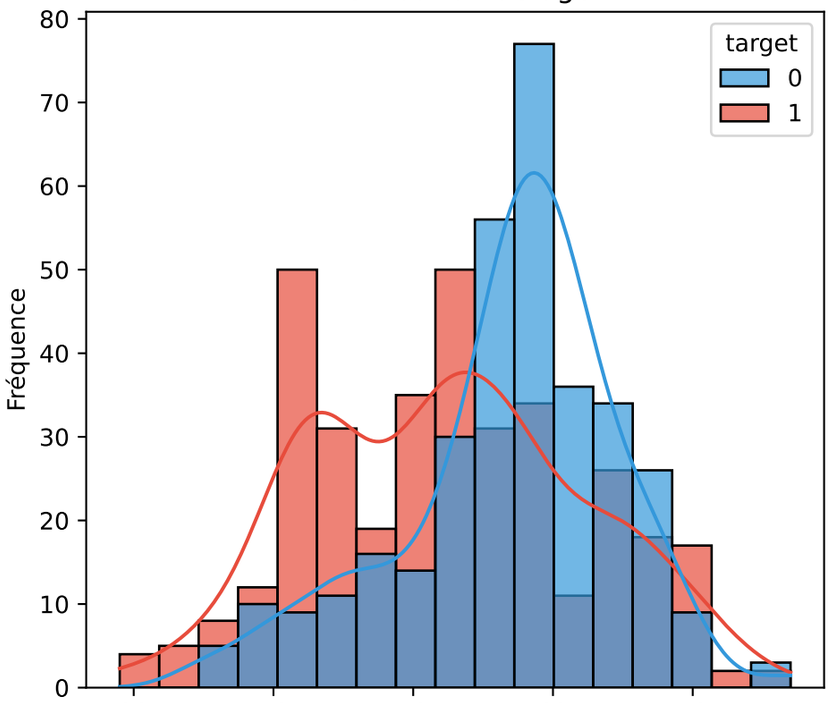
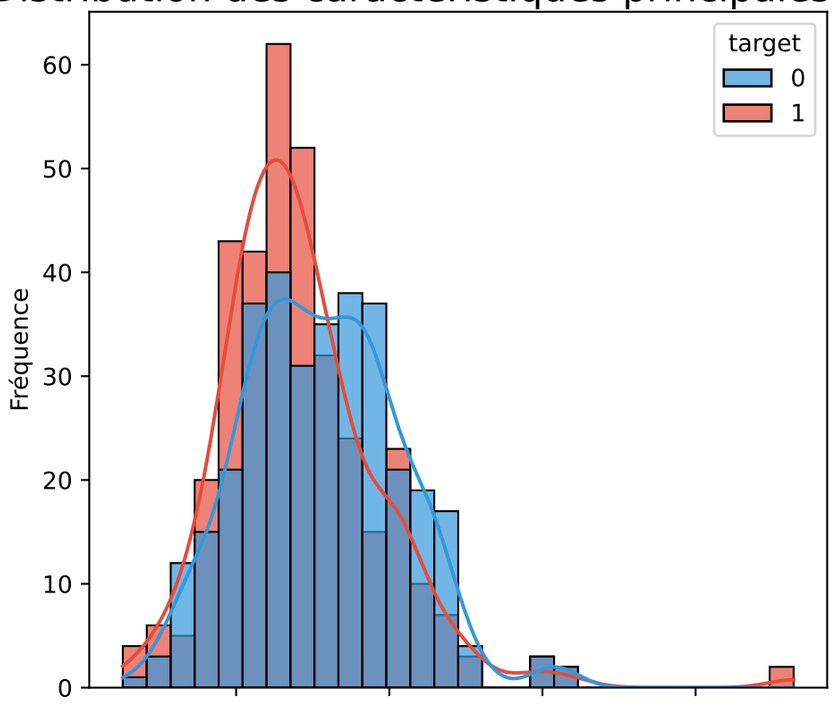
Analyse et Prédiction des Maladies Cardiaques

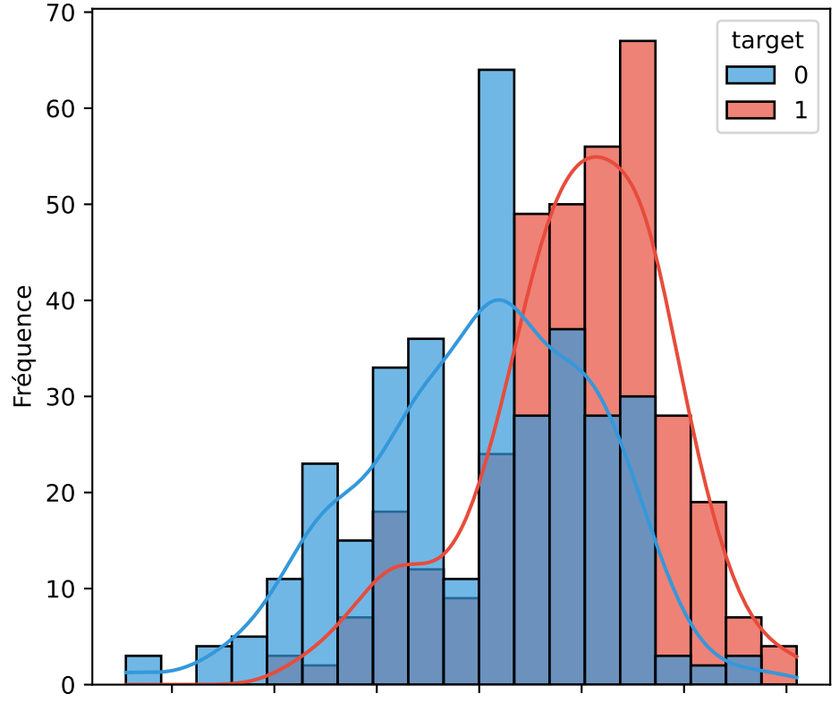
Rapport des Visualisations (Version Corrigée)

