Module : L'apprentissage statique

**Rapport de projet :**

**L'implémentation des algorithmes de l'apprentissage statique**

Table des matières

[1. Problematique: 5](#_Toc188515403)

[2. Description du dataset : 6](#_Toc188515404)

[3. Pré-traitement des données : 7](#_Toc188515405)

[3.1Histogrammes de couleurs \& Local Binary Patterns : 7](#_Toc188515406)

[3.2 HOG (Histogram of Oriented Gradients): 8](#_Toc188515407)

[3.3 Ondelettes : 9](#_Toc188515408)

[4 k-plus proches voisins : 11](#_Toc188515409)

[4.1 Definition 11](#_Toc188515410)

[4.2 Processus d'entrainement 12](#_Toc188515411)

[4.2.1 Résultat avec HOG : 13](#_Toc188515412)

[4.2.2 Resultat avec Histogrammes de couleurs: 14](#_Toc188515413)

[4.2.3 Resultat avec les ondelettes : 15](#_Toc188515414)

[4.3 Résumé des Résultats de K-NN : 16](#_Toc188515415)

[5 Support Vector Machine (SVM) : 17](#_Toc188515416)

[5.1 Definition : 17](#_Toc188515417)

[5.2 Processus d'entrainement : 18](#_Toc188515418)

[5.2.1 Resultat avec Histogrammes de couleurs : 19](#_Toc188515419)

[6 Decision Tree : 20](#_Toc188515420)

[6.1 Definition : 20](#_Toc188515421)

[6.2 Processus d'entrainement : 21](#_Toc188515422)

[6.3 Le resultat : 21](#_Toc188515423)

7 CONCLUSION:……………………………………………………………………………………………………………….22

# Problematique:

a détection précoce des maladies des plantes est cruciale pour améliorer les rendements agricoles et limiter les pertes économiques. Cependant, l’identification manuelle des maladies est souvent longue et sujette à des erreurs. Ce projet vise à développer un modèle d'apprentissage automatique capable de reconnaître différents types de feuilles et de détecter leur état de santé (sain ou malade). En utilisant une approche automatisée, ce modèle peut aider les agriculteurs et chercheurs à surveiller efficacement la santé des plantes



Figure1- Plante Malade

# Description du dataset :

Le dataset utilisé pour l'entraînement du modèle a été importé depuis la plateforme Kaggle. Ce dataset contient environ 87 000 images RGB de feuilles de cultures, regroupant des spécimens sains et malades classés en 38 catégories différentes.

En raison des contraintes matérielles et du temps de calcul élevé requis, nous avons sélectionné un sous-ensemble représentatif des données. Ainsi, nous avons utilisé 1 000 images par classe pour l'entraînement et 300 images par classe pour la validation et les tests.

