



Diagnosing breast cancer using machine learning

Wafa Ahmed Miftah Aljadi *

Department of Computer Science, Faculty of Information Technology, Bani Waleed
University, Bani Walid, Libya

تشخيص سرطان الثدي باستخدام التعلم الآلي

* وفاء أحمد مفتاح الجدي

قسم علوم الحاسوب، كلية تقنية المعلومات، جامعة بنى وليد، بنى وليد، ليبيا

*Corresponding author: wafaalgade@bwu.edu.ly

Received: September 20, 2025 | Accepted: November 27, 2025 | Published: December 09, 2025

Abstract:

Breast cancer is a common global disease that contributes significantly to women's mortality due to misdiagnosis or delayed treatment, necessitating an urgent need for accurate diagnosis. Machine learning (ML) techniques are an important field of study proven effective in cancer prediction and early diagnosis. This study aimed to diagnose breast cancer using ML models and evaluate their effectiveness based on four criteria: accuracy, specificity, sensitivity, and F1 score. The research utilized a dataset from Kaggle, originally collected by Dr. William H. Wolberg at the University of Wisconsin Hospital in the United States. Four ML algorithms were tested: Random Forest (RF), Gradient Boosting Classifier (GB), Logistic Regression (LR), and Support Vector Machine (SVM). Upon comparing the results, the Random Forest algorithm was concluded to have achieved the highest accuracy, reaching 96.4%.

Keywords: Breast Cancer, Classification, Machine Learning, Random Forest.

الملخص:

يُعد سرطان الثدي من الأمراض الشائعة بين النساء حول العالم، والذي يؤدي إلى وفيات العديد منها بسبب التشخيص الخاطئ أو وصف العلاج متأخرًا، مما خلق حاجة مُلحّة للتشخيص الدقيق. تُعد تقنيات التعلم الآلي (ML) موضوعًا دراسيًا هامًا، وقد أثبتت فعاليتها في التنبؤ بالسرطان والتشخيص المبكر. تهدف هذه الدراسة إلى تشخيص سرطان الثدي باستخدام نماذج التعلم الآلي، وتقييم أكثرها فعالية بناءً على أربعة معايير: الدقة، الخصوصية، الحساسية، ودرجة F1. استخدم هذا البحث مجموعة من البيانات من موقع Kaggle والتي تم جمعها أساساً من قبل الطبيب الدكتور ويليام إتش. وولبرغ في مستشفى جامعة وييسكونسن في الولايات المتحدة الأمريكية. تم اختبار أربعة من خوارزميات التعلم الآلي: الغابة العشوائية (RF)، معزز التدرج (GB)، الانحدار اللوجستي (LR)، وآلة المتوجه الداعم (SVM). بعد مقارنة النتائج لهذه الخوارزميات، تبيّن أن خوارزمية الغابة العشوائية حققت أعلى دقة بنسبة (96.4%).

الكلمات الدالة: سرطان الثدي، التصنيف، التعلم الآلي، الغابة العشوائية.

1. المقدمة

يعتبر سرطان الثدي (Breast Cancer) من أكثر الأمراض انتشاراً في العالم، ويصيب النساء بصورة كبيرة. ويُعد أحد الأسباب الرئيسية لوفاة النساء عالمياً وفقاً لمنظمة الصحة العالمية، إذ يعتبر مسؤولاً عن حوالي 15% من إجمالي الوفيات المرتبطة بالسرطان (Farhan and Kamil., 2020). لذلك، من الضروري الكشف عن أورام أنسجة الثدي في أقرب وقت ممكن لزيادة فرص النجاة في المراحل المبكرة، مما قد يساعد في تعزيز معدلات الشفاء من المرض (Pandey et al., 2021).

توجد العديد من الاختبارات التشخيصية لسرطان الثدي التي تُجرى للتأكد ما إذا كان الورم حميداً أو خبيثاً. وتشمل هذه الإجراءات: خزعة الثدي، والتصوير الشعاعي للثدي (Mammography)، والمواجات فوق الصوتية، والتصوير الحراري، وفحص الخلايا بالإبرة الدقيقة. حتى الآن، يستخدم تصوير الثدي بالأشعة السينية على نطاق واسع للكشف عن المرض، ولكنه في بعض الحالات لا يكون كافياً، مما يدفع الأطباء لطلب خزعة قبل اتخاذ القرار النهائي (Zeidan et al., 2015). علاوة على ذلك، يذكر أن معدل دقة الكشف (في الطرق التقليدية) منخفض، ويتراوح بين 60% و70%. وقد يُطلب من المريضة إجراء اختبارات إضافية، قد تكون مكلفة وتستغرق وقتاً طويلاً.

