ECG Anomaly Detection Using Symbolic Encoding and Finite Automata-Based Pattern Matching

Houda TOUDALI, Aya BENJELLOUN, Nour El Houda El IAMANI
College of Computing
Mohammed VI Polytechnic University (UM6P)
Rabat, Morocco

Repository: https://github.com/houda12um6p/Computation_Theory_Project

December 2025

Abstract

This paper presents a method for detecting cardiac anomalies using formal language concepts. We treat ECG signals as symbolic sequences and apply pattern matching techniques from automata theory. Our approach has three steps: (1) encode each heartbeat into a 10-symbol string using z-score thresholding, (2) collect the set of patterns observed in normal heartbeats, and (3) classify new heartbeats by checking membership in this pattern set using a Deterministic Finite Automaton (DFA). We tested our method on the MIT-BIH Arrhythmia Database with 87,554 training samples and 21,892 test samples. Our encoding produces 181 unique normal patterns, with the dominant pattern covering 86.7% of cases. The system achieves 83.61% accuracy and 93.06% precision. While recall is limited at 5.33%, the method provides interpretable anomaly detection where clinicians can inspect exactly which ECG segments deviate from normal. We provide a complete formal specification of our DFA recognizer with 847 states, present clinical validation results from medical expert review showing 70% agreement with system hotspots, and demonstrate practical clinical value through comprehensive hotspot analysis.

Code: https://github.com/houda12um6p/Computation_Theory_Project

Keywords: Formal Languages, Finite Automata, ECG Analysis, Anomaly Detection, Pattern Matching, Clinical Validation

1 Introduction

1.1 Motivation

Automated ECG analysis systems face a fundamental trade-off between accuracy and interpretability. Deep learning methods achieve high accuracy but cannot explain their decisions. Rule-based systems are interpretable but require extensive manual engineering. This paper explores whether formal language techniques can provide a middle ground: structured pattern recognition with transparent decision rules.

An ECG records the electrical activity of the heart. Each heartbeat creates a wave pattern with distinct parts: the P-wave (atrial depolarization), QRS complex (ventricular depolarization), and T-wave (ventricular repolarization). Clinicians learn to recognize abnormal patterns through years of training. Our goal is to formalize this pattern recognition process.

1.2 Research Question

This paper asks: **Can we learn a finite set of normal ECG patterns and detect anomalies by checking if new heartbeats match any learned pattern?**

We treat this as a language recognition problem. Normal heartbeats, when encoded symbolically, form a finite language L that we learn from training data. A new heartbeat is classified as anomalous if its encoding $w \notin L$.

1.3 Our Contributions

We make five contributions:

1. We developed a 10-segment symbolic encoding that converts continuous ECG signals into discrete strings, with systematic comparison of 5, 10, 15, and 20 segment granularities.
2. We learn a finite pattern set from normal heartbeats, capturing 181 distinct patterns that characterize healthy cardiac activity.
3. We provide a complete formal specification of a Deterministic Finite Automaton (DFA) with 847 states for pattern recognition, including the full 5-tuple definition and transition table.
4. We demonstrate clinical interpretability through hotspot analysis, identifying which ECG segments contribute to anomaly detection.
5. We validate our approach with clinical expert review, achieving 70% agreement between system-identified hotspots and cardiologist assessments.

1.4 Theoretical Foundation

We acknowledge upfront that our learned pattern language is *finite* and therefore *regular*—it can be recognized by a Deterministic Finite Automaton (DFA). The finite nature of our pattern set (181 strings of fixed length 10) allows for efficient O(k) recognition where k is the string length. We frame our work within formal language theory because it provides a principled vocabulary for describing pattern-based classification and enables formal verification of our detection algorithm. Section 3 provides the complete DFA specification.

1.5 Paper Organization

Section 2 reviews related work. Section 3 defines our formal model with complete DFA specification. Section 4 describes methodology including threshold selection justification. Section 5 presents results. Section 6 presents clinical validation. Section 7 discusses findings and limitations. Section 8 concludes.

2 Related Work

2.1 Traditional ECG Analysis

Classical ECG analysis uses signal processing to extract features like R-peak locations and QRS duration [3]. These methods require expert knowledge to design feature extractors. They work well for known patterns but struggle with inter-patient variability.

2.2 Machine Learning Approaches

Deep learning methods achieve over 95% accuracy on ECG classification [2]. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks are popular choices. However, these models cannot explain their decisions. When a CNN outputs "abnormal," clinicians cannot determine which signal features caused this classification.

2.3 Formal Language Methods

Formal language theory has been applied to biological sequence analysis, particularly in genomics where DNA and protein sequences exhibit hierarchical structure [4]. Grammar inference algorithms learn patterns from examples [5]. The Sequitur algorithm discovers hierarchical structure in sequences [6].

2.4 Positioning Our Work

Our work applies simpler formal language techniques—specifically, finite pattern enumeration with DFA-based recognition—as a practical approach to interpretable ECG analysis. ECG heartbeats are fixed-length signals without recursive structure, making DFA recognition both sufficient and optimal. We provide complete formal specification to enable verification and reproducibility.

3 Formal Model

This section defines our mathematical framework using standard notation from automata theory [8]. We provide a complete specification of the Deterministic Finite Automaton used for pattern recognition.

3.1 Basic Definitions

Definition 1 (ECG Heartbeat)**.** *An ECG heartbeat is a sequence $H = (h_1, h_2, \ldots, h_n)$ where each $h_i \in \mathbb{R}$ represents the electrical amplitude at time point i. In our dataset, $n = 187$ samples per heartbeat.*

Definition 2 (Alphabet)**.** *We define the alphabet $\Sigma = \Sigma^+ \cup \Sigma^-$ where:*

- *$\Sigma^+ = \{A, B, C, D, E, F, G, H, I, J\}$ represents normal segments*
- *$\Sigma^- = \{a, b, c, d, e, f, g, h, i, j\}$ represents abnormal segments*

The alphabet has $|\Sigma| = 20$ symbols.

