



Sciences Economiques et Sociales de la Santé  
& Traitement de l'Information Médicale

[sesstim.univ-amu.fr](http://sesstim.univ-amu.fr)

**Nicolas NGO**  
Ingénieur d'études  
UMR 1252 SESSTIM

**Construction de méthodes de sélection de variables en utilisant un nouvel indice de séparabilité : la gamma-metric.**

**Septembre 2020**



**[Cliquez ici pour voir l'intégralité des ressources associées à ce document](#)**

# Construction de méthodes de sélection de variables en utilisant l'indice de séparabilité : la $\gamma$ -metric

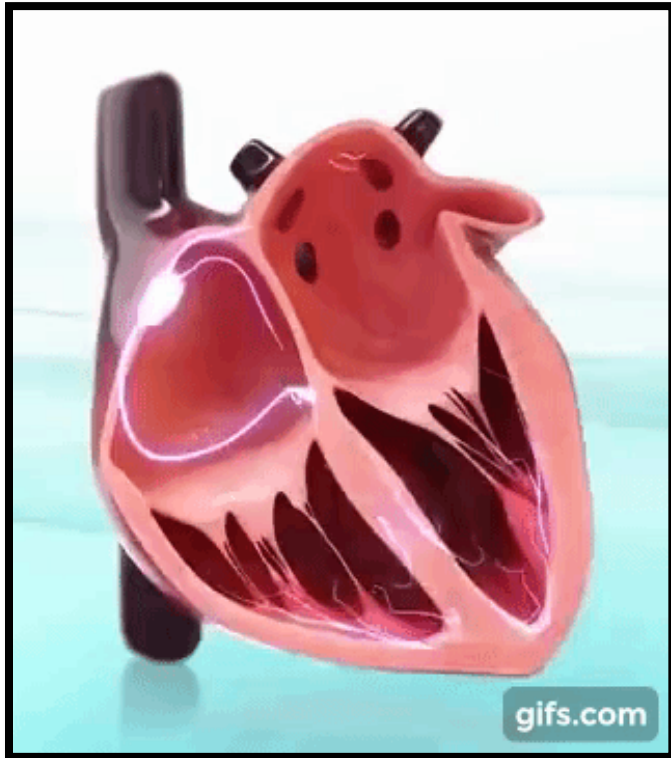
Ngo N., Michel P., Pons JF., Delliaux S., Giorgi R.

# Plan

- Contexte
- Objectif
- Matériel et méthodes
- Résultats
- Discussion et conclusions

# Contexte

## *Rythme sinusal*

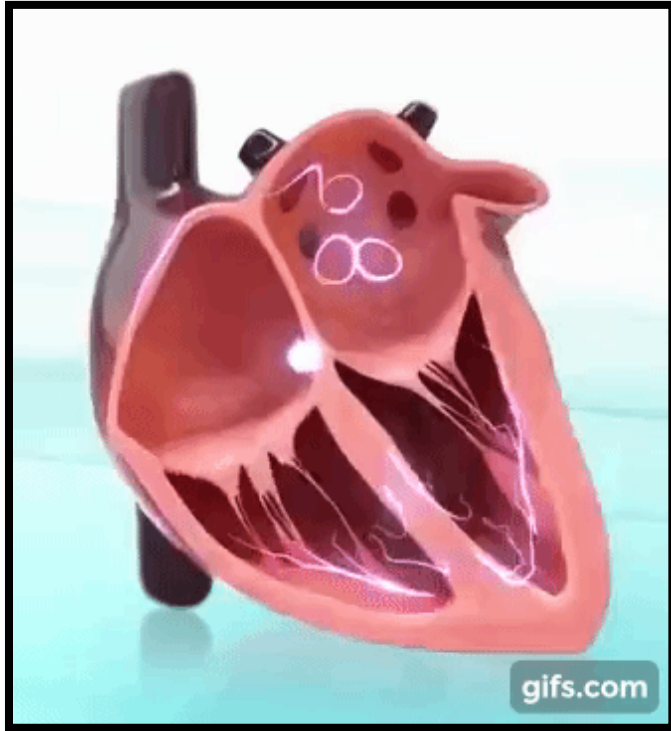


- Courant partant du nœud sinusal,
  - Il se propage dans les oreillettes et les ventricules,
  - Provoque la contraction des chambres du cœur,
  - Le sang est éjecté vers les organes du corps
- 
- Contractions ordonnées et à un rythme régulier

