

تقييم أداء خوارزمية الغابات العشوائية في تصنيف أورام الثدي اعتماداً على الخصائص النسيجية للصورة

هيفاء طه عبد

رسل حسن كطيف

قسم الاحصاء/كلية الادارة والاقتصاد/الجامعة المستنصرية

haefaa_adm@uomustansiriyah.edu.iq

rsisii34344@gmail.com

مستخلص البحث:

يهدف هذا البحث إلى تقويم كفاءة منهج تصنيفي يعتمد على تحليل الصور الطبية في تشخيص أورام الثدي والتمييز بين الأنماط الحميدة والخبيثة. اعتمدت الدراسة على مجموعتين من الصور التشريحية الرقمية، بلغ حجم الأولى (10,000) صورة والثانية (2,400) صورة، بعد إخضاعهما لسلسلة من خطوات المعالجة المسبقة التي شملت رفع جودة الصور، وتوحيد أبعادها، واستخدام تقنيات تحليل النسيج لاستخراج خصائص تصف البنية الدقيقة للورم. جرى بعد ذلك بناء نموذج للتصنيف باستخدام منهج الغابات العشوائية، وتقدير أدائه بالاستعانة بمجموعة شاملة من المؤشرات الإحصائية. أظهرت النتائج قدرة مرتفعة للنموذج على الفصل بين الفئتين، مع تحقيق مستويات عالية من الدقة واستقرار واضح في الأداء رغم اختلاف حجم العينات. كما بينت النتائج أن خوارزمية الغابات العشوائية تتمتع بمتانة إحصائية وأداء ثابت، في حين تسهم العينات الأكبر في تعزيز قوة التمييز وتحسين استقرار المؤشرات. وتبرز هذه النتائج الدور المهم للمعالجة المسبقة واستخراج السمات النسيجية في رفع جودة التشخيص، بما يدعم إمكانية توظيف هذا الأسلوب في التطبيقات الطبية المعنية بتحديد طبيعة الأورام.

الكلمات المفتاحية: التصنيف، الغابات العشوائية، صور نسيجية، أورام الثدي.

1- المقدمة

تعدّ أورام الثدي من أكثر الحالات المرضية انتشاراً، وتكتسب أهمية خاصة نظراً لأثرها الصحي والاجتماعي الواسع. ويُعد التشخيص المبكر العامل الأكثر تأثيراً في تحسين فرص العلاج وتقليل نسب الوفيات، مما يستدعي تطوير طرائق تحليل دقيقة وموثوقة تسهم في دعم القرار الطبي. ومع التزايد الكبير في استخدام الصور الطبية في تشخيص الأورام، برزت الحاجة إلى اعتماد أساليب تحليل كفوءة قادرة على استخلاص معلومات دقيقة من الصور واستثمارها في بناء نماذج تصنيف تعتمد على الدقة والموضوعية. وفي هذا السياق، تبرز أهمية تحليل النسيج داخل الصورة، إذ يوفر هذا التحليل مؤشرات كمية تعبر عن الاختلافات الدقيقة بين الأنسجة السليمة والمصابة. ومن بين الطرائق الشائعة في هذا المجال، تُعد الأنماط الثنائية المحلية إحدى الأدوات الفعالة لالتقاط ميزات نسيجية قادرة على وصف البنية الدقيقة للورم. ويتيح هذا النوع من الخصائص إمكانية بناء نماذج تصنيف تعتمد على البيانات المستخلصة من الصورة وليس على التقييم البصري. ارتكز هذا البحث على استخدام تحليل النسيج مع تطبيق منهج الغابات العشوائية بوصفه أحد الأساليب الإحصائية القادرة على التعامل مع البيانات ذات البنية المتنوعة. وتم استخدام مجموعتين من الصور التشريحية الرقمية، الأولى كبيرة الحجم والثانية أصغر، بهدف دراسة تأثير حجم البيانات على أداء النموذج وقدرته على التعميم. وتم تقييم النموذج باستخدام مجموعة من المؤشرات الإحصائية التي تعكس دقته وكفاءته واستقراره. وتسهم نتائج هذا البحث في توفير إطار عملي لتحليل صور أورام الثدي وتصنيفها بدقة عالية، بما يدعم التوجهات الحديثة نحو اعتماد الأساليب الكمية في التشخيص الطبي، ويعزز قاعدة العمل في الدراسات التطبيقية التي تستفيد من الصور الطبية ذات البنية النسيجية المعقدة.

2- هدف البحث

يهدف هذا البحث إلى تطوير وتقويم نموذج تصنيفي يعتمد على تحليل الصور الطبية من أجل التمييز بين أورام الثدي الحميدة والخبيثة، وذلك من خلال تطبيق منهج الغابات العشوائية على صور تمت معالجتها واستخراج خصائص نسيجية منها، مع دراسة تأثير حجم العينة في دقة النموذج واستقراره، ومقارنة نتائجه على عينتين مختلفتين من حيث العدد والتنوع.

3- خوارزمية الغابات العشوائية Random Forest Algorithm

تُعد الغابة العشوائية إحدى النماذج المتقدمة في التعلم الآلي (Machine Learning) التي تُستخدم على نطاق واسع في مهام التصنيف (Classification) والتنبؤ (Prediction). وتعتمد فكرتها الأساسية على مفهوم أشجار القرار (Decision Trees)، إذ يتم إنشاء مجموعة كبيرة من أشجار القرار بشكل عشوائي ثم دمج نتائجها لإنتاج القرار النهائي الأكثر دقة واستقراراً [Dietterich, (2000)] [Salman et al, (2024)]

تم تقديم خوارزمية الغابات العشوائية لأول مرة على يد Leo Breiman عام 2001، مستلهماً أفكاره من أعمال Geman & Amit، حيث اعتُبرت امتداداً لفكرة التكديس (Stacking) لتصبح منافساً قوياً لخوارزميات التعزيز. [Breiman, (2001)]

لتدريب خوارزميات التعلم الآلي ونماذج الذكاء الاصطناعي، لا بد من توفر كمية كافية من البيانات عالية الجودة، إذ تُعد البيانات حجر الأساس في بناء النماذج وتحسين أدائها. تساعد هذه البيانات في صقل الخوارزميات وتعزيز كفاءة البرامج والأجهزة، وتقييم سلوك المستخدمين، وتحديد الأنماط واتخاذ القرارات والتنبؤ المستقبلي، مما يؤدي في النهاية إلى رفع كفاءة النموذج ودقته.

يتم إنشاء كل شجرة قرار داخل الغابة باستخدام عينات عشوائية من البيانات عبر أسلوب يُعرف باسم الاختيار بالعينة Bootstrap Sampling، حيث تُستخدم مجموعات فرعية مختلفة من البيانات لبناء كل شجرة. بعد ذلك، تُدمج نتائج جميع الأشجار من خلال آلية التصويت (Voting) أو المتوسط (Averaging) للوصول إلى التنبؤ النهائي الأكثر موثوقية.

إحدى المزايا البارزة للغابات العشوائية هي قدرتها العالية على التعامل مع البيانات غير المتوازنة (Imbalanced Data) والبيانات التي تحتوي على قيم مفقودة (Missing Values). كما أنها تخفف من مشكلة التحيز الناتج عن اختيار المتغيرات (Variable Selection Bias) التي تواجه بعض النماذج الأخرى. إضافة إلى ذلك، تُعد الغابات العشوائية فعالة جداً في تقليل مشكلة الإفراط في الملاءمة (Overfitting)، إذ يتم تدريب العديد من أشجار القرار على عينات مختلفة من البيانات، مما يعزز من قدرة النموذج على التعميم والتنبؤ على بيانات جديدة لم تُشاهد سابقاً.