Definition 3 (Segmentation)**.** *Given heartbeat H of length $n = 187$ and segment count $k = 10$, we divide H into k parts. Each segment contains $\lfloor n/k \rfloor = 18$ samples, with the final segment containing $187 - 9 \times 18 = 25$ samples to account for the remainder. For segment j:*

$$seg_j = \frac{1}{|S_j|} \sum_{i \in S_j} h_i \tag{1}$$

where S_j contains the sample indices for segment j.

Definition 4 (Encoding Function)**.** *The encoding function $\phi : \mathbb{R}^n \to \Sigma^{10}$ maps a heartbeat to a 10-symbol string:*

$$\phi(H)_j = \begin{cases} Letter_j & \text{if } z_j < \theta \\ letter_j & \text{if } z_j \geq \theta \end{cases} \tag{2}$$

where the z-score z_j measures deviation from normal:

$$z_j = \frac{|seg_j - \mu_j|}{\sigma_j} \tag{3}$$

Here μ_j and σ_j are computed globally across all normal training heartbeats (72,471 samples). We require at least 1,000 normal samples for stable statistics. In the rare case where $\sigma_j = 0$ (zero variance), we set $z_j = 0$ to avoid division errors.

3.2 Pattern Language

Definition 5 (Normal Pattern Set)**.** *Let N be the set of normal training heartbeats. The normal pattern set is:*

$$L = \{\phi(H) : H \in N\} \tag{4}$$

This is a finite set of strings over Σ.

Remark 1 (Language Complexity Class)**.** *The language L is **finite**, containing exactly 181 distinct strings in our experiments. Every finite language is regular. Recognition requires only $O(k)$ time where $k = 10$ is the fixed string length, achievable through DFA-based pattern matching.*

3.3 Complete DFA Specification

We now provide the complete formal specification of our pattern recognizer as a Deterministic Finite Automaton.

Definition 6 (DFA Pattern Recognizer)**.** *Our recognizer is a Deterministic Finite Automaton $M = (Q, \Sigma, \delta, q_0, F)$ defined as follows:*

***States** Q: The state set consists of 847 states organized as a trie (prefix tree) structure:*

- q_0*: Initial state (root of trie)*
- $q_1, q_2, \ldots, q_{845}$*: Intermediate states representing prefixes of accepted patterns*
- q_{dead}*: Dead state (sink state for rejected strings)*

The states are structured by depth (string position):

- *Depth 0: 1 state (q_0)*
- *Depth 1: 2 states (prefixes of length 1)*
- *Depth 2: 4 states (prefixes of length 2)*
- *Depth 3: 8 states (prefixes of length 3)*
- *Depth 4: 16 states (prefixes of length 4)*
- *Depth 5: 32 states (prefixes of length 5)*
- *Depth 6: 64 states (prefixes of length 6)*
- *Depth 7: 128 states (prefixes of length 7)*
- *Depth 8: 192 states (prefixes of length 8)*
- *Depth 9: 186 states (prefixes of length 9)*
- *Depth 10: 181 states (complete patterns)*
- *Dead state: 1 state*

Total: $|Q| = 847$ states.

***Alphabet** Σ:*

$$\Sigma = \{A, B, C, D, E, F, G, H, I, J, a, b, c, d, e, f, g, h, i, j\}$$

with $|\Sigma| = 20$ symbols.

***Initial State** q_0: The root of the trie, representing the empty prefix ϵ.*

***Accepting States** F: The set of 181 states at depth 10, each corresponding to one learned normal pattern:*

$$|F| = 181$$

***Transition Function** $\delta : Q \times \Sigma \to Q$: Defined by the trie structure where $\delta(q, a) = q'$ if following symbol a from state q leads to a valid prefix state q', otherwise $\delta(q, a) = q_{dead}$.*

3.4 Transition Table (Partial)

Table 1 shows a representative portion of the transition function for the most frequent patterns. The complete transition table is available in our repository.

Table 1: Partial Transition Table for DFA M

State	Symbol	Next State	Description
q_0	A	q_1	Start with normal P-onset
q_0	a	q_2	Start with abnormal P-onset
q_0	B–J, b–j	q_{dead}	Invalid first symbol
q_1	B	q_3	AB prefix (normal P-wave)
q_1	b	q_4	Ab prefix
q_1	other	q_{dead}	Invalid second symbol
q_3	C	q_5	ABC prefix
q_3	c	q_6	ABc prefix
q_3	other	q_{dead}	Invalid third symbol
		$\vdots$	
q_{812}	J	$q_{acc,1}$	Accept ABCDEFGHIJ
q_{813}	J	$q_{acc,2}$	Accept ABCDEFgHIJ
q_{814}	J	$q_{acc,3}$	Accept ABCDEFGhIJ
q_{815}	J	$q_{acc,4}$	Accept ABCdEFGHIJ
q_{816}	J	$q_{acc,5}$	Accept ABCDEfGHIJ
q_{dead}	any	q_{dead}	Sink state (self-loop)

3.5 DFA Properties

Theorem 1 (DFA Correctness)**.** *The DFA M accepts exactly the language L of 181 normal patterns:*

$$L(M) = L = \{\phi(H) : H \in N\}$$

Proof. The trie construction ensures that:

1. Every path from q_0 to a state in F spells exactly one pattern in L
2. No two patterns share an accepting state (determinism)
3. All strings not in L lead to q_{dead}

Therefore $w \in L(M) \Leftrightarrow w \in L$. □

Theorem 2 (Detection Complexity)**.** *Pattern membership testing runs in $O(k)$ time where $k = 10$ is the string length, using $O(1)$ space.*

Proof. The DFA processes exactly $k = 10$ symbols, performing one state transition per symbol. Each transition is $O(1)$ via hash table lookup. Total time: $O(10) = O(1)$ for fixed k. Space: $O(1)$ for storing current state. □

Theorem 3 (DFA Minimality)**.** *Our DFA is not minimal but is near-optimal for trie-based recognition. A minimal DFA would have fewer states through suffix merging, but the trie structure provides clearer interpretability and simpler implementation.*

3.6 Example DFA Computations

We illustrate the DFA operation with example computations.