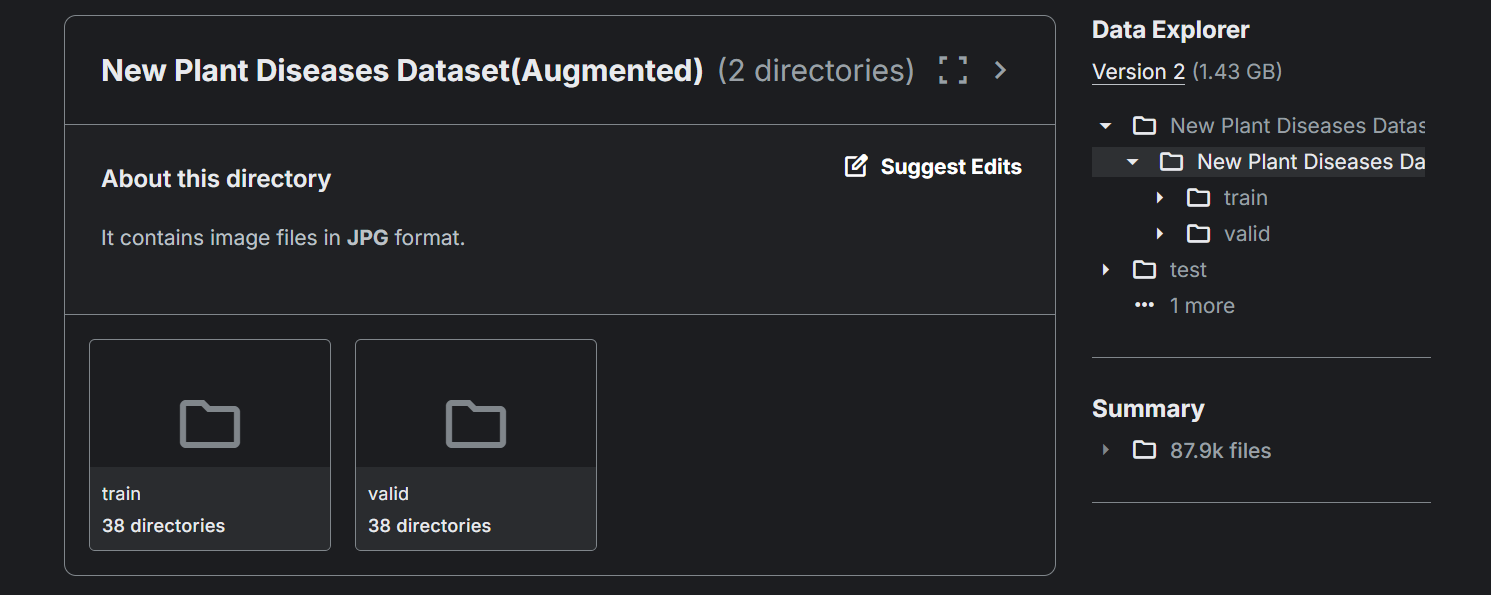


Figure2- Kaggele dataset

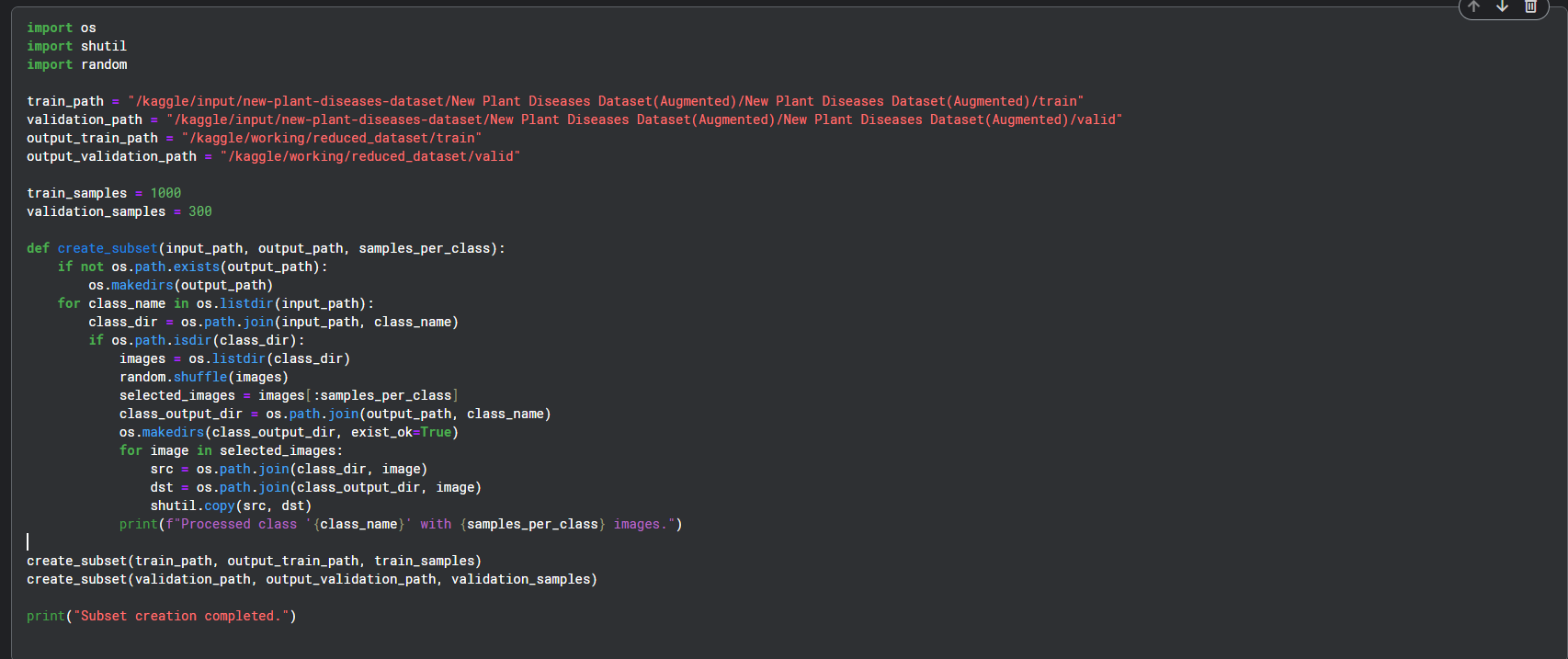


Figure3- Trainning dataset

# Pré-traitement des données :

Afin de garantir de bons résultats pour les modèles entraînés, plusieurs opérations de prétraitement ont été appliquées au dataset. Ces étapes incluent la normalisation des dimensions des images et l’extraction des caractéristiques à l’aide de différentes méthodes, comparées entre elles pour évaluer leur efficacité.

## 3.1Histogrammes de couleurs \& Local Binary Patterns :

Les histogrammes de couleurs sont une méthode statistique utilisée pour extraire des caractéristiques visuelles d'une image,Ils représentent la distribution des couleurs présentes dans une image. combinés à des descripteurs de texture comme les Local Binary Patterns (LBP), permettent de capturer à la fois les informations globales (couleurs) et locales (textures) des feuilles. Cette combinaison enrichit les caractéristiques extraites, ce qui peut améliorer la précision du modèle.