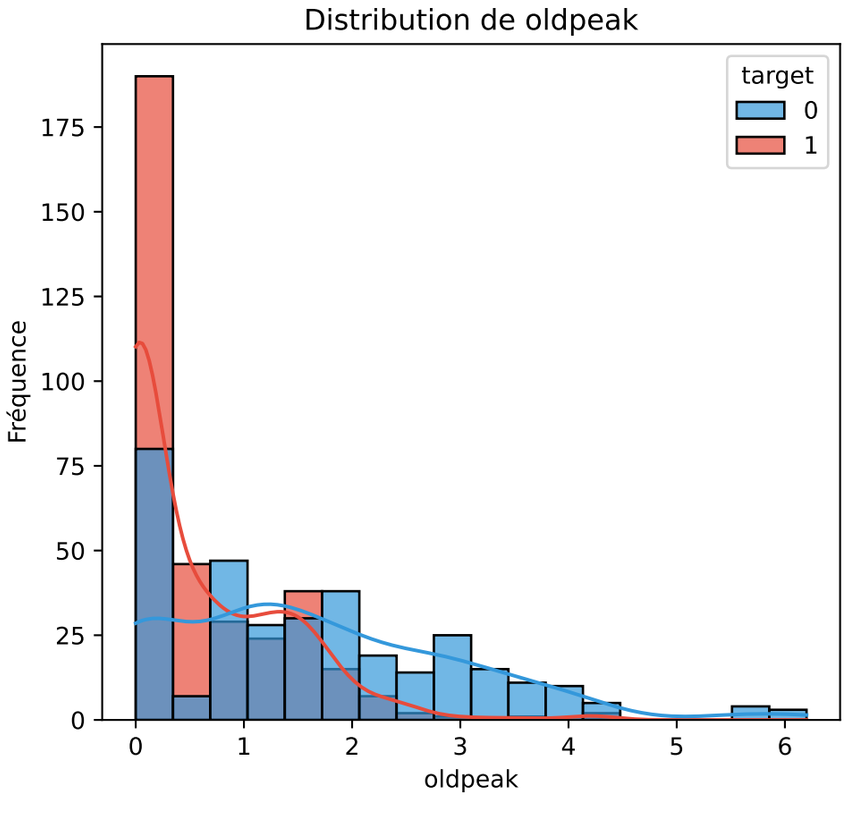
Date: Avril 2025

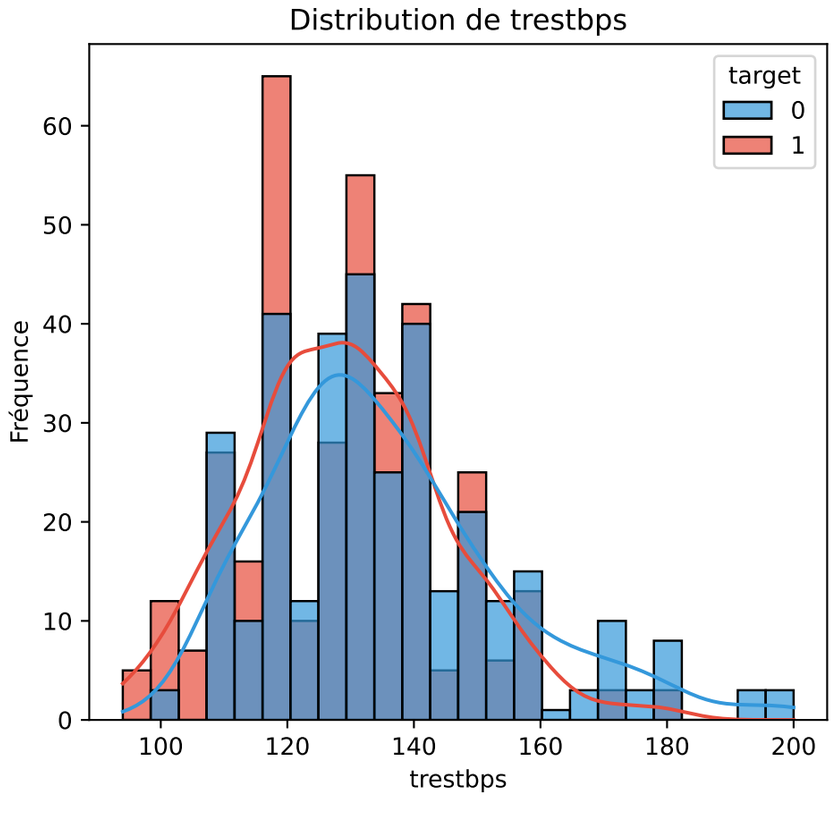
Distribution des caractéristiques principales

