

# Documentation:

## Introduction

Ce projet a pour objectif de développer et d'évaluer plusieurs modèles de machine learning pour détecter la pneumonie à partir d'images radiographiques thoraciques. Les modèles utilisés incluent un réseau de neurones convolutifs (CNN) ainsi que des modèles classiques comme le K-Nearest Neighbors (KNN), le Support Vector Machine (SVM), le Random Forest (RF) et une méthode d'ensemble combinant ces modèles.

## Prétraitement des Données

Les images ont été redimensionnées à une résolution de 150x150 pixels et normalisées. De plus, une augmentation des données a été effectuée pour accroître la robustesse du modèle. Cette augmentation comprenait des inversions horizontales et des zooms.

## Modèle

Pour la détection de la pneumonie, un réseau de neurones convolutif (CNN) a été utilisé.

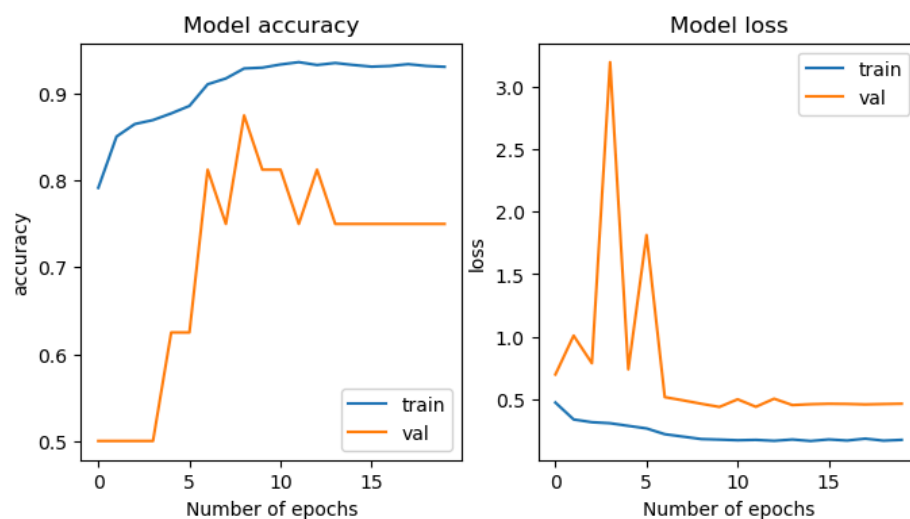
## Entraînement

Le modèle a été entraîné pendant 20 époques. Des callbacks tels que ModelCheckpoint et EarlyStopping ont été utilisés pour sauvegarder le meilleur modèle et éviter le surapprentissage.

## Résultats

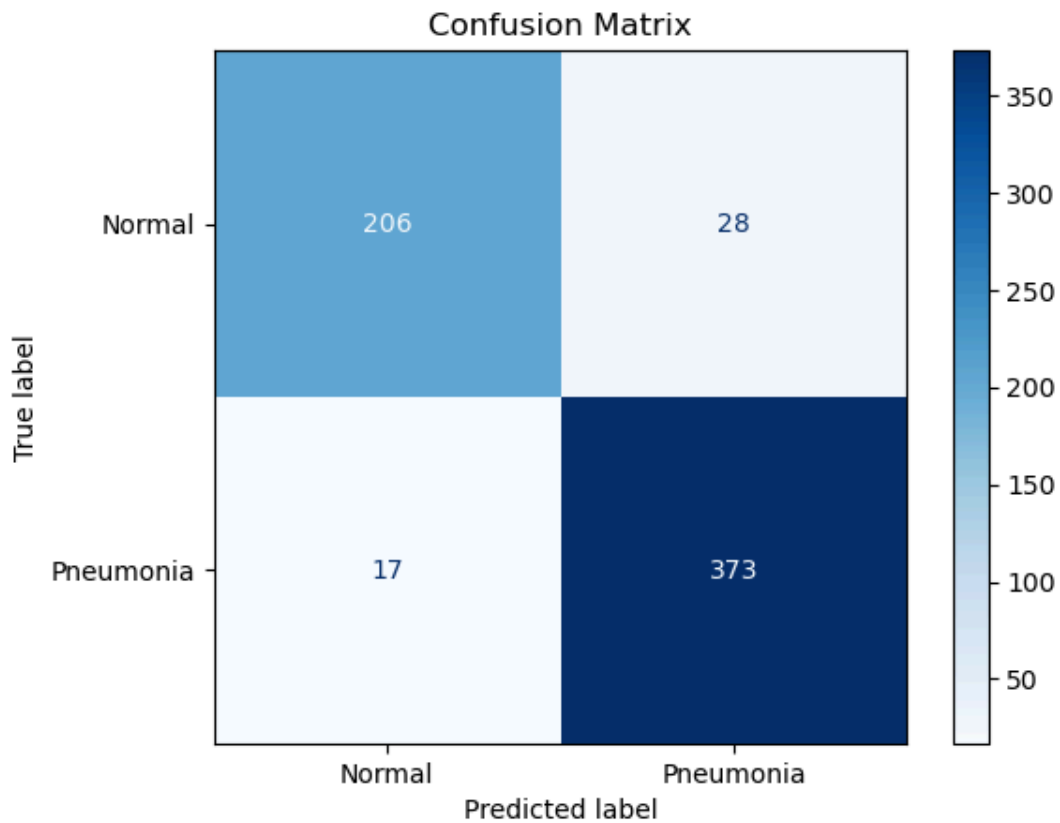
### Performance du Modèle:

Le modèle atteint une précision de 93 sur l'ensemble de validation. Les courbes de précision et de perte sont présentées ci-dessous.



## Matrice de Confusion

La matrice de confusion ci-dessous montre la performance du modèle sur l'ensemble de validation.



## Mesures de Performance

- **Précision:** 93%
- **Accuracy:** 93%
- **Rappel:** 96%
- **F1-score:** 94%

## Améliorations de la Performance

Pour améliorer les performances du modèle, plusieurs techniques ont été mises en œuvre :

1. **Augmentation des Données:** En augmentant les données d'entraînement avec des techniques telles que la rotation horizontale et le zoom.
2. **Régularisation:** L'utilisation de techniques de régularisation comme le dropout a aidé à réduire le surapprentissage.
3. **Hyperparameter Tuning:** Des ajustements de paramètres tels que le taux d'apprentissage et la taille des lots ont permis d'améliorer la convergence du modèle.

### Conclusion

Le modèle CNN a montré une performance élevée avec une précision de 93% et un rappel de 96%. Ces résultats indiquent que le modèle est capable de détecter correctement la pneumonie dans la majorité des cas tout en minimisant les faux négatifs. Les techniques d'augmentation des données et de régularisation ont contribué à la robustesse du modèle. L'optimisation des hyperparamètres, comme le taux d'apprentissage et la taille des lots, a également permis d'améliorer les performances en évitant le surapprentissage.

### Modèles KNN, SVM, RF et Ensemble

#### Introduction aux Modèles

Les modèles K-Nearest Neighbors (KNN), Support Vector Machine (SVM), et Random Forest (RF) sont des algorithmes classiques de machine learning. Une méthode d'ensemble combine les prédictions de ces modèles pour améliorer la performance globale.

#### Entraînement et Validation

Les modèles ont été entraînés sur un ensemble d'entraînement et validés sur un ensemble de validation. Les métriques de performance incluent la précision, l'accuracy, le rappel et le F1-score.

