

استخدام الذكاء الصناعي (أشجار القرار Decision Trees) في التعرف على

مرضى سرطان الثدي

* أم.د.م. علي أحمد عجيب *

* د.م. منهل صالح المحمد *

* م. ديانا حسين الحسن *

(تاريخ الإيداع 5 / 7 / 2021 . قُبِلَ للنشر في 24 / 11 / 2021)

□ ملخّص □

يعتبر سرطان الثدي من أخطر الأمراض التي تهدد حياة المرأة، حيث أنه يصيب ما يقرب 10% من النساء في جميع أنحاء العالم في مرحلة ما من حياتهن. يلعب التعلم الآلي في هذا المجال دوراً حيوياً في تشخيص الأمراض بشكل عام ومنها السرطانات، حيث يستخدم كطريقة فعالة لتصنيف البيانات. استخدمت في هذا البحث تقنية شجرة القرار للتعرف على مرضى سرطان الثدي، واستخدمت الحزمة البرمجية Matlab R2018b لبرمجة النظام المقترح، حيث أظهرت نتائج مشجعة للتعرف على مرضى السرطان. كما أجريت مجموعة من التجارب المقارنة بين النموذج المقترح في هذا البحث مع نفس النموذج باستخدام مفهوم الشبكات العصبونية الاصطناعية الذكية ((Back Propagation Neural Network(BPNN))، ونموذج خوارزمية الدعم الآلي ((Support Vector Machine (SVM))، ومصنف الجار الأقرب (K Nearest Neighbors (KNN))، حيث تبين أن مصنف شجرة القرار المستخدم في النظام المقترح يسلك سلوكاً فريداً أدى إلى اختياره في هذا البحث، حيث أن دقة التعرف تزداد مع زيادة عدد الأشخاص، على خلاف باقي المصنفات. **كلمات مفتاحية:** الذكاء الصناعي، شجرة القرار، مرض سرطان الثدي.

*أستاذ دكتور مساعد في كلية الهندسة المعلوماتية في أكاديمية الأسد للهندسة العسكرية.

**دكتور مهندس في كلية الهندسة المعلوماتية في أكاديمية الأسد للهندسة العسكرية.

Using artificial intelligence (Decision Trees) to identify breast cancer patients

Ali Ahmed Ajeeb *

Manhal Saleh Al-Muhammad **

Diana Hussein Al-Hassan ***

(Received 5 / 7 / 2021 . Accepted 24 / 11 / 2021)

□ ABSTRACT □

Breast cancer is considered one of the most dangerous diseases that threaten a woman's life, as it affects approximately 10% of women worldwide at some point in their lives.

Machine learning plays a vital role in this field in diagnosing diseases in general, including cancers, as it is used as an effective way to classify data.

In this research, decision tree technology was used to identify breast cancer patients, and the Matlab R2018b software package was used to program the proposed system, as it showed encouraging results for identifying cancer patients.

A set of experiments were also conducted comparing the model proposed in this research with the same model using the Back propagation Neural Network (BPNN), the Support Vector Machine SVM model, and the K Nearest Neighbors (KNN) classifier. Where it was found that the decision tree classifier used in the proposed system behaves a unique behavior that led to its selection in this research, as the accuracy of the identification increases with the increase in the number of people, unlike the rest of the classifiers.

Keywords Artificial intelligence, decision tree, breast cancer.

*Assistant Professor, Informatics College, AL-Assad Academy, For Military Engineering

**PhD, Informatics College, AL-Assad Academy, For Military Engineering

***Engineer, Informatics College, AL-Assad Academy, For Military Engineering

1- مقدمة (Introduction) :

يعتبر مرض السرطان السبب الثاني لحالات الوفاة على مستوى العالم، ومن بين جميع أنواع السرطان يعتبر سرطان الثدي (Breast Cancer) هو الأكثر شيوعاً بين النساء على مستوى العالم وفقاً لأرقام منظمة الصحة العالمية، حيث تسبب في وفاة 71000 امرأة في عام 2015 [1].

لتوفير التشخيص الدقيق لسرطان الثدي هناك حاجة لأنظمة الكشف بمساعدة الكمبيوتر (Computer Aided Detection (CAD) والتي تستخدم نهج التعلم الآلي. تساعد أنظمة الـ CAD في اكتشاف سرطان الثدي في مرحلة مبكرة، مما يزيد معدل البقاء على قيد الحياة للمصابين بهذا المرض [3]، [2].

هناك العديد من الأعمال التي نفذت في مجال التعليم الآلي ومنها:

نشر الباحث Shweta Kharya في عام 2014 بحثاً لتطوير نظام التوقع الاحتمالي لسرطان الثدي باستخدام مصنفات (Naive Bayes)، والتي يمكن استخدامها في اتخاذ القرار الصحيح، حيث حققت دقة أقصاها 93% [4]. قدمت الباحثة هبة أسري وآخرون مقالة في عام (2016) حيث قارنوا بين أداء خوارزميات مختلفة للتعلم الآلي وهي:

(SVM, Decision Tree (C4.5), Naive Bayes (NB), KNN)

والهدف الرئيسي للمقالة هو تقييم مدى صحة تصنيف البيانات فيما يتعلق بكفاءة وفعالية كل خوارزمية من حيث الدقة والحساسية والخصوصية، حيث أظهرت النتائج التجريبية أن SVM يعطي أعلى دقة وهي [5] (97.13%).

قدم الباحثان (Garima Verma and Hemraj Verma) في عام (2019) دراسة للتنبؤ بنوع سرطان الثدي، تم استخدام خوارزمية دعم القرار (SVM)، وأظهرت النتائج أن الدقة في SVM (90.3%) [6].

قدم الباحث Prabhjot Kaur وآخرون في عام (2019) مقالة تقترح سمات جديدة للتغلب على القيود الموجودة، والتي ستزيد من فعالية أنظمة التنبؤ بالسرطان، وتمت مقارنة كفاءة السمات الحالية والمقترحة من خلال معالجة مجموعات بيانات باستخدام خوارزميات استخراج البيانات بالاعتماد على أداة WEKA، وتم استخدام الخوارزميات التالية (Naive Bayes, KStar, Bagging, Decision Tree). أفاد التحليل التجريبي الذي تم إجراؤه في الورقة البحثية عن تحسن في كفاءة التنبؤ بالسرطان مقارنة بأنظمة التنبؤ الحالية [7].

قدم الباحث أسامه رزق شاهين وآخرون عام (2012) مقالة بعنوان اكتشاف أورام الثدي من صور الأشعة السينية باستخدام خوارزمية K-means، تم في هذا البحث اقتراح تطوير جديد للخوارزمية K-means للكشف المبكر عن سرطان الثدي من خلال تجزئة صورة الأشعة السينية للثدي، ويهدف المقترح إلى تقسيم صورة الثدي لتجميع تلك البيكسل التي تظهر مماثلة للأنسجة وفقاً لعدد معين من التكتلات حيث يتم تحديد عدد من المجموعات وفقاً لعدد من الإشارات في الرسم البياني للصورة (histogram). تم اختبار الطريقة المقترحة على عدة صور مأخوذة من قواعد بيانات نظام تقرير صور الثدي BIRADS وتم مقارنة النتائج مع التصنيف القياسي (Mammogram MIAS) [8].

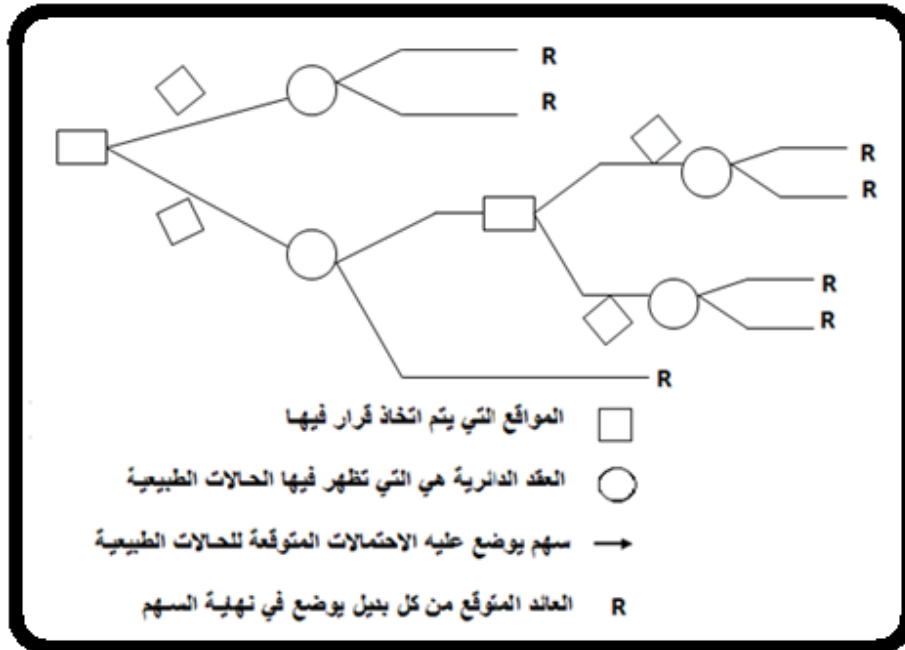
2- أهمية البحث:

تتلخص أهمية البحث في النقاط التالية:

- (1) التعرف على شجرة القرار.
- (2) دعم القرار في التعرف على مرضى سرطان الثدي.
- (3) توضيح بنية النظام المقترح للتعرف على مرضى سرطان الثدي.

3- دراسة عامة حول شجرة القرار:

شجرة القرار هي مصنف على شكل هيكل شجرة حيث كل عقدة هي إما عقدة ورقة تشير إلى قيمة السمة (الهدف)، أو عقدة قرار، وتستخدم شجرة القرار عندما يكون هناك صعوبة أمام متخذ القرار ببناء جدول النتائج الشرطية سواء كان معبراً عن هذه النتائج بالأرباح أو الخسائر أو المنفعة، حيث تشير المربعات في هذه الشجرة إلى المواقع التي يتخذ فيها القرار أما العقد الدائرية فتشير إلى المواقع التي تظهر فيها الحالات الطبيعية وهناك أسهم تصل ما بين المربعات ويوضع عليها الاحتمالات المتوقعة للحالات الطبيعية أما القيم النهائية المعروفة بالعوائد (نتائج القرار) توضع في نهاية الأسهم المعبرة عن كل نتيجة نحصل عليها من كل حالة من الحالات الطبيعية[9]. يوضح الشكل (1) مخططاً بيانياً عاماً لتمثيل شجرة القرار.



الشكل (1): مخطط بياني عام لتمثيل شجرة القرار [9].

3-