

MAI 2025

RAPPORT SUR L'ÉVALUATION COMPARATIVE DE MODÈLES DEEP LEARNING ET LEUR IMPACT ÉCOLOGIQUE



Étudiante : Hafsa Moumni

Formation : Master Data Science – Ynov Campus Lyon

Date : Mai 2025

Outil : CodeCarbon



Objectif du projet

Ce projet a pour objectif de comparer plusieurs modèles d'apprentissage supervisé (SVM, CNN, DenseNet, VGG) sur des critères à la fois de performance (précision) et de responsabilité environnementale (temps d'entraînement et émissions carbone).

L'expérimentation a été menée localement, sans GPU, avec un processeur Intel i7 sur Windows 11, en utilisant CodeCarbon pour mesurer l'impact énergétique.

Données & Méthodologie

- **Dataset** : Images en niveaux de gris de taille 28x28, répliquées en 3 canaux pour VGG (format final : 32x32x3)
- **Taille du dataset** : 25 000 images
- **Tâche** : Classification multi-classes
- **Mesures effectuées** :
 - Précision sur jeu de test
 - Temps d'entraînement
 - Émissions carbone (kg CO₂eq)

Modèles évalués

Modèle	Description
SVM	Modèle linéaire de base
DenseNet	Réseau convolutif léger
CNN	Réseau personnalisé avec 3 couches
VGG-Pretrained	VGG16 avec poids ImageNet (non entraînable)
VGG-Untrained	Architecture VGG entraînée from scratch

Résultats

Modèle	Précision	Temps (s)	Emissions (kg CO ₂ eq)
SVM	0.184	72.38	0.000135
DenseNet	0.309	5.98	0.000011
CNN	0.507	17.19	0.000032
VGG-Pretrained	0.558	301.62	0.000563
VGG-Untrained	0.522	24.02	0.000045

