
Détection de Pneumonie à l'aide de SVM

Introduction

Ce document décrit l'utilisation d'un classifieur SVM pour détecter la pneumonie à partir d'images de radiographies pulmonaires. Les images sont pré-traitées pour isoler les poumons et sont ensuite utilisées pour entraîner et tester le modèle.

Importation des bibliothèques et définition des chemins

Dans cette section, nous importons les bibliothèques nécessaires pour le traitement des images, l'entraînement du modèle et l'évaluation des performances. Nous définissons également les chemins vers les ensembles de données d'entraînement et de test.

Définition des fonctions utilitaires

Nous définissons ici deux fonctions utilitaires :

1. **Une fonction pour mesurer le temps d'exécution des fonctions**, ce qui nous permet d'optimiser les performances.
 2. **Une fonction pour charger les images à partir de deux dossiers différents**, les prétraiter et les étiqueter. Ces fonctions sont cruciales pour assurer que les images sont correctement préparées pour l'entraînement du modèle.
-

Fonction de recadrage des images

Cette fonction recadre la zone des poumons dans une image pour se concentrer uniquement sur la région d'intérêt. Le but est d'éliminer les parties de l'image qui ne sont pas pertinentes pour la détection de la pneumonie, ce qui permet d'améliorer les performances du modèle en réduisant le bruit.

Chargement des images et des étiquettes

Nous chargeons ici les images et les étiquettes d'entraînement et de test à l'aide des fonctions définies précédemment. Le but est de préparer les données qui seront utilisées pour l'entraînement et l'évaluation du modèle. Nous utilisons deux ensembles de données : un pour l'entraînement et un pour le test, afin de pouvoir évaluer les performances du modèle sur des données qu'il n'a pas vues auparavant.

Création et entraînement du modèle SVM

Nous créons un pipeline avec la mise à l'échelle des données et SVM, puis entraînons le modèle avec les données d'entraînement. Le but de cette étape est de construire un modèle capable de distinguer les radiographies normales des radiographies indiquant une pneumonie.

Prédiction et évaluation du modèle

Nous effectuons des prédictions sur les données de test et affichons le rapport de classification. Cette étape permet d'évaluer les performances du modèle en termes de précision, rappel et score F1, et de vérifier comment le modèle se comporte sur les nouvelles données.

Résultats de la classification

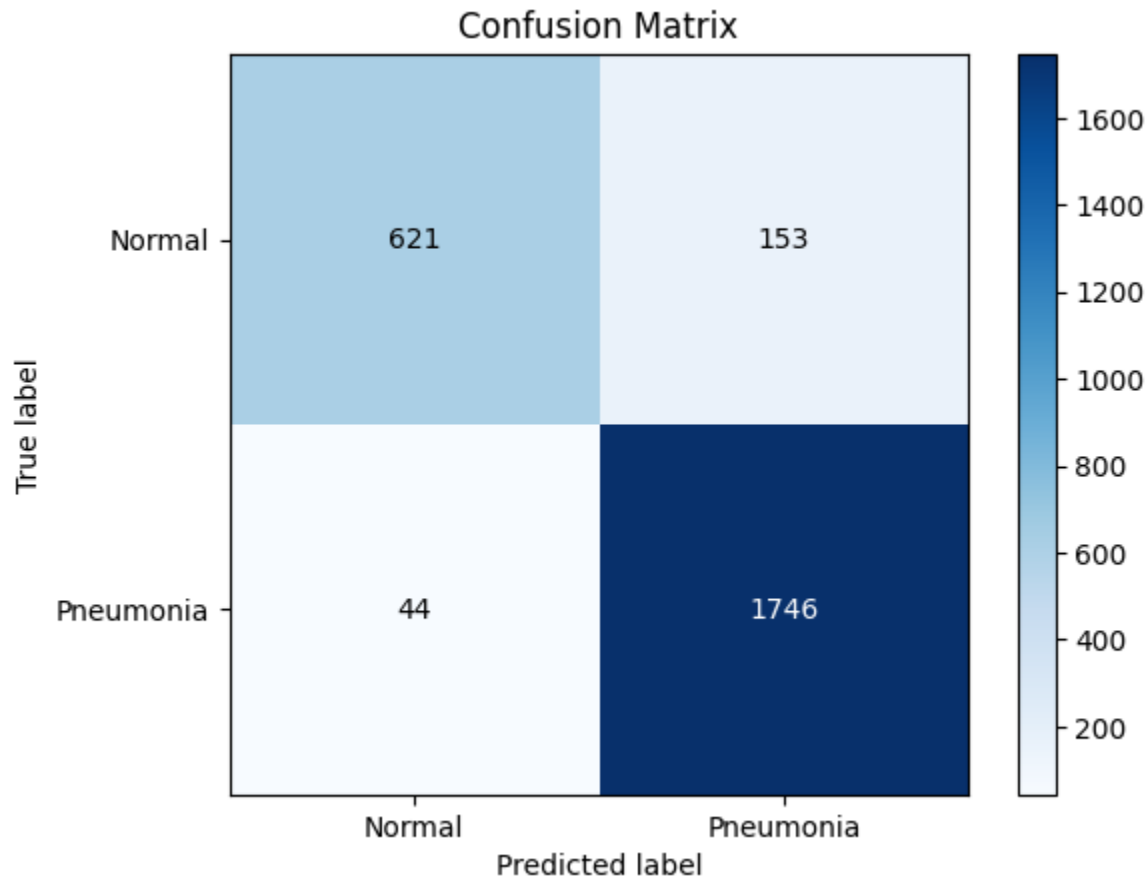
Voici les résultats obtenus après avoir effectué un grid search pour tester plus de 600 combinaisons de paramètres pour obtenir le meilleur modèle :