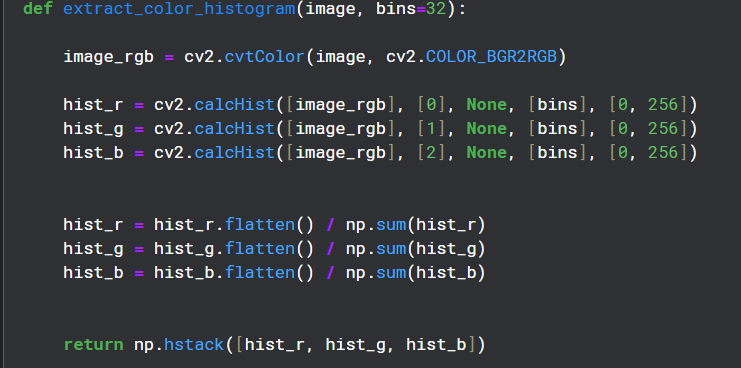


Figure4 - Histogrammes de couleurs

## 3.2 HOG (Histogram of Oriented Gradients):

Cette méthode extrait les contours et les gradients directionnels des images, ce qui est particulièrement utile pour capturer les formes et les structures des feuilles.

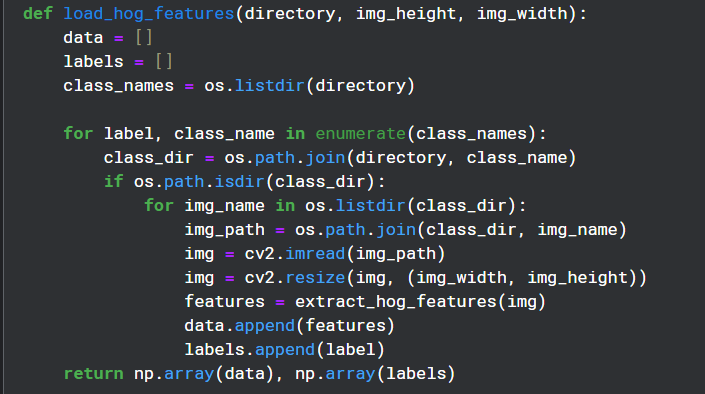


Figure5 - Histogrammes de couleurs

## 3.3 Ondelettes :

Les ondelettes sont des fonctions mathématiques qui possèdent deux caractéristiques essentielles : **la localisation dans le temps (ou l'espace)** et **la capacité à être étendues ou compressées**. En d'autres termes, une ondelette est une fonction qui oscille et qui peut être déplacée dans le temps ou l'espace, tout en conservant une forme similaire. Ces propriétés permettent aux ondelettes de capturer à la fois les détails fins (hautes fréquences) et les grandes structures (basses fréquences) d'un signal ou d'une image. La transformation en ondelettes est obtenue en appliquant une série de convolutions et de translations de la fonction ondelette, ce qui permet de décomposer un signal en plusieurs niveaux de détails à différentes échelles.

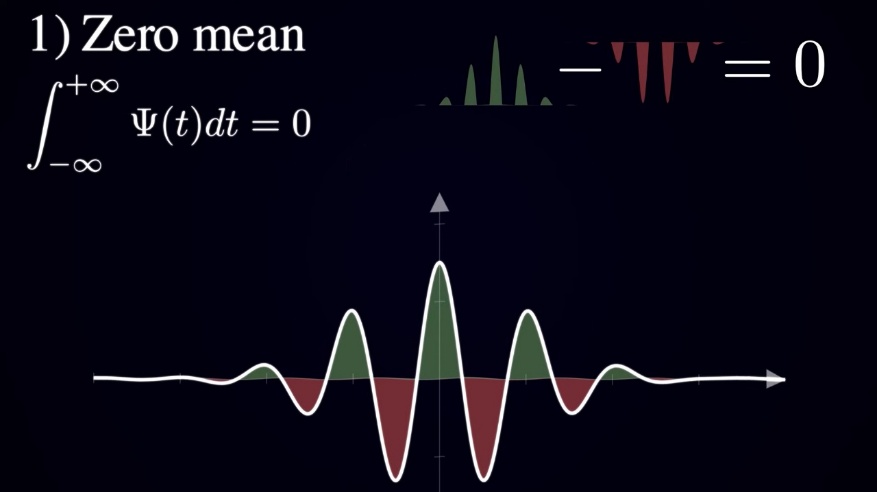
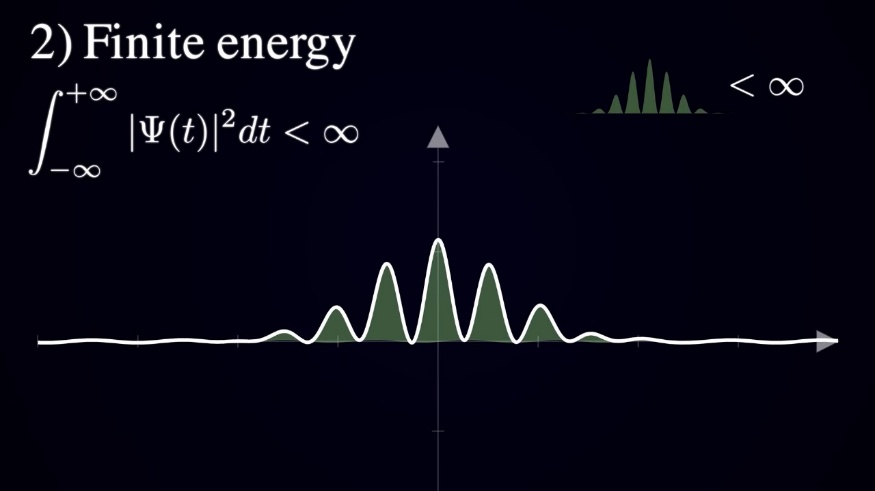
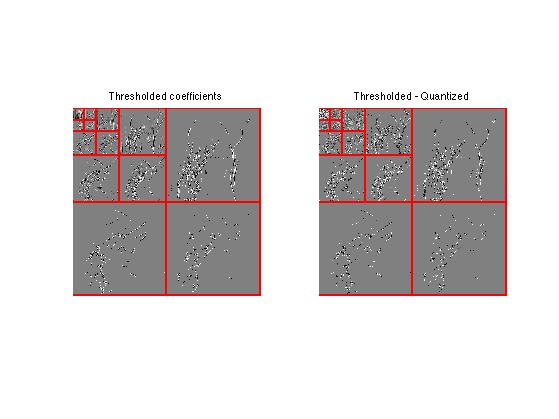