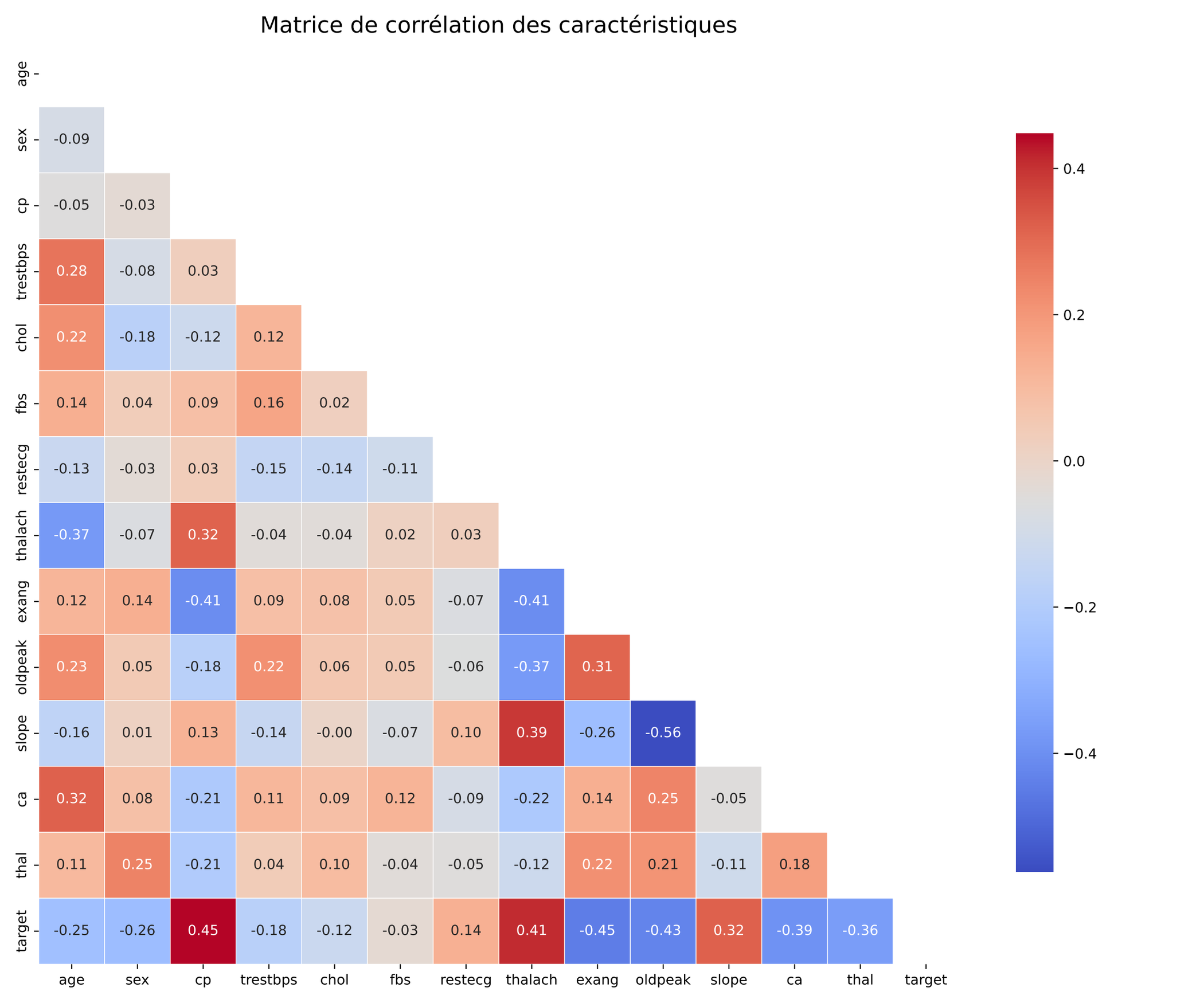


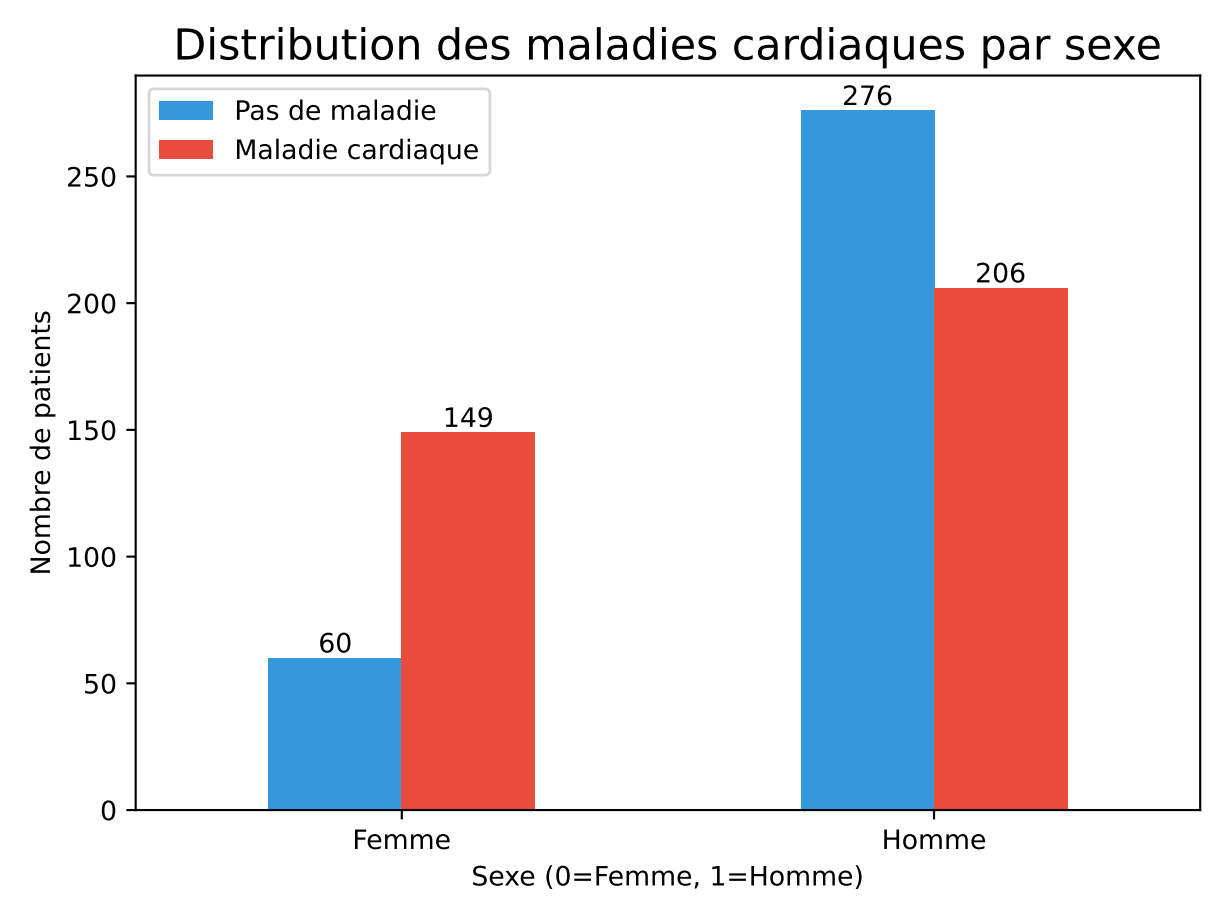


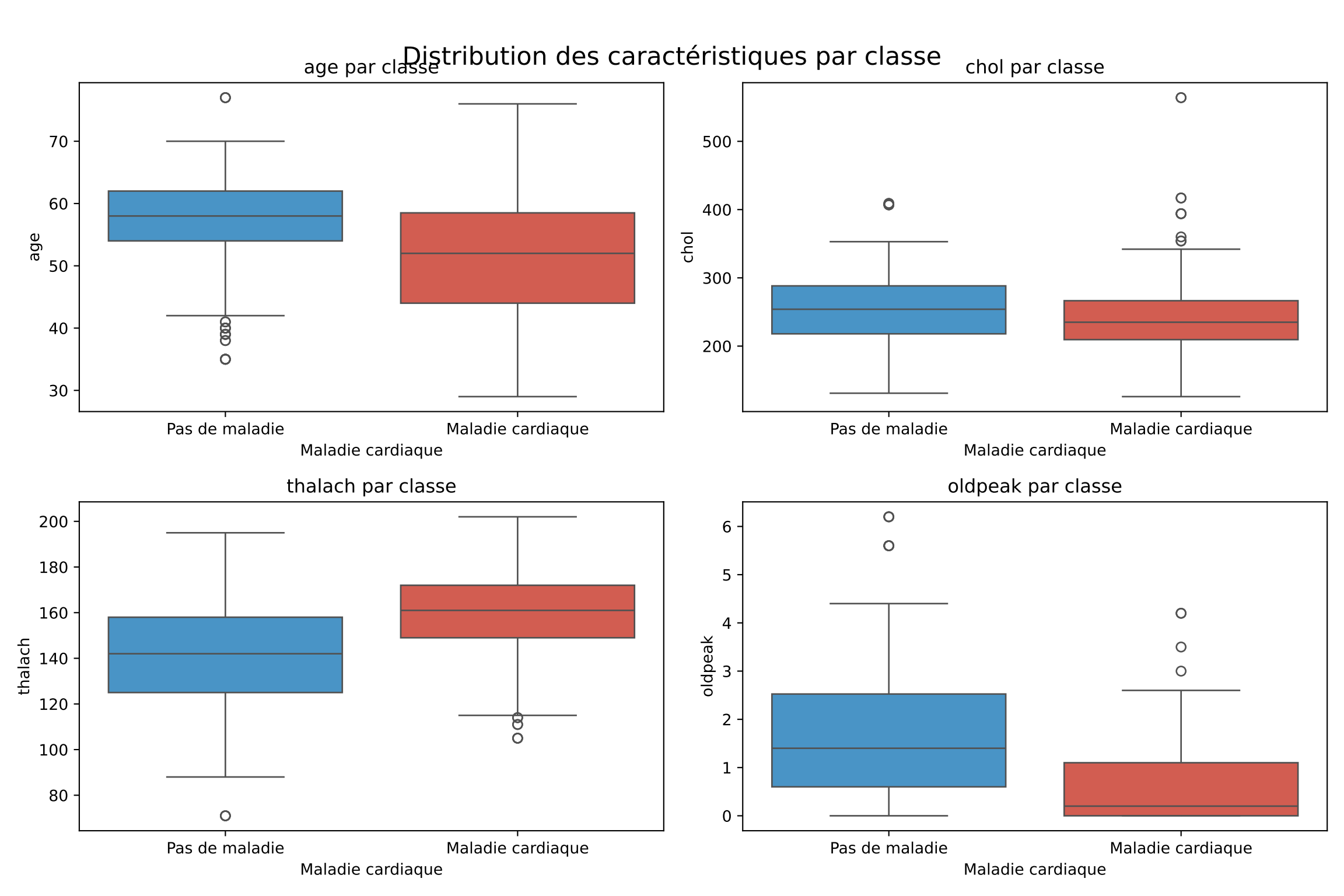




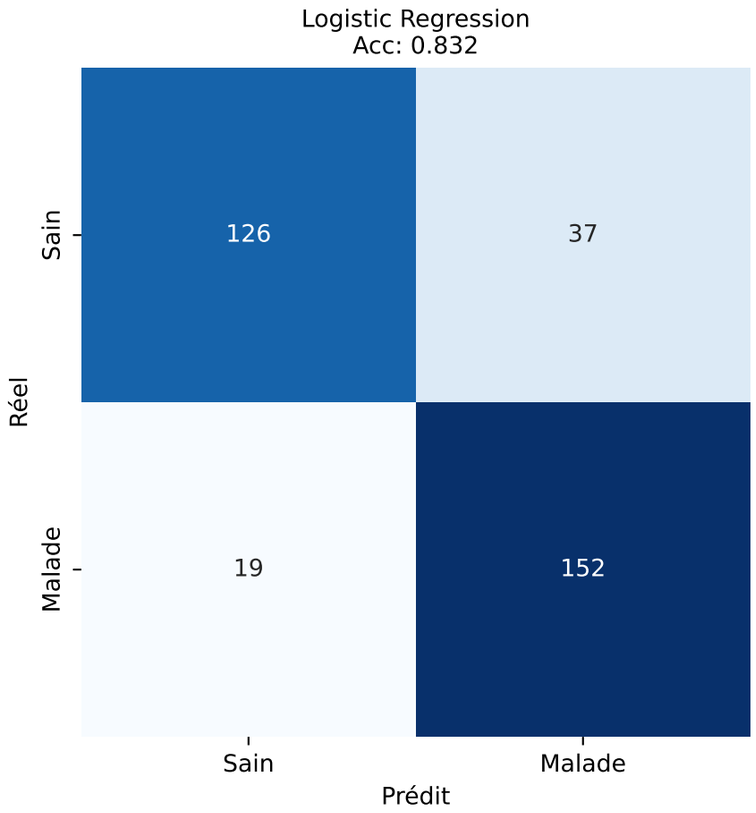


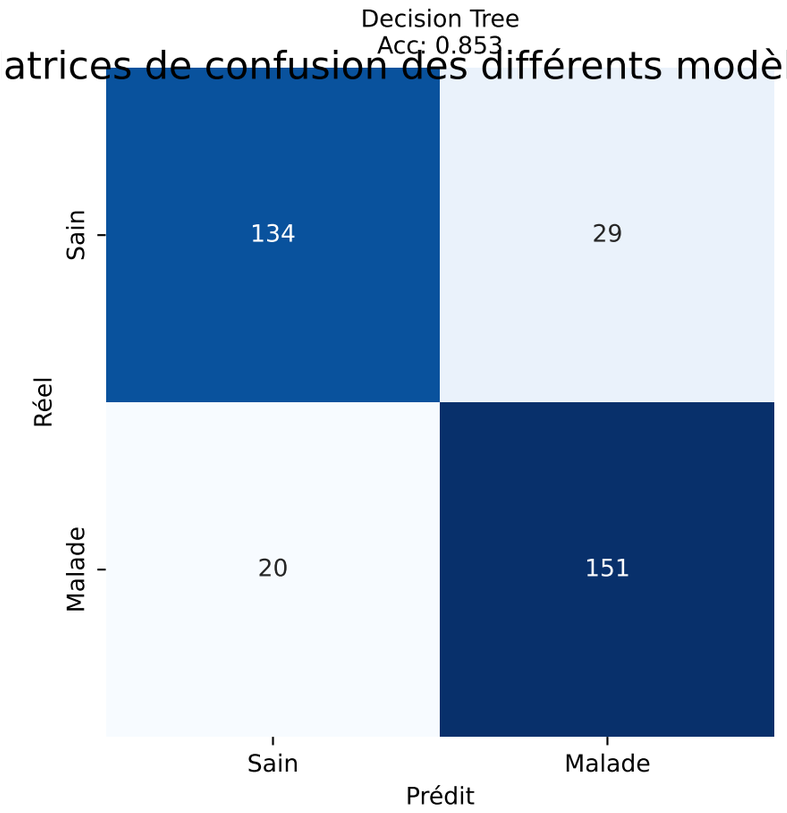

Matrice de corrélation des caractéristiques
age
sex -0.09
cp -0.05 -0.03
trestbps 0.28 -0.08 0.03
chol 0.22 -0.18 -0.12 0.12
fbs 0.14 0.04 0.09 0.16 0.02
restecg -0.13 -0.03 0.03 -0.15 -0.14 -0.11
thalach -0.37 -0.07 0.32 -0.04 -0.04 0.02 0.03
exang 0.12 0.14 -0.41 0.09 0.08 0.05 -0.07 -0.41
oldpeak 0.23 0.05 -0.18 0.22 0.06 0.05 -0.06 -0.37 0.31
