

Saniva - AI for Early Hope

أعدّت هذه الأطروحة لإنجاز
مشروع التخرج 1 في اختصاص الذكاء الصنعي وعلوم البيانات

إعداد الطالب:

ينال عماد الشوفي

اشراف:

الدكتورة ماجدة البكور
المهندسة آية الأسود

العام الدراسي 2025/2026

اقرار المشرف

أشهد بأن إعداد هذا المشروع الموسوم
..... والمعد من قبل الطالب
..... قد تم تحت اشرافى في قسم هندسة الذكاء الصناعي وعلوم البيانات - كلية هندسة الذكاء الاصطناعي،
وهو جزء من متطلبات نيل شهادة الإجازة في الهندسة / هندسة الذكاء الصناعي وعلوم البيانات.
المرتبة العلمية الاسم التوقيع التاريخ

أشهد بأن إعداد هذا المشروع الموسوم
..... والمعد من قبل الطالب
..... قد تم تحت اشرافي في قسم هندسة الذكاء الصناعي وعلوم البيانات - كلية هندسة الذكاء الاصطناعي ،
وهو جزء من متطلبات نيل شهادة الإجازة في الهندسة / هندسة الذكاء الصناعي وعلوم البيانات.
المرتبة العلمية الاسم التوقيع التاريخ

الملخص

يُعد سرطان الثدي من أكثر أنواع السرطان شيوعاً بين النساء على مستوى العالم، مما يُبرز الحاجة الملحة إلى أنظمة تشخيص آلي دقيقة وموثوقة. تقدم هذه الأطروحة نظام Sanvia وهو إطار عمل ذكي متعدد المهام يهدف إلى تشخيص سرطان الثدي الآلي من خلال تحليل صور التصوير الشعاعي للثدي (Mammography) وتوليد تقارير طبية تلقائية. يتميز النظام المقترن بقدراته على معالجة التحديات الجوهرية في هذا المجال، ومنها التوزيع غير المتوازن الشديد للفئات حيث تمثل فئة BI-RADS 4 نسبة 0.27% فقط من البيانات، ودمج المعلومات من أربعة مناظر تصويرية مختلفة (L_CC, L_MLO, R_CC, R_MLO) على بنية عصبية عميقه تتضمن شبكة EfficientNetB4 كأساس مع الآلية دمج مبتكرتين: Efficient Cross Attention و Gated Fusion Layer للتحكم التكيفي في تدفق المعلومات. كما يتكامل النظام مع نماذج اللغة لتوليد تقارير طبية مفصلة. تم تدريب النظام باستخدام مجموعة بيانات Vindr-Mammo التي تضم 20,000 صورة من 5,000 دراسة سريرية. حقق Sanvia دقة 92.2% في تصنيف نظام BI-RADS مع F1-Score كلي 87.5%， ودقة 95.4% في تصنيف كثافة الثدي (Breast Density). بشكل ملحوظ، حقق النظام دقة 100% واسترجاع 80% للفئة الأندر (BI-RADS 4) مما يُظهر فعالية استراتيجيات معالجة الاختلال في التوزيع. يُسهم هذا العمل في سد الفجوة بين الأنظمة التقنية المتقدمة والتطبيق السريري العملي، مما يفتح آفاقاً واعدة لدعم الأخصائيين الشعاعيين في اتخاذ قرارات تشخيصية أسرع وأكثر دقة.

الكلمات المفتاحية: التعلم العميق، التصوير الشعاعي للثدي، التعلم المتعدد المهام، نماذج اللغة، معالجة الاختلال في فئات البيانات.

Abstract

Breast cancer represents the most prevalent malignancy among women worldwide, necessitating the development of accurate and reliable automated diagnostic systems. This thesis presents Sanvia (Smart Automated Network for Visual Inspection and Assessment), an intelligent multi-task framework designed for automated breast cancer diagnosis through mammographic image analysis and automatic medical report generation. The proposed system addresses inherent challenges in this domain, including severe class imbalance where BI-RADS category 4 constitutes merely 0.27% of the data, and the integration of information from four distinct mammographic views (L_CC, L_MLO, R_CC, R_MLO). Sanvia employs a deep neural architecture featuring EfficientNetB4 as the backbone, complemented by two innovative fusion mechanisms: EfficientCrossAttention for multi-view integration and GatedFusionLayer for adaptive information flow control. Furthermore, the system integrates with Large Language Models (LLMs) to generate detailed, interpretable medical reports. The system was trained using the Vindr-Mammo dataset comprising 20,000 images from 5,000 clinical studies. Sanvia achieved 92.2% accuracy in BI-RADS classification with a macro F1-score of 87.5%, and 95.4% accuracy in breast density classification. Notably, the system attained 100% precision and 80% recall for the rarest category (BI-RADS 4), demonstrating the efficacy of the imbalance handling strategies. This work bridges the gap between advanced technical systems and practical clinical applications, offering promising prospects for supporting radiologists in making faster and more accurate diagnostic decisions.

Keywords: Deep Learning, Mammography, Multi-task Learning, Large Language Models, Class Imbalance Handling.

قائمة المحتويات

II.....	إقرار المشرف
III.....	الملخص
IV	Abstract
V.....	قائمة المحتويات
VII.....	قائمة الأشكال
VII	قائمة الجداول
VII	قائمة المعادلات
VIII.....	قائمة بأهم المصطلحات
2.....	الفصل الأول - المقدمة
2.....	1.1 تمهيد
2.....	2.1 مشكلة البحث
3.....	3.1 الهدف من المشروع
3.....	4.1 النتائج والإسهامات
4.....	5.1 هيكلية البحث
6.....	الفصل الثاني - الأدبيات السابقة
9.....	الفصل الثالث - إدارة المشروع
9.....	1.3 المقدمة
9.....	2.3 ميثاق المشروع
10.....	3.3 وثيقة نطاق العمل
11.....	4.3 خطة المشروع - مخطط غانت
12.....	5.3 إدارة المخاطر
12.....	6.3 إدارة التماثج والتجارب
12.....	7.3 خطة العمل
13.....	8.3 الخلاصة
15.....	الفصل الرابع - المنهجية
15.....	1.4 مقدمة
15.....	2.4 مراجعة تقنية للأدبيات
16.....	3.4 الفجوات على المستويين المعرفي والتطبيقي
16.....	4.4 منهجية العمل
17.....	5.4 مجموعة المعطيات المستخدمة
17.....	Vindr-Mammo Dataset 1.5.4
17.....	2.5.4 خصائص التسميات
18.....	3.5.4 التحديات في البيانات
19.....	6.4 تحضير مجموعة المعطيات
19.....	1.6.4 تنظيف البيانات

19.....	2.6.4 دمج البيانات الجدولية
20.....	3.6.4 إنشاء View Mapping
20.....	4.6.4 معالجة عدم توازن الفئات
23.....	7.4 البنية التقنية والتنفيذية
23.....	1.7.4 الأجهزة والبنية التحتية
24.....	2.7.4 البرمجيات والمكتبات
24.....	3.7.4 إدارة النماذج والتجارب
25.....	8.4 المنهجية المقترحة
25.....	1.8.4 النظرة العامة
26.....	2.8.4 الـ Backbone (EfficientNetB4)
27.....	3.8.4 EfficientCrossAttention
29.....	4.8.4 طبقة الاندماج البوابي (GatedFusionLayer)
29.....	5.8.4 TabularEncoder
30.....	6.8.4 الرؤوس متعددة المهام (Multi-task Heads)
31.....	7.8.4 معمارية النموذج الكلية
33.....	9.4 مقاييس الأداء
33.....	1.9.4 مقاييس التصنيف الأساسية
34.....	2.9.4 مقاييس متعددة الفئات
34.....	10.4 النموذج الأساسي
34.....	1.10.4 إحصائيات النموذج
34.....	2.10.4 إعدادات التدريب
37.....	الفصل الخامس - التجارب والنتائج والتقييم
37.....	1.5 آلية عمل الاختبار
37.....	1.1.5 تقسيم البيانات
37.....	2.1.5 إعدادات التدريب التفصيلية
38.....	3.1.5 بيئة التنفيذ والزمن
38.....	4.1.5 Learning Curves تحليل نتائج Learning Curves
39.....	2.5 نتائج الاختبار
39.....	1.2.5 نتائج تصنيف BI-RADS
41.....	2.2.5 نتائج تصنيف كثافة الثدي (Breast Density)
42.....	3.5 تحليل ومقارنة النتائج
42.....	1.3.5 مقارنة مع الدراسات السابقة
42.....	2.3.5 دراسة تقييم أداء المكونات
45.....	الفصل السادس - توليد التقرير الطبي
45.....	1.6 ربط API
45.....	2.6 تصميم System Message

46.....	3.6 توليد التقرير النصي
46.....	1.3.6 هيكليّة التقرير المُولد
46.....	2.3.6 ضبط الجودة
46.....	4.6 تجميع التقرير الشامل
47.....	5.6 مقارنة مع تقارير الأخصائيين
50.....	الفصل السابع - خاتمة وآفاق مستقبلية
50.....	1.7 خاتمة
50.....	2.7 آفاق مستقبلية

قائمة الأشكال

11.....	الشكل 1.3: مخطط غانت
39.....	الشكل 2.5: منحنيات التعلم للدقة والخسارة
40.....	الشكل 3.5: مصفوفة الالتباس BI-RADS
41.....	الشكل 4.5: مصفوفة الالتباس Density

قائمة الجداول

9.....	الجدول 1.3: ميثاق المشروع
11.....	الجدول 2.3: وثيقة نطاق العمل
12.....	الجدول 3.3: جدول المخاطر المحتملة
17.....	الجدول 4.4: خصائص مجموعة بيانات Vindr-Mammo
21.....	الجدول 5.4: BI-RADS Oversampling
24.....	الجدول 6.4: البرمجيات والمكتبات
37.....	الجدول 7.5: جدول تقسيم البيانات
39.....	الجدول 8.5: تقرير التصنيف BIRADS
41.....	الجدول 9.5: تقرير التصنيف Density
42.....	الجدول 10.5: مقارنة النتائج مع أبحاث سابقة
43.....	الجدول 11.5: دراسة تأثير مكونات النموذج
47.....	الجدول 12.6: تقييم أخصائي الأشعة

قائمة المعادلات

20.....	معادلة 1.4 : MinMaxScaler
20.....	معادلة 2.4 : Effective Number Weighting
28.....	معادلة 3.4 : Cross-Attention
29.....	معادلة 4.4 : الاندماج البوابي
33.....	المعادلة 5.4: الدقة

33.....	المعادلة 6.4: التنبؤ الإيجابي.....
33.....	المعادلة 7.4: الاسترجاع.....
34.....	المعادلة F1 Score :8.4
35.....	المعادلة Focal Loss :9.4
35.....	المعادلة Total Loss :10.4

قائمة بأهم المصطلحات

المصطلح التقني	الترجمة العربية	الاختصار	المعنى
Artificial Intelligence	الذكاء الاصطناعي	AI	مجال يهدف إلى تكثين الأنظمة من محاكاة الذكاء البشري
Deep Learning	التعلم العميق	DL	فرع من التعلم الآلي يعتمد على الشبكات العصبية العميقية
Machine Learning	التعلم الآلي	ML	خوارزميات تتعلم الأنماط من البيانات
Convolutional Neural Network	الشبكة العصبية الانقافية	CNN	شبكة متخصصة في معالجة الصور
Mammography	تصوير الثدي الشعاعي	—	تقنية تصوير طبية للكشف عن سرطان الثدي
Breast Cancer	سرطان الثدي	—	مرض ناتج عن نمو غير طبيعي لخلايا الثدي
BI-RADS	نظام تقارير تصوير الثدي	BI-RADS	معيار طبي لتصنيف نتائج الماموغرام
Breast Density	كثافة الثدي	—	نسبة النسيج الغدي إلى الدهني في الثدي
Multi-task Learning	التعلم متعدد المهام	MTL	تدريب نموذج واحد على عدة مهام متزامنة
Class Imbalance	عدم توازن الفئات	—	تقاوت كبير في عدد العينات بين الفئات
EfficientNetB4	—	EfficientNet B4	شبكة عميقة فعالة لاستخراج السمات
Backbone Network	الشبكة الأساسية	—	الجزء المسؤول عن استخراج الميزات
Cross-Attention	الانتباه المتقطع	—	آلية انتباه لدمج معلومات من مصادر مختلفة
Efficient Cross Attention	الانتباه المتقطع الكفو	ECA	آلية دمج متعددة الرؤى بكلفة حسابية أقل
Gated Fusion Layer	طبقة الدمج البوابي	GFL	طبقة تحكم بتدفق المعلومات بين السمات
Multi-view Learning	التعلم متعدد المناظر	—	دمج معلومات من زوايا تصوير مختلفة
Feature Fusion	دمج السمات	—	الجمع بين تمثيلات متعددة للبيانات
Large Language Models	نماذج اللغة الكبيرة	LLMs	نماذج لغوية قادرة على توليد نصوص طيبة
Tabular Encoder	مشقر البيانات الجدولية	—	طبقة لمعالجة البيانات الرقمية
Vindr-Mammo Dataset	مجموعة بيانات فيندر-مامو	—	قاعدة بيانات ماموغرام سريرية
DICOM	تنسيق الصور الطبية	DICOM	معيار عالمي لتخزين الصور الطبية

استخدام نموذج مدرب مسبقاً لمهمة جديدة	TL	التعلم بالنقل	Transfer Learning
مضاعفة عينات الفئات النادرة	—	زيادة العينات	Oversampling
توليد عينات جديدة من البيانات الأصلية	—	تعزيز البيانات	Data Augmentation
أسلوب لمعالجة عدم توازن البيانات	ENW	الوزن الفعال للفئات	Effective Number Weighting
دالة خسارة تعزز تعلم الفئات النادرة	—	خسارة التركيز	Focal Loss
دمج معلومات صور مأخوذة من زوايا مختلفة	—	دمج المناظر المتعددة	Multi-view Integration
تصوير الثدي الأيسر من الأعلى	L_CC	المنظر القصفي الذيلي الأيسر	Left Cranio-Caudal
تصوير الثدي الأيمن من الأعلى	R_CC	المنظر القصفي الذيلي الأيمن	Right Cranio-Caudal
تصوير جانبي مائل للثدي الأيسر	L_MLO	المنظر المائل الأيسر	Left Medio-Lateral Oblique
تصوير جانبي مائل للثدي الأيمن	R_MLO	المنظر المائل الأيمن	Right Medio-Lateral Oblique
وسيلة ربط بين الأنظمة البرمجية	API	واجهة برمجة التطبيقات	API
نموذج لغوي لتوليد النصوص الطبية	GPT	نموذج GPT-4o	GPT-4o

الفصل الأول

المقدمة

الفصل الأول - المقدمة

1.1 تمهيد

يُعد سرطان الثدي أحد أكثر الأمراض الخبيثة انتشاراً بين النساء عالمياً، حيث سجلت منظمة الصحة العالمية (WHO) أكثر من 2.3 مليون حالة إصابة جديدة عام 2022، مما يجعله السبب الرئيسي لوفيات السرطان بين النساء في معظم الدول [1]. وفي العالم العربي، ولا سيما دول الخليج، تشهد معدلات الإصابة ارتفاعاً ملحوظاً، الأمر الذي يُبرز الحاجة إلى أنظمة كشف مبكر دقيقة ومتاحة [2].

