

Détection du cancer du sein à l'aide de l'apprentissage par transfert en Réseaux de neurones convolutifs

SanaUllah Khan¹, Naveed Islam¹ and Joel J. P. C Rodrigues¹

¹ (titre de l'article:) A novel deep learning based framework for the detection and classification of breast cancer using transfer learning

¹ Auteur du Rapport: Fabienne Janvier

¹ INSTITUT FRANCOPHONE INTERNATIONNAL (IFI)

¹ SUPERVISEUR : Dr. HO Tuong Vinh

Reception date of the manuscript: 28/Mars/2019

Acceptance date of the manuscript: 28/mMars/2019

Publication date: 28/Mars/2019

Abstract— Le cancer du sein est l'un des principaux types de cancer qui cause la mortalité chez les femmes dans le monde. La détection précise du cancer du sein à l'aide d'algorithmes automatisés reste un problème dans la littérature. Bien qu'une pléthore de travaux ait tenté de résoudre ce problème, une solution exacte reste à trouver. Dans cet article, pour résoudre ce problème, ils ont proposé un nouveau cadre basé sur la notion d'apprentissage approfondi pour la détection et la classification du cancer du sein sur les images de cytologie mammaire en utilisant le concept d'apprentissage par transfert. Contrairement aux paradigmes d'apprentissage classiques, qui se développent et se produisent de manière isolée, l'apprentissage par transfert vise à utiliser les connaissances acquises lors de la résolution d'un problème dans un autre problème connexe. À l'aide de l'architectures CNN préformées ils ont extraire les caractéristiques à savoir googleNet, Visual Geometry Group Network (VGGNet) et Residual Networks (ResNet), qui sont introduites dans une couche entièrement connectée pour la classification des cellules malignes et bénignes.

Keywords— Apprentissage approfondi, Smart pattern recognition, Tranfert learning, Cancer du Sein

I. INTRODUCTION

Le cancer du sein est le type de cancer invasif le plus fréquent chez les femmes du monde entier. Aux États-Unis, on estime à 0,25 million le nombre de nouveaux cas de cancer du sein et à 0,04 million le nombre de décès en 2016. Dans les pays développés, environ une femme sur huit développera cette maladie au cours de sa vie. On a diagnostiqué deux méthodes couramment utilisées pour faire la détection du cancer du sein : la mammographie et la biopsie.

Dans la mammographie, on dit que les images spécifiques du sein sont utilisées pour détecter les symptômes précoces du cancer chez les femmes par radiographie, en utilisant la mammographie et le taux de mortalité a diminué. Pour la biopsie c'est une autre méthode de diagnostic efficace et précise pour la détection du cancer du sein. Dans cette approche, un échantillon de tissu provenant d'une région affectée du sein est analysé au microscope par un pathologiste pour la détection et la classification de la tumeur. Actuellement, la biopsie joue un rôle essentiel dans le cancer du sein ainsi que dans d'autres types de diagnostic du cancer. Grâce à la biopsie, le pathologiste peut déterminer deux types de lésions :

bénignes et malignes. La lésion bénigne n'est pas cancéreuse ; il s'agit en effet d'une anomalie des cellules épithéliales, et la plupart de ces anomalies ne peuvent pas devenir une source de cancer du sein. Les cellules malignes ou cancéreuses sont ces types de cellules, qui commencent à se diviser anormalement et se développent de façon irrégulière. L'analyse manuelle des images microscopiques est une tâche très complexe et difficile en raison de l'apparence irrégulière des cellules bénignes et malignes.

Au cours des dernières années, Bien que l'application CNN fonctionne très bien sur de grands ensembles de données mais Afin d'obtenir une plus grande précision de reconnaissance et de réduire les coûts de calcul, le concept d'apprentissage par transfert peut être exploité pour améliorer les performances des différentes applications CNN en combinant leurs connaissances et aussi pour pallier les lacunes des systèmes existants de détection et de classification des images cytologiques.

II. CONTEXTE

Le concept d'apprentissage contextuel donne une nouvelle orientation à l'apprentissage par transfert dans la détection et la classification du cancer du sein. La combinaison d'une architecture CNN multiple augmente les performances de l'apprentissage par transfert et peut remplacer l'utilisation

d'une architecture CNN traditionnelle à modèle unique. De même, la combinaison de ResNet50, InceptionV2 et InceptionV3 est pré-formée sur ImageNet, ce qui a permis d'obtenir un modèle rapide et précis pour la classification d'images à base de cellules.

III. OBJECTIFS

Dans le cadre propose l'objectif de cet article est de:

- Fournir un cadre basé sur une architecture d'apprentissage approfondie pour la détection et la classification des cellules dans les images de la cytologie mammaire afin de donner une bonne direction vers des produits de qualité pour une bonne diagnostique.
- Analyser le concept d'apprentissage par transfert sur trois différentes architectures d'apprentissage approfondi.
- Fournir une analyse comparative de chaque architecture d'apprentissage approfondi en ce qui concerne la précision dans le contexte de l'apprentissage par transfert.
- disposer un système automatique pour obtenir des résultats efficaces et très précis, afin d'assurer l'uniformité des résultats bonne orientation vers des produits qualitatifs pour les diagnostics.