➔ Rythme sinusal (RS)

# Contexte

## *Fibrillation atriale*

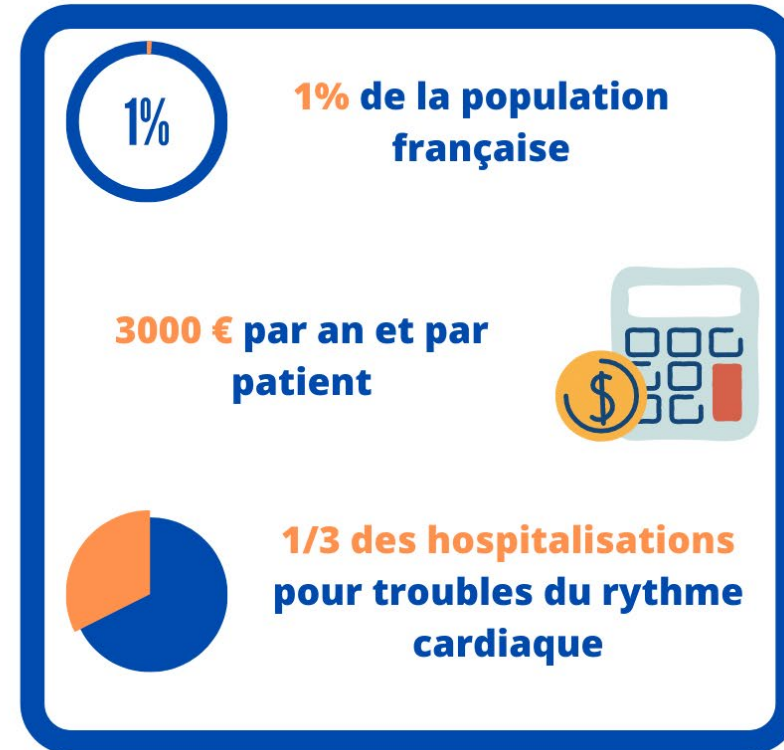
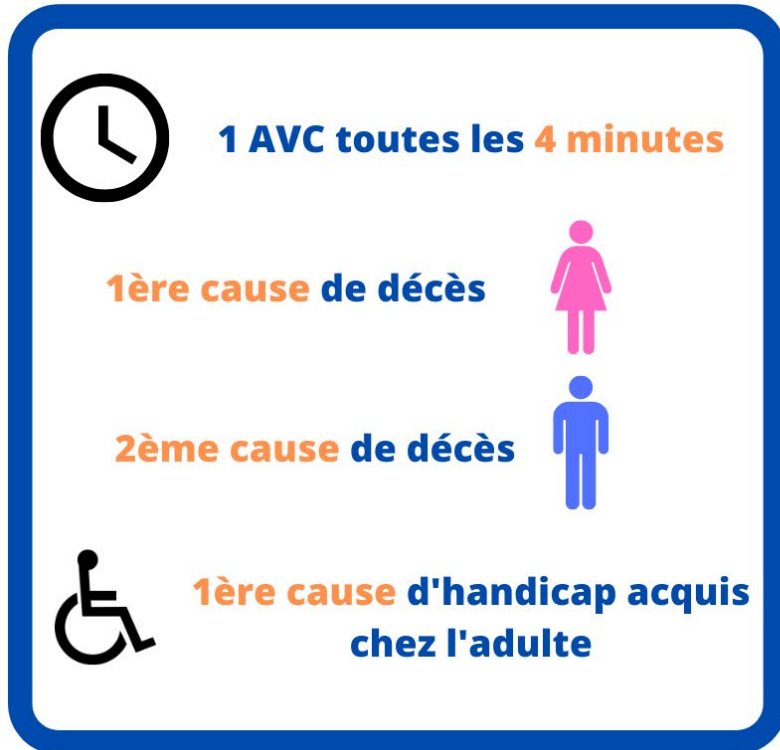


- Courant partant de plusieurs points dans les oreillettes,
  - Provoque des contractions chaotiques
  - Perte d'efficacité de l'évacuation du sang
- 
- Contractions irrégulières et généralement plus rapides

➔ Fibrillation atriale (FA)

