Syrian Arab Republic
Ministry of Higher Education
Syrian Virtual University

التنبؤ بالأورام الخبيثة في سرطان الثدي باستخدام أدوات الذكاء الاصطناعي(أداة تعلم الآلة)

Prediction of Malignant Tumors In Breast Cancer using Artificial Intelligence Tools (Machine Learning)

A thesis submitted as a fulfilment of requirements for Master's degree in Bioinformatics

By

Raghad Abdulaziz
Raghad_159886

Supervised by

Dr. Yanal Alkudsy

2023-2022 م

المحتويات

قائمة الأشكال:

قائمة الجداول:

قائمة المصطلحات والاختصارات:

المصطلح	المعنى باللغة العربية	المعنى باللغة الإنكليزية
WHO	منظمة الصحة العالمية	World Health Organization
IDC	سرطان الأقنية الغازي	Invasive ductal carcinoma
ILC	السرطان الفصيصي الغازي	Invasive lobular carcinoma
DCIS	سرطان الأقنية الموضعي	Ductal carcinoma in situ
LCIS	السرطان الفصيصي الموضعي	Lobular carcinoma in situ
ML	تعلم الآلة	Machine learning
AI	الذكاء الصنعي	Artificial intelligence
LOG	الانحدار اللوجستي	Logistic Regression:
KNN	الجار الأقرب	K-nearest neighbor
SVM	ألة المتجه الداعم	Support Vector Machine
DT	شجرة القرار	Decision Tree
RF	الغابة العشوائية	Random Forest

الملخص:

يعتبر سرطان الثدي أحد أكثر الأمراض شيوعاً بين النساء في جميع أنحاء العالم. تم إجراء العديد من الدراسات للتنبؤ بتشخيص سرطان الثدي. ومع ذلك، فإنّ معظم هذه التحليلات استخدمت الأساليب الإحصائية. لذلك، تهدف هذه الدراسة إلى استخدام تقنيات التعلم الآلي لبناء نماذج عالية الدقة والحساسية للكشف عن الأورام الخبيثة في سرطان الثدي بناءً على العديد من المتغيرات، وذلك من أجل التمكن من التدخل السريع في بروتوكول علاج المريض لتقليل الوفيات قدر الإمكان.

تم استخدام مجموعة بيانات من Kaggle بعد معالجتها وتصورها، حيث تألفت مجموعة البيانات النهائية من 569 عينة21 دخل، وخرج وحيد (ورم خبيث وورم حميد).

أظهرت دراستنا أن جميع خوارزميات التعلم الآلي حققت دقة مثالية أكبر من 99% وفق النهج الأول (مجموعة الاختبار= 25%)، حيث احتلت شجرة القرار والانحدار اللوجستي والغابة العشوائية المرتبة الأولى بدقة 100%، تليها بقية الخوارزميات بنسبة 99.3%.

كما وجدنا أن الدقة انخفضت قليلاً في العديد من الخوارزميات وفقاً للمنهج الثاني (مجموعة الاختبار= 40%) لتصل إلى 99.56%. أما عند تحسين البارامترات، زادت دقة خوارزمية متجه آلة الدعم من 99.56% إلى 100%. ويمكن وصف أداء هذا المصنف بأنه متوازن.

ختاماً. تؤكد هذه الدراسة على أهمية اختيار خوارزميات التصنيف المناسبة للتنبؤ بنتائج مرضى سرطان الثدي. تساهم هذه النتائج في مجال تشخيص سرطان الثدي وتوفر رؤى لتحسين استراتيجيات العلاج الشخصية

Abstract:

Breast cancer is one of the most common diseases in women worldwide. Many studies have been conducted to predict the prognosis of breast cancer. However, most of these analyses were predominantly performed using basic statistical methods. There for, this study aims to use machine learning techniques to build high accuracy and sensitivity models for detecting malignancy of breast cancer based on many variables in order to be able to intervene quickly in the patient's treatment protocol to reduce mortality as much as possible.

We utilized a dataset from Kaggle after processing and visualizing it. The final dataset consisted of 569 samples, 21 inputs, and one output (malignant tumor and benign tumor).

Our study showed that all machine learning algorithms achieved perfect accurac greater than 99% according to the first approach (testing set= 25%), where the decision tree, logistic regression, and random forest ranked first with an accuracy of 100%, followed by the rest of the algorithms at 99.3%.

We also found that the accuracy decreased slightly in many algorithms according to the second approach (testing set= 40%) to reach 99.56%. Moreover, when optimizing hyperparameters, the accuracy of the SVM increased from 99.56% to 100%. The performance of this classifier can be described as balanced.

In conclusion. this study underscores the importance of selecting appropriate classification algorithms for predicting breast cancer patient outcomes. These findings contribute to the field of breast cancer prognosis and provide insights for improving personalized treatment strategies.

الفصل الأول: مقدمة
Introduction

1.1. مقدمة:

يعتبر سرطان الثدي السرطان الأكثر شيوعاً لدى الإناث وهو أيضاً السبب الرئيسي عندهنّ حول العالم. تم تشخيص حوالي 1.38 مليون حالة جديدة من سرطان الثدي في عام 2008 بنسبة وصلت إلى 50٪ تقريباً من إجمالي حالات السرطان. تحدث غالبية الوفيات في البلدان النامية بنسبة قدّرت بحوالي 60%.

يوجد حالياً فرق كبير في معدلات البقاء على قيد الحياة لسرطان الثدي في جميع أنحاء العالم، حيث يصل معدل البقاء على قيد الحياة لمدة 5 سنوات إلى 80٪ في البلدان المتقدمة بينما لا يتعدى 40٪ في البلدان النامية. تواجه العديد من البلدان قيوداً على الموارد والبنى التحتية التي تحدّ من القدرة على تحسين إنذار سرطان الثدي ومدة البقيا على قيد الحياة [1].

عادة ما يتم استخدام استراتيجيات مختلفة لتدبير سرطان الثدي مثل العلاج الهرموني، العلاج الشعاعي، الجراحة، والعلاج الكيميائي، وإن الهدف الأساسي دائماً هو تحسين نوعية الحياة ومعدل البقاء على قيد الحياة [2].

- مشكلة البحث:

يعتبر سرطان الثدي السبب الرئيسي لحدوث الوفيات عند الإناث، وتعتبر الأورام الخبيثة بشكل عام قاتلة، إذ عادة ما يكون معدل نموها أعلى بكثير من الأورام الحميدة. لذلك، إن الكشف المبكر على نوع الورم إن كان خبيثاً أو حميداً يعتبر أساسياً لتحديد العلاج المناسب للمرضى المصابين بسرطان الثدي.

- مبررات البحث:

يعتبر التنبؤ بسرطان الثدي من المهام الصعبة في تحليل البيانات الطبية، حيث يحتاج الأطباء وأخصائيو علم الأورام إلى العديد من الأدوات لاتخاذ القرار والتمييز بين الورم الخبيث والحميد. لذلك، فإن استخدام خوارزميات التعلم الآلي (ML) سيكون مفيداً جداً في شأن اتخاذ القرار بشأن التنبؤ بسرطان الثدي.

- هدف البحث:

إنّ الهدف من هذا البحث هو إيجاد خوارزميات ذات دقة وحساسية عالية قادرة على التنبؤ بسرطان الثدي الخبيث بناء على العديد من السمات المميزة للورم مثل نصف القطر، الملمس، المحيط،

مساحة المنطقة، النعومة، الاكتناز، التقعر، النقاط المقعرة، التجانس والبعد الكسري. وذلك من أجل التمكّن من التدخّل السريع لإيجاد التدبير الأمثل لعلاج الورم في الوقت المناسب والتقليل من الوفيّات قدر الإمكان.

2.1. الوبائية:

يحدث سرطان الثدي بشكل رئيسي عند النساء في منتصف العمر وكبار السن. يقدّر متوسط العمر أثناء تشخيص سرطان الثدي 62 عاماً. وهذا يعني أن نصف النساء المصابات بسرطان الثدي يبلغن من العمر 62 عاماً أو أقل عند تشخيصهن. هناك عدد قليل جداً من النساء المصابات بسرطان الثدي ممن يقل عمرهن عن 45 عاماً [3].

احتمال وفاة المرأة بسبب سرطان الثدي هو حوالي حوالي 2.5٪. انخفضت معدلات الوفيات بسرطان الثدي بشكل ملحوظ منذ عام 1989، حيث وصل هذا الانخفاض إلى 43٪ وذلك حتى عام 2020. يعتقد أن انخفاض معدلات الوفيات هو نتيجة لاكتشاف سرطان الثدي في وقت مبكر بسبب زيادة الوعي حول السرطان والفحص المبكر، بالإضافة إلى تحسين العلاجات. مع ذلك، فقد تباطأ هذا الانخفاض قليلاً في السنوات الأخيرة [3].

كانت تقديرات جمعية السرطان الأمريكية لسرطان الثدي في الولايات المتحدة لعام 2023 هي [3]:

- تشخيص حوالي 297.790 حالة جديدة من سرطان الثدي الغازي لدى النساء invasive breast cancer.
- تشخيص حوالي 55,720 حالة جديدة من سرطان الأقنية الموضعي (DCIS) ductal carcinoma in situ.
- موت حوالي 43.700 امرأة بسبب سرطان الثدي.

3.1. تشريح الثدي:

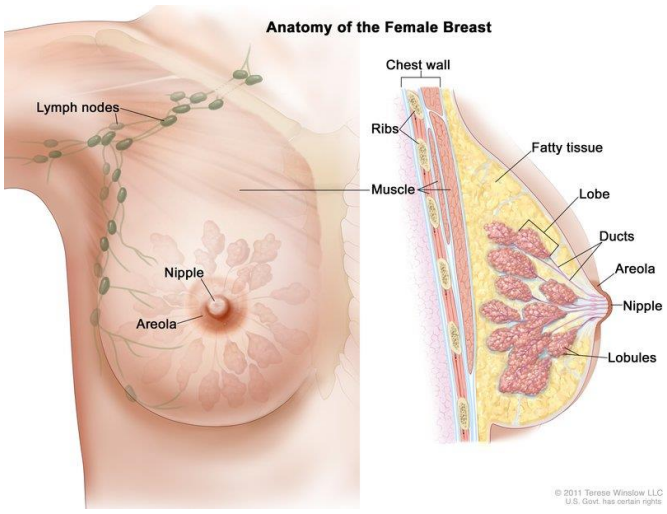
يملك كلا الذكور والإناث ثديين. الثدي هو عبارة عن نسيج دهني، وعادة ما يحتوي ثدي الأنثى على عدد أكبر من الأنسجة الغدية مقارنة بالذكور. يحتوي ثدي الأنثى على 12–20 فص lobes التي تنقسم بدورها إلى فصيصات أصغر lobules. ترتبط هذه الفصوص والفصيصات عبر قنوات الحليب.

يتم تغذية الأنسجة الدهنية عن طريق شبكة من الأعصاب والأوعية الدموية والأوعية الليمفاوية والغدد الليمفاوية، وتتكون أيضاً من أنسجة ضامة ليفية وأربطة.

يعتبر الثدي عضو حساس جداً للتغيرات الهرمونية في الجسم، حيث يخضع لتغييرات دورية متزامنة مع الدورة الشهرية، وتمتلك معظم الإناث ثدياً أصغر قليلاً من الآخر.

إن تحفيز الحلمة يعزز إفراز البرولاكتين من الغدة النخامية. إنّ هذا الهرمون يؤثر على الرحم ويمكن أن يسبب تقلصات، بالإضافة إلى تحفيز إفراز الحليب بعد الولادة.

تتغلغل قنوات الحليب البالغ عددها 15–25 نحو قاعدة الحلمة، وتعمل على نقل الحليب نحو الحلمة. يتواجد حول الحلمة هالة كروية ويتراوح قطرها بين 15 و60 ملم. تتواجد أسفل الحلمة ألياف العضلات الملساء بشكل دائري وشعاعي في النسيج الضام الكثيف وطولياً بجانب القنوات اللبنية لتصل إلى الحلمة. هذه الألياف العضلية هي السبب في إفراز الحليب وانتصاب الحلمة وانقباضها [4].



الشكل 1: تشريح الثدي [5].

4.1. سرطان الثدي:

يبدأ سرطان الثدي في أنسجة الثدي عندما تنمو الخلايا بشكل لا يمكن السيطرة عليه، مما يؤدي في النهاية إلى تكوين كتلة أو ورم.

تتمتع الخلايا في أجسامنا بدورة خلوية طبيعية: فهي تنمو وتنقسم، وتصنع نسخاً من نفسها حسب الحاجة لتحل محل الخلايا القديمة أو غير الطبيعية. وكما أنها تتلقى إشارات للنمو، فإنها تتلقى إشارات للموت أيضاً عند تعرضها للتلف.