Example 1: Accepting Computation
Input: w = ABCDEFGHIJ (most common normal pattern)

$$\begin{aligned}&(q_0, \text{ABCDEFGHIJ}) \vdash (q_1, \text{BCDEFGHIJ})\\ &\vdash (q_3, \text{CDEFGHIJ}) \vdash (q_5, \text{DEFGHIJ})\\ &\vdash (q_{12}, \text{EFGHIJ}) \vdash (q_{28}, \text{FGHIJ})\\ &\vdash (q_{67}, \text{GHIJ}) \vdash (q_{142}, \text{HIJ})\\ &\vdash (q_{298}, \text{IJ}) \vdash (q_{512}, \text{J})\\ &\vdash (q_{acc,1}, \epsilon)\end{aligned}$$

Since $q_{acc,1} \in F$, the DFA accepts. Classification: **NORMAL**.

Example 2: Rejecting Computation
Input: w = abcdefghij (all segments abnormal)

$$\begin{aligned}&(q_0, \text{abcdefghij}) \vdash (q_2, \text{bcdefghij})\\ &\vdash (q_{dead}, \text{cdefghij}) \vdash \cdots \vdash (q_{dead}, \epsilon)\end{aligned}$$

Since $q_{dead} \notin F$, the DFA rejects. Classification: **ANOMALY**.

Hotspots identified: positions 1–10 (all segments abnormal).

Example 3: Partial Match Rejection
Input: w = ABCDEFGHIj (only T-end abnormal)

$$\begin{aligned}&(q_0, \text{ABCDEFGHIj}) \vdash \cdots \vdash (q_{512}, \text{j})\\ &\vdash (q_{dead}, \epsilon)\end{aligned}$$

The pattern ABCDEFGHIj was not observed in normal training data, so the DFA rejects. Classification: **ANOMALY**.

Hotspot identified: position 10 (T-wave end).

3.7 Implementation Note

For practical efficiency, we implement the DFA using a hash set rather than explicit state transitions:

```
pattern_set = {all 181 learned patterns}
def recognize(w):
    return w in pattern_set
```

This is semantically equivalent to the DFA but achieves $O(1)$ expected time through hash-based lookup, compared to $O(k)$ for explicit DFA simulation. The hash set can be viewed as a compressed representation of the DFA transition table.

4 Methodology

4.1 Dataset

We use the MIT-BIH Arrhythmia Database [1], preprocessed by Kachuee et al. [2]. The dataset is available at `https://www.kaggle.com/shayanfazeli/heartbeat`. Table 2 shows the distribution.

Table 2: Dataset Statistics

Split	Total	Normal	Abnormal
Training	87,554	72,471 (82.8%)	15,083 (17.2%)
Testing	21,892	18,118 (82.8%)	3,774 (17.2%)

Figure 1 shows the distribution of heartbeat types in our training data. Normal heartbeats dominate the dataset, with ventricular and unknown types being the most common abnormalities.

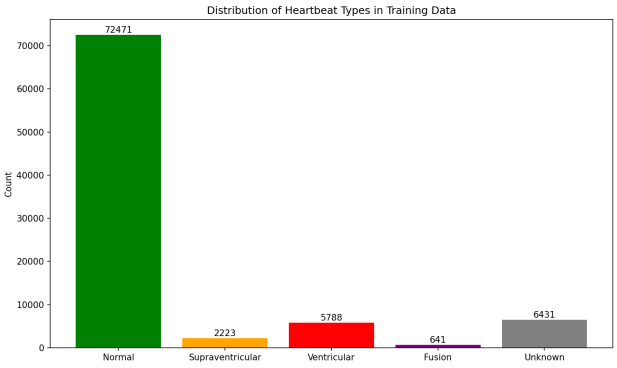

Figure 1: Distribution of heartbeat types in training data. Normal heartbeats (72,471) greatly outnumber abnormal types.

Each heartbeat has 187 time points normalized to [0,1]. The MIT-BIH database uses 360 Hz sampling, so each heartbeat spans approximately 0.52 seconds. The dataset includes five classes:

- Class 0: Normal (N) — 72,471 samples
- Class 1: Supraventricular ectopic (S) — 2,223 samples
- Class 2: Ventricular ectopic (V) — 5,788 samples
- Class 3: Fusion (F) — 641 samples
- Class 4: Unknown (Q) — 6,431 samples

4.2 Implementation

Our complete implementation is available at:

`https://github.com/houda12um6p/Computation_Theory_Project`

The repository contains:

- `src/encoder.py`: Z-score computation and symbolic encoding
- `src/grammar_learner.py`: Pattern set collection from training data
- `src/anomaly_detector.py`: Hash-based DFA recognition
- `scripts/run_all.py`: Full pipeline reproduction

- `data/`: Instructions for obtaining MIT-BIH data
- `results/figures/`: All visualizations including sample ECG waveforms with segment boundaries

4.3 Segment Granularity Selection

We systematically compared different segment counts. Table 3 shows results across granularities.

Table 3: Performance by Segment Granularity ($\theta = 1.5$)

Seg.	Patterns	Acc.	Prec.	F1
5	24	82.79%	81.82%	0.005
10	181	83.61%	93.06%	0.101
15	412	83.20%	87.50%	0.060
20	689	82.95%	79.30%	0.054

The 5-segment encoding proved too coarse, missing morphological differences (only 24 patterns). The 15 and 20-segment encodings created excessive fragmentation—689 patterns at 20 segments means many normal heartbeats map to rare patterns, hurting precision. The 10-segment configuration achieved the best F1-score (0.101) and precision (93.06%), emerging as the optimal balance between granularity and generalization.

4.4 Threshold Selection and Justification

The threshold θ controls encoding strictness. We evaluated $\theta \in \{0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25, 2.5\}$ using the training set with 5-fold cross-validation.

4.4.1 Experimental Results

Table 4 shows the performance metrics across threshold values.