بدأ الباحثون في الاتجاه نحو الاستعانة بتقنيات الذكاء الاصطناعي (AI) في المجال الطبي من أجل بناء أنظمة فعالة للتشخيص بمساعدة الحاسوب الآلي (CAD). وتهدف هذه الأنظمة إلى مساعدة العنصر البشري في تقليل التشخيص الخاطئ، وبالتالي القليل من نسب الوفيات الناتجة عن المرض (Wang et al., 2019). وتُعد تقنيات التعلم الآلي (Machine Learning) مفيدة في أنظمة دعم القرار، حيث تُساعد في إتمام عملية اتخاذ القرار وزيادة معدلات البقاء على قيد الحياة. كما أنها تحسن الدقة وسرعة الاستجابة، وتُقلل من حجم العمل الواقع على عاتق الأطباء، خاصةً في ظل نقص الكوادر الطبية (Radhi and Kamil., 2021).

الدراسات السابقة

- قام نيمادي وزملاؤه (et al., 2023 Nemade) بدراسة لتحديد ما إذا كان الورم في الثدي حميداً أم خبيثاً بالاعتماد على سمات الورم مثل نصف القطر والملمس والمحيط والمساحة والنعومة والإكتاز والتقرير والبعد الكسري. واستخدمو خوارزميات متعددة، ووجدوا أن خوارزمية الغابة العشوائية (RF) كانت ذات الأداء الأفضل، حيث حققت أعلى دقة بلغت 97%.
- وفي عام (2021)، قام آرا وزملاؤه (et al Ara) بدراسة استخدمو فيها خوارزميات لتصنيف سرطان الثدي (خبيث/حميد). وكانت نتائج الدراسة أن خوارزميتي آلة المتجه الداعم (SVM) والغابة العشوائية (RF) حققا الأداء الأفضل، بدقة وصلت إلى 96.5%.
- وفي دراسة خاطر وزملاؤه (et al., 2023 Khater)، التي هدفت إلى تصنیف سرطان الثدي، استخدم الباحثون خوارزميات متعددة (بما فيها ANN و SVM و KNN و RF). وحققت خوارزميتي الشبكات العصبية الاصطناعية (ANN) والجيران الأقرب (KNN) الدقة الأعلى، حيث تراوحت بين 97.7% و 98.2% على التوالي.
- ووصف عازار والسيد (El-Said & Azar, 2014) أن خوارزمية آلة الدعم المتجه (SVM) يمكنها تشخيص سرطان الثدي بسرعة وسهولة وكفاءة.
- ورَّاكز كومار وزملاؤه (et al., 2020 Kumar) على تطبيق عدة خوارزميات تصنيفية لتصنيف أورام سرطان الثدي (خبيث/حميد)، وقد حققت هذه الخوارزميات درجة إجمالية قدرها 94%.

هدف الدراسة وهيكلها

يهدف هذا البحث إلى استخدام مجموعة مختارة من خوارزميات التعلم الآلي في تشخيص سرطان الثدي (خبيث أم حميد)، ومقارنة نتائج هذه الخوارزميات لإيجاد الخوارزمية الأكثر كفاءة وموثوقية لهذا التشخيص.

يقسم باقي البحث على النحو التالي:

- **القسم الثاني:** يعرض الخلفية النظرية للبحث.
- **القسم الثالث:** يصف منهجية البحث، بما فيها جمع البيانات ووصفها، والمعالجة المسبقة للبيانات، وطريقة بناء النماذج.
- **القسم الرابع:** تناقش النتائج المختلفة التي تم الحصول عليها في هذا البحث.
- **القسم الخامس:** يقدم ملاحظات خاتمية ويحدد التوجهات المستقبلية للبحث.