إنّ الخلايا السرطانية لا تتصرف مثل الخلايا السليمة، حيث إن هذه الخلايا الشاذة لا تبقى على قيد الحياة فحسب، بل تنقسم وتتكاثر أيضاً على الرغم من أن الجسم لا يحتاج إليها. وهذا يخلق المزيد من الخلايا الشاذة المشابهة، والتي تشكل بعد ذلك ورماً.

يمكن أن يكون الورم غير سرطاني (حميد) أو سرطاني (خبيث). تتكون الأورام الحميدة من خلايا تشبه إلى حد كبير الخلايا الطبيعية، وتنمو ببطء، ولا تغزو الأنسجة القريبة أو تنتشر إلى أجزاء أخرى من الجسم.

بينما يمكن للأورام الخبيثة، إذا تركت دون علاج، أن تنتشر خارج الورم الأصلي إلى مناطق أخرى من الجسم.

يبدأ سرطان الثدي إما في خلايا الغدد المنتجة للحليب (وتسمى الفصيصات) أو في الممرات التي تفرز الحليب من الفصيصات إلى الحلمة (وتسمى القنوات). وفي حالات أقل شيوعاً، يمكن أن يبدأ سرطان الثدي في الأنسجة الضامة الدهنية والليفية للثدي (وتسمى الأنسجة اللحمية).

اعتمادًا على مرحلة سرطان الثدي، يمكن للخلايا السرطانية أن تغزو أنسجة الثدي السليمة القريبة وتشق طريقها إلى العقد الليمفاوية تحت الإبط.

الغدد الليمفاوية هي أعضاء صغيرة تقوم بتصفية المواد الغريبة في الجسم. فإن وصلت الخلايا السرطانية إلى العقد الليمفاوية، فيمكنها الانتقال عبر السائل الليمفاوي إلى أجزاء أخرى من الجسم.

يحدث سرطان الثدي دائماً بسبب حدوث خطأ في المادة الوراثية (يُسمى الشذوذ الجيني). ومع ذلك، فإن 5% إلى 10% فقط من حالات السرطان مرتبطة بالشذوذات الجينية الموروثة من أحد الوالدين. حوالي 85% من حالات سرطان الثدي ناجمة عن تشوهات وراثية نتيجة التقدم في السن وتدهور الحياة بشكل عام [6].

بالنسبة لأنواع سرطان الثدي: هناك العديد من الأنواع المختلفة لسرطان الثدي، ويتم تحديد كل منها تبعاً للموقع الذي يبدأ فيه النمو في الثدي، مدى نمو السرطان أو انتشاره، أو تبعاً لسمات معينة تؤثر على كيفية تصرف السرطان. إن معرفة نوع سرطان الثدي الذي يتم تشخيصه يساعد على اختيار أفضل خيارات العلاج. تبعاً لموقع السرطان يمكن تقسيم إلى سرطان غازي أو سرطان غير غازي.

أما سرطان الثدي لدى الذكور أمر نادر الحدوث، لكنه يحدث. يتم تشخيص أقل من 1% من جميع حالات سرطان الثدي لدى الرجال. معظم سرطانات الثدي عند الذكور هي سرطانات الأقنية الغازية، تشمل الأنواع ما يلي [4, 7]:

1.4.1. سرطان الثدي الغازي:

عندما يطلق على سرطان الثدي اسم غازي (أو متسلل)، فهذا يعني أنه انتشر في أنسجة الثدي المحيطة. يتم تحديد النوعين الأكثر شيوعاً من سرطان الثدي الغازي من خلال مكان بدء نموهما في الثدي وهما:

1.1.4.1. سرطان الأقنية الغازي Invasive ductal carcinoma (IDC):

يبدأ في قنوات الحليب، وهي الأنابيب التي تحمل الحليب من الفصيصات إلى الحلمة. وهو النوع الأكثر شيوعاً من سرطان الثدي؛ حوالي 80٪ من جميع حالات سرطان الثدي هي سرطانات الأقنية الغازي.

2.1.4.1. السرطان الفصيصي الغازي Invasive lobular carcinoma (ILC):

يبدأ في الفصيصات، وهي الغدد الموجودة في الثدي والتي تنتج الحليب. وهو ثاني أكثر أنواع سرطان الثدي شيوعاً؛ حوالي 10% من جميع حالات سرطان الثدي الغازية هي سرطانات مفصصة غازية.

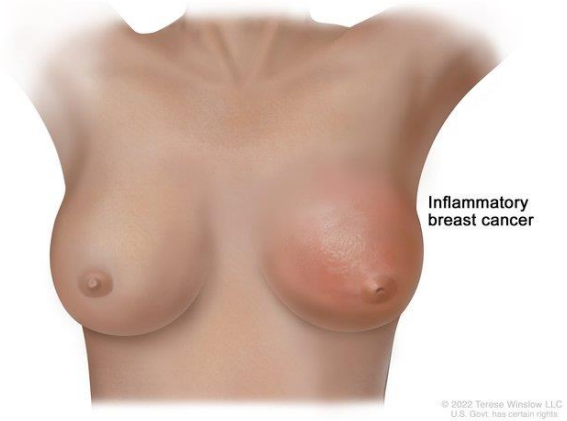
تتميز بعض أنواع سرطان الثدي الغازي بسمات تؤثر على كيفية تطورها وكيفية علاجها.

3.1.4.1. سرطان الثدي الثلاثي السلبي Triple-negative breast cancer:

هو نوع عدواني من سرطان الثدي الغازي الذي يكون اختباره سلبياً لمستقبلات هرمون الاستروجين ومستقبلات البروجسترون ولا يحتوي على بروتينات HER2 إضافية. حوالي 12% من جميع حالات سرطان الثدي الغازية تكون سلبية ثلاثية.

4.1.4.1. سرطان الثدي الالتهابي Inflammatory breast cancer:

هو نوع نادر وعدواني من سرطان الثدي الغازي. حوالي 1% من جميع حالات سرطان الثدي في الولايات المتحدة هي سرطان الثدي الالتهابي.



الشكل 2: سرطان الثدي الالتهابي [8].

5.1.4.1. سرطان الثدي النقائلي Metastatic breast cancer:

يُطلق عليه أيضاً سرطان الثدي في المرحلة الرابعة، هو سرطان ثدي غازي انتشر إلى مناطق بعيدة عن الثدي، مثل العظام، الكبد، الرئتين، الدماغ. يمكن أن يعود سرطان الثدي إلى مناطق أخرى من الجسم بعد أشهر أو سنوات من العلاج (يسمى التكرار النقائلي).

5.1.4.1. سرطان الثدي المتكرر Recurrent breast cancer:

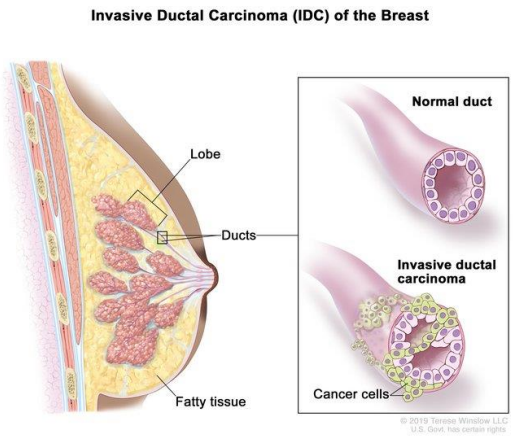
هو سرطان الثدي الغازي الذي يعود بعد أشهر أو سنوات من العلاج. يمكن أن يتكرر سرطان الثدي في نفس الثدي (تكرار موضعي)، أو العقد الليمفاوية القريبة في الإبط أو الترقوة (تكرار إقليمي)، أو في جزء آخر من الجسم (تكرار نقائلي أو بعيد).

2.4.1. سرطان الثدي غير الغازي:

أي أن الورم لم ينتشر خارج أنسجة الثدي. هناك نوعان رئيسيان من سرطان الثدي غير الغازي:

1.2.4.1. سرطان الأقنية الموضعي Ductal carcinoma in situ (DCIS):

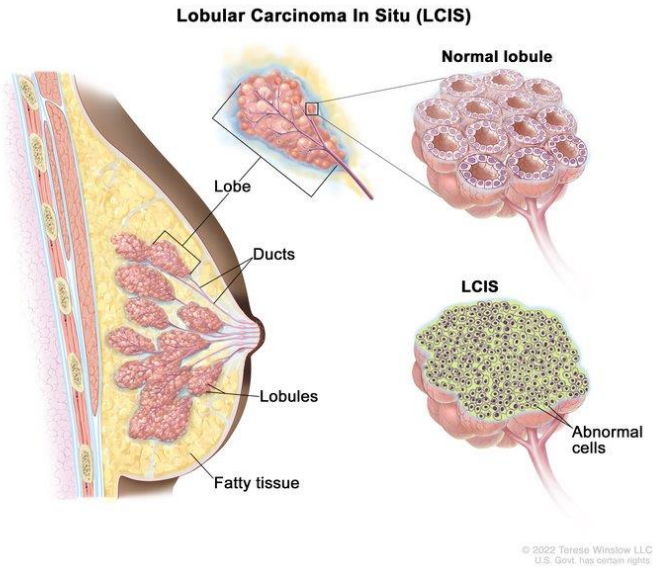
هو سرطان غير جراحي لم ينتشر خارج قنوات الحليب. لا يشكل DCIS تهديداً للحياة ولكنه يعتبر عامل خطورة لتطور سرطان الثدي الغازي. حوالي 16% من جميع أنوع سرطان الثدي المشخصة هي سرطان القنوات الموضعي (DCIS).



الشكل 3: سرطان الأقنية الموضعي [9].

2.2.4.1. السرطان الفصيصي الموضعي Lobular carcinoma in situ (LCIS):

هو سرطان ثدي غير جراحي لم ينتشر خارج الفصيصات التي بدأ فيها. على الرغم من اسمه، فإن LCIS هو حالة ثدي حميدة وليس سرطان ثدي حقيقي.



الشكل 4: السرطان الفصيصي الموضعي [10].

5.1. مراحل سرطان الثدي:

تشمل مراحل سرطان الثدي ما يلي [11]:

1.5.1. المرحلة 0:

تكون هذه المرحلة غير غازية للورم. يبدأ الورم في النمو ولم يتم العثور على دليل على وجوده الغزو في الأنسجة المحيطة. من الأمثلة على هذه المرحلة هو سرطان الخلايا القنوية (DCIS).

2.5.1. المرحلة 1:

توصف هذه المرحلة بأنها سرطان ثدي غازي، والغزو المجهري ممكن في هذه المرحلة. يصنف بدوره إلى فئتين هما المرحلة A1 و B1.

المرحلة A1 تصف الورم الذي يصل حجمه إلى 2 سم ولا تشارك فيه العقد الليمفاوية، بينما يصف الشكل B1 الخلايا السرطانية الأكبر حجماً، وتم العثور على 0.2 ملم منه في العقدة الليمفاوية.

3.5.1. المرحلة 2:

تحتوي المرحلة الثانية أيضاً على شكلين A2 وB2.

تصف المرحلة A2 أن الورم موجود في العقد الليمفاوية الإبطية أو في الغدد الليمفاوية. يمكن أن يكون الورم أصغر أو أكبر من 2 سم ولكن لا يزيد عن 5 سم. ومع ذلك، خلال المرحلة B2 يمكن أن يكون حجم الورم أكبر من 5 سم ولكنه لا يستطيع الوصول إلى الغدد الليمفاوية الإبطية.

4.5.1. المرحلة 3:

يتم تقسيمها إلى ثلاث فئات فرعية هي A3، B3 وC3.

المرحلة A3 لا يتم العثور على الورم في الثدي ولكن يمكن العثور عليه في الغدد الليمفاوية الإبطية.

يصف الشكل B3 أن الورم يمكن أن يكون بأي حجم ويمكن العثور عليه في العقد الليمفاوية الإبطية.

يمكن اعتبار المرحلة B3 التهابية حيث يكون جلد الثدي محمر ومتورم. لكن المرحلة C3 تصف انتشار ورم يصل إلى 10 أو أكثر من 10 عقد ليمفاوية إبطية.

5.5.1. المرحلة 4:

هي المرحلة المتقدمة والمنتشرة من السرطان وتصف هذه المرحلة انتشار المرض إلى الأعضاء أخرى في الجسم (الرئتين، العظام، الكبد، الدماغ).