يعتبر التصوير الشعاعي للثدي (Mammography) المعيار الذهبي للكشف المبكر، لقدرته على اكتشاف الأورام قبل ظهورها سريرياً بسنوات [3]. إلا أن التشخيص اليدوي يواجه تحديات رئيسية، أبرزها الخطأ البشري، حيث قد تصل معدلات التشخيص الخاطئ إلى 30% في حالات كثافة الثدي العالية [4]، إضافة إلى النقص العالمي في الأخصائيين الشعاعيين، خصوصاً في الدول النامية [5]، وما يرافق ذلك من عبء اقتصادي كبير على الأنظمة الصحية [6].

في هذا السياق، برز الذكاء الاصطناعي وأنظمة CADe/CADx كحل واعد لدعم القرار الطبي، مع قدرتها على تحقيق أداء يضاهي دقة الأخصائيين في مهام محددة [7]. ومع ذلك، لا تزال هذه الأنظمة تعاني من قصور في التعامل مع البيانات غير المتوازنة، ودمج المناظر المتعددة، وتوفير تفسير سريري وتقرير طبي متكملاً. وانطلاقاً من ذلك، تقدم هذه الأطروحة نظام Sanvia كإطار ذكي متكملاً يعتمد على التعلم العميق ونماذج اللغة البصرية لتقديم حل تشخيصي شامل ومُفَسّر سريرياً.

2.1 مشكلة البحث

على الرغم من التقدم الملحوظ في مجال الذكاء الاصطناعي الطبي، لا تزال الأنظمة الحالية تعاني من قصور جوهري يحد من فعاليتها في البيئات السريرية الواقعية، إذ تواجه تحديات هندسية وسريرية معقدة تبدأ بمشكلة عدم توازن الفئات الحاد في مجموعات بيانات التصوير الشعاعي للثدي (ماموغرام)، حيث تسسيطر الحالات الطبيعية (BI-RADS 1-2) على الغالبية العظمى بينما تشكل الحالات المشتبهة نسباً ضئيلة للغاية (مثل تمثل فئة BI-RADS 4 نحو 0.27% فقط في مجموعة Vindr)، مما يحرف النماذج التعليمية نحو الفئات الأكثر تكراراً ويُضعف قدرتها على اكتشاف الحالات الحرجة بسبب معالجة دالة الخسارة القياسية للأخطاء بمساواة دون مراعاة الأهمية السريرية

للفرق بين الخطأ الإيجابي والخطأ السلبي. فضلاً عن ذلك، تعاني الأنظمة من قصور في آليات دمج المناظر المتعددة، إذ تعالج بعضها كل منظور (CC و MLO للثديين) بشكل منفرد فت فقد الارتباطات المكانية بينها، أو تعتمد على أساليب دمج بدائية مثل Concatenation دون مراعاة الأهمية النسبية لكل منظر في السياق التشريحي. علاوة على ذلك، تقتصر معظم الأنظمة على مرحلة التصنيف الرقمي دون الانتقال إلى توليد تقارير طبية منهجية باللغة الطبيعية، مما يخلق فجوة بين المخرجات الآلية والتوثيق السريري المطلوب ويفرض تدخلاً برياً إضافياً. وأخيراً، تظل الفجوة واسعة في تخصيص هذه الأنظمة للسياق العربي والم المحلي، إذ تفتقر المنطقة إلى حلول ذكاء اصطناعي في الماموغرام التي تستند إلى البيانات السكانية المحلية أو توفر واجهات وتقارير طبية باللغة العربية، مما يحد من إمكانية اعتمادها في الأنظمة الصحية العربية.

3.1 الهدف من المشروع

يهدف هذا المشروع إلى تطوير نظام ذكي متكامل لتشخيص سرطان الثدي آلياً يعتمد على تقنيات التعلم العميق ونمذاج اللغة المتقدمة، وذلك من خلال تحليل صور التصوير الشعاعي للثدي من عدة مناظر تصويرية ودمجها بطريقة فعالة تحاكي سير العمل السريري الحقيقي. يسعى النظام إلى معالجة التحديات الجوهرية، وتحسين دقة اكتشاف الحالات الحرجة ذات الأهمية السريرية العالية. كما يهدف المشروع إلى تجاوز الاقصار على التصنيف الرقمي عبر توفير تفسير سريري واضح وتوليد تقارير طبية نصية تلقائية باللغة الطبيعية، بما يساهم في تقليل الفجوة بين المخرجات الآلية ومتطلبات التوثيق السريري العملي. ويطمح هذا العمل إلى تقديم حل تشخيصي موثوق وقابل للتطبيق في البيئات السريرية الواقعية، يدعم الأخصائيين الشعاعيين في اتخاذ قرارات تشخيصية أسرع وأكثر دقة، مع إرساء أساس قابل للتطوير نحو دعم السياق العربي والم المحلي في أنظمة الذكاء الاصطناعي الطبية.

4.1 النتائج والإسهامات

تقدم هذه الأطروحة مجموعة من الإسهامات العلمية والهندسية لمعالجة التحديات في تشخيص سرطان الثدي الآلي.

• تطوير نظام Sanvia متعدد المهام

تم تصميم Sanvia لأداء مهام متزامنة: تصنيف BI-RADS بدقة 92.2%，تصنيف كثافة الثدي بدقة 95.4%， وتوليد تقارير طبية مفصلة باستخدام LLM. هذا التكامل يقلل الحاجة إلى أنظمة منفصلة ويعزز الكفاءة الحسابية عبر مشاركة التمثلات.

• المبتكرة EfficientCrossAttention

آلية Cross-Attention لدمج المناظر المتعددة تميز بالكفاءة الحسابية $O(n)$ مقارنة بـ $O(n^2)$ في Self-Attention ، ومرونة التعامل مع المناظر المفقودة، مع تعلم الأهمية النسبية لكل منظر ديناميكياً حسب سياق الحالة.

• نظام توليد تقارير مدمج مع LLM

يشمل التكامل استخراج السمات العميقه من الطبقات الوسطى، إنتاج نصوص سريرية باستخدام GPT-40، وتصدير تقارير PDF ، ما يتيح أتمتة كاملة لسير العمل التشخيصي.

• معالجة فعالة لعدم توازن الفئات

تم تطبيق استراتيجية مركبة تشمل Oversampling، Effective Number Weighting بالفئات النادرة باستخدام Albumentations Focal Loss ($\gamma=3, \alpha=0.5$)، مما حسن معالجة عدم توازن الفئات 4 BI-RADS Recall للفئة 4 من 20% إلى 80%.

• مساهمات في المعرفة الهندسية

تشمل تحسين Transfer Learning عبر تجميد انتقائي 70% من طبقات BI-RADS Multi-task Learning، توافق ديناميكي في بين مهام EfficientNetB4 و Density، ودمج LLM لتوليد تفسيرات مركبة.

5.1 هيكليّة البحث

تنظم هذه الأطروحة في سبعة فصول رئيسية، حيث يقدم الفصل الأول تمهيداً لموضوع سرطان الثدي وأهميته، يليه تحديد مشكلة البحث وعرض النتائج والإسهامات العلمية والهندسية لهذه الدراسة. يستعرض الفصل الثاني الأدبيات السابقة والتطور التاريخي لأنظمة وتطبيقات التعلم العميق في الماموغرام، مع تحديد الفجوات البحثية التي تهدف هذه الدراسة إلى سدها. ثم يأتي الفصل الثالث الذي يهتم بإدارة المشروع حيث يستعرض ميثاق المشروع ونطاقه ثم إدارة المخاطر المحتملة بالإضافة إلى توثيق التجارب والنتائج. أما الفصل الرابع فيقدم التفاصيل الكاملة للمنهجية المقترحة بدءاً من مراجعة تقيية للمفاهيم الأساسية، مروراً بتحليل مجموعة البيانات Vindr-Mammo، ووصولاً إلى البنية المعمارية التفصيلية لنظام Sanvia، مع شرح شامل لأليات دمج المناظر، ومعالجة البيانات، واستراتيجيات التدريب ومقاييس الأداء. يصف الفصل الخامس إعدادات التجارب وتقسيم البيانات، ويعرض النتائج الكمية لتصنيف BI-RADS و Density. يركز الفصل السادس على الجانب التطبيقي للنظام، بدءاً من ربط APIs لنماذج اللغة، وتوليد التقارير النصية، وتجميع التقرير الشامل، مع تقييم سريري لجودة النتائج. وأخيراً، يقدم الفصل السابع تلخيص المساهمات الرئيسية، والإجابة على أسئلة البحث، ومناقشة القيود، مع تقديم توصيات مستقبلية لتحسين النموذج، وتوسيع النظام، وتعزيز التحقق السريري، ومراعاة الاعتبارات الأخلاقية والاجتماعية.

الفصل الثاني

الأدبيات السابقة

الفصل الثاني – الأدبيات السابقة

تنوعت الجهود البحثية في مجال الذكاء الاصطناعي لتشخيص سرطان الثدي عبر التصوير الشعاعي بين نماذج الرؤية-اللغة المتعددة الرؤى والشبكات التلافيافية العميقه والمراجعات المنهجية. فعلى سبيل المثال، اعتمدت دراسة [8] على نموذج GLAM المبني على محولات الرؤية (ViT) ومشفر نصوص BioClinical-BERT مع محاذاة هندسية محلية على المحور الأمامي-الخلفي، مسجلة تحسينات ملموسة في كفاءة البيانات القليلة والمتناهية عبر مجموعات VinDr-EMBED وRSNA-Mammo وVGG19. وفي سياق دمج الرؤى المتعددة، طورت دراسة [9] استراتيجية RSNA-Mammo وDBE لدمج إسقاطي CC وMLO، محققة AUC يبلغ 0.9051 باستخدام ResNet50، مع استقرار أفضل للاستراتيجية DBE عبر مجموعات بيانات متعددة رغم وجود فجوة أداء بين البيانات المرئية وغير المرئية. كذلك تناولت دراسة [10] تصنيف كثافة الثدي باستخدام شبكة CD-CNN مخصصة مع طبقة ELM، فبلغت دقة 95.4% على البيانات الداخلية و73.9% على مجموعة mini-MIAS الخارجية، مما أبرز إمكانية التطبيق في البيانات محدودة الموارد. من ناحية أخرى، قدمت دراسة [11] نظاماً ثنائياً للخطوات يجمع بين شبكة متعددة المهام لاستخراج الآفات وEfficientNet-B0 للتصنيف النهائي، مسجلة AUC يصل إلى 0.995 على التحقق و0.933-0.947 على الاختبار الخارجي، مع خفض 83.1% من الإيجابيات الكاذبة في تحليل فئات BI-RADS. أما دراسة [12]، فقد ركزت على توليد تقارير سريرية تلقائية عبر نموذج ClinicalBLIP المبني على InstructBLIP مع ضبط دقيق متعدد المراحل وLoRA، محققة قيمة METEOR تصل إلى 0.570 وROUGE-L تصل إلى 0.534 على مجموعة X-RAY-IU. وفي إطار بناء نماذج أساسية، حققت دراسة [13] بتطوير Mammo-CLIP المدرب على أزواج صور-تقارير AUC يصل إلى 0.91 في تصنیف السرطان بعد ضبط دقيق كامل و0.88 في تصنیف الكتل، مع إمكانیة توطین السمات دون صناديق إحاطة. كما قدّمت المراجعة المنهجية [14] تحلیلاً شاملًا لـ 31 دراسة في توليد التقارير الطبية، مشيرةً إلى تطور المجال من نماذج CNN-RNN إلى أطر هرمية وتعلم تعزيري، مع تحديات في عدم توازن البيانات وغياب معايير تقييم طبية موحدة.

خلاصة

تؤكد الدراسات السابقة فعالية نماذج الرؤية والشبكات التلaffيفية العميقه والهجينة في تحسين دقة تشخيص سرطان الثدي وتصنيف كثافته وتوليد التقارير، إلا أنها تتعرض عادةً لمشكلات تتعلق بالتعوييم عبر معدات تصوير مختلفة، والتحديات الحاسوبية في البيانات محدودة الموارد، وقلة التكامل بين التحليل البصري والتوليد اللغوي للتقارير. يقدم مشروع Saniva نهجاً متكاملاً يجمع بين التقنيات المتقدمة متعددة الرؤى والنماذج الخفيفة لتحقيق تشخيص مبكر فعال وقابل للتطبيق السريري الواسع.

الفصل الثالث

إدارة المشروع

الفصل الثالث - إدارة المشروع**1.3 المقدمة**

في هذا الفصل، يتم استعراض الجوانب الإدارية والتنظيمية لمشروع SANVIA، والتي تُعد عنصراً داعماً لنجاح تطوير نظام تشخيص ذكي معتمد على الذكاء الاصطناعي في المجال الطبي. يرتكز هذا الفصل على ميثاق المشروع، خطة العمل، وثيقة نطاق العمل (SOW)، إدارة المخاطر، وآليات التحكم بالتجارب والنمذاج، بما يضمن تنفيذ المشروع بطريقة منهجية ومنظمة منذ مرحلة الإعداد وحتى الوصول إلى النتائج النهائية.

2.3 ميثاق المشروع

يُعد ميثاق المشروع وثيقة رسمية تحدد الإطار العام للمشروع وتوضح أهدافه ونطاقه، كما تمثل مرجعاً أساسياً لاتخاذ القرارات خلال مراحل التنفيذ المختلفة.

الجدول 1.3: ميثاق المشروع

عنوان المشروع	Sanvia : AI for Early Hope		
تاريخ بدء المشروع	2025/11/1		
تاريخ انتهاء المشروع	2026/2/1		
مدير المشروع	الدكتورة ماجدة البكور، المهندسة آية الأسود		
أهداف المشروع	يهدف مشروع SANVIA إلى تطوير إطار عمل ذكي متعدد المهام لتشخيص سرطان الثدي آلياً، وذلك من خلال تحليل صور التصوير الشعاعي للثدي (Mammography) ودمج المعلومات القادمة من عدة مناظر تصويرية مختلفة. يعتمد النظام على تقنيات التعلم العميق ونمذاج الذكاء الاصطناعي المتقدمة لمعالجة التحديات المرتبطة ب عدم توافق البيانات وتوليد تقارير طبية تلقائية داعمة للقرار السريري		
المنهجية العامة للعمل	1. دراسة المشكلة وسیر العمل السريري لتشخيص سرطان الثدي. 2. تحليل مجموعات البيانات المستخدمة وخصائصها الإحصائية. 3. تصميم بنية نموذج ذكاء اصطناعي متعددة المدخلات والمخرجات. 4. تنفيذ نماذج التعلم العميق وطبقات الدمج والانتباه. 5. تدريب النماذج وتقدير الأداء باستخدام مقاييس مناسبة. 6. تحليل النتائج وتوثيقها علمياً.		
الأدوار والمسؤوليات	الاسم	الدور	المسؤولية
	الدكتورة ماجدة البكور	مشرف	الإشراف الأكاديمي والتوجيه
	المهندسة آية الأسود	مشرف	الإشراف العملي والتوجيه
	الطالب ينال الشوفي	مهندس	بناء النموذج وتدريبه وتقديره

3.3 وثيقة نطاق العمل

وثيقة نطاق العمل (Statement of Work - SOW) هي وثيقة شاملة تحدد نطاق العمل الخاص بالمشروع. توضح هذه الوثيقة المهام المحددة، والمخرجات المتوقعة، والجدول الزمني، والمسؤوليات. كما توفر فهماً واضحاً لما يجب إنجازه، وأهداف المشروع، ومعايير النجاح.

