République Algérienne Démocratique et Populaire
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique
École Supérieure en Informatique - 08 Mai 1945 - Sidi Bel Abbès

Mémoire de Fin d'étude
Présentée pour l'obtention du **diplôme** de **Master**
En : Informatique
Spécialité : Système d'Information et Web (SIW)

Thème

Breast density prediction using deep learning methods

Présenté par :

- Boudinar Mohamed El Amine
- Abdelali Mohamed Amine

Soutenue publiquement le 04/07/2023, devant le jury composé de :

Mr	Khaldi Miloud	Maître de Conférences B	Président
Mme	Dif Nassima	Maître de Conférences B	Encadrante
Mme	Bousmaha Rabab	Maître Assistant B	Examinateur1
Mr	Kechar Mohamed	Maître de Conférences A	Examinateur2

Abstract

Predictive medicine is instrumental in assisting doctors to evaluate patients and help them prevent the risk of breast cancer disease where they can make accurate predictions based on breast examination and analysis, allowing for early detection and timely intervention. This depends on multiple factors such as age, lifestyle, family history, and breast density which is the most important factor. This thesis showcases the current state of research on the impact of applying artificial intelligence techniques and deep learning in the detection of breast density. The initial part of our discussion focuses on breast cancer, particularly in the assessment of breast density and the challenges radiologists face in this field. We then delve into the field of deep learning, highlighting the latest studies in this rapidly evolving field, with a specific emphasis on its impact on the medical sector. Although deep learning's potential to improve medical diagnostics is promising, its reliance on precise and accurate data is vital, as it has a direct impact on patient outcomes. While some countries have already begun to integrate deep learning techniques into their medical practices, there is still much progress to be made. However, early results have been promising, and the potential for continued advancements in this field is optimistic. The second part of this dissertation is dedicated to the latest research in analyzing complex mammogram images using cutting-edge deep-learning techniques, which involves the integration of modern databases and various perspectives from researchers in the field. We review the most recent literature on this subject and explore the diverse theories and approaches adopted by scientists to address the challenges posed by these complex images.

KeyWords: Breast Cancer, Breast Density, Risk Factors, Deep Learning, BI-RADS, CNN

Résumé

La médecine prédictive aide les médecins à évaluer les patientes et à les encourager à prévenir le risque de cancer du sein en faisant des prédictions précises basées sur l'analyse des images medicales , permettant ainsi une détection précoce et une intervention opportune. Cela dépend de nombreux facteurs, tels que l'âge, le nature du vie, les antécédents familiaux et la densité mammaire, qui est le facteur le plus important.Cette thèse présente l'état actuel de la recherche sur l'impact de l'application de techniques d'intelligence artificielle et d'apprentissage profond dans la detection de la densité mammaire.La première partie de notre discussion se concentre sur le cancer du sein, en particulier sur l'évaluation de la densité mammaire et les défis auxquels sont confrontés les radiologues dans ce domaine. Nous plongeons ensuite dans le domaine de l'apprentissage profond, mettant en évidence les dernières études dans ce domaine en constante évolution, avec un accent particulier sur son impact sur le secteur médical. Bien que le potentiel de l'apprentissage profond pour améliorer le diagnostic médical soit prometteur, sa dépendance à des données précises et exactes est vitale, car elle a un impact direct sur les résultats des patients. Bien que certains pays aient déjà commencé à intégrer des techniques d'apprentissage profond dans leurs pratiques médicales, il reste encore beaucoup de progrès à faire. Cependant, les premiers résultats ont été prometteurs, et le potentiel de progrès continus dans ce domaine est optimiste. La deuxième partie de cette thèse est consacrée aux dernières recherches sur l'analyse d'images complexes de mammographie en utilisant des techniques d'apprentissage profond de pointe, qui implique l'intégration de bases de données modernes et de diverses perspectives des chercheurs dans le domaine. Nous passons en revue la littérature la plus récente sur ce sujet et explorons les diverses théories et approches adoptées par les scientifiques pour relever les défis posés par ces images complexes.