6.1. الأعراض والتغييرات المرافقة:

وفقا لجمعية السرطان الأمريكية، فإن أي من التغييرات غير الطبيعية التالية في الثدي يمكن أن تكون من أعراض سرطان الثدي [12]:

- ✓ تورم الثدي بأكمله أو جزء منه
- ✓ تهيج الجلد
- ✓ ألم في الثدي
- ✓ ألم في الحلمة أو انقلاب الحلمة إلى الداخل
- ✓ احمرار أو تقشر أو سماكة في جلد الحلمة أو الثدي
- ✓ إفرازات من الحلمة غير حليب الثدي
- ✓ كتلة في منطقة الإبط

على الرغم من أن أعراض سرطان الثدي تختلف بشكل كبير، إلا أن العديد من أنواع سرطان الثدي ليس لها أعراض واضحة على الإطلاق.

في بعض الحالات، قد تكون الكتلة صغيرة جداً بحيث لا يمكنك الشعور بها. في كثير من الأحيان، تظهر منطقة غير طبيعية لدى اجراء تصوير للثدي بالأشعة السينية (الأشعة السينية للثدي)، مما يدفع إلى إجراء اختبارات إضافية.

وفي حالات أخرى، تكون العلامة الأولى لسرطان الثدي عبارة عن كتلة جديدة في الثدي يمكن جسها. من المرجح أن تكون الكتلة غير المؤلمة، الصلبة، وذات الحواف غير المستوية سرطاناً، لكن في بعض الأحيان يمكن أن تكون الكتلة طرية، ناعمة، ومستديرة.

إن إجراء فحص ذاتي شهري للثدي هو أفضل طريقة لملاحظة أي تغييرات في الثدي. ومن المهم أن يقوم الطبيب بفحص أي تغييرات تظهر في الثدي في أسرع وقت ممكن.

7.1. عوامل الخطورة:

سرطان الثدي هو مرض معقد يمكن أن يتأثر بعدد من العوامل المختلفة. من بين العوامل الرئيسية التي تزيد من احتمال إصابة النساء بسرطان الثدي هي العوامل الوراثية، حيث يمكن أن تزيد وجود تاريخ عائلي للمرض من خطر الإصابة. عوامل الهرمونات أخرى تلعب أيضاً دوراً هاماً، حيث يمكن أن يزيد ارتفاع مستويات هرمون الاستروجين في الجسم من خطر الإصابة بسرطان الثدي. العوامل البيئية مثل التدخين والتغذية غير الصحية ونقص النشاط البدني يمكن أن تزيد أيضاً من خطر الإصابة بالمرض [13].

8.1. العلاج:

1.8.1. الجراحة:

تستخدم الجراحة لإزالة الأورام السرطانية من الثدي أو لإزالة الثدي بشكل جزئي أو كلي، ويعتمد نطاق العملية على مرحلة المرض، حجم الورم، ومكانه. قد تتضمن هذه العمليات إزالة العقد اللمفاوية المجاورة أيضاً. يمكن أن تتضمن الجراحة استخدام تقنيات متقدمة مثل الجراحة بالتنظير أو إعادة بناء الثدي بعد إزالته جزئياً. يتعاون الأطباء والجراحون لاختيار النهج الجراحي المناسب لكل حالة مرضية، ويسعون إلى الحفاظ على الجودة والجمال والراحة للمريضة خلال هذا العلاج.

تأتي الجراحة كجزء من العلاج المتعدد الأوجه الذي يمكن أن يشمل العلاج الإشعاعي والكيميائي والهرموني بعد العملية لضمان أقصى استفادة من العلاج وزيادة فرص النجاح في معالجة سرطان الثدي [14, 15].

2.8.1. العلاج الكيميائي:

هو إحدى الأساليب الرئيسية للتعامل مع هذا المرض. يعتمد العلاج الكيميائي على استخدام أدوية مضادة للسرطان لتدمير ومكافحة الخلايا السرطانية في الثدي والحد من انتشارها إلى مناطق أخرى في الجسم. يمكن تناول هذه الأدوية عن طريق الفم أو عن طريق الوريد، ويتم تخصيص نوع وجرعة الكيميائي بناءً على نوع السرطان ومرحلته واحتياجات المريضة.

يمكن أن يترافق العلاج الكيميائي مع آثار جانبية مؤقتة مثل فقدان الشعر، التعب، والغثيان، ولكن عادة ما تكون هذه الأعراض مؤقتة ويمكن التعامل معها من خلال الرعاية الطبية [14, 15].

3.8.1. العلاج الشعاعي:

يعتمد هذا النهج على استخدام أشعة عالية الطاقة لاستهداف وتدمير الخلايا السرطانية في منطقة الثدي المصابة. يهدف العلاج بالإشعاع إلى الحد من نمو الورم السرطاني ومنع انتشاره إلى الأنسجة المجاورة، وذلك بعد الجراحة لإزالة الورم أو قبلها لتقليل حجم الورم. تُجرى عمليات الإشعاع بعناية فائقة لتقليل التأثير على الأنسجة الصحية المحيطة بالورم، ويتم تحديد جرعة الإشعاع وجدولها وفقاً لمرحلة المرض واحتياجات المريضة.

يمكن أن تترافق جلسات العلاج بالإشعاع مع آثار جانبية مؤقتة مثل التعب واحمرار جلد الثدي، وتتراجع هذه الأعراض بعد انتهاء العلاج. يلعب العلاج الشعاعي دوراً مهماً في تحقيق الشفاء وزيادة فرص النجاح في علاج سرطان الثدي[16].

4.8.1. العلاج الهرموني:

هو إحدى استراتيجيات العلاج المهمة للنساء اللاتي يعانين من أنواع معينة من سرطان الثدي. يعتمد هذا النهج على منع تأثير الهرمونات الأنثوية مثل الاستروجين على نمو وتطور الخلايا السرطانية. يتم ذلك من خلال إما تقليل إنتاج هذه الهرمونات في الجسم أو منع تفاعلها مع الخلايا السرطانية.

يتم استخدام العلاج الهرموني في الغالب لعلاج سرطان الثدي الذي يكون إيجابي لمستقبل الاستروجين(ER-positive) أو إيجابي لمستقبل للبروجسترون (PR-positive).

تُعتبر هذه العلاجات فعّالة في تقليل نسبة عودة الورم وتحسين نتائج العلاج، ويتم تحديد نوع العلاج ومدى فعاليته بناءً على خصائص المريضة ونوع الورم ومرحلته [14, 15].

5.8.1. العلاج المستهدف:

هو استراتيجية حديثة تعتمد على فهم عميق للخصائص الجينية والجزيئية للورم الثدي. يهدف هذا النوع من العلاج إلى استخدام الأدوية والعلاجات المستهدفة لاستهداف خلايا الورم بدقة، مما يقلل من تأثيرها على الأنسجة السليمة المحيطة. يمكن أن تشمل تقنيات العلاج المستهدف تثبيط البروتينات الضارة أو إيقاف الإشارات الخاصة بنمو الورم.

يتيح هذا النهج للأطباء تخصيص العلاج بشكل أكثر فعالية لاحتياجات كل مريض بناءً على السمات الفردية لورمهم، مما يعزز من فعالية العلاج ويقلل من الآثار الجانبية [17].

6.8.1. العلاج المناعي:

يعتمد هذا العلاج على تعزيز جهاز المناعة للجسم لمكافحة الخلايا السرطانية. تتضمن تقنيات العلاج المناعي استخدام الأجسام المضادة واللقاحات التي تستهدف البروتينات الموجودة على سطح الخلايا السرطانية، مما يعزز استجابة المناعة ضد الورم. يعتبر هذا النهج واعداً بسبب قدرته على تحسين نتائج علاج سرطان الثدي وزيادة فرص البقاء على قيد الحياة للمرضى، وذلك دون الحاجة إلى تدمير الأنسجة السليمة بنفس القدر الذي يحدث في بعض العلاجات الأخرى [18].

9.1. تعلم الآلة: مقدمة عامة

التعلم الآلي هو فرع من أفرع الذكاء الصنعي AI (Artificial intelligence) 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. يشمل ذلك مجموعة متنوعة من التقنيات مثل الشبكات العصبية الاصطناعية، والتعلم العميق، والتصنيف الآلي وغيرها.

يُستخدم التعلم الآلي في مجموعة واسعة من التطبيقات مثل التعرف على الصور، معالجة اللغة الطبيعية، التنبؤ، التحكم في الروبوتات، الطب والكثير من المجالات الأخرى. يشهد هذا المجال تطوراً مستمراً ويعد أحد أهم التكنولوجيات في العصر الحديث، حيث يسهم في تحسين أداء الأنظمة وتحسين القرارات بشكل كبير [19].

10.1. أنواع تعلم الألة:

يوجد العديد من طرق تعلم الآلة من أهمها:

1. التعلم الموجّه supervised
2. التعلّم غير الموجّه unsupervised
3. التعلّم المعزّز reinforcement

1.10.1. التعلم الموجه:

يعد التعلم الموجه أحد أبسط أنواع التعلم الآلي. في هذا النوع، يتم تدريب خوارزمية التعلم الآلي على مجموعة من البيانات. تعد مجموعة بيانات التدريب هذه جزءاً صغيراً من مجموعة البيانات الأكبر وتعمل على إعطاء الخوارزمية فكرة أساسية عن المشكلة والحل ونقاط البيانات التي يجب التعامل

معها. ثم تعثر الخوارزمية بعد ذلك على العلاقات بين البارامترات المعطاة، وتؤسس بشكل أساسي علاقة السبب والنتيجة بين المتغيرات في مجموعة البيانات. في نهاية التدريب، يكون لدى الخوارزمية فكرة عن كيفية عمل البيانات والعلاقة بين المدخلات والمخرجات [19].

2.10.1. التعلم غير الموجه:

في هذه الطريقة يتم التعلم بدون وجود مجموعة تدريب، حيث تدرك الخوارزمية العلاقات بين نقاط البيانات بطريقة مجردة، دون الحاجة إلى مدخلات من البشر [19].

3.10.1. التعلم المعزز:

يحاكي التعلم المعزز طريقة تعلم البشر من البيانات في حياتهم. إنه يتميز بخوارزمية تعمل على تحسين نفسها وتتعلم من المواقف الجديدة باستخدام طريقة التجربة والخطأ.

استناداً إلى المفهوم النفسي للتكييف، يعمل التعلم المعزز من خلال وضع الخوارزمية في بيئة عمل مفسِّر ونظام مكافأة. في كل تكرار للخوارزمية، يتم إعطاء نتيجة الإخراج للمفسِّر، والذي يقرر ما إذا كانت النتيجة مواتية أم لا. في حالة عثور البرنامج على الحل الصحيح، يقوم المفسِّر بتعزيز الحل من خلال تقديم مكافأة للخوارزمية.، وإذا كانت النتيجة غير مناسبة، تضطر الخوارزمية إلى التكرار حتى تحصل على نتيجة أفضل [19].

11.1. خوارزميات التصنيف:

يمكن تعريف عمليّة التصنيف ضمن بيئة تعلّم الآلة بأنّها عمليّة توزيع البيانات والتي تسمّى بيانات التدريب، ضمن فئات مختلفة حسب خاصيتها المشتركة بالاعتماد على خوارزميّات متعدّدة. تدرج عمليّة التصنيف ضمن عمليّات التعلّم الموجّه، وتعتبر عمليّة التصنيف أساسا لعمليّة التنبّؤ من خلال النماذج التي يتمّ بناؤها أثناء عمليّة التصنيف والمرتبطة بنوع المصنّف المستخدم، وفيما يلي نستعرض المصنّفات التي تمّ استخدامها ضمن هذا البحث.

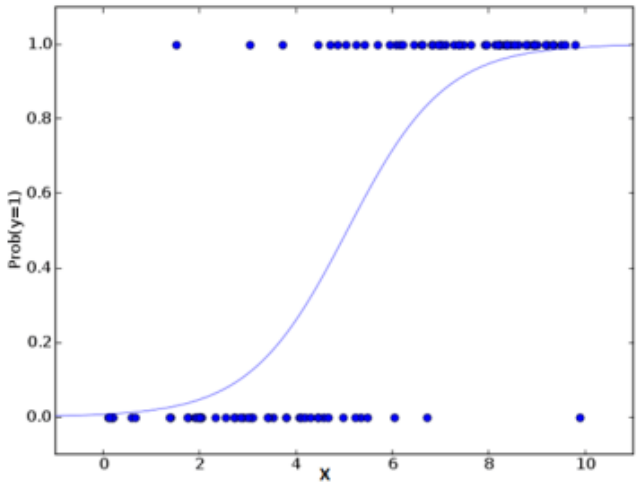
1.11.1.الانحدار اللوجستي Logistic Regression:

يستخدم الانحدار اللوجستي LOG الدالة السينية sigmoid function، والتي تأخذ أي قيمة حقيقية بين صفر و واحد. يتم تعريفها رياضيا على الشكل التالي:

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

حيث t هو الدخل.

إذا رسمناها بيانياً، فسيكون الرسم البياني على شكل منحنى حرف S لذلك تسمى بالدالة السينية (الشكل 5) [20, 21].