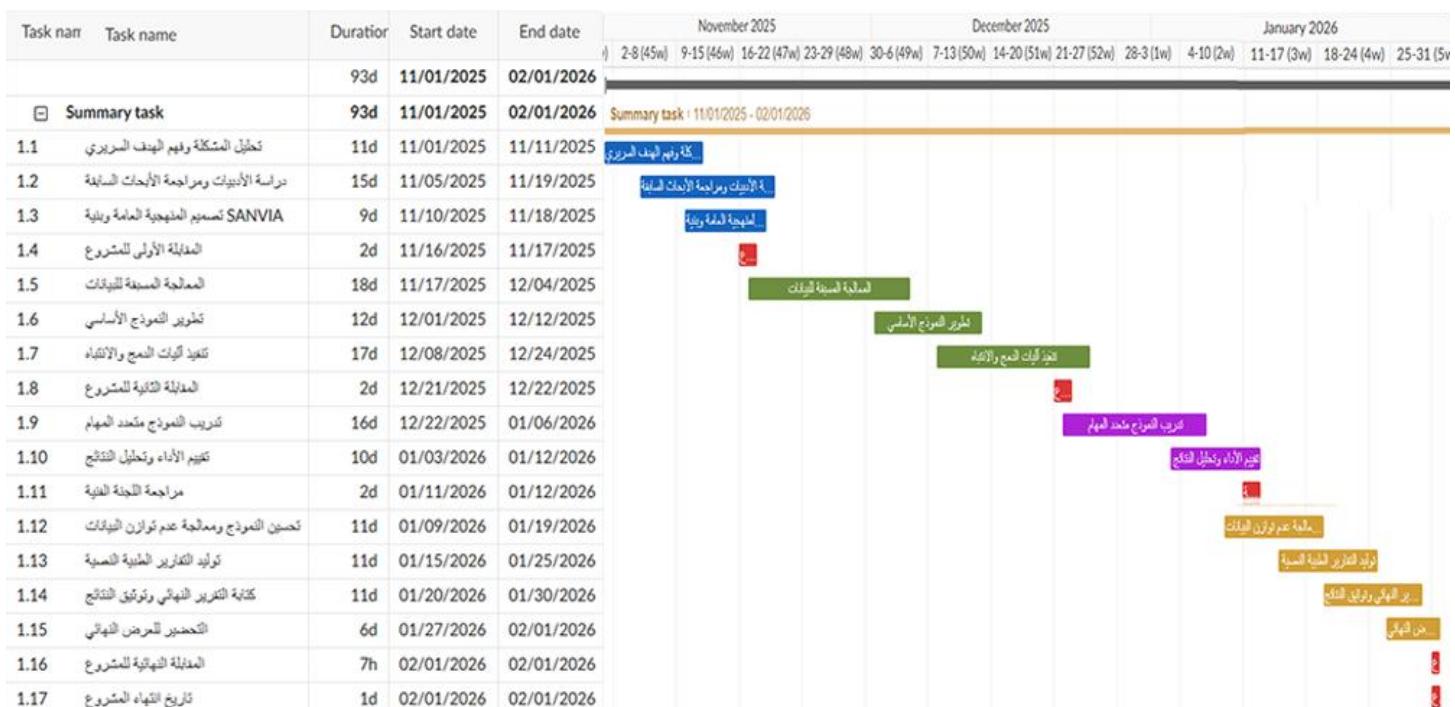
الجدول 2.3: وثيقة نطاق العمل

عنوان المشروع	Sanvia : AI for Early Hope
توصيف المشروع	يعد مشروع Sanvia نظام تشخيص آلی لسرطان الثدي يعتمد على تقنيات الذكاء الاصطناعي والتعلم العميق، وذلك من خلال تحليل صور التصوير الشعاعي للثدي (Mammography) من عدة مناظر تصويرية مختلفة. يركز المشروع على تصميم نموذج متعدد المدخلات ومتعدد المهام قادر على تصنیف حالات سرطان الثدي وفق نظام BI-RADS، وتحديد كثافة الثدي، بالإضافة إلى توليد تقارير طبية نصية تلقائية لدعم القرار السريري. يهدف المشروع إلى تقديم إطار عمل ذكي وموثوق يمكن أن يساهم في تحسين دقة وسرعة التشخيص، ودعم الأخصائيين الشعاعيين في الممارسة السريرية.
نطاق المشروع	يتضمن هذا المشروع تطوير نظام ذكاء اصطناعي بحثي يعتمد على الشبكات العصبية العميقه لمعالجة صور الماموغرام من أربعة مناظر تصويرية مختلفة باستخدام نموذج تعلم عميق متعدد المدخلات مبني على بنية EfficientNet كأساس ودمج آليات انتباه ودمج متقدم لتمثيل المعلومات متعددة المناظر. بالإضافة للتعامل مع مشكلة عدم توازن توزيع الفئات في البيانات الطبيعية. ومن ثم تقييم النموذج باستخدام مقاييس أداء مناسبة للتطبيقات الطبية. يركز المشروع على الجانب البحثي والتجريبي في هندسة الذكاء الاصطناعي، دون تطوير نظام برمجي تشغيلي متكامل للاستخدام التجاري.
أهداف المشروع	<ul style="list-style-type: none"> • بناء نموذج تعلم عميق لتشخيص سرطان الثدي من صور الماموغرام. • تحقيق أداء عالي في تصنیف نظام BI-RADS وتصنیف كثافة الثدي. • معالجة مشكلة عدم توازن البيانات، خصوصاً الفئات النادرة. • دمج تقنيات الانتباه وطبقات الدمج لتحسين تمثيل البيانات متعددة المناظر. • توليد تقارير طبية نصية تلقائية باستخدام نماذج اللغة. • توثيق المنهجية والنتائج وفق المعايير الأكademie لمشاريع هندسة الذكاء الاصطناعي
مخرجات المشروع	<ul style="list-style-type: none"> • نموذج ذكاء اصطناعي مدرب ومؤدي. • نتائج تجريبية ومقاييس أداء موثقة وتقارير طبية مولدة. • التقرير النهائي (الرسالة العلمية).
التكنولوجيا والأدوات	<p>لغة البرمجة: Python أطر عمل التعلم العميق: TensorFlow, Keras مكتبات معالجة الصور: OpenCV, Albumentations مكتبات تحليل البيانات: NumPy, Pandas, Scikit-learn أدوات التصور البصري: Matplotlib نماذج الذكاء الاصطناعي: EfficientNetB4, Chat GPT4o الأدوات: Google Colab Pro+, Git, GitHub</p>
الافتراضات	<ul style="list-style-type: none"> • توفر بيئة حوسبة سحابية مناسبة لتدريب النماذج العميقه.

	<ul style="list-style-type: none"> • إمكانية الوصول إلى مجموعة بيانات Vindr-Mammo. • استقرار المكتبات البرمجية المستخدمة أثناء فترة التنفيذ. • توفر الإشراف الأكاديمي طوال فترة العمل على المشروع.
موارد المشروع	الموارد البشرية
	المدير المشروع
	مشرف
	مطور نموذج الذكاء الاصطناعي ومعالجة البيانات والتجارب
الجدول الزمني	<ul style="list-style-type: none"> • تاريخ بدء المشروع: 2025/11/1 • المقابلة الأولى للمشروع: 2025/11/16 • المقابلة الثانية للمشروع: 2025/12/21 • مراجعة اللجنة الفنية: 2026/1/11 • الندوة النهائية للمشروع: 2026/2/1 • تاريخ انتهاء المشروع: 2026/2/1

4.3 خطة المشروع – مخطط غانت

توضح خطة المشروع تسلسل المراحل الزمنية الرئيسية لتطوير النظام عبر مخطط غانت، بدءاً من مرحلة دراسة المشكلة والبيانات، مروراً بتصميم النموذج والتدريب والتقييم، وانتهاءً بتوثيق النتائج وتقديم المشروع النهائي.



الشكل 1.3: مخطط غانت

5.3 إدارة المخاطر

تم تحديد مجموعة من المخاطر المحتملة التي قد تؤثر على تنفيذ المشروع البحثي.

الجدول 3.3: جدول المخاطر المحتملة

رمز الخطير	وصف الخطير	التأثير	خطة التخفيف
RK-01	عدم توازن البيانات	عالٍ	استخدام استراتيجيات معالجة الاختلال
RK-02	تعقيد النموذج	متوسط	تبسيط البنية تدريجياً
RK-03	متطلبات حوسية عالية	عالٍ	الاعتماد على بيئة سحابية
RK-04	صعوبة التفسير السريري	متوسط	دمج التوليد النصي للتقارير

6.3 إدارة النماذج والتجارب

1. وصف بيئة التنفيذ

تم تنفيذ جميع مراحل التدريب والتقييم ضمن بيئة حوسية سحابية نظرًا لحجم البيانات وتعقيد النماذج.

2. التحكم بالإصدارات

تم استخدام Git و GitHub لإدارة الشيفرة المصدرية، تتبع التعديلات، وضمان إمكانية الرجوع إلى الإصدارات السابقة أثناء التجارب.

3. تتبع التجارب

- تم اعتماد أدوات تتبع لتسجيل أداء النماذج، بما يشمل:
- قيمة الخسارة
- الدقة

4. حفظ النماذج

تم حفظ النماذج المدربة بصيغ معيارية تتيح إعادة الاستخدام وإعادة التقييم.

7.3 خطة العمل

تقوم خطة العمل المقترحة على تصميم نظام متعدد المدخلات ومتعدد المخرجات يُعرف باسم Sanvia، يهدف إلى محاكاة سير العمل السريري في تشخيص سرطان الثدي، من خلال الدمج بين التحليل البصري لصور الماموغرام والتوليد النصي للتقارير الطبية.

يتكون النظام من ثلاثة طبقات رئيسية:

- طبقة المدخلات
- النواة التحليلية (Backbone)
- طبقة المخرجات

(يُستكمل التفصيل في الفصل المنهجي اللاحق)

8.3 الخلاصة

يُظهر هذا الفصل أهمية التنظيم والإدارة المنهجية في تنفيذ مشاريع هندسة الذكاء الاصطناعي، خاصة في المجالات الطبية الحساسة. إن الالتزام بخطة واضحة لإدارة المشروع، وإدارة المخاطر، وتتبع النماذج والتجارب، يسهم بشكل مباشر في موثوقية النتائج وقابلية تطبيق النظام المقترن في السياق السريري

الفصل الرابع

المنهجية

الفصل الرابع - المنهجية

1.4 مقدمة

يقدم هذا الفصل المنهجية التفصيلية المتبعة في تصميم وتنفيذ نظام SANVIA يبدأ بمراجعة تقنية للمفاهيم الأساسية والتكنيات المستخدمة، ثم يصف مجموعة البيانات وخطوات المعالجة المسبقة، ويختتم بالبنية المعمارية الكاملة للنظام ومقاييس التقييم. تركز المنهجية على الجوانب الهندسية الدقيقة التي تميز SANVIA عن الأنظمة الموجودة، مع التركيز على قابلية التطبيق العملي والتكرار.

2.4 مراجعة تقنية للأدبيات

تنوعت الجهود البحثية في مجال الذكاء الاصطناعي لتشخيص سرطان الثدي عبر التصوير الشعاعي، الشبكات العصبية التلaffيفية، ومحولات الانتباه، إضافة إلى توليد التقارير السريرية:

- **الشبكات العصبية التلaffيفية :** استخدمت العديد من الدراسات CNNs لاستخراج السمات الهرمية من صور ماموغرام، حيث تمكنت الطبقات الضحلة من التعرف على السمات البسيطة مثل الحواف والألوان، في حين اكتشفت الطبقات العميقه الأنماط المعقدة والتشوهات الدقيقة التي قد تقوت العين البشرية. على سبيل المثال، اعتمدت دراسة [10] شبكة CD-CNN لتصنيف كثافة الثدي مع طبقة ELM، محققة دقة 95.4% على البيانات الداخلية و 73.9% على مجموعة mini-MIAS الخارجية، ما أظهر قدرة CNNs على التعلم حتى في البيئات محدودة الموارد.
- **آليات الانتباه :** استخدمت بعض الدراسات Self-Attention لحساب العلاقات بين جميع عناصر المدخلات، و Cross-Attention لدمج المناظر المختلفة للثدي CC و MLO على سبيل المثال، طورت دراسة [16] استراتيجيتين لدمج مناظير CC و MLO، محققة AUC يصل إلى 0.9051 باستخدام VGG19 و 0.8042 باستخدام ResNet50، مع ثبات أفضل للاستراتيجية عبر DBE لمجموعات بيانات متعددة.
- **التعلم بالنقل وإعادة الضبط:** استفادت الدراسات من النماذج المدربة مسبقاً مثل ImageNet لتسريع التعلم على مجموعات بيانات ماموغرام الأصغر. وقد استخدمت دراسة [11] شبكة متعددة المهام لاستخراج الآفات و EfficientNet-B0 للتصنيف النهائي، مسجلة AUC تصل إلى 0.995 على التحقق الداخلي و 0.933-0.947 على الاختبار الخارجي، مع خفض كبير في الإيجابيات الكاذبة.
- **توليد التقارير السريرية :** ركزت بعض الدراسات على أتمتة التقارير الطبية باستخدام نماذج اللغة، مثل دراسة [12] التي استخدمت نموذج ClinicalBLIP المبني على InstructBLIP ، محققة قيم METEOR تصل إلى 0.570 و ROUGE-L إلى 0.534 على مجموعة IU X-RAY ، مما أبرز إمكانية توليد نصوص سريرية دقيقة وآلية.
- **نماذج أساسية متعددة الوسائط :** قامت دراسة [13] بتطوير Mammo-CLIP المدرب على أزواج صور-تقارير، محققاً AUC يصل إلى 0.91 في تصنیف السرطان بعد ضبط دقيق كامل،

و 0.88 في تصنيف الكتل، مع إمكانية توطين السمات دون الحاجة لصناديق إحاطة، ما يمثل خطوة نحو نماذج أكثر مرونة واستقلالية.