# Contexte

## Enjeux de la détection de la FA



# Contexte

## Diagnostic



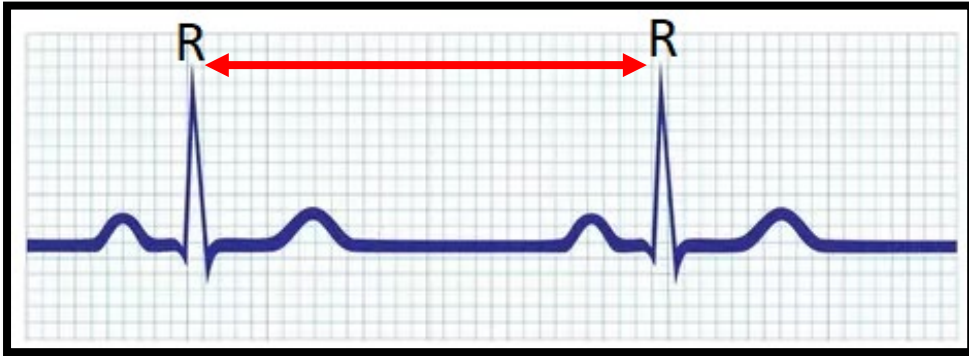
- Détecter le plus tôt possible la fibrillation

➡ Électrocardiogramme (ECG)

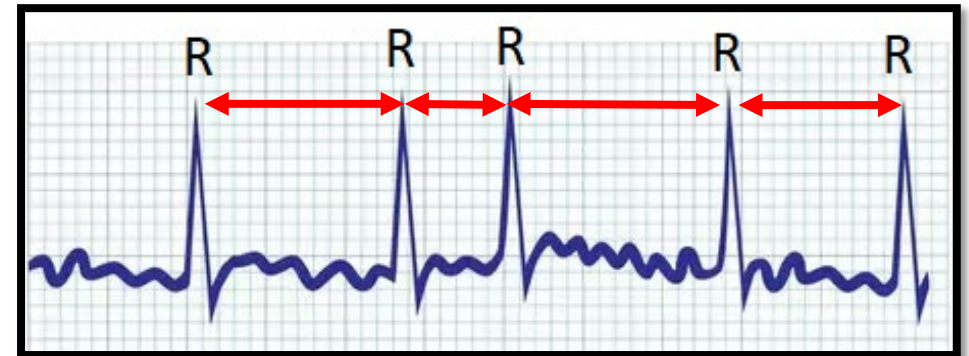
# Contexte

## *De l'ECG au modèle de classification*

Rythme sinusal



Fibrillation atriale



- Construction de variables à partir des intervalles RR
- Sélection de variables
- Modèle de classification



# Objectif

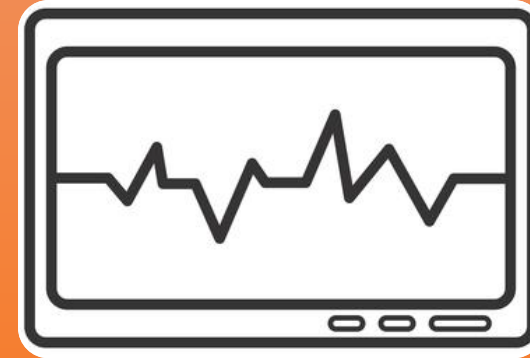
Utiliser la  $\gamma$ -metric comme fonction d'évaluation dans des algorithmes de sélection de variables, au vu de détecter la fibrillation atriale à partir de données d'électrocardiogramme.