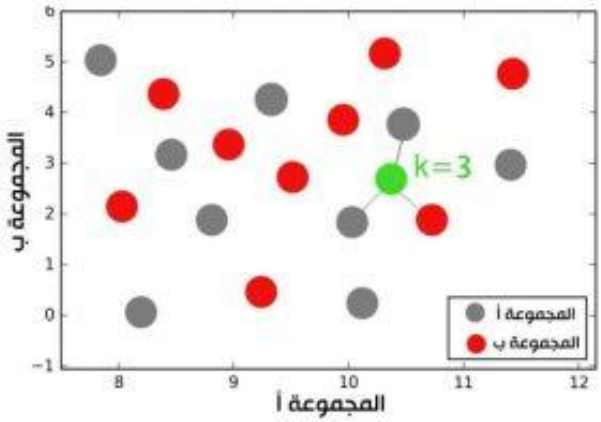
الشكل 5: مثال على التنبؤ باحتمالية الفئة باستخدام الانحدار اللوجستي [21].

2.11.1. خوارزمية الجار الأقرب K-nearest neighbor:

تعتبر خوارزمية الجار الأقرب kNN إحدى أهم وأبسط خوارزميات تعلم الآلة الموجه. كما أن لها القدرة على التعامل مع البيانات الشاذة أو القيم المتطرفة بكفاءة عالية .

يعتمد مبدأ عمل هذه الخوارزمية على حساب المسافة الاقليدية بين النقاط، حيث كلما قلت المسافة بين نقطتين زادت إحتمالية إنتماء النقطتين بعضهما لبعض و من هنا جاء اسم الخوارزمية، و يشير الحرف k إلى عدد العينات التي سيتم تصنيف نقطة ما بناء على المسافات بينها و بين جيرانها الذين يبلغ عددهم k.

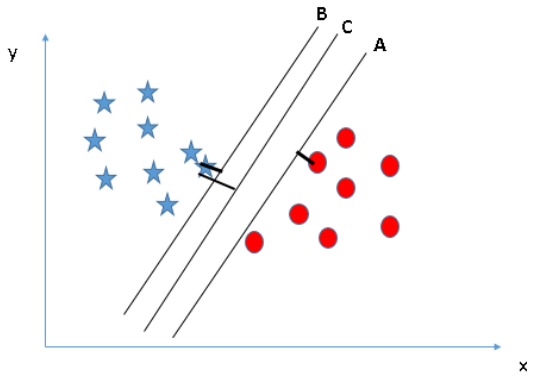
إذا افترضنا أن k = 3 فإن الخوارزمية تقيس المسافة بين النقطة المستهدفة وأقرب ثلاث نقاط إليها فإذا كانت أقرب نقطتين تنتمي للمجموعة أ والنقطة الثالثة لوحدها تنتمي للمجموعة ب فإن النقطة المستهدفة سيتم تصنيفها على أساس أنها تنتمي للمجموعة أ (شكل 6) [22].



الشكل 6: مثال على التنبؤ باحتمالية الفئة باستخدام KNN [22].

3.11.1. ألة المتجه الداعم Support Vector Machine:

خوارزمية آلة المتجه الداعم SVM هي خوارزمية تعلم آلي خاضع للإشراف يمكن استخدامها في مسائل التصنيف Classification أو التنبؤ. ومع ذلك، فإنها تستخدم في الغالب في مسائل التصنيف. في خوارزمية SVM، نرسم كل عنصر من عناصر البيانات كنقطة في الفضاء ذو بعد n (حيث n هو عدد السمات Features) مع قيمة كل سمة هي قيمة إحداثيات معينة. ثم، نقوم بإجراء التصنيف من خلال إيجاد المستو الفائق Hyper-plane الذي يميز الفئتين جيداً. تتمثل الفكرة الرئيسية لخوارزمية آلة المتجه الداعم SVM في العثور على حد القرار (المستوى الفائق) الذي يفصل إلى أقصى حد بين الفئات المختلفة في مساحة الميزة. يتم اختيار هذا المستوى الفائق ليكون المسؤول عن زيادة المسافة بين أقرب نقاط كل فئة. هذه النقاط الأقرب تعرف بمتجهات الدعم، وتعرف المسافة بينها بالهامش (الشكل 7) [19].



الشكل 7: مثال على التنبؤ باحتمالية الفئة باستخدام SVM [19].

- طريقة النواة Kernel method:

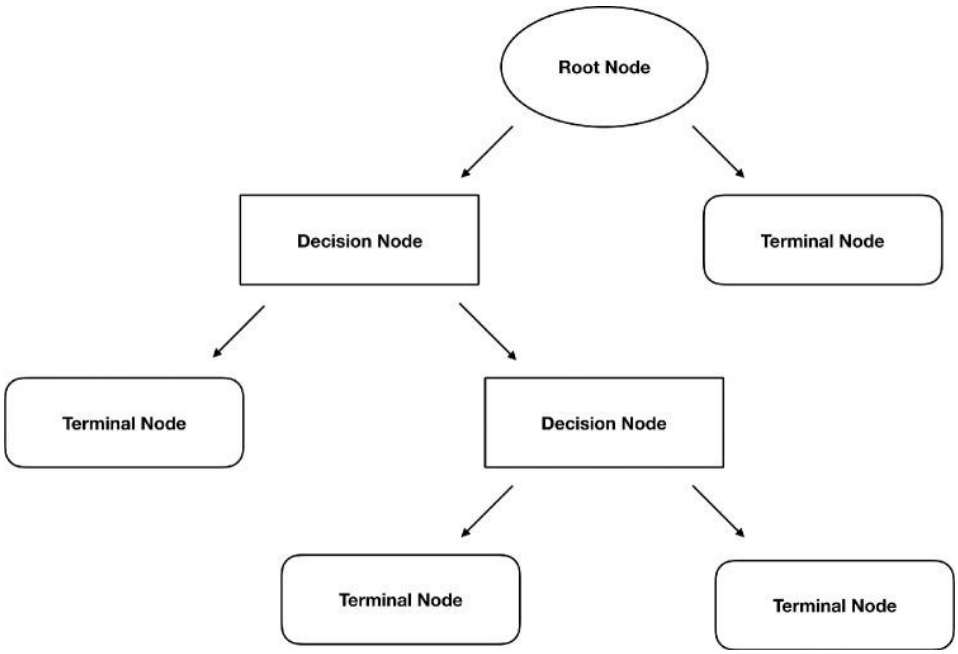
في الممارسة العملية، يمكن لخوارزمية آلة المتجه الداعم SVM معالجة البيانات غير القابلة للفصل خطياً عن طريق تعيين بيانات الإدخال إلى مساحة ذات أبعاد أعلى باستخدام طريقة النواة. يسمح هذا لخوارزمية آلة المتجه الداعم SVM بالعثور على المستوى الفائق الذي يمكنه من فصل الفئات.

طريقة النواة Kernal method هي عبارة عن أداة تعمل على تحويل المساحات ذي الأبعاد القليلة وتحوليها إلى مساحات ذي أبعاد متعددة (مثلا تحويل المستويات ذي البعدين الى مستويات ذي ثلاثة او اربعة أبعاد)، أي أنها تحول المسألة التي لا يمكن فصلها إلى مسألة قابلة للفصل. تفيد هذه الطريقة في الغالب في مسائل التصنيف غير الخطية. تتضمن بعض دوال النواة الشائعة دالة الأساس الخطية linear ومتعددة الحدود polynomial والشعاعية (RBF) والنواة السينية Sigmoid kernels.

4.11.1. خوارزمية شجرة القرار Decision Tree:

خوارزمية شجرة القرار DT هي خوارزمية من خوارزميات تعلم الآلة تقوم بتحليل البيانات وتقسيمها إلى فئات (أو تصنيفات) باستخدام سلسلة من القرارات المستندة إلى القيم الموجودة في مجموعة من المتغيرات. ويتم تمثيل هذه القرارات في شكل هيكل شجري، حيث يتم تقسيم البيانات على شكل فروع (Branches) وتصنيفات (Leaves)، ويتم اتخاذ القرارات بناءً على معايير محددة.

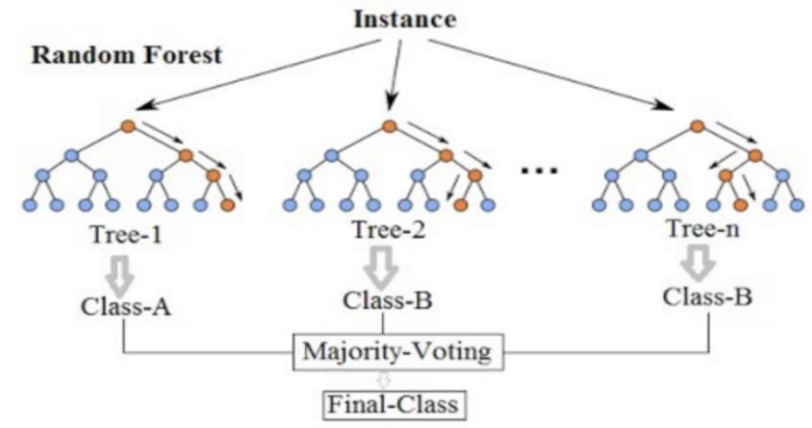
تبدأ شجرة القرار بعقدة رئيسيّة تدعى الجذر Root، مجموعة عقد داخليّة Nodes، ومجموعة عقد نهائيّة Terminals، بحيث يتضمّن كلّ من الجذر ومجموعة العقد الداخليّة القاعدة التي تحدّد مسار الفروع المرتبطة ممّا يسمح في النهاية إلى الوصول إلى النتيجة النهائيّة (الشكل 8) [24 ,23].



الشكل 8: هيكل شجرة القرار [22].

5.11.1. الغابة العشوائية Random Forest:

تتكون خوارزمية الغابة العشوائية RF، من عدد كبير من DT الفردية التي تعمل كمجموعة. حيث تنبثق كل شجرة فردية في الغابة العشوائية من توقع الصنف ويصبح الصنف الذي يحصل على أكبر عدد من الأصوات هو توقع النموذج (شكل 9) [25].



الشكل 9: هيكل الغابة العشوائية.

12.1. معايير التقييم:

من أهم معايير التقييم الأكثر شيوعاً المستخدمة في مسائل التصنيف [26]:

1.12.1. الدقة Accuracy:

الدقة هي مقياس التقييم الأكثر استخداماُ في مسائل التصنيف. حيث تقيس نسبة نقاط البيانات المصنفة بشكل صحيح من إجمالي عدد نقاط البيانات، يتم حسابها على النحو التالي:

$$\text{الدقة} = (TP + TN + FP + FN) / (TP + TN)$$

حيث:

- TP: هو عدد الإيجابيات الحقيقية
- TN: هو عدد السلبيات الحقيقية
- FP: هو عدد الإيجابيات الخاطئة
- FN: هو عدد السلبيات الخاطئة
-

2.12.1. الإنضباط Precision:

يعرف بأنه نسبة التنبؤات الإيجابية الحقيقية من جميع التوقعات الإيجابية، حيث يقيس قدرة النموذج على التنبؤ بالعينات الإيجابية بشكل صحيح، يتم حسابه على النحو التالي:

$$\text{الإنضباط} = TP / (TP + FP)$$

3.12.1. الاستدعاء Recall:

معروف أيضاً باسم الحساسية، هو نسبة التنبؤات الإيجابية الصحيحة من جميع العينات الإيجابية الفعلية. كذلك يقيس قدرة النموذج على تحديد العينات الإيجابية بشكل صحيح، ويتم حسابه على النحو التالي:

$$\text{الاستدعاء} = TP / (TP + FN)$$

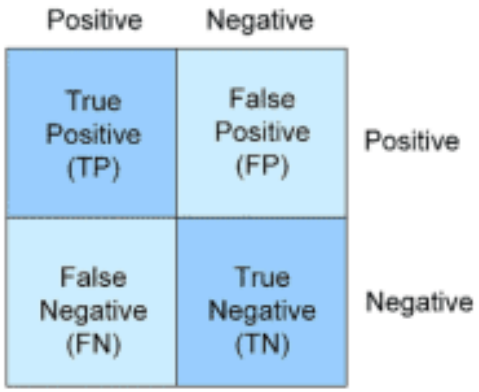
4.12.1. درجة F1:

مقياس درجة F1 هي المتوسط التوافقي للدقة والاستدعاء. كذلك، تعتبر درجة F1 مقياس شائع الاستخدام عندما يكون كل من الدقة والاستدعاء مهمين. يتم حساب درجة F1 على النحو التالي:

$$\text{درجة F1} = \frac{1}{\left(\frac{1}{\text{الإستدعاء}} + \frac{1}{\text{الإنضباط}}\right)\frac{1}{2}}$$

5.12.1. مصفوفة الشك Confusion matrix:

مصفوفة الارباك هي جدول يلخص أداء نموذج التصنيف. على سبيل المثال، تعرض مصفوفة الإرباك عدد الإيجابيات الصحيحة والسلبيات الصحيحة والإيجابيات الخاطئة والسلبيات الخاطئة. من ناحية أخرى تعد هذه المصفوفة مفيدة لتصور أداء النموذج وتحديد مجالات التحسين.