Mots clés: Cancer du sein, Densité mammaire, Facteurs de risque, Apprentissage en profondeur, BI-RADS, CNN

الملخص

يفيد الطِبُّ التَنبئيُّ في تَقْيِيمِ المَرْضَى وَمُساعَدَتِهِمْ عَلَى الوقَايَةِ مِنْ مَخَاطِرِ الإِصَابَةِ بِسَرَطَانِ الثَّدِيِّ، حَيْثُ يَمكُنُهُمْ إِجْرَاءُ تَنَبُّؤَاتٍ دَقِيقَةٍ بِنَاءً عَلَى تَحْلِيلِ صُورِ الثدي بالأَشْعَّةِ، مِمَّا يُسَمِّحُ بِالكَشْفِ الْمُبَكِّرِ وَالتَّدْخُلِ فِي الْوَقْتِ الْمُنَاسِبِ. هَذَا يَعْتَمِدُ عَلَى عَوَامِلَ عَدِيدَةٍ مُثَلَّ الْعُمْرِ وَنَمْطِ الْمَعِيْشَةِ وَتَارِيخِ الْعَائِلَةِ وَكَثَافَةِ الثَّدِيِّ، وَهُوَ الْعَامِلُ الأَكْثَرُ أَهَمِّيَّةً. فِي هَذِهِ الْأَطْرُوحَةِ . سَنُقَدِّمُ نَتَائِجَ الْحَالَةِ الرَاهِنَةِ فِي الْبَحْثِ حَوْلَ تَأْثِيرِ تَطْبِيقِ تَّقَنِيَّاتِ التَّعَلُّمِ الْعَمِيقِ فِي تَحْلِيلِ الصُّوَرِ الطِّبِّيَّةِ لِلثَّدِيِّ. يَتَمَحَّور الْجُزْءُ الْأَوَّلُ مِنْ مُنَاقَشَتِنَا سَرَطَانِ الثَّدِيِّ وَتَقْيِيمِ خَطَرِ الْإِصَابَةِ بِهِ. كَمَا سَنُسَلِّطُ الضَّوْءَ عَلَى تَحَدِّيَاتِ التَّشْخِيصِ الَّتِي تُوَاجِهُ أَطْبَاءَ الْأَشْعَّةِ فِي تَقْيِيمِ كَثَافَةِ الثَّدِيِّ .ثُمَّ سَنَتَحَدَّثُ بِالتَفْصِيلِ عَنْ مَجَالِ التَّعَلُّمِ الْعَمِيقِ وَتَأْثِيرِهِ فِي تَحْلِيلِ الصُّوَرِ الطِّبِّيَّةِ، حَيْثُ سَنُعْرِضُ أَحْدَثَ الدِّرَاسَاتِ وَالْأَبْحَاثِ فِي هَذَا الْمَجَالِ. وَ ننوه أَنَّ نَجَاحَ التَّعَلُّمِ الْعَمِيقِ يَعْتَمِدُ بِشَكْلٍ كَبِيرٍ عَلَى تَوْفُرِ بَيَانَاتٍ دَقِيقَةٍ وَمُوَثَّقَةٍ التي بدورها تَؤَثِّرُ بِشَكْلٍ مُبَاشَرٍ عَلَى نَتَائِجِ الْمَرْضَى . يَتَضَمَّنُ الْجُزْءُ الثَّانِي أَحَدَثَ الْأَبْحَاثِ فِي تَحْلِيلِ الصُّوَرِ الْمُعْقَدَةِ لِلثَّدِيِّ بِاسْتِخْدَامِ تَقَنِيَّاتِ التَّعَلُّمِ الْعَمِيقِ. سَنُقَدِّمُ مُرَاجَعَةً شاملة لِلْأَدَبِيَّاتِ الْحَدِيثَةِ فِي هَذَا الْمَجَالِ وَنَسْتَكْشِفُ مُخْتَلَفَ النَّظَرِيَّاتِ وَالْمَقَارَبَاتِ الَّتِي يَعْتَمِدُهَا الْبَاحِثُونَ لِلْتَّعَامُلِ مَعَ المشاكل الَّتِي تُوَاجِهُهُم. في الأخير نَهْدَفُ مِنْ خَلَالِ هَذِهِ الْأَطْرُوحَةِ إِلَى تَوْفِيرِ نَظَرَةٍ عَامَّةٍ عَنْ إِنْجَازَاتِ تقنيات التَّعَلُّمِ الْعَمِيقِ فِي تفسير صور الثَّدِيِّ وَمَدَى تَأْثِيرِهَا فِي التَّشْخِيصِ الطِّبِّيِّ .