# Matériel et méthodes

## *Données*



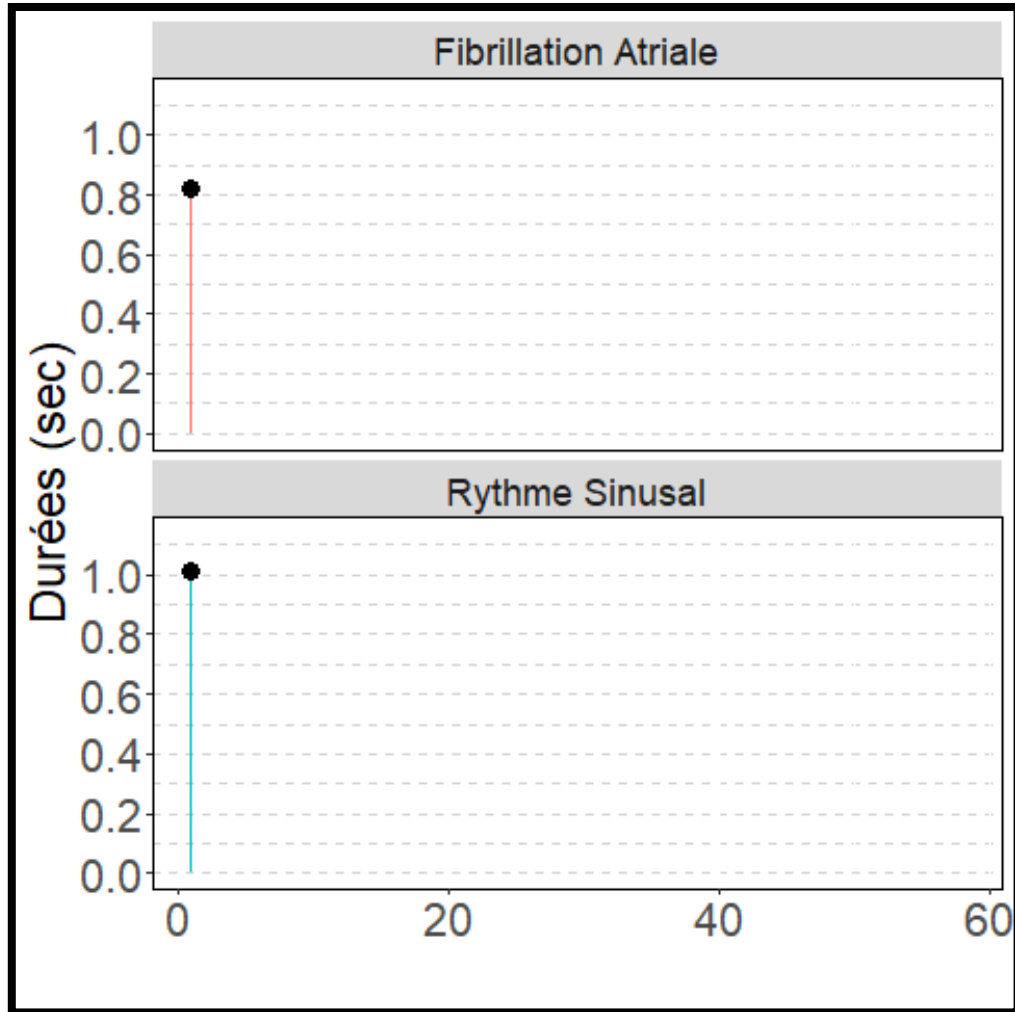
- 50 000 séries d'intervalles RR d'une minute
- 97 patients
- 5,8% des séries correspondent à un épisode de FA



- Enregistrements d'Holter de La Timone
- 34 patients
- 40 000 séries d'intervalles RR
- 26,7% des séries correspondent à un épisode de FA

# Matériel et méthodes

## *Construction de variables*



- Exploiter les différences entre un signal de FA et de RS
  - Variables dédiées à l'analyse de la variabilité du rythme cardiaque
  - Variables liées à la monotonie
  - Variables liées à l'autocorrélation de la série
- Utilisation des taux d'accroissement des séries d'intervalles RR
  - Moyenne, écart-type, étendue, coefficient de variation etc...
  - Variabilité des taux d'accroissement