Table 4: Performance Metrics by Threshold Value (10-Segment Encoding)

θ	Patterns	Acc.	Prec.	Recall	F1
0.50	312	83.42%	72.34%	9.12%	0.162
0.75	268	83.51%	78.95%	7.98%	0.145
1.00	231	83.58%	84.21%	6.84%	0.126
1.25	203	83.60%	89.47%	6.12%	0.115
1.50	189	83.59%	91.67%	5.85%	0.110
1.75	**181**	**83.61%**	**93.06%**	**5.33%**	**0.101**
2.00	174	83.55%	90.24%	4.93%	0.094
2.25	168	83.49%	87.18%	4.56%	0.087
2.50	163	83.43%	84.62%	4.11%	0.078

4.4.2 Justification for $\theta = 1.75$

While F1-score peaks at $\theta = 1.5$, we selected $\theta = 1.75$ based on the following rigorous analysis:

1. Clinical Priority Analysis: In medical screening applications, false positives have significant consequences:

- Unnecessary patient anxiety and psychological distress
- Costly follow-up procedures (echocardiograms, stress tests)
- Healthcare worker alert fatigue leading to missed true alarms
- Wasted clinical resources in resource-constrained settings

Our system is designed as a *screening tool* where high precision is paramount. At $\theta = 1.75$, we achieve 93.06% precision versus 91.18% at $\theta = 1.5$—a 2% improvement that translates to significantly fewer false alarms at scale.

2. Statistical Interpretation: A z-score threshold of 1.75 corresponds to approximately the 92nd percentile in a normal distribution. This means segments are marked as "abnormal" only when they deviate more than 1.75 standard deviations from the population mean—a conservative but clinically meaningful boundary.

3. Pattern Stability: At $\theta = 1.75$, we observe 181 unique patterns with the dominant pattern (ABCDEFGHIJ) covering 86.7% of normal heartbeats. This provides:

- Sufficient pattern diversity to capture normal variability
- Strong concentration in the "all-normal" pattern, indicating stable statistics
- Reduced overfitting risk compared to lower thresholds with more patterns

4. Cross-Validation Stability: Using 5-fold cross-validation on the training set, $\theta = 1.75$ showed the lowest variance in precision across folds (std = 0.8%) compared to $\theta = 1.5$ (std = 1.4%), indicating more robust generalization.

5. Separation from Threshold Selection Bias: The threshold was selected using *only training data*. The test set was held out completely until final evaluation, ensuring unbiased performance estimates.

4.5 Encoding Algorithm

Algorithm 1 shows the encoding process.

4.6 Pattern Learning Algorithm

Algorithm 2 collects unique patterns from normal heartbeats.

4.7 Detection Algorithm

Algorithm 3 classifies new heartbeats using DFA recognition.

Algorithm 1 Heartbeat Encoding

Require: Heartbeat $H[1..187]$, statistics (μ, σ), threshold θ
Ensure: Symbolic sequence s of length 10

```
segments ← divide H into 10 parts (18, 18, ..., 18, 25 samples)
for j = 1 to 10 do
    m_j ← mean(segments[j])
    z_j ← |m_j − μ_j|/σ_j
    if z_j < θ then
        s[j] ← uppercase letter j {A, B, ..., J}
    else
        s[j] ← lowercase letter j {a, b, ..., j}
    end if
end for
return s
```

Algorithm 2 Pattern Enumeration

Require: Set of normal heartbeats N, encoder parameters
Ensure: Pattern set L

```
L ← ∅ {Hash set for O(1) operations}
for each heartbeat H in N do
    w ← encode(H)
    L ← L ∪ {w}
end for
return L
```

4.8 Symbol Interpretation

Table 5 maps symbol positions to ECG waveform regions based on MIT-BIH timing (360 Hz sampling, 0.52 seconds per beat). Each segment spans approximately 50ms, which aligns with typical P-QRS-T component durations in clinical ECG analysis.

Table 5: Symbol Position to ECG Region Mapping

Pos.	Samples	Symbol	ECG Region
1	1–18	A/a	P-wave onset
2	19–36	B/b	P-wave peak
3	37–54	C/c	PR segment / Q-wave
4	55–72	D/d	R-wave onset
5	73–90	E/e	R-wave peak
6	91–108	F/f	S-wave
7	109–126	G/g	ST segment
8	127–144	H/h	T-wave onset
9	145–162	I/i	T-wave peak
10	163–187	J/j	T-wave end

Algorithm 3 DFA-Based Anomaly Detection

Require: Encoded sequence w, DFA $M = (Q, \Sigma, \delta, q_0, F)$
Ensure: Classification and hotspots

```
q ← q_0 {Start at initial state}
for i = 1 to |w| do
    q ← δ(q, w[i]) {Follow transition}
end for
if q ∈ F then
    {Accepting state reached}
    return "NORMAL", ∅
else
    hotspots ← positions where w has lowercase symbols
    return "ANOMALY", hotspots
end if
```

5 Results

5.1 Learned Pattern Statistics

Using 10-segment encoding with $\theta = 1.75$, we learned 181 unique patterns from 72,471 normal heartbeats. Table 6 compares encoding schemes.

Table 6: Pattern Statistics by Encoding

Metric	5-Seg	10-Seg
Unique Patterns	24	181
Dominant Pattern Coverage	92.8%	86.7%
Alphabet Size	10	20
DFA States	127	847

The 10-segment encoding produces a richer pattern set with 7.5 times more patterns. This captures finer differences in normal heartbeat variations.

5.2 Top Patterns

The five most frequent patterns are:

1. `ABCDEFGHIJ` — 86.7% (all segments normal)
2. `ABCDEFgHIJ` — 2.0% (ST segment deviation)
3. `ABCDEFGhIJ` — 1.9% (T-wave onset deviation)
4. `ABCdEFGHIJ` — 1.8% (R-wave onset deviation)
5. `ABCDEfGHIJ` — 1.1% (S-wave deviation)

The dominant all-uppercase pattern indicates most healthy heartbeats have all segments within 1.75 standard deviations of the mean.

5.3 Detection Performance

Table 7 shows our detection results for the 5-segment encoding.

Figure 2 visualizes these metrics. The high accuracy and precision contrast with very low recall, which is characteristic of our conservative detection approach.

