

Classification de l'état de santé du fœtus

basée sur des données de Cardiotocographie

Timofey ABRAMOV ▪ Yazan EL MAHMOUD ▪ Selima KHESSAIRI
Projet SY09 — Printemps 2025

Contexte & Problématique

Enjeu de santé publique

Selon l'ONU, près de **300 000 femmes** décèdent chaque année de complications liées à la grossesse. La cardiotocographie (CTG) permet de monitorer la santé fœtale, mais son interprétation reste complexe et subjective.

Objectif de l'étude

- ✓ Prédire efficacement la santé fœtale à partir de données CTG
- ✓ Améliorer la détection précoce des situations à risque
- ✓ Réduire la subjectivité de l'interprétation médicale

⚠ Déséquilibre des classes

- Normal : **77,84%** (1655 cas)
- Suspect : **13,87%** (295 cas)
- Pathologique : **8,27%** (176 cas)

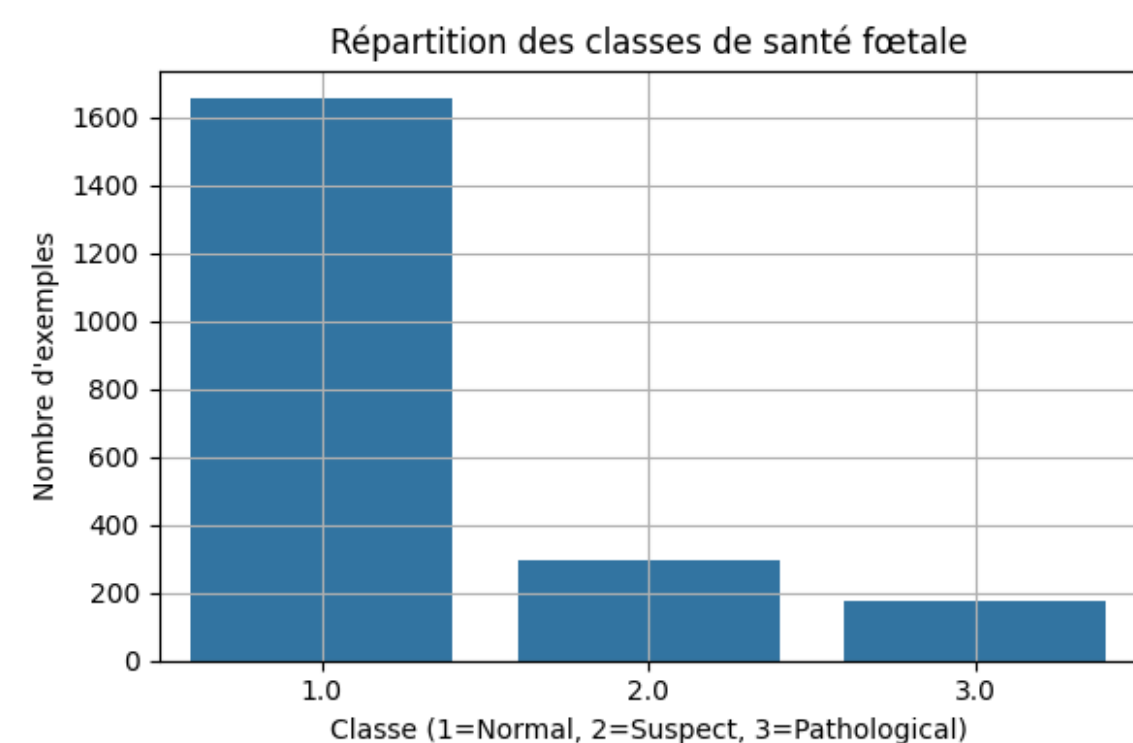


Figure 1: Distribution de la variable cible (n=2126)

Jeu de Données

Fetal Health Classification Dataset

- 2126 enregistrements CTG
- 21 variables descriptives numériques
- Données complètes (pas de valeurs manquantes)
- Données cliniques réelles collectées en milieu hospitalier

Catégories de variables

- Rythme cardiaque fœtal** : valeur basale, variabilité, accélérations/décélérations
- Mesures utérines** : contractions, durée, valeur maximale
- Autres indicateurs** : mouvements fœtaux, activité anormale