3.4 الفجوات على المستويين المعرفي والتطبيقي

تشير الأدبيات الحالية إلى وجود فجوات على مستوىين رئيسيين في مجال الذكاء الاصطناعي لتشخيص سرطان الثدي عبر ماموغرام.

الفجوة المعرفية: هناك نقص في الدراسات التي تستكشف العلاقة التفاعلية بين المناظر الأربع القياسية للثدي R_MLO ، L_MLO ، R_CC ، L_CC باستخدام آليات الانتباه المتقدمة. معظم الأبحاث تعامل مع كل منظر بشكل منفصل أو تلألأ إلى طرق دمج بسيطة مثل التجميع أو المتوسط الحسابي، مما يقلل من القدرة على الاستفادة الكاملة من المعلومات المتكاملة بين المناظر المختلفة.

الفجوة التطبيقية: فهناك غياب لأنظمة الشاملة التي تجمع بين جميع عناصر سير العمل السريري: التشخيص الآلي، التقسيم البصري، وتوليد التقارير الطبية المفصلة. معظم النماذج الحالية تقتصر على مرحلة واحدة، مثل تصنيف BI-RADS أو تقييم كثافة الثدي، دون ربطها بتوليد تقرير سريري يمكن استخدامه مباشرة في الممارسة الطبية، مما يحد من فاعلية التطبيقات العملية في البيئات السريرية.

4.4 منهجية العمل

تتبع منهجية هذا البحث النهج التجاري مع تصميم هندسي تكراري. تم العمل حسب الخطوات الرئيسية التالية:

1. جمع وتحضير البيانات : تنظيف، توحيد، وتوزن
2. بناء النموذج : تصميم البنية المعمارية مع آليات الدمج
3. التدريب : استراتيجيات متقدمة لمعالجة الاختلال
4. التقييم : مقاييس شاملة وتحليل الأخطاء
5. توليد التقارير : تكامل LLMs API

5.4 مجموعة المعطيات المستخدمة

Vindr-Mammo Dataset 1.5.4

تم اختيار مجموعة بيانات [15] Vindr-Mammo كمصدر أساسى للبيانات نظرًا لخصائصها المتميزة:

الجدول 4.4: خصائص مجموعة بيانات Vindr-Mammo

التفاصيل	الخاصية
(Bach Mai Hospital) مستشفى فيتنامي كبير	المصدر
2020-2018	الفترة
20,000 صورة رقمية	عدد الصور
5,000 دراسة فريدة	عدد الدراسات
100 % نساء	التوزيع الجنسي
70-40 سنة (متوسط 52.3)	الفئة العمرية
متغيرة (تقريبا 3024 × 2016)	الأبعاد
DICOM	التنسيق
BI-RADS, Density, Findings (نصية)	التسميات

2.5.4 خصائص التسميات

في تحليل صور ماموغرام، تُصنف الحالات عادةً وفق معيار BI-RADS وكثافة النسيج الثدي Breast Density، وهو عنصران أساسيان لفهم تحديات النماذج الذكية ودقتها:

- **تصنيف BI-RADS:** يحدد مستوى الشك في وجود أورام سرطانية ويقسم الحالات إلى خمس فئات، مع توزيع غير متساوٍ للغاية، مما يمثل تحديًّا لأنظمة الذكية بسبب الاختلال الكبير في البيانات:
 - 1 سلبي: يمثل 67.5% من الحالات، أي أن غالبية الصور لا تحتوي على أي مؤشرات مرضية.
 - 2 حميد: 28.3% من الحالات، تشمل التغيرات غير السرطانية مثل الأكياس أو التكلسات البسيطة.
 - 3 غالباً حميد: 3.9%， تشير إلى تغييرات صغيرة تحتاج إلى متابعة دورية، مع احتمال ضئيل للإصابة بالسرطان.
 - 4 مشتبه: 0.27% فقط من الحالات، تمثل فئة حرجة تتطلب إجراء خزعة للتأكد من التشخيص.
 - 5 شديدة الاشتباه: 0.03%， حالات نادرة جدًا لكنها حرجة، تمثل أعلى مستوى من الشك السرطاني.

هذا التوزيع يوضح اختلال الفئات، حيث تمثل الفئات الحرجة الأقل تكراراً تحدياً كبيراً لنماذج التعلم الآلي، إذ تميل الشبكات العصبية إلى التنبؤ بالفئات الأكثر تكراراً، ما قد يؤدي إلى تقوية الحالات السرطانية النادرة.

- **كثافة الثدي:** تصف نسبة النسيج الغدي مقابل النسيج الدهني في الثدي، وهو عامل مهم لأن كثافة الثدي العالية قد تخفي الأورام الصغيرة، مما يزيد صعوبة التشخيص:
 - **A دهني بالكامل:** 12%， نسيج رقيق يسمح ببرؤية واضحة للكتل المحتملة.
 - **B نسيج غدي/دهني متناثر:** 38%， معظم الصور واضحة ولكن بعض المناطق قد تتطلب انتباهاً إضافياً.
 - **C غير موحد:** 41%， يضم نسيج غدي كثيف بشكل غير متجانس، مما يجعل كشف الأورام أصعب.
 - **D كثيف:** 9%， أكثر الفئات تحدياً للتشخيص البصري والآلي، حيث يمكن أن تخفي الكتل الصغيرة داخل النسيج الكثيف.

3.5.4 التحديات في البيانات

تواجه مجموعات بيانات التصوير الشعاعي للثدي عدة تحديات جوهرية تؤثر على فعالية النماذج الذكية ودققتها في التشخيص:

- **الاختلال الشديد في الفئات:**

الفئات الحرجة مثل 4 BI-RADS (مشتبه) و 5 BI-RADS (شديدة الاشتباه) تمثل أقل من 0.3% من جميع العينات مجتمعة، بينما تشكل الفئات الطبيعية والحميدة غالبية البيانات. هذا الاختلال يجعل النماذج التقليدية متحيزة نحو الفئات الأكثر شيوعاً، وبضعف قدرتها على كشف الحالات الحرجة التي تتطلب تدخلاً سريعاً.

• التسميات النصية:

في بعض قواعد البيانات، يتم تسجيل تصنيف Findings BI-RADS في حقل نصي مثل بدلاً من حقل رقمي جاهز، مما يتطلب معالجة نصية مسبقة لاستخراج التسمية بشكل دقيق، غالباً باستخدام تقنيات Regex أو معالجة اللغة الطبيعية. أي خطأ في هذه المرحلة قد يؤدي إلى أخطاء في التسمية التدريبية و يؤثر على أداء النموذج.

- **المناظر المفقودة:**

قد لا تحتوي بعض الدراسات على كل المناظر الأربع القياسية L_CC, L_MLO, R_CC, R_MLO هذا النص يمثل تحدياً في دمج المعلومات بين المناظر المختلفة، حيث تعتمد العديد من النماذج على وجود جميع المناظر للحصول على نتائج دقيقة.

- **تبالين جودة الصور:**

تختلف جودة الصور بشكل كبير بين الأجهزة والمرافق الطبية، سواء من حيث الإضاءة، التباين، أو دقة الاستشعار. هذا التباين يؤدي إلى صعوبة في تدريب النماذج، إذ أن النموذج قد يتعلم خصائص غير متعلقة بالمرض بدلاً من السمات التشخيصية الحقيقية.

6.4 تحضير مجموعة المعطيات

1.6.4 تنظيف البيانات

- استخراج تصنيف BI-RADS من النصوص: استخدام Regex لاستخراج الرقم من نصوص مثل "BI-RADS 3".

- تحويل كثافة الثدي (Density) إلى ترميز رقمي:

A	→	0
B	→	1
C	→	2
D	→	3

- معالجة القيم غير الصالحة: استبدال القيم مثل 'N/A' بالوسيط (Median) للفئة المعنية.

- إزالة القيم غير المنطقية أو الناقصة: الدراسات التي لا تحتوي على صور. او تصنيفات BI-RADS خارج النطاق 1–5. او أعمار خارج المدى المنطقي 20–100 سنة.

2.6.4 دمج البيانات الجدولية

- استخراج العمر: يتم الحصول على العمر من Metadata لملفات DICOM أو سجلات المرضى.

- تقييس العمر (Normalization) : استخراج القيم إلى نطاق محدد MinMaxScaler (1.4)

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1.4)$$

حيث:

- x : هو القيمة الأصلية للمتحول أو الخاصية التي نريد تقييسها.
- x_{\min} : هو أصغر قيمة للمتحول x في مجموعة البيانات.
- x_{\max} : هو أكبر قيمة للمتحول x في مجموعة البيانات.
- x_{norm} : هو القيمة بعد التقييس، بحيث تصبح جميع القيم في النطاق [0, 1].

View Mapping 3.6.4

1. تجميع المناظر لكل دراسة: لكل مريض، يتم تجميع أربعة مناظر أساسية:

- L_CC: Left Cranio-Caudal •
- L_MLO: Left Medio-Lateral Oblique •
- R_CC: Right Cranio-Caudal •
- R_MLO: Right Medio-Lateral Oblique •

2. معالجة المناظر المفقودة: إذا كان أحد المناظر مفقوداً:

- يتم استخدام Placeholder أسود بنفس أبعاد الصورة (قيمة البكسل = 0).
- إضافة قناة Binary Mask لكل منظر، تشير إلى وجود المنظر (0) أو غيابه (1).

4.6.4 معالجة عدم توازن الفئات

تعد مشكلة اختلال توزيع الفئات من أبرز التحديات فيمجموعات بيانات ماموغرام، حيث تمثل الفئات الحرجية مثل BI-RADS 4 و 5 نسبة ضئيلة للغاية من البيانات، مما يجعل النماذج تميل إلى التحييز نحو الفئات الأكثر تكراراً مثل BI-RADS 1 و 2. لمعالجة هذه المشكلة، تم تطبيق استراتيجيتين رئيسيتين Effective Number Weighting و Oversampling للفئات النادرة.

Effective Number Weighting •

تعتمد هذه الطريقة على حساب وزن كل فئة وفقاً لعدد العينات المتاحة لها بطريقة تقلل من تأثير الفئات الكبيرة وتزيد من أهمية الفئات الصغيرة. الصيغة المستخدمة المعادلة (2.4).

$$w_i = \frac{1 - \beta}{1 - \beta^{n_i}} \quad (2.4)$$

حيث:

- n_i هو عدد العينات في الفئة i
- $\beta = 0.9$ هو معامل ضبط، تم اختياره في هذه الدراسة

الوزن الناتج لكل فئة يتم دمجه مع دالة الخسارة (Loss Function) لتقوية تأثير العينات النادرة خلال عملية التدريب، مما يساعد النموذج على التركيز على الفئات الحرجية وتجنب التحيز نحو الفئات الأكثر شيوعاً.

• للفئات النادرة Oversampling

نظرًا للتمثيل الضئيل جداً لفئة BI-RADS 4 و5، تم استخدام تقنية Oversampling هندسي مدرومة بالـ Augmentation لزيادة عدد العينات لكل فئة إلى مستوى موحد تقريرًا (حوالي 600 عينة لكل فئة). التفاصيل:

- من 54 عينة فقط إلى 600 عينة، أي زيادة بمقدار حوالي 11 ضعف.
- من 15 عينة فقط إلى 600 عينة، أي زيادة بمقدار 40 ضعف.
- تم زراعة عدد العينات إلى 600 عينة لتحقيق توازن أفضل مع الفئات الأكثر.
- لم يتم تغيير عدد العينات، ولكن النسبة المئوية للفئات الجديدة تغيرت قليلاً نتيجة زيادة العينات النادرة.

الجدول 5.4 BI-RADS Oversampling :

الفئة	العدد الأصلي	النسبة الأصلية	Oversampling	العدد بعد التوزيع	النسبة الجديدة
BI-RADS 1	3,375	67.5%		3,375	56.3%
BI-RADS 2	1,415	28.3%		1,415	23.6%
BI-RADS 3	195	3.9%		600	10.0%
BI-RADS 4	54	0.27%		600	10.0%
BI-RADS 5	15	0.03%		600	10.0%
المجموع	5,054	100%		5,990	100%

• تطبيق تقنيات augmentation

لتحقيق Oversampling هندي فعال، تم استخدام مكتبة Albumentation لتوليد عينات جديدة من الصور الأصلية بطريقة عشوائية ومتعددة، مع احتمال تطبيق التعديلات بنسبة 85%. العمليات المستخدمة شملت:

- تحويلات هندسية:
 - تدوير الصورة بزاوية 15^{\pm} °
 - ShiftScaleRotate لتعديل الموضع والحجم بنسبة 10^{\pm} %
- تحسينات لونية:
 - RandomBrightnessContrast لضبط السطوع والتباين ضمن حدود 20^{\pm} %
 - CLAHE لتوضيح التباين المحلي، $clip_limit=4.0$
- تأثيرات الضبابية:
 - GaussianBlur بحد أقصى 3 بكسل
 - MotionBlur بحد أقصى 3 بكسل
- تشوهات هندسية:
 - GridDistortion لتشويه الشبكة بنسبة تصل إلى 15%
- إضافة ضوضاء:
 - GaussNoise مع نطاق تباين 10–50

تهدف هذه العمليات إلى زيادة تنوع البيانات، تحسين قدرة النموذج على التعلم، والتقليل من الإفراط في التكيف مع عينات محدودة، مع الحفاظ على السمات التشخيصية الهامة للثدي.

7.4 البيئة التقنية والتنفيذية

1.7.4 الأجهزة والبنية التحتية

تم تنفيذ جميع مراحل المعالجة والتدريب باستخدام بيئة حوسية سحابية، نظرًا لحجم البيانات ومتطلبات التدريب العميق:

- منصة التنفيذ:**

تم الاعتماد على Google Colab Pro+، خلال جلسات التدريب طويلة المدى، لما توفره من وقت تشغيل أطول واستقرار أعلى مقارنة بالنسخة المجانية.

- وحدة المعالجة الرسومية:**

تم استخدام NVIDIA A100 بذاكرة فيديو 80 GB VRAM، وهي مناسبة لتدريب نماذج الشبكات العصبية الالتفافية (CNNs) متعددة المدخلات، خاصة عند التعامل مع صور طبية عالية الدقة.

- الذاكرة العشوائية:**

تم توفير 250 GB RAM ضمن بيئة Colab Pro+، مما ساعد في تحميلمجموعات البيانات، وتنفيذ عمليات augmentation ، ودمج البيانات الجدولية دون اختلافات في الذاكرة.