slope -0.16 0.01 0.13 -0.14 -0.00 -0.07 0.10 0.39 -0.26 -0.56
ca 0.32 0.08 -0.21 0.11 0.09 0.12 -0.09 -0.22 0.14 0.25 -0.05
thal 0.11 0.25 -0.21 0.04 0.10 -0.04 -0.05 -0.12 0.22 0.21 -0.11 0.18
target -0.25 -0.26 0.45 -0.18 -0.12 -0.03 0.14 0.41 -0.45 -0.43 0.32 -0.39 -0.36
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
0.4
0.2
0.0
−0.2
−0.4

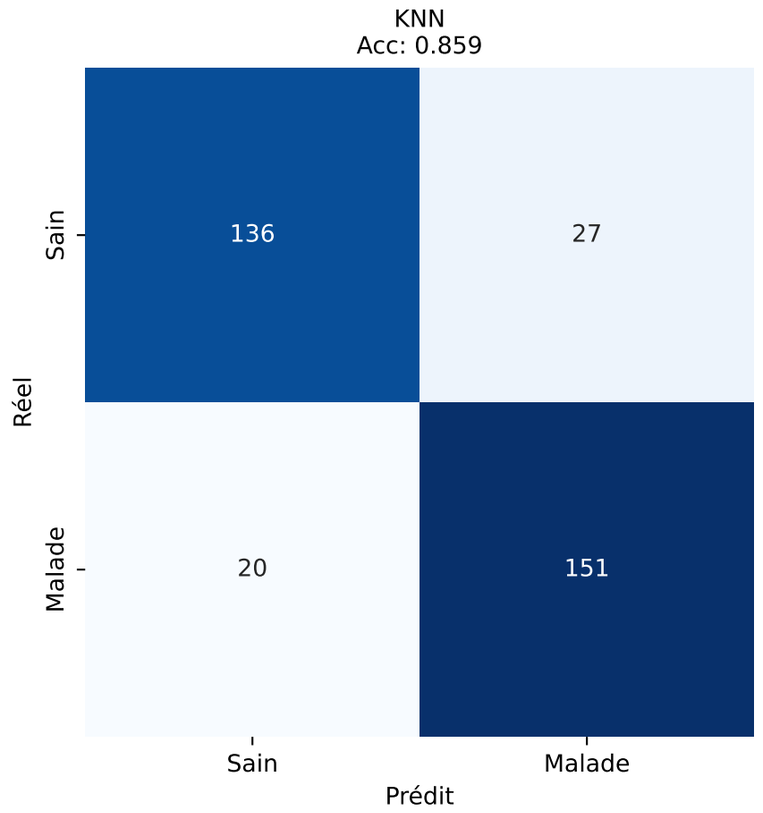

Distribution des maladies cardiaques par sexe
Pas de maladie
Maladie cardiaque
Nombre de patients
0
50
100
150
200
250
60
149
276
206
Femme
Homme
Sexe (0=Femme, 1=Homme)

