

# Algoritmi de Reprezentare Rară, Învățarea Dictionarelor și Reconstrucție Rară: Detectarea și Clasificarea Aritmiilor EKG pe 12 Derivații

Tudor Pistol și Teofil Simiraș

January 16, 2025

## Contents

## 1 TITLU ȘI CUPRINS

**Titlu propus pentru proiect:**

*„Algoritmi de Reprezentare Rară, Învățarea Dictionarelor și Reconstrucție Rară: Detectarea și Clasificarea Aritmiilor EKG pe 12 Derivații”*

**Cuprins (planificat pentru prezentare):**

1. Introducere și Context
2. Problema Abordată
3. Justificarea Problemei
4. Abordare Tehnică Propusă
5. Tehnologii și Biblioteci Folosite
6. Rezultate Așteptate și Metodologie de Evaluare
7. Concluzii și Direcții Potențiale de Extindere

## 2 PROBLEMA ABORDATĂ ÎN PROIECT

În cadrul acestui proiect, **intenționăm să detectăm și să clasificăm aritmiile cardiace** (cum ar fi fibrilația atrială, extrasistole etc.) în semnalele EKG cu **12 derivații**. Se știe că:

- Datele EKG de 12 derivații oferă o imagine foarte completă a activității electrice a inimii, dar sunt și mai voluminoase.
- Analiza clasică necesită implicarea intensivă a specialiștilor (cardiologi), iar variabilitatea umană poate duce la erori.

**Obiectiv principal:** Să dezvoltăm (în următoarea perioadă) un *pipeline automat* care, după preprocesarea și segmentarea bătailor cardiace, să aplice **Reprezentarea Rară (Sparse Coding)** pentru extragerea de caracteristici și apoi să folosească un **clasificator** (SVM, Logistic Regression etc.) pentru a identifica automat bătăile normale față de cele care prezintă aritmii cardiace.

## 3 JUSTIFICAREA PROBLEMEI ABORDATE

### 1. De ce e importantă?

- Aritmiile cardiace sunt frecvente și pot fi critice dacă nu sunt depistate la timp.
- Un sistem semi-automat sau automat pentru detecția aritmiilor cardiace reduce timpul de diagnostic și crește acuratețea.

### 2. Context și ce problemă rezolvă?

- În spitale se adună zilnic sute/mii de EKG-uri. Un algoritm robust ajută la trierea rapidă a pacienților care au nevoie de investigații suplimentare.
- În dispozitive portabile (Holter EKG, wearables), un algoritm cu cost computațional relativ scăzut poate alerta medicul sau pacientul în timp real.

### 3. Unde poate fi folosit?

- Clinici, centre de cardiologie, laboratoare de cercetare care lucrează cu analiza semnalelor cardiace.
- În aplicații de telemedicină și monitorizare la distanță (conectate la cloud).

## 4 CUM ESTE ABORDATĂ PROBLEMA TEHNIC

### 4.1 Preprocesare Semnal EKG

**Filtrare (Butterworth și Chebyshev):**

- Deoarece la curs am discutat despre filtre **Butterworth** și **Chebyshev**, vom folosi un *band-pass* (0.5–40 Hz) pentru a reține doar componentele relevante ale semnalului EKG.
- Butterworth oferă o tranziție mai lină și nu introduce ripple în banda de trecere.
- Chebyshev (dacă e nevoie de roll-off mai abrupt) introduce ripple, dar separă mai ferm zona filtrată.

**Segmentare pe bătăi cardiace:**

- Identificăm complexul QRS cu un algoritm (ex. Pan-Tompkins).
- Extragem o fereastră fixă (ex. 100 ms înainte și 300 ms după R-peak) pentru fiecare bătaie.

### Normalizare:

- Aducem amplitudinile în același interval (ex.  $[-1, 1]$ ) pentru a facilita învățarea.

## 4.2 Reprezentare Rară (Sparse Representation)

**Definiție:** Reprezentarea rară presupune exprimarea unui semnal  $y \in \mathbb{R}^m$  ca o combinație liniară a unui *număr mic de atomi* dintr-un dicționar  $D \in \mathbb{R}^{m \times n}$ :

$$y = Dx,$$

unde doar câțiva coeficienți din  $x$  sunt nenuli (suport rar).

### Probleme de optimizare pentru Sparse Coding

- Criteriu al erorii:

$$\min_x \|x\|_0 \quad \text{s.t.} \quad \|y - Dx\| \leq \epsilon.$$

- Criteriu al rarității (sparsity):

$$\min_x \|y - Dx\|^2 \quad \text{s.t.} \quad \|x\|_0 \leq s.$$

## 4.3 Algoritmi pentru Sparse Coding

### OMP (Orthogonal Matching Pursuit):

- (a) Este un algoritm *greedy* foarte utilizat.
- (b) Construiește iterativ suportul vectorului rar  $x$ :
  - (i) Inițializează reziduul  $e = y$  și suportul  $S = \emptyset$ .
  - (ii) Găsește atomul  $k$  care are cea mai mare corelație cu reziduul:

$$k = \arg \max_{j \notin S} |e^T d_j|.$$

- (iii) Actualizează suportul  $S = S \cup \{k\}$ . Apoi rezolvă problema de minim pe suportul  $S$ :

$$x_S = (D_S^T D_S)^{-1} D_S^T y,$$

iar restul componentelor  $x$  sunt zero.