Variables les plus corrélées à la cible

- prolongued_decelerations (0.48)
- abnormal_short_term_variability (0.47)
- percentage_of_time_with_abnormal_long_term_variability (0.43)

Analyse Exploratoire

Analyse en Composantes Principales

- Les 2 premières composantes expliquent **45,55%** de la variance
- Séparation partielle des classes observée
- Chevauchement important entre classes 'suspect' et autres

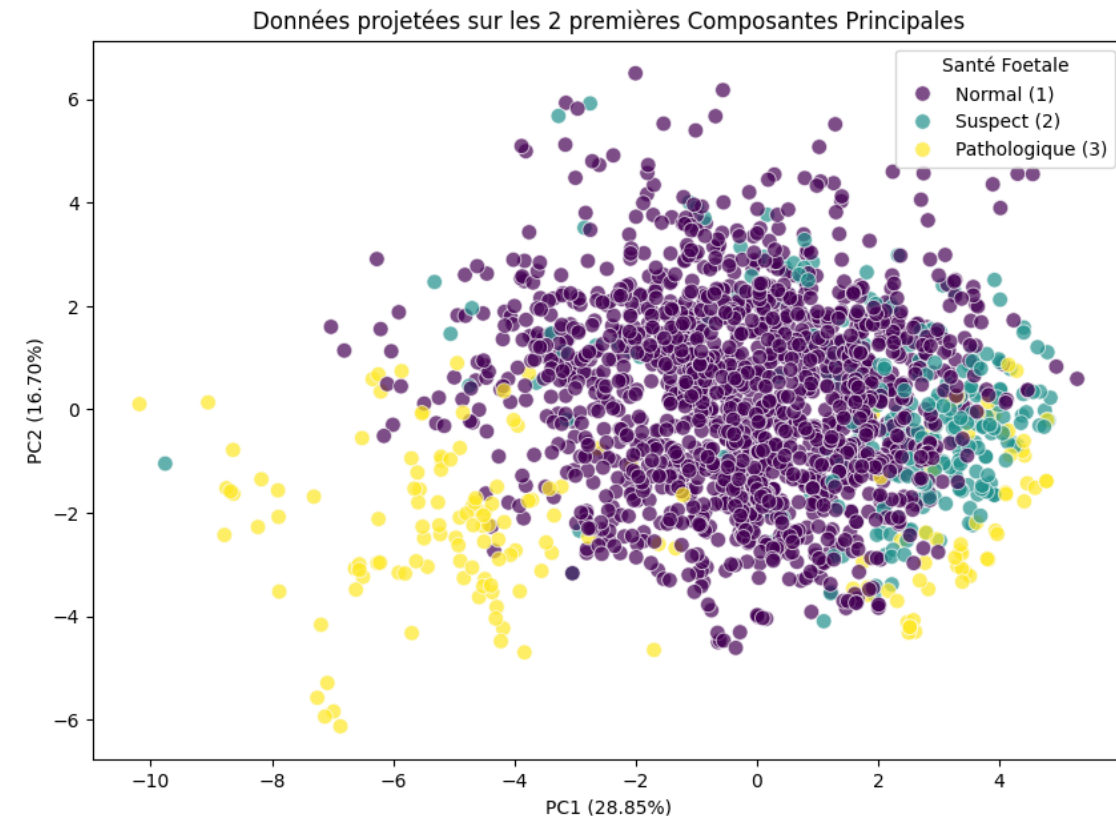


Figure 2: Projection ACP - Premier plan factoriel

Approches Non Supervisées

✗ Échec des méthodes non supervisées

- K-means** : Indice de Rand ajusté très faible (0.045)
- CAH** : Indice de Rand ajusté très faible (0.15)
- Causes** : Chevauchement des classes, données non sphériques, haute dimensionnalité