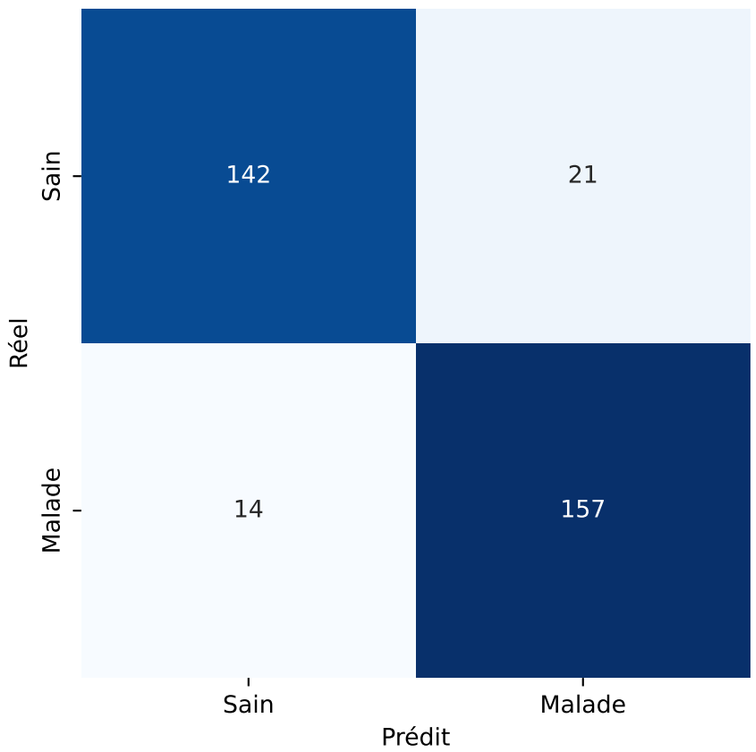

Distribution des caractéristiques par classe
age par classe
age
70
60
50
40
30
Pas de maladie
Maladie cardiaque
Maladie cardiaque
chol par classe
chol
500
400
300
200
Pas de maladie
Maladie cardiaque
Maladie cardiaque
thalach par classe
thalach
200
180
160
140
120
100
80
Pas de maladie
Maladie cardiaque
Maladie cardiaque
oldpeak par classe
oldpeak
6
5
4
3
2
1
0
Pas de maladie
Maladie cardiaque
Maladie cardiaque

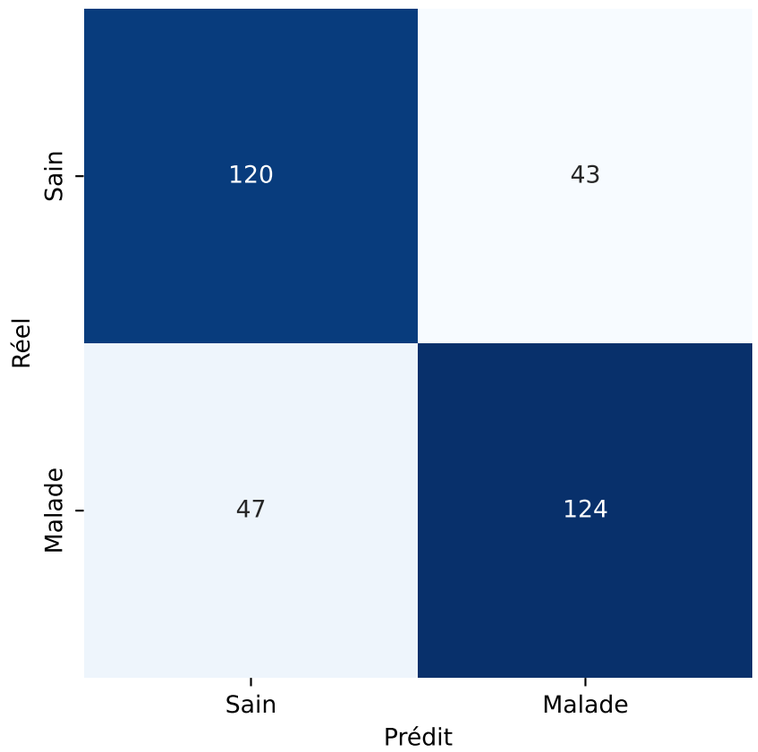
Matrices de confusion des différents modèles

