

Détection Automatique de la Pneumonie à partir d'Images Radiographiques via Apprentissage Machine

Projet ZOIDBERG2.0 — Epitech Nancy – 07/2026

GOUIA Hichem, DENIS Melvyn, SERRIER Mathéo

Résumé

Dans ce projet, nous explorons différentes techniques d'apprentissage automatique et profond pour automatiser la détection de la pneumonie à partir d'images radiographiques thoraciques.

Nous avons appliqué plusieurs méthodes classiques (SVM, Random Forest, etc.) et des architectures de deep learning (ResNet18, U-Net). Une analyse approfondie des données a été réalisée, notamment via PCA.

Les résultats montrent que le modèle SVM avec noyau linéaire offre la meilleure précision générale, tandis que le modèle ResNet18 excelle en classification profonde.

1. Introduction

La pneumonie est une infection pulmonaire potentiellement mortelle, particulièrement dangereuse chez les enfants et les personnes âgées. La détection rapide et précise de cette pathologie à partir d'images X-ray est donc cruciale. Dans ce contexte, le projet ZOIDBERG2.0 vise à automatiser cette tâche grâce à des méthodes d'intelligence artificielle.

2. Données et Prétraitement

Les données sont constituées de radiographies thoraciques, réparties dans trois dossiers : train, validation et test. Chaque dossier contient deux classes : NORMAL et PNEUMONIA.

Les images ont été redimensionnées, normalisées, puis encodées numériquement pour être utilisées par les modèles.

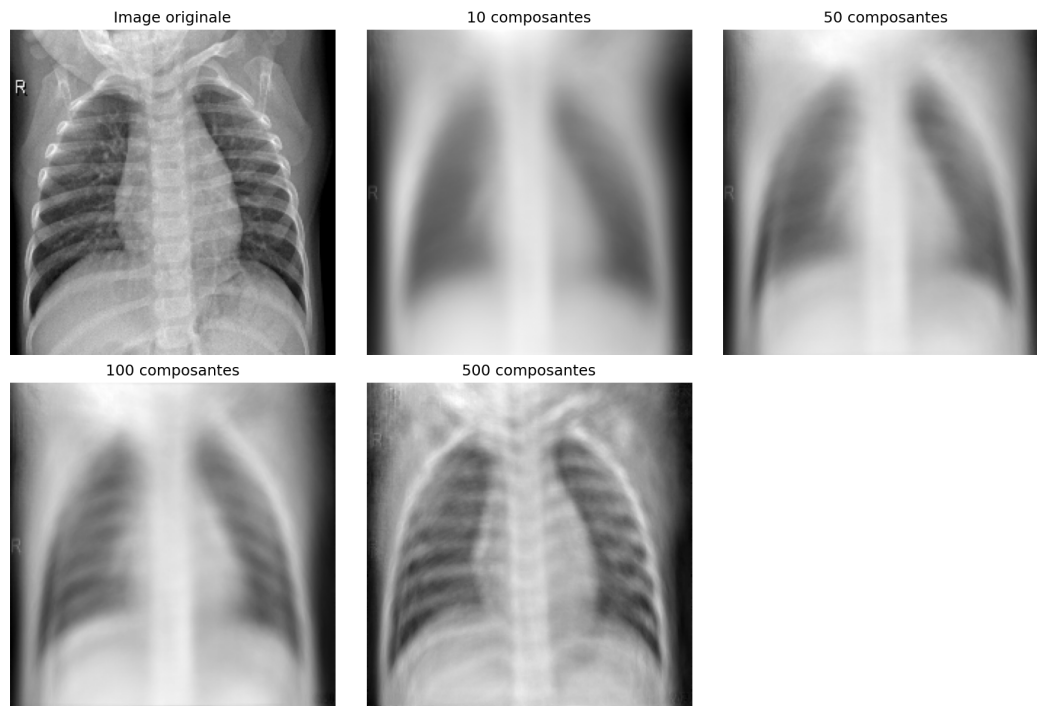


Figure 1 : Image originale et reconstructions selon le nombre de composantes principales.

3. Analyse Exploratoire

Une analyse PCA a été utilisée pour réduire la dimensionnalité et visualiser la structure des données.

