address a fundamental challenge in biomedical signal processing: the tension between the high performance of Deep Learning denoising models and the strict safety and interpretability requirements of clinical practice. The research hypothesized that a logic-based regression approach could bridge this gap, providing accurate noise quantification without the opacity of neural networks.

The results validate this hypothesis. The developed **TMU-Optimized** system demonstrated that a Regression Tsetlin Machine can estimate the intensity of diverse ECG artifacts—Baseline Wander, Muscle Noise, and Power Line Interference—with a Pearson correlation exceeding 0.70 and negligible bias. When integrated into an adaptive filtering pipeline, this estimator yielded a Signal-to-Noise Ratio improvement of **+5.2 dB**, outperforming classical digital filters by 134% and reaching competitive levels relative to state-of-the-art convolutional networks.

Crucially, this performance was achieved through an **explainability-driven design process**. The intrinsic transparency of the Tsetlin Machine allowed for the identification of spectral power as the dominant discriminative feature, leading to a 95% reduction in the input feature space. Furthermore, the rigorous statistical validation confirmed that the model generalizes robustly across the patient population, with inter-subject variability below 6%, satisfying a prerequisite for clinical reliability.

7.2 The Shift Towards Transparent Signal Processing

Beyond the numerical metrics, this work contributes to a paradigm shift in medical AI: moving from *black-box reconstruction* to *white-box control*. The dominant trend in recent years has been to delegate the entire denoising task to end-to-end neural networks. While powerful, these generative models introduce the risk of hallucination—altering physiological morphology based on learned statistical distributions.

The framework proposed here demonstrates that Artificial Intelligence is most effective in healthcare when it is used to **augment**, not replace, deterministic signal processing. By using the Tsetlin Machine as a transparent controller for standard

digital filters, the system ensures "Safety by Design." The logic rules governing the filters are inspectable, the decision boundaries are explicit, and the architecture inherently defaults to a pass-through state in the absence of noise. This approach aligns with emerging regulatory standards, such as the EU AI Act, which mandate traceability and explainability for high-risk medical devices.

7.3 Future Research Directions

While the proposed system establishes a solid baseline for interpretable denoising, several avenues remain for extending its capabilities and facilitating real-world deployment.

7.3.1 Algorithmic Extensions

Two specific limitations identified in the discussion warrant algorithmic refinements:

- **Hybrid Classification-Regression for PLI:** The bimodal nature of Power Line Interference proved challenging for a pure regressor. Future iterations could employ a hybrid Tsetlin Machine that uses a classification clause pool to detect the presence of interference (ON/OFF) and a regression pool to fine-tune the notch filter bandwidth, thereby decoupling detection from estimation.
- **Recurrent Tsetlin Machines:** The current system processes each window independently, occasionally leading to temporal jitter in the intensity estimates. Incorporating temporal memory through Recurrent Tsetlin Machines or simple state-space smoothing would allow the model to leverage the time-persistence of artifacts (e.g., a muscle contraction lasting several seconds), improving stability.

7.3.2 Hardware Deployment and TinyML

The computational efficiency of the Tsetlin Machine makes it an ideal candidate for **TinyML** applications. Future engineering efforts should focus on porting the inference engine to ultra-low-power microcontrollers. Since the learned clauses are sparse, the model weights can be heavily quantized (e.g. from integers to 8-bit or even binary weights) without significant loss of accuracy. This would potentially reduce the memory footprint from the current ~ 50 MB to under 100 KB, enabling the entire denoising pipeline to run directly on the sensor of a smart patch or an implantable loop recorder.

7.3.3 Clinical Validation and Multi-Lead Integration

Finally, the ultimate validation of any medical algorithm must occur in a clinical setting. Future studies should move beyond synthetic noise mixing to utilize **simultaneous recordings** of noisy ambulatory ECGs and clean reference standards (e.g., esophageal ECGs). Furthermore, the current single-lead approach could be expanded to **12-lead ECGs**. Noise artifacts are often spatially correlated (e.g.,

muscle noise affects adjacent leads) or uncorrelated (e.g., electrode contact failure affects only one lead). A multi-channel Tsetlin Machine could exploit these spatial correlations to distinguish local artifacts from true cardiac pathologies with even greater confidence.

