



جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Madame Sara EL-ATEIF

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le Samedi 20 Juillet 2024 à 11h00 au Grand Amphi à l'ENSIAS

Intitulé de la thèse

**Multimodal Deep Learning in Medical Imaging: Applications
to Eye Disease and COVID-19 Diagnosis**

Président :

Pr. Mohamed Essaaidi, PES, École Marocaine des Sciences de l'Ingénieur

Directeur de thèse :

Pr. Ali Idri, PES, ENSIAS, Université Mohammed V de Rabat

Rapporteurs :

Pr. Sanaa El Fkihi, PES, ENSIAS, Université Mohammed V de Rabat

Pr. Atman Jbari, PES, ENSAM, Université Mohammed V de Rabat

Pr. Younes Jabrane, PES, ENSA, Université Cadi Ayyad, Marrakech

Examineurs :

Pr. Mohammed Radouane, PH, ENSIAS, Université Mohammed V de Rabat

Pr. Ibtissam Abnane, PH, ENSIAS, Université Mohammed V de Rabat

Résumé: Les maladies oculaires et la COVID-19 représentent des défis sanitaires mondiaux majeurs, affectant des millions de personnes à travers le monde.

Les maladies oculaires, telles que la rétinopathie diabétique (RD), la dégénérescence maculaire liée à l'âge et le glaucome, sont des causes principales de déficience visuelle et de cécité. Au Maroc seulement, plus de 2 millions de citoyens sont diabétiques, ce qui les expose à un risque de développer une RD [1]. La détection et le traitement précoces sont essentiels pour prévenir la perte de vision, mais les méthodes diagnostiques traditionnelles peuvent être inaccessibles et inexactes, en particulier dans les milieux à ressources limitées.

La COVID-19, une maladie respiratoire causée par le virus SRAS-CoV-2, a provoqué une morbidité et une mortalité importantes à travers le monde. L'étalon-or du diagnostic, les tests RT-PCR, peuvent être coûteux, prendre du temps et avoir une sensibilité limitée, en particulier aux stades précoces de la maladie [2]. Les radiographies pulmonaires et les tomodensitométries sont également utilisées pour le diagnostic, mais leur sensibilité et leur accessibilité peuvent être limitées, et elles peuvent ne pas être en mesure de distinguer la pneumonie COVID-19 des autres types de pneumonie [3]. La propagation rapide de la COVID-19 a submergé les systèmes de santé à l'échelle mondiale, soulignant le besoin d'outils de diagnostic rapides et fiables.

L'apprentissage profond (DL) est devenu un outil puissant pour l'analyse d'images médicales, offrant le potentiel d'un diagnostic automatisé et objectif des maladies oculaires et de la COVID-19. Cependant, la plupart des recherches existantes se concentrent sur des approches mono-modalité, qui n'analysent qu'un seul type d'image médicale à la fois. Cela peut limiter la précision du diagnostic, car différentes modalités peuvent fournir des informations complémentaires cruciales pour un diagnostic précis. L'apprentissage multimodal, qui combine des informations provenant de plusieurs modalités, a le potentiel d'améliorer considérablement la précision et la robustesse du diagnostic [4, 5]. Alors que certaines recherches ont exploré l'apprentissage multimodal pour le diagnostic des maladies oculaires et de la COVID-19, la plupart des études se sont concentrées sur le développement de nouvelles méthodes DL et n'explorent généralement qu'une seule approche de fusion (précoce, conjointe ou tardive) [6–10]. De plus, il y a un manque de recherche comparant les performances des CNN traditionnels avec les nouveaux modèles de transformateurs de vision dans le contexte de l'apprentissage multimodal pour ces maladies.

Cette thèse aborde ces lacunes en : (1) réalisant une comparaison complète de différents modèles DL, architectures et approches d'apprentissage de modalité (mono-modalité et fusion multimodale précoce, conjointe et tardive) sur divers ensembles de données pour le diagnostic des maladies oculaires et de la COVID-19. (2) Évaluant les performances de ces modèles en utilisant des mesures standard et des tests statistiques pour identifier les modèles

les plus performants avec des différences statistiquement significatives. (3) Examinant l'explicabilité des modèles développés pour accroître la confiance et faciliter leur adoption clinique.

Ces travaux contribuent au domaine de l'analyse d'images médicales en démontrant le potentiel de l'apprentissage profond multimodal pour le diagnostic des maladies oculaires et de la COVID-19. Les modèles et les connaissances développés peuvent contribuer à l'amélioration de la pratique clinique et ouvrir la voie à de nouvelles recherches dans ce domaine.

Mots-clés: Apprentissage profond, Apprentissage multimodale, Maladies des Yeux, COVID-19, diagnostique, Analyse image médicale.

Abstract: Eye diseases and COVID-19 pose significant global health challenges, affecting millions worldwide. Eye diseases, such as diabetic retinopathy (DR), age-related macular degeneration, and glaucoma, are leading causes of vision impairment and blindness. In Morocco alone, over 2 million citizens are diabetic, putting them at risk of developing DR [1]. Early detection and treatment are crucial for preventing vision loss, but traditional diagnostic methods can be inaccessible and inaccurate, particularly in resource-constrained settings.

COVID-19, a respiratory disease caused by the SARS-CoV-2 virus, has caused significant morbidity and mortality worldwide. The gold standard for diagnosis, RT-PCR tests, can be expensive, time-consuming, and have limited sensitivity, particularly in the early stages of the disease [2]. Chest X-rays and Computed Tomography scans are also used for diagnosis, but their sensitivity and accessibility can be limited, and they may not distinguish COVID-19 pneumonia from other types of pneumonia [3]. The rapid spread of COVID-19 has overwhelmed healthcare systems globally, highlighting the need for fast and reliable diagnostic tools.

Deep learning (DL) has emerged as a powerful tool for medical image analysis, offering the potential for automated and objective diagnosis of both eye diseases and COVID-19. However, most existing research focuses on mono-modality approaches, which analyze only one type of medical image at a time. This can limit diagnostic accuracy, as different modalities can provide complementary information that is crucial for accurate diagnosis. Multimodal learning, which combines information from multiple modalities, has the potential to significantly improve diagnostic accuracy and robustness [4, 5]. While some research has explored multimodal learning for eye diseases and COVID-19 diagnosis, most studies have focused on developing new DL methods and typically explore only one fusion approach (early, joint, or late) [6–10].

Additionally, there is a lack of research comparing the performance of traditional CNNs with the newly developed vision transformer models in the context of multimodal learning for these diseases. This thesis addresses these



جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

gaps by (1) conducting a comprehensive comparison of different DL models, architectures, and modality learning approaches (mono-modality and multimodal early, joint, and late fusion) on various datasets for both eye disease and COVID-19 diagnosis.(2) Evaluating the performance of these models using standard metrics and statistical tests to identify the best-performing models with statistically significant differences. (3) Examining the explainability of the developed models to increase trust and facilitate their clinical adoption.

This research contributes to the field of medical image analysis by demonstrating the potential of deep multimodal learning for eye disease and COVID-19 diagnosis. The developed models and insights can contribute to improved clinical practice and pave the way for further research in this area.

Keywords: Deep learning, multimodal learning, eye diseases, COVID-19, diagnosis, medical image analysis.

