

Classification d'Articles Alimentaires pour le Tri Automatisé

Rapport de Résultats et Recommandations

À l'attention de : Direction des Opérations & Logistique

Date : 30 novembre 2025

Résumé Exécutif

L'objectif de ce projet était de développer une intelligence artificielle capable de reconnaître automatiquement 36 variétés de fruits et légumes pour automatiser le tri en entrepôt.

Nos tests ont comparé deux approches. L'approche classique (MobileNetV2) atteint **65%** de précision et reste limitée par le manque de données. L'approche innovante utilisant un modèle de vision avancé (CLIP) atteint **97%** de précision immédiate.

Recommandation : Déployer la solution basée sur CLIP pour garantir la fiabilité opérationnelle immédiate.

Table des matières

1	Contexte et Enjeux	2
2	Méthodologie Adoptée	2
2.1	Stratégie A : Amélioration du modèle classique (MobileNetV2)	2
2.2	Stratégie B : Utilisation d'un modèle de fondation (LLM/VLM)	2
3	Analyse des Résultats	2
3.1	Diagnostic du modèle MobileNetV2	2
3.2	Performance du modèle CLIP	3
4	Impact Opérationnel et Risques	3
5	Recommandations	3

1 Contexte et Enjeux

La grande distribution fait face à des pertes significatives dues aux erreurs de tri manuel des fruits et légumes (stockage en mauvaises conditions, erreurs d'étiquetage). L'automatisation par vision par ordinateur vise à :

- Réduire les erreurs humaines d'identification.
- Accélérer la cadence de tri sur les convoyeurs.
- Diminuer le gaspillage alimentaire.

Ce rapport présente les résultats de la seconde phase de prototypage, visant à améliorer les performances initiales qui étaient insuffisantes.

2 Méthodologie Adoptée

Face au défi principal du projet le très faible nombre d'images disponibles pour l'entraînement (environ 10 photos par fruit ou légume) nous avons testé deux stratégies distinctes.

2.1 Stratégie A : Amélioration du modèle classique (MobileNetV2)

Nous avons tenté d'optimiser le réseau de neurones standard utilisé lors de la première session.

- **Augmentation des données** : Nous avons artificiellement multiplié les images en appliquant des rotations, des zooms et des variations de contraste pour "forcer" le modèle à mieux généraliser.
- **Régularisation** : Des techniques mathématiques (Dropout, L2) ont été ajoutées pour empêcher le modèle d'apprendre par cœur les images d'entraînement sans comprendre les concepts.

2.2 Stratégie B : Utilisation d'un modèle de fondation (LLM/VLM)

Nous avons utilisé une technologie de rupture : le modèle CLIP (Vision-Language Model).

- **Principe** : Contrairement au modèle classique qui doit tout apprendre de zéro, ce modèle a déjà "vu" des millions d'images sur internet. Il sait déjà à quoi ressemble une "pomme" ou un "navet".
- **Méthode** : Nous ne l'avons pas entraîné (méthode dite *Zero-Shot*), nous lui avons simplement demandé de classer nos images en se basant sur ses connaissances préalables.

3 Analyse des Résultats

Le tableau ci-dessous compare les performances des deux solutions testées sur les données de validation (images jamais vues par le modèle lors de l'entraînement).

Modèle	Précision (Accuracy)	Fiabilité	Ressources Requises
MobileNetV2 (Optimisé)	≈ 60% - 65%	Moyenne	Faibles (Léger, embarquable)
CLIP (VLM)	97.18%	Excellent	Élevées (Nécessite GPU)

TABLE 1 – Comparaison des performances des modèles

3.1 Diagnostic du modèle MobileNetV2

Malgré les améliorations techniques, ce modèle souffre de surapprentissage (Overfitting). Il obtient 99% de réussite sur les images qu'il connaît, mais échoue sur les nouvelles images.

- **Cause :** 10 images par catégorie sont insuffisantes pour qu'un réseau classique apprenne les variations infinies de la nature (forme, couleur, lumière).
- **Erreurs types :** Confusion entre *Radis* et *Carotte* (formes allongées similaires), ou *Ail* et *Oignon*.

3.2 Performance du modèle CLIP

Le modèle CLIP a démontré une robustesse exceptionnelle avec plus de 97% de réussite immédiate. Il distingue sans effort les nuances subtiles entre des légumes visuellement proches grâce à sa compréhension sémantique de l'image.

4 Impact Opérationnel et Risques

L'utilisation du modèle classique (65% de précision) présenterait des risques opérationnels majeurs :

- **Erreurs de stock :** 1 article sur 3 serait mal classé, faussant les inventaires.
- **Impact financier :** Risque de vendre un produit cher au prix d'un produit bon marché (ex : confusion entre variétés de pommes).

L'utilisation du modèle CLIP (97% de précision) réduit le taux d'erreur à moins de 3%, ce qui est acceptable pour un processus industriel assisté.

5 Recommandations

Sur la base des résultats obtenus, nous formulons les recommandations suivantes pour la mise en production :

1. **Choix Technologique :** Adopter l'approche VLM (CLIP) pour le prototype. C'est la seule solution viable actuellement compte tenu du faible volume de données client.
2. **Infrastructure :** Prévoir une infrastructure serveur adaptée (GPU) car ce modèle est plus lourd à faire tourner que la solution classique.
3. **Stratégie Data (Long terme) :** Si le client souhaite absolument utiliser une solution légère (embarquée sur de petites caméras), il est impératif de lancer une campagne de collecte de données pour atteindre au moins 100 à 200 images par fruit. Cela permettrait de ré-entraîner le modèle MobileNetV2 efficacement dans un second temps.

Créé à partir des analyses du notebook `Project_Food_classification_v2.ipynb`.