➔ 111 variables explicatives construites

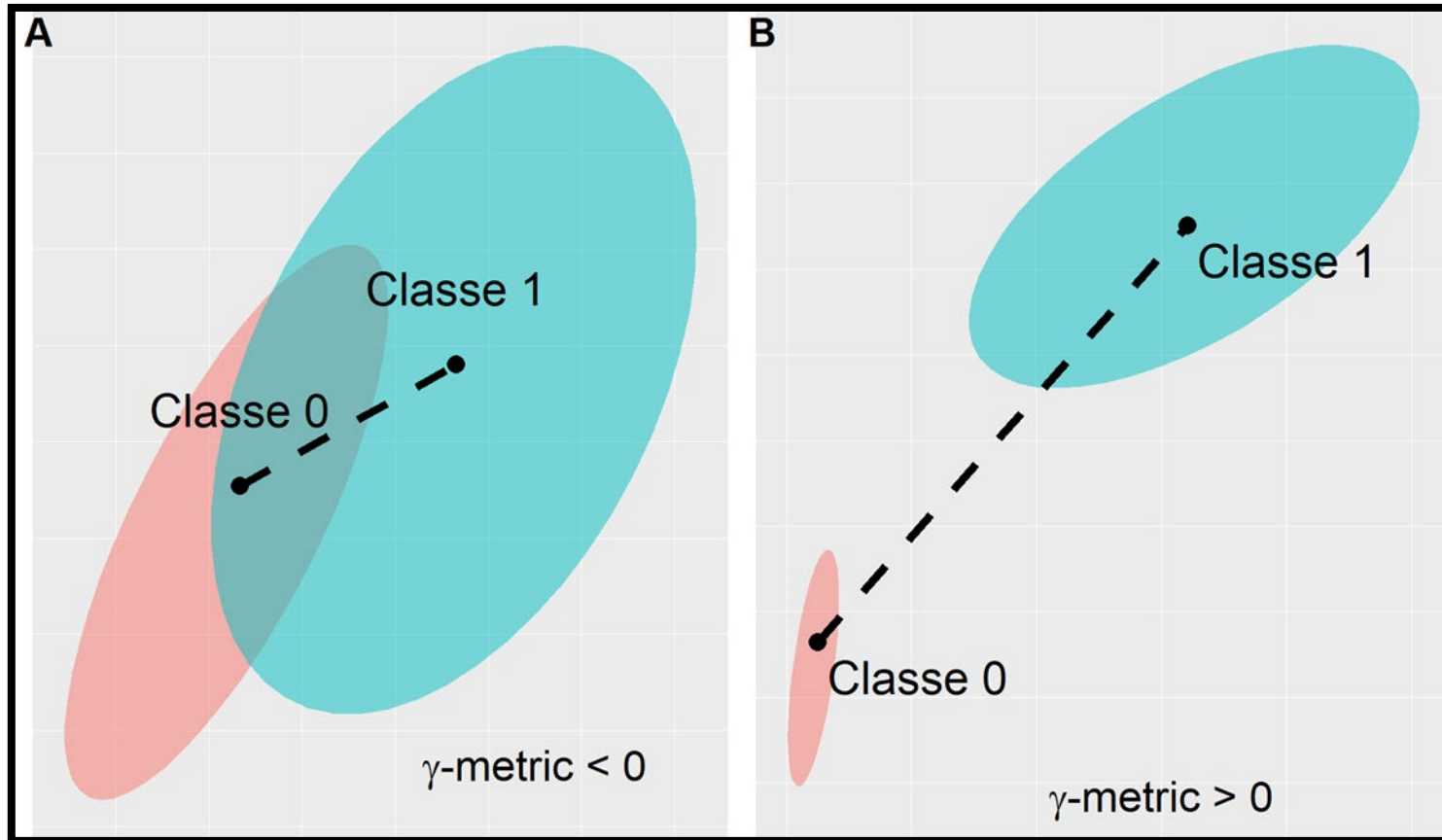
# Matériel et méthodes

## *Sélection de variables*

- Algorithme de sélection de variables :
  - Type d'approche
  - Direction de recherche
  - Fonction d'évaluation
  - Critère d'arrêt
- Recherche exhaustive trop coûteuse ( $2^p - 1$  combinaisons possible)
- Direction de recherche **Forward, Backward, Hill climbing** et **Best first**
- Utilisation de la  $\gamma$ -metric comme fonction d'évaluation

# Matériel et méthodes

## $\gamma$ -metric



- Construction d'une matrice de variance-covariance par classe,
- Construction d'une ellipsoïde par classe à l'aide des vecteurs et valeurs propres de ces matrices,
- Mesure de la séparation entre le centre des ellipsoïdes

