

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 d'ingénieur d'état

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 a leading cause of mortality among women, and precise lesion segmentation is crucial for early diagnosis and treatment planning.

This work introduces a novel CAD system that integrates weakly supervised learning relying solely on bounding-box annotations and federated learning to enable privacy preserving, collaborative model training across multiple institutions.

The primary goal is to develop a robust CAD system that processes mammography images, produces accurate breast lesion segmentation using weakly supervised learning, and enables privacy-preserving, multi-site model training through federated learning.

To achieve this, we trained a state-of-the-art dual-framework network, which comprises a primary network and an ancillary network that requires only bounding-box annotations rather than full pixel-level masks. The pretrained weights are then shared across clients via FedAvg. The model was trained on the CBIS-DDSM mammography dataset. Under decentralized training, the model achieves a validation Dice score of 0.68 and an IoU of 0.61.

The integration of federated learning with weak supervision offers a unique combination of efficiency and privacy. By requiring only bounding-box annotations, it drastically reduces radiologist workload compared to traditional pixel-level segmentation methods, while its federated learning framework enables multiple institutions to collaboratively improve model accuracy without sharing sensitive patient data.

Keywords: Breast Cancer, Lesion Segmentation Weakly-Supervised Learning, Federated Learning, CBIS-DDSM Dataset, Mammography, Privacy-Preserving, Computer-Aided Diagnosis (CAD) .

Résumé

Le cancer du sein demeure l'une des principales causes de mortalité chez les femmes, et une segmentation précise des lésions est essentielle pour un diagnostic précoce et une planification thérapeutique efficace.

Ce travail propose un nouveau système d'aide au diagnostic (CAD) intégrant l'apprentissage faiblement supervisé reposant uniquement sur des annotations sous forme de boîtes englobantes et l'apprentissage fédéré, afin de permettre un entraînement collaboratif des modèles respectueux de la confidentialité entre plusieurs institutions.

L'objectif principal est de développer un système CAD robuste capable de traiter les mammographies, de produire une segmentation précise des lésions mammaires grâce à l'apprentissage faiblement supervisé, et de permettre un entraînement multi-site préservant la vie privée via l'apprentissage fédéré.

Pour ce faire, nous avons entraîné un réseau dual de pointe, composé d'un réseau principal et d'un réseau auxiliaire, qui ne requièrent que des annotations par boîtes englobantes plutôt que des masques pixel-niveau complets. Leurs poids pré-entraînés sont ensuite partagés entre les clients grâce à l'algorithme FedAvg. Le modèle a été entraîné sur le jeu de données de mammographies CBIS-DDSM. En condition d'entraînement décentralisé, il atteint un score de Dice de 0,68 et un IoU de 0,61 en validation.

L'intégration de l'apprentissage fédéré avec une supervision faible offre une combinaison unique d'efficacité et de respect de la confidentialité. En ne nécessitant que des annotations sous forme de boîtes englobantes, elle réduit considérablement la charge de travail des radiologues par rapport aux méthodes classiques de segmentation pixel-niveau, tandis que son cadre fédéré permet à plusieurs institutions d'améliorer collaborativement la précision du modèle sans partager de données sensibles des patients.

Mots-clés: Cancer du sein, segmentation de lésions, apprentissage faiblement supervisé, apprentissage fédéré, CBIS-DDSM, mammographie, préservation de la confidentialité, système d'aide au diagnostic (CAD).

الملخص

يُعد سرطان الثدي من أبرز أسباب الوفيات بين النساء، وتُعتبر عملية تقسيم (Segmentation) الآفات بدقة أمرًا بالغ الأهمية للتشخيص المبكر والتخطيط العلاجي.

يقدم هذا العمل نظامًا جديدًا للتشخيص بمساعدة الحاسوب (CAD) يدمج بين التعلم تحت الإشراف الضعيف — المعتمد فقط على التعليقات التوضيحية بصناديق الإحاطة (Bounding Boxes) — والتعلم الفيدرالي، مما يتيح تدريبًا تعاونيًا للنماذج عبر مؤسسات متعددة مع الحفاظ على خصوصية البيانات.

الهدف الأساسي هو تطوير نظام CAD قوي قادر على معالجة صور الماموغرافيا، وإنتاج تقسيم دقيق لآفات الثدي باستخدام التعلم تحت الإشراف الضعيف، وتمكين التدريب اللامركزي متعدد المواقع مع الحفاظ على خصوصية البيانات عبر التعلم الفيدرالي.

ولتحقيق ذلك، قمنا بتدريب شبكة حديثة ثنائية البنية تتكون من شبكة رئيسية وأخرى مساعدة، لا تحتاج إلا إلى صناديق الإحاطة بدلاً من الأقنعة الكاملة على مستوى البكسل. يتم بعد ذلك مشاركة الأوزان المدربة مسبقًا بين العملاء باستخدام خوارزمية FedAvg. وقد تم تدريب النموذج على مجموعة بيانات الماموغرافيا CBIS-DDSM. تحت بيئة التدريب اللامركزي، حقق النموذج نتيجة $Dice = 0.68$ و $IoU = 0.61$ على بيانات التحقق.

إن دمج التعلم الفيدرالي مع الإشراف الضعيف يقدم مزيًا فريدًا من الكفاءة والحفاظ على الخصوصية؛ إذ إن الاكتفاء بالتعليقات بصناديق الإحاطة يقلل بشكل كبير من عبء عمل أخصائي الأشعة مقارنة بطرق التقسيم التقليدية على مستوى البكسل، بينما يتيح الإطار الفيدرالي لمؤسسات متعددة تحسين دقة النموذج بشكل تعاوني دون مشاركة بيانات المرضى الحساسة.

الكلمات المفتاحية: سرطان الثدي، تقسيم الآفات، التعلم تحت إشراف ضعيف، التعلم الفيدرالي، مجموعة بيانات CBIS-DDSM، الماموغرافيا، الحفاظ على الخصوصية، التشخيص بمساعدة الحاسوب (CAD).