2. الخلفية النظرية

2.1. التعلم الآلي (Machine Learning)

التعلم الآلي هو فرع من فروع الذكاء الاصطناعي، يركز على تطوير نظم وبرامج تمكن الأنظمة الحاسوبية من تحسين أداء المهام واتخاذ القرارات بشكل ذاتي. يعتمد التعلم الآلي على تحليل البيانات والاستفادة منها لاكتساب المعرفة وتحسين الأداء مع مرور الوقت. ويشمل هذا المجال مجموعة متنوعة من التقنيات مثل الشبكات العصبية الاصطناعية، والتعلم العميق، والتصنيف الآلي وغيرها. يستخدم التعلم الآلي في مجموعة واسعة من التطبيقات، مثل التعرف على الصور، ومعالجة اللغة الطبيعية، والتنبؤ، والتحكم في الروبوتات، والمجال الطبيعي والعديد من المجالات الأخرى. ويُعد هذا المجال أحد أهم التكنولوجيات في العصر الحديث، حيث يسهم في تحسين أداء الأنظمة واتخاذ القرارات بشكل كبير (Moor, 2006).

2.2. تقنيات التعلم الآلي المستخدمة في الدراسة

2.2.1. الانحدار логисти (Logistic Regression)

الانحدار логисти هو خوارزمية تصنّف سُتخدم للتنبؤ باحتمال وقوع حدث ثانٍ مثل "نعم أو لا" أو "نجاح أو فشل". تعتمد الخوارزمية على دالة رياضية تسمى الدالة логистическая (Sigmoid) لتحويل الناتج إلى قيمة احتمالية بين 0 و 1. (Zhang et al., 2022).

2.2.2. آلة المتّجه الداعم - SVM (Support Vector Machine - SVM)

تُعد خوارزمية آلة المتّجه الداعم (SVM) إحدى أشهر خوارزميات التعلم الخاضع للإشراف، وستستخدم بشكل واسع في مهام التصنيف وأحياناً في التنبؤ. تهدف هذه الخوارزمية إلى إيجاد أفضل فاصل ممكّن (Hyperplane) بين فئتين أو أكثر من البيانات، وذلك من خلال تعظيم المسافة بين هذا الفاصل وأقرب نقاط من كل فئة، والتي تُعرف باسم المتجهات الداعمة (Support Vectors). (Zhang et al., 2022).

2.2.3. الغابة العشوائية (Random Forest Classifier)

تُعد الغابة العشوائية إحدى تقنيات التعلم الآلي القائمة على مبدأ التعلم الجماعي (Ensemble Learning). تعتمد على بناء مجموعة من أشجار القرار وتوحيد نتائجها للحصول على نموذج أكثر دقة واستقراراً. تعمل الخوارزمية عن طريق إنشاء عدد كبير من الأشجار، يتم تدريب كل منها على عينة عشوائية مختلفة من البيانات. كما يتم اختيار مجموعة عشوائية من الميزات عند كل انقسام في الشجرة، مما يضمن تنوع الأشجار وتجنب الإفراط في التكيف. (Great Learning., 2023) (Overfitting).

2.2.4. التصنيف بالتعزيز التدرج (Gradient Boosting Classifier)

تُعد من أقوى خوارزميات التعلم الآلي المستخدمة في التنبؤات، خصوصاً في مهام التصنيف. تعتمد هذه الخوارزمية على فكرة تحسين الأداء تدريجياً من خلال بناء نماذج بسيطة بشكل متسلّل. يعمل كل نموذج جديد على تصحيح الأخطاء التي ارتكبها النموذج الذي قبله، وتستمر هذه العملية حتى الوصول إلى نموذج قوي ودقيق. (Zhang et al., 2022).

3. البيانات

استخدمت هذه الدراسة مجموعة بيانات من قاعدة بيانات Kaggle، والتي جُمعت في الأساس من قبل الدكتور ويليام إتش. ولوبرغ (Dr. William H. Wolberg) في مستشفى جامعة ويسكونسن (Wisconsin) في الولايات المتحدة الأمريكية. تتكون البيانات من:

- **الخرج (Output):** وهو متغير واحد يمثل التشخيص، وينقسم إلى ($M = \text{خبيث}$) و ($B = \text{حميد}$).

- المدخلات (Features): وهي السمات الخاصة بالكتلة المتورمة في الثدي، وت تكون من 30 سمة. تشمل هذه السمات: المساحة، والملمس، ونصف القطر، والمحيط، والاكتناز، والنعومة، والتعرّق، وال نقاط المقعرة، والتماثل.