IV. PROBLEMATIQUE

Dans l'article ils doivent résoudre le problème de sur ajustement des données. Bien que l'application CNN fonctionne très bien sur les grands ensembles de données, elle ne permet pas de réaliser des gains significatifs sur les petits ensembles de données Si la taille de l'ensemble de données cible est plus petite et similaire à l'ensemble des données d'entraînement alors le risque de sur-ajustement est élevé, si la taille de l'ensemble de données cible est plus grande et similaire à l'ensemble des données d'entraînement le risque de sur-ajustement est faible et il nécessite que le réglage fin du modèle pré-entraîne.

V. APPRENTISSAGE PAR TRANSFERT

L'apprentissage par transfert (transfer learning, en anglais) est l'une des méthodes les plus connues de l'apprentissage automatique qui permet d'apprendre les connaissances de base appliquées pour résoudre un problème et de les réutiliser pour d'autre problème pertinent. L'apprentissage par transfert est ce que nous faisons tous les jours, c'est profiter d'un apprentissage acquis précédemment, pour, par analogie, résoudre un problème similaire mais différent. Par exemple, il est plus facile d'apprendre à piloter une moto si on sait déjà faire du vélo.

VI. ÉTUDE DES TRAVAUX CONNEXE

En tant que branche de la technique d'apprentissage machine, l'apprentissage approfondi a récemment révolutionné les domaines de la représentation d'images et de la reconnaissance d'objets. Cependant, il nécessite des échantillons de formation à grande échelle, comme dans ImageNet, pour le réglage et l'optimisation des hyper paramètres. L'apprentissage par transfert est l'amélioration de l'apprentissage dans une nouvelle tâche par le transfert de connaissances que nous trouvons à travers ses travaux con-

nexes :

- La classification des lésions de masse mammaire dans les mammographies par apprentissage par transfert. publier en janvier 2017 à Hong Kong. Fan Jiang et Hui Lui on exploré la technique d'apprentissage par transfert avec GoogleNet pour résoudre le problème des lésions de masse mammaire en mammographie et il a obtenu de meilleur performance pour un taux d'AUC de 0.88
- Transfer histopathologie base sur l'apprentissage de classification des images pour la détection du cancer du sein. Publier en 2018. Deniz et Coll on utiliser une méthode comme l'apprentissage par transfert pour adapter un modèle de CNN préformé au problème pose. Les modèles AlexNet et Vgg16 sont pris en compte dans le travail présenté pour l'extraction des caractéristiques des images et AlexNet est utilisé pour un réglage plus fin.
- Détection du cancer du sein à l'aide de l'apprentissage par transfert dans les réseaux neuronaux convolutif. Publier en 2017. Shuyue Guan et Murray Loew on teste trois(3) méthodes et Utiliser l'apprentissage par transfert dans CNN est une solution prometteuse pour la détection du cancer du sein. Nos résultats montrent que le modèle CNN préformé (VGG-16) peut extraire automatiquement caractéristiques des images mammographies et un bon classificateur NN peuvent être formés par ces fonctionnalités sans fournir ses caractéristiques.

Nous avons passé en revue plusieurs études récentes étroitement liées à la nôtre. Ces études ont appliqué l'apprentissage par transfert de CNN à détecter un cancer du sein / une anomalie sur la base d'une mammographie. En comparaison avec ces études, nous avons utilisé beaucoup plusieurs images mammographies pour la formation et le test du CNN classificateurs et un modèle pré-entraîné distinct.

VII. SOLUTION PROPOSÉE

Dans cet article La solution propose est base sur l'architecture CNN, est expliqué pour la détection et la classification des cellules malignes et bénignes dans les images de cytologie mammaire, Afin d'obtenir une plus grande précision de reconnaissance et de réduire les coûts de calcul, le concept d'apprentissage par transfert peut être exploité pour améliorer les performances des différentes architectures CNN en combinant leurs connaissances GoogLeNet, VGGNet et ResNet pour extraire séparément les différentes caractéristiques de bas niveau. Les étapes de leurs architectures proposées :

- Prétraitements des données et traitement d'augmentation.
- Architecture CNN pré-formée pour l'extraction de fonctionnalités.