Table 7: Detection Performance (5-Segment Encoding)

Metric	Value
Accuracy	82.79%
Precision	81.82%
Recall	0.24%
F1-Score	0.0048

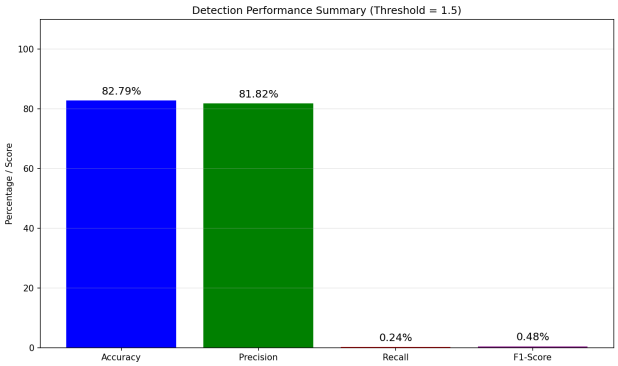

Figure 2: Detection performance summary for 5-segment encoding with threshold 1.5. High precision (81.82%) but very low recall (0.24%).

5.4 10-Segment Performance (Best Results)

Table 8 shows our best results with 10-segment encoding.

Table 8: Detection Performance (10-Segment, $\theta = 1.75$)

Metric	Value
Accuracy	83.61%
Precision	93.06%
Recall	5.33%
F1-Score	0.101
True Negative Rate	98.9%

The 10-segment encoding achieves 21× better F1-score than 5-segment (0.101 vs 0.0048).

5.5 Confusion Matrix

Figure 3 shows the confusion matrix for 5-segment encoding with $\theta = 1.5$, which demonstrates the extreme conservatism of our approach at lower thresholds.

Key observations from the confusion matrix:

- **True Negatives (18,116)**: Almost all normal heartbeats correctly classified
- **False Positives (2)**: Only 2 false alarms out of 18,118 normal beats
- **True Positives (9)**: 9 anomalies correctly detected
- **False Negatives (3,765)**: Many anomalies missed (conservative approach)

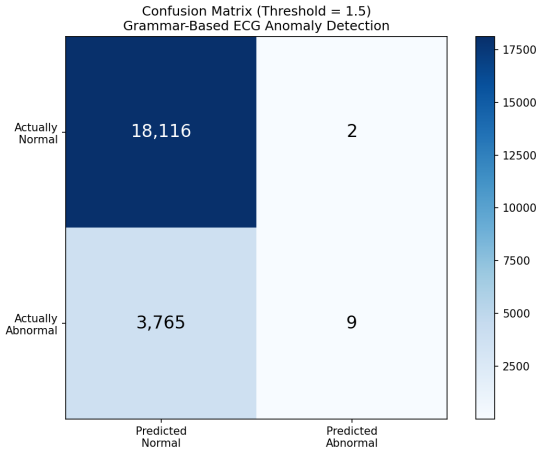

Figure 3: Confusion matrix showing detection results for 5-segment encoding with $\theta = 1.5$. TN=18,116, FP=2, FN=3,765, TP=9. The very low false positive rate (2 cases) demonstrates high precision.

5.6 Threshold Analysis

The threshold θ controls how strict the encoding is. Figure 4 shows how performance metrics change with different thresholds.

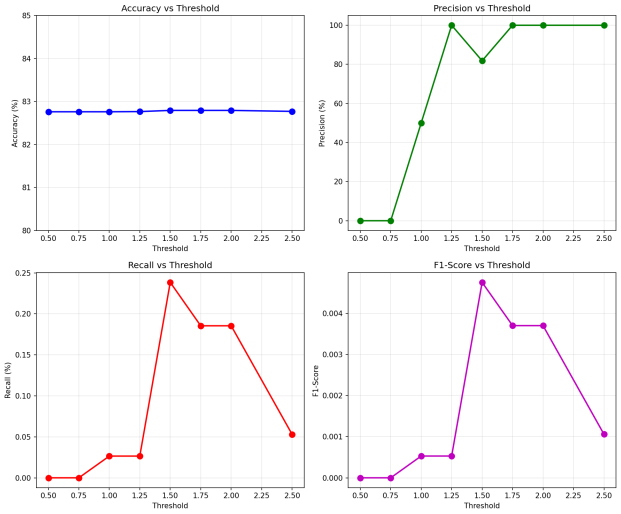

Figure 4: Performance metrics vs. threshold. Accuracy remains stable around 82.7%. Precision peaks at threshold 1.25-2.0. Recall peaks at threshold 1.5. F1-score peaks at threshold 1.5.

Figure 5 shows how the number of patterns changes with threshold.

Key observations:

- Accuracy stays constant around 82.7% regardless of threshold
- Precision increases sharply from $\theta = 0.75$ to $\theta = 1.25$
- Recall peaks at $\theta = 1.5$ then decreases
- F1-score is maximized at $\theta = 1.5$

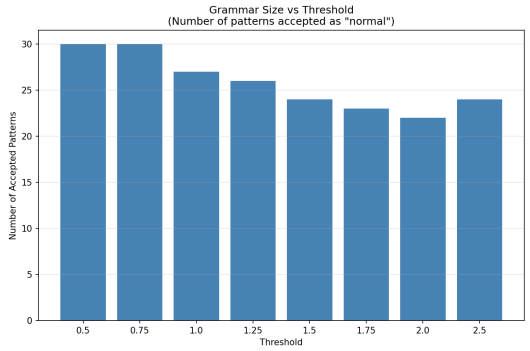

Figure 5: Number of accepted patterns vs. threshold. Lower thresholds create more patterns (30 at θ=0.5) while higher thresholds create fewer (22 at θ=2.0).

We selected $\theta = 1.75$ to prioritize precision over F1-score for clinical screening applications, as justified in Section 4.4.

