Introduction Générale

1. Introduction

Le traitement et l'analyse d'images médicales est un domaine en pleine expansion où le problème de la recherche automatique d'informations contenues dans les images médicales s'avère un besoin pressant. En effet, la grande diversité des dispositifs d'imagerie médicale, la difficulté d'interprétation de ces images ainsi que leur grand nombre, engendrent un travail fastidieux pour ceux qui doivent les interpréter. Pour traiter cette grande masse d'informations, les médecins s'orientent actuellement vers l'utilisation des systèmes d'aide à l'analyse et à l'interprétation de ces images. Cette analyse vise à faciliter le diagnostic posé par le praticien et à le rendre le plus précis et le plus fiable possible.

2. Problématique

Le secteur de la santé est assez différent des autres secteurs. Les attentes des personnes sont de plus haut niveau en termes de services dans ce domaine indépendamment du coût. Une grande partie du budget est consacrée à ces services, mais les attentes des gens ne sont toujours pas satisfaites. Le domaine médical a diverses modalités qui produisent une grande quantité d'images médicales. Ces images sont lues ou interprétées par des experts humains. Mais l'interprétation par des experts humains à ses limites pour des raisons telles que la complexité de l'image, la subjectivité de l'image, les variations entre les différents interprètes et la fatigue. Il faut donc une méthode d'interprétation plus appropriée. Le *« deep learning »* a donné des solutions réussies pour de nombreuses applications du monde réel. Il fournit également une solution pour l'interprétation d'images médicales avec une grande précision. Mais, la pénétration de l'apprentissage en profondeur dans l'industrie des soins de santé est assez lente par rapport aux autres problèmes du monde réel en raison de la sensibilité du domaine.

3. Motivations

Le cancer du sein reste le cancer le plus fréquent chez la femme et l'une des premières causes de décès parmi les cancers gynécologiques. En Algérie et dans le monde la situation est tout à fait dramatique. En effet, une femme sur 10 est touchée par cette maladie durant son existence. Il se place, tous sexes confondus, au 2e rang de tous les cancers, après le cancer de la prostate. Beaucoup d'efforts ont été effectués sous forme de programmes de dépistage pour une prévention et une détection précoce. L'imagerie des mammographies est l'une des modalités la plus utilisée. Cette technique à laquelle nous nous sommes intéressés dans ce travail, est la technique de référence incontournable pour l'exploration du sein, la plus performante en matière de surveillance et de dépistage précoce du cancer du sein. Cependant, chaque année, des millions de mammographies sont produites dans le monde pour un dépistage précoce d'un cancer du sein, ou pour établir un diagnostic afin de contrôler une action thérapeutique qui peuvent de manière significative augmenter la possibilité de la survie pour des patients. Ces images de mammographies fournissent des informations radiologiques

sur la glande mammaire qui sont, malheureusement, extrêmement difficiles à exploiter par l'expert humain qui s'appuie sur sa vision et son expérience. L'élaboration de logiciels dédiés à l'analyse de ces mammographies, doit permettre d'optimiser leur exploitation, pour le plus grand bénéfice du patient et du médecin.

4. Contribution

Au contraire des articles sur « Breast Cancer Detection » qui se focalisent sur la tâche de la classification des masses tous en négligeant les calcifications pour élever l'accuracy des résultats ici on va leurs prendres en considération. Nos contributions sont les suivantes :

La stratégie composite dont la tâche en 4 classes peut être décomposée en deux sous-tâches : la classification (mass-calcification) et le diagnostic (benign-malignant). Le 1er modèle exploite le Transfer Learning sur l'architecture VGG16 CNN (pré-entraîné sur ImageNet le réseau qui a remporté l'ILSVRC ImageNet Large Scale Visual Recognition Challenge en 2014.) pour la classification des anomalies (mass-calcification) sur des images de mammographie prises à partir du jeu de données CBIS DDSM) a obtenu une précision de **91,37%** sur l'ensemble de test. Puis le 2ème modèle DCNN pour la tâche de classification à 2 classes (Benign/Malignant) a obtenu une précision de **69,05%** sur l'ensemble de test.

L'idée est de former deux modèles indépendamment, chacun se concentrant sur une seule sous-tâche ; les deux réseaux fonctionneront alors en parallèle, chacun produisant la moitié de la prédiction complète à 4 classes. La logique derrière cette approche est que chaque réseau se spécialise pour résoudre un sous-problème plus petit et potentiellement plus facile, obtenant ainsi de meilleurs résultats que de résoudre le problème dans son ensemble d'un seul coup.

Donc l'adoption de la stratégie composite de 2 modèles pour la tâche à 4 classes se traduit par une précision de **62,5%** sur l'ensemble de test. En comparant le résultat par exemple le travail 2019 de Nhung Le, Nithish Addepalli et Ravi Choudhary nommé « *Breast Cancer Detection Through High Resolution Screening Mammograms* » de Center For Data Science, New York University, a obtenu une meilleure précision de 60% avec 4 classes.

La plus part des travaux de détection de cancer de sein font la classification des images et parce que CNN identifie uniquement la classe de l'objet, pas l'emplacement de l'objet dans une image. On a opté pour mask R-CNN dans la dernière expérience.

L'utilisation brûte de Mask R-CNN pour la segmentation et la détection des masses est nouvelle et en tant qu'évaluation des performances, la précision est de **65,23%** sur l'ensemble de test avec une sensitivité (TPR) de **80,51%**. Dans l'évaluation de la segmentation des régions du cancer du sein, l'indice Dice est de **68.10%.** Ces résultats indiquent que notre schéma proposé peut être utile pour la détection du cancer du sein.

5. L'organisation du manuscrit

Ce manuscrit est composé d'une introduction générale et trois chapitres. Le premier chapitre introduit les concepts relatifs aux contexte médical du cancer de sein, l'anatomie du sein, les types des tumeurs et leurs symptômes en passant par le dépistage du cancer du sein ainsi que quelques bases de mammographie connues.

Dans le deuxième chapitre, nous définissons les grandes lignes des techniques utilisées pour le domaine d'apprentissage profond « *Deep Learning* » et « *Breast Cancer Detection* » ainsi que la présentation des tableaux d'état de l'art des techniques utilisées pour le diagnostique du cancer de sein.

Le troisième chapitre est consacré à notre contribution par l'étude de différentes techniques de classification et enfin une trop nouvelle technique qui englobe la segmentation, classification et la localisation des masses. Donc nous présentons les ressources et outils software ainsi que hardware utilisées pour l'implémentation et la validation de nos approches. Nous définissions les architectures DCNN utilisées dans chacune des expériences et la discussion des résultats obtenus avec les analyses et les synthèses appropriées.

Enfin nous clôturons ce manuscrit par une conclusion sur les travaux réalisés ainsi que nos perspectives futures.