Classe	Précision	Rappel	F1-score	Support
0	0.93	0.80	0.86	774
1	0.92	0.98	0.95	1790
Précision globale	0.92			2564
Moyenne macro	0.93	0.89	0.90	2564
Moyenne pondérée	0.92	0.92	0.92	2564

Ces résultats montrent une bonne performance générale du modèle, avec une précision globale de 92%. Avant d'obtenir ces résultats, nous avons testé plus de 600 combinaisons de paramètres pour trouver les meilleures hyperparamètres.

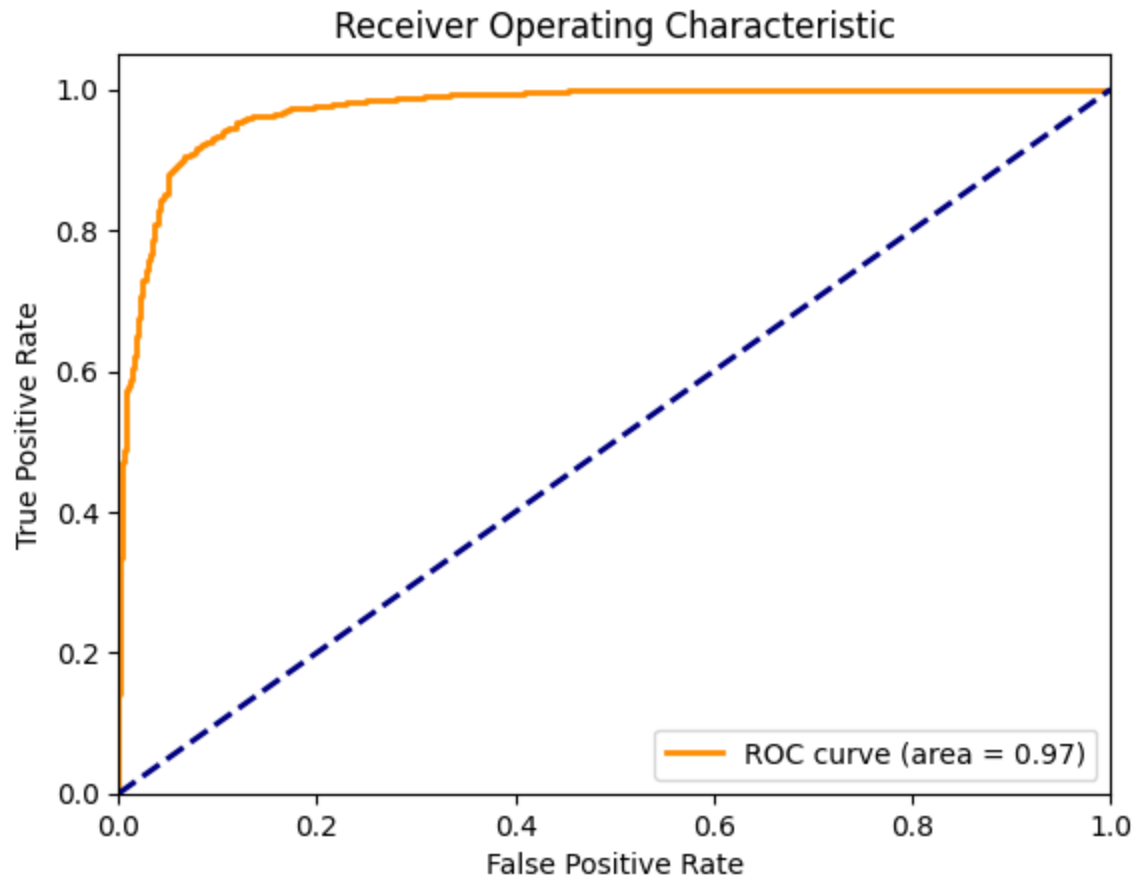
Matrice de confusion

Nous traçons la matrice de confusion pour visualiser les performances du modèle. La matrice de confusion permet de voir le nombre de vraies et fausses prédictions pour chaque classe, ce qui aide à identifier les domaines où le modèle peut s'améliorer.



Courbe ROC

Nous calculons et traçons la courbe ROC pour évaluer les performances du modèle en termes de taux de faux positifs et de vrais positifs. La courbe ROC et l'aire sous la courbe (AUC) sont des métriques importantes pour évaluer la qualité du modèle, surtout dans les cas de classification binaire. La courbe ROC montre la capacité du modèle à distinguer entre les classes. Une AUC de 0,97 indique une excellente capacité de classification.



Conclusion

Le traitement des images et la réduction des données d'entraînement ont joué un rôle crucial dans l'optimisation du modèle SVM pour la détection de la pneumonie. En isolant la région d'intérêt et en éliminant le bruit, nous avons pu améliorer la qualité des données fournies au modèle, ce qui a permis d'augmenter la précision et la fiabilité des prédictions. Cette approche a également contribué à réduire la complexité du modèle et à accélérer le processus d'entraînement, démontrant ainsi l'importance de la préparation et du prétraitement des données dans les projets de machine learning.