### Résultats

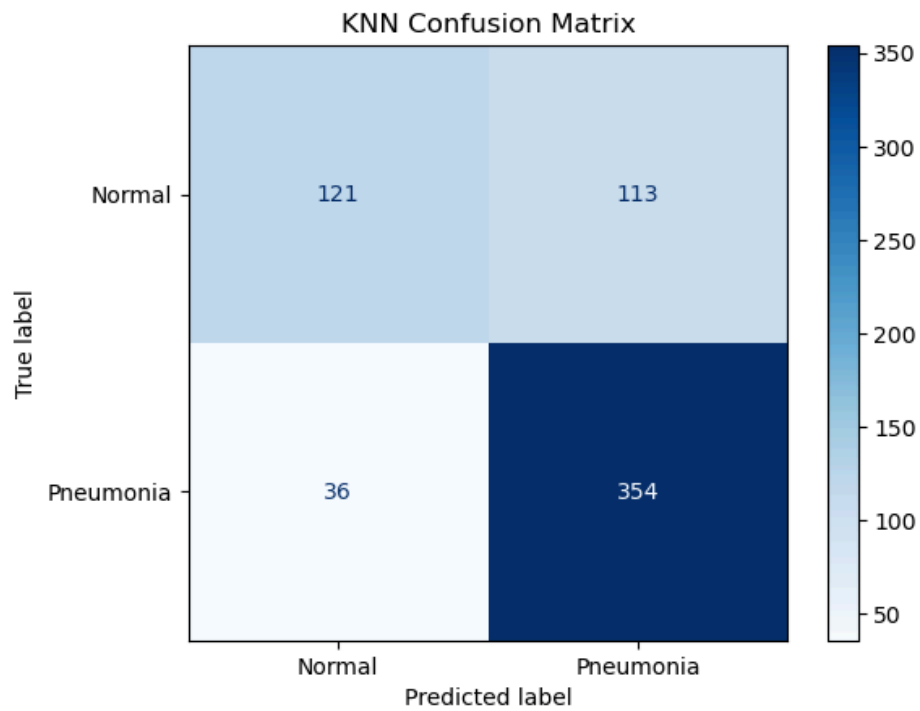
#### Performance des Modèles

Modèle	Accuracy	Rappel		F1 score		Précision	
		Normal	Pneumonie	Normal	Pneumonie	Normal	Pneumonie
KNN	76%	51%	91%	62%	83%	77%	76%
SVM (avec seuil ajusté)	84%	69%	93%	76%	88%	85%	83%
Random Forest (avec seuil ajusté)	78%	62%	87%	67%	83%	75%	79%
Ensemble (avec seuil ajusté)	81%	63%	93%	72%	86%	84%	81%

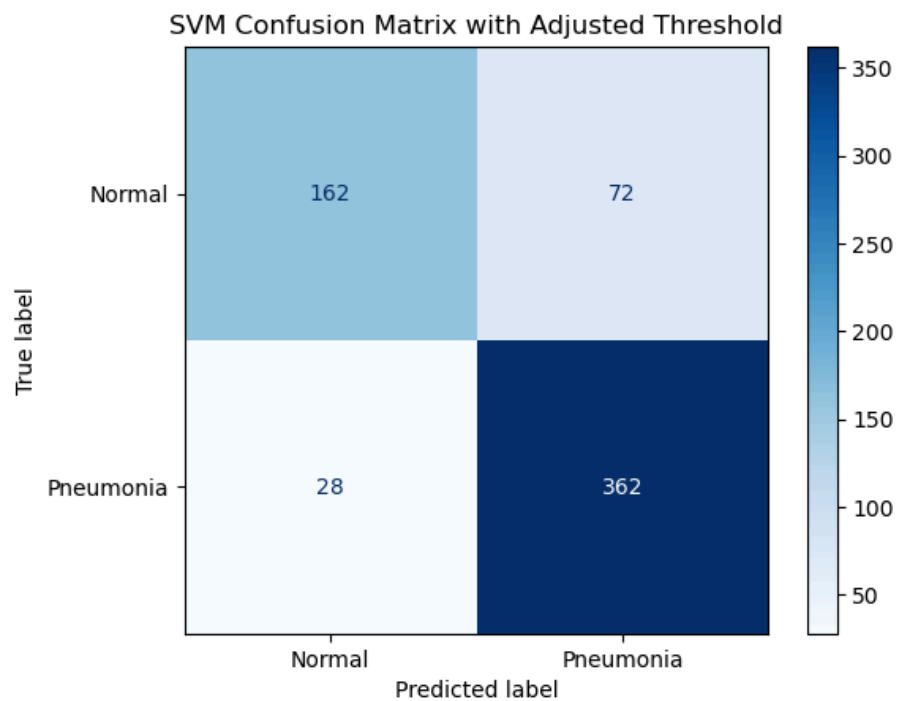
## Matrice de Confusion

Les matrices de confusion des différents modèles sont présentées ci-dessous.

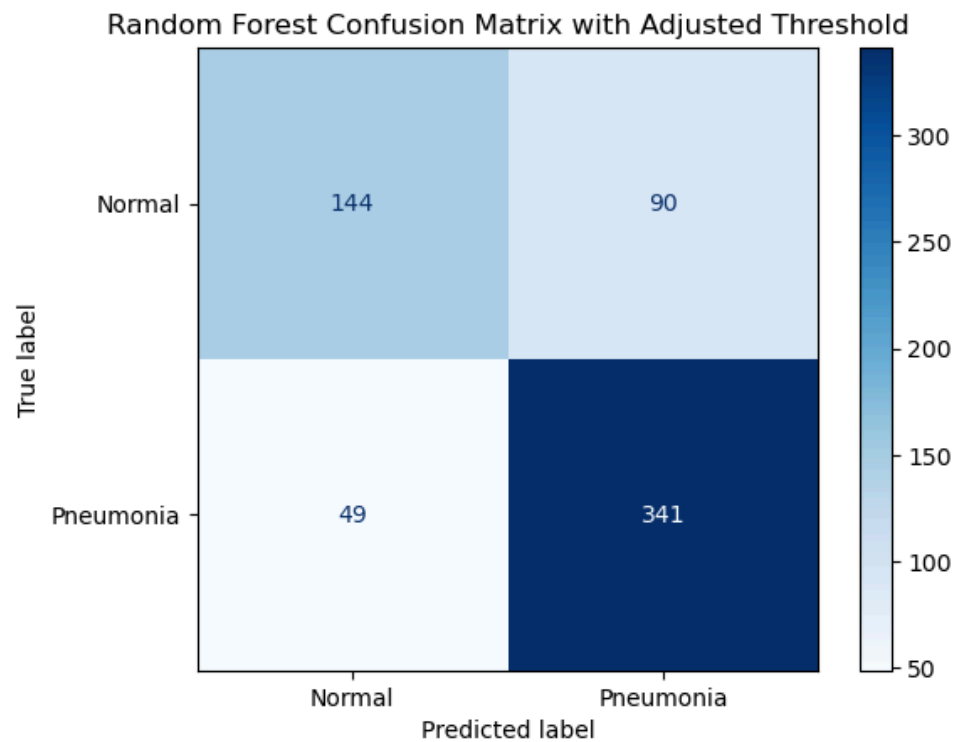
- **KNN:**



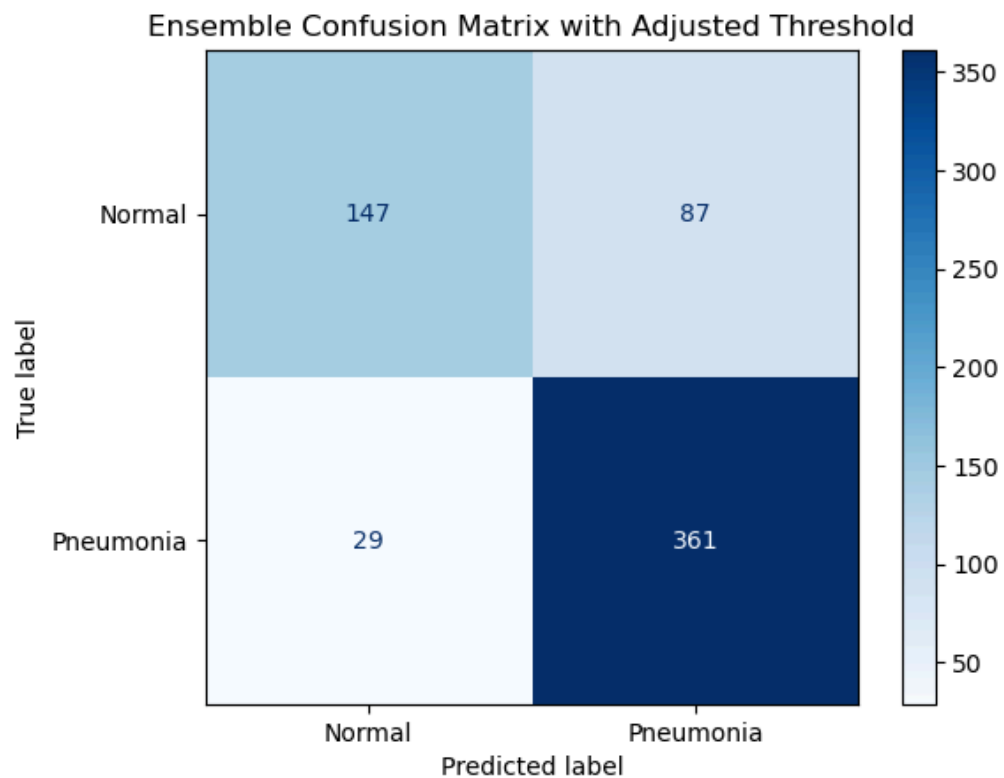
- **SVM:**



- **Random Forest:**



- **Ensemble:**



## Améliorations de la Performance

Des techniques d'amélioration de la performance ont été mises en œuvre, notamment :