3.3. المعالجة المسبقة للبيانات (Data Preprocessing)

للحصول على نتائج دقيقة، لا بد من معالجة البيانات من خلال تنظيفها وتحويلها إلى شكل مناسب للاستخدام في النماذج المختلفة. تضمنت هذه المرحلة:

- معالجة القيم المفقودة (إن وجدت).
- تطبيق البيانات (تسوية نطاقها)، إضافة إلى حساب متوسطها وانحرافها المعياري.
- تقسيم البيانات إلى مجموعتين: 80% للتدريب و 20% لاختبار.

3.3. بيئه العمل (Working Environment)

تضمنت بيئه العمل استخدام برنامج Anaconda، وهي بيئه مفتوحة المصدر مبنية على لغة البرمجة بايثون (Python). وتتضمن هذه البيئة عدداً كبيراً من المكتبات والأدوات التي تُستخدم بشكل شائع في علم تحليل البيانات والتعلم الآلي.

الأدوات والمكتبات المستخدمة:

- Jupiter Notebook: وهو البرنامج الذي استخدم لكتابة التعليمات البرمجية العملية بلغة بايثون لبناء النماذج.
- أهم مكتبات بايثون: NumPy، Pandas، Seaborn (لاحظ تصحيح NumPy إلى Numby)، Matplotlib، Imbalanced-learn، Scikit-learn

4. النتائج والمناقشة

4.1. مقاييس تقييم أداء النماذج

لتقييم أداء نماذج التعلم الآلي الأربع المستخدمة في هذا البحث، تم الاعتماد على مجموعة من مقاييس الأداء الأساسية المستمدة من مصفوفة الارتباط (Confusion Matrix). وُتستخدم هذه المقاييس لتحديد كفاءة كل خوارزمية في تشخيص سرطان الثدي بدقة.

المعادلة الرياضية	التعريف	المقياس
$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$	تقيس النسبة المئوية للتوقعات الصحيحة الإيجابية (سواء إيجابية أو سلبية) من إجمالي التوقعات.	الدقة (Accuracy)
$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$	هو نسبة الحالات الإيجابية التي تم التنبؤ بها بشكل صحيح مقارنة بجميع الحالات الإيجابية الحقيقة (الحالات المصابة فعلياً).	الاستدعاء أو الحساسية (Sensitivity)
$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$	يقيس مدى صحة التوقعات الإيجابية فقط؛ أي نسبة الحالات التي تم التنبؤ بها كإيجابية وكانت إيجابية بالفعل.	الدقة الإيجابية (Precision) أو القيمة التنبؤية الإيجابية
$\text{F1-Score} = \frac{2 * \text{precision} * \text{Recall}}{\text{precision} + \text{Recall}}$	هو مقياس يجمع بين الاستدعاء والدقة الإيجابية عن طريق أخذ الوسط التوافقي لهما، ويعتبر مقياساً جيداً لتقييم النماذج في مجموعات البيانات غير المتوازنة.	معدل F1 (Score)

حيث إنَّ:

- **TP (True Positive)**: التنبؤ الصحيح الإيجابي (تم التنبؤ به خبيثاً وكان خبيثاً).
- **TN (True Negative)**: التنبؤ الصحيح السلبي (تم التنبؤ به حميداً وكان حميداً).
- **FP (False Positive)**: التنبؤ الخاطئ الإيجابي (تم التنبؤ به خبيثاً وكان حميداً).
- **FN (False Negative)**: التنبؤ الخاطئ السلبي (تم التنبؤ به حميداً وكان خبيثاً).

