

الجزائرية الديمقراطية الشعبية الجمهورية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة العليا للإعلام الآلي - 08 ماي 1945 - بسبدي بلعباس

Ecole Supérieure en Informatique

Mai 1945- Sidi Bel Abbès 08-



MEMOIRE

En Vue de l'obtention du diplôme de **Master**

Filière : **Informatique**

Spécialité : Intelligence artificielle et sciences de données (IASD)

Thème

**Computer Aided Diagnosis System for Breast Cancer Segmentation
using Weakly Supervised Learning and Federated Learning**

Présenté par : Seddik Dounia

Soutenu le : **28/09/2025**

Devant le jury composé de :

- | | |
|------------------------------|------------------------------|
| - Mme Imene SAIDI | Président |
| - Mme.Nassima DIF | Encadreur |
| - M.Sidi Mohammed BENSLIMANE | Co-Encadreur |
| - M.Mohammed BEKKOUCHE | Examinateur |
| - M.Mohammed KECHAR | Représentant de l'incubateur |

Année Universitaire : 2024 / 2025

Summary

Breast cancer remains the most prevalent malignancy among women globally, with early detection through mammographic screening being critical for improving survival outcomes. The accuracy of radiological interpretation and effectiveness of computer-aided diagnosis systems are paramount, requiring precise mass localization to enhance patient prognosis. However, developing robust CAD systems faces significant challenges including the need for extensive pixel-level annotations, inter-institutional data heterogeneity, and privacy constraints that limit collaborative learning across healthcare networks.

This dissertation focuses on breast cancer segmentation addressing the fundamental challenge of reducing annotation burden by implementing weakly supervised learning. While implementing deep learning models additional hurdles arise, such as dealing with insufficient data and a limited generalization ability. These challenges can be addressed in this work through the integration of federated learning protocols by enabling collaborative model improvement across multiple healthcare institutions. Federated learning emerges as a valuable tool to preserve patient privacy while improving model generalization.

Keywords: Breast Cancer, Weakly-Supervised Learning, Federated Learning, Medical Image Segmentation, Mammography, Privacy-Preserving Machine Learning, Computer-Aided Diagnosis (CAD).

Résumé

Le cancer du sein demeure la malignité la plus répandue chez les femmes à l'échelle mondiale, et le dépistage précoce par mammographie est essentiel pour améliorer les chances de survie. L'exactitude de l'interprétation radiologique et l'efficacité des systèmes d'aide au diagnostic (CAD) sont primordiales, nécessitant une localisation précise des masses afin d'optimiser le pronostic des patientes. Toutefois, le développement de systèmes CAD robustes fait face à des défis majeurs, notamment la nécessité d'annotations exhaustives au niveau des pixels, l'hétérogénéité des données entre institutions et les contraintes de confidentialité qui limitent l'apprentissage collaboratif au sein des réseaux de santé.

Cette thèse se concentre sur la segmentation du cancer du sein en abordant le défi fondamental de la réduction de la charge d'annotation grâce à l'apprentissage faiblement supervisé. Lors de la mise en œuvre de modèles d'apprentissage profond, d'autres obstacles apparaissent, tels que l'insuffisance des données et une capacité limitée de généralisation. Ces défis peuvent être surmontés dans ce travail grâce à l'intégration de protocoles d'apprentissage fédéré, permettant l'amélioration collaborative des modèles entre plusieurs institutions de santé. L'apprentissage fédéré s'impose ainsi comme un outil précieux pour préserver la confidentialité des patients tout en améliorant la capacité de généralisation des modèles.

Mots-clés: Cancer du sein, Apprentissage faiblement supervisé, Apprentissage fédéré, Segmentation d'images médicales, Mammographie, Apprentissage automatique respectueux de la confidentialité, Systèmes d'aide au diagnostic (CAD).

الملخص

يُعدّ سرطان الثدي أكثر الأورام الخبيثة شيوعاً بين النساء على مستوى العالم، ويُعتبر الكشف المبكر عبر الفحص الشعاعي للثدي أمراً بالغ الأهمية لتحسين معدلات البقاء على قيد الحياة. إن دقة التفسير الإشعاعي وفعالية أنظمة التشخيص بمساعدة الحاسوب (CAD) أمران أساسيان، إذ يتطلبان تحديداً دقيقاً للكامل لتعزيز تشخيص المريضة وتوقع مسار المرض. ومع ذلك، يواجه تطوير أنظمة CAD قوية تحديات كبيرة، من بينها الحاجة إلى كم هائل من التعليقات التوضيحية على مستوى البكسل، والتباين بين بيانات المؤسسات المختلفة، بالإضافة إلى قيود الخصوصية التي تحدّ من إمكانيات التعلم التعاوني عبر شبكات الرعاية الصحية.

تركز هذه الأطروحة على تجزئة سرطان الثدي من خلال معالجة التحدي الجوهرى المتمثل في تقليل عبء التعليقات التوضيحية باستخدام التعلم تحت الإشراف الضعيف. وأثناء تطوير نماذج التعلم العميق، تبرز عقبات إضافية مثل قلة البيانات وصعوبة تعميم النماذج. يمكن مواجهة هذه التحديات في هذا العمل من خلال دمج بروتوكولات التعلم الفيدرالي التي تتيح تحسين النماذج بشكل تعاوني عبر مؤسسات صحية متعددة. ويُعدّ التعلم الفيدرالي أداة قيمة للحفاظ على خصوصية المرضى مع تحسين قدرة النماذج على التعميم.

الكلمات المفتاحية: سرطان الثدي، التعلم تحت الإشراف الضعيف، التعلم الفيدرالي، تجزئة الصور الطبية، التصوير الشعاعي للثدي، التعلم الآلي المحافظ على الخصوصية، التشخيص بمساعدة الحاسوب (CAD).