Figure6 - Les caracrestiques mathématiques

Les ondelettes sont un outil puissant pour l'extraction de caractéristiques d'images, car elles permettent une analyse simultanée des informations spatiales et fréquentielles. Le processus consiste à décomposer une image en différentes sous-bandes, chacune représentant des niveaux spécifiques de détails. Cette décomposition est réalisée à l'aide de filtres passe-bas et passe-haut, qui isolent respectivement les formes générales (basses fréquences) et les détails fins comme les bords et les textures (hautes fréquences). Les coefficients résultants, obtenus à partir des sous-bandes, décrivent les caractéristiques importantes de l'image, comme sa texture, ses contours ou ses variations d'intensité. Ces coefficients peuvent ensuite être utilisés dans des applications variées, telles que la classification d'images, la compression ou encore la reconnaissance de motifs.



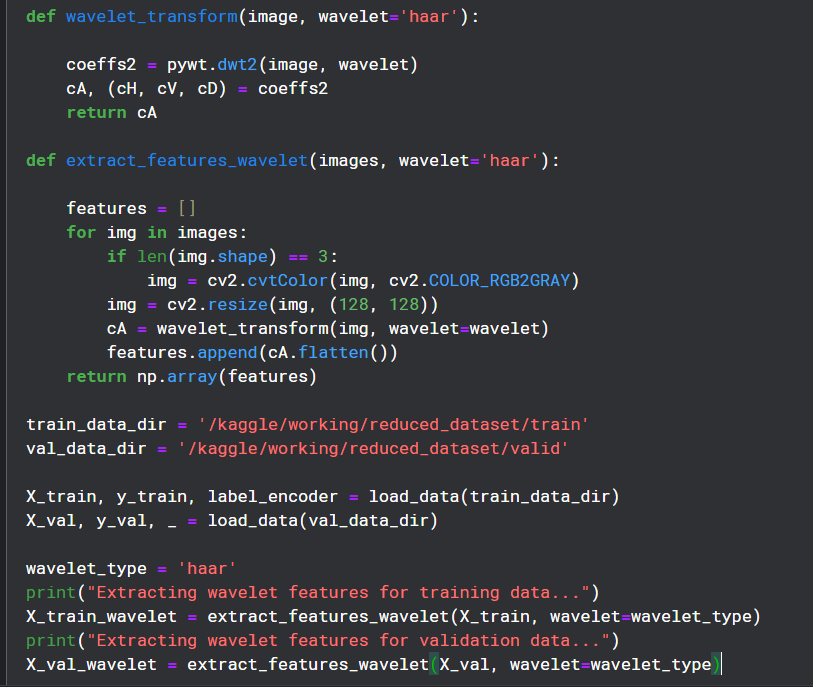
 Figure5 - traitement d’images avec Les ondelettes

Figure6- Implementation de les ondelettes

# k-plus proches voisins :

## 4.1 Definition

Le k-NN (k-Plus Proches Voisins) est un algorithme de classification supervisée qui attribue une étiquette à un nouvel échantillon en fonction de la majorité des étiquettes de ses k voisins les plus proches, déterminés par une mesure de distance (souvent la distance Euclidienne).

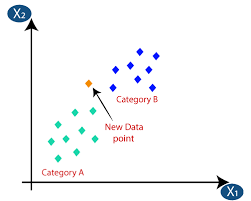


Figure7 - K-NN Algorithm

## 4.2 Processus d'entrainement

* Extraire les données des images nécessaires à l'apprentissage en utilisant HOG,Histogrammes de couleurs ,Ondelettes.
* Réduire la dimensionnalité des caractéristiques HOG pour éviter le surapprentissage et améliorer l'efficacité en utilisant ACP.
* Initialisation du Classifieur k-NN.
* Définir la grille des hyperparamètres.

