

# Rapport de Progrès : Détection Automatique de Tumeurs Mammaires par Ultrasons

## Une Approche Basée sur l'Intelligence Artificielle

|                                 |                      |                         |
|---------------------------------|----------------------|-------------------------|
| <b>Sosane Mahamoud Houssein</b> | <b>Zeï nab Touré</b> | <b>Abidé Badjoudoum</b> |
| HOUS92310307                    | TOUZ63280208         | BADA09349800            |
| <i>hous44@uqo.ca</i>            | <i>touz08@uqo.ca</i> | <i>bada20@uqo.ca</i>    |

25 mars 2025

## 1 Introduction

Ce document présente l'avancement de notre projet de détection de tumeurs mammaires à partir d'images ultrasonores. Conformément à notre proposition initiale, nous avons complété la phase d'analyse exploratoire des données (EDA) et préparé le pipeline de prétraitement.

## 2 État d'Avancement

### 2.1 Données Utilisées

Les résultats proviennent de données récentes<sup>1</sup>.

Le jeu de données contient :

- 780 images ultrasonores (format PNG)
- Résolution moyenne :  $500 \times 500$  pixels
- Répartition initiale :
  - Bénignes : 437 images (56%)
  - Malignes : 210 images (27%)
  - Normales : 133 images (17%)

### 2.2 Analyse Exploratoire (EDA)

Les résultats clés de notre analyse :

#### 2.2.1 Caractéristiques des Images

TABLE 1 – Statistiques des dimensions d'images

|         | Hauteur (px) | Largeur (px) |
|---------|--------------|--------------|
| Minimum | 310          | 190          |
| Maximum | 719          | 1048         |

---

1. Al-Dhabyani, W. et al. (2020). *Dataset of breast ultrasound images*. Data in Brief, 28, 104863. DOI : 10.1016/j.dib.2019.104863.

### 2.2.2 Analyse des Masques

TABLE 2 – Superficie tumorale moyenne (en pixels)

| Classe | Moyenne | Médiane | Min | Max     |
|--------|---------|---------|-----|---------|
| Bénin  | 20,734  | 10,263  | 804 | 209,121 |
| Malin  | 43,376  | 34,433  | 569 | 167,411 |
| Normal | 0       | 0       | 0   | 0       |

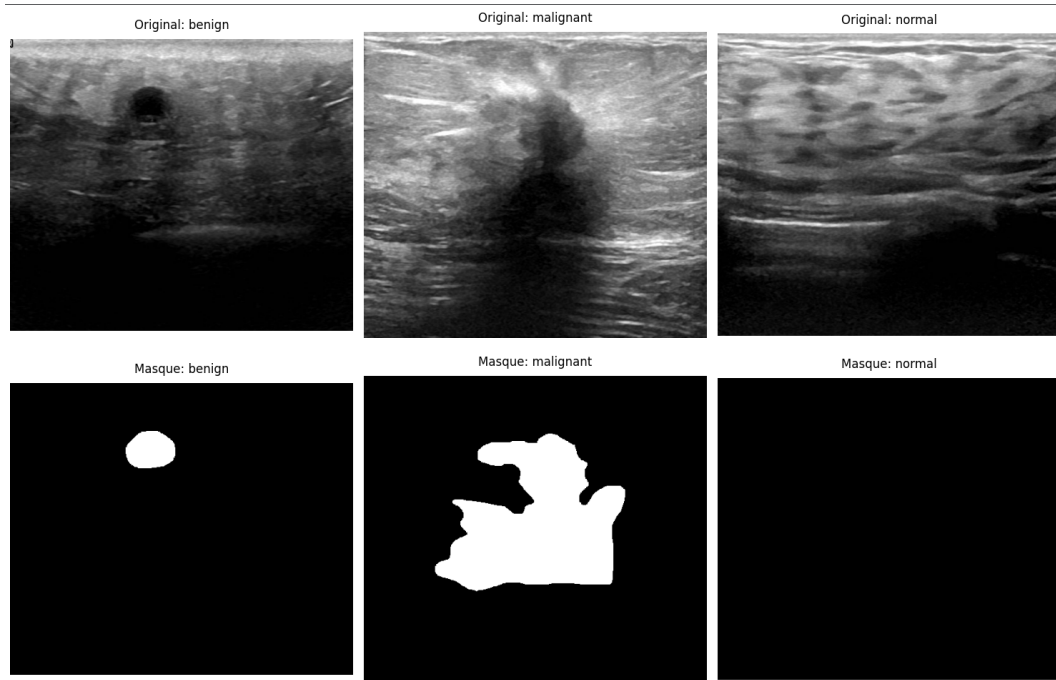


FIGURE 1 – Exemples d’images avec leurs masques correspondants

## 3 Adaptations Méthodologiques

Suite à l’EDA, nous avons ajusté notre approche :

### 3.1 Résolution des Problèmes Identifiés

— **Déséquilibre de classes** : Implémentation d’une fonction de coût pondérée

$$w_c = \frac{N}{C \times N_c} \quad (1)$$

où  $N$ =total,  $C$ =classes,  $N_c$ =échantillons par classe

— **Normalisation des dimensions** : Redimensionnement à  $256 \times 256$  px avec préservation du ratio d’aspect

— **Utilisation des masques** : Intégration comme 4ème canal d’entrée

### 3.2 Pipeline de Prétraitement

1. Chargement des paires (image, masque)

2. Redimensionnement avec remplissage intelligent
3. Augmentation de données :
  - Rotation aléatoire ( $\pm 15^\circ$ )
  - Retournement horizontal
  - Ajustement de contraste
4. Normalisation :  $\mu = [0.485, 0.456, 0.406]$ ,  $\sigma = [0.229, 0.224, 0.225]$

## 4 Prochaines Étapes

- Implémentation des architectures CNN (ResNet18, EfficientNet)
- Expérimentation avec des mécanismes d'attention
- Évaluation comparative des modèles
- Optimisation des hyperparamètres

## Annexe

### Exemple de Code

```
# Chargement des données
dataset = BreastUltrasoundDataset(
    image_paths=X_train,
    mask_paths=mask_train,
    labels=y_train,
    transform=train_transform
)
```