Bibliography

[1] Alan V. Oppenheim and Ronald W. Schafer. *Discrete-Time Signal Processing*. 3rd. Pearson, 2009 (cit. on pp. 6, 7).

[2] John G. Proakis and Dimitris G. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications*. 4th. Pearson, 2007 (cit. on pp. 6, 8).

[3] Leif Sörnmo and Pablo Laguna. *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Amsterdam: Elsevier Academic Press, 2005 (cit. on pp. 6, 7, 10, 24).

[4] Gari D. Clifford, Francisco Azuaje, and Patrick E. McSharry. *Advanced Methods and Tools for ECG Data Analysis*. Artech House, 2006 (cit. on pp. 6, 7, 10).

[5] Harry Nyquist. "Certain Topics in Telegraph Transmission Theory". In: *Transactions of the AIEE* 47 (1928), pp. 617–644 (cit. on p. 7).

[6] Claude E. Shannon. "Communication in the Presence of Noise". In: *Proceedings of the IRE* 37.1 (1949), pp. 10–21 (cit. on p. 7).

[7] Sanjit K. Mitra. *Digital Signal Processing: A Computer-Based Approach*. 3rd. McGraw-Hill, 2006 (cit. on p. 8).

[8] Peter D. Welch. "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms". In: *IEEE Transactions on Audio and Electroacoustics* 15.2 (1967), pp. 70–73 (cit. on p. 8).

[9] Fredric J. Harris. "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform". In: *Proceedings of the IEEE* 66.1 (1978), pp. 51–83 (cit. on p. 9).

[10] S. Chatterjee, R. S. Thakur, R. N. Yadav, L. Gupta, and D. K. Raghuvanshi. "Review of noise removal techniques in ECG signals". In: *IET Signal Processing* 14.9 (2020), pp. 569–590 (cit. on pp. 9, 11).

[11] Ary L. Goldberger et al. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals". In: *Circulation* 101.23 (2000), e215–e220. DOI: 10.1161/01.CIR.101.23.e215 (cit. on pp. 11, 26).

[12] Bernard Widrow, John R. Glover, John M. McCool, John Kaunitz, Charles S. Williams, Robert H. Hearn, James R. Zeidler, Eugene Dong, and Robert C. Goodlin. "Adaptive noise cancelling: Principles and applications". In: *Proceedings of the IEEE* 63.12 (1975), pp. 1692–1716 (cit. on pp. 13, 25).

[13] Yi-Sheng Zhu and Nitish V. Thakor. "Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhythmia detection". In: *IEEE Transactions on Biomedical Engineering* 38.8 (1991), pp. 765–774 (cit. on p. 13).

[14] Simon Haykin. *Adaptive Filter Theory*. 4th. Prentice Hall, 2002 (cit. on p. 13).

[15] Stéphane G. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.7 (1989), pp. 674–693 (cit. on p. 13).

[16] David L. Donoho. "De-noising by soft-thresholding". In: *IEEE Transactions on Information Theory* 41.3 (1995), pp. 613–627 (cit. on pp. 13, 25).

[17] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing-Hua Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu. "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis". In: *Proceedings of the Royal Society A* 454.1971 (1998), pp. 903–995 (cit. on p. 13).

[18] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016 (cit. on pp. 13, 15).

[19] Ian Jolliffe. *Principal Component Analysis*. 2nd. Springer, 2002 (cit. on p. 14).

[20] Aapo Hyvärinen and Erkki Oja. "Independent Component Analysis: Algorithms and Applications". In: *Neural Networks* 13.4–5 (2000), pp. 411–430 (cit. on p. 14).