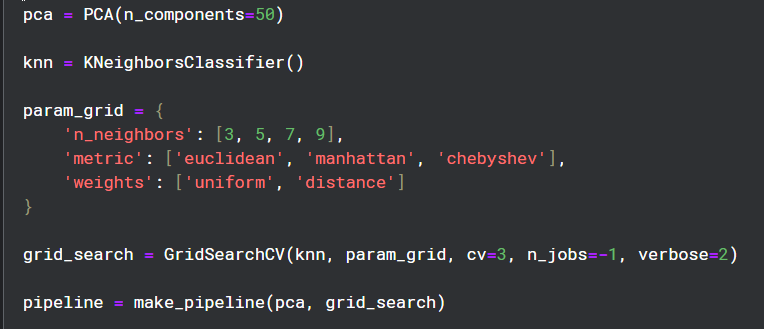


Figure8 - Processus d'entrainement pour K-nn

### 4.2.1 Résultat avec HOG :

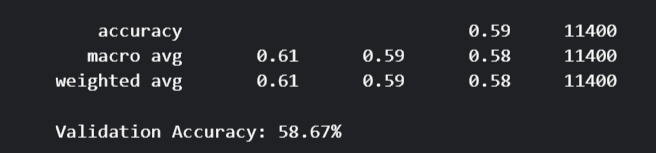


Figure9 - K-NN & HOG Accuracy

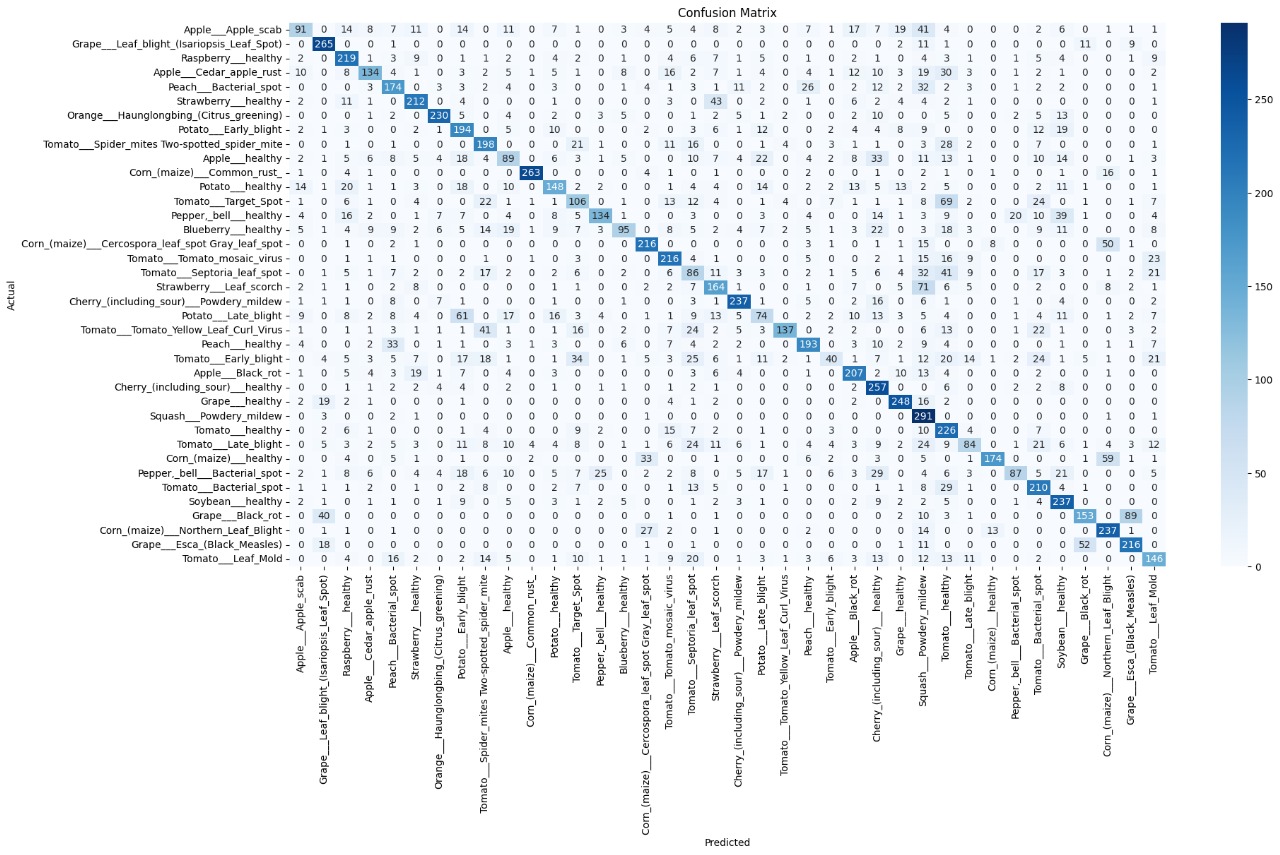


Figure10 - K-NN & HOG Matrice de Confusion

### 4.2.2 Resultat avec Histogrammes de couleurs:

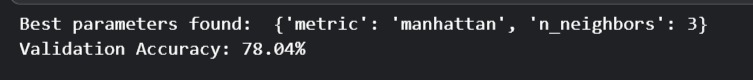


Figure11- K-NN & HRGB Accuracy

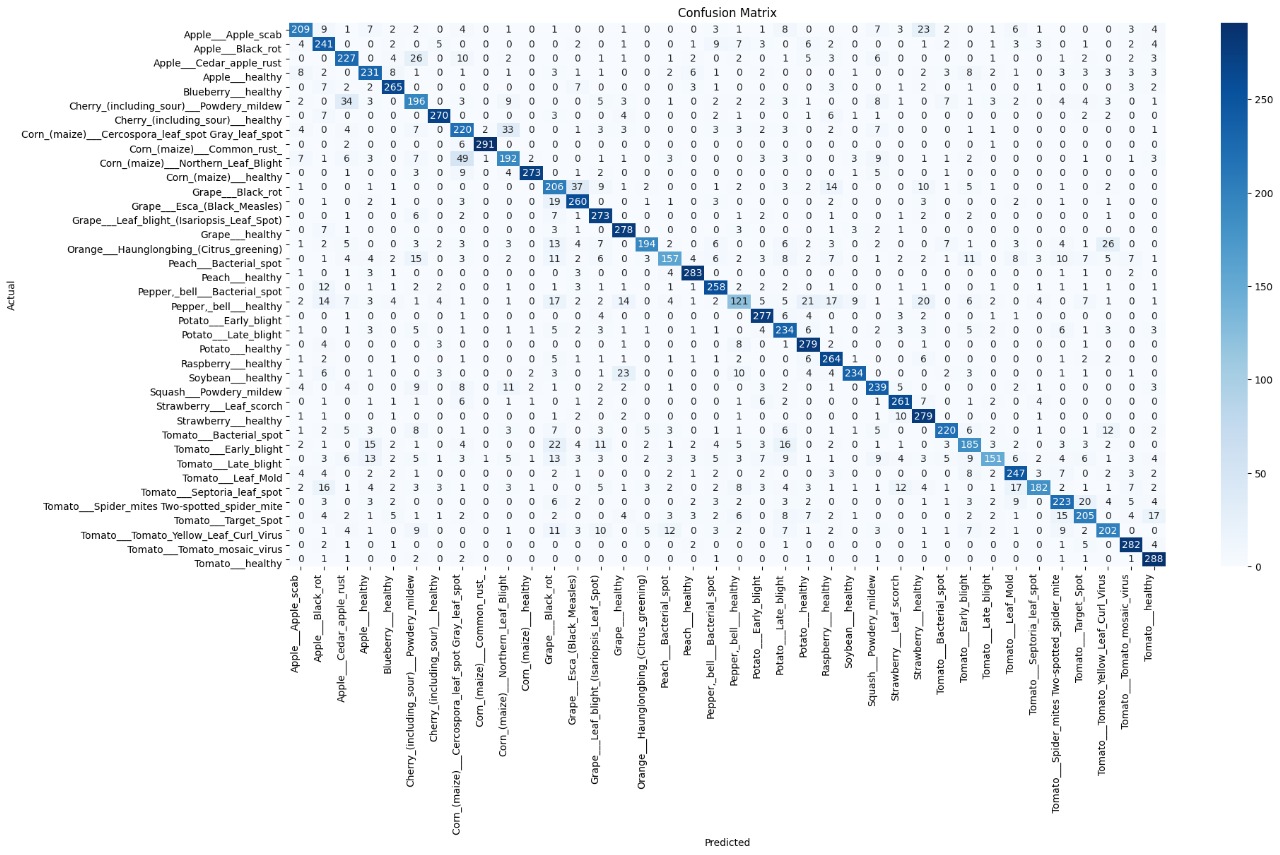


Figure12 - K-NN & HRGB Matrice de Confusion

### 4.2.3 Resultat avec les ondelettes :

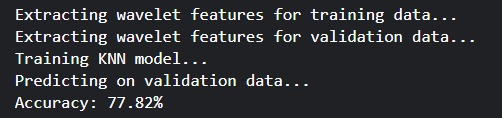


Figure13 - K-NN & Wavelets Accuracy

