



# Rapport sur la constitution de notre base de données sur les types de pâtes - Deep Learning

Mathieu Campan, Priscilia Gonthier, Sarah Abou Daya,  
Amine Benchaaboun, Simon Hautesserres  
Groupe L1 - 3

Département Sciences du Numérique  
2022-2023

## Table des matières

|     |   |   |
|-----|---|---|
| 1   | Sujet choisi : . . . . .                            | 3 |
| 1.1 | Explication illustrée du sujet : . . . . .          | 3 |
| 1.2 | Lien vers la base de données sur GitHub : . . . . . | 3 |
| 2   | Méthodologie : . . . . .                            | 4 |
| 2.1 | Acquisition des données : . . . . .                 | 4 |
| 2.2 | Répartition des données : . . . . .                 | 6 |
| 3   | Pronostic : . . . . .                               | 6 |

# 1 Sujet choisi :

## 1.1 Explication illustrée du sujet :

Le sujet de notre projet porte sur la classification de différentes variétés de pâtes à travers une base de données partagée en 10 classes.

Nous avons choisi ce sujet en raison de notre amour pour les pâtes et de notre envie de pouvoir différencier simplement les différents types de pâtes. Nous avons également été inspiré par les nombreux plats de pâtes différents existants dans le monde entier et avons été curieux de découvrir les différences subtiles entre les variétés.

En utilisant l'apprentissage profond pour classer les pâtes, nous espérons non seulement en apprendre plus sur les différentes variétés, mais également aider d'autres amateurs de pâtes à mieux comprendre et apprécier la richesse de la cuisine italienne.



FIGURE 1 – "Nous aimons les pâtes"

## 1.2 Lien vers la base de données sur GitHub :

Voici le lien pour accéder à notre base de données :

<https://github.com/MathieuCampan/DeepPasta>

Nous avons également développé un script pour charger nos données dans Google Colab :

[https://colab.research.google.com/drive/15\\_nELci0gMwr5nG-QgX3FOeascwhj2mH?usp=sharing](https://colab.research.google.com/drive/15_nELci0gMwr5nG-QgX3FOeascwhj2mH?usp=sharing)

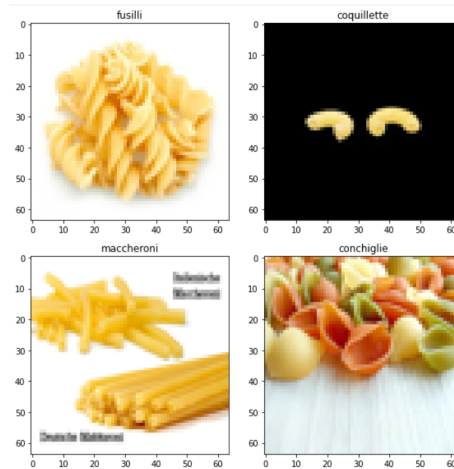


FIGURE 2 – "Exemple de Données Récupérées"

## 2 Méthodologie :

### 2.1 Acquisition des données :

Pour ce faire, nous avons collecté des données pour chaque classe selon les différentes caractéristiques des pâtes telles que leur forme, leur texture, leur couleur, leur ingrédient de base, etc...

Ces données ont été récupérées par scraping depuis le moteur de recherche google. Nous avons utilisé un script python nommé Google-Image-Scraper, accessible sur github.

Pour chaque classe, nous avons récupéré 100 images pour avoir un total de 1000 images pour notre base de données.

Nous nous sommes répartis 2 classes par personnes, ce qui a fait 200 images à récupérer chacun.

Chacune des images a été sélectionnée pour former au mieux une base d'apprentissage cohérente et complète, notamment en variant le nombre (1 ou 2 pâtes à un plat complet de pâtes nature) mais aussi en y ajoutant quelques ingrédients ou des sauces.

Nous avons également sélectionné les images de bonne qualité et pris en priorité les formats de type .jpg pour faciliter le traitement à venir.

Pour l'annotation des images, nous avons simplement donné comme nom le type de pâtes suivi du numéro de l'image pour pouvoir les différencier rapidement.

Pour ce qui est de la détection de doublons, nous avons utilisé le logiciel gratuit Awesome Duplicate Photo Finder, qui a permis de facilement détecter et supprimer les doublons.

|                |                                    |
|----------------|------------------------------------|
| ■ cappelletti  | dossier Cappelletti                |
| ■ conchiglie   | dossiers renommés pour homogénéité |
| ■ coquillettes | Delete README.md                   |
| ■ farfalle     | dossiers renommés pour homogénéité |
| ■ fusilli      | dossiers renommés pour homogénéité |
| ■ lasagne      | Delete README.md                   |
| ■ maccheroni   | dossiers renommés pour homogénéité |
| ■ penne        | dossiers renommés pour homogénéité |
| ■ spaghetti    | dossiers renommés pour homogénéité |
| ■ tagliatelle  | dossier tagliatelle                |
| 📄 README.md    | dossier img                        |