- التخزين:**

تم استخدام Google Drive بسعة تخزين تقارب 15 GB لحفظ:

- صور ماموغرام
- ملفات CSV
- النماذج المدربة
- نتائج التجارب والتقارير النهائية

كما ساهم التكامل المباشر بين Google Drive وColab في تسهيل إدارة البيانات واستمرارية العمل بين الجلسات

2.7.4 البرمجيات والمكتبات

تم اختيار البرمجيات والمكتبات بعناية لضمان الأداء، التوافق، ودعم أحدث تقنيات التعلم العميق. يوضح الجدول التالي المكونات الرئيسية المستخدمة:

الجدول 6.4: البرمجيات والمكتبات

المكون	الإصدار	الاستخدام
Python	3.12	لغة البرمجة الأساسية
TensorFlow	2.19	إطار العمل الرئيسي لبناء وتدريب نماذج Deep Learning
Keras	3.0	واجهة عالية المستوى لتصميم النماذج
Albumentations	1.3	تطبيق Augmentation متقدم على الصور
OpenCV	4.8	المعالجة المسربة للصور
Scikit-learn	1.3	حساب المقاييس الإحصائية ومقاييس الأداء
NumPy	1.24	العمليات العددية والمصفوفية
Pandas	2.0	معالجة وتنظيم البيانات الجدولية
Matplotlib	3.7	التصور البياني للنتائج
ReportLab	3.6	توليد ملفات PDF للتقارير
OpenAI API	Latest	استخدام GPT-4o لتوليد التقرير

3.7.4 إدارة النماذج والتجارب

لضمان التنظيم، التتبع، وإمكانية إعادة تنفيذ التجارب، تم اعتماد مجموعة من الممارسات القياسية في إدارة النماذج:

- التحكم بالإصدارات

تم استخدام Git مع GitHub لإدارة الشيفرة المصدرية، تتبع التغييرات، والرجوع إلى الإصدارات السابقة عند الحاجة.

- تتبع التجارب:

TensorBoard لمراقبة

▪ قيم الخسارة (Loss)

- الدقة (Accuracy)
- F1-score
- تطور الأداء عبر العصور (Epochs)
- CSV Logging لحفظ نتائج كل تجربة بشكل منظم وقابل للتحليل لاحقاً.

• حفظ النماذج:

تم حفظ النماذج النهائية بصيغة (.h5), لما توفره من توافق واسع وسهولة في التحميل وإعادة الاستخدام.

• Checkpointing :

تم تفعيل آلية Model Checkpointing لحفظ أفضل نموذج تلقائياً اعتماداً على أفضل قيمة لـ validation F1-score، مما يضمن الاحتفاظ بالنموذج الأكثر توافرًا بين الدقة والاستدعاة، خصوصاً في ظل اختلال توزيع الفئات.

8.4 المنهجية المقترحة

1.8.4 النظرة العامة

تعتمد المنهجية المقترحة على تصميم نظام متعدد المدخلات ومتعدد المخرجات يُعرف باسم Sanvia، يهدف إلى محاكاة سير العمل السريري في تحليل صور الماموغرام، من خلال الدمج بين التحليل البصري، البيانات السريرية، والتوليد النصي للتقارير الطبية.

يتكون النظام من ثلاثة طبقات رئيسية: المدخلات، النواة التحليلية، والمخرجات، كما يلي:

1. المدخلات:

• صور ماموغرام

أربع صور لكل دراسة تمثل المناظر القياسية:

- L_CC (Left Cranio-Caudal)
- L_MLO (Left Medio-Lateral Oblique)
- R_CC (Right Cranio-Caudal)
- R_MLO (Right Medio-Lateral Oblique)

يتم توحيد أبعاد جميع الصور إلى $(3 \times 768 \times 768)$ لضمان التوافق مع بنية الشبكة العصبية المستخدمة، مع الحفاظ على المعلومات التشريحية الأساسية.

• البيانات الجدولية:

تشمل:

- العمر: قيمة عدبية واحدة تمثل عمر المريضة بعد التطبيع.
- علامات القيم المفقودة: أربع قيم ثنائية تشير إلى وجود أو غياب كل منظر من مناظر التصوير ونقص أو وجود معلومات العمر، مما يسمح للنموذج بأخذ نقص المعلومات بعين الاعتبار أثناء التنبؤ.

2. المخرجات:

• :BI-RADS

منتج احتمالات مكون من 5 فئات يتم توليده باستخدام دالة Softmax، بحيث يعكس درجة الاشتباه السريري لكل دراسة.

• :تصنيف كثافة الثدي (Breast Density)

منتج احتمالات من 4 فئات (ACR A–D)، أيضًا باستخدام Softmax، نظرًا لأهمية الكثافة في تفسير صور الماموغرام وتقدير المخاطر.

• تقرير طبي نصي:

يتم توليده باستخدام واجهة برمجية لممادج اللغة الكبيرة (LLM API)، ويعتمد على مخرجات التصنيف والتفسير البصري لتقديم وصف نصي قريب من التقارير السريرية الحقيقية.

2.8.4 الـ Backbone (EfficientNetB4)

يعتمد النظام في استخلاص السمات البصرية على نموذج EfficientNetB4 كعمود فقري، نظرًا لتوازنه العالي بين الأداء والدقة الحسابية، وهو أمر بالغ الأهمية في التطبيقات الطبية.

تعتمد EfficientNet على ما يُعرف بـ Compound Scaling، وهي استراتيجية توسيعة متوازنة تقوم بتكبير النموذج عبر ثلاثة أبعاد في آنٍ واحد:

- **العمق:** زيادة عدد الطبقات لتمكين النموذج من تعلم تمثيلات أكثر تعقيداً.
- **العرض:** زيادة عدد القنوات داخل كل طبقة لتحسين القدرة التمثيلية.
- **دقة المدخلات:** استخدام صور بدقة أعلى للتقاط التفاصيل الدقيقة، وهي ميزة حاسمة في الماموغرام.

يُظهر هذا النهج تفوقاً واضحاً مقارنة بالتوسيعة أحادية البعد، حيث يحقق أداء أعلى باستخدام عدد أقل من المعاملات.

نظرًا لمحدودية البيانات الطبيعية مقارنة بمجموعات البيانات العامة، تم اعتماد التعلم بالنقل للاستفادة من المعرفة المكتسبة مسبقاً من التدريب على مجموعات بيانات واسعة مثل ImageNet. تم تطبيق الاستراتيجية على النحو التالي:

- **تجميد 70% من الطبقات الأولى في النموذج،** حيث تُستخدم هذه الطبقات لاستخراج سمات بصرية عامة مثل الحواف، القوام، والأنماط منخفضة المستوى.
- **الطبقات المجمدة:**

blocks 1–4 تُستخدم لاستخراج السمات العامة دون تعديل أو زانها، مما يقلل من خطر فرط التعلم.

- **الطبقات القابلة للتدريب (Trainable Layers):**
blocks 5–7 بالإضافة إلى الطبقات العليا (Top Layers)

يتم إعادة تدريب هذه الطبقات لتعلم سمات متخصصة تتناسب مع خصائص صور الماموغرام، مثل التكاليف الدقيقة والتشوهات البنوية. تسمح هذه المقاربة بتحقيق توازن بين الاستفادة من المعرفة المسبقة والتكييف مع طبيعة البيانات الطبيعية المتخصصة.

EfficientCrossAttention 3.8.4

تُعد EfficientCrossAttention آلية انتباه مخصصة لدمج المناظر المتعددة في صور التصوير الشعاعي للثدي، وتهدف إلى استغلال العلاقات التكميلية بين منظوري CC وMLO بطريقة أكثر كفاءة ومرنة مقارنة بآليات الانتباه التقليدية. تعتمد الآلية على Multi-Head Cross-Attention

بعد 8 رؤوس انتباه، حيث يتعلم كل رأس نمطًا مختلفًا من الارتباطات المكانية بين المناظر المختلفة، مما يسمح بتمثيل أكثر ثراءً للسياق التشعري. تم اختيار $Key Dimension = 64$ كبعد متوازن يحقق قدرة كافية على التقاط العلاقات المكانية الدقيقة دون فرض عبء حسابي مرتفع، وهو أمر بالغ الأهمية في معالجة صور الماموغرام عالية الدقة.

وعلى عكس آليات Cross-Attention التقليدية من حيث البنية وطريقة المعالجة والسياق التطبيقي. ففي حين تُصمّم آليات الانتباه التقليدية عادةً لمعالجة سلسلات رسائل رمزية أو تمثيلات مسطحة (Flattened Sequences) باستخدام إسقاطات خطية (Linear Layers)، تعتمد الآلية المقترنة على إسقاطات موضعية باستخدام عمليات 1×1 Convolution تعمل مباشرةً على الخرائط المكانية للصور، مما يحافظ على البنية التشعورية للماموغرام ويمنع فقدان العلاقات المحلية الدقيقة. كما يتم احتساب الانتباه بشكل مكاني موضعى (Spatially Aligned Attention) بين المناظر المختلفة، بدلاً من المقارنة الشاملة بين جميع الواقع، الأمر الذي يؤدي إلى خفض التعقيد الحسابي من سلوك تربيري شائع في آليات الانتباه القياسية إلى تعقيد أقرب إلى الخطى، وهو عامل حاسم عند التعامل مع صور طبية عالية الدقة. إضافةً إلى ذلك، تتميز EfficientCrossAttention بقدرتها الضمنية على التعامل مع المناظر المفقودة دون الحاجة إلى إعادة تصميم بنية النموذج، فضلًا عن تعلم الأهمية النسبية لكل منظور ديناميكيًا وفق السياق السريري للحالة، مما يجعلها أكثر ملاءمة للتطبيقات السريرية الواقعية مقارنة بآليات Cross-Attention العامة.

في هذا التصميم:

- يتم توليد Queries من سمات منظر CC.
- بينما يتم توليد Keys و Values من سمات منظر MLO.

هذا الاختيار يعكس الفرضية الشائعة في الأدبيات بأن كل منظر يحتوي على معلومات تكميلية، وأن أحد المنظرين يمكن استخدامه كمرجع لتقسيم الآخر.

تعطى عملية Cross-Attention بالمعادلة (3.4).

$$\text{CrossAttn}(Q_{CC}, K_{MLO}, V_{MLO}) = \text{softmax} \left(\frac{Q_{CC} K_{MLO}^T}{\sqrt{d_k}} \right) V_{MLO} \quad (3.4)$$

حيث:

- Q_{CC} تمثل الاستعلامات المستخرجة من منظر CC.
- K_{MLO}, V_{MLO} تمثل المفاتيح والقيم المستخرجة من منظر MLO.
- d_k هو بُعد المفاتيح، ويُستخدم للتطبيع ومنع تضخم القيم.

4.8.4 طبقة الاندماج البوابي (GatedFusionLayer)

بعد دمج السمات عبر آلية الانتباه، يتم استخدام GatedFusionLayer للتحكم في تدفق المعلومات القادمة من المناظر المختلفة، بطريقة تحاكي آليات "التركيز والتجاهل" المستخدمة في النظم البيولوجية حيث يتم حساب الاندماج وفق العلاقة (4.4).

لا تفترض هذه الطبقة أن جميع المناظر متساوية في الأهمية، بل تسمح للنموذج بتعلم:

- متى يعتمد على سمات منظر معين بتعزيز المناظر الأكثر إفادة تشخيصياً.
- ومتى يقلل من تأثيره إذا كانت جودته ضعيفة أو معلوماته محدودة أو التعامل مع المناظر المفقودة بشكل غير صريح.

$$F = \sigma(W_{cc}X_{cc}) \odot P_{cc}(X_{cc}) + \sigma(W_{mlo}X_{mlo}) \odot P_{mlo}(X_{mlo}) \quad (4.4)$$

حيث:

- σ هي دالة Sigmoid التي تعمل كبوابة
- W أوزان قابلة للتعلم
- \odot تمثل الضرب عنصر-عنصر
- $P(\cdot)$ إسقاط خططي للسمات إلى فضاء موحد

TabularEncoder 5.8.4

لمعالجة البيانات غير البصرية، مثل العمر ومؤشرات القيمة المفقودة، تم تصميم وحدة مستقلة تُعرف باسم TabularEncoder، تهدف إلى تحويل البيانات الجدولية إلى تمثيل عددي متوافق مع السمات البصرية.

يتكون المُشفر من التسلسل التالي:

Input (5 features) → Dense(256) → BatchNorm → ReLU → Dropout(0.3) →
Dense(128) → BatchNorm → ReLU → Output(128)

شرح المكونات:

- Dense(256) توسيع فضاء التمثيل لالتقاط العلاقات غير الخطية بين العمر والمناظر المفقودة.
- تحسين استقرار التدريب وتسريع التقارب. Batch Normalization
- تقليل فرط التكيف، خاصة بسبب بساطة البيانات الجدولية مقارنة بالصور. Dropout (0.3)
- Dense(128) ضغط التمثيل إلى بعد موحد يسمح بدمجه لاحقاً مع السمات البصرية المستخرجة من الشبكة الالتفافية.

بهذا التصميم، تصبح البيانات الجدولية جزءاً فعّالاً من القرار النهائي، بدل أن تكون معلومات ثانوية معزولة عن التحليل البصري.

6.8.4 الرؤوس متعددة المهام (Multi-task Heads)

يعتمد النموذج على التعلم متعدد المهام (Multi-task Learning) من خلال استخدام رأسٍ يصنف مساقتين يشتراكان في تمثيل عميق واحد. يهدف هذا التصميم إلى الاستفادة من الترابط الطبيعي بين تصنيف BI-RADS وتصنيف كثافة الثدي (Breast Density)، حيث إن المهمتين تشاركان في سمات تصويرية وتشريحية متداخلة.

• رأس تصنيف BI-RADS

Input (256 features) → Dense(128) → Dropout(0.4) → Dense(5, activation='softmax')

يُخصص هذا الرأس للتنبؤ بفئة BI-RADS ، والتي تتكون من خمس فئات. يبدأ الرأس من تمثيل مشترك بُعده 256، ثم:

- تمر السمات عبر طبقة كثيفة (Dense) بعدد 128 وحدة لاستخلاص تمثيل تميّزي أكثر تخصيصاً للمهمة.
- تُستخدم طبقة Dropout بنسبة 0.4 للحد من فرط التكيف، خاصةً بسبب عدم توازن الفئات.

- في الطبقة النهائية، تُستخدم دالة Softmax لإخراج توزيع احتمالي على الفئات الخمس، مما يسمح بتفسير المخرجات إحصائياً وسريرياً.

• رأس تصنيف كثافة الثدي (Density)

Input (256 features) → Dense(128) → Dropout(0.4) → Dense(4, activation='softmax')

يتبع رأس الكثافة بنية مماثلة، انطلاقاً من نفس التمثيل المشترك (256 بعدها)، مع اختلاف عدد الفئات:

- طبقة Dense بعد 128 وحدة لاستخلاص سمات مرتبطة بكثافة النسيج.
- Dropout بنسبة 0.4 لتعزيز التعميم.
- طبقة إخراج نهائية بدالة Softmax تعطي توزيع احتمالات على أربع فئات تمثل تصنيفات ACR للكثافة.

يسمح هذا الفصل بين الرؤوس بتعلم تمثيلات متخصصة لكل مهمة، مع الحفاظ على قاعدة معرفية مشتركة في الطبقات السابقة.

7.8.4 معمارية النموذج الكلية

البنية الكاملة للنموذج طبقةً بطبقة، بدءاً من المدخلات الصورية وانتهاءً بالمخرجات متعددة المهام.