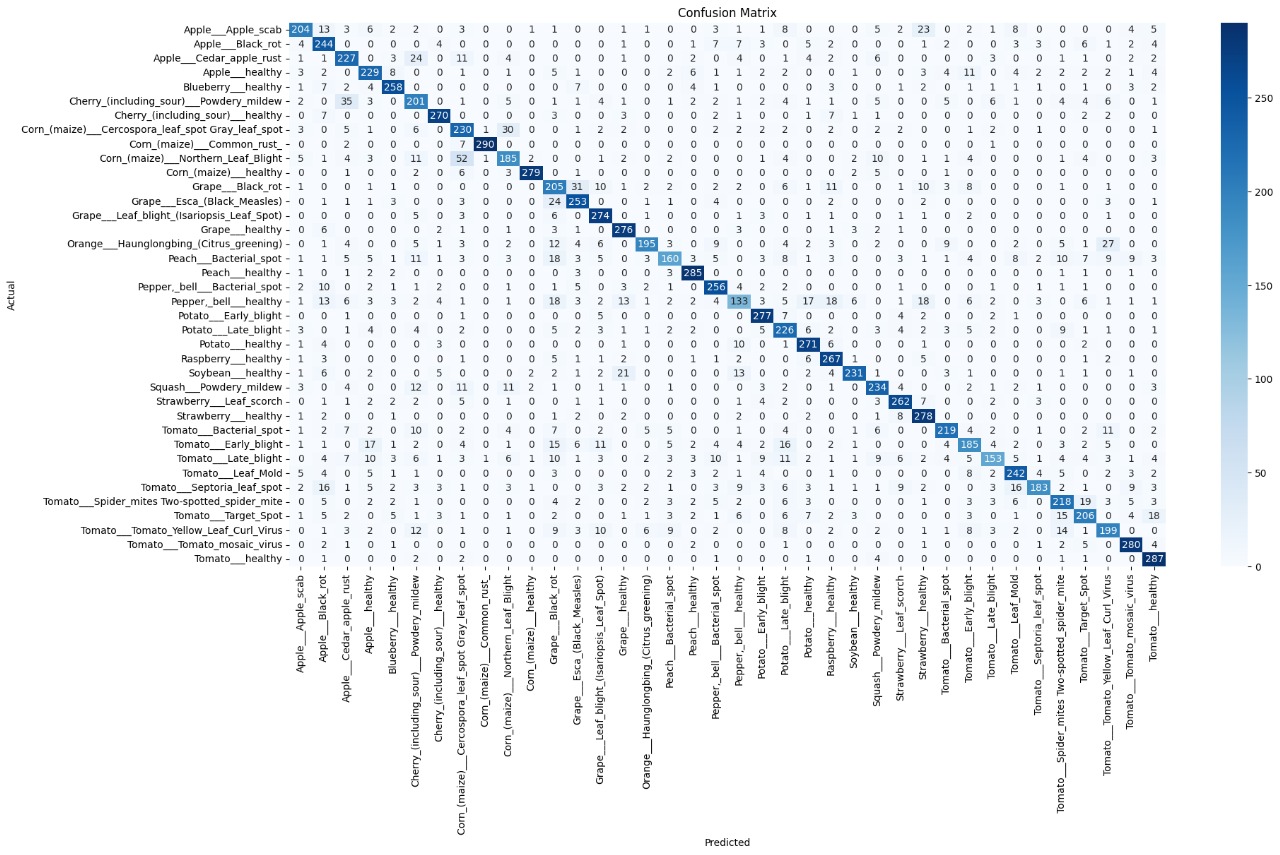


Figure14 - K-NN & Wavelets Matrice de Confusion

## 4.3 Résumé des Résultats de K-NN :

Parmi ces approches, l'utilisation des histogrammes de couleurs s'est avérée être la plus performante avec un taux de précision de **78,12%,** surpassant les autres méthodes.

Cette performance s'explique par la capacité des histogrammes à capturer efficacement les caractéristiques globales de couleur, qui jouent un rôle clé dans la classification des images dans ce contexte.

# Support Vector Machine (SVM) :

## Definition :

Est un algorithme d'apprentissage supervisé utilisé pour les tâches de classification et de régression. Elle est particulièrement efficace pour traiter des données dans des espaces de haute dimension et lorsque le nombre de dimensions est supérieur au nombre d'échantillons.

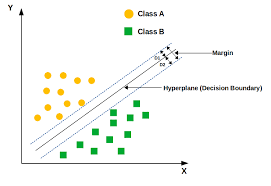


Figure15 - SVM Algorithm

## Processus d'entrainement :

* Extraire les données des images nécessaires à l'apprentissage en utilisantHistogrammes de couleurs.
* Initialisation du du modèle SVM.
* Définir la grille des hyperparamètres.
* Entraînement du modèle et Extraction des meilleurs hyperparamètres.