تُستخدم الغابات العشوائية على نطاق واسع في العديد من المجالات التطبيقية مثل تصنيف الصور، التنبؤ بالأسعار، التحليل المالي، النمذجة البيئية والطبية، والتعلم الإشرافي (Supervised Learning). يمكن تصنيف عملية استخراج البيانات (Data Mining) إلى نوعين رئيسيين

[Cutler et al, (2011)]:

1. الاستخراج الوصفي (Descriptive Data Mining): يهدف إلى وصف البيانات واكتشاف الأنماط والعلاقات بينها.

2. الاستخراج التنبؤي (Predictive Data Mining): يعتمد على دراسة البيانات التاريخية لتحديد الاتجاهات المستقبلية والتنبؤ بها.

يتم إنشاء النماذج التنبؤية من خلال تحليل الخصائص التنبؤية (Predictor Variables) بهدف توفير فرضيات تساعد في اتخاذ قرارات مستقبلية مبنية على أسس إحصائية دقيقة. وتعتمد دقة هذه

النماذج على طرائق تقدير الخطأ (Error Estimation Techniques) التي تُستخدم لتقييم كفاءة النموذج.

تُعد خوارزمية الغابات العشوائية إحدى أساليب التعلم الآلي الفعّالة، حيث تُستخدم النماذج غير الخاضعة للإشراف (Unsupervised Learning) بشكل رئيسي في تحليل البيانات الوصفية، بينما تُوظف النماذج الخاضعة للإشراف (Supervised Learning) في مهام التنبؤ والنمذجة التنبؤية [Duda et al, (2001)][James et al, (2013)].

تتألف الغابة العشوائية من مجموعة كبيرة من أشجار القرار (Decision Trees)، حيث يتم تكوين كل شجرة عبر أخذ عينات عشوائية من البيانات باستخدام أسلوب Bootstrap Sampling، مع اختيار مجموعة فرعية عشوائية من السمات (Features) عند كل عقدة في الشجرة. يضمن هذا النهج تقليل التحيز وتقليل التباين في النماذج. [Breiman, (2001)]. تتميز خوارزمية الغابات العشوائية بعدة مزايا مقارنة بالطرائق الأخرى مثل Bagging و Boosting، وتشمل:

1. تحقيق دقة تنبؤية عالية جداً على مختلف مجموعات البيانات.
 2. القدرة على التعامل بكفاءة مع قواعد البيانات الضخمة (Large Datasets).
 3. إمكانية معالجة عدد كبير من المتغيرات المستقلة دون الحاجة لحذف أي منها.
 4. تقديم تقديرات غير متحيزة لخطأ التعميم (Unbiased Estimation of Generalization Error).
 5. الأداء الفعّال في بيئات الحوسبة المتوازية (Parallel Computing).
- من الناحية الإحصائية، توفر الغابات العشوائية أدوات تحليلية قوية تشمل:
- a. قياس أهمية المتغيرات (Variable Importance Measures)، لتحديد أثر كل متغير على التنبؤ النهائي.
 - b. الترجيح التفاضلي الطبقي (Differential Class Weighting)، لمعالجة مشكلات عدم توازن البيانات.
 - c. استكمال القيم المفقودة (Missing Value Imputation)، للتعامل مع البيانات الناقصة.
 - d. سهولة التفسير والتصور البياني (Interpretability and Visualization)، ما يدعم فهم سلوك النموذج.

يمكن استخدام الغابات العشوائية في معالجة كل من المتغيرات الفئوية (Categorical Variables) والمستمرة (Continuous Variables)، سواء في مهام التصنيف (Classification) أو الانحدار (Regression). وتتميز هذه التقنية بالعديد من الخصائص الحسابية والإحصائية المهمة:

- a) التعامل المتزامن مع مهام الانحدار والتصنيف.
- b) سرعة نسبية في عمليات التدريب والتنبؤ.
- c) الاعتماد على عدد محدود من معلمات الضبط (Hyperparameters).
- d) احتواء النموذج على تقدير مدمج لخطأ التعميم.
- e) ملائمتها للمشكلات عالية الأبعاد (High-Dimensional Problems).
- f) سهولة التنفيذ على أنظمة متعددة النوى. [Louppe, (2014)] (Parallel Execution)

1. تصويت الغابة للتصنيف: (Majority Voting)

$$y^{\wedge} = mode\{h_1(x), h_1(x), \dots, h_B(x)\} \quad (1)$$

حيث $h_i(x)$ هو تنبؤ الشجرة i ، و B هو عدد الأشجار في الغابة.

2. تقدير متوسط خطأ الانحدار: (Regression)

$$\hat{y} = \frac{1}{B} \sum_{i=1}^B h_i(x) \quad (2)$$

3. أهمية المتغيرات: (Variable Importance via Gini Impurity)

$$VI_j = \sum_{t \in T_j} \Delta_i(t) \quad (3)$$

حيث $\Delta_i(t)$ هو الانخفاض في معيار Gini Impurity عند تقسيم العقدة t باستخدام المتغير j ، و T_j تمثل مجموعة العقد التي استخدم فيها المتغير j .

1-3 آلية عمل خوارزمية الغابات العشوائية

(Algorithmic Procedure of Random Forests)

تعمل خوارزمية الغابات العشوائية وفق سلسلة من الخطوات المنظمة لضمان تنبؤ دقيق وموثوق، وهي كما يلي:

1- اختيار العينات العشوائية (Bootstrap Sampling)

يتم إنشاء مجموعة تدريب عشوائية (B) من البيانات الأصلية باستخدام العينة مع الإرجاع (Sampling with Replacement). تحتوي كل مجموعة على عدد بيانات يقارب حجم البيانات الأصلية، لكن بعض الملاحظات قد تتكرر عدة مرات بينما يتم استبعاد أخرى. يتيح هذا الأسلوب تنوع البيانات المستخدمة في تدريب كل شجرة وتقليل التحيز. [Breiman, (2001)].

2- بناء اشجار القرار (Decision Tree Construction)

لكل عينة عشوائية يتم إنشاء شجرة قرار مستقلة، وفق الخطوات التالية:

a) عند كل عقدة، لا تُستخدم جميع المتغيرات، بل تُختار مجموعة فرعية عشوائية من المتغيرات ($mtry$ (mtry)) هو عدد المتغيرات (السمات) التي تختارها الخوارزمية بصورة عشوائية عند كل عقدة داخل كل شجرة أثناء عملية الانقسام (Node Split). عادةً، يتم تحديد أفضل تقسيم باستخدام مقاييس نقاء العقد، مثل:

a. مؤشر جيني (Gini Index) في خوارزمية CART. (عتمد خوارزمية CART على مبدأ إنشاء أشجار قرار ثنائية تعتمد على معيار نقاء يُعرف بمؤشر جيني، الذي يقيس درجة التجانس داخل العقد. في كل خطوة، تبحث الخوارزمية عن المتغير والقيمة الحدية اللذين يُنتجان أفضل فصل بين المجموعات، بحيث يصبح كل فرع أكثر تجانساً قدر الإمكان. وبخلاف ID3، تُنتج CART تقسيمات ثنائية فقط، كما يمكن استخدامها في كل من التصنيف والانحدار. وتتميز الخوارزمية بمرونتها وسرعتها وقدرتها على التعامل مع البيانات المستمرة والفئوية على حد سواء.)

b. مقياس الانتروبيا (Entropy) في خوارزمية ID3 أو C4.5. (تعتمد خوارزمية ID3 على بناء شجرة القرار من خلال اختيار المتغير الذي يحقق أعلى قيمة للكسب المعلوماتي، وهو المقياس الذي يحدد قدرة المتغير على تقليل الإنتروبيا داخل البيانات. تبدأ الخوارزمية بحساب الإنتروبيا العامة للمجموعة، ثم تُقيم كل متغير من حيث مدى مساهمته في تنظيم البيانات وتقليل درجة عدم التجانس. بعدها تختار المتغير الأكثر تمييزاً وتستخدمه لتقسيم البيانات إلى فروع، ثم تكرر العملية نفسها في كل فرع حتى الوصول إلى عقد نهائية يكون فيها القرار واضحاً. تمتاز ID3 ببساطتها ووضوح منهجها، وهي مناسبة لمهام التصنيف ذات البيانات الفئوية.)

c) لا تُجرى عملية التقليم (Pruning) داخل كل شجرة، إذ يتم التحكم في النموذج ضمناً عند دمج نتائج جميع الأشجار.