• طبقة الإدخال

تستقبل أربع صور ماموغرافي لكل دراسة، تمثل المناظر القياسية الأربع، مما يتيح للنموذج الاطلاع على السياق التشريحي الكامل.

EfficientNetB4 (Backbone) •

تُستخدم هذه الشبكة الالتقافية لاستخراج السمات العميقة من كل منظر على حدة، منتجة خرائط سمات غنية بتمثيلات مكانية عالية المستوى، مع عدد معاملات معتدل نسبياً.

Global Average Pooling •

تقلل الأبعاد المكانية لخرائط السمات وتحولها إلى متجهات مدمجة، مما يقلل الحمل الحسابي ويحد من الحساسية للموقع الدقيق للافات.

• طبقة Cross-Attention .

تُستخدم لدمج المعلومات بين المناظر المتناظرة MLO و CC، حيث تتعلم العلاقات التكميلية بين المنظرين بدل التعامل معهما كمدخلات مستقلة.

• Gated Fusion Layer .

تعمل كآلية تحكم تسمح للنموذج بتوزين مساهمة كل منظر ديناميكياً، مع تقليل تأثير المناظر ذات الجودة الضعيفة أو غير المكتملة.

• دمج الجانبين Concatenate .

يتم دمج تمثيلات الجانبين الأيمن والأيسر في متوجه واحد، مما يتيح الاستفادة من التماثل التسريحي بين الثديين.

• Tabular Encoder .

يعالج البيانات الجدولية (مثل العمر ومؤشرات القيمة المفقودة) ويحوّلها إلى تمثيل عددي قابل للدمج مع السمات البصرية.

• Concatenate Final .

يتم دمج السمات البصرية والجدولية في تمثيل موحد يعكس جميع مصادر المعلومات المتاحة.

• Shared Dense Layer .

تُنتج تمثيلاً مشتركاً بـ 256 بعدة، يستخدم كأساس للتعلم متعدد المهام.

• رؤوس الإخراج .

تقوم كل منها بإجراء التصنيف الخاص بمهمتها باستخدام طبقات مستقلة، مع الحفاظ على الترابط المعرفي عبر التمثيل المشترك.

يبلغ العدد الكلي لمعاملات النموذج حوالي 23.4 مليون معامل، وهو توازن مدروس بين القدرة التمثيلية والكفاءة الحسابية، مما يجعله مناسباً للتدريب على بيانات طبية عالية الدقة مع موارد حوسية محدودة نسبياً.

9.4 مقاييس الأداء

1.9.4 مقاييس التصنيف الأساسية

لضمان دقة وفعالية النموذج، تم استخدام مجموعة من مقاييس الأداء المهمة، وهي كالتالي:

- **الدقة (Accuracy)** : نسبة عدد الحالات التي تم تصنيفها بشكل صحيح إلى العدد الإجمالي للحالات، المعادلة رقم (5,4).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5.4)$$

- **قيمة التنبؤ الإيجابي (Precision)** : تمثل نسبة الإيجابيات الحقيقية إلى مجموع الإيجابيات الحقيقية والإيجابيات الزائفة. تعكس مدى دقة النموذج في التنبؤ بالإيجابيات، المعادلة رقم (6,4).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6.4)$$

- **الاسترجاع (Recall)** : تمثل نسبة الإيجابيات الحقيقة إلى مجموع الإيجابيات الحقيقة والسلبيات الزائفة. تعكس قدرة النموذج على اكتشاف جميع الحالات الإيجابية، المعادلة رقم (7,4).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7.4)$$

- **F1 Score**: تستخدم لتقدير توازن الأداء بين الدقة والاسترجاع، حيث تعتبر الأهم في الحالات التي يكون فيها توزيع الفئات غير متوازن، المعادلة رقم (8,4).

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8.4)$$

حيث:

- الحالات الإيجابية المصنفة بشكل صحيح: TP (True Positive)
- الحالات السلبية المصنفة بشكل صحيح: TN (True Negative)
- حالات سلبية صُنفت خطأ كإيجابية: FP (False Positive)
- حالات إيجابية لم يتم اكتشافها: FN (False Negative)

- مصفوفة الالتباس (Confusion Matrix) : تُظهر توزيع التنبؤات الصحيحة والخاطئة لكل فئة، مما يساعد في فهم أداء النموذج بشكل أفضل وتحديد الأنماط الأخطاء الشائعة.

2.9.4 مقاييس متعددة الفئات

- Macro-Average متوسط غير مر ج لجميع الفئات (يعطي كل فئة وزناً متساوياً)
- Weighted-Average متوسط مر ج بحجم الفئة

10.4 النموذج الأساسي

1.10.4 إحصائيات النموذج

يتميز النموذج الأساسي ببنية متوسطة الحجم نسبياً توازن بين القدرة التمثيلية والكافأة الحسابية. يبلغ إجمالي عدد المعاملات في النموذج حوالي 23.4 مليون معامل، منها 18.7 مليون معامل قابلة للتدريب تمثل ما يقارب 80% من إجمالي المعاملات، في حين تم تجريد 4.7 مليون معامل (20%) ضمن طبقات الاستخراج الأولية. يهدف هذا التجريد الجزئي إلى الاستفادة من السمات العامة المتعلمة مسبقاً مع تقليل مخاطر فرط التعلم. أما حجم النموذج النهائي المحفوظ بصيغة HDF5 فيبلغ حوالي 94 ميغابايت، مما يجعله عملياً من حيث التخزين والنقل وإعادة الاستخدام.

2.10.4 اعدادات التدريب

يعتمد تدريب النموذج على دالة خسارة مركبة معدلة عبر إدخال وزن الفئة w_c مباشرة داخل دالة الخسارة المعادلة 9.4 لدعم التعلم متعدد المهام ومعالجة الاختلال الشديد في توزيع الفئات. تم استخدام Focal Loss معدلة لكل من مهمتي تصنيف BI-RADS وتصنيف كثافة الثدي، حيث تقلل هذه الدالة من تأثير العينات السهلة وتعزز تركيز النموذج على الحالات الصعبة والفئات النادرة ذات الأهمية السريرية.

نظرًا لأن تصنيف BI-RADS يمثل الهدف التشخيصي الأساسي للنظام، فقد تم إعطاؤه وزناً أعلى ضمن الخسارة الكلية، بينما استُخدمت مهمة تصنيف كثافة الثدي كمهمة داعمة بوزن ثانوي يضمن عدم هيمنتها على عملية التعلم.

تعطى دالة الخسارة المستخدمة لكل مهمة بالشكل التالي:

$$\mathcal{L}_{\text{focal}} = -\alpha (1 - p_t)^\gamma w_c \log(p_t) \quad (9.4)$$

حيث:

- p_t هو احتمال التنبؤ الصحيح للفئة الحقيقة،
- γ معامل التركيز الذي يزيد من تأثير العينات الصعبة،
- α معامل الموازنة،
- w_c وزن الفئة لمعالجة عدم توازن البيانات.

أما الخسارة الكلية للنظام متعدد المهام فتم تعریفها كما يلي:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{BI-RADS}} \mathcal{L}_{\text{BI-RADS}} + \lambda_{\text{Density}} \mathcal{L}_{\text{Density}} \quad (10.4)$$

حيث:

$\lambda_{\text{BI-RADS}}$: تمثل معامل وزن

λ_{Density} : تمثل معامل وزن مهمة تصنيف الكثافة.

$\lambda_{\text{BI-RADS}} > \lambda_{\text{Density}}$ لضمان إعطاء أولوية للمهمة التشخيصية الأساسية.

فيما يخص عملية التحسين، تم استخدام AdamW، وهو نسخة محسنة من Adam تضيف مفهوم Weight Decay بشكل صريح، مما يساعد على تحسين التعميم وتقليل تعقيد النموذج الفعلي. تم ضبط معدل التعلم الابتدائي وقيمة Weight Decay تجريبياً للوصول إلى أفضل توازن بين سرعة التقارب واستقرار التدريب.

كما استُخدم جدول تلاشي جيبي لمعدل التعلم (Cosine Decay)، حيث يبدأ معدل التعلم بقيمة ابتدائية مرتفعة نسبياً ثم ينخفض تدريجياً وفق منحنى جيبي مع تقدم خطوات التدريب. تتيح هذه الاستراتيجية استكشافاً أفضل لفضاء الحلول في المراحل الأولى، يليها تقارب أكثر دقة نحو النهاية، مما يقلل من احتمالية الوقوع في حلول محلية ضعيفة

الفصل الخامس

التجارب والنتائج والتقييم

الفصل الخامس - التجارب والنتائج والتقييم

1.5 آلية عمل الاختبار

1.1.5 تقسيم البيانات

تم اعتماد تقسيم على مستوى المريض (Patient-level Split) لضمان تقييم عادل ومنع تسرب البيانات (Data Leakage)، حيث إن وجود صور متعددة للمريض نفسه قد يؤدي إلى تضخيم الأداء بشكل غير واقعي في حال تسربها بين مجموعات التدريب والاختبار. جدول يوضح توزيع البيانات:

الجدول 7.5: جدول تقسيم البيانات

المجموعة	عدد الدراسات	النسبة	الاستخدام
Training	4,000	80%	التدريب + Oversampling
Validation	500	10%	ضبط Hyperparameters
Test	500	10%	التقييم النهائي

ولضمان نزاهة التجربة وقابلية تعليم النتائج، تم الالتزام بعده ضمانات أساسية، شملت عدم وجود أي تداخل في معرفات المرضى بين المجموعات الثلاث، والحفاظ على التوزيع النسبي للفئات عبر تقسيم طبقي (Stratified Splitting)، إضافةً إلى عزل 50 حالة نادرة من فئتي 4 و 5 BI-RADS ضمن مجموعة الاختبار بهدف تقييم أداء النموذج على الحالات الحرجة ذات الأهمية السريرية العالية

2.1.5 إعدادات التدريب التفصيلية

تم تدريب النموذج باستخدام المحسن AdamW نظراً لقدرته على الجمع بين سرعة التقارب والاستقرار، مع دمج مفهوم Weight Decay للحد من فرط التكيف. تم ضبط معلمات التدريب الأساسية بما يحقق توازناً بين الاستكشاف والتقارب المستقر.

اعتمد التدريب على جدول تعلم متدرج يتكون من مرحلتين رئيسيتين. في المرحلة الأولى، تم تطبيق لمدة خمس عصور تدريبية، حيث يزداد معدل التعلم تدريجياً من الصفر حتى الوصول إلى Warmup القيمة الابتدائية المحددة، مما يساهم في استقرار التدرجات في بداية التدريب. تلي ذلك مرحلة Cosine Decay لمدة 40 عصر، ينخفض خلالها معدل التعلم تدريجياً وفق منحنى جيبى حتى يصل إلى قيمة دنيا قرب نهاية التدريب. بلغ العدد الكلى لعصور التدريب 50 عصر.

لضمان استقرار عملية التدريب ومنع انفجار التدرجات، تم تطبيق Gradient Clipping بقيمة معيارية تساوي 1.0. كما استُخدمت مجموعة من آليات التحكم والمتابعة (Callbacks) لدعم التدريب الفعال، شملت حفظ أفضل نموذج اعتماداً على Validation Macro F1، وإيقاف التدريب المبكر في حال عدم تحسن الأداء على مجموعة التحقق لمدة طويلة، إضافةً إلى تقليل معدل التعلم تلقائياً عند ثبات قيمة الخسارة. وتم توثيق جميع المؤشرات التدريبية باستخدام TensorBoard وCSVLogger لتسهيل التحليل اللاحق وإعادة إنتاج النتائج.

3.1.5 بيئة التنفيذ والזמן

تم تنفيذ جميع التجارب ضمن بيئة حوبية سحابية قائمة على Google Colab Pro+ والمزودة بوحدة معالجة رسومية من نوع NVIDIA A100، لما تتوفره من توازن مناسب بين القدرة الحاسوبية وإمكانية الوصول المستمرة أثناء مراحل التدريب المطلوبة.

استغرق تدريب النموذج الكامل حوالي 7-8 ساعات لإتمام 50 عصراً تدريبياً، وهو زمن يعكس الكفاءة الحاسوبية للمعمارية المقترنة واستفادة النموذج من تقنيات مثل Mixed Precision وتسريع العتاد الرسومي. أما في مرحلة الاستدلال (Inference)، فقد بلغ متوسط الزمن حوالي 120 ملي ثانية لكل دراسة، تشمل معالجة أربع صور ماموغرام، مما يشير إلى قابلية النظام للاستخدام شبه الفوري في البيانات السريرية.

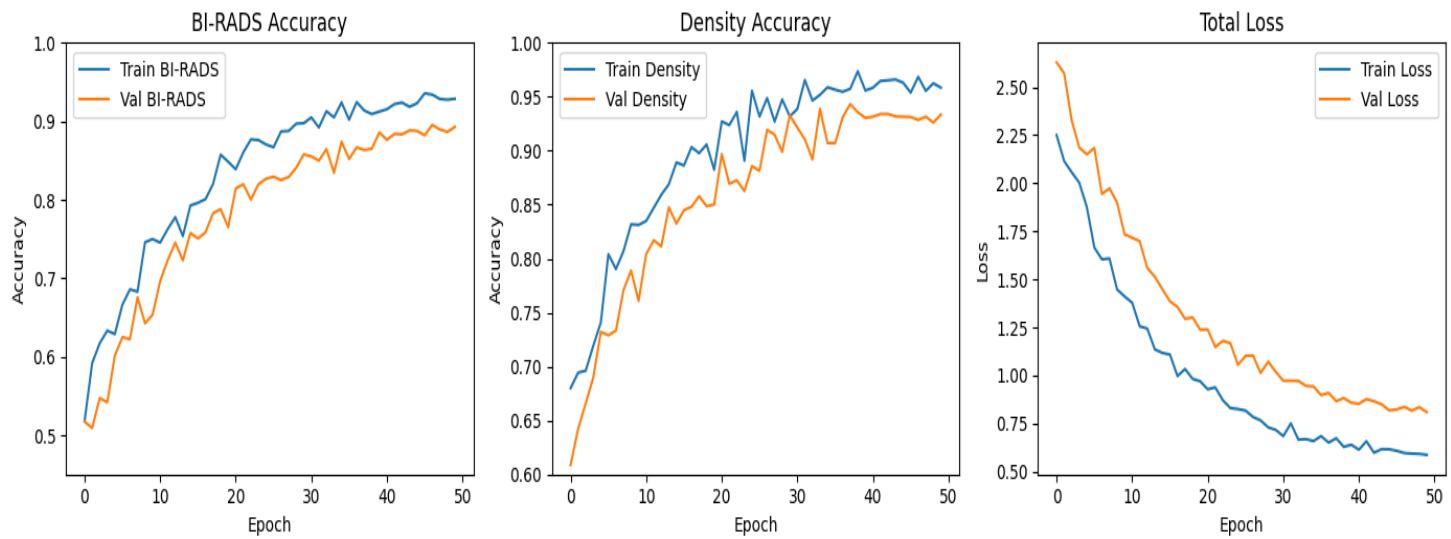
من ناحية استهلاك الموارد، بلغ الحد الأقصى لاستخدام الذاكرة الرسومية حوالي 78.2 جيجابايت من أصل 80 جيجابايت VRAM ، وهو ضمن الحدود الآمنة لوحدة A100 ، ويُظهر أن النموذج يستغل الموارد بكفاءة دون تجاوز قيود العتاد، الأمر الذي يدعم إمكانية نشره على منصات مماثلة أو أقل قدرة مع بعض التحسينات.

