

Résumé

Le cancer du poumon est l'une des maladies cancérigènes les plus courantes et sa prévalence continue d'augmenter. La détection précoce est essentielle pour améliorer les chances de survie des patients. Notre projet vise à mettre en place un système d'aide à la détection précoce de cette maladie dans les centres médicaux. L'idée centrale consiste à implémenter deux architectures, à savoir Mask R-CNN et U-Net, afin de déterminer laquelle offre les meilleurs résultats. Ensuite, nous extrairons les caractéristiques de Haralick en utilisant la matrice GLCM et en combinant différents indices. Nous effectuerons également plusieurs expérimentations en utilisant différentes matrices GLCM. Enfin, nous utiliserons l'algorithme K-means pour classer les nodules comme étant bénins ou malins. Notre système sera testé sur des bases de données publiques réputées et fiables.

Mots Clée : cancer du poumon, nodules, Mask R-CNN, U-Net, caractéristiques de Haralick, GLCM, K-means.

Abstract

Lung cancer is one of the most common cancerous diseases, and its prevalence continues to increase. Early detection is crucial to improve the chances of patient survival. Our project aims to implement an early detection system for this disease in medical centers. The central idea is to implement two architectures, namely Mask R-CNN and U-Net, to determine which one offers the best results. Then, we will extract Haralick features using the GLCM matrix and combining different indices. We will also perform several experiments using different GLCM matrices. Finally, we will use the K-means algorithm to classify nodules as benign or malignant. Our system will be tested on reputable and reliable public datasets.

Keywords: Lung cancer, nodules, Mask R-CNN, U-Net, Haralick features, GLCM, K-means.

Table des matières

| | |
|--|-----------|
| Table des figures | i |
| Liste des tableaux | iv |
| Liste des Algorithmes | v |
| Nomenclature | vi |
| 1 Généralités et état de l'art | 2 |
| 1.1 Introduction | 2 |
| 1.2 Généralités sur le cancer du poumon | 2 |
| 1.2.1 Les Poumons | 2 |
| 1.2.2 Les nodules pulmonaires | 3 |
| 1.2.3 Le cancer | 4 |
| 1.2.4 Le cancer du poumon | 4 |
| 1.2.4.1 Cancer non à petites cellules (NSCLC) | 4 |
| 1.2.4.2 Cancer à petites cellules (SCLC) | 4 |
| 1.2.5 Quelques statistiques sur le cancer du poumon | 4 |
| 1.2.6 L'imagerie médicale | 5 |
| 1.2.7 La tomographie par ordinateur | 5 |
| 1.3 L'apprentissage en profondeur (Deep Learning) | 6 |
| 1.3.1 Définition | 6 |
| 1.4 Les réseaux de neurones artificiels | 6 |
| 1.4.1 Définition | 6 |
| 1.4.2 Les composants basiques d'un neurone artificiel | 7 |
| 1.4.3 Réseau de neurones à propagation avant (FNN) | 10 |
| 1.4.4 Réseaux de neurones récurrents (RNN) | 10 |
| 1.4.5 Réseaux de neurones convolutifs (CNN) | 11 |
| 1.4.6 La détection du cancer du poumon | 12 |