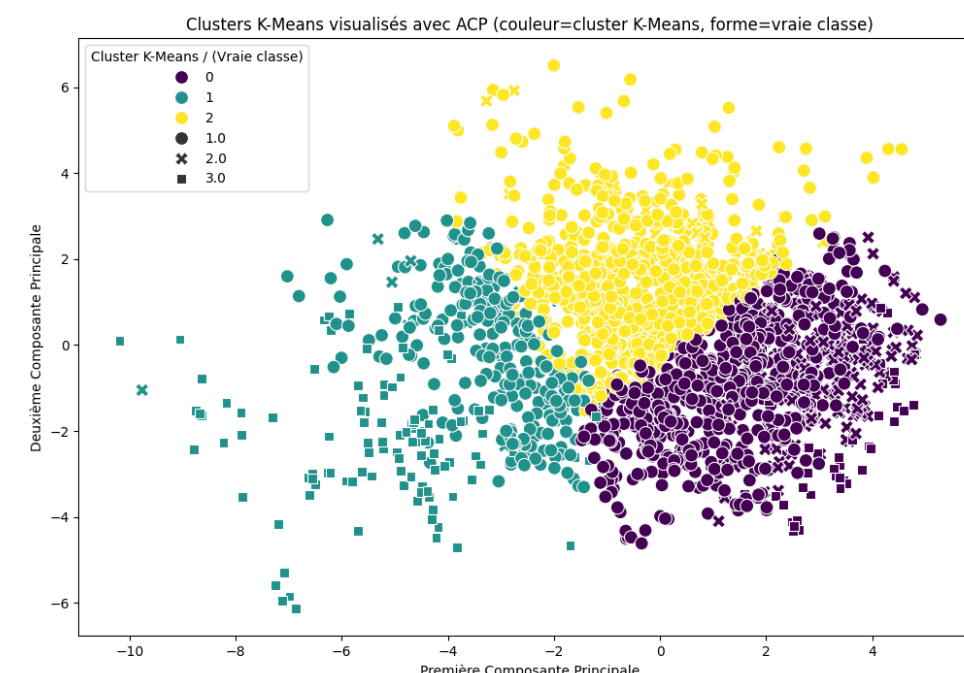


Figure 3: Comparaison K-means vs vraies classes

Approches Supervisées

Stratégie d'évaluation

- ✓ Données standardisées et stratifiées
- ✓ Focus sur la détection des cas pathologiques
- ✓ Prise en compte des coûts d'erreur asymétriques

Méthodes testées : K Plus Proches Voisins (k=4 optimal) • Naïve Bayes Gaussien
Analyse Discriminante (LDA/QDA) • Régression Logistique

Résultats des Modèles

Performances par algorithme

Modèle	Accuracy	Prec. Patho.	Rappel Patho.
Rég. Logistique	88,5%	88%	66%
KPPV (k=4)	90,8%	84%	77%
Arbres	89,9%	81%	83%
RF	92%	87%	87%

Performances par algorithme

Modèle	Accuracy	Prec. Patho.	Rappel Patho.
LDA	85,9%	67%	46%
QDA	81,2%	76%	37%
Naive Bayes	81,0%	53%	46%

Analyse détaillée - Régression Logistique

Table 1: Matrice de confusion - Régression Logistique

	Prédit Normal	Prédit Suspect	Prédit Pathologique
Vrai Normal	314	17	1
Vrai Suspect	16	40	3
Vrai Pathologique	3	9	23

Points forts de la régression logistique multinomiale

- ✓ Excellente reconnaissance des cas normaux (95% rappel)
- ✓ Précision remarquable sur les cas pathologiques (88%)
- ⚠ Mauvaise reconnaissance des cas suspects

Approche Coût-Sensible

Table 2: Matrice de coûts utilisée

	Prédit Normal	Prédit Suspect	Prédit Pathologique
Vrai Normal	0	0.5	2
Vrai Suspect	1	0	1
Vrai Pathologique	5	4	0

Stratégie médicale adaptée

Privilégier les faux positifs plutôt que les faux négatifs critiques. Mieux vaut suspecter à tort une pathologie que de la manquer.

Conclusions & Perspectives

Contributions majeures

- ✓ **Solution efficace** pour un problème médical complexe
- ✓ Approche **interprétable** pour les praticiens
- ✓ Méthodologie **adaptée aux enjeux cliniques** (coûts asymétriques)
- ✓ Validation rigoureuse sur **données réelles**

⚠ Limites actuelles

- Performance modérée sur les cas suspects
- Besoin de validation prospective
- Test de Wald non effectué

Références

- [0] ONU (2024). Statistiques sur la mortalité maternelle mondiale.
- [0] Fetal Health Classification Dataset. UCI Machine Learning Repository.
- [0] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *JMLR*, 12, 2825–2830.