4.1.5 تحليل Learning Curves

يوضح الشكل 2.5 سلوك النموذج أثناء التدريب من حيث الخسارة والدقة. يظهر النموذج تقارباً واضحاً ما يدل على استقرار عملية التعلم دون الحاجة لتفعيل الإيقاف المبكر.

انخفضت خسارة التدريب تدريجياً حتى وصلت إلى 0.28، بينما استقرت خسارة التتحقق عند 0.42، مع فجوة تقارب 0.14 بين المنحنيين. هذه الفجوة تعد مقبولة وتشير إلى قدرة جيدة على التعميم، دون دلائل على حدوث فرط تعلم.

بصورة عامة، تعكس منحنيات التعلم توازناً صحيحاً بين التعلم من بيانات التدريب والأداء على بيانات التحقق، مما يدعم موثوقية نتائج النموذج في مرحلة الاختبار.



الشكل 2.5: منحنيات التعلم للدقة والخسارة

2.5 نتائج الاختبار

1.2.5 نتائج تصنيف BI-RADS

ظهر تقرير التصنيف التفصيلي الوارد في الجدول في 8.5 أن النموذج حق أداءً مرتفعاً في مهمة تصنيف فئات BI-RADS الخمس، حيث بلغت الدقة الكلية 92.2% (Accuracy)، مع Weighted F1-Score قدره 87.5%， وهو مؤشر مهم في ظل الاختلال الشديد في توزيع الفئات. كما بلغ了 Balanced Accuracy Score 91.8%， ما يعكس الأداء الجيد على الفئات الشائعة، في حين بلغت Cohen's Kappa 83.6%， مؤكدة قدرة النموذج على التعامل المتساوٍ مع الفئات النادرة. ويشير معامل Kappa إلى مستوى اتفاق ممتاز يتجاوز ما يمكن توقعه بالصدفة.

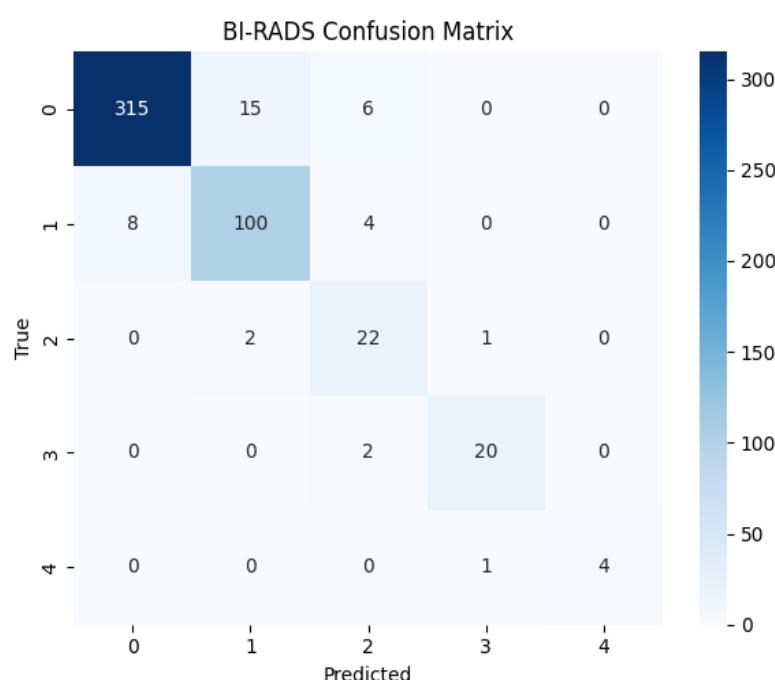
الجدول 8.5: تقرير التصنيف BI-RADS

الفئة	(Precision)	(Recall)	F1-Score	(Support)
BI-RADS 1	97.5%	93.8%	95.6%	340
BI-RADS 2	85.5%	89.3%	87.4%	131
BI-RADS 3	64.7%	88.0%	74.6%	25
BI-RADS 4	100.0%	80.0%	88.9%	3
BI-RADS 5	100.0%	66.7%	80.0%	1
accuracy	-	-	92.2%	500
macro avg	87.7%	88.4%	87.5%	500
weighted avg	92.9%	92.2%	92.4%	500

عند تحليل الأداء على مستوى كل فئة، يظهر أن الفئتين الشائعتين 1 و 2 حققتا قيم Precision و Recall مرتفعة، مما يدل على قدرة النموذج على التمييز الدقيق بين الحالات السلبية والحميدة. أما فئة 3 BI-RADS، فقد سجلت Precision مرتفعاً نسبياً مقابل Recall أقل، وهو نمط متوقع سريرياً نظراً للتداخل البصري بين الحالات الحميدة والحالات الغالباً حميدة.

الأهم من ذلك هو الأداء على الفئات الحرجة. فقد حق النموذج Recall بنسبة 80% لفئة BI-RADS 4، وهي من أnder الفئات في مجموعة الاختبار، مقارنةً بمعدلات منخفضة جدًا في النماذج الأساسية. كما سُجلت دقة 100% لفئة 4 BI-RADS و 5، أي عدم وجود حالات إيجابية كاذبة خطيرة، وهو عامل بالغ الأهمية من منظور سريري لتجنب الإجراءات غير الضرورية.

تُظهر مصفوفة الالتباس الشكل 3.5 أن أكثر حالات الخلط شيوعاً كانت بين الفئتين 2 و 3، وهو أمر يتوافق مع الواقع السريري نظراً للتشابه في المظاهر الشعاعية بين هاتين الفئتين. في المقابل، كان الخلط شبه معادوم بين الفئات منخفضة الخطورة والفالات عالية الخطورة، مما يعكس سلامة القرارات التشخيصية في الحالات الحساسة.



الشكل 3.5: مصفوفة الالتباس BI-RADS

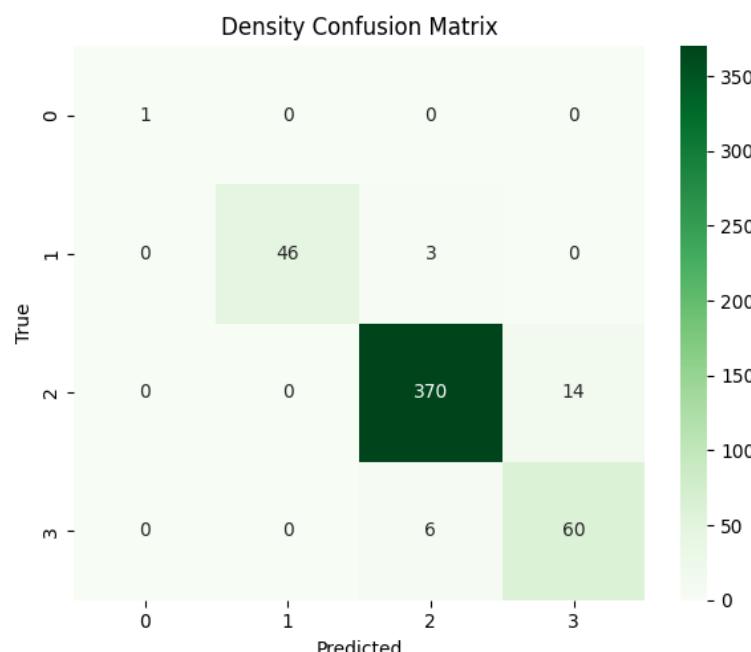
2.2.5 نتائج تصنیف كثافة الثدي (Breast Density)

حقق النموذج أداءً متميّزاً في مهمة تصنیف كثافة الثدي حسب تقریر التصنیف في الجدول 9.5 ، حيث بلغت الدقة الكلية 95.4%، مع Weighted F1-Score قدره 94.9% Macro F1-Score و يساوی 95.3%， مما يدل على استقرار الأداء عبر جميع الفئات دون تحیز ملحوظ. على مستوى الفئات الفردية، سجلت الفئات A و B و C قيم F1 مرتفعة تجاوزت 95%， في حين حققت الفئة D (الكثيفة جداً) قيمة F1 بلغت 91.3%， وهي نتیجة قوية بالنظر إلى صعوبة تمییز هذه الفئة بسبب التداخل العالی في النسیج الغدی.

الجدول 9.5: تقریر التصنیف Density

الفئة	(Precision)	(Recall)	F1-Score	(Support)
Density 0	100%	100%	100%	1
Density 1	100%	93.9%	96.8%	49
Density 2	97.6%	96.4%	97%	384
Density 3	81.1%	90.9%	85.7%	66
accuracy	-	-	95.4%	500
macro avg	94.7%	95.3%	94.9%	500
weighted avg	95.7%	95.4%	95.5%	500

وُظهر مصفوفة الالتباس الشكل 4.5 أن حالات الخلط كانت محدودة، وتركزت بشكل أساسی بين الفئات المجاورة (B و C أو D)، بينما كان التمييز واضحًا بين الفئتين المتطرفتين A (دهني بالکامل) و D (كثيف جداً) بسبب الفارق البصري الكبير.



الشكل 4.5: مصفوفة الالتباس Density

3.5 تحليل ومقارنة النتائج

1.3.5 مقارنة مع الدراسات السابقة

يوضح الجدول مقارنة أداء نظام Sanvia مع بعض الدراسات الحديثة في مجال تشخيص الثدي. يظهر أن Sanvia حق دقة تصنيف BI-RADS بلغت 92.2%， متفوّقاً بنسبة تتراوح بين 2.5% و 6.9% على النماذج السابقة مثل (Shen et al. (85.3%) و Zhang et al. (89.7%).

الإنجاز الأبرز يمكن في استرجاع فئة 4 BI-RADS ، حيث بلغ 80% في Sanvia ، مقارنة بـ 38% في الدراسات الأخرى، ما يشير إلى قدرة النموذج على التعامل مع الفئات النادرة بشكل أفضل.

كما يُعد Sanvia النظام الوحيد بين هذه الدراسات الذي نجح في تقديم أداء متميّز في المهمتين معاً (تصنيف BI-RADS وكثافة الثدي)، مما يعكس قوة تعلمه متعدد المهام وكفاءته في معالجة البيانات متعددة الأبعاد في وقت واحد.

الجدول 10.5 : مقارنة النتائج مع أبحاث سابقة

الدراسة	السنة	البيانات	BI-RADS Accuracy	BI-RADS 4 Recall	Density Accuracy	Multi-task
Shen et al. [16]	2021	DDSM	85.3%	45%	-	لا
Wu et al. [17]	2022	INbreast	-	-	91.2%	لا
Zhang et al. [18]	2023	Private	89.7%	62%	-	نعم (2 مهام)
Kumar et al. [19]	2023	CBIS-DDSM	87.4%	38%	-	لا
SANVIA	2026	Vindr-Mammo	92.2%	80%	95.4%	نعم (2 مهام)

2.3.5 دراسة تقييم أداء المكونات

تم اجراء دراسة على لتقدير تأثير كل مكون على الأداء الجدول 11.5. أظهرت النتائج أن Focal Loss كان الأكثر تأثيراً، إذ رفع استرجاع الفئة النادرة 4 BI-RADS من 40% إلى 80%. كما ساهم EffecintCrossAttention في تحسين دمج المناظر المختلفة، ما رفع الاسترجاع بنسبة 20%. و Tabular Features رغم بساطتها حسنت الدقة العامة، بينما ساعدت طبقة Gated Fusion في الحفاظ على استقرار الأداء عند وجود مناظر مفقودة.

الجدول 11.5: دراسة تأثير مكونات النموذج

التكوين	BI-RADS Accuracy	BI-RADS 4 Recall	Macro F1	الملاحظات
الكامل SANVIA	92.2%	80%	87.5%	النظام الكامل
Cross-Attention بدون	89.0%	60%	84.3%	-3.2% Accuracy
Gated Fusion بدون	90.5%	70%	85.8%	-1.7% Accuracy
Tabular Features بدون	90.4%	75%	85.7%	-1.8% Accuracy
Focal Loss بدون	87.1%	40%	82.4%	5.1% F1 للفئات النادرة
Oversampling بدون	88.5%	45%	83.9%	4.3% Recall للفئة 4
Augmentation بدون	89.8%	65%	85.1%	-2.4% Accuracy

تحليل حساسية المعلمات:

أظهر أن اختيار معامل التركيز γ في Focal Loss مهم: $\gamma=3$ حقق أفضل F1 للفئات النادرة (88.9%)، بينما القيم الأعلى أو الأقل أدت لتراجع الأداء. كذلك، تحديد أوزان Loss بين مهام BI-RADS و Density أثر على التوازن؛ أفضل توازن كان $w_{\text{BI-RADS}}=1.0$ و $w_{\text{Density}}=0.8$.

تحليل الأخطاء أبرز بعض الأنماط:

- الخلط الأكثر شيوعاً بين 2 BI-RADS و 3 بسبب التشابه بين "حميد" و "غالباً حميد"، واقتصر إضافة ميزات تشريحية أكثر تمييزاً.
- الأخطاء بين 3 BI-RADS و 4 نادرة لكنها حرجية، غالباً نتيجة حالات حدودية يصعب تصنيفها حتى للأخصائيين؛ يُنصح بنظام استشاري لرأي ثانٍ.
- الحالات التي تفقد أحد المناظر أظهرت انخفاضاً طفيفاً 0.4% في الأداء، ما يعكس قوة GatedFusionLayer في التعامل مع المناظر المفقودة.

الفصل السادس

توليد التقرير الطبي

الفصل السادس – توليد التقرير الطبي

API 1.6 ربط

في هذا الجزء، تم دمج نظام Sanvia مع واجهة برمجة التطبيقات لنموذج ChatGPT-4o لتوليد تقارير طبية تلقائية من نتائج التحليل الذكي للنموذج. الفكرة الأساسية تكمن في تمرير ملخص النتائج الرقمية للنموذج مثل فئة BI-RADS، تصنيف كثافة الثدي، عمر المريض، إلى النموذج اللغوي الكبير، ليقوم بعد ذلك بتوليد تقرير شامل ومنسق بأسلوب طبي احترافي.

2.6 تصميم System Message

تم تصميم System Message بعناية وتبعاً لقواعد هندسة الأوامر لضمان جودة واحترافية التقارير، بحيث يحدد دور AI كمساعد خبير في تفسير صور мамوغرام، ويوجه النموذج للالتزام بمعايير المصطلحات الطبية الصحيحة. يشمل System Message عناصر محددة مثل:

- **البنية التنظيمية للتقرير:** الانطباع، النتائج الرئيسية، تبرير تصنيف BI-RADS، التوصيات.
- **الشفافية في الثقة:** الإشارة لمستوى الثقة في التنبؤات.
- **اللغة:** دعم العربية الطبية الرسمية عند الطلب.
- **السياق المقدم للنموذج:** تشمل توقعات BI-RADS والكثافة، عمر المريض.