| | | |
|----------|---|-----------|
| 1.5 | Travaux connexes | 13 |
| 1.5.1 | Central focused convolutional neural networks : Developing a data-driven model for lung nodule segmentation | 13 |
| 1.5.2 | Multi-view Deep Convolutional Neural Networks for Lung Nodule Segmentation | 13 |
| 1.5.3 | Pulmonary Nodule Detection Based on ISODATA-Improved Faster RCNN and 3D-CNN with Focal Loss | 14 |
| 1.5.4 | S4ND : Single-Shot Single-Scale Lung Nodule Detection | 14 |
| 1.6 | Conclusion | 15 |
| 2 | Conception du système | 16 |
| 2.1 | Introduction | 16 |
| 2.2 | Schéma global du système | 17 |
| 2.3 | La base de données | 17 |
| 2.4 | Pré-traitement des données | 18 |
| 2.4.1 | Sélection des coupes | 18 |
| 2.4.2 | L'extraction du parenchyme pulmonaire | 19 |
| 2.4.2.1 | Échelle de Hounsfield (HU) | 19 |
| 2.4.2.2 | Binarisation des images par seuillage | 20 |
| 2.4.2.3 | Suppression des taches liées au bord du CT-scan | 21 |
| 2.4.2.4 | Conserver les deux plus grandes régions | 22 |
| 2.4.2.5 | Remplissage des petits trous et séparation des poumons | 22 |
| 2.4.2.6 | L'enveloppe convexe et dilatation | 23 |
| 2.4.3 | Normalisation des coupes | 25 |
| 2.5 | La segmentation 2D des nodules pulmonaires | 25 |
| 2.5.1 | Segmentation sémantique et Segmentation d'instance | 26 |
| 2.5.1.1 | Segmentation sémantique | 26 |
| 2.5.1.2 | Segmentation d'instance | 26 |
| 2.5.2 | Segmentation avec U-Net | 26 |
| 2.5.2.1 | Préparation des données | 26 |
| 2.5.2.2 | Architecture du modèle | 27 |
| 2.5.3 | Segmentation avec Mask R-CNN | 28 |
| 2.5.3.1 | Préparation des données | 28 |
| 2.5.3.2 | Architecture du modèle | 28 |
| 2.6 | Classification des nodules | 31 |
| 2.6.1 | La matrice de co-occurrence et les indices de Haralick | 31 |
| 2.6.2 | K-means | 33 |

| | | |
|----------|--|-----------|
| 2.7 | Conclusion | 35 |
| 3 | Résultats et implémentations | 36 |
| 3.1 | Introduction | 36 |
| 3.2 | Environnements de travail | 36 |
| 3.2.1 | Environnement matériel | 36 |
| 3.2.1.1 | Google Colaboratory | 37 |
| 3.2.1.2 | Kaggle | 37 |
| 3.2.2 | Environnement logiciel | 37 |
| 3.2.2.1 | Python | 37 |
| 3.2.2.2 | SimpleITK | 38 |
| 3.2.2.3 | NumPy | 38 |
| 3.2.2.4 | OpenCV | 38 |
| 3.2.2.5 | TensorFlow | 39 |
| 3.2.2.6 | Keras | 39 |
| 3.2.2.7 | Scikit-learn | 39 |
| 3.2.2.8 | Scikit-image | 40 |
| 3.2.2.9 | Tkinter | 40 |
| 3.2.2.10 | Visual Studio Code | 40 |
| 3.3 | Présentation de l'application | 41 |
| 3.3.1 | Écran d'accueil | 41 |
| 3.3.2 | Écran de chargement du fichier | 41 |
| 3.3.3 | Écran de visualisation du scan | 42 |
| 3.3.4 | Bouton de détection | 43 |
| 3.4 | Page des résultats finaux | 44 |
| 3.4.0.1 | Liste des nodules détectés | 46 |
| 3.5 | Tests et résultats | 49 |
| 3.5.1 | Critères d'évaluation | 49 |
| 3.5.1.1 | Accuracy | 49 |
| 3.5.1.2 | La précision | 50 |
| 3.5.1.3 | rappel | 50 |
| 3.5.1.4 | Le score de Dice | 50 |
| 3.5.1.5 | Intersection sur Union | 50 |
| 3.5.1.6 | courbe ROC | 50 |
| 3.5.2 | L'évaluation des modèles de segmentation | 51 |
| 3.5.2.1 | Courbes ROC des résultats des modèles | 52 |
| 3.5.2.2 | Affichage des résultats de la segmentation des modèles | 53 |

| | | |
|-------|---|-----------|
| 3.5.3 | L'évaluation de la classification | 54 |
| 3.6 | Conclusion | 57 |
| | Bibliographie | 59 |