- (iv) Recalculează reziduul:  $e = y - D_S x_S$ .
  - (v) Se oprește când  $\|e\| \leq \epsilon$  sau suportul atinge dimensiunea maximă  $s$ .

## 4.4 Antrenarea Dicționarelor (Dictionary Learning)

**Scop:** Obținerea unui dicționar  $D$  care să ofere reprezentări rare cât mai bune pentru setul de semnale EKG de antrenare  $Y$ :

$$\min_{D, X} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \|x_\ell\|_0 \leq s, \quad \forall \ell.$$

- $Y \in \mathbb{R}^{m \times N}$  – setul de antrenare (bătăi EKG).
- $D \in \mathbb{R}^{m \times n}$  – dicționarul.
- $X \in \mathbb{R}^{n \times N}$  – coeficienții rari (fiecare coloană e un  $x_\ell$ ).

#### Algoritmul K-SVD (pe scurt):

1. *Sparse coding*: Se calculează coeficienții  $x_i$  cu un algoritm (ex. OMP).
2. *Actualizarea dicționarului*:
  - Se calculează eroarea reziduală  $F = Y - DX$ .
  - Pentru fiecare atom  $d_j$ , se extrage doar partea de reziduu unde  $d_j$  e folosit; se aplică SVD pe acest subset și se obțin noile valori pentru  $d_j$  și coeficienții corespunzători.

### 4.5 Clasificarea Bătăilor EKG (Aritmii vs. Normale)

- După calcularea coeficienților rari ( $x$ ) pentru fiecare bătaie, concatenăm/folosim acești coeficienți ca *feature vector*.
- **Algoritm de clasificare**:
  1. **SVM (ex. kernel RBF)**: adesea folosit pentru probleme de clasificare binară (normal vs. aritmie).
  2. **Regresie Logistică**: metodă mai simplă, interpretabilă.
  3. (Opțional) **Random Forest, XGBoost** sau **Rețele Neuronale** pentru performanțe avansate.
- Se împarte setul de date (bătăi EKG) în **train/test**, se antrenează clasificatorul pe setul **train** și se evaluează pe **test**.

### 4.6 Aplicații ale Reprezentării Rare în EKG

- **Clasificare (arie principală)**: Bătăi normale vs. aritmii, folosind coeficienții rari.
- **Denoising**: Zgomotul din semnal poate fi redus prin reconstrucție.
- **Compression / Inpainting**: Reconstrucția zonelor lipsă din EKG.

## 5 REZULTATE AȘTEPTATE ȘI METODOLOGIE DE EVALUARE

### 1. Scenariu de Antrenare/Test

- Setul de date EKG se împarte în bătăi cardiace normale și bătăi cu aritmii.
- Se antrenează dicționarul (K-SVD) și clasificatorul (SVM / Logistic Regression etc.) pe setul de antrenare, se validează și se testează ulterior.

## 2. Metrici de evaluare

- Acuratețe, Sensibilitate (Recall), Specificitate, F1-score, ROC/AUC.
- Evaluarea reconstrucției:  $\|Y - DX\|_F$  (dacă urmărim și calitatea reconstrucției).

## 3. Rezultate așteptate

- Acuratețe de peste 90% în detecția aritmiilor.
- Reducerea semnificativă a zgomotului prin sparse coding.
- O separare clară a bătailor anormale față de cele normale în spațiul coeficienților rari.

# 6 CONCLUZII ȘI DIRECȚII POTENȚIALE DE EXTINDERE

## Concluzie Principală

Reprezentarea rară (Sparse Coding) permite extragerea de trăsături esențiale din semnalul EKG pe 12 derivații, facilitând **clasificarea aritmiilor cardiace** cu acuratețe ridicată și oferind posibilități de *denoising*, *inpainting* și *compresie*.

## Direcții posibile de extindere

- **Abordare multi-clasă:** tipuri variate de aritmii (fibrilație atrială, flutter, bloc AV etc.).
- **Implementare embedded:** pe dispozitive Holter cu latență redusă.
- **Sparse Autoencoder:** în loc de OMP + K-SVD, putem folosi un autoencoder cu constrângeri de raritate.

## Bibliografie

- [1] *A large scale 12-lead electrocardiogram database for arrhythmia study*, <https://physionet.org/content/ecg-arrhythmia/1.0.0/>
- [2] *Heart Arrhythmias*, [https://www.physio-pedia.com/Heart\\_Arrhythmias](https://www.physio-pedia.com/Heart_Arrhythmias)
- [3] *Cursul de Procesarea Semnalelor*, <https://cs.unibuc.ro/~crusu/ps/index.html>
- [4] *Cursul de Calcul Numeric*, <https://numeric.cs.unibuc.ro/cni.html>