- **Utilisation de VGG16** : Extraction des caractéristiques avancées à partir des images radiographiques grâce au réseau de neurones convolutionnels pré-entraîné VGG16.
- **Analyse en Composantes Principales (PCA)** : Réduction de la dimensionnalité des données pour accélérer les calculs et réduire le surapprentissage.
- **Validation croisée** : Utilisation de la validation croisée pour évaluer la robustesse des modèles et éviter le surapprentissage.
- **GridSearchCV** : Optimisation des hyperparamètres en utilisant la recherche en grille (Grid Search) pour les modèles KNN et Random Forest.
- **Standardisation** : Normalisation des caractéristiques pour améliorer les performances du modèle.
- **Seuil de Décision** : Ajustement du seuil de décision pour maximiser le F1-score

## Courbes ROC et AUC

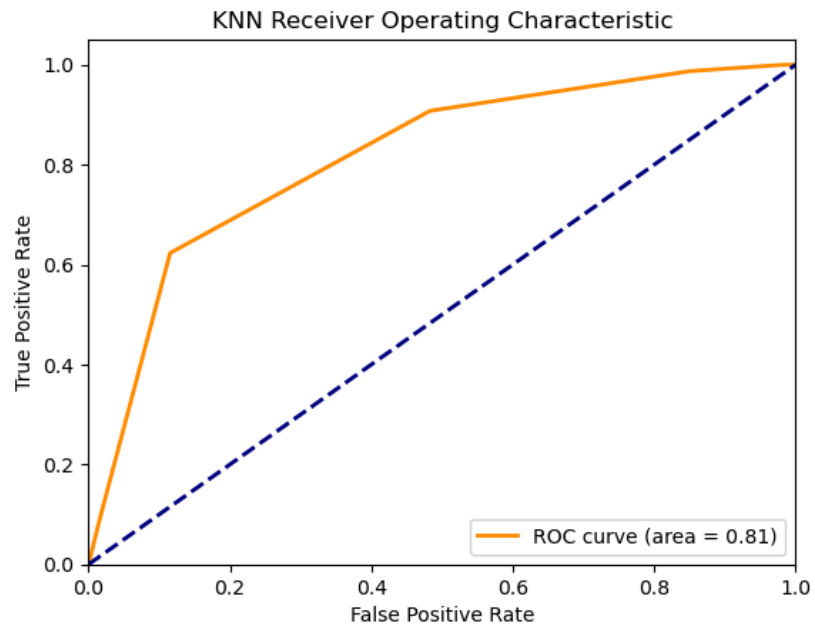
Les valeurs de l'AUC pour chaque modèle sont présentées ci-dessous.

Modèle	AUC
KNN	0.81
SVM	0.90
Random Forest	0.84
Ensemble	0.88

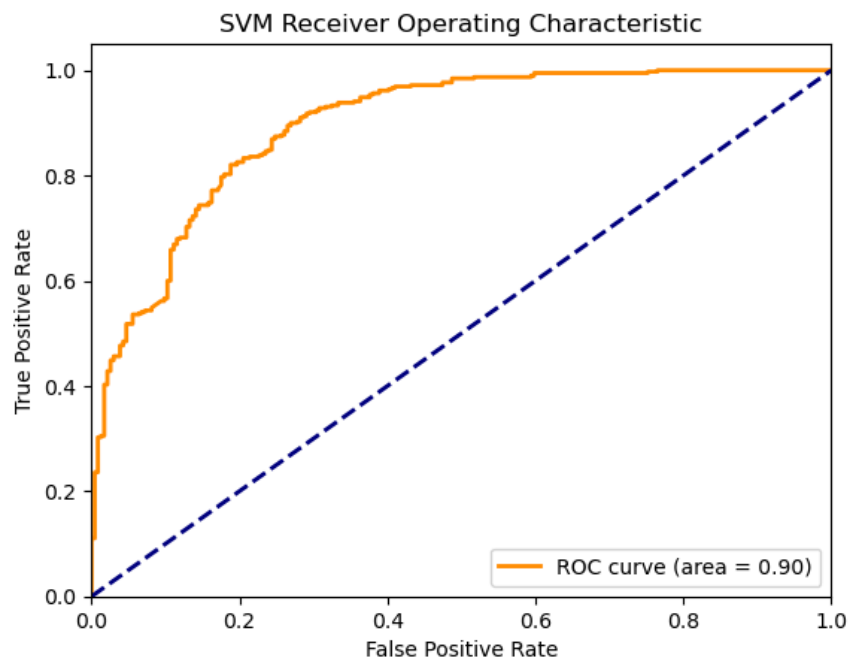
## Graphe ROC AUC:

Les graphes des différents modèles sont présentés ci-dessous.