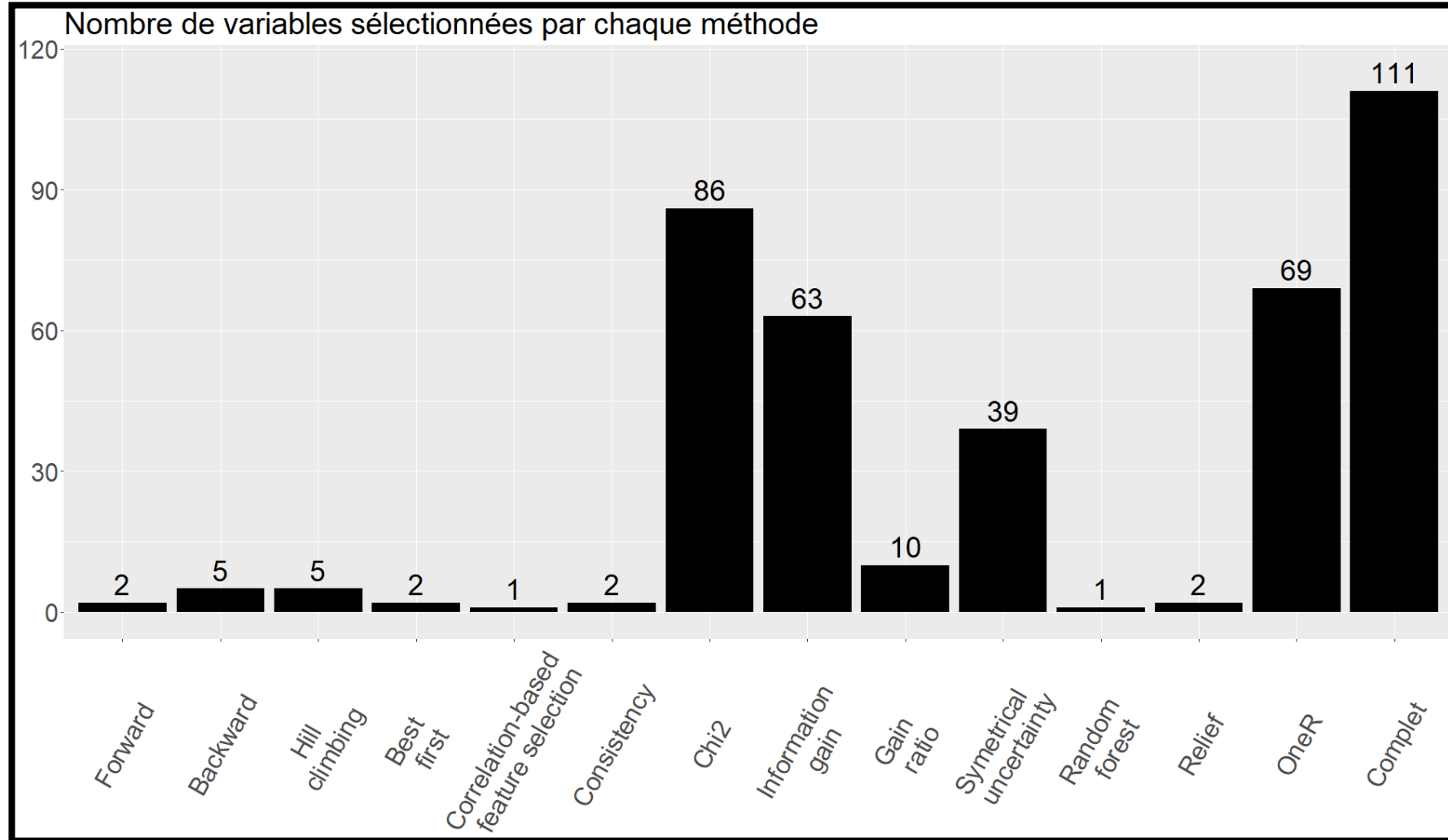
# Matériel et méthodes

## *Plan d'analyse*

1. Séparation du jeu de données (train/test)
2. Sélection de variables (13 méthodes)
3. Construction de modèles de classification à partir des variables sélectionnées par chaque méthode (13 modèles + 1 modèle complet)
4. Calcul des indicateurs de performance sur l'échantillon d'apprentissage
5. Calcul des indicateurs de performance sur l'échantillon de validation externe