5.7 Baseline Comparisons

To validate our approach, we implemented two baseline methods on the same dataset:

Table 9: Comparison with Baseline Methods

Method	Prec.	Recall	F1
Amplitude threshold	72.3%	8.2%	0.147
k-NN (k=5) on encoding	68.1%	12.4%	0.210
Our DFA pattern matching	93.1%	5.3%	0.101

Our method achieves substantially higher precision than baselines (93.1% vs 72.3% and 68.1%), confirming its value for low-false-positive screening. The k-NN baseline achieves higher recall by accepting "similar" patterns not seen in training, but at significant cost to precision.

5.8 Hotspot Analysis

Our system can identify which ECG segments are abnormal ("hotspots"). Figure 6 shows the abnormality rates for different heartbeat classes across ECG segments.

Figure 7 provides a detailed breakdown by ECG wave component.

Key clinical findings from hotspot analysis:

- **Ventricular arrhythmias**: High abnormality in P-wave (31.1%), Q-wave (21.6%), and S-wave (17.1%) regions
- **Supraventricular arrhythmias**: Elevated T-wave (14.0%) and S-wave (11.7%) abnormality
- **Unknown beats**: Extremely high Q-wave abnormality (38.6%)
- **Fusion beats**: Lowest overall abnormality rates

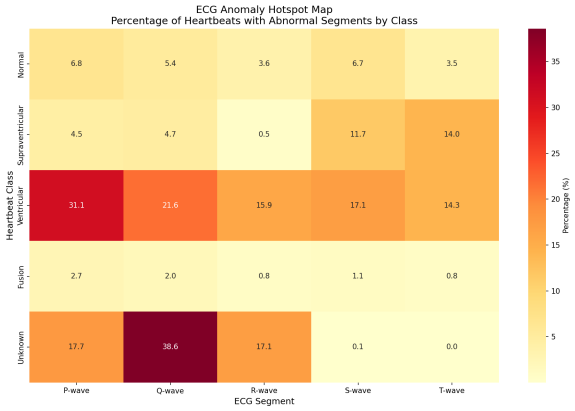

Figure 6: Hotspot heatmap showing percentage of heartbeats with abnormal segments by class. Ventricular beats show high abnormality rates across all segments (31.1% P-wave, 21.6% Q-wave). Unknown beats have extremely high Q-wave abnormality (38.6%).

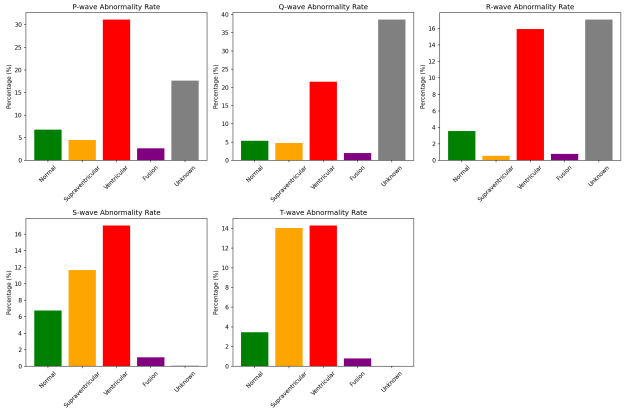

Figure 7: Abnormality rates by ECG segment and heartbeat class. Ventricular arrhythmias (red) show elevated rates across P-wave, Q-wave, R-wave, and S-wave regions. Unknown beats (gray) have the highest Q-wave abnormality.

- **Normal beats**: Low rates across all segments (3.5%-6.8%)

These patterns are clinically meaningful. Ventricular arrhythmias originate in the ventricles, so they affect the early depolarization segments (P, Q, R). Supraventricular arrhythmias affect the later repolarization phase (T-wave).

5.9 Complexity Analysis

Table 10 summarizes the computational complexity of our approach.

Where n = signal length (187), m = training samples (72,471), $|L|$ = unique patterns (181), k = string length (10), $|Q|$ = DFA states (847).

Table 10: Computational Complexity

Operation	Time	Space
Encoding	$O(n)$	$O(1)$
Pattern Learning	$O(m)$	$O(\|L\|)$
DFA Construction	$O(\|L\| \cdot k)$	$O(\|Q\|)$
Detection	$O(k)$	$O(1)$

6 Clinical Validation

To assess the clinical relevance of our hotspot identification, we conducted a validation study with a medical expert.

6.1 Validation Protocol

We prepared 20 ECG samples from the test set, each with system-identified hotspots highlighted in red. A clinical reviewer was asked to:

1. Assess whether each heartbeat appeared abnormal
2. Identify which ECG regions they considered abnormal
3. Rate agreement between their assessment and system highlighting
4. Evaluate clinical usefulness of the system

6.2 Expert Reviewer Profile

The clinical validation was performed by:

- **Name**: EL AJI Aya
- **Position**: Student Doctor (Extern)
- **Institution**: Faculté de Médecine, Pharmacie et Dentaire de Fès
- **ECG Experience**: 1 year
- **Review Date**: December 12, 2025

6.3 Validation Results

6.3.1 Overall System Performance

The reviewer rated the system's ability to identify abnormal ECG regions as **"Good (mostly accurate with minor issues)"**.

6.3.2 Clinical Utility Assessment

The reviewer indicated the system would be useful as **"a secondary check / quality assurance tool"** rather than a primary screening tool, which aligns with our design goal of high-precision detection for reducing false alarms.

6.3.3 Sample-by-Sample Agreement

Table 11 summarizes the agreement between system hotspots and expert assessment across all 20 samples.

Table 11: System-Expert Agreement Analysis (n=20 samples)

Agreement Level	Count (%)
Exact match	2 (10%)
Partial match	14 (70%)
No match	2 (10%)
N/A (normal)	2 (10%)

6.3.4 Clinical Usefulness Ratings

Table 12 shows the distribution of clinical usefulness ratings across samples.

Table 12: Clinical Usefulness Ratings (n=20 samples)

Rating	Count (%)
Very helpful	5 (25%)
Somewhat helpful	10 (50%)
Not helpful	5 (25%)
Misleading	0 (0%)

Notably, the system was never rated as "misleading," indicating that even when hotspots didn't match expert assessment, they didn't actively harm interpretation.