3-دمج نتائج الاشجار (Aggregation): بعد بناء كل الأشجار، يتم دمج النتائج لإنتاج التنبؤ النهائي:

في التصنيف (Classification): يُعتمد على تصويت الأغلبية (Majority Voting) بين جميع الأشجار إذا كان لدينا مجموعة من أشجار القرار T_1, T_2, \dots, T_B تم إنشاؤها من بيانات التدريب $D = (x_i, y_i)_{i=1}^n$ ، فإن التنبؤ النهائي في الغابة العشوائية يُحسب كما يلي [Dietterich, (2000)] [James et al, (2013)]:

$$\hat{y} = \{T_b(x)\}_{b=1}^B \quad (4)$$

في الانحدار (Regression): يُحسب المتوسط الحسابي لقيم جميع الأشجار

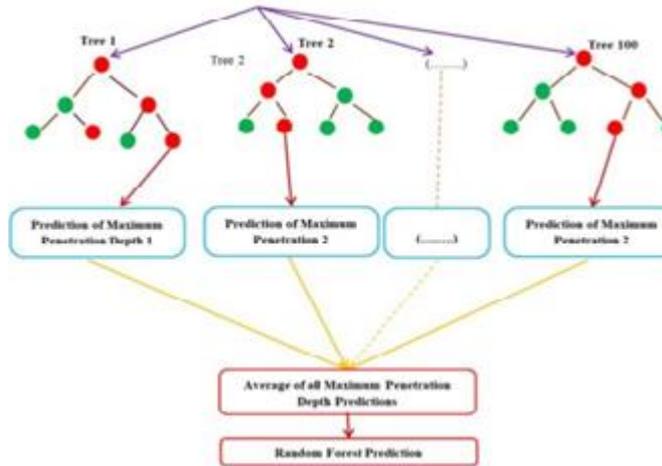
$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (5)$$

حيث $T_b(x)$ (هو التنبؤ الناتج عن الشجرة b ، و \hat{y} هو التنبؤ النهائي للغابة.

3-2 تقدير خطأ العينة خارج الحقيقية (Out of Bag Error Estimation)

نظرًا لأن كل شجرة تُبنى باستخدام عينة Bootstrap، فإن حوالي ثلث البيانات لا تُستخدم في تدريب الشجرة، وتُعرف باسم عينات خارج الحقيقية (Out-of-Bag Samples, OOB). يمكن استخدام هذه العينات لتقدير دقة النموذج بشكل غير متحيز، دون الحاجة إلى مجموعة اختبار منفصلة، عبر الخطوات التالية [Cutler et al, (2011)] [Louppe, (2014)].

1. لكل مشاهدة x_i ، يتم تحديد الأشجار التي لم تُستخدم في تدريبها.
 2. تمرير x_i عبر هذه الأشجار للحصول على مجموعة من التنبؤات.
 3. حساب متوسط أو نمط هذه التنبؤات لتقدير الفئة أو القيمة المتوقعة لـ y_i .
 4. مقارنة التنبؤات بالقيم الحقيقية لحساب متوسط الخطأ الإجمالي للنموذج.
- توفر هذه الطريقة تقديرًا دقيقًا وغير متحيز لدقة النموذج، وغالبًا ما تكون نتائجها مشابهة تقريبًا لطرائق التحقق المتقاطع (Cross-Validation) مع تكلفة حسابية أقل.



الشكل (1): البنية العامة لخوارزمية الغابة العشوائية

3-3 تقدير خطأ العينة (Out-of- Error)

يُعتبر تقدير الخطأ خارج العينة من أبرز خصائص الغابة العشوائية، ويُحسب كالتالي [Quinlan, (1987)]:

1. لكل مشاهدة x_i ، يتم تحديد الأشجار التي لم تُستخدم في تدريبها (أي لم تُشمل في عينة Bootstrap).

2. يتم تمرير x_i عبر تلك الأشجار فقط للحصول على مجموعة من التنبؤات.
3. يتم حساب متوسط (أو نمط) تلك التنبؤات لتقدير الفئة أو القيمة المتوقعة لـ x_i .
4. يُحسب خطأ التنبؤ الإجمالي من خلال مقارنة هذه القيم بالقيم الحقيقية y_i .
تُعطي هذه الطريقة تقديرًا غير متحيز لدقة النموذج، وغالبًا ما تكون مساوية تقريبًا لنتائج التحقق المتقاطع (Cross-Validation)، ولكن بكلفة حسابية أقل.

4- الجانب التطبيقي

1-4 الحصول على البيانات

يُظهر اختيار مجموعتين من الصور الطبية، إحداهما كبيرة الحجم نسبيًا (10,000 صورة) والأخرى متوسطة الحجم (2,400 صورة)، توجّهًا منهجيًا يهدف إلى اختبار النموذج في بيانات بيئات مختلفة من حيث الحجم وتنوع الأنماط البصرية. ويعكس التوزيع المتوازن بين فئتي الأورام الحميدة والخبيثة في كل مجموعة تحكمًا منهجيًا مهمًا يقلل من التحيز الاحتمالي للنموذج، إذ إن عدم التوازن يؤدي عادة إلى انحياز النموذج نحو الفئة الأكثر تمثيلًا ويضعف حساسيته تجاه الفئة الأقل ظهورًا.

كما أن التقسيم العشوائي للبيانات إلى 80% للتدريب و20% للاختبار يضمن وجود حجم كافٍ لبناء نموذج قوي من جهة، وقدرة على تقييم أدائه على بيانات غير مرئية من جهة أخرى، وهو ما يحقق شرط التعميم (Generalization) الضروري للتطبيقات الطبية الحساسة. إن الحفاظ على توازن الفئات داخل مجموعتي التدريب والاختبار يُعد مؤشرًا إيجابيًا يُنتظر معه تحقيق معدلات دقة واسترجاع مرتفعة في مراحل التصنيف اللاحقة، خصوصًا عند التعامل مع نماذج تعتمد على الأنماط البصرية الدقيقة.

2-4 المعالجة المسبقة للبيانات

تُعد المعالجة المسبقة من أكثر المراحل تأثيرًا في جودة التنبؤ، ويظهر من الإجراءات المتبعة أن عملية التحضير تمت وفق أسس علمية دقيقة: هناك عدة خطوات تعمل على إعداد بيانات نظيفة ومتوازنة وذات تنوع بصري مناسب، ما يمهد لبناء نموذج قادر على التقاط الاختلافات الدقيقة بين الأورام الحميدة والخبيثة، وهي اختلافات غالبًا ما تكون subtle وصعبة الاكتشاف دون معالجة مسبقة دقيقة.

1- توحيد حجم الصور

إعادة قياس الصور إلى أبعاد ثابتة (3×224×224) يحقق شرط التجانس بين المدخلات، مما يسهل عمليات حساب المصفوفات داخل النماذج ويمنع اختلاف الأبعاد من التأثير في حسابات المرشحات أو عمليات الالتفاف (Convolution). هذا التوحيد ضروري خصوصًا في البيانات الطبية التي تخضع لتفاوت ملحوظ في أجهزة التصوير وجودة الإخراج.

2- التحويل إلى التدرج الرمادي

إزالة المعلومات اللونية للصور غير المخصصة للتحليل العميق خطوة تساهم في تقليل حجم البيانات بنسبة كبيرة، مما يقلل الحمل الحوسبي دون التأثير في المعلومات التشخيصية، لأن تمييز الأورام يعتمد في الغالب على أنماط النسيج وليس على الألوان. هذا يقلل الضوضاء ويزيد من استقرار النموذج.

3- توسيع البيانات (Data Augmentation)

اعتماد تدوير الصور بزوايا 90°، 180°، و270° يُعد من الأساليب الفعالة لزيادة حجم البيانات الاصطناعية، مما يعزز قدرة النموذج على التعميم عبر رؤية الورم من اتجاهات متعددة. وتُظهر الدراسات أن زيادة التنوع البصري يقلل من خطر overfitting، وهو ما يُتوقع معه تحسن في دقة النموذج على بيانات الاختبار.

4-التطبيع (Normalization)

تطبيع قيم البكسلات إلى المجال [0,1] يضيف استقرارًا عدديًا أثناء عمليات التدريب، ويسهم في تسريع التقارب (Convergence) داخل النماذج الإحصائية والعميقة على حد سواء. كما يقلل تأثير الاختلافات في الإضاءة والحدة بين الصور.

5-عمليات القلب العشوائي

يضيف القلب الأفقي والعمودي تنوعًا بصريًا إضافيًا، ويكسر الأنماط المتكررة التي قد يعتمد عليها النموذج بشكل غير مرغوب. هذا التنوع يُتوقع أن يرفع حساسية النموذج تجاه الملامح التشريحية بغض النظر عن اتجاه ظهورها في الصورة الأصلية.

4-3 استخراج الميزات باستخدام LBP

يمثل اعتماد الأنماط الثنائية المحلية (Local Binary Patterns – LBP) نهجًا يدويًا فعالًا لاستخراج السمات الدقيقة من الصور الطبية، خصوصًا عند التعامل مع أنسجة ذات تباينات طفيفة مثل أنسجة الثدي.

تحول هذه الطريقة كل بكسل إلى قيمة ثنائية اعتمادًا على المقارنة بينه وبين جيرانه، مما يسمح بالكشف عن الملامح الدقيقة في النسيج، مثل الخشونة، الانتظام، أو التليّف؛ وهي سمات تُعد مؤشرات مهمة في التمييز بين الأورام.

ويُتوقع أن يؤدي استخدام LBP إلى:

- (a) تحسين دقة التصنيف عندما تكون العينات غنية بالملامح النسيجية.
- (b) تعزيز قدرة النماذج الإحصائية مثل الانحدار اللوجستي والغابات العشوائية على التعامل مع البيانات ذات الأبعاد العالية.
- (c) تقليل أثر اختلاف الإضاءة، لأن تمثيل LBP يعتمد على العلاقات المحلية بين البكسلات وليس على قيمها المطلقة.

4-4 خوارزميات التصنيف**اولا: تحليل نتائج عينة 10,000 صورة**

بعد الانتهاء من مراحل جمع البيانات ومعالجتها واستخراج الميزات النسيجية منها، تم تطبيق خوارزمية تصنيف تهدف إلى التمييز بين أورام الثدي الحميدة والخبيثة باستخدام مجموعة البيانات التي تضم (10,000) صورة. وقد اختيرت خوارزمية الغابات العشوائية (Random Forest) لما تمتاز به من كفاءة عالية في التعامل مع البيانات متعددة الأبعاد وقدرتها على تقليل الخطأ الناتج عن التحيز أو التباين في التصنيف.

تم إعداد نموذج الغابات العشوائية وفق المعلمات الآتية:

- (a) عدد الأشجار = 250 (Number of Trees)
- (b) أقصى عمق للشجرة = 9 (Maximum Depth)
- (c) أقصى عدد للميزات = (Maximum Features) الجذر التربيعي لعدد المتغيرات المدخلة (≈ 16)

بلغ زمن تدريب النموذج (26.44004) ثانية، في حين بلغ زمن التنفيذ (0.255608) ثانية. وقد أظهرت نتائج التدريب والاختبار أداءً متميزًا كما هو موضح في الجدول (1)

جدول (1) مؤشرات الأداء الإحصائية لبيانات التدريب والاختبار في خوارزمية الغابات العشوائية

المؤشر	بيانات التدريب	بيانات الاختبار
Accuracy	0.9945	0.9475
Precision	0.998488	0.964694
Recall	0.9905	0.929
F1	0.994478	0.94651
Specificity	0.9985	0.966
NPV	0.990575	0.931533
FPR	0.0015	0.034
LR+	660.3333	27.32353
LR-	0.009514	0.073499
DOR	69404.51	371.7539
MCC	0.989032	0.895613
Youden	0.989	0.895
ROC AUC	0.999906	0.991292
Average Precision	0.999907	0.991855
Cross Entropy	0.154687	0.250308
Brier Score	0.028514	0.063016
AUC Δ	0.499906	0.491292
wilson CI 95% low	0.992625	0.936839
wilson CI 95% high	0.9959	0.956445

1- تحليل النتائج

توضح النتائج أن النموذج أظهر أداءً قويًا جدًا في كل من بيانات التدريب والاختبار. بلغت الدقة (Accuracy) في التدريب (0.9945) وفي الاختبار (0.9475)، مما يدل على أن النموذج متوازن ولا يعاني من الإفراط في التعلم. كما حققت الدقة الإيجابية (Precision) قيمة مرتفعة في التدريب (0.998) وفي الاختبار (0.965)، ما يعكس قدرة النموذج على تقليل الأخطاء في التنبؤ بالحالات الإيجابية (الإيجابيات الكاذبة).

كذلك كانت الحساسية (Recall) عالية في كلتا المرحلتين (0.9905 و 0.929)، وهو ما يعني أن النموذج تمكن من الكشف عن معظم الحالات المصابة بدقة عالية. وتؤكد القيمة المرتفعة للمعدل التوافقي (F1 = 0.994) في التدريب والاختبار (0.947) على وجود توازن جيد بين الدقة الإيجابية والحساسية، مما يعزز موثوقية النموذج.

أظهرت النوعية (Specificity) قيمة مرتفعة (0.9985 في التدريب و 0.966 في الاختبار)، مما يشير إلى قدرة النموذج على تحديد الحالات السليمة بدقة عالية، في حين بلغت القيمة التنبؤية السالبة (NPV) (0.9906) في التدريب و 0.9315 في الاختبار، وهو ما يدل على موثوقية عالية في استبعاد الحالات غير المصابة.

أما بالنسبة للمؤشرات التشخيصية، فقد بلغت نسبة الاحتمال الموجبة (LR⁺) (660.33) في التدريب و 27.32 في الاختبار، وهي قيم مرتفعة تؤكد قدرة النموذج على تعزيز احتمال التشخيص الصحيح

للحالات الإيجابية. كما كانت نسبة الاحتمال السالبة (LR^-) منخفضة (0.0095) في التدريب و0.0735 (في الاختبار)، مما يدل على انخفاض احتمالية الخطأ في استبعاد الحالات الإيجابية. وبلغت النسبة التشخيصية (69404.51) (DOR) في التدريب و371.75 (في الاختبار)، وهي قيم مرتفعة تدل على قوة التمييز العالية للنموذج بين الفئتين. كما بينت النتائج أن معامل ارتباط ماثيوز (MCC) بلغ (0.989) في التدريب و0.896 (في الاختبار)، وهو مؤشر إحصائي قوي على جودة التصنيف وتوازن النموذج. وأكد مؤشر يودن (0.989) (Youden's Index) في التدريب و0.895 (في الاختبار) هذا التوازن بين الحساسية والنوعية. من الناحية الإحصائية، أظهرت مساحة تحت المنحنى (ROC) (AUC) قيمة مرتفعة جداً (0.9999) في التدريب و0.9913 (في الاختبار)، مما يعكس قدرة شبه مثالية للنموذج على التمييز بين الفئات. كما كانت الدقة المتوسطة (Average Precision) مرتفعة أيضاً (0.9999 و0.9919)، مما يدعم كفاءة النموذج في التصنيف. أما القيم المنخفضة لكل من خسارة الانتروبيا المتقاطعة ($Cross Entropy = 0.1547$) في التدريب و0.2503 (في الاختبار) ومقياس براير (Brier Score = 0.0285 و0.0630) فهي تعكس دقة التنبؤات الاحتمالية للنموذج واستقرارها. وأخيراً، فإن قيم فرق مساحة المنحنى (0.4999) (AUC Δ) في التدريب و0.4913 (في الاختبار) وفترات الثقة الضيقة وفق طريقة (Wilson) تشير إلى أن النتائج دقيقة وذات موثوقية عالية.

2- تفسير الأخطاء

على الرغم من الأداء المرتفع للنموذج في مختلف المؤشرات، إلا أن بعض الأخطاء ظهرت بنسب محدودة يمكن تفسيرها من خلال المقاييس الإحصائية. فقد بلغت نسبة الإيجابيات الكاذبة 0.0015 في بيانات التدريب و0.034 في الاختبار، وهي قيم منخفضة تُظهر أن معظم الحالات السليمة جرى تصنيفها بشكل صحيح، إلا أن الزيادة الطفيفة في بيانات الاختبار تُعزى إلى تنوع الصور الجديدة ووجود فروق دقيقة في الخصائص النسيجية لم تُسجل في بيانات التدريب. كما عكست نسبة الاحتمال السالبة قيمة صغيرة جداً (0.009514) و (0.073499) مما يشير إلى محدودية الخطأ في استبعاد الحالات المصابة، إلا أن الارتفاع البسيط في الاختبار يدل على وجود بعض الحالات التي حملت خصائص غامضة أو متداخلة، فظهرت بصورة أقرب للحالات السليمة. وتُفسر هذه الأخطاء أيضاً من خلال قيم الانتروبيا المتقاطعة ومقياس براير، إذ أظهرت ارتفاعاً طفيفاً في الاختبار مقارنة بالتدريب، مما يعكس زيادة بسيطة في عدم اليقين عند تقدير الاحتمالات. وبشكل عام، تبقى هذه الأخطاء ضمن الحدود الطبيعية المتوقعة عند انتقال النموذج من البيانات التدريبية إلى بيانات جديدة، ولا تؤثر في قوة الأداء أو موثوقية النتائج.

3- تحليل مصفوفة الارتباك (Confusion Matrix Analysis)

تعد مصفوفة الارتباك من أهم الأدوات الإحصائية المستخدمة في تقييم أداء نماذج التصنيف، إذ تتيح بيان عدد المشاهدات المصنفة تصنيفاً صحيحاً أو خاطئاً في كل فئة، مما يساعد على فهم نقاط القوة والضعف في النموذج. وفيما يلي عرض لمصفوفة الارتباك الخاصة ببيانات التدريب والاختبار لنموذج الغابات العشوائية.

أولاً: مصفوفة الارتباك لبيانات التدريب

يبين الجدول (2) مصفوفة الارتباك الخاصة ببيانات التدريب، حيث يُلاحظ ما يأتي:

- عدد المشاهدات السلبية (الأورام الخبيثة) التي تم التنبؤ بها بشكل صحيح بلغ (3994).
- عدد المشاهدات الإيجابية (الأورام الحميدة) التي تنبأ بها النموذج بشكل صحيح بلغ (3962).
- عدد الحالات السلبية الكاذبة (False Negatives = 6)، أي الحالات التي تنبأ فيها النموذج بوجود ورم خبيث بينما الحقيقة تشير إلى أنه حميد.

d. عدد الحالات الإيجابية الكاذبة (False Positives = 38) ، أي الحالات التي تنبأ فيها النموذج بوجود ورم حميد بينما الحقيقة تشير إلى أنه خبيث.

جدول (2) مصفوفة الارتباك لبيانات التدريب باستخدام خوارزمية الغابات العشوائية

الفئة الحقيقية/ الفئة المتنبأ بها	خبيث	حميد
خبيث	3994	6
حميد	38	3962

تُظهر النتائج دقة عالية في تصنيف الفئتين، إذ استطاع النموذج تصنيف معظم الأورام الخبيثة والحميدة بشكل صحيح. كما أن عدد الأخطاء منخفض للغاية، مما ينسجم مع المؤشرات الإحصائية التي قدمها النموذج (الحساسية، النوعية، F1، MCC وغيرها). يشير انخفاض السلبية الكاذبة إلى قدرة مرتفعة على كشف الأورام الحميدة بدقة، في حين أن انخفاض الإيجابية الكاذبة يدل على موثوقية التنبؤ بعدم وجود سرطان خبيث.

ثانياً: مصفوفة الارتباك لبيانات الاختبار

يبين الجدول (3) مصفوفة الارتباك الخاصة ببيانات الاختبار، حيث يُلاحظ ما يأتي:

a. عدد المشاهدات السلبية (الأورام الخبيثة) التي تم التنبؤ بها بشكل صحيح بلغ (966)

b. عدد المشاهدات الإيجابية (الأورام الحميدة) التي تم التنبؤ بها بشكل صحيح بلغ (929)

c. عدد الحالات السلبية الكاذبة (False Negatives = 34) ، أي الحالات التي تنبأ فيها النموذج بوجود ورم خبيث بينما الحقيقة تشير إلى أنه حميد.

d. عدد الحالات الإيجابية الكاذبة (False Positives = 71) ، أي الحالات التي تنبأ فيها النموذج بوجود ورم حميد بينما الحقيقة تشير إلى أنه خبيث.

جدول (3) مصفوفة الارتباك لبيانات الاختبار باستخدام خوارزمية الغابات العشوائية

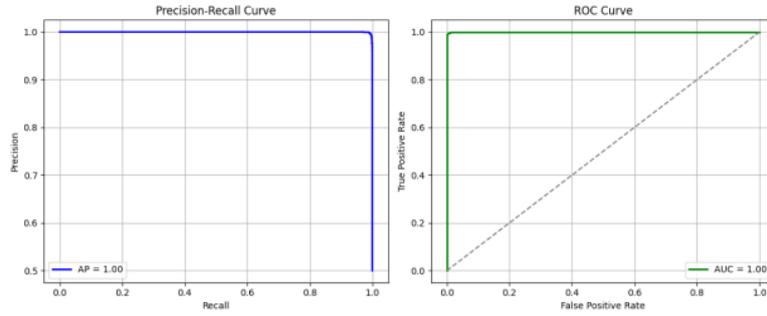
الفئة الحقيقية/ الفئة المتنبأ بها	خبيث	حميد
خبيث	966	34
حميد	71	929

ينتضح من مصفوفة الاختبار أن النموذج حافظ على أداء عالٍ، رغم أن عدد الأخطاء أعلى قليلاً مقارنة ببيانات التدريب، وهو أمر طبيعي ومتوقع عند اختبار النموذج على بيانات جديدة. النموذج أظهر قدرة ممتازة على التمييز بين الأورام، كما أن الأعداد المرتفعة للتصنيف الصحيح تدعم قدرة النموذج على التعميم وعدم انحيازه لبيانات التدريب فقط.

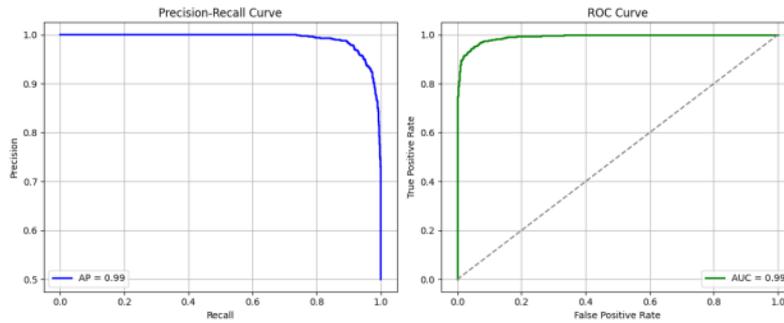
ثالثاً: منحنيات التشخيص للنموذج

a. الشكل (2) (يعرض منحنى (ROC إلى اليمين) ومنحنى (AP إلى اليسار) لبيانات التدريب والاختبار، إذ تظهر مساحة كبيرة تحت المنحنى، مما يدل على قدرة عالية للنموذج على الفصل بين الفئتين.

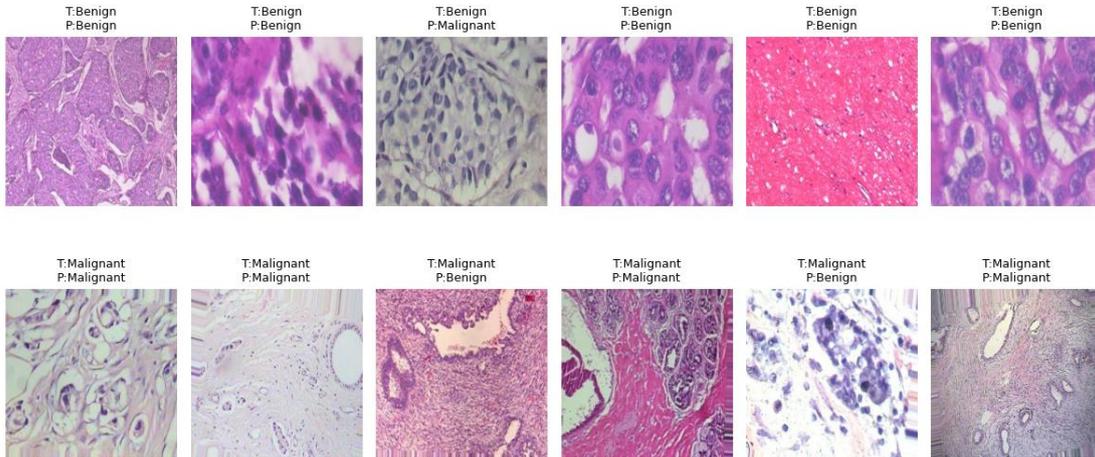
b. الشكل (3) يعرض مجموعة من الصور؛ الصف الأول يظهر عينات لأورام حميدة مع قيم التنبؤ الخاصة بها، أما الصف الثاني فيعرض عينات لأورام خبيثة والقيم التنبؤية المقابلة لها.



الشكل (2) يمثل منحى ROC على اليمين ومنحى AP على اليسار لبيانات التدريب باستعمال الغابات العشوائية



الشكل (3) يمثل منحى ROC على اليمين ومنحى AP على اليسار لبيانات التدريب باستعمال الغابات العشوائية



الشكل (4) الصف الأول يمثل صور لورم حميد قيمة التنبؤ الخاصة بها و الصف الثاني يمثل صور لورم خبيث و القيمة المتنبأ بها باستعمال الغابات العشوائية
 ثانياً: نتائج الاختبار على العينة 2400 صورة

بعد تحليل نتائج العينة الأولى المكونة من (10,000) صورة، تمت إعادة تطبيق خوارزمية الغابات العشوائية على العينة الثانية التي تضم (2,400) صورة، وذلك بهدف التحقق من قدرة النموذج على التعميم عند التعامل مع بيانات مستقلة، وفحص مدى ثبات الأداء الإحصائي للنموذج عند تنويع العينات. وقد جرى استخدام إجراءات المعالجة المسبقة واستخراج الميزات نفسها لضمان ثبات خطوات التحليل.

تم إعداد النموذج بالاعتماد على المعلومات الآتية:

- a. عدد الأشجار 100 :
b. أقصى عمق للشجرة 6 :
c. أقصى عدد للميزات في كل تقسيم: الجذر التربيعي لعدد المتغيرات (16 ≈)
وقد أجريت عملية التدريب والاختبار باستعمال مجموعة الصور المعالجة، واستخلصت مؤشرات الأداء الإحصائية كما هو موضح في الجدول (4).

جدول (4)

مؤشرات الأداء لبيانات التدريب والاختبار باستخدام خوارزمية الغابات العشوائية لعينة 2400

المؤشر	بيانات التدريب	بيانات الاختبار
Accuracy	0.983854	0.9375
Precision	0.988433	0.948718
Recall	0.979167	0.925
F1	0.983778	0.936709
Specificity	0.988542	0.95
NPV	0.97936	0.926829
FPR	0.011458	0.05
LR+	85.45455	18.5
LR-	0.021075	0.078947
DOR	4054.818	234.3333
MCC	0.967751	0.875274
Youden	0.967708	0.875
ROC AUC	0.998747	0.987188
Average Precision	0.998843	0.986932
Cross Entropy	0.247946	0.332209
Brier Score	0.05663	0.089952
AUCΔ	0.498747	0.487188
wilson CI 95% low	0.977174	0.912179
wilson CI 95% high	0.988602	0.955873

1- تحليل نتائج أداء النموذج

أظهرت النتائج أن النموذج حافظ على مستوى أداء قوي عند تطبيقه على العينة الثانية، إذ بلغت الدقة (Accuracy) 0.9839 في التدريب و 0.9375 في الاختبار، وهو أداء مرتفع يشير إلى قدرة جيدة على التعميم مع انخفاض طفيف ومنطقي عند الانتقال لبيانات جديدة.

كما بلغت الدقة الإيجابية (Precision) 0.988 في التدريب و 0.949 في الاختبار، وهو ما يدل على قدرة جيدة على تقليل الإيجابيات الكاذبة. أما الحساسية (Recall) فقد بلغت 0.979 في التدريب و 0.925 في الاختبار، مما يعكس قدرة مناسبة على كشف الحالات المصابة بدقة. بلغ معامل F1 0.984 في التدريب و 0.937 في الاختبار، وهو مؤشر على توازن جيد بين الدقة الإيجابية والحساسية في كلتا الحالتين. وفيما يتعلق بالنعوية (Specificity)، فقد سجلت 0.9885 في التدريب و 0.95 في الاختبار، وهو ما يشير إلى دقة عالية في تحديد الحالات السليمة، بينما بلغت القيمة التنبؤية السلبية (NPV) 0.979 في التدريب و 0.927 في الاختبار، مما يدل على موثوقية جيدة في استبعاد الحالات غير المصابة. كما كان معدل الإيجابيات الكاذبة (FPR) منخفضاً في التدريب (0.011) وارتفع قليلاً في الاختبار إلى (0.05) ضمن مستوى مقبول للتعميم. من جهة أخرى، حققت المؤشرات التشخيصية أداءً قوياً؛ إذ بلغت نسبة الاحتمال الموجبة (LR+) 18.5 في الاختبار، وهي قيمة جيدة تشير إلى قوة الدليل التشخيصي في حالة التنبؤ الإيجابي، بينما كانت نسبة الاحتمال السالبة (LR-) منخفضة (0.079)، مما يعزز القدرة على استبعاد الحالات السلبية بشكل دقيق. كما سجلت النسبة التشخيصية (DOR) قيمة عالية بلغت 234.33، مما يعكس تمييزاً قوياً بين الفئتين. بلغ معامل ارتباط ماثيوز (MCC) 0.968 في التدريب و 0.875 في الاختبار، وهي قيم مرتفعة تؤكد جودة التوازن في التصنيف. كما قدم مؤشر Youden قيمة مقاربة 0.968 و 0.875، ما يشير إلى تكامل الأداء بين الحساسية والنعوية.

وعلى مستوى المؤشرات الرسومية، حقق النموذج مساحة تحت المنحنى ROC AUC بلغت 0.9987 في التدريب و 0.9872 في الاختبار، وهي قيم عالية جداً تدل على أداء شبه مثالي في التمييز بين الفئات. أما خسارة الانتروبيا المتقاطعة ومقياس براير فقد كانت منخفضة نسبياً (0.2479 و 0.0566 في التدريب، و 0.3322 و 0.0899 في الاختبار)، مما يعكس جودة التنبؤات الاحتمالية وثبات النموذج. كما أن نطاقات الثقة (Wilson CI 95%) جاءت ضيقة نسبياً، مما يعزز موثوقية النتائج على مستوى التقدير الإحصائي.

2- مصفوفة الارتباك لعينة 2400 صورة

تُعد مصفوفة الارتباك من أهم المقاييس التشخيصية في تقييم أداء نماذج التصنيف، إذ تُمكن من معرفة عدد الحالات المصنّفة بشكل صحيح أو خاطئ في كل فئة. وفيما يلي عرض النتائج الخاصة بنموذج الغابات العشوائية عند تطبيقه على العينة الثانية.

أولاً: مصفوفة الارتباك لبيانات التدريب العينة الثانية

يوضح الجدول (5) مصفوفة الارتباك لبيانات التدريب، ويتبين منها ما يأتي:

- عدد المشاهدات السلبية (أورام خبيثة) التي تم التنبؤ بها بشكل صحيح (949): صورة.
- عدد المشاهدات الإيجابية (أورام حميدة) التي تم التنبؤ بها بشكل صحيح (940): صورة.
- عدد الحالات السلبية الكاذبة: (False Negatives = 11) وهي الحالات التي تنبأ فيها النموذج بوجود ورم خبيث بينما الحقيقة تشير إلى ورم حميد.
- عدد الحالات الإيجابية الكاذبة: (False Positives = 20) وهي الحالات التي تنبأ فيها النموذج بوجود ورم حميد بينما الحقيقة تشير إلى ورم خبيث.

جدول (5) مصفوفة الارتباك لبيانات التدريب باستخدام خوارزمية الغابات العشوائية

القيمة الحقيقية/ القيمة المتنبأ بها	خبث	حميد
خبث	949	11
حميد	20	940

تشير القيم أعلاه إلى دقة عالية للنموذج في تصنيف كل من الأورام الحميدة والخبثية، مع نسبة منخفضة من الأخطاء، وهو ما ينسجم مع مؤشرات الأداء الإحصائية التي سبق عرضها. ويعكس العدد المحدود لكل من السلبية والإيجابية الكاذبة كفاءة جيدة للنموذج أثناء التدريب.

ثانياً: مصفوفة الارتباك لبيانات الاختبار العينة 2400

يبين الجدول (6) مصفوفة الارتباك الخاصة ببيانات الاختبار، ويلاحظ ما يأتي:

- عدد المشاهدات السلبية (أورام خبيثة) المتنبأ بها بشكل صحيح (228): صورة.
- عدد المشاهدات الإيجابية (أورام حميدة) المتنبأ بها بشكل صحيح (222): صورة.
- عدد السلبية الكاذبة: (False Negatives = 12) حالات صُنفت خطأ كأورام خبيثة بينما هي حميدة.
- عدد الإيجابية الكاذبة: (False Positives = 18) حالات صُنفت خطأ كأورام حميدة بينما هي خبيثة.

جدول (6) مصفوفة الارتباك لبيانات الاختبار باستخدام خوارزمية الغابات العشوائية

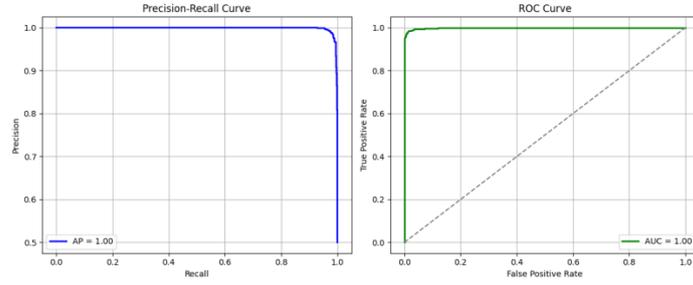
القيمة الحقيقية/ لقيمة المتنبأ بها	خبث	حميد
خبث	228	12
حميد	18	222

يتضح من مصفوفة الاختبار أن النموذج حافظ على قدرة جيدة في التنبؤ الصحيح، رغم ارتفاع طفيف في عدد الأخطاء مقارنة ببيانات التدريب، وهو أمر منطقي عند التعامل مع بيانات جديدة لم يسبق للنموذج رؤيتها. وتبقى النسب العامة للتصنيف الصحيح مرتفعة، مما يؤكد كفاءة النموذج في التعميم.

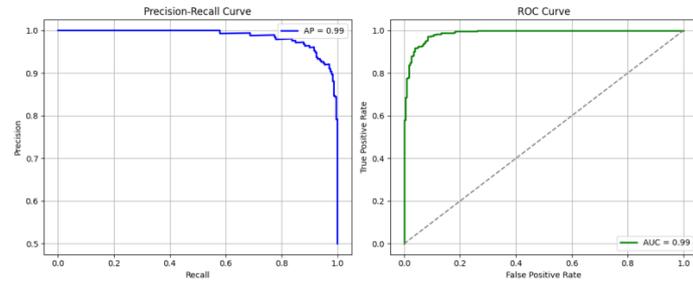
ثالثاً: المنحنيات التشخيصية للعينة الثانية

- الشكل (5): يمثل منحنى (ROC على اليمين) ومنحنى (Average Precision – AP على اليسار) لبيانات التدريب، حيث تظهر مساحة كبيرة تحت المنحنى، مما يدل على الأداء القوي للنموذج.
- الشكل (6): يعرض منحنى ROC ومنحنى AP لبيانات الاختبار، والتي تبين ثبات الأداء مع العينة المستقلة.

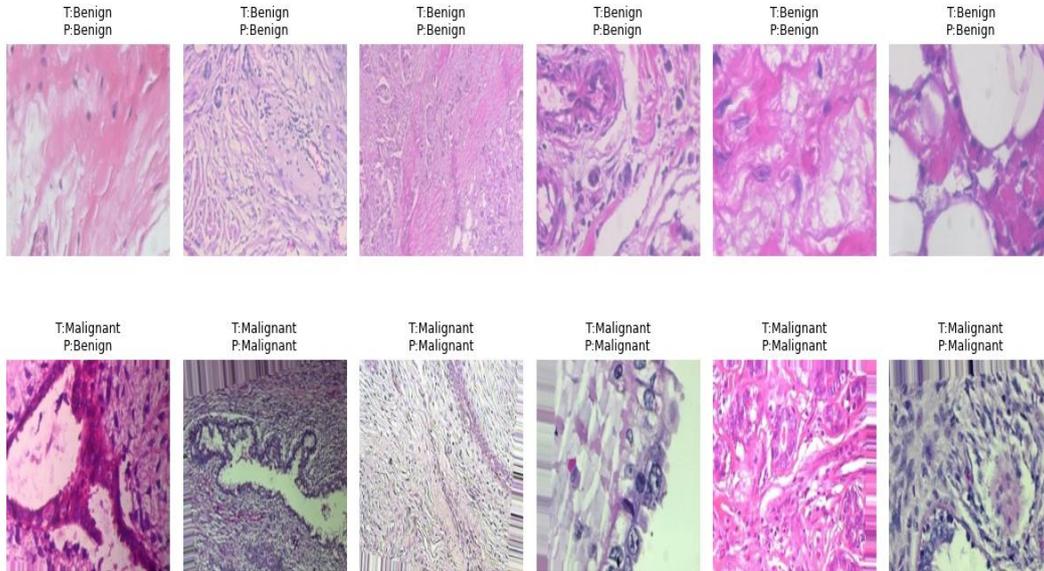
الشكل (7) يحتوي على مجموعة من الصور؛ الصف الأول يعرض عينات أورام حميدة مع القيم التنبؤية المرتبطة بها. الصف الثاني يعرض صور أورام خبيثة مع قيم التنبؤ المقابلة. ويهدف هذا الشكل إلى بيان كيفية استجابة النموذج للعينات البصرية من كلا الفئتين.



الشكل (5) يمثل منحى ROC على اليمين ومنحى AP على اليسار لبيانات التدريب باستعمال الغابات العشوائية



الشكل (6) يمثل منحى ROC على اليمين ومنحى AP على اليسار لبيانات الاختبار باستعمال الغابات العشوائية



الشكل (7) الصف الأول يمثل صور لورم حميد قيمة التنبؤ الخاصة بها والصف الثاني يمثل صور لورم خبيث والقيمة المتنبأ بها باستعمال الغابات العشوائية
ثالثاً: مقارنة اداء خوارزمية الغابات العشوائية بين العينتين
تم إجراء مقارنة شاملة لأداء خوارزمية الغابات العشوائية باستخدام مجموعتي البيانات المكونتين من (10,000) صورة و(2,400) صورة لأورام الثدي، وذلك بهدف تقييم قدرة النموذج على التنبؤ بدقة، ومدى استقراره، وفعاليته في التعميم عند التعامل مع عينات مختلفة الحجم.

جدول (10)

مقارنة مؤشرات أداء خوارزمية الغابات العشوائية بين العينتين (10,000، 2400) صورة

المؤشر	FR (10000)	FR (2400)
Accuracy	0.9475	0.9375
Precision	0.964694	0.948718
Recall	0.929	0.925
F1	0.94651	0.936709
Specificity	0.966	0.95
NPV	0.931533	0.926829
FPR	0.034	0.05
LR+	27.32353	18.5
LR-	0.073499	0.078947
DOR	371.7539	234.3333
MCC	0.895613	0.875274
Youden	0.895	0.875
ROC AUC	0.991292	0.987188
Average Precision	0.991855	0.986932
Cross Entropy	0.250308	0.332209
Brier Score	0.063016	0.089952
AUCΔ	0.491292	0.487188
wilson CI 95% low	0.936839	0.912179
wilson CI 95% high	0.956445	0.955873

1-تحليل نتائج المقارنة

تكشف نتائج الجدول عن تفوق واضح لأداء النموذج عند استخدام العينة الأكبر (10,000 صورة)، إذ حققت معظم المؤشرات الأساسية، مثل الدقة (Accuracy)، والدقة الإيجابية (Precision)، والمساحة تحت منحنى ROC، قيمة أعلى مقارنة بالعينة الأصغر. ويعكس هذا التفوق الدور الإحصائي المهم لحجم العينة في تعزيز قدرات النموذج على التعميم ودقة التنبؤ. تفسير ذلك يعود إلى أن العينة الكبيرة توفر تنوعاً أكبر في الأنماط النسيجية داخل الصور، مما يمكن النموذج من تعلم خصائص مميزة أكثر اتساعاً، ويقلل من احتمالية الانحياز. كما ترتفع قيم المؤشرات التشخيصية LR+ و DOR عند استخدام العينة الأكبر، مما يدل على قوة أعلى للنموذج في التمييز بين الأورام الحميدة والخبيثة. وعلى الرغم من هذا التفوق العددي، إلا أن العينة الصغيرة أظهرت أداءً مستقرًا وقويًا، إذ بقيت معظم المؤشرات ضمن مستويات مرتفعة، ما يعكس قابلية النموذج للعمل بكفاءة حتى عند توفر بيانات محدودة. ويلاحظ تقارب قيم الحساسية (Recall) والقيمة التنبؤية السالبة (NPV) بين العينتين، مما يوضح أن قدرة النموذج على اكتشاف الحالات المصابة واستبعاد السليمة قد حافظت على مستوى

جيد. وبشكل عام، تؤكد هذه النتائج أن خوارزمية الغابات العشوائية تتمتع بمتانة إحصائية واضحة، وأداء ثابت في مختلف أحجام العينات، إلا أن العينة الأكبر تمنح النموذج قوة أعلى في التمييز واستقراراً أكبر في تقديرات المؤشرات.

5- الخاتمة

يهدف هذا البحث إلى تقييم كفاءة خوارزمية الغابات العشوائية في تمييز أورام الثدي الحميدة والخبيثة بالاعتماد على صور تشريحية معالجة ومُستخلصة الميزات باستخدام الطرائق اليدوية (Handcrafted Features) وبالأخص أنماط LBP. ومن خلال تطبيق الخوارزمية على عينتين مختلفتي الحجم؛ العينة الأولى المكونة من (10,000) صورة والعينة الثانية المكونة من (2,400) صورة، أمكن تحليل أداء النموذج بدقة عالية ووفق مجموعة واسعة من المؤشرات التشخيصية والإحصائية. أظهرت نتائج العينة الكبيرة تفوقاً واضحاً في أغلب مؤشرات الأداء، ولا سيما الدقة (Accuracy)، والدقة الإيجابية (Precision)، ومنطقة تحت المنحنى ROC AUC، ومؤشر F1، مما يعكس قوة نموذج الغابات العشوائية عند توفر بيانات أكبر وأكثر تنوعاً، ويُحسن قدرته على التعميم وتقليل احتمالية الانحياز. أما نتائج العينة الأصغر فقد حافظت على مستويات أداء مرتفعة وثابتة، مؤكدة قدرة الخوارزمية على العمل بكفاءة حتى مع البيانات المحدودة، رغم الانخفاض النسبي المتوقع في بعض المؤشرات مثل LR+ و DOR نتيجة قلة عدد العينات التدريبية. كما بينت مصفوفات الارتباك في كلا العينتين فعالية النموذج في الحد من الإيجابيات الكاذبة والسلبيات الكاذبة، مع تحقيق توازن جيد بين الحساسية والنوعية، وهو ما تعززه القيم المرتفعة لمعامل ماثيوز (MCC) ومؤشر يودن (Youden). وتؤكد المؤشرات الاحتمالية مثل Brier Score و Cross Entropy جودة التنبؤات الاحتمالية للنموذج واستقرار نتائجه. وبناءً على المقارنة الشاملة، يتضح أن خوارزمية الغابات العشوائية تُعد خياراً موثوقاً وفعالاً لتصنيف أورام الثدي اعتماداً على ميزات نسيجية مستخرجة يدوياً، كما أن أدائها يستفيد بشكل ملحوظ من زيادة حجم العينة. وتُبرز هذه النتائج أهمية استخدام مجموعات بيانات كبيرة قدر الإمكان، إضافة إلى أهمية إجراء معالجة مسبقة دقيقة واستخراج ميزات ملائمة لتحقيق أعلى مستويات الدقة. وتفتح هذه الدراسة المجال لاعتماد منهجيات هجينة مستقبلية، تجمع بين الغابات العشوائية وطرائق التعلم العميق أو خوارزميات تحسين متقدمة، بما يعزز القدرة على تحليل صور أكثر تعقيداً وزيادة كفاءة النماذج في التطبيقات الطبية الذكية.

Performance Evaluation of the Random Forest Algorithm in Classifying Breast Tumors Based on Histological Image Feature

Haifa Taha Abd

Rasel Hassan Katif

haefaa_adm@uomustansiriyah.edu.iq

rsisii34344@gmail.com

Department of statistic/college of Administration and
Economics/Mustansiriyah University

Abstract

This study aims to evaluate the efficiency of a classification approach based on medical image analysis for diagnosing breast tumors and distinguishing between benign and malignant patterns. The analysis relied on two sets of digital histological images, comprising 10,000 images in the first set and 2,400 images in the second, after subjecting them to a series of preprocessing steps that included image enhancement, dimension standardization, and the use of texture-analysis techniques to extract features that describe the tumor's micro-structural characteristics. A classification model was then constructed using the Random Forest method, and its performance was assessed through a comprehensive set of statistical indicators. The findings demonstrate a high capability of the model to differentiate between the two classes, achieving elevated accuracy levels and displaying stable performance despite the variation in sample size. The results further reveal that the Random Forest algorithm exhibits statistical robustness and consistent performance, while larger samples contribute to stronger discriminatory power and improved stability of the evaluation metrics. Overall, the study underscores the significant role of preprocessing and texture-feature extraction in enhancing diagnostic quality, supporting the potential use of this approach in medical applications concerned with tumor characterization.

Key word: classification, Random Frost, Texture Image, Breast Tumors.