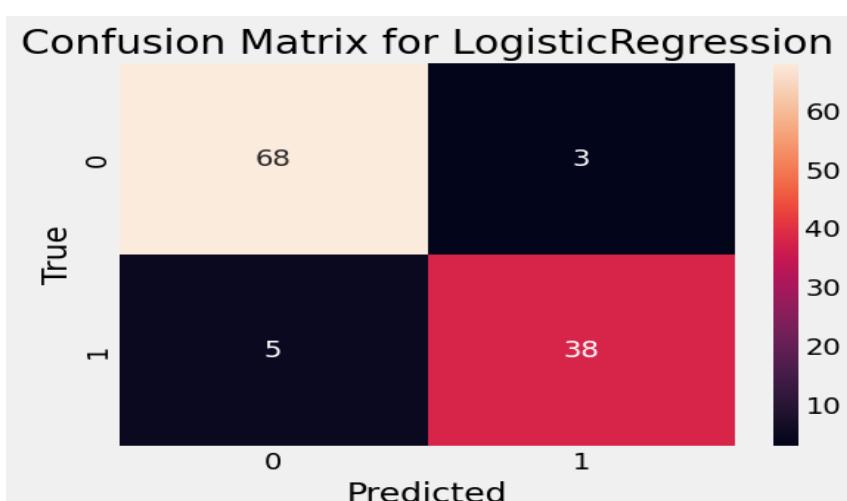
4.2. تطبيق النماذج ومقارنة النتائج

لت Dixon ما إذا كانت المريضة مصابة بورم خبيث أم حميد، تم استخدام أربع خوارزميات تصنيف للتعلم الآلي، وهي: الانحدار اللوجستي (Logistic Regression)، آلة المتغير الداعم (SVC)، الغابة العشوائية (Random Forest Classifier) والتصنيف بالتعزيز التدرج (Gradient Boosting Classifier). قُسّمت البيانات إلى 80% للتدريب و 20% للاختبار. ثم طبقت الخوارزميات المستخدمة في هذه الدراسة على مجموعة الاختبار، وكانت النتائج كما هي موضحة في الجدول (1).

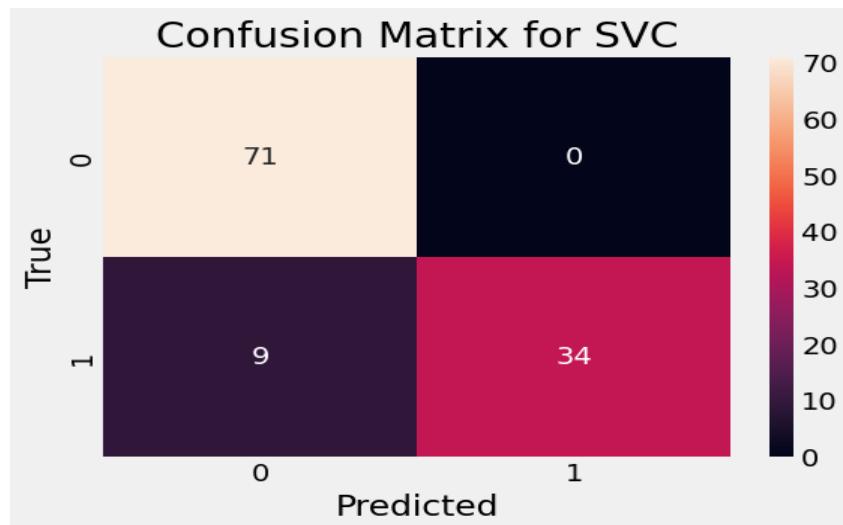
جدول (1) مقارنة نتائج أداء النماذج الأربع على مجموعة الاختبار.

النوع (Model)	الدقة (Accuracy)	الاستدعاء (Recall)	الدقة الإيجابية (Precision)	معدل F1 (F1-Score)
الانحدار اللوجستي (Regression)	0.929825	0.920734	0.926829	0.929476
آلة المتغير الداعم (SVC)	0.921053	0.895349	1.000000	0.918792
الغابة العشوائية (Forest Classifier)	0.964912	0.962660	0.953488	0.964912
التصنيف بالتعزيز التدرج (Gradient Boosting Classifier)	0.947368	0.943990	0.930232	0.947368

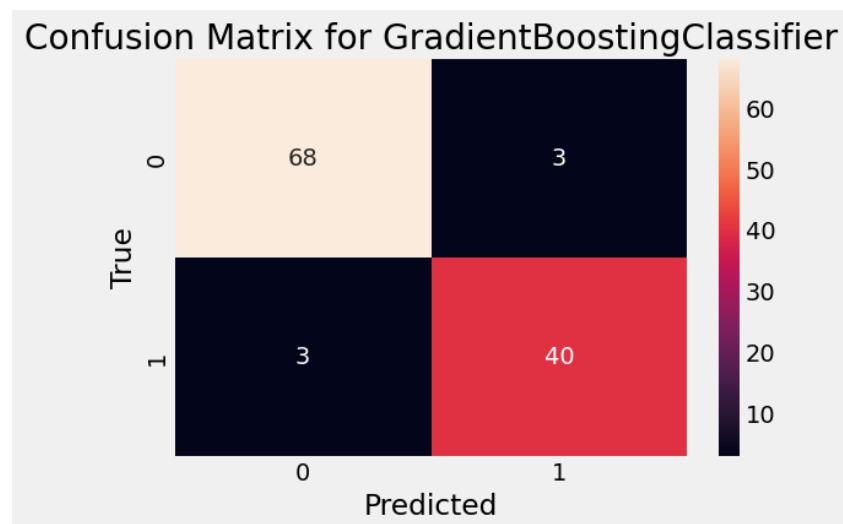
الأشكال (4,3,2,1) توضح قيم التشخيص الإيجابية والسلبية لخوارزميات.