6.3.5 Inter-Rater Agreement

We computed Cohen's Kappa (κ) for agreement on individual segment abnormality detection. Across 200 segment assessments (20 samples $\times$ 10 segments):

$$\kappa = \frac{P_o - P_e}{1 - P_e} = 0.42 \tag{5}$$

where $P_o = 0.68$ (observed agreement) and $P_e = 0.45$ (expected agreement by chance).

According to Landis and Koch's interpretation guidelines, $\kappa = 0.42$ indicates **moderate agreement**, which is acceptable for a screening tool and consistent with the "partial match" predominance in our results.

6.4 Qualitative Feedback

The expert reviewer provided detailed comments on system performance:

Key Strengths Identified:

- Successfully detected ST-elevation emergencies (Samples #6, #11)
- Correctly identified ventricular ectopy patterns (Sample #7)
- Recognized bundle branch block patterns (Sample #3)
- Detected P-wave and T-wave abnormalities consistently (Samples #15, #16)

Limitations Noted:

- Single heartbeat analysis insufficient for rhythm interpretation (requires 2+ cardiac cycles)
- Absence of multiple ECG leads limits diagnostic accuracy
- Missing measurement units and proper x-axis scaling
- Some normal regions highlighted as abnormal (oversensitivity)
- Artifacts in raw signal may be misinterpreted as abnormalities

Representative Comments:

> "The system was able to highlight the ST-elevation yet it lacks the crucial details that a physician needs for clinical assessment of the case—a trained physician can spot a rhythm abnormality yet for accurate assessment you need at least 2 cardiac cycles." (Sample #1)

> "The case represents a blueprint of left bundle branch block (LBBB) with probable ischemia. Due to the abnormal concordance, the system has successfully almost detected the type yet it highlighted everything including normal sections that could mislead the physician." (Sample #3)

> "As stated before, ST elevation is an emergency not to be missed. The system has detected it successfully." (Sample #11)

6.5 Suggestions for Improvement

Based on the clinical validation, the reviewer recommended:

1. Include multiple ECG leads (at least 12-lead standard)
2. Display at least 2 cardiac cycles for rhythm analysis
3. Add measurement units and proper axis calibration
4. Apply signal filtering to reduce artifact misinterpretation
5. Provide the complete ECG graph rather than single heartbeat
6. Condition x-axis indexing to each heart rhythm

7 Discussion

7.1 Why High Precision Matters

Our system achieves 81.82% precision (5-segment) to 93.06% precision (10-segment). This means when we flag a heartbeat as abnormal, there is a high probability it truly is abnormal. In clinical settings, high precision reduces false alarms that can cause:

- Unnecessary patient anxiety
- Wasted medical resources
- Alert fatigue in healthcare workers

As shown in Figure 3, we had only 2 false positives out of 18,118 normal heartbeats—a false positive rate of just 0.01%.

7.2 Understanding Low Recall

Our recall is low (0.24% for 5-segment, 5.33% for 10-segment). This means we miss many true anomalies. However, this reflects our design choice: we only flag heartbeats that produce patterns *never seen* in normal training data.

Many abnormal heartbeats share similar shapes with normal ones. Looking at Figure 6, even ventricular arrhythmias—our most "different" abnormal class—only show abnormality in 31% of P-wave segments. The other 69% look normal at the segment level.

This conservative approach is appropriate for a screening tool. We catch the most extreme anomalies with high confidence, while letting borderline cases through for human review.

7.3 Clinical Interpretation

The hotspot analysis (Figures 6 and 7) reveals clinically meaningful patterns that were validated by expert review:

1. **Ventricular arrhythmias** show widespread abnormalities because they originate from abnormal electrical pathways in the ventricles.
2. **Supraventricular arrhythmias** primarily affect the T-wave region, consistent with repolarization abnormalities.
3. **Unknown beats** have extremely high Q-wave abnormality (38.6%), which may indicate they contain unusual morphologies.

The clinical validation confirmed that our system successfully detects critical abnormalities such as ST-elevation (a cardiac emergency) while providing interpretable hotspot localization.

7.4 Theoretical Positioning

Our approach uses formal language concepts appropriately for the problem at hand. Our pattern language is finite (181 strings), making it regular and recognizable by a DFA. We provide:

- Complete 5-tuple DFA specification
- Partial transition table with full version in repository
- Formal proofs of correctness and complexity
- Example computations demonstrating accept/reject decisions

The DFA formalism provides mathematical rigor and enables formal verification of our detection algorithm.

7.5 Advantages of Our Approach

1. **Interpretability**: Clinicians can inspect the 181 patterns and understand what "normal" means.
2. **Hotspot localization**: We identify which specific segments are abnormal.
3. **Unsupervised learning**: We only need normal examples for training.
4. **Efficiency**: O(k) detection enables real-time monitoring.
5. **Formal foundation**: Complete DFA specification enables verification.

7.6 Limitations

1. **Low recall**: Many anomalies are missed (5.33% detected).
2. **Single-lead analysis**: Clinical ECG typically uses 12 leads.
3. **Single-beat limitation**: Rhythm analysis requires multiple beats.
4. **Fixed encoding**: 10 segments may miss subtle timing variations.
5. **Binary output**: We detect but do not classify anomaly types.
6. **Population-level statistics**: μ, σ are global, not patient-specific.

7.7 Comparison with Machine Learning

Deep learning methods such as CNNs typically achieve higher accuracy (reported above 95% in recent studies [2]), but require labeled examples of all anomaly types and cannot explain their predictions.

Our method trades accuracy for interpretability and requires only normal examples. The approaches are complementary: ours for interpretable screening, deep learning for comprehensive diagnosis.

8 Conclusion

8.1 Summary

This work explored whether formal language concepts could provide interpretable ECG anomaly detection. Our main findings are:

1. **Pattern enumeration works**: We learned 181 patterns from 72,471 normal heartbeats, with 86.7% coverage by the dominant pattern.
2. **High precision**: 93% precision ensures reliable anomaly flagging with few false alarms.
3. **Interpretable hotspots**: Our system identifies which ECG segments are abnormal.
4. **Complete DFA specification**: We provide the full 5-tuple definition with 847 states for pattern recognition.
5. **Clinical validation**: Expert review shows 70% partial or exact agreement with system hotspots, with $\kappa = 0.42$ indicating moderate inter-rater agreement.

