Scanpath Pattern Recognition for ECG Interpretation
Probabilistic Finite Automata Approach

Nassim Lazrek nassim.lazrek@um6p.ma
Omar Ait Said omar.aitsaid@um6p.ma
Ilyass Skiriba ilyass.skiriba@um6p.ma
Mohammed VI Polytechnic University – Computational Theory Fall 2025

---

1 Summary

This project develops a scanpath pattern recognition system for ECG interpretation using **Probabilistic Finite Automata (PFA)**. The system distinguishes expert from novice readers, quantifies expertise through entropy measures, generates realistic scanpaths, and predicts sequence continuations. Our approach models sequential visual behavior as a first-order Markov process over the 12-lead ECG state space, providing an interpretable alternative to black-box methods.

2 Chosen Formal Model: Probabilistic Finite Automata

We adopt a PFA defined as $\mathcal{A} = (Q, \Sigma, T, \pi_0)$ where:

- $Q = \Sigma = \{\text{I, II, III, aVR, aVL, aVF, V1–V6}\}$: 12 ECG leads as states/alphabet
- $T \in [0, 1]^{12\times12}$: Transition matrix, $T_{ij} = P(X_{t+1} = j|X_t = i)$
- $\pi_0 \in [0, 1]^{12}$: Initial distribution, $\pi_0(i) = P(X_1 = i)$

Key assumption: First-order Markov property—next fixation depends only on current lead. Justified by: (1) cognitive locality of visual attention, (2) anatomical constraints on lead transitions, (3) saccade mechanics based on current position.

Advantages: Interpretable states (ECG leads), $O(n)$ inference, native generation capability, provable complexity guarantees.

3 Research Question

To what extent can expert ECG interpretation strategies be modeled as first-order Markov processes? Specifically: (1) Can PFAs classify expert vs. novice scanpaths? (2) Does transition entropy quantify expertise? (3) Can PFAs generate realistic expert patterns? (4) How accurately can PFAs predict sequence continuations?

4 Preliminary Approach

4.1 Model Architecture

Two separate PFAs: $\mathcal{A}_{\text{exp}}$ (trained on expert data) and $\mathcal{A}_{\text{nov}}$ (trained on novice data).

4.2 Core Algorithms

1. **Training**: MLE with Laplace smoothing: $\hat{T}_{ij} = (C_{ij} + \alpha)/(\sum_k C_{ik} + \alpha|Q|)$
2. **Probability**: $P(S|\mathcal{A}) = \pi_0(s_1) \cdot \prod_{t=1}^{n-1} T_{s_t,s_{t+1}}$ in $O(n)$ time
3. **Classification**: Log-likelihood ratio $\Lambda(S) = \log P(S|\mathcal{A}_{\text{exp}}) - \log P(S|\mathcal{A}_{\text{nov}})$; Expert if $\Lambda > 0$
4. **Generation**: Ancestral sampling from learned distributions
5. **Completion**: Viterbi-style dynamic programming for optimal continuation

4.3 Expertise Quantification

Transition entropy: $H(\mathcal{A}) = -\sum_i \pi(i) \sum_j T_{ij} \log_2 T_{ij}$

Hypothesis: Expert PFAs have lower entropy (more structured patterns) than novice PFAs.

5 Dataset Plan

Primary: Search for real ECG eye-tracking datasets (OpenNeuro, OSF, Zenodo, research publications).

Fallback: Synthesized dataset (200 scanpaths: 100 expert, 100 novice) based on clinical guidelines:

- *Expert*: 75% start at Lead II, sequential V1→V6 transitions, inferior clustering (II↔III↔aVF)
- *Novice*: Uniform initial distribution, higher entropy, less structured transitions

6 Expected Outcomes and Success Criteria

Deliverable	Success Criteria
PFA-based model	Classification accuracy >90%, significant entropy difference between expert/novice
Technical paper	Complete formal model, proofs, experiments following course guidelines
GitHub repository	Clean code, documentation, visualizations, reproducible results
Generation quality	Perplexity <5.0, patterns match expert characteristics
Completion accuracy	Top-3 accuracy >80% for next-fixation prediction

7 Potential Challenges

- **Data access**: Real clinical scanpaths may be unavailable; synthetic data limits ecological validity
- **Markov limitation**: First-order assumption may miss long-range dependencies in expert strategies
- **Granularity**: Lead-level model doesn't capture intra-lead patterns (P-wave vs. QRS)
- **Synthetic separability**: High accuracy on synthetic data may not generalize to real clinical data

8 Timeline

Nov 18–20	Team formation, proposal submission
Nov 25–Dec 3	Implementation, first draft
Dec 7	Final deliverables due
Dec 9–11	Oral presentations
Jan 22, 2026	CHI 2026 Posters submission

GitHub: `https://github.com/p3w-p3w-alpha/Computational-Theory-UM6PCC`