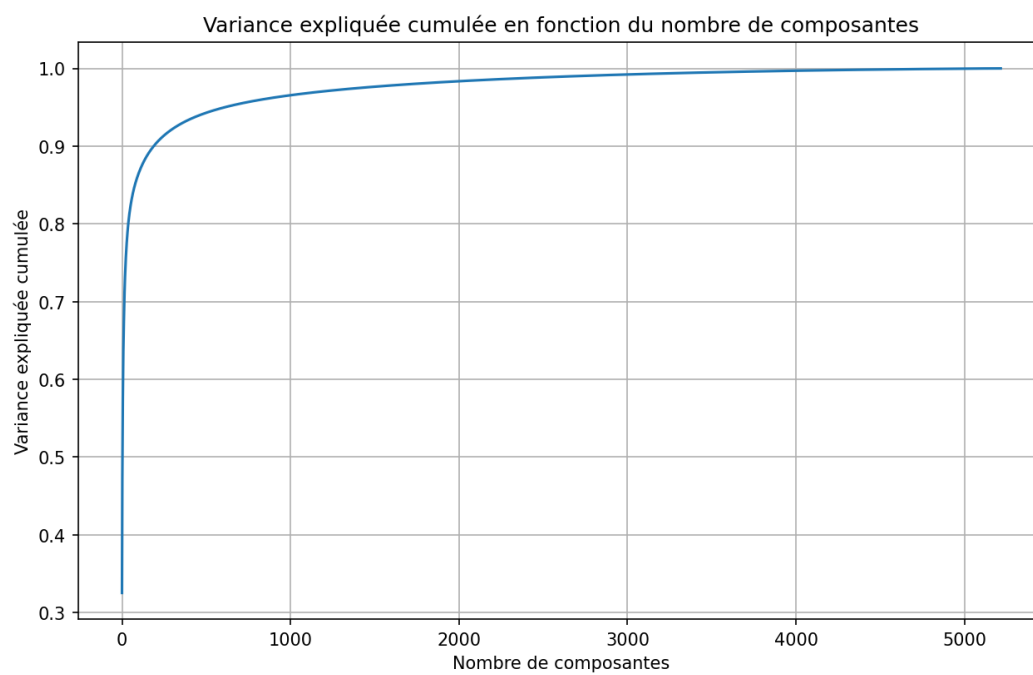


Figure 2 : Variance expliquée cumulée selon le nombre de composantes principales.

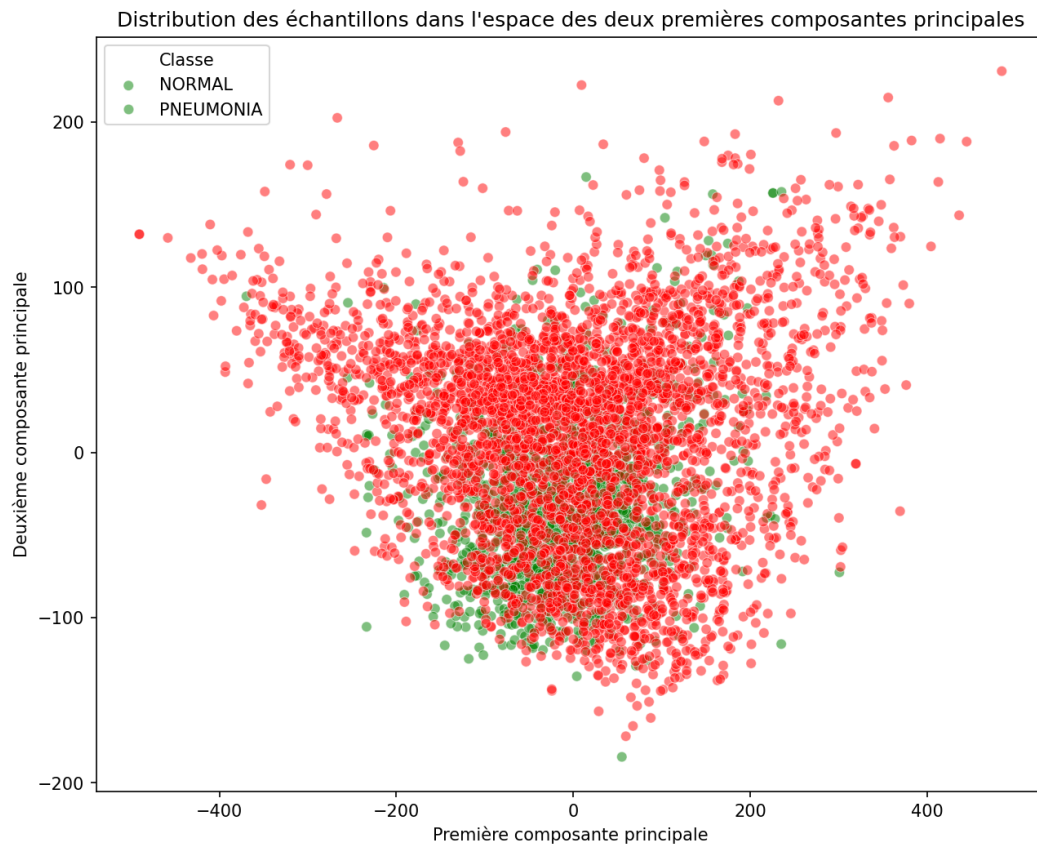


Figure 3 : Projection des échantillons selon les deux premières composantes principales.

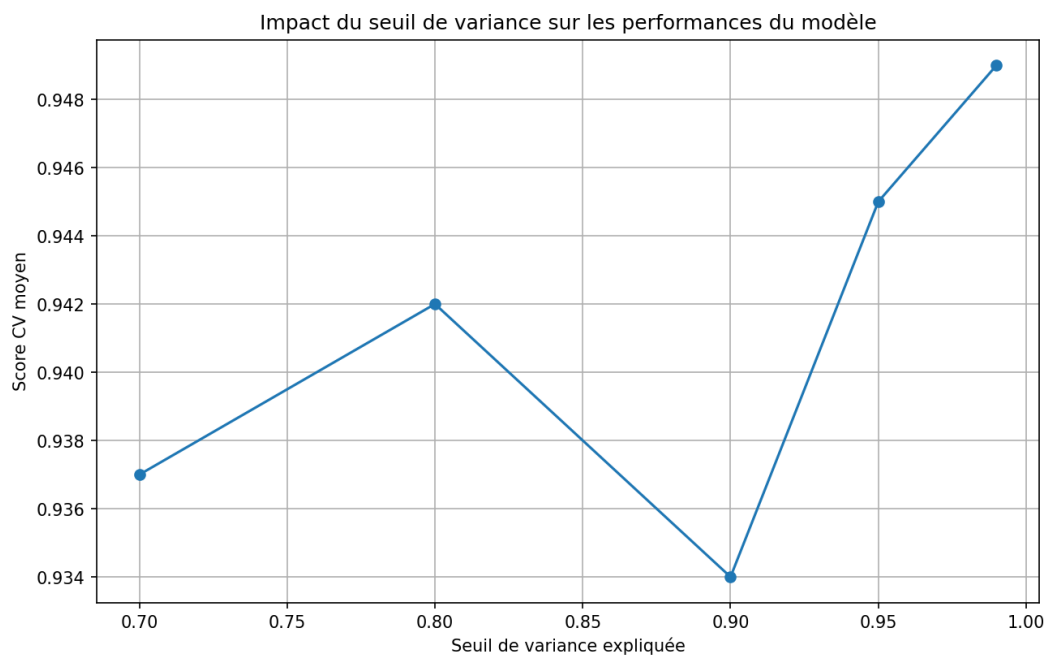


Figure 4 : Impact du seuil de variance sur les performances (score CV).

4. Méthodes d'Apprentissage Machine

Nous avons comparé plusieurs modèles classiques : régression logistique, random forest, gradient boosting et SVM. Les scores de validation croisée ainsi que les précisions sur le jeu de validation ont été comparés.

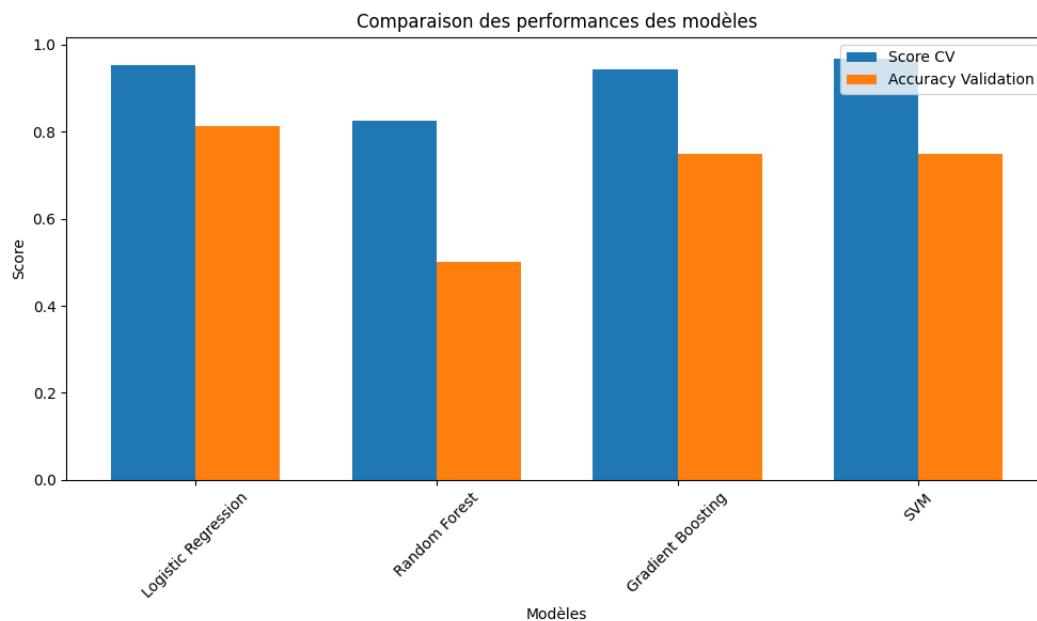


Figure 5 : Comparaison des performances des modèles classiques.

Le modèle SVM avec noyau linéaire a montré les meilleures performances globales, avec un score de validation croisée moyen de 0.96.

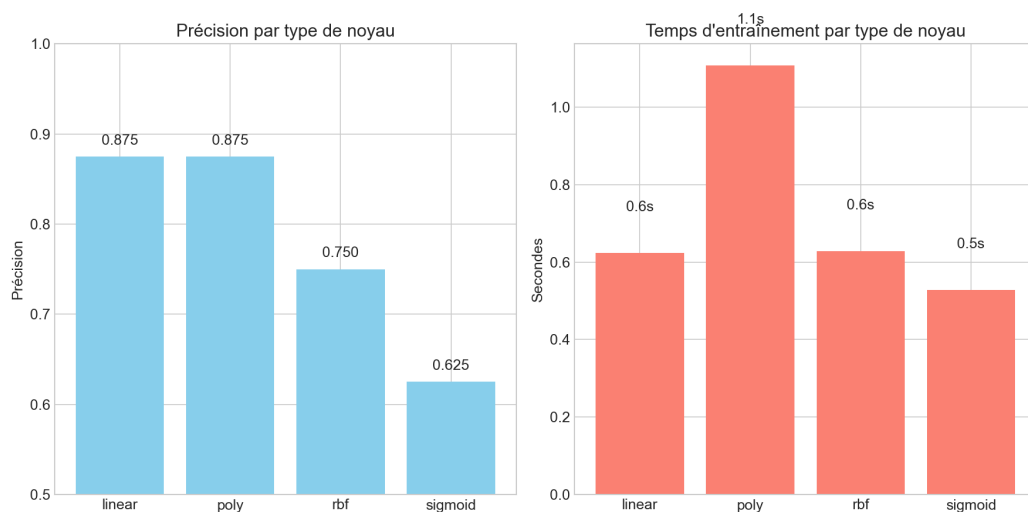


Figure 6 : Précision et temps d'entraînement par type de noyau SVM.

Voici les performances détaillées du meilleur SVM (noyau linéaire) :

```
RÉSUMÉ DES PERFORMANCES DU MODÈLE SVM
-----
Meilleur noyau: linear
Paramètres optimaux: {'C': 0.1, 'break_ties': False, 'cache_size': 200, 'class_weight': 'balanced', 'coef0': 0.0, 'decision_function_shape': 'raw'}
Précision sur test: 0.7628
F1-score pondéré: 0.7286
AUC ROC: 0.8910
Sensibilité (détection PNEUMONIA): 0.9949
Spécificité (détection NORMAL): 0.3761
Temps d'inférence moyen: 0.00000 secondes par image
```

Figure 7 : Rapport de classification détaillé du SVM.

5. Apprentissage Profond

Nous avons testé deux modèles profonds : ResNet18 pour la classification et U-Net pour la segmentation. Le modèle ResNet18 a été entraîné sur les images prétraitées et a montré de très bons résultats avec une précision de 85,6%.

```
Précision sur test: 0.7997

Rapport de classification détaillé:

```

	precision	recall	f1-score	support
NORMAL	0.90	0.53	0.66	234
PNEUMONIA	0.77	0.96	0.86	390
accuracy			0.80	624
macro avg	0.83	0.74	0.76	624
weighted avg	0.82	0.80	0.78	624

Figure 8 : Résultats du modèle ResNet18.

```

Précision sur le test: 0.8558

Rapport de classification:

```

	precision	recall	f1-score	support
NORMAL	0.94	0.65	0.77	234
PNEUMONIA	0.82	0.98	0.89	390
accuracy			0.86	624
macro avg	0.88	0.82	0.83	624
weighted avg	0.87	0.86	0.85	624

Figure 9 : Exemple de sortie du modèle U-Net (segmentation).

6. Conclusion

Ce projet a permis de mettre en œuvre des techniques classiques et profondes pour la détection automatique de la pneumonie.

Le SVM linéaire s'est révélé le plus performant en termes de simplicité et d'efficacité. ResNet18 confirme l'intérêt du deep learning pour ce type de tâche.

Des perspectives d'amélioration incluent l'augmentation de données, l'ajout de classes intermédiaires (pneumonie bactérienne vs virale), et l'explicabilité des modèles.