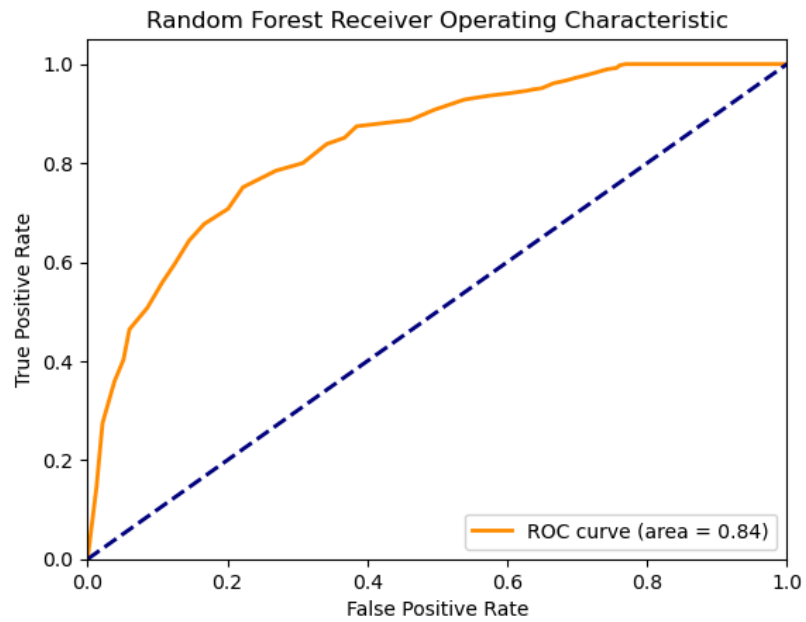
- **KNN:**



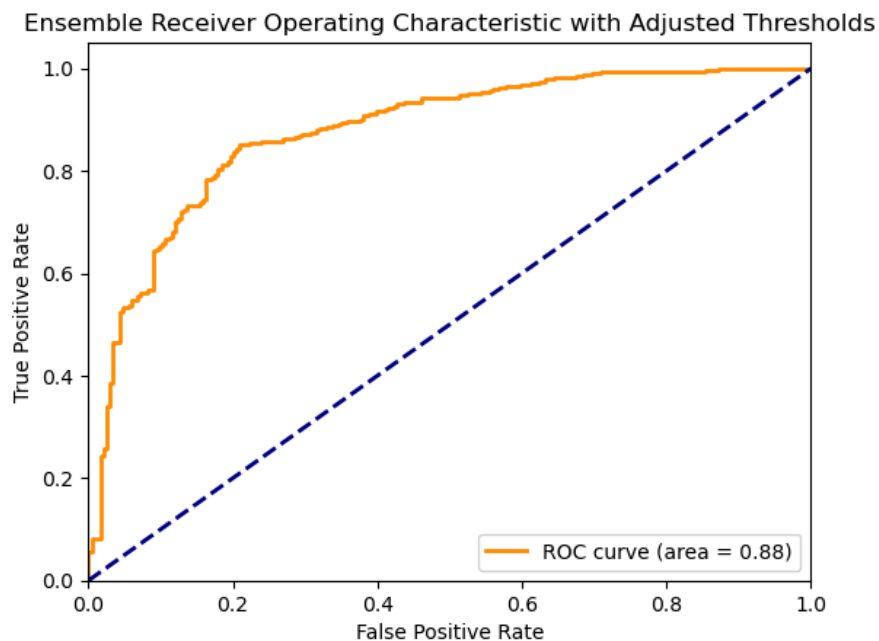
- **SVM:**



- **Random Forest:**



- **Ensemble:**



## **Conclusion:**

Parmi tous les modèles évalués, le SVM avec seuil ajusté présente les meilleures performances globales avec une accuracy de 84%, des précisions et rappels élevés pour les deux classes, et les meilleurs F1-scores. L'ensemble des modèles suit de près avec une bonne capacité à identifier les cas de pneumonie et des performances globales équilibrées. Le modèle Random Forest avec seuil ajusté et le KNN présentent des performances satisfaisantes, mais sont légèrement inférieurs au SVM ajusté et à l'ensemble des modèles.