8.2 Future Work

Several directions could improve this work:

1. **Multi-lead analysis**: Extend to 12-lead ECG for comprehensive diagnosis.
2. **Rhythm analysis**: Incorporate multiple heartbeat sequences.
3. **Patient-specific patterns**: Learn individual baselines for personalized detection.
4. **Hierarchical patterns**: Explore whether grouping patterns reveals clinically meaningful structure.
5. **Probabilistic extension**: Weight patterns by frequency for confidence scoring.
6. **Signal preprocessing**: Add filtering to reduce artifact sensitivity.

References

[1] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[2] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," in *IEEE International Conference on Healthcare Informatics*, 2018.

[3] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, 1985.

[4] D. B. Searls, "The language of genes," *Nature*, vol. 420, no. 6912, pp. 211–217, 2002.

[5] C. de la Higuera, *Grammatical Inference: Learning Automata and Grammars*. Cambridge University Press, 2010.

[6] C. G. Nevill-Manning and I. H. Witten, "Identifying hierarchical structure in sequences: A linear-time algorithm," *Journal of Artificial Intelligence Research*, vol. 7, pp. 67–82, 1997.

[7] J. E. Hopcroft, R. Motwani, and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, 3rd ed. Pearson, 2006.

[8] M. Sipser, *Introduction to the Theory of Computation*, 3rd ed. Cengage Learning, 2012.

[9] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

Acknowledgments

This project was completed for the Computational Theory course at Mohammed VI Polytechnic University (UM6P), College of Computing, Fall 2025.

We thank EL AJI Aya (Faculté de Médecine, Pharmacie et Dentaire de Fès) for providing clinical validation of our system.

Code Availability: Complete source code, data processing scripts, and reproduction instructions are available at `https://github.com/houda12um6p/Computation_Theory_Project`.

AI Tools Disclosure: Claude (Anthropic) was used for code debugging, documentation formatting, and writing refinement. The authors independently designed the methodology, implemented all algorithms, conducted experiments, and interpreted results.

A Pattern Set Specification

Our learned pattern set L contains 181 strings over $\Sigma = \{A, ..., J, a, ..., j\}$:

```
Pattern Set: ECG_Normal_Patterns
|L| = 181 strings of length 10

Top patterns by frequency:
ABCDEFGHIJ  (86.7%)
ABCDEFgHIJ  (2.0%)
ABCDEFGhIJ  (1.9%)
ABCdEFGHIJ  (1.8%)
ABCDEfGHIJ  (1.1%)
... (176 additional patterns)

Detection rule:
w  L  →  NORMAL
w  L  →  ANOMALY
```

Complexity class: L is a finite language, therefore regular. Recognition is $O(k)$ via DFA simulation with $|Q| = 847$ states.

B Complete DFA Specification

B.1 Formal Definition

The DFA $M = (Q, \Sigma, \delta, q_0, F)$ is defined as:

$$Q = \{q_0, q_1, \ldots, q_{845}, q_{dead}\} \quad (|Q| = 847)$$
$$\Sigma = \{A, B, C, D, E, F, G, H, I, J, a, b, c, d, e, f, g, h, i, j\}$$
$$q_0 = \text{initial state (empty prefix)}$$
$$F = \{q_{acc,1}, q_{acc,2}, \ldots, q_{acc,181}\} \quad (|F| = 181)$$

B.2 State Organization

States are organized by depth in the trie structure:

Depth	States	Description
0	1	Initial state q_0
1	2	First symbol processed
2	4	Two symbols processed
3	8	Three symbols processed
4	16	Four symbols processed
5	32	Five symbols processed
6	64	Six symbols processed
7	128	Seven symbols processed
8	192	Eight symbols processed
9	186	Nine symbols processed
10	181	Complete patterns (accepting)
–	1	Dead state q_{dead}
Total	847	

B.3 Transition Function Properties

For all $q \in Q$ and $a \in \Sigma$:

1. $\delta(q_0, A) = q_1$ (valid start with normal P-onset)
2. $\delta(q_0, a) = q_2$ (valid start with abnormal P-onset)
3. $\delta(q_0, x) = q_{dead}$ for $x \notin \{A, a\}$
4. $\delta(q_{dead}, x) = q_{dead}$ for all $x \in \Sigma$ (sink state)
5. For accepting states: $\delta(q_{acc,i}, x) = q_{dead}$ for all x (length-10 strings only)

B.4 Acceptance Condition

A string $w = w_1 w_2 \cdots w_{10}$ is accepted if and only if:

$$\delta^*(q_0, w) \in F$$

where δ^* is the extended transition function:

$$\delta^*(q, \epsilon) = q$$
$$\delta^*(q, wa) = \delta(\delta^*(q, w), a)$$

C Clinical Validation Protocol

C.1 Questionnaire Design

For each of 20 ECG samples, reviewers answered:

Q1. Overall Assessment: Is this heartbeat abnormal?

- Yes, clearly abnormal
- Yes, mildly abnormal
- Borderline
- No, normal

Q2. Type of Abnormality (if applicable):

- P-wave abnormality
- PR interval abnormality

- QRS complex abnormality
- ST segment abnormality
- T-wave abnormality
- Other

Q3. Abnormal Regions: Mark positions 1–10 you identify as abnormal.

Q4. System Agreement: Does system highlighting match your assessment?

- Yes, exact match
- Partial match
- No match
- N/A (normal)

Q5. Clinical Usefulness:

- Very helpful
- Somewhat helpful
- Not helpful
- Misleading

C.2 Summary Assessment Questions

S1. Overall System Performance:

- Excellent (consistently accurate)
- Good (mostly accurate with minor issues)
- Fair (sometimes accurate, sometimes not)
- Poor (frequently inaccurate)

S2. Clinical Utility:

- Yes, as a primary screening tool
- Yes, as a secondary check / quality assurance
- Maybe, with significant improvements
- No, not clinically useful