الشكل 10: مصفوفة الشك [26].

الفصل الثاني: الدراسات المرجعية
Review of literature

1.2. مقدمة عامة:

استخدام تعلم الآلة في التنبؤ بنوع سرطان الثدي يمثل نقلة مهمة في مجال الرعاية الصحية. يعتمد هذا النهج الحديث على قدرة الأنظمة الذكية على تحليل البيانات السريرية والتشخيصية بشكل دقيق لتوقع احتمالية الإصابة بسرطان الثدي. يعتمد هذا التنبؤ على استخدام البيانات الكبيرة وتقنيات التعلم العميق للكشف المبكر عن التغييرات المرتبطة بالورم، مما يمنح الأطباء والمرضى أدوات فعالة لتخطيط وتوجيه العلاج بشكل أكثر دقة. في هذا السياق، يمكن لتعلم الآلة أن يلعب دوراً مهماً في تحسين فرص العلاج وزيادة معدلات البقاء على قيد الحياة لمرضى سرطان الثدي.

من أهم الدراسات المرجعية التي استخدمت تعلم الآلة للتنبؤ بسرطان الثدي نذكر ما يلي:

2.2. **الدراسة الأولى:**

بعنوان:

Machine Learning Techniques for Breast Cancer Prediction

تم اجراؤها من قبل Nemade وزملائه عام 2023 [27]، حيث تم استخدام هذه الدراسة لتحديد ما إذا كان الورم حميداً أم خبيثاً اعتماداً على نصف القطر، الملمس، المحيط، المساحة، النعومة، الاكتناز، التقعر، التجانس، والبعد الكسري للورم.

تم استخدام خوارزميات تصنيف مختلفة بما فيها (Naïve Bayes (NB، الانحدار اللوجستي (LR) (SVM)، (K-Nearest Neighbor (KNN، (DT)، و (RF) على مجموعة بيانات السرطان، وتم بعد ذلك تقييم أداء هذه الخوارزميات.

وجد أن شجرة القرار كانت ذو الأداء الأفضل بأعلى دقة بلغت 97% من بين جميع النماذج.

3.2. **الدراسة الثانية:**

بعنوان:

Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms

تم اجراؤها من قبل Ara S وزملائه عام 2021 [28]، حيث تم استخدام في هذه الدراسة خوارزمية SVM، LR، KNN، DT، وRF وذلك من أجل تصنيف الورم الخبيث والحميد بالاعتماد على عدة معايير ورمية أهمها نصف القطر، الملمس، المحيط، المساحة، النعومة، الاكتناز، التقعر، التجانس، والبعد الكسري للورم.

تم مقارنة دقة جميع الخوارزميات من أجل إيجاد النموذج الذي يحقق الدقة الأفضل.

وجد أن SVM وRF حققا الأداء الأفضل من بين جميع الخوارزميات بدقة وصلت حتى 96.5%.

4.2. الدراسة الثالثة:

بعنوان:

An Explainable Artificial Intelligence Model for the Classification of Breast Cancer

تم اجراؤها من قبل Khater T وزملائه عام 2023 [29]، حيث تم استخدام KNN، ANN، XG boost، SVM، وRF.

كان الهدف الأساسي من هذه الدراسة هو تصنيف الورم إلى حميد أو خبيث بالاعتماد على بيانات رقم 1 تابعة لـ Wisconsin خاصة بسرطان الثدي وبيانات رقم 2 تابعة لـ Wisconsin خاصة بتشخيص سرطان الثدي.

حققت خوارزميتي KNN وANN الدقة الأعلى بدقة تراوحت 97.7% و98.2% على التوالي للبيانات رقم 1، ودقة تراوحت 98.6% و94.4% للبيانات رقم 2.

5.2. الدراسة الرابعة:

بعنوان:

Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques

تم اجراؤها من قبل Islam M وزملاؤه عام 2020 [30].

كان الهدف من هذه الدراسة هي الكشف المبكر عن سرطان الثدي حيث تم استخدام العديد من الخوارزميات وهي SVM، ANN، وLOG باستخدام مجموعة بيانات سرطان الثدي في ولاية ويسكونسن من قاعدة بيانات التعلم الآلي UCI.

قياس أداء الخوارزميات باستخدام الدقة، الحساسية، النوعية، والدقة، F1.

أظهرت النتائج أن الشبكات العصبية الاصطناعية ANN حصلت على أعلى مستوى من الدقة الحساسية وF1 بنسبة وصلت حتى 98.57%، 97.82%، و% 98.90 على التوالي، في حين كانت خوارزمية SVM في المرتبة الثانية بنسبة وصلت إلى 97.14%، 95.65%، و% 97.77.

6.2. الدراسة الخامسة:

بعنوان:

Early predictive model for breast cancer classification using blended ensemble learning

تمت هذه الدراسة من قبل R M وزملائه عام 2023 [31]، وكان الهدف منها هو بناء نموذج قادر على التنبؤ المبكر بسرطان الثدي للحد من الفيات قدر الإمكان.

تم استخدام في هذه الدراسة التعلم الجماعي وهي تقنية تعمل على اتباع خمس خوارزميات للتعلم الآلي وهي SVM، KNN، LOG، RF، وDT كمتعلمين أساسيين وتم مقارنة دقة جميع المتعلمين الأساسيين المدمجين (فردياً) والنتيجة النهائية للتعلم الجماعي في هذه الدراسة باستخدام العديد من مقاييس الأداء وهي الدقة والاستدعاء والدقة ودرجة f1 للتنبؤ المبكر بسرطان الثدي. لوحظ وجود تحسن ملحوظ بنسبة 98.14 % في نموذج التعلم الجماعي مقارنة بالمتعلمين الأساسيين.

الفصل الثالث: المواد وطرائق البحث
Material and Methods

1.3. عيّنة الدراسة:

تم استيراد البيانات من قاعدة البيانات Kaggle والتي تم جمعها في الأساس من قبل الطبيب Dr. William H. Wolberg في مستشفى جامعة Wisconsin في الولايات المتحدة الأمريكية.

استخدم الطبيب وولبرغ لجمع هذه البيانات عينات سوائل من المرضى الذين تم تشخيصهم بوجود كتل صلبة في الثدي واستعان ببرنامج Xcyt وهو برنامج حاسوبي لإجراء الرسوم قادر على إجراء تحليل للسمات الخلوية بناءً على المسح الرقمي. يستخدم البرنامج خوارزمية ملائمة لحساب عشر سمات features من كل خلية في العينة، ثم يحسب القيمة المتوسطة والقيمة القصوى والخطأ المعياري لكل سمة، فنحصل في النتيجة على 30 متجهاً ذو قيمة حقيقية.

تتكون البيانات من 569 عينة لمريضات مصابات بسرطان الثدي.

المتغير الهدف (y): هو التشخيص (M= خبيث، B= حميد)

و10 متغيرات (X) تصف نواة الخلايا:

1. نصف القطر (متوسط المسافات من المركز إلى النقاط الموجودة على المحيط)
2. الملمس (الانحراف المعياري لقيم التدرج الرمادي)
3. المحيط
4. مساحة المنطقة
5. النعومة (الاختلاف المحلي في أطوال نصف القطر)
6. الاكتناز (التراص) (المحيط ^ 2 / المساحة – 1.0)
7. التقعر (شدة التقعر من المحيط)
8. النقاط المقعرة (عدد الأجزاء المقعرة)
9. التجانس
10. البعد الكسري

ولكل سمة ثلاثة مقاييس:

أ. المتوسط

ب. الخطأ المعياري

ج. القيمة القصوى (متوسط القيم الثلاث الكبرى)

2.3. الدراسة العمليّة:

تمّ العمل على مرحلتين:

1.2.3. المرحلة الأولى:

تضمّنت استخدام البرنامج الإحصائي الـ SPSS الإصدار 26.0 لتوصيف متغيّرات الدراسة ودراسة الارتباطات والعلاقات الهامّة بين المتغيّرات.

تم استخدام معامل ارتباط بيرسون لدراسة الارتباط بين المتغيرات الكمية من أجل حذف المتغيّرات المرتبطة ارتباطاً قوياً وذلك لتجنّب التكرار وانحياز الخوارزميّة لمتغيّر معيّن، حيث تمّ دراسة الارتباطات بين متوسطات المتغيرات العشرة، وتمّ تحديد الارتباطات الهامة عند عتبة عامل ارتباط بيرسون> 0.9، وتم حذف المتغيرات المرتبطة مع قيمة الخطأ المعياري والقيمة القصوى التي تقابلها.

ثمّ استخدم اختبار T-Student Test لمقارنة متوسّطات المتغيّرات الكميّة بين نوعي السرطان.

عدّت الفروق هامّة إحصائياً عند عتبة الدلالة ($P<0.05$) في الاختبارات المطبّقة.

2.2.3. المرحلة الثانية:

تضمّنت العمل على Google Colaboratory وهو منصّة برمجيّة تستخدم لكتابة التعليمات البرمجيّة العلميّة بلغة البايثون لبناء النماذج المختلفة.

تمّ في هذه المرحلة استخدام البيانات الناتجة عن المعالجة في المرحلة الأولية، حيث تمّ توحيد البيانات Standardization ; حيث تستخدم هذه التقنيّة لتجنّب إعطاء وزن أكبر للبيانات ذات القيمة الأكبر لذلك يتمّ تحويل كل المتغيّرات إلى مقياس موحّد جديد متوسطه 0 وانحرافه المعياري 1.

تمّ استخدام العديد من المكتبات الخاصّة بلغة البايثون، نستعرض في الجدول 1 أهمّ هذه المكتبات:

جدول 1: أهمّ المكتبات المستخدمة في المشروع.

المكتبة	الوصف
Numpy	مكتبة متخصّصة في الحوسبة العلميّة بلغة البايثون، وتحتوي على تشكيلة متنوّعة من الأدوات والتقنيّات التي من الممكن استخدامها لحلّ المشاكل الرياضيّة.
Pandas	مكتبة متخصّصة بإجراء ما يسمّى بـ data Munging، والمقصود به هو إجراء تغييرات على بيانات أساسيّة غير مرتّبة Raw DATA بحيث ينتج عن هذا التغيير تحويل البيانات إلى شكل آخر يمكن فهمه والتعامل معه وهو Data frame.
SKlearn	مكتبة رئيسيّة تستخدم عادة في مشاريع تعلّم الآلة، تحتوي على العديد من الخوارزميات والطرق المستخدمة في مجال تعلّم الآلة مثل التصنيف Classification، بالإضافة لاستخدامها في مرحلة معالجة البيانات وتقييم النماذج.
seaborn	مكتبة مخصّصة للرسوم البيانيّة وبناء واجهات متقدّمة.

تم بعد ذلك ببناء النموذج باستخدام عدة خوارزميات منها: خوارزمية الجار الأقرب k-nearest neighbors، الانحدار اللوجستي Logistic Regression (LR)، آلة متجه الدعم Support Vector Machine (SVM)، شجرة القرار Decision Tree (DT)، والغابة العشوائية Random Forest (RF).

وفي النهاية تم استخدام مصفوفة الشك (الارتباك) لتقييم أداء الخوارزميات.

حيث تم العمل وفق منهجين:

المنهج الأول: تم تقسيم البيانات إلى مجموعة تدريب 75% ومجموعة اختبار 25%

المنهج الثاني: تم تقسيم البيانات إلى مجموعة تدريب 60% ومجموعة اختبار 40% (نظراً للدقة العالية التي تم الحصول عليها في المنهج الأول وذلك للتأكد من عدم وجود overfitting)

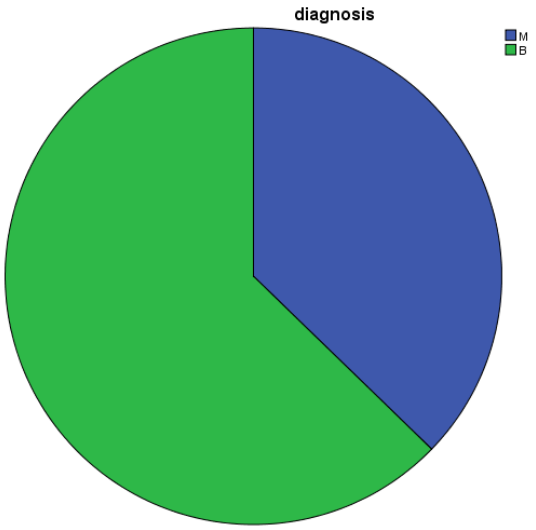
الفصل الرابع: النتائج والمناقشة
Results and discussion

1.4. نتائج المرحلة الأولى:

كانت نسبة المصابين بورم خبيث في دراستنا أقل من المصابين بورم حميد بنسبة بلغت 37%، بينما بلغت نسبة المصابين بورم حميد 63% (جدول 2).