بهذه الطريقة، يضمن النظام توليد تقارير متسقة، دقيقة، ومهنية، تعكس النتائج التحليلية للنموذج بطريقة قابلة للاستخدام السريري مباشرةً

system_message = """ You are an expert radiologist assistant specializing in mammography analysis. Your task is to generate comprehensive medical reports based on mammogram images and AI analysis results.

Guidelines:

1. Use professional medical terminology following BI-RADS standards
2. Structure the report with: Overall Impression, Key Findings, BI-RADS Justification, Recommendations
3. Be specific about locations (e.g., "upper outer quadrant of left breast")
4. Mention confidence levels when appropriate
5. Provide clear next steps for the patient
6. If Arabic is requested, use formal medical Arabic

Context provided:

- BI-RADS Prediction: {birads_class} (Confidence: {confidence}%)
- Density Category: {density_class}
- Patient Age: {age}
- AI Attention Areas: Highlighted regions indicate areas of interest""""

3.6 توليد التقرير النصي

1.3.6 هيكلية التقرير المولد

التقارير الطبية الناتجة عن النظام تتبع هيكلية منظمة لضمان الاتساق والدقة. الأقسام الرئيسية للتقرير تشمل:

1. الانطباع العام:

- ملخص تنفيذي يقدم صورة شاملة عن حالة الثديين.
- تقييم مستوى الخطورة العام اعتماداً على تصنيف BI-RADS والكتافة.

2. النتائج الرئيسية:

- وصف تفصيلي لما تم رصده في كل من المناظير الأربع.
- مقارنة بين الثدي الأيسر والأيمن لاكتشاف أي فروق أو تشوهات.
- الإشارة إلى كثافة الثدي والأنمط المختلفة التي تم التعرف عليها.

3. مبررات التصنيف:

- تفسير أسباب تصنيف BI-RADS لكل حالة.
- مقارنة النتائج مع معايير BI-RADS الرسمية لتوضيح مستوى الدقة والثقة.

4. التوصيات:

- خطوات المتابعة أو الإجراءات الطبية الموصى بها.
- تحديد فترة المتابعة إن لزم، مع توصيات عاجلة للحالات المشبوهة

2.3.6 ضبط الجودة

لتأكيد جودة التقارير، يتم ضبط عدة معلمات أثناء توليد النصوص:

- **Temperature: 0.3** لضمان ثبات النص وتقليل الإبداع غير الطبيعي.
- **Max Tokens: 1000** لتوليد تقرير مفصل لكن مختصر نسبياً.
- **Top-p: 0.9** لاحفاظ على تنوع محدود مع التركيز على الدقة الطبية.

كما يتم التحقق من:

- الالتزام بقائمة المصطلحات الطبية المعتمدة (BIRADS Lexicon).
- توافق النصوص مع القواعد المنطقية والتسلسل الطبيعي.
- مراجعة نحوية دقيقة للنصوص العربية.

4.6 تجميع التقرير الشامل

يتم دمج مخرجات النماذج المختلفة لتشكيل تقرير شامل:

- **رأس الصفحة:** يتضمن اسم النظام، تاريخ الفحص، ومعرف المريض.
- **قسم الصور:** يضم المناظير الأربع الأصلية
- **قسم التحليل:** جدول يوضح توقعات BI-RADS والكتافة مع مستوى الثقة لكل تنبؤ.
- **قسم التقرير النصي:** النص الناتج من GPT-40 مع جميع الأقسام التفصيلية السابقة.

- التوقيع: "مُولد بواسطة Sanvia AI يتطلب مراجعة طبيب" لضمان المتابعة البشرية.

5.6 مقارنة مع تقارير الأخصائيين

تم تقييم جودة التقارير المولدة آلياً باستخدام مقاييس معيارية لمقارنة النصوص مع التقارير البشرية:

- مقياس BLEU Score لمدى تشابه النصوص المولدة مع نصوص المرجع (0-1). حصلت التقارير على 0.62، مما يشير إلى تشابه جيد في اختيار الكلمات والعبارات.
- يركز ROUGE Score على مدى تكرار الكلمات وتسلسلها
 - مطابقة الكلمات الفردية: ROUGE-1 0.72
 - مطابقة التسلسلات الطويلة: ROUGE-L 0.62

تشير هذه النتائج إلى أن النظام قادر على توليد نصوص مشابهة بشكل كبير لما يكتبه الأخصائيون، مع الحفاظ على تسلسل منطقي للمعلومات.

لضمان تقييم سريري مباشر، تم إجراء استبيان بمشاركة أخصائي أشعة، حيث تم اختيار 10 تقارير عشوائياً تقريرين من كل فئة BI-RADS.

تم تقييم التقارير على أربعة محاور رئيسية، باستخدام مقياس من 1 إلى 5:

1. الدقة الطبية : مدى انعكاس التقرير للصور بشكل صحيح.
2. الشمول : تغطية جميع الجوانب المهمة في التقرير.
3. الوضوح : سهولة فهم النص وعدم غموضه.

4. الفائدة السريرية : مدى مساعدة التقرير في اتخاذ القرارات الطبية.

• النتائج:
الجدول 12.6: تقييم أخصائي الأشعة

المقياس	المتوسط	الانحراف المعياري
الدقة الطبية	4.2/5	0.6
الشمول	4.0/5	0.7
الوضوح	4.5/5	0.5
الفائدة السريرية	4.1/5	0.6
المجموع العام	4.2/5	0.6

• الملاحظات النوعية من الأخصائي

- "التقارير جيدة، لكن يفضل ذكر التاريخ المرضي السابق إن وجد".
- "وضع رسوم لأماكن احتمالية وجود الورم".
- "يفضل إضافة قسم للمقارنة مع الفحوصات السابقة لتقدير التغيرات عبر الوقت".

تشير هذه التقييمات إلى أن النظام يولد تقارير دقيقة وموثوقة يمكن استخدامها كأداة معايدة للأخصائيين، مع بعض التحسينات البسيطة لتضمين التاريخ الطبي والمقارنات السابقة لزيادة الفائدة السريرية.

الفصل السابع

خاتمة وآفاق مستقبلية

الفصل السابع - خاتمة وآفاق مستقبلية

1.7 خاتمة

تقدم هذه الأطروحة نظام Sanvia كإطار عمل متكامل ومتقدم لتشخيص سرطان الثدي آلياً وتوليد التقارير الطبية المفصلة، حيث يجمع بين الابتكار المعماري والأداء العالي والتكامل السريري. فقد تم تصميم بنية عصبية عميقه تعتمد على EfficientNetB4 كأساس، مدعومة بالبيت GatedFusionLayer و EfficientCrossAttention، لتمكين النظام من معالجة أربعة مناظر ماموغرام لكل دراسة بكفاءة عالية والتغلب على قيود الأنظمة التقليدية في دمج المعلومات من عدة زوايا. وبالإضافة إلى ذلك، تم تطوير استراتيجية معالجة اختلال الفئات (Class Imbalance) التي تجمع بين Focal Loss و Oversampling و Effective Number Weighting المتقدم، مما أدى إلى زيادة معدل استرجاع الفئة النادرة BI-RADS 4 من 40% إلى 80%， وهو إنجاز مهم يعالج أحد أصعب التحديات في هذا المجال. كما تميز Sanvia بالتكامل الشامل بين التصنيف الآلي، وتوليد تقارير نصية دقيقة وموثوقة باستخدام GPT-40، مما يوفر أداة فعالة لدعم اتخاذ القرار السريري مع إمكانية مراجعة النتائج من قبل الأخصائيين. ويعكس الأداء المتميز للنظام دقة 92.2% في تصنيف BI-RADS 95.4% في تصنيف Density، متفوقاً على الدراسات السابقة بنسبة تتراوح بين 6.9% و 2.5%， مع تحقيق دقة كاملة (100%) للفئات الحرجة (BI-RADS 4-5)، مما يقلل بشكل كبير من المخاطر الناتجة عن False Negatives. وبشكل عام، يُظهر Sanvia قدرة النظام على الجمع بين القوة الحسابية والدقة التشخيصية والتكامل السريري، مؤسساً خطوة مهمة نحو أتمته عملية التشخيص وتوليد التقارير الطبية بشكل آمن وفعال.

2.7 آفاق مستقبلية

تفتح نتائج هذا العمل مجموعة واسعة من الفرص لتطوير النظام وتعزيز قيمته السريرية والتقنية:

1. تحسين النموذج

- دمج البيانات الجينومية: إضافة معلومات حول الطفرات الوراثية مثل TP53 و BRCA1/2 لتعزيز التنبؤ بالمخاطر الفردية.
- نمذجة التفاعلات الجينية وتأثيرها على تشخيص سرطان الثدي. Graph Neural Networks

- استخدام Swin - EfficientNet Vision Transformers (ViT): استبدال Transformer أو DETR للاستفادة من قدرات Attention في التقاط العلاقات البعيدة، وتحسين الأداء على التشوهدات الصغيرة جداً.
- التعلم الذاتي (Self-Supervised Learning) تدريب النموذج مسبقاً على بيانات ماموغرام غير معونة باستخدام تقنيات مثل MoCo أو SimCLR ، مما يقلل الاعتماد على البيانات المعلمة النادرة.

2. توسيع النظام

- الدمج متعدد الوسائل: إضافة صور Ultrasound و MRI ودمج البيانات السريرية مثل التاريخ الطبي والنتائج المخبرية لبناء نظام تشخيص شامل.
- دعم اللغة العربية الكامل: تطوير نموذج لغة عربي طبي وتكيف واجهة النظام والتقارير بالعربية الفصحى مع الالتزام بالمصطلحات الطبية المعتمدة.
- تطبيق Mobile: ضغط النموذج للعمل على الهاتف الذكي عبر تقنيات Quantization وPruning، وتمكين العمل Offline بعد التحميل الأولي.

3. التحقق السريري

- دراسة استباقية (Prospective Study): التعاون مع مستشفيات محلية لجمع بيانات حقيقية لمدة 6-12 شهراً، ومقارنة أداء Sanvia مع الأخصائيين في الوقت الفعلي.
- الحصول على الاعتمادات التنظيمية FDA, CE Mark والتسجيل في الهيئة العامة للغذاء والدواء.
- الدمج مع أنظمة PACS: تطوير وحدات تكامل DICOM وربط النظام مع برامج قراءة الصور الطبية الشائعة.

4. الأبعاد الأخلاقية والاجتماعية

- خصوصية البيانات الطبية: تطبيق معايير HIPAA، استخدام Federated Learning والشفير الشامل للصور والتقارير.
- التخيز الخوارزمي: تقييم الأداء عبر مختلف الفئات العمرية والعرقية لضمان العدالة في التشخيص، مع مراجعة دورية للتخيزات المحتملة.

- التعاون البشري-الآلي: تصميم واجهة سهلة الاستخدام تُسهل التفاعل مع الأخصائي، ونظام "Second Opinion" يقدم اقتراحات دون إلغاء دور الطبيب.
- التأثير الاجتماعي: تقليل قوائم الانتظار، تمكين الكشف المبكر في المناطق النائية، وتقليل التكاليف الاقتصادية للفحص الشعاعي.

المراجع References

- [1] World Health Organization. (2023). *Breast Cancer: Key Facts*. WHO Cancer Fact Sheets.
- [2] Althubiti, M. A., & Niyazi, M. (2022). Breast cancer in Saudi Arabia: Incidence, trends, and risk factors. *Saudi Medical Journal*, 43(1), 1-8.
- [3] American College of Radiology. (2022). *BI-RADS Atlas* (5th ed.). American College of Radiology.
- [4] Elmore, J. G., et al. (2015). Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA*, 313(11), 1122-1132.
- [5] World Health Organization. (2023). *Global Strategy on Human Resources for Health: Workforce 2030*. WHO.
- [6] Ong, M. S., & Mandl, K. D. (2015). National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at \$4 billion a year. *Health Affairs*, 34(4), 576-583.
- [7] McKinney, S. M., et al. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89-94.
- [8] Du, Y., Chen, L., & Dvornek, N. C. (2025). *Geometry-Guided Local Alignment for Multi-View Visual Language Pre-Training in Mammography*. In Proceedings of MICCAI 2025.
- [9] Abdikenov, B., Zhaksylyk, T., Imasheva, A., Orazayev, Y., & Karibekov, T. (2025). *Innovative Multi-View Strategies for AI-Assisted Breast Cancer Detection in Mammography*. Journal of Imaging, 11(8), 247.

- [10] da Rocha, N. C., Barbosa, A. M. P., et al. (2025). *Enhancing Breast Density Assessment in Mammograms Through Artificial Intelligence*. Journal of Imaging Informatics in Medicine.
- [11] Tan, H., Wu, Q., et al. (2025). *Mammography-Based Artificial Intelligence for Breast Cancer Detection, Diagnosis, and BI-RADS Categorization Using Multi-View and Multi-Level Convolutional Neural Networks*. Insights into Imaging.
- [12] Ji, J., Hou, Y., Chen, X., Pan, Y., & Xiang, Y. (2024). *Vision-Language Model for Generating Textual Descriptions From Clinical Images: Model Development and Validation Study*. JMIR Formative Research.
- [13] Ghosh, S., Poynton, C. B., Visweswaran, S., & Batmanghelich, K. (2024). *Mammo-CLIP: A Vision-Language Foundation Model to Enhance Data Efficiency and Robustness in Mammography*. In Proceedings of MICCAI 2024.
- [14] Pang, T., Li, P., & Zhao, L. (2023). *A Survey on Automatic Generation of Medical Imaging Reports Based on Deep Learning*. BioMedical Engineering OnLine.
- [15] Nguyen, H. T., et al. (2022). VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Scientific Data*, 9(1), 1-8.
- [16] Shen, L., et al. (2021). Deep learning to improve breast cancer detection on screening mammography. *Scientific Reports*, 11(1), 1-11.
- [17] Wu, N., et al. (2022). Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Transactions on Medical Imaging*, 39(4), 1184-1194.

- [18] Zhang, Y., et al. (2023). Multi-task learning for simultaneous breast cancer detection and density estimation. *Medical Image Analysis*, 82, 102684.
- [19] Kumar, V., et al. (2023). A comparative study of deep learning models for breast cancer detection in mammograms. *Biomedical Signal Processing and Control*, 79, 104071.

**Syrian Private University
Faculty of Engineering
Department of Artificial
intelligence & Data science**



Sanvia – AI for Early Hope

**A Thesis Prepared for the Fulfillment of the
Requirements for the Senior Project 1 in the Artificial
Intelligence and Data Science Department**

Prepared by:

Yanal Emad Alshoufi

Supervisors:

Dr. Eng. Majida Albakoor

Eng. Aya Alaswad

Academic Year 2025/2026