الكلمات المفتاحية: سرطان الثدي ، كثافة الثدي ، عوامل الخطر ، التعلم العميق ، الشبكات العصبة التلافيفية ، نظام بيانات تقارير تصوير الثدي

Acknowledgements

First of all, thank you Allah for giving us the courage and strength to complete this thesis.

We would also like to thank our supervisor Dr. Dif Nassima for her valuable advice, corrections, and suggestions that greatly contributed to the completion of this work. Also, we would like to express our sincere gratitude to Dr. Djazia Benchuk for her valuable support throughout this research, especially in the medical field. Her continuous encouragement contributed to this work.

In addition, we would like to thank the president of the jury as well as the members of the jury who agreed to review and evaluate our research. Their expertise and feedback improve the quality of our work.

We commend the (ESI-SBA) School and its dedicated mentors for their academic support in providing us with a solid foundation to overcome the challenges posed by this research. We also thank all the staff members for providing us with a comfortable environment for our research.

Dedication

To my dearest father, Ben Mhamed, my beloved mother, Zouaia Mama, and my wonderful little brother, Khaled. The three most important people in my life. Today, I want to express my deepest gratitude and appreciation for your support and relentless hard work. It is through your love and guidance, alongside the blessings of God, that I stand in the position I am today. Words cannot fully capture the extent of your encouragement and affection toward me. From every corner and at all times, I am honored to convey my heartfelt thanks to each of you. I particularly mention my mother, who is battling a grave illness. Despite this, she has never faltered in her love and care for our family. Her strength, patience, and resilience inspire me every time. I pray with all my heart that Allah will grant her healing. This work is dedicated to all of you as a symbolic gift, representing my immense gratitude.

I also extend my gratitude to my beloved family, who played an active role in my achievements, never failing to support me.

Additionally, I want to express my sincere thanks to my dear friends, with whom I shared countless moments and joyous days: Nadjib, Adel, Yannis, Seddik, Mukhtar, Aymen, and all my friends. Your support during these times means the world to me.

Last but not least, I cannot overlook the contributions of my friend and business partner, Amine, whose role was instrumental in the success of this work.

Standing here, I say to all of you: Thank you, and may God grant us success and contentment in all that we pursue.

Boudinar.

Acronyms

DL	*Deep Learning*
WHO	*World Health Organization*
AJHS	*Algerian Journal of Health Sciences*
BRCA	*Breast Cancer Gene*
MLO	*Mediolateral Oblique*
CC	*Cranio Caudal*
ACR	*American College of Radiology*
BI-RADS	*Breast Imaging-Reporting and Data System*
MRI	*Magnetic resonance imaging*
AI	*Artificial intelligence*
MIAS	*Mammographic Image Analysis Society*
PGM	*Portable Gray Map*
DDSM	*Digital Database for Screening Mammography*
IRMA	*Image Retrieval in Medical AppliCation*
CBIS-DDSM	*Curated Breast Imaging Subset of DDSM*
AOUP	*Azienda Ospedaliero Universitaria Pisana*
FFDM	*Full Field Digital Mammography*
SFM	*Screen Film Mammography*
CNN	*Convolutional neural network*

DICOM	*Digital Imaging and Communications in Medicine*
PNG	*Portable Network Graphics*
RBG	*Red Green Blue*
SGD	*Stochastic Gradient Descent*
GAP	*Global Average Pooling*
VS-LEF	*View Specific Late Evidential Fusion*
FLEF	*Feature Late Evidential Fusion*
LEF	*Late Evidential Fusion*
VS-FLEF	*View Specific Future Late Evidential Fusion*
DCN	*Deep Convolution Network*
FC	*Fully Connected*
DC	*Dilated Convolution*
CA	*Channel Wise Attention*
NAG	*Nesterov Accelerated Gradient*
TP	*True Positive*
FP	*False Positive*
FN	*False Negative*
TN	*True Negative*
OCA	*Overall Classification Accuracy*
ICA	*Individual Classification Accuracy*
RC	*Recall*
SP	*Specifity*
PR	*Precision*
AUC	*Area Under Curve*
SSE	*The Sum of Squared Errors*