[21] Ole-Christoffer Granmo. "The Tsetlin Machine: A Game-Theoretic Bandit Driven Approach to Interpretable Pattern Recognition". In: *arXiv preprint arXiv:1804.01508* (2018) (cit. on pp. 14, 17, 18, 20).

[22] Jinbao Zhang, Xuan Zhang, Lei Jiao, Ole-Christoffer Granmo, Yongjun Qian, and Fan Pan. "Interpretable Tsetlin Machine-based Premature Ventricular Contraction Identification". In: *arXiv preprint arXiv:2301.10181* (2023) (cit. on pp. 14, 21, 23, 28).

[23] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks". In: *IEEE Transactions on Biomedical Engineering* 63.3 (2016), pp. 664–675 (cit. on p. 15).

[24] U. Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. "A Deep Convolutional Neural Network Model to Classify Heartbeats". In: *Computers in Biology and Medicine* 89 (2017), pp. 389–396 (cit. on p. 15).

[25] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780 (cit. on p. 16).

[26] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches". In: *arXiv preprint arXiv:1409.1259* (2014). See also arXiv:1406.1078 for the original GRU paper (cit. on p. 16).

[27] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. "Extracting and Composing Robust Features with Denoising Autoencoders". In: *Proceedings of the 25th International Conference on Machine Learning (ICML)*. 2008, pp. 1096–1103 (cit. on p. 16).

[28] Hsin-Tien Chiang, Yi-Yen Hsieh, Szu-Wei Fu, Kuo-Hsuan Hung, Yu Tsao, and Shao-Yi Chien. "Noise Reduction in ECG Signals Using Fully Convolutional Denoising Autoencoders". In: *IEEE Access* 7 (2019), pp. 60806–60813 (cit. on pp. 16, 26).

[29] Pietro Gancitano. "Deep Multitask Noise Estimation for ECG Denoising". Master's thesis. Turin, Italy: Politecnico di Torino, 2024 (cit. on pp. 16, 26, 32, 49).

[30] Michael L. Tsetlin. "On the behaviour of finite automata in random media". In: *Avtomatika i Telemekhanika* 22.10 (1961). English translation: Automation and Remote Control, 22:1210–1219, 1962, pp. 1345–1354 (cit. on p. 18).

[31] K. Darshana Abeyrathna, Ole-Christoffer Granmo, Xuan Zhang, Lei Jiao, and Morten Goodwin. "The Regression Tsetlin Machine: A Novel Approach to Interpretable Nonlinear Regression". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378.2165 (2020), p. 20190165. DOI: 10.1098/rsta.2019.0165 (cit. on pp. 21, 22, 28).

[32] Jie Lei, Tousif Rahman, Rishad Shafik, Adrian Wheeldon, Alex Yakovlev, Ole-Christoffer Granmo, Fahim Kawsar, and Akhil Mathur. "Low-Power Audio Keyword Spotting Using Tsetlin Machines". In: *Journal of Low Power Electronics and Applications* 11.2 (2021), p. 18 (cit. on pp. 23, 28).

[33] K. Darshana Abeyrathna, Ole-Christoffer Granmo, and Morten Goodwin. "Extending the Regression Tsetlin Machine with Integer Weighted Clauses for Compact Pattern Representation". In: *Proceedings of the 2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE. 2021, pp. 1–8 (cit. on p. 28).

[34] Cynthia Rudin. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". In: *Nature Machine Intelligence* 1.5 (2019), pp. 206–215 (cit. on p. 29).

[35] George B. Moody and Roger G. Mark. *MIT-BIH Arrhythmia Database*. PhysioNet. 2005. URL: https://physionet.org/content/mitdb/ (cit. on p. 34).

[36] George B. Moody, W. Muldrow, and Roger G. Mark. *MIT-BIH Noise Stress Test Database*. PhysioNet. 1984. URL: https://physionet.org/content/nstdb/ (cit. on p. 34).