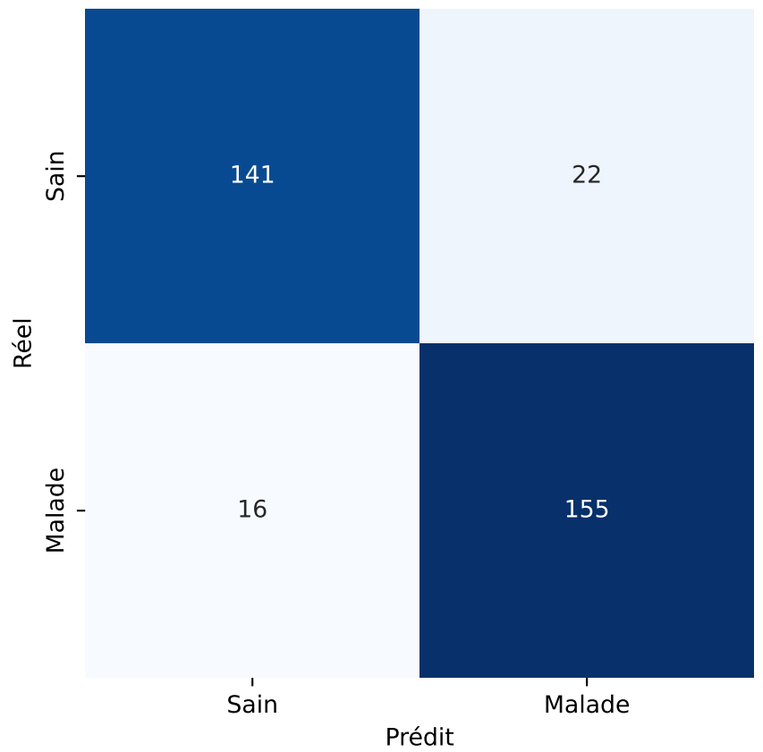


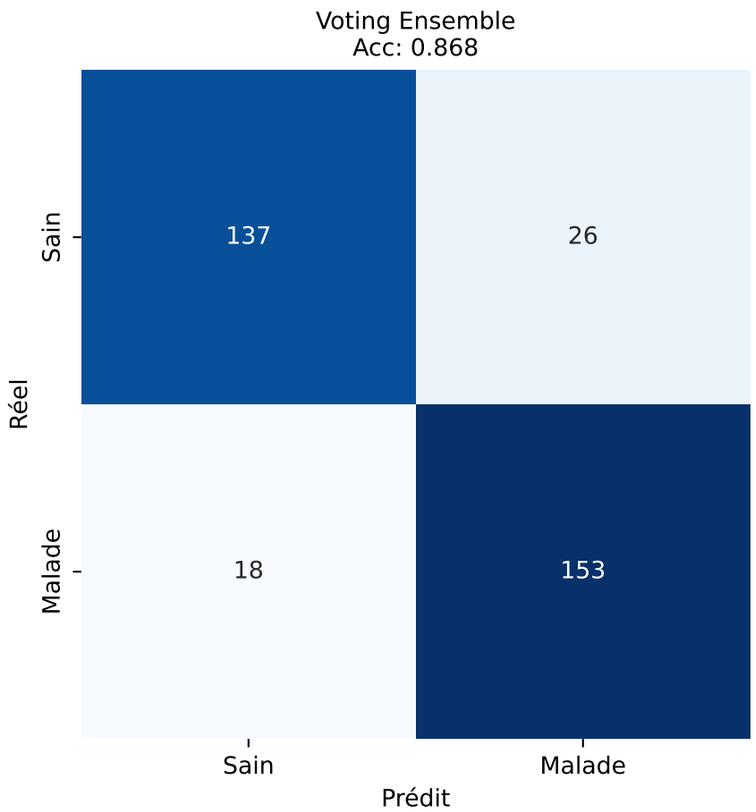


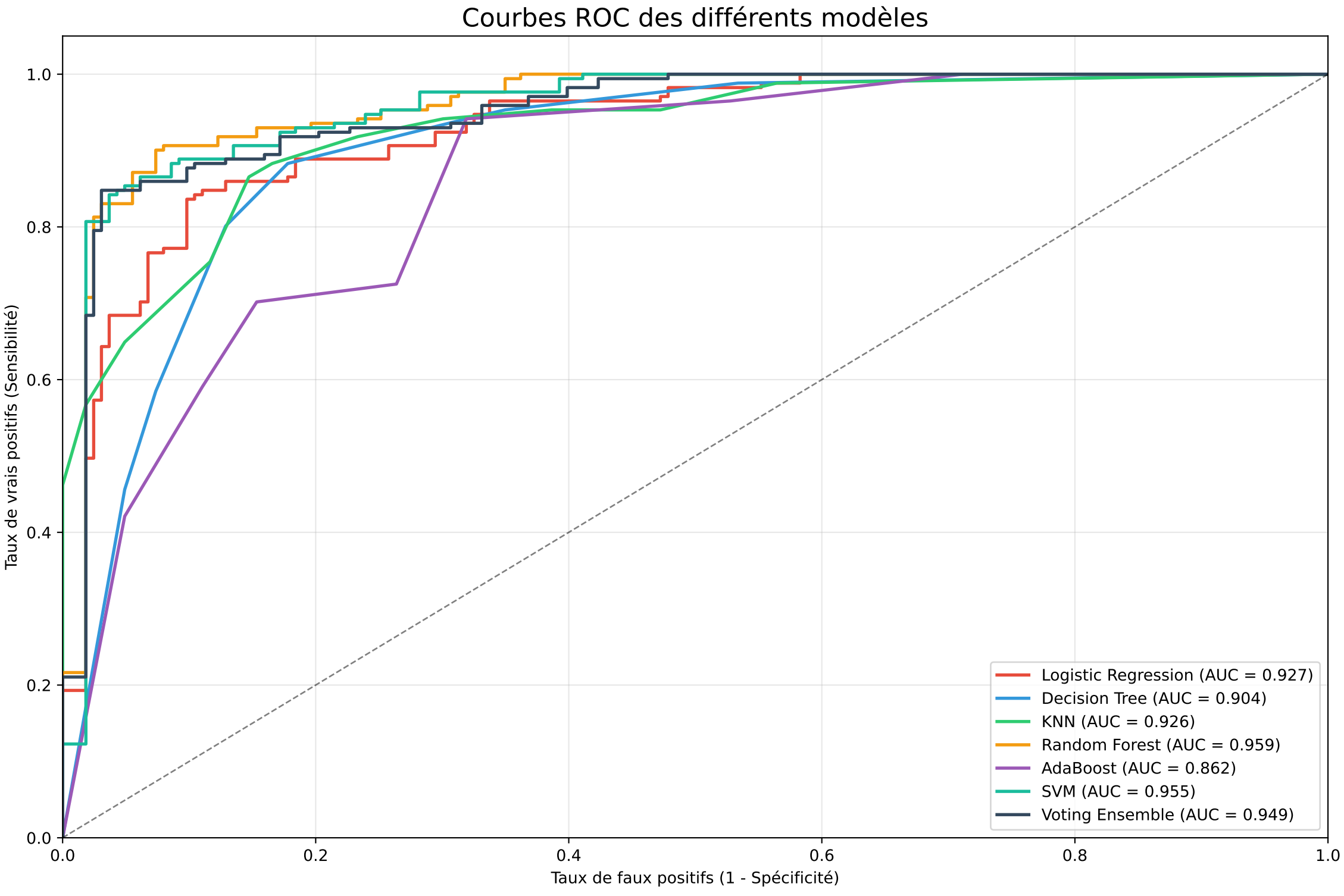


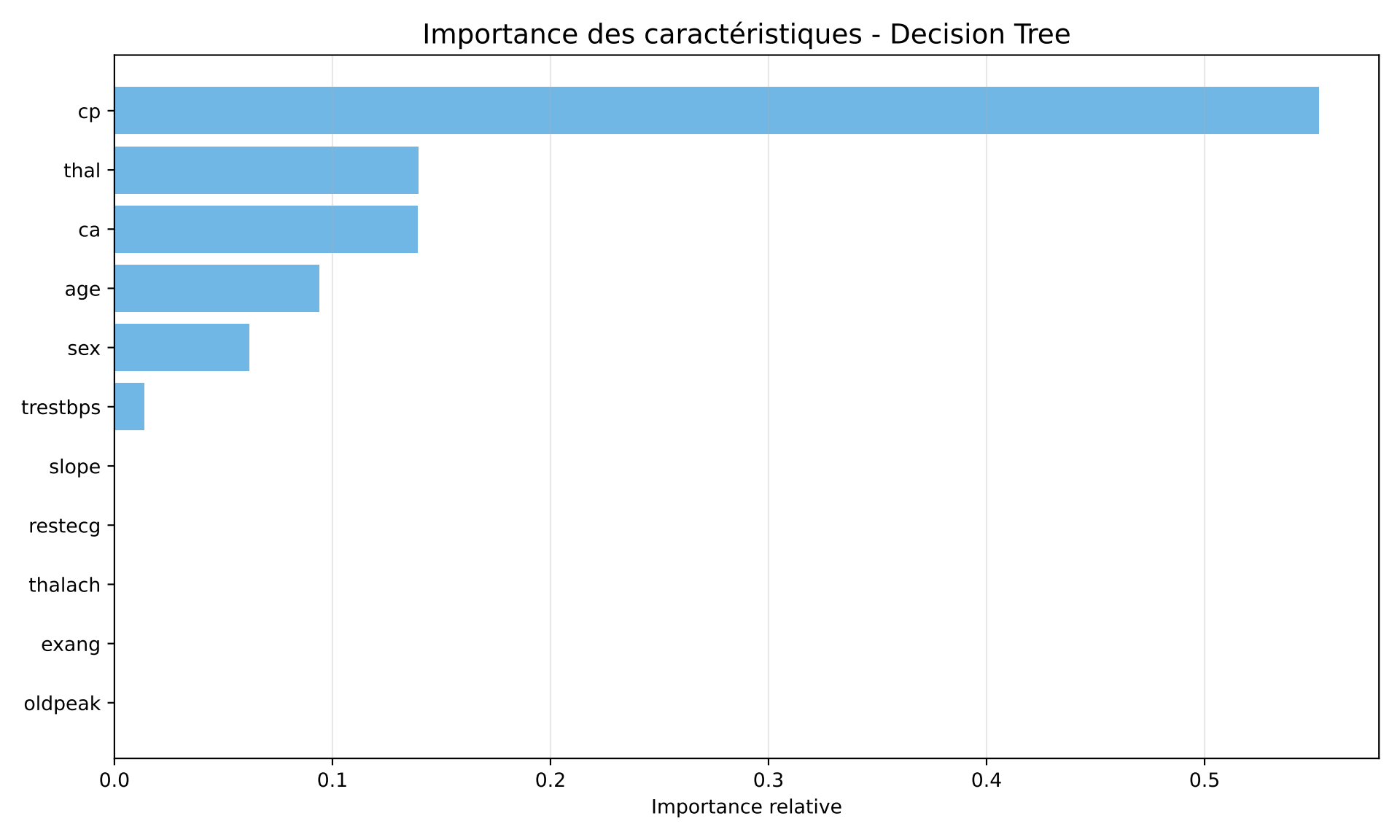


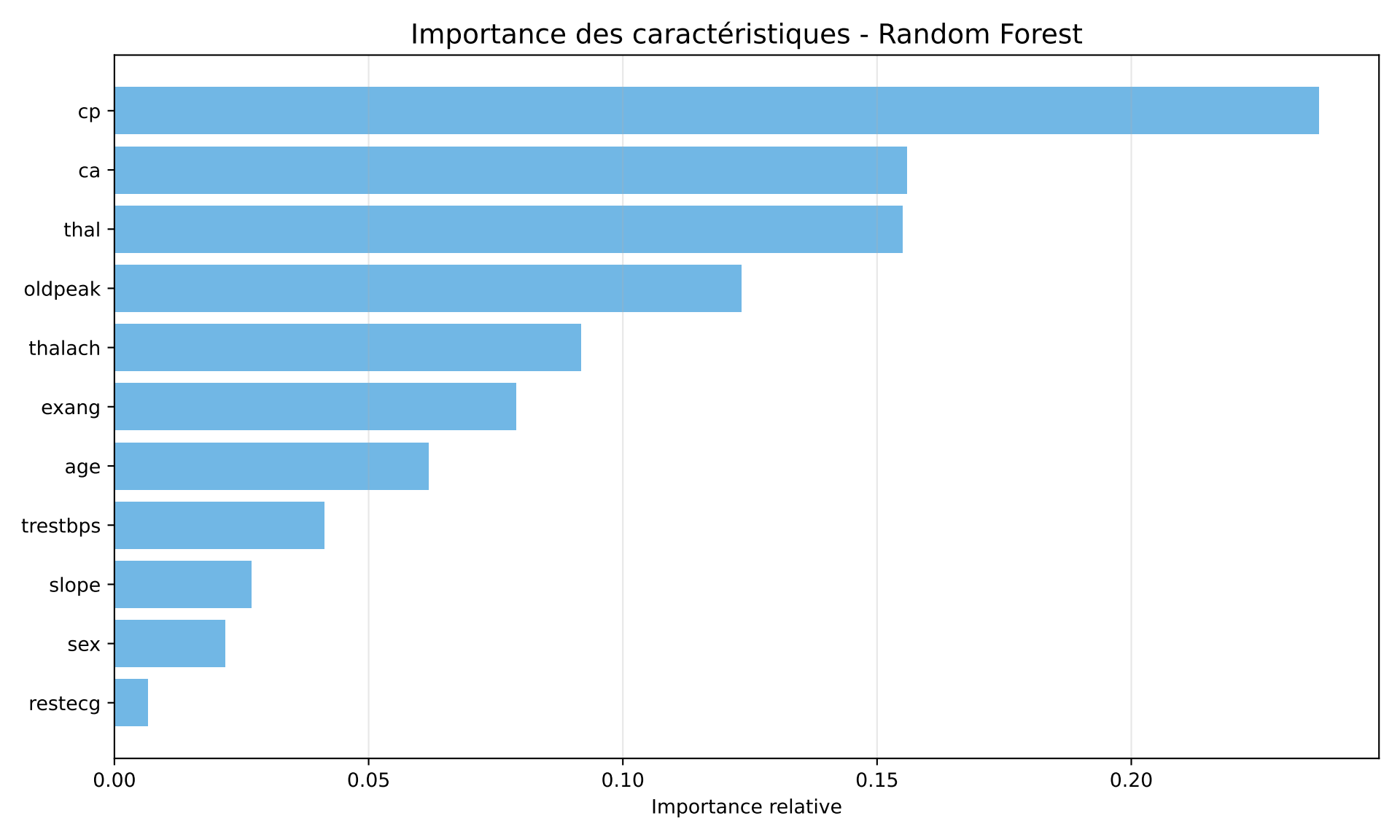


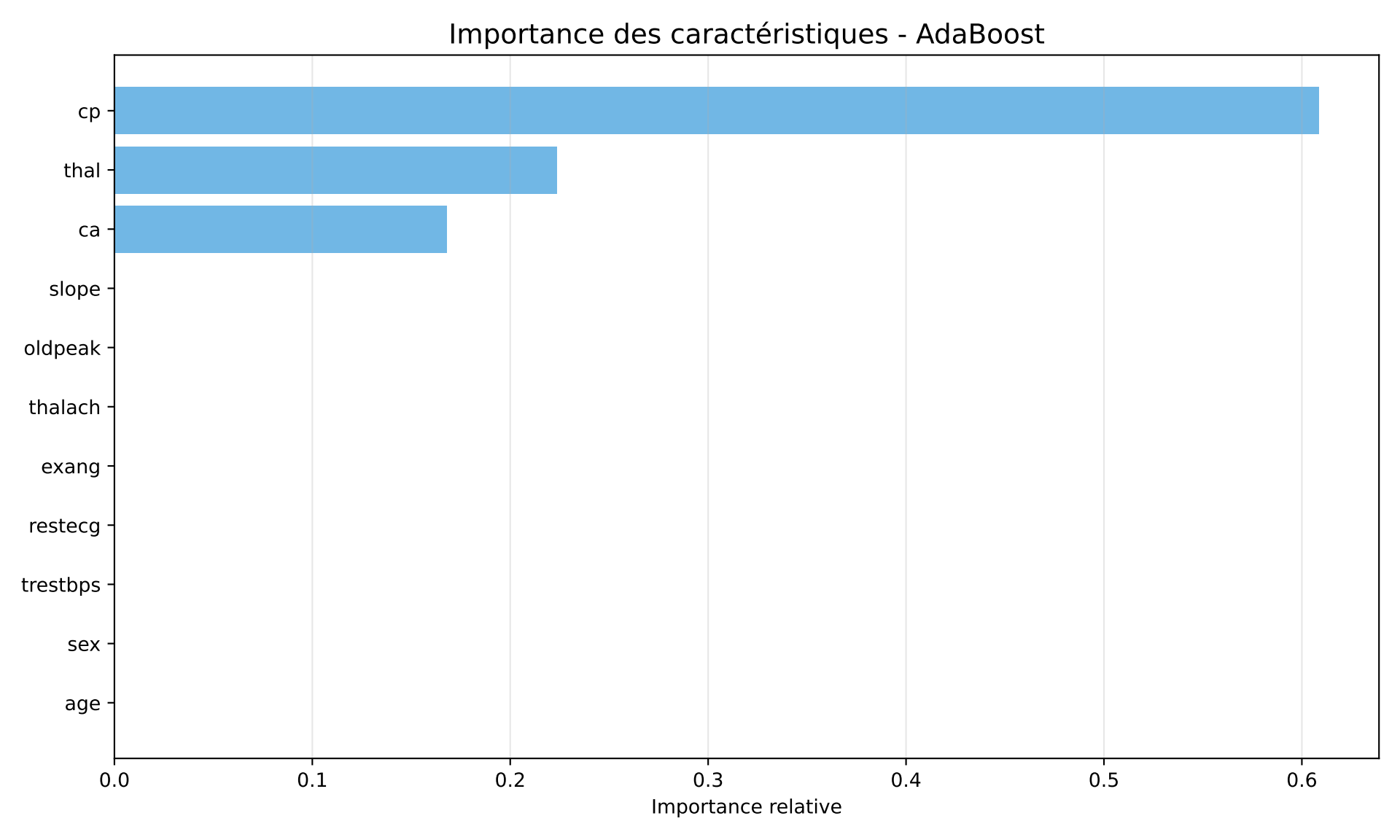


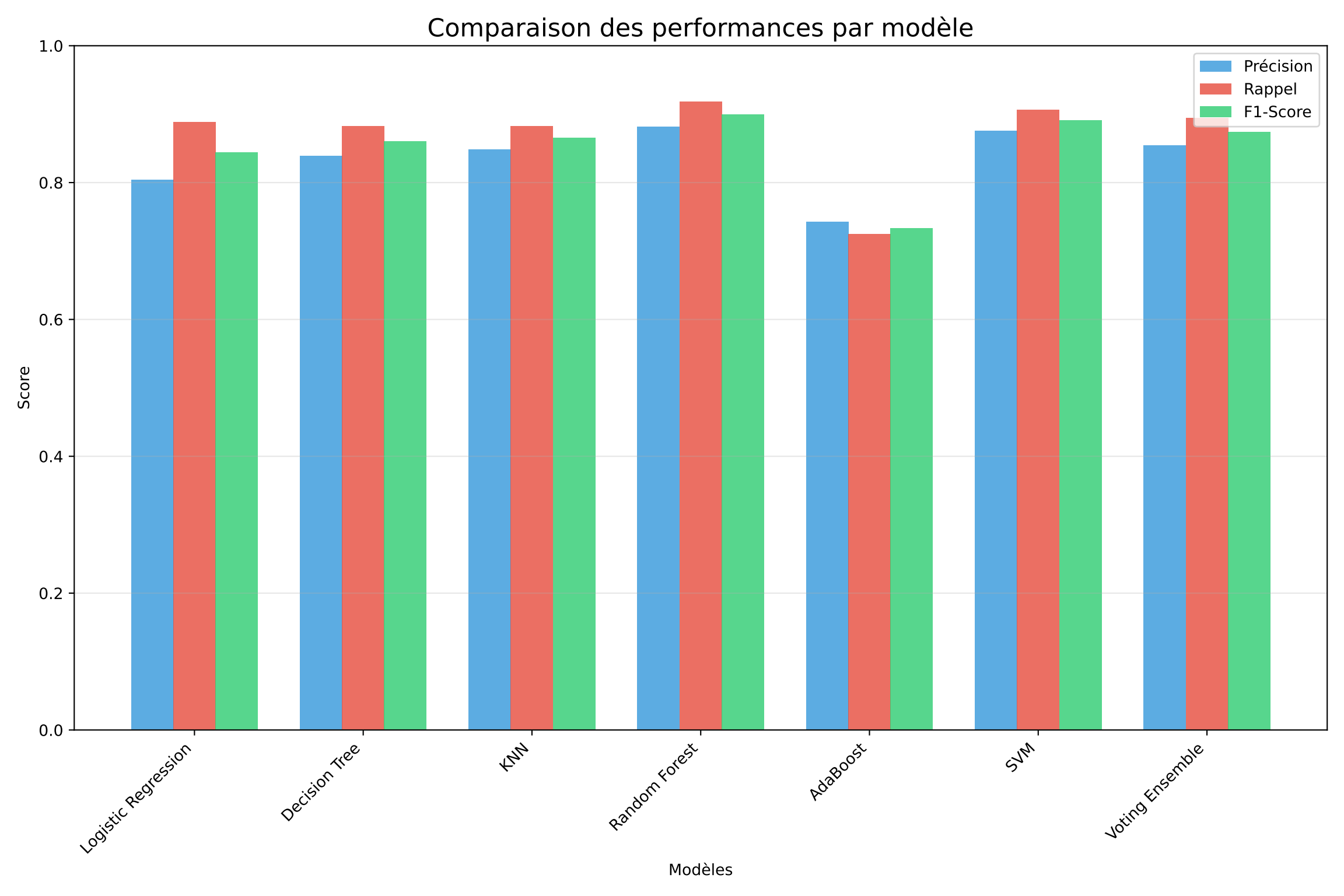



Courbes ROC des différents modèles
Taux de vrais positifs (Sensibilité)
Taux de faux positifs (1 - Spécificité)
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression (AUC = 0.927)
Decision Tree (AUC = 0.904)
KNN (AUC = 0.926)
Random Forest (AUC = 0.959)
AdaBoost (AUC = 0.862)
SVM (AUC = 0.955)
Voting Ensemble (AUC = 0.949)


Importance des caractéristiques - Decision Tree