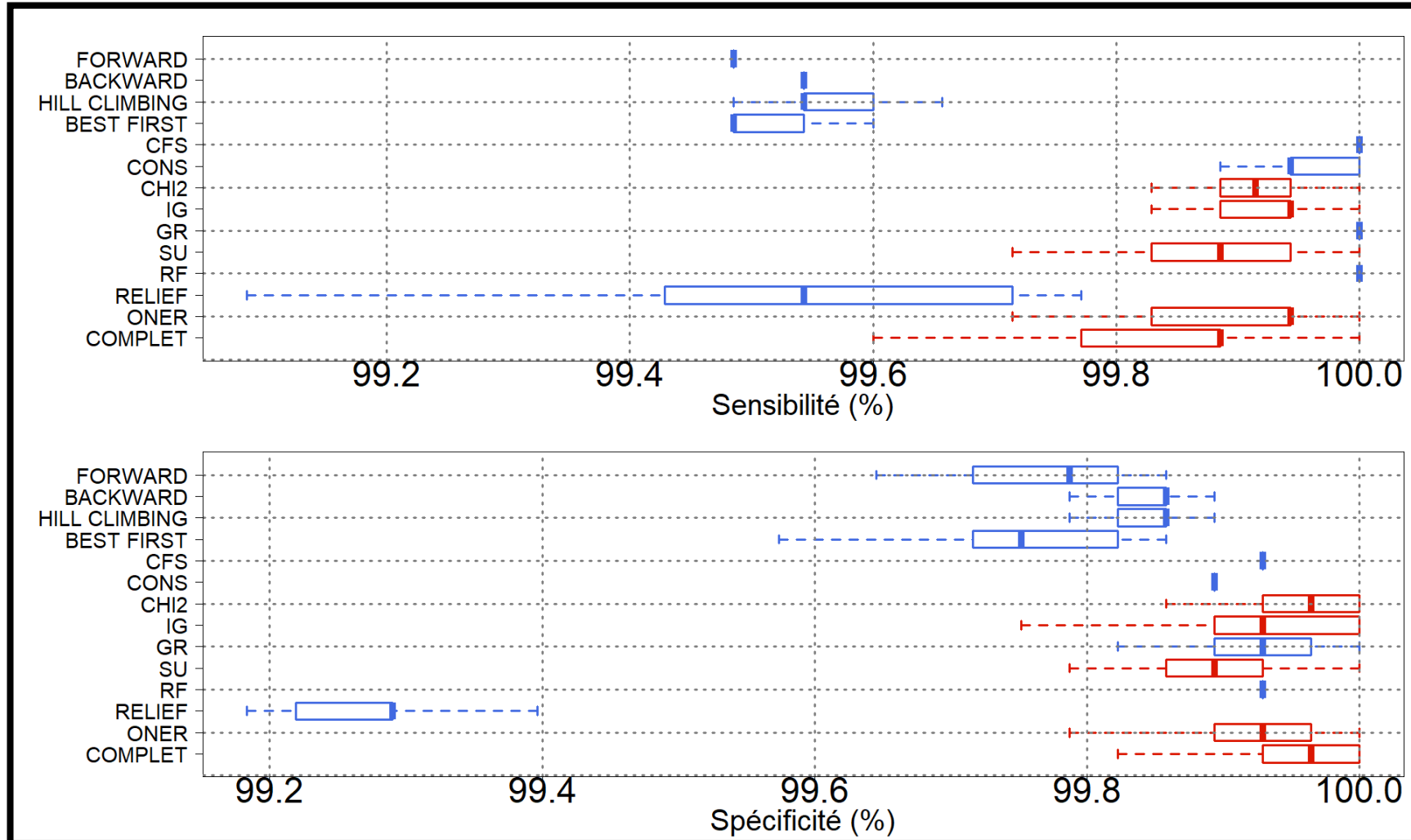
# Résultats

## Sélection de variables



# Résultats

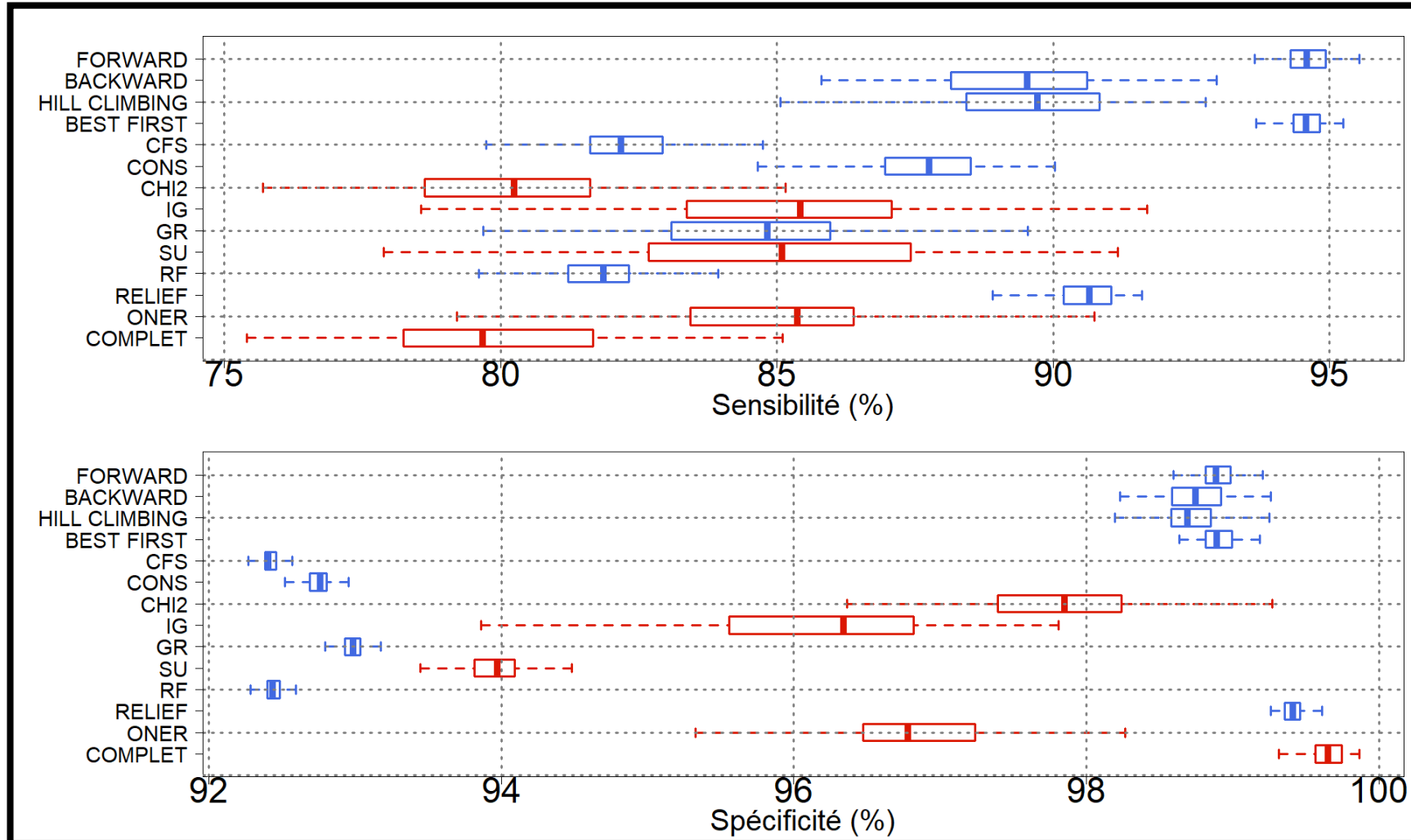
## Classification (apprentissage)





# Résultats





## Classification (validation externe)



# Discussions et conclusion

- Réduction de la dimension (2 à 5 variables sélectionnées)
- Bonnes performances en classification :
  - Sensibilité et spécificité supérieures à 99% sur l'échantillon d'apprentissage
  - Backward et Hill climbing ont obtenu la meilleure sensibilité sur l'échantillon de validation externe

# Références

-  GOLDBERGER, AL, AMARAL, LA, GLASS, L et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000 ;101 :e215–e220.
-  MARK, R, SCHLUTER, P, MOODY, G, DEVLIN, P et CHERNOFF, D. An annotated ECG database for evaluating arrhythmia detectors. In : *IEEE Transactions on Biomedical Engineering*. T. 29. 8. IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC 345 E 47TH ST, NEW YORK, NY ... 1982 :600–600.
-  PONS, JF, HADDI, Z, DEHARO, JC et al. Heart rhythm characterization through induced physiological variables. *Scientific Reports* jan. 2017 :23.
-  MICHEL, P, PONS, JF, DELLIAUX, S et GIORGI, R. A multivariate filter approach for feature selection in classification. Submitted in *Statistical Methods in Medical Research*. 2018.

Merci pour votre attention !

Contact : [nicolas.NGO@univ-amu.fr](mailto:nicolas.NGO@univ-amu.fr)