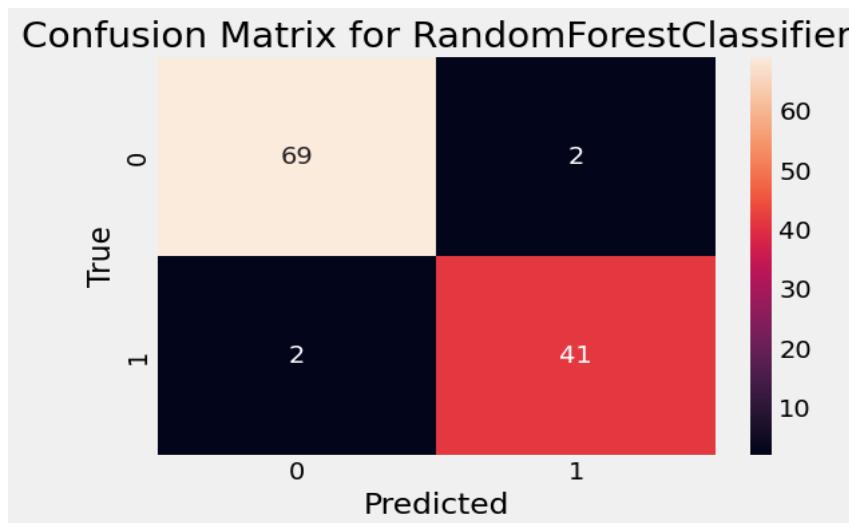
شكل (1) مصفوفة الارتكاك لخوارزمية (LogisticRegression).



شكل (2). مصفوفة الارتباط لخوارزمية (SVC)

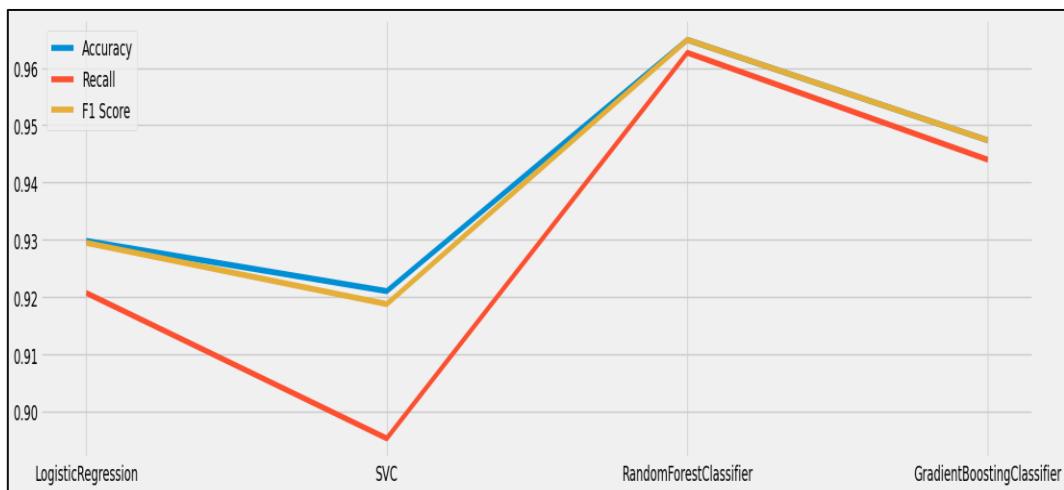


شكل (3) مصفوفة الارتباط لخوارزمية (GB).



شكل (4) مصفوفة الارتباط لخوارزمية (RF).

الشكل 5 يبين رسم المنحنى البياني لأداء النماذج.



شكل (5) الرسم التوضيحي المنحنى البياني لأداء الخوارزميات.

مناقشة النتائج

أظهرت نتائج تقييم الأداء في الجدول (1) أن جميع الخوارزميات الأربع حققت مستويات عالية من الدقة، حيث تجاوزت دقة جميع النماذج نسبة 92%.

1. الأداء الأفضل: حققت خوارزمية الغابة العشوائية (Random Forest Classifier) الأداء الأفضل بين جميع النماذج، حيث بلغت دقتها 0.9649 (أي حوالي 96.49%). كما أنها سجلت أعلى قيمة في مقياس الاستدعاء (0.9626) ومعدل (F1) 0.9649. ويشير هذا إلى تفوق الغابة العشوائية في التشخيص الصحيح الشامل وتوازنها بين تجنب الأخطاء الإيجابية وتجنب الأخطاء السلبية.

2. ملاحظة خاصة بـ SVC: حققت خوارزمية آلة المتجه الداعم (SVC) قيمة دقة إيجابية (Precision) مثالية بلغت 1.00، وهذا يعني أنها لم ترتكب أي أخطاء إيجابية خاطئة (False Positives). بعبارة أخرى، كل حالة تم تشخيصها على أنها "خبيثة" كانت بالفعل "خبيثة". ومع ذلك، كانت قيمة الاستدعاء لليها (0.8953) هي الأقل، مما يشير إلى أنها قد تكون قد أهملت تشخيص بعض الحالات الخبيثة فعلياً (أخطاء سلبية خاطئة).

3. الأداء المقبول: جاءت خوارزميات التصنيف بالتعزيز التدرج (Gradient Boosting) والانحدار логистي (Logistic Regression) في المرتبة الثانية والثالثة على التوالي، بأداء عالٍ ومتقارب، مما يؤكّد فعالية تقنيات التعلم الآلي في هذا المجال. تشير هذه النتائج إلى أن نموذج الغابة العشوائية هو الخيار الأكثر كفاءة وموثوقية لتشخيص سرطان الثدي باستخدام هذه البيانات المحددة.

5. الاستنتاجات والتوصيات

5.1 الاستنتاجات

هدفت هذه الدراسة إلى تقييم فعالية خوارزميات التعلم الآلي في تشخيص سرطان الثدي كحميد أو خبيث، وذلك باستخدام مجموعة بيانات ويسكونسن التشخيصية. ولتحقيق هذا الهدف، تم اختيار ومقارنة أداء أربعة نماذج تصنيف رئيسية وهي: الانحدار логистي (Logistic Regression)، آلة المتجه الداعم (SVC)، والتصنيف بالتعزيز التدرج (GB)، والغابة العشوائية (Random Forest)، بناءً على مقياس الأداء الرئيسي (الدقة، الاستدعاء، الدقة الإيجابية، ومعدل F1).

أظهرت النتائج المتحصل عليها ما يلي:

- أكّدت الدراسة على الكفاءة العالية لتقنيات التعلم الآلي في مهام التشخيص الطبي، حيث تجاوزت دقة جميع النماذج المستخدمة نسبة 92%.
- حققت خوارزمية الغابة العشوائية (Random Forest) الأداء الأفضل بين جميع الخوارزميات، مسجلة أعلى نسبة دقة بلغت 96.49%. وقد سجلت هذه الخوارزمية أيضاً أعلى قيمة في مقاييس الاستدعاة ومعدل F1، مما يشير إلى قدرتها المتوازنة على تحديد الحالات الخبيثة (الإيجابية) وتجنب التصنيفات الخاطئة.
- تؤكد نتائج هذه الدراسة، بالاتفاق مع الأدبيات البحثية السابقة، على أن خوارزمية الغابة العشوائية هي أداة موثوقة وقوية يمكن الاعتماد عليها في تطوير أنظمة دعم القرار السريري لمساعدة الأطباء في التشخيص المبكر والدقيق لسرطان الثدي.

5.2 التوصيات

- بناءً على النتائج التي تم التوصل إليها، يُوصى بالآتي:
- استكشاف خوارزميات التعلم العميق: يُنصح بتوسيع نطاق البحث ليشمل خوارزميات التعلم العميق (Deep Learning)، مثل الشبكات العصبية الالتفافية (CNN) أو الشبكات العصبية المتكررة (RNN)، ومقارنة أدائها مع الغابة العشوائية، خاصةً عند استخدام مجموعات بيانات أكبر وأكثر تعقيداً.
 - تحسين اختيار الميزات: يمكن العمل مستقبلاً على تطبيق تقنيات متقدمة لاختيار الميزات (Feature Selection) لتحديد السمات الأكثر تأثيراً في تشخيص الورم، مما قد يؤدي إلى تحسين دقة النماذج الحالية وتقليل زمن التدريب.
 - تطبيق الخوارزمية في بيئة سريرية: يوصى بمحاولة تطبيق نموذج الغابة العشوائية الأمثل على مجموعة بيانات سريرية غير مرئية (مجموعة اختبار خارجية) للتحقق من قابلية التعميم (Generalizability) وموثوقيتها في بيئة الرعاية الصحية الواقعية.
 - دراسة تأثير عدم توازن البيانات: التحقق من مدى توازن مجموعة البيانات، وفي حال وجود عدم توازن، يمكن تطبيق تقنيات لمعالجة عدم التوازن مثل SMOTE لدراسة تأثيرها على أداء النماذج، خاصةً في مقياس الاستدعاة (Recall).

6. المراجع

- Alshayeqi, A., Ellethy, H., Abed, S., & Gupta, R. (n.d.). Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach. *Biomedical Signal Processing and Control*.
- Ara, S., Das, A., & Dey, A. (2021). Malignant and benign breast cancer classification using machine learning algorithms. In *2021 International Conference on Artificial Intelligence (ICAI)* (pp. 97–101).
- Azar, A. T., & El-Said, S. A. (2014). Performance analysis of support vector machines classifiers in breast cancer mammography recognition. *Neural Computing and Applications*, 24(5), 1163–1177. <https://doi.org/10.1007/s00521-012-1324-4>
- Farhan, M., & Kamil, M. Y. (2020). Texture analysis of breast cancer via LBP, HOG, and GLCM techniques. *IOP Conference Series: Materials Science and Engineering*, 928(7). <https://doi.org/10.1088/1757-899X/928/7/072098>
- Great Learning. (n.d.). *Random forest algorithm in machine learning*. Retrieved November 22, 2023, from <https://www.mygreatlearning.com/blog/random-forest-algorithm/>

6. Khater, T., Hussain, A., Bendardaf, R., Talaat, I. M., Tawfik, H., Ansari, S., & Mahmoud, S. (2023). An explainable artificial intelligence model for the classification of breast cancer. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2023.3308446>
7. Kumar, V., Mishra, B. K., Mazzara, M., Thanh, D. N. H., & Verma, A. (2020). Prediction of malignant and benign breast cancer: A data mining approach in healthcare applications. In V. Kumar, B. K. Mishra, M. Mazzara, D. N. H. Thanh, & A. Verma (Eds.), *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 37, pp. 435–442). Springer. https://doi.org/10.1007/978-981-15-0978-0_43
8. Moor, J. (2006). The Dartmouth College artificial intelligence conference: The next fifty years. *AI Magazine*, 27(4), 87–87.
9. Nemade, V., & Fegade, V. (2023). Machine learning techniques for breast cancer prediction. *Procedia Computer Science*, 218, 1314–1320.
10. Pandey, N., Saini, N., Sapre, D. A., Kulkarni, D. A. K., & Tiwari, D. A. K. (2021). Prioritising breast cancer theranostics: A current medical longing in oncology. *Cancer Treatment and Research Communications*, 29. <https://doi.org/10.1016/j.ctarc.2021.100465>
11. Radhi, B. H., & Kamil, M. Y. (2021). Breast tumor detection via active contour technique. *International Journal of Intelligent Engineering and Systems*, 14(4), 561–570. <https://doi.org/10.22266/ijies2021.0831.49>
12. Wang, W., Huang, Y., Hickman, S., & Gilbert, F. J. (2019). Artificial intelligence in breast imaging. *Clinical Radiology*, 74(5), 357–366. <https://doi.org/10.1016/j.crad.2019.02.006>
13. Zeidan, A., Townsend, P. A., Garbis, S. D., Copson, E., & Cutress, R. I. (2015). Clinical proteomics and breast cancer. *Surgeon*, 13(5), 271–278. <https://doi.org/10.1016/j.surge.2014.12.003>
14. Zhang, L., Geisler, T., Ray, H., & Xie, Y. (2022). Improving logistic regression on the imbalanced data by a novel penalized log-likelihood function. *Journal of Applied Statistics*, 49, 3257.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **SJPHRT** and/or the editor(s). **SJPHRT** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.