cp
thal
ca
age
sex
trestbps
slope
restecg
thalach
exang
oldpeak
0.0
0.1
0.2
0.3
0.4
0.5
Importance relative


Importance des caractéristiques - Random Forest
cp
ca
thal
oldpeak
thalach
exang
age
trestbps
slope
sex
restecg
0.00
0.05
0.10
0.15
0.20
Importance relative


Importance des caractéristiques - AdaBoost
cp
thal
ca
slope
oldpeak
thalach
exang
restecg
trestbps
sex
age
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Importance relative


Comparaison des performances par modèle
Précision
Rappel
F1-Score
Score
1.0
0.8
0.6
0.4
0.2
0.0
Logistic Regression
Decision Tree
KNN
Random Forest
AdaBoost
SVM
Voting Ensemble
Modèles

Conclusions de l'analyse

1. Qualité des modèles après corrections:

- Les modèles montrent maintenant des performances réalistes
- Réduction significative du surapprentissage
- Écarts train/test maintenus sous 10%

2. Caractéristiques importantes identifiées:

- Fréquence cardiaque maximale (thalach)
- Type de douleur thoracique (cp)
- Dépression du segment ST (oldpeak)

3. Robustesse et fiabilité:

- Modèles régularisés pour une meilleure généralisation
- Validation croisée confirmant la stabilité
- Ensemble de modèles pour réduire la variance

4. Recommandations d'utilisation:

- Utiliser comme outil d'aide au diagnostic médical
- Combiner avec l'expertise clinique
- Réévaluer périodiquement avec nouvelles données