FIGURE 3 – Dossiers des Classes de pâtes GitHub

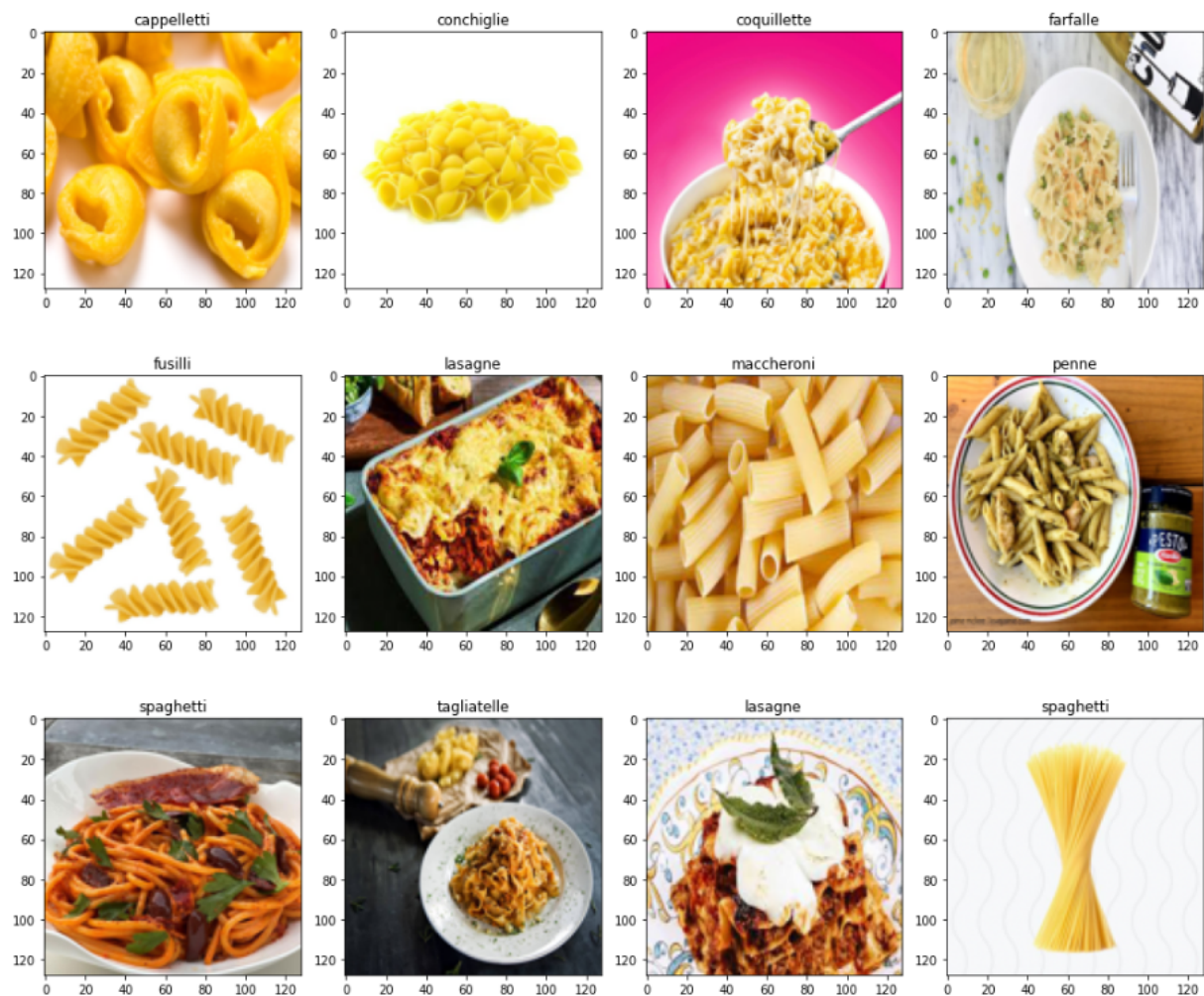


FIGURE 4 – Images issues de notre base de données

## 2.2 Répartition des données :

Pour la répartition des données, nous avons décidés de consacrer 60% des images à l'apprentissage, 20% à la validation et 20% au test.

Dans la base d'apprentissage, nous avons mis des images qui montrent un type de pâtes sous différentes formes et perspectives. Pour les tests et la validation, nous avons mis des images un peu plus compliquées pour vérifier si l'algorithme va bien reconnaître les types. Ces bases comportent notamment des images de plats complets de pâtes, avec de la sauce ou en salade.

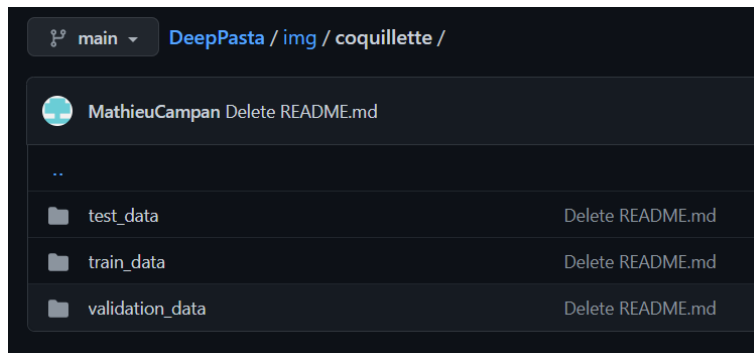


FIGURE 5 – Dossiers des Bases d'Entrainement, de Test et de Validation GitHub

## 3 Pronostic :

Le problème initial n'est pas très compliqué car les classes sont facilement reconnaissables. L'une des difficultés potentielles que nous pourrions rencontrer est due à la pixélisation des images de notre base de données. En effet, nous avons constaté qu'il était compliqué de distinguer à l'oeil nu les différents types de pâtes sur des plats complexes, comme des grattins. Nous pourrions pallier à cela en augmentant la résolution des images, ce qui augmenterait également le temps de calcul.

Cela reste cependant faisable dans un temps raisonnable pour ce projet.