Figure16 - Processus d'entrainement pour SVM

### Resultat avec Histogrammes de couleurs :

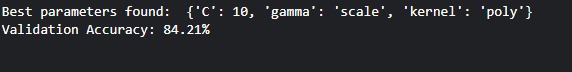


Figure17- SVM & HOG Accuracy

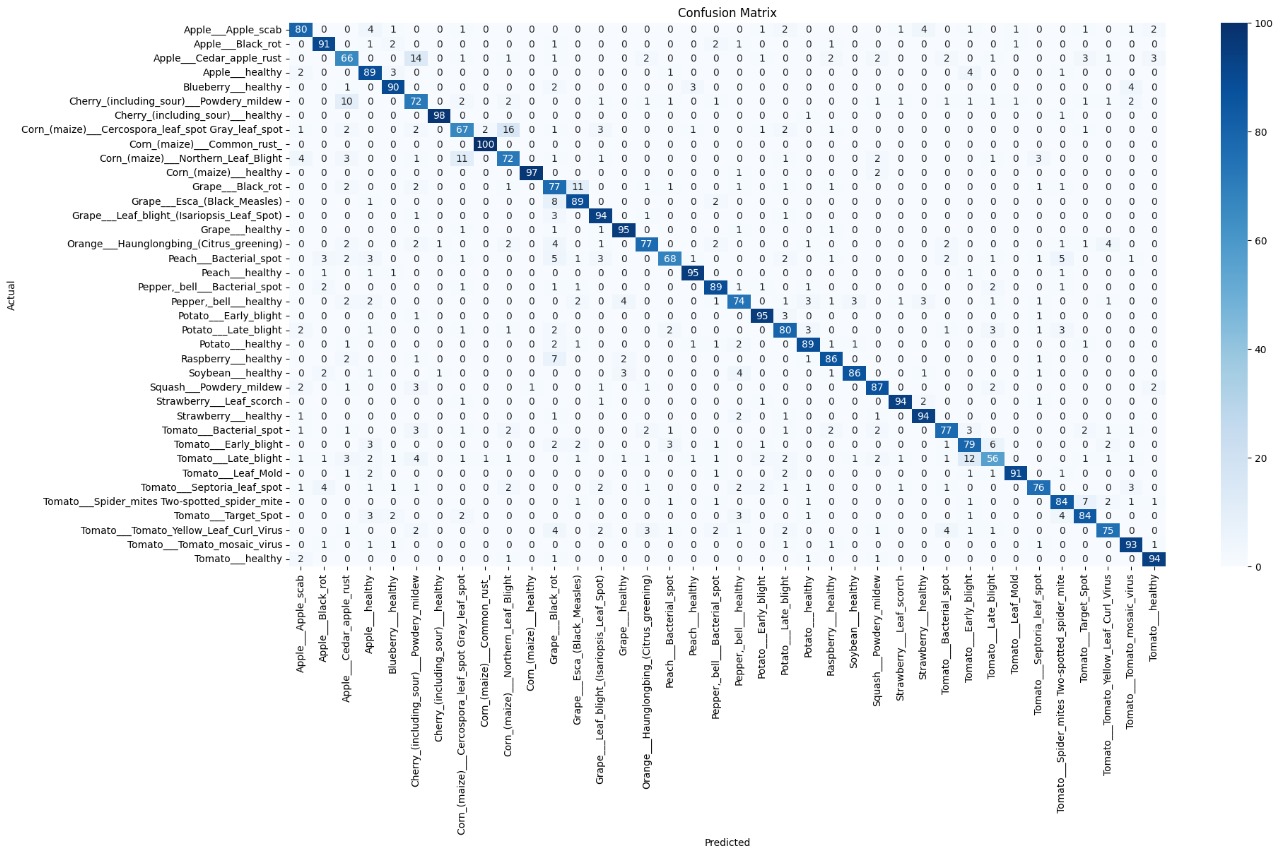


Figure18- SVM & HOG Confusion Matrice

# Decision Tree :

## Definition :

Un arbre de décision est un algorithme d’apprentissage supervisé utilisé pour la classification et la régression. Il fonctionne en divisant les données en sous-groupes basés sur des conditions ou des questions successives (appelées nœuds). Chaque branche de l’arbre représente une décision, et chaque feuille correspond à un résultat ou une catégorie

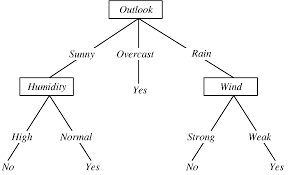


Figure19- Decision Tree Algorithm

## Processus d'entrainement :

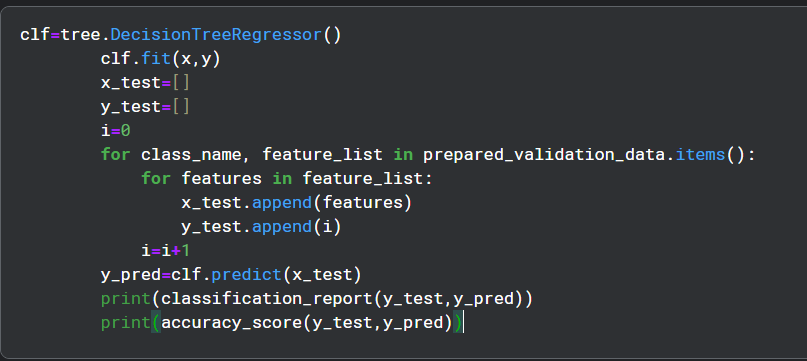
* Extraire les données des images nécessaires à l'apprentissage en utilisant Histogrammes de couleurset.
* Extraire les informations de texture en utilisant LBP.
* Initialisation du du modèle Decision Tree.

Figure20 - Processus d'entrainement pour Decision Tree

## Le resultat :

Validation Accuracy: **74.84%**

# CONCLUSION :

SVM est le choix optimal pour ce projet en termes de précision. Cependant, k-NN combiné aux histogrammes de couleurs reste une alternative intéressante, notamment pour des systèmes nécessitant des implémentations simples. L’arbre de décision, bien que moins performant, offre l’avantage d’une interprétation claire et rapide. Enfin, l’utilisation de HOG n’a pas donné des résultats convaincants dans ce cas, ce qui pourrait être lié à des caractéristiques mal adaptées ou à des limitations dans l’extraction des contours pour ce dataset.