جدول 2: نسبة المصابين بورم حميد وخبيث في الدراسة.

		Frequency	Percent
Valid	M	212	37.3
	B	357	62.7
	Total	569	100.0



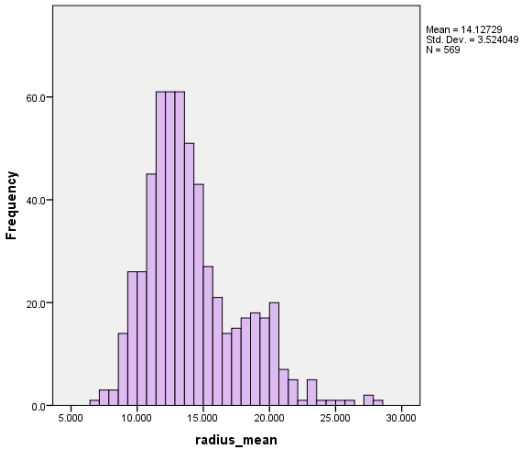
الشكل 11: نسبة المصابين بورم حميد وخبيث في الدراسة.

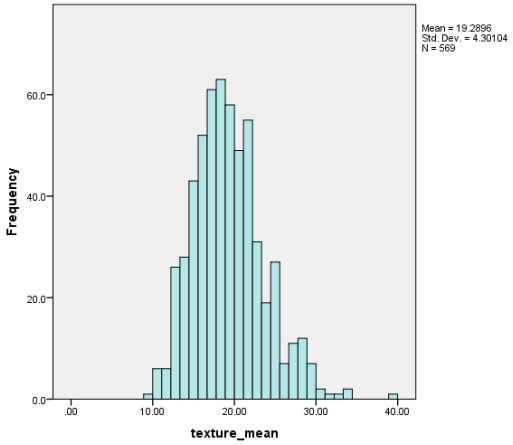
بلغ متوسط نصف قطر الكتل في العينة 14.13 بانحراف معياري قدره 3.52، بينما بلغ متوسط الخطأ المعياري لأنصاف أقطار الكتل 0.4. بلغ متوسط محيط الكتل 91.97 بانحراف قدره 351.9، بينما بلغ متوسط المساحة 654.89، ويوضح الجدول التالي توصيف إحصائي لباقي المتغيرات الكمية المدروسة في العينة.

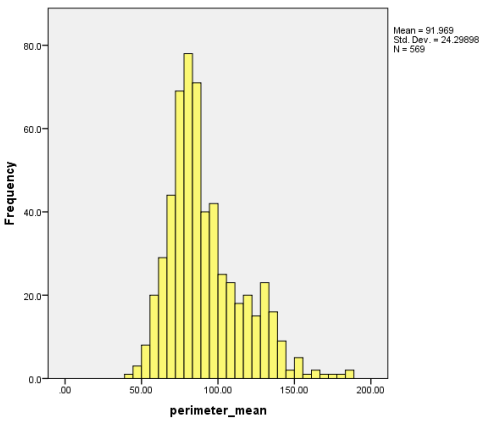
جدول 3: توصيف احصائي للمتغيرات الكمية في الدراسة.

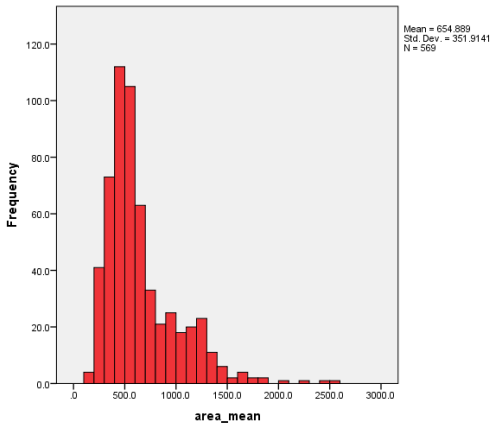
	Minimum	Maximum	Mean	Std. Deviation
radius_mean	6.981	28.110	14.12729	3.524049
texture_mean	9.71	39.28	19.2896	4.30104
perimeter_mean	43.79	188.50	91.9690	24.29898
area_mean	143.5	2501.0	654.889	351.9141
smoothness_mean	.05263	.16340	.0963603	.01406413
compactness_mean	.01938	.34540	.1043410	.05281276
concavity_mean	.000000	.426800	.08879932	.079719809
concave points_mean	.000000	.201200	.04891915	.038802845
symmetry_mean	.1060	.3040	.181162	.0274143
fractal_dimension_mean	.04996	.09744	.0627976	.00706036
radius_se	.1115	2.8730	.405172	.2773127
texture_se	.3602	4.8850	1.216853	.5516484
perimeter_se	.7570	21.9800	2.866059	2.0218546
area_se	6.802	542.200	40.33708	45.491006
smoothness_se	.001713	.031130	.00704098	.003002518
compactness_se	.002252	.135400	.02547814	.017908179
concavity_se	.000000	.396000	.03189372	.030186060
concave points_se	.000000	.052790	.01179614	.006170285
symmetry_se	.007882	.078950	.02054230	.008266372
fractal_dimension_se	.000895	.029840	.00379490	.002646071
radius_worst	7.930	36.040	16.26919	4.833242
texture_worst	12.02	49.54	25.6772	6.14626
perimeter_worst	50.41	251.20	107.2612	33.60254
area_worst	185.2	4254.0	880.583	569.3570
smoothness_worst	.07117	.22260	.1323686	.02283243
compactness_worst	.02729	1.05800	.2542650	.15733649

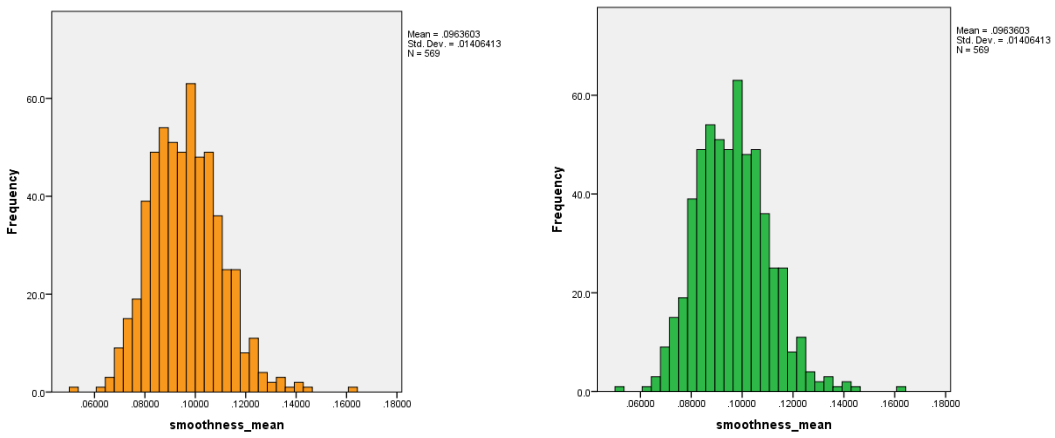
concavity_worst	.000000	1.252000	.27218848	.208624281
concave points_worst	.000000	.291000	.11460622	.065732341
symmetry_worst	.1565	.6638	.290076	.0618675
fractal_dimension_worst	.05504	.20750	.0839458	.01806127











الشكل 12: رسم بياني لتوزع أهم المتغيرات الكمية في الدراسة.

بعد دراسة الارتباط بين متوسطات المتغيرات لوحظ وجود ارتباط قوي جدا (pearson correlation>0.9) (جدول 4) بين كل من:

نصف قطر الكتلة مع المحيط (0.99**) ومع المساحة (0.98**).

محيط الكتلة والمساحة (0.98**).

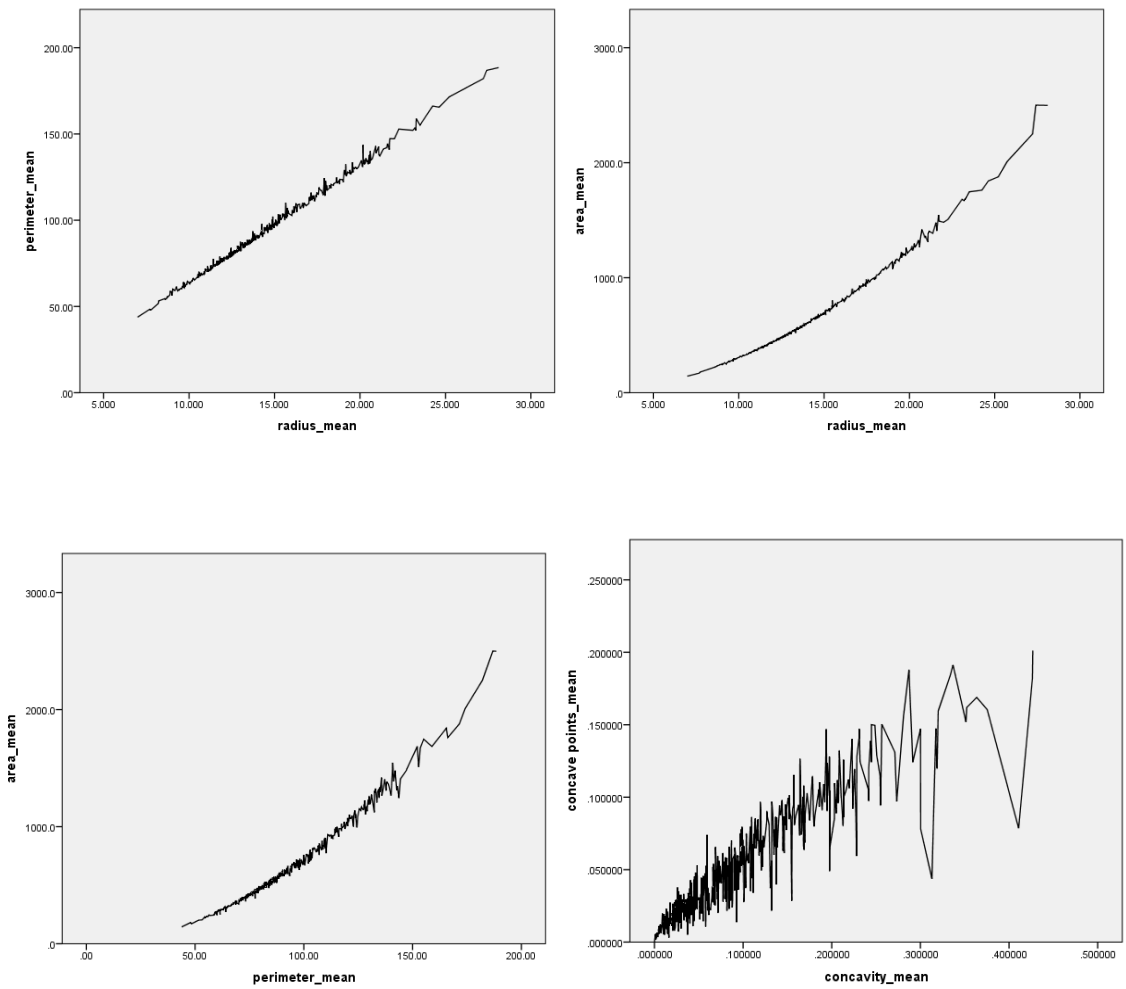
التقعر مع عدد الأجزاء المقعرة (0.92**).

لذلك تم حذف المتغيرات التالية (المحيط، المساحة، والتقعر)

جدول 4: دراسة الارتباطات بين المتغيرات الكمية.