VGG-Pretrained est le plus précis

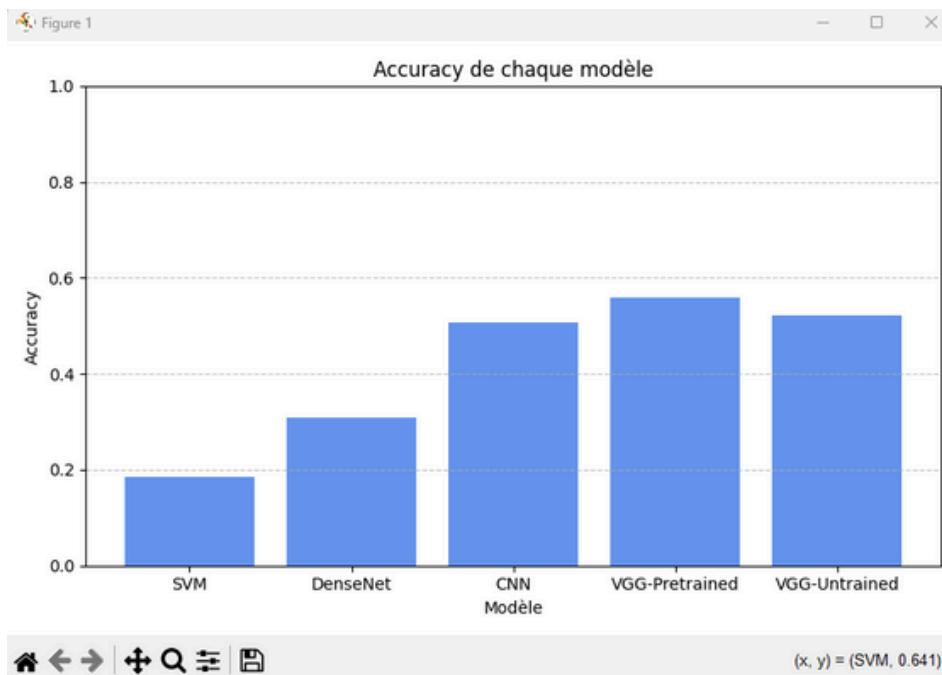
DenseNet est le plus économe en énergie

VGG génère ~10x plus d'émissions que DenseNet

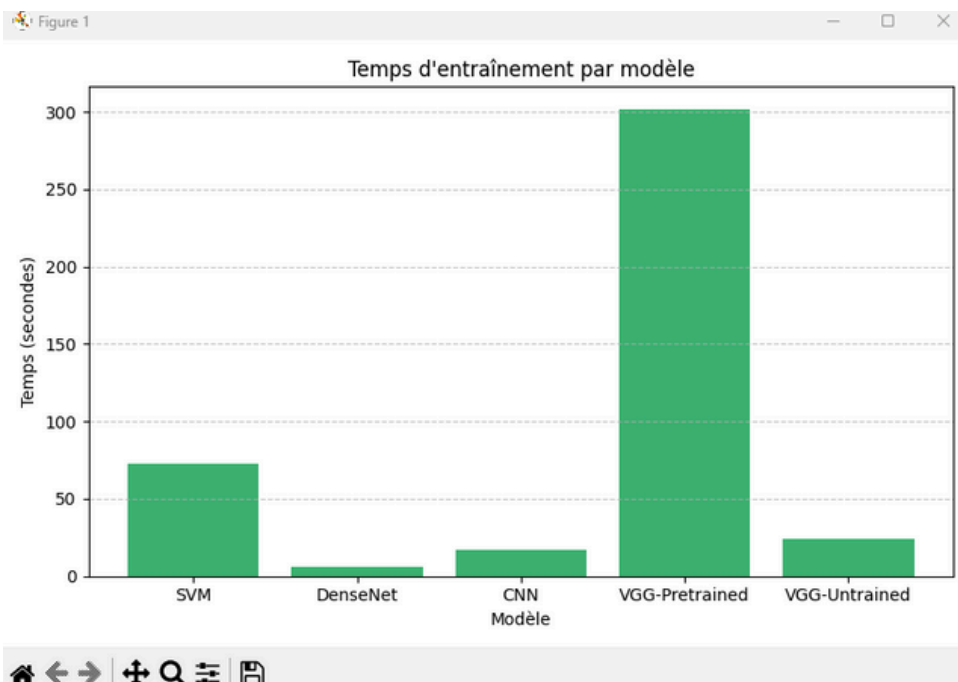
Visualisations

Les graphes suivants ont été générés à partir du script `visualize_results.py` :

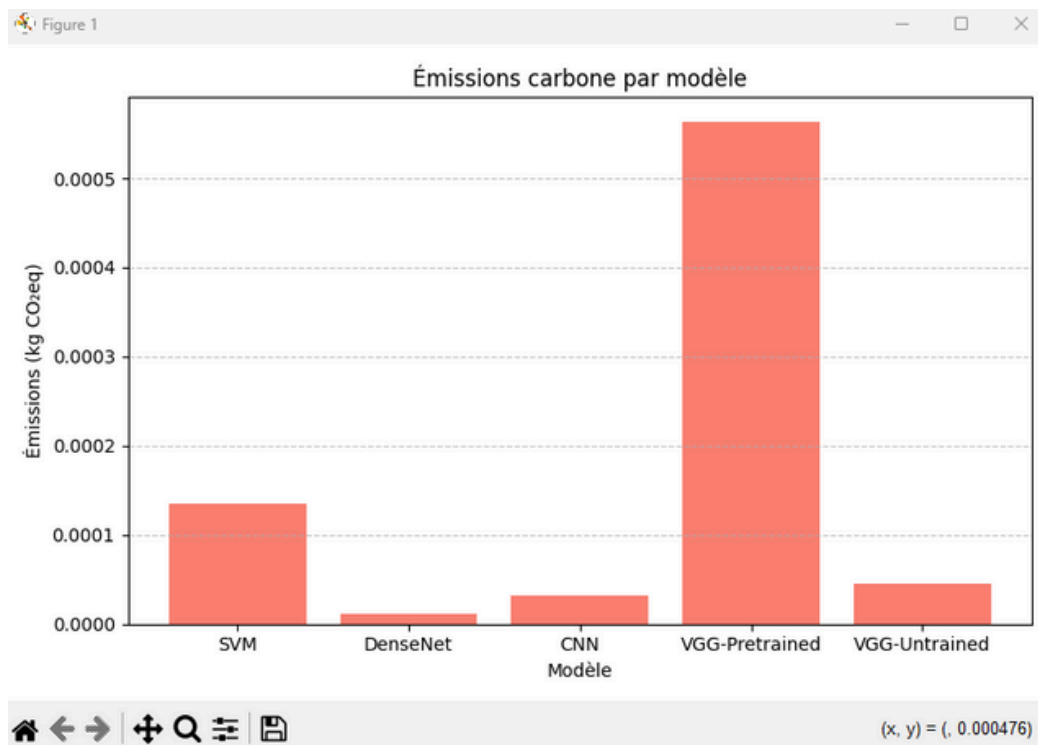
- Précision par modèle



- Temps d'entraînement



- Émissions carbone



Ils ont été intégrés au poster Canva pour appuyer la comparaison.

Choix techniques :

Pourquoi GridSearch n'a pas été utilisé ?

Initialement, l'objectif était de rester sobre énergétiquement.

- Pas de GridSearch automatique au départ.

Au lieu de cela, plusieurs configurations simples ont été testées manuellement pour chaque modèle.

Ensuite, un script d'optimisation via *GridSearchCV* a été préparé pour les 3 meilleurs modèles afin de tester des hyperparamètres, mais cette partie a été optionnelle selon les ressources disponibles.



Conclusion

*Les modèles pré-entraînés comme **VGG** offrent des performances optimales mais avec un coût environnemental important.*

***DenseNet** représente un excellent compromis entre précision et responsabilité écologique.*

Cette étude met en avant l'importance de choisir les bons modèles pas uniquement en fonction de leur performance, mais aussi en fonction de leur coût énergétique, dans une logique de Data Science responsable.