VIII. IMPLEMENTATION ET EXPERIMENTATION

Pour évaluer la performance du cadre présenté, ils ont utilisé deux bases de données d'image microscopique du sein. Données de référence de Standard et l'autre est développé localement à l'hôpital LRH de Peshawar, au Pakistan. Ils ont appliqué la technique d'augmentation par la mise à l'échelle, rotation, translation et la modélisation des couleurs pour produire un total de 8000 images. Pour former l'architecture ils

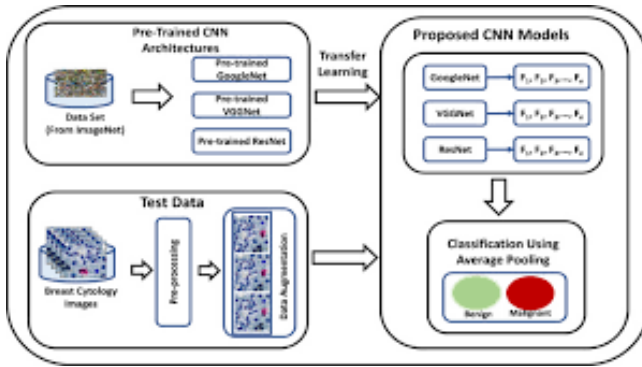


Fig. 1: Schéma du cadre d'apprentissage profond proposé.

ont utilisées 6000 images et 2000 images pour tester le modèle forme. La taille des images étaient 100x, 140x, 200x, 500x sont utilisées pour l'évaluation du taux d'accuracy pendant l'exécution, 75% de l'ensemble des données sont utilisées à des fins de formation, 25% des données sont utilisées pour tester la précision de l'architecture proposée.

Lors de l'expérimentation de l'approche proposée la répartition est effectuée selon trois procédures différentes, les données sont séparées par des ensembles de données de formation et d'essai : 90%-10%, 80%-20%, 70%-30%. 80%-20% signifie que 80% des données sont utilisées pour la formation, tandis que les 10% restants sont utilisés pour tester les archives de CNN.

IX. RESULTAT ET EVALUATION

À partir d'un seul CNN les résultats obtenus sont comparés avec les caractéristiques combinées et des différentes techniques d'extraction, individuellement chaque architecture donne un résultat moyen de classification 93,5%, 94,15%, 94,35% et le cadre proposé donne une précision de 97,525% donc ces résultats montrent que le cadre proposé est plus performant en termes de précision de détection et de classification du cancer du sein.

Classifier	Training-testing data splitting	Class type	Precision	Recall	F1 score	Accuracy	Average Accuracy
GoogleNet	90%-10%	B	0.93	0.94	0.94	93.67%	93.22%
		M	0.96	0.94	0.95		
	80%-20%	B	0.93	0.93	0.93	93.00%	
		M	0.93	0.94	0.93		
	70%-30%	B	0.96	0.9	0.93	93.00%	
VGGNet	90%-10%	M	0.92	0.98	0.95	93.67%	94.00%
		B	0.9	0.97	0.94		
	80%-20%	M	0.9	0.91	0.92	96.00%	
		B	0.97	0.96	0.95		
	70%-30%	M	0.95	0.93	0.92	92.33%	
ResNet	90%-10%	B	0.91	0.92	0.94	98.00%	94.89%
		M	0.9	0.99	0.96		
	80%-20%	B	0.97	0.98	0.99	96.00%	
		M	0.99	0.98	0.99		
	70%-30%	B	0.9	0.98	0.99	90.67%	
Proposed Framework	90%-10%	M	0.9	0.92	0.9	97.00%	97.67%
		B	0.91	0.98	0.99		
	80%-20%	B	0.96	0.97	0.98	97.67%	
		M	0.95	0.96	0.98		
	70%-30%	B	0.97	0.99	0.97	98.33%	

TABLE 1: ANALYSE COMPARATIVE BASÉE SUR LE SPLITTING DU CADRE PROPOSÉ AVEC D'AUTRES ARCHITECTURES CNN.

X. CONCLUSIONS

The conclusions should present a review of the key points of the article with special emphasis on the analysis and discussion of the results that were made in the previous sections and in the applications or extensions of them. You should not

reproduce the summary in this section or repeat paragraphs already included in another sections of the work.

XI. PERSPECTIVES

Pour nos études futures, nous pourrions essayer de reconnaître les zones anormales dans des images mammographiques. En utilisant le CNN, nous pourrions reconnaître les anomalies sur les images mammographiques et dessiner des limites sur des zones automatiquement. Nous pourrions utiliser d'autres modèles préformés, et comparer leurs performances comme InceptionResNetV2 (572 couches), InceptionV3 (159 couches), ResNet50 (168 couches). Il sera intéressant de voir les performances de la détection du cancer du sein en utilisant des CNN très profonds.

[1]. [1] [2] [3] [4]

REFERENCES

- [1] C. P. L. H. F.A. Spanhol, L.S. Oliveira, "A dataset for breast cancer histopathological image classification." IEEE Trans. Biomed. Eng. 63 (7), 2016, pp. 1455–1462.
- [2] L. Heutte, "A dataset for breast cancer histopathological image classification." IEEE Trans. Biomed. Eng. 63 (7), 2016, pp. 1455–1462.
- [3] A. D. S. E. A. M. S. Vesal, N. Ravikumar, "Classification of breast cancer histology images using transfer learning," *International Conference Image Analysis and Recognition*, vol. 18, no. 1, p. 812–819, 2018.
- [4] C. P. L. H. F.A. Spanhol, L.S. Oliveira, *A dataset for breast cancer histopathological image classification.* : IEEE Trans. Biomed. Eng. 63 (7), 2016.