**BreastScreen-AI :**

**Un outil de détection précoce du cancer du sein à Djibouti par l’Intelligence Artificielle**

**Concept et plan de mise en œuvre du projet**

**Membres de l’équipe : Yacin MOUHOUMED ELMI**

1. **Vue d'ensemble du projet**

Le cancer du sein est la première cause de mortalité par cancer chez les femmes, notamment dans les pays à faibles ressources comme Djibouti. Le manque d'équipements spécialisés (comme les mammographes) et le nombre limité de radiologues compliquent le diagnostic précoce. Le projet **BreastScreen-AI** vise à combler ce déficit à travers une solution d’intelligence artificielle (IA) capable de détecter automatiquement les lésions mammaires sur des échographies, plus accessibles localement. En alignement avec les Objectifs de Développement Durable (ODD), notamment l’ODD 3 (Bonne santé et bien-être) et l’ODD 9 (Innovation et infrastructures), ce projet cherche à améliorer l'accès au dépistage et à renforcer les capacités locales via la technologie.

1. **Objectifs**

* Créer un modèle de deep learning (CNN) capable de classer automatiquement des échographies mammaires (normal, bénin, malin).
* Intégrer ce modèle dans une interface conviviale pour les professionnels de santé.
* Former les utilisateurs à l’usage de cet outil dans les hôpitaux et centres de santé de Djibouti.
* Déployer l’application finale sur **Gradio App** via **Hugging Face Spaces**, permettant un accès web sécurisé.

1. **Contexte**

Plusieurs études démontrent le potentiel des CNN (Convolutional Neural Networks) dans la détection du cancer du sein. Cependant, la plupart utilisent des données provenant de contextes européens ou asiatiques. Djibouti présente des réalités techniques et socio-sanitaires différentes, notamment un accès limité à la mammographie. L’utilisation d’images échographiques, disponibles localement, combinée à une approche par apprentissage profond adaptée, constitue une solution réaliste et innovante. Les initiatives similaires n’incluent pas encore l’intégration Gradio pour une accessibilité immédiate des modèles IA par les professionnels médicaux.

1. **Méthodologie**

Nous utiliserons une approche d’apprentissage supervisé avec des réseaux convolutifs (CNN).  
Les étapes principales sont :

* Prétraitement des données (resize, normalisation)
* Entraînement d’un modèle CNN léger (MobileNetV2 ou EfficientNet)
* Évaluation sur des images échographiques annotées
* Conversion du modèle en format exportable (H5 ou SavedModel)
* Déploiement sur **Gradio (Hugging Face Spaces)** pour interface web

1. **Diagramme de conception architecturale**

Voici l’architecture globale :

***Échographie → Prétraitement → Modèle CNN → Prédiction → Interface Gradio → Résultat***

**Composants :**

* Module de chargement d’image (Upload ou webcam)
* Bloc de traitement (resize + normalisation)
* Prédiction par modèle CNN
* Interface utilisateur avec Gradio (Hugging Face)

1. **Sources de données**

Les données initiales proviennent du dataset public **BUSI (Breast Ultrasound Images Dataset)** contenant 780 échographies classées (normal, bénin, malin). Les images sont prétraitées (256x256 pixels, normalisation des pixels, format RGB). À terme, des partenariats seront établis avec les hôpitaux djiboutiens pour obtenir des données locales (avec approbation éthique).

1. **Revue de littérature (synthèse)**

La littérature confirme la supériorité des CNN pour la classification d'images médicales. Les modèles comme MobileNetV2 et EfficientNet offrent un bon compromis entre performance et vitesse. Les études existantes révèlent toutefois un manque de représentativité des données africaines, justifiant le besoin d'un projet comme BreastScreen-AI. (cf. Zhang et al., Mohamed et al., 2023 ; Muduli et al., 2022)

**Plan de mise en œuvre**

* 1. **Technology Stack**

|  |  |
| --- | --- |
| **Composant** | **Outil choisi** |
| Langage | Python |
| Framework IA | TensorFlow / Keras |
| Dataset | BUSI (images échographiques) |
| Interface | Gradio |
| Déploiement | Hugging Face Spaces |
| Environnement | Google Colab + GitHub |
| Outils annexes | NumPy, OpenCV, Matplotlib, Pandas |

* 1. **Chronologie du projet (exemple)**

|  |  |
| --- | --- |
| **Semaines** | **Activité** |
| Semaine 1 | Recherche & revue de littérature |
| Semaine 2 | Préparation des données (nettoyage, étiquetage) |
| Semaine 3 | Développement du modèle CNN |
| Entraînement & validation |
| Test sur données de test (x\_test) |
| Semaine 4 | Déploiement sur Gradio (Hugging Face) |
| Évaluation, rapport & présentation finale |

* 1. **Étapes importantes**
* Finalisation du modèle CNN avec une bonne précision
* Démonstration fonctionnelle de l’interface Gradio ( en cours)
* Présentation finale devant le jury avec démonstration en direct ( à la fin de la formation)
  1. **Défis et atténuations**

|  |  |
| --- | --- |
| **Défis** | **Stratégie d’atténuation** |
| Faible quantité de données | Utilisation de data augmentation et fine-tuning |
| Accès limité aux données locales | Partenariat avec les hôpitaux et obtention d'accords éthiques dans le futur |
| Risque de surapprentissage | Régularisation (dropout), cross-validation |
| Déploiement web | Utilisation de Gradio + Hugging Face pour simplifier |

* 1. **Considérations éthiques**
* **Confidentialité des données** : Anonymisation des images utilisées
* **Consentement éclairé** : Accord des patients pour les images locales (dans le futur)
* **Biais algorithmiques** : Validation croisée avec des jeux diversifiés
* **Transparence** : Code et modèle disponibles sur GitHub (open access)
  1. **Références**
* Dhungel, N. et al. (2017). *A deep learning method for the detection of masses in mammograms*. Medical Image Analysis.
* Muduli, D. et al. (2022). *Breast cancer detection using transfer learning on MRI*. CMIG.
* Mohamed, I. et al. (2023). *Lightweight CNN for breast cancer detection*. arXiv.
* Zhang, Q. et al. (2021). *EfficientNet-based automated breast cancer diagnosis*. IEEE Access.
* BUSI Dataset (2020). *Breast Ultrasound Images Dataset* - Open Access.