Correlations											
		radius	texture	perimeter	area	smoothness	compactness	concavity	concave points	symmetry	Fractal dimension
radius	Corr	1	.324**	.998**	.987**	.171**	.506**	.677**	.823**	.148**	−.312**
	Sig		.000	.000	.000	.000	.000	.000	.000	.000	.000
texture	Corr	.324**	1	.330**	.321**	−.023−	.237**	.302**	.293**	.071	−.076−

	Sig.	.000		.000	.000	.578	.000	.000	.000	.089	.068
pe-rimeter	Corr	.998**	.330**	1	.987**	.207**	.557**	.716**	.851**	.183**	-.261-**
	Sig.	.000	.000		.000	.000	.000	.000	.000	.000	.000
area	Corr	.987**	.321**	.987**	1	.177**	.499**	.686**	.823**	.151**	-.283-**
	Sig.	.000	.000	.000		.000	.000	.000	.000	.000	.000
smoothness	Corr	.171**	-.023-	.207**	.177**	1	.659**	.522**	.554**	.558**	.585**
	Sig	.000	.578	.000	.000		.000	.000	.000	.000	.000
com-pact-ness	Corr	.506**	.237**	.557**	.499**	.659**	1	.883**	.831**	.603**	.565**
	sig	.000	.000	.000	.000	.000		.000	.000	.000	.000
con-cavity	Corr	.677**	.302**	.716**	.686**	.522**	.883**	1	.921**	.501**	.337**
	Sig.	.000	.000	.000	.000	.000	.000		.000	.000	.000
con-cave points	Corr	.823**	.293**	.851**	.823**	.554**	.831**	.921**	1	.462**	.167**
	Sig.	.000	.000	.000	.000	.000	.000	.000		.000	.000
sym-metry	Corr	.148**	.071	.183**	.151**	.558**	.603**	.501**	.462**	1	.480**
	Sig.	.000	.089	.000	.000	.000	.000	.000	.000		.000
Fractal dimen-sion	Corr	.312-**	.076-	.261-**	.283-**	.585**	.565**	.337**	.167**	.480**	1
	Sig.	.000	.068	.000	.000	.000	.000	.000	.000	.000	



شكل 13: رسم بياني يمثل الارتباط القوي بين (المحيط، المساحة، ونصف القطر)، وبين شدة التقعر وعدد نقاط التقعر.

تم بعد ذلك دراسة العلاقة بين متغيرات الدراسة ونوع الورم (خبيث أو حميد)، وذلك لمعرفة المتغيرات ذات العلاقة الأقوى بتحديد نوع الورم.

كما نلاحظ من الجدول التالي كان نصف قطر، مساحة، ومحيط الكتل الخبيثة أكبر مقارنة بالكتل الحميدة، وكان لجميع المتغيرات المدروسة علاقة دالة إحصائياً بنوع الورم ($P<0.05$) باستثناء بعض المتغيرات مثل متوسط البعد الكسري، الخطأ المعياري للملمس، النعومة، التجانس، البعد الكسري ($P>0.05$).

تشير العلاقات الهامة إحصائياً التي لوحظت في غالبية متغيرات الدراسة إلى أهمية هذه المتغيرات في تصنيف الورم كحميد أو خبيث.

جدول 5: جدول يوضح العلاقة بين وجود ورم حميد أو خبيث مع متغيرات الدراسة.

	diagnosis	N	Mean	Std. Deviation	sig
radius_mean	B	357	12.14652	1.780512	.000
	M	212	17.46283	3.203971	
texture_mean	B	357	17.9148	3.99512	.000
	M	212	21.6049	3.77947	
perimeter_mean	B	357	78.0754	11.80744	.000
	M	212	115.3654	21.85465	
area_mean	B	357	462.790	134.2871	.000
	M	212	978.376	367.9380	
smoothness_mean	B	357	.0924776	.01344608	.000
	M	212	.1028985	.01260824	
compactness_mean	B	357	.0800846	.03374995	.000
	M	212	.1451878	.05398750	
concavity_mean	B	357	.04605762	.043442151	.000
	M	212	.16077472	.075019328	
concave points_mean	B	357	.02571741	.015908778	.000
	M	212	.08799000	.034373909	
symmetry_mean	B	357	.174186	.0248068	.000
	M	212	.192909	.0276381	
fractal_dimension_mean	B	357	.0628674	.00674734	.760
	M	212	.0626801	.00757332	
radius_se	B	357	.284082	.1125696	.000
	M	212	.609083	.3450386	

texture_se	B	357	1.220380	.5891797	.843
	M	212	1.210915	.4831781	
perimeter_se	B	357	2.000321	.7711692	.000
	M	212	4.323929	2.5685457	
area_se	B	357	21.13515	8.843472	.000
	M	212	72.67241	61.355268	
smoothness_se	B	357	.00719590	.003060610	.110
	M	212	.00678009	.002890430	
compactness_se	B	357	.02143825	.016351511	.000
	M	212	.03228117	.018387190	
concavity_se	B	357	.02599674	.032918236	.000
	M	212	.04182401	.021603428	
concave points_se	B	357	.00985765	.005708625	.000
	M	212	.01506047	.005517362	
symmetry_se	B	357	.02058381	.006998539	.877
	M	212	.02047240	.010064888	
fractal_dimension_se	B	357	.00363605	.002938219	.063
	M	212	.00406241	.002041498	
radius_worst	B	357	13.37980	1.981368	.000
	M	212	21.13481	4.283569	
texture_worst	B	357	23.5151	5.49395	.000
	M	212	29.3182	5.43480	
perimeter_worst	B	357	87.0059	13.52709	.000
	M	212	141.3703	29.45706	

area_worst	B	357	558.899	163.6014	.000
	M	212	1422.286	597.9677	
smoothness_worst	B	357	.1249595	.02001347	.000
	M	212	.1448452	.02186983	
compactness_worst	B	357	.1826725	.09217998	.000
	M	212	.3748241	.17037198	
concavity_worst	B	357	.16623772	.140367741	.000
	M	212	.45060557	.181506723	
concave points_worst	B	357	.07444434	.035797374	.000
	M	212	.18223731	.046307790	
symmetry_worst	B	357	.270246	.0417448	.000
	M	212	.323468	.0746850	
fractal_dimension_worst	B	357	.0794421	.01380405	.000
	M	212	.0915300	.02155289	

2.4. نتائج المرحلة الثانية:

بعد الانتهاء من التوصيف الاحصائي تمّ العمل على Google Colab لبناء النماذج المختلفة.

تتألف مجموعة البيانات النهائيّة من 569 عيّنة، 21 دخل، وخرج وحيد ثنائي التصنيف (ورم خبيث وورم حميد).

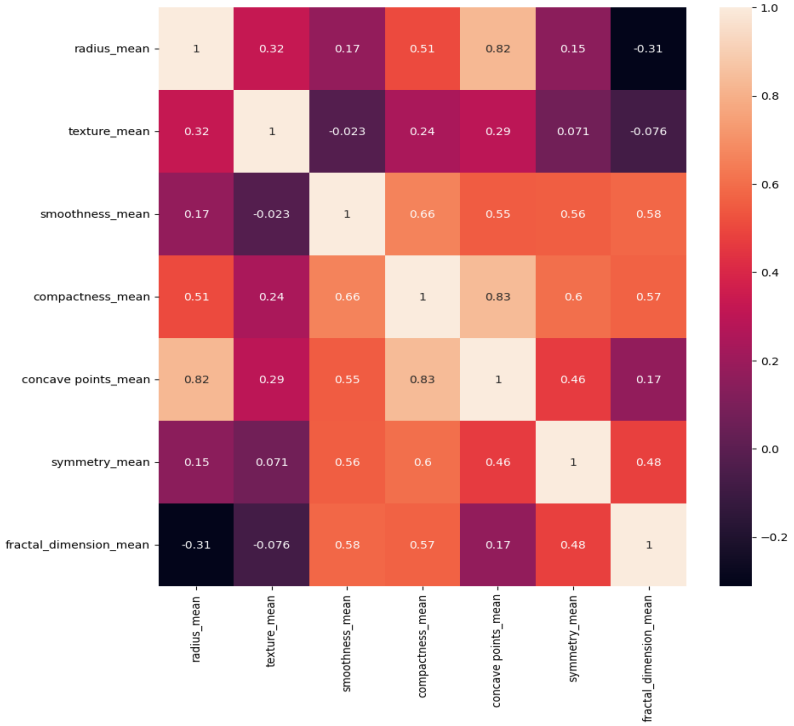
كانت جميع المتغيرات ذات طبيعة رقمية حيث تم معالجتها خلال المرحلة الأولى وترميز المتغيرات الاسمية، ولم يكن في العينة أي قيم مفقودة.

لمعرفة القيم المفقودة في عينة الدراسة تم استخدام التعليمة التالية:

```
df.iloc[:,:].isna().sum()
```

تمّ بعد ذلك إعادة رسم العلاقة بين متوسطات المتغيّرات بعد حذف المتغيرات المرتبطة، وتم ذلك باستخدام خريطة الحرارة heatmap كما هو موضح في الشكل 14، ومن أجل ذلك تم استخدام التعليمة التالية:

```
sns.heatmap(df.iloc[:,1:8].corr(),annot=True)
```



شكل 14: Heatmap للمتغيّرات المدروسة.

من أجل تقسيم البيانات إلى مجموعة تدريب 75% ومجموعة اختبار 25%، تم استخدام التعليمة التالية:

```
from sklearn.model_selection import train_test_split
X_train,X_test, y_train,y_test=
train_test_split(X,y,test_size=0.25, random_state= 0)
```

وتم استخدام خوارزميات التصنيف المختلفة بعد استيرادها باستخدام التعليمات التالية:

```
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestClassifier
```

```
from sklearn.neighbors import KNeighborsClassifier
```

أخيراً، تم تقييم أداء الخوارزميات باستخدام مصفوفة الشك باستخدام التعليمة التالية:

```
cm = confusion_matrix(y_test, y_pred2)
print(cm)
accuracy_score(y_test,y_pred)
```

1.2.4. نتائج المنهج الأول (Testing set= 0. 25):

تمكّنت جميع خوارزميات تعلّم الآلة من التنبّؤ بشكل ممتاز فيما كان الورم حميد أو خبيث، حيث كانت دقة الخوارزميّات متقاربة أكبر من 99% كما هو موضّح في الجدول 6، ويمكن تفسير ذلك بأنّ جميع المتغيّرات التي تمّ استخدامها للتنبّؤ بسرطان الثدي هي متغيرات ذات درجة عالية من الأهمية وهذا ما تم اثباته من خلال المنهج الأول لدى دراسة العلاقة بين المتغيرات ونوع سرطان الثدي.

احتلّت شجرة القرار، الانحدار اللوجستي، والغابة العشوائية المرتبة الأولى بدقة وصلت لـ 100% يليها باقي الخوارزميات 99.3%.

توافقت هذه النتيجة مع نتائج العديد من الدراسات، والتي أثبتت جدارة الخوارزميات المستخدمة في التنبؤ بسرطان الثدي [34–32].

عادة ما يحدث Overfitting عندما يكون لدينا نموذج مرن للغاية ممّا يفسر الدقّة العالية للخوارزميات بشكل عام لذلك تم اللجوء إلى تغيير بارامترات الخوارزميات لمعرفة مدى تأثرها على الدقة.

تم تغيير عمق الشجرة (max_depth) في خوارزميّة الغابة العشوائيّة (7، 10، و20)، إلا أن الخوارزمية حققت الدقة ذاتها، وهي من النتائج المتوقّعة حيث عادة ما تزداد دقة الخوارزميّة بزيادة عمق الأشجار، وفي دراستنا هذه بلغت الدقة 100% منذ البداية.

عند النظر إلى خوارزمية الجار الأقرب، وجدنا أن دقة التصنيف بلغت 99.3% في حال كان عدد الجيران (n_neighbors= 7,15)، تم تغيير هذه المعلّمة عدة مرات حتى الوصول إلى الدقة الأمثل 100% في حالة عدد الجيران 5.

جدول 6: دقة خوارزميات التصنيف حسب المنهج الأول.

		Accuracy Test=0.25
Logistic regression		%100
Support vector machine		%99.3
Decision tree		%100
Random forest	max_depth= 7	%100
	max_depth= 10	%100
	max_depth= 20	%100
KNN	n_neighbors= 5	%100
	n_neighbors= 7	%99.3
	n_neighbors= 15	%99.3

للتعمّق أكثر في أداء خوارزميّات التصنيف المستخدمة يجب العودة إلى مصفوفات الشك الخاصّة بكلّ واحدة. لوحظ أنّ مصفوفة الشك الخاصّة بخوارزمية DT، LOG، وRF قامت بالتصنيف بالشكل الأفضل بدون وجود أيّ أخطاء جدول 7، 8، و9، حيث تنبأت بـ 49 قيمة صحيحة على أنها ورم خبيث، و94 قيمة صحيحة على أنها ورم حميد.

جدول 7: مصفوفة شك خوارزميّة LOG الناتجة عن تطبيق المنهج الأوّل.

LOG	**التصنيف المتوقّع**	
	خبيث	**حميد**

0	49	**خبيث**	**التصنيف الحقيقي**
94	0	**حميد**	

جدول 8: مصفوفة شك خوارزميّة DT الناتجة عن تطبيق المنهج الأوّل.

التصنيف المتوقّع		**DT**	
حميد	**خبيث**		
0	49	**خبيث**	**التصنيف الحقيقي**
94	0	**حميد**	

جدول 9: مصفوفة شك خوارزميّة RF الناتجة عن تطبيق المنهج الأوّل.

التصنيف المتوقّع		**RF**	
حميد	**خبيث**	(7,10,20)	
0	49	**خبيث**	**التصنيف الحقيقي**
94	0	**حميد**	

أمّا بالنسبة لخوارزمية SVM ذات الدقة الأدنى فنلاحظ من الجدول 10 أنّ مقدار الخطأ في التوقّع كان قليل، حيث تم التوقع بقيمة واحدة خاطئة على أنها ورم خبيث في حين أنها ورم حميد، بينما توقعت بـ 49 قيمة صحيحة على أنها ورم خبيث، و93 قيمة صحيحة على أنها ورم حميد.

جدول 10: مصفوفة شك خوارزميّة SVM الناتجة عن تطبيق المنهج الأوّل.

SVM		التصنيف المتوقّع	
		خبيث	حميد
التصنيف الحقيقي	خبيث	49	0
	حميد	1	93

وأخيراً، بالنسبة لمصفوفة شك خوارزمية الجار الأقرب، بالرغم من أن دقتها هي مماثلة لدقة SVM، إلا أنها أخطأت بالتوقع بشكل مغاير لها، حيث توقعت بقمة خاطئة واحدة على أنها ورم حميد في حين أنها في الحقيقة ورم خبيث، وتنبأت بـ48 قيمة بشكل صحيح على أنها ورم خبيث، و94 قيمة صحيحة على أنها ورم حميد.

جدول 11: مصفوفة شك خوارزميّة KNN الناتجة عن تطبيق المنهج الأوّل.

KNN		التصنيف المتوقّع	
(7,15)		خبيث	حميد
التصنيف	خبيث	48	1

94	0	**حميد**	

2.2.4. نتائج المنهج الثاني (Testing set = 0.40):

انخفضت الدقّة بشكل خفيف في العديد من الخوارزميّات لتصل إلى 99.56%، بينما بقيت 100% لدى بقية الخوارزميّات كما هو موضّح في الجدول 12.

جدول 12: المقارنة بين المنهج الأوّل والمنهج الثاني من حيث دقّة الخوارزميّات.

		Accuracy Test=0.25	Accuracy Test=0.40
Logistic regression		%100	%100
Support vector machine		%99.3	99.56%
Decision tree		%100	%100
Random forest	max_depth= 7	%100	%100
	max_depth= 10	%100	%100
	max_depth= 20	%100	%100
KNN	n_neighbors= 5	%100	99.56%
	n_neighbors= 7	%99.3	99.56%
	n_neighbors= 15	%99.3	99.56%

وبالعودة إلى مصفوفات الشك، لوحظ أن عدد القيم الخاطئة التي قامت خوارزميّة KNN بالتنبّؤ بها هي نفسها في الحالات الثلاث جدول 13، ويمكن تفسير انخفاض الدقة نتيجة ازدياد العدد في مجموعة التدريب، حيث تم التنبؤ بقيمة خاطئة واحدة على أنها ورم حميد وهي في الأساس ورم خبيث.

جدول 13: مصفوفة شك خوارزميّة KNN الناتجة عن تطبيق المنهج الثاني.

KNN		التصنيف المتوقّع	
		خبيث	حميد
التصنيف الحقيقي	خبيث	79	**1**
	حميد	0	**148**

أما بالنسبة لمصفوفة شك SVM، فكانت القيمة الخاطئة التي تنبأت بها معاكسة لمصفوفة KNN، حيث تنبأت بقيمة خاطئة وحيدة على أنها ورم خبيث وهي في الأساس ورم حميد.

جدول 14: مصفوفة شك خوارزميّة SVM الناتجة عن تطبيق المنهج الثاني.

SVM		التصنيف المتوقّع	
		خبيث	حميد
التصنيف الحقيقي	خبيث	79	**0**
	حميد	1	147

نظراً لتفاوت عدد البيانات في كل صفّ من صفوف الخرج ولإعادة تأكيد عدم وجود overfitting للبيانات، تمّ اللجوء إلى التحقق المتقاطع k-fold cross-validation وgrid search لكل من خوارزميتي LOG وSVM، حيث تعتبر هذه من الاستراتيجيات الهامة المتبعة التي تقوم بإيجاد أفضل البارامترات التي تحقّق أفضل أداء.

وجد أنّ البارامترات التالية {'C': 10, 'gamma': 0.01, 'kernel': 'rbf'} تحقّق أفضل دقّة تدريب 100%.

(تقوم kernal function بتحويل مجموعة بيانات التدريب إلى بيانات بأبعاد أكبر لجعلها قابلة للفصل خطيّاً. الوظيفة الأساسية لمصنف SV هي وظائف الأساس الشعاعي والتي يرمز لها عادة rbf (radial basis functions)، بينما ترمز c لقوة التنظيم أمّا gamma هو معامل النواة لـ rbf).

بقيت دقة LOG مساوية لـ 100%، بينما ارتفعت دقة SVM من 99.56% إلى 100% ويمكن وصف أداء هذا المصنّف بالمتوازن، وهذا متوقّع نظراً لقوّة أداء هذا المصنّف. يمثّل الجدول التالي تقرير عن معايير تقييم دقّة خوارزميّة SVM.

جدول 15: تقييم دقة خوارزميّة SVM.

classification report				
	precision	recall	f1-score	support
Malignant	1.00	1.00	1.00	80
benign	1.00	1.00	1.00	148
accuracy			1.00	228
macro avg	1.00	1.00	1.00	228
weighted avg	1.00	1.00	1.00	228

3.4. خاتمة عامة:

في الختام، تسلط نتائج دراستنا الضوء على الأداء المتقارب لخوارزميات التصنيف في التنبؤ بنتائج مرضى سرطان الثدي. أظهرت غالبية الخوارزميات دقة ممتازة دون أي تنبؤات خاطئة، حيث يمكن اعتبار هذه الخوارزميات موثوقة للتنبؤ بنتائج سرطان الثدي.

4.4. التوصيات:

1. الاهتمام بسرطان الثدي على مستوى الجهات الحكومية والخاصة.

2. متابعة استخدام أدوات المعلوماتية الحيوية بشكل أدق للكشف عن أورام الثدي.
3. تعزيز دور المعلوماتية الحيوية خصوصاً على المستوى الطبي والمشافي لمساعدة الأطباء خلال مرحلة العلاج.
4. اقتراح برامج أدق وأعمق مع إمكانية جعلها متاحة لجميع العامة على شكل تطبيقات على الهاتف المحمول.
5. اجراء دراسات أكبر وإدخال بيانات تشمل المجتمع السوري.
6. تقديم خلاصة البحث للمجتمع سواء بتشخيص سرطان الثدي أو استخدام أدوات المعلوماتية الحيوية.

الفصل الخامس: المراجع
References

المراجع:

1. Coleman MP, Quaresma M, Berrino F, et al (2008) Cancer survival in five continents: a worldwide population-based study (CONCORD). Lancet Oncol 9:730–756
2. Reeder JG, Vogel VG (2008) Breast cancer prevention. Cancer Treat Res 141:149–164
3. Breast Cancer Statistics | How Common Is Breast Cancer? | American Cancer Society. https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html. Accessed 25 Oct 2023
4. Akram M, Iqbal M, Daniyal M, Khan AU (2017) Awareness and current knowledge of breast cancer. Biol Res. https://doi.org/10.1186/S40659-017-0140-9
5. Breast, Female, Anatomy: Image Details - NCI Visuals Online. https://visualsonline.cancer.gov/details.cfm?imageid=7127. Accessed 12 Nov 2023
6. What Is Breast Cancer? An Overview. https://www.breastcancer.org/about-breast-cancer. Accessed 12 Nov 2023
7. Reis-Filho JS, Lakhani SR (2008) Breast cancer special types: why bother? J Pathol 216:394–398
8. Breast Cancer, Inflammatory: Image Details - NCI Visuals Online. https://visualsonline.cancer.gov/details.cfm?imageid=7199. Accessed 12 Nov 2023
9. Invasive Ductal Carcinoma Of The Breast: Image Details - NCI Visuals Online. https://visualsonline.cancer.gov/details.cfm?imageid=12799. Accessed 12 Nov 2023
10. Lobular Carcinoma In Situ: Image Details - NCI Visuals Online. https://visualsonline.cancer.gov/details.cfm?imageid=9311. Accessed 12 Nov 2023
11. Teichgraeber DC, Guirguis MS, Whitman GJ (2021) Breast Cancer Staging: Updates in the AJCC Cancer Staging Manual, 8th Edition, and Current Challenges for Radiologists, From the AJR Special Series on Cancer Staging. AJR Am J Roentgenol 217:278–290

12. Signs and Symptoms of Breast Cancer. https://www.breastcancer.org/signs-symptoms. Accessed 12 Nov 2023
13. Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP (2017) Risk Factors and Preventions of Breast Cancer. Int J Biol Sci 13:1387–1397
14. Shenkier T, Weir L, Levine M, Olivotto I, Whelan T, Reyno L (2004) Clinical practice guidelines for the care and treatment of breast cancer: 15. Treatment for women with stage III or locally advanced breast cancer. C Can Med Assoc J 170:983–994
15. Mohiuddin JJ, Deal AM, Carey LA, Lund JL, Baker BR, Zagar TM, Jones EL, Marks LB, Chen RC (2016) Neoadjuvant Systemic Therapy Use for Younger Patients with Breast Cancer Treated in Different Types of Cancer Centers Across the United States. J Am Coll Surg 223:717-728.e4
16. Kim KJ, Huh SJ, Yang JH, et al (2005) Treatment results and prognostic factors of early breast cancer treated with a breast conserving operation and radiotherapy. Jpn J Clin Oncol 35:126–133
17. Lu B, Natarajan E, Balaji Raghavendran HR, Markandan UD (2023) Molecular Classification, Treatment, and Genetic Biomarkers in Triple-Negative Breast Cancer: A Review. Technol Cancer Res Treat. https://doi.org/10.1177/15330338221145246
18. Emens LA, Loi S (2023) Immunotherapy Approaches for Breast Cancer Patients in 2023. Cold Spring Harb Perspect Med. https://doi.org/10.1101/CSHPERSPECT.A041332
19. Moor J (2006) The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. AI Mag 27:87–87
20. Zhang L, Geisler T, Ray H, Xie Y (2022) Improving logistic regression on the imbalanced data by a novel penalized log-likelihood function. J Appl Stat 49:3257
21. (PDF) استخدام الانحدار اللوجستي متعدد الاستجابة لتحديد العوامل المؤثرة على مرض العيون. https://www.researchgate.net/publication/331062654_astkhdam_alanhdar_allwjsty_mtdd_alastjabt_lthdyd_alwaml_almwthrt_ly_mrd_alywn. Accessed 22 Nov 2023
22. Abu Alfeilat HA, Hassanat ABA, Lasassmeh O, Tarawneh AS, Alhasanat MB, Eyal Salman HS, Prasath VBS (2019) Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. Big data 7:221–248
23. Chrimes D (2023) Using Decision Trees as an Expert System for

Clinical Decision Support for COVID-19. Interact J Med Res 12:e42540

24. Podgorelec V, Kokol P, Stiglic B, Rozman I (2002) Decision trees: an overview and their use in medicine. J Med Syst 26:445–463
25. Random forest Algorithm in Machine learning | Great Learning. https://www.mygreatlearning.com/blog/random-forest-algorithm/. Accessed 22 Nov 2023
26. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Peter Campbell J (2020) Introduction to Machine Learning, Neural Networks, and Deep Learning. Transl Vis Sci Technol. https://doi.org/10.1167/TVST.9.2.14
27. Nemade V, Fegade V (2023) Machine Learning Techniques for Breast Cancer Prediction. Procedia Comput Sci 218:1314–1320
28. Ara S, Das A, Dey A (2021) Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms. 2021 Int Conf Artif Intell ICAI 2021 97–101
29. Khater T, Hussain A, Bendardaf R, Talaat IM, 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
30. Islam MM, Haque MR, Iqbal H, Hasan MM, Hasan M, Kabir MN (2020) Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN Comput Sci 1:1–14
31. Mahesh TR, Vinoth Kumar V, Vivek V, Karthick Raghunath KM, Sindhu Madhuri G (2022) Early predictive model for breast cancer classification using blended ensemble learning. Int J Syst Assur Eng Manag. https://doi.org/10.1007/S13198-022-01696-0
32. Ganggayah MD, Taib NA, Har YC, Lio P, Dhillon SK (2019) Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Med Informatics Decis Mak 2019 191 19:1–17
33. Montazeri M, Montazeri M, Montazeri M, Beigzadeh A (2016) Machine learning models in breast cancer survival prediction. Technol Health Care 24:31–42
34. Ganggayah MD, Taib NA, Har YC, Lio P, Dhillon SK (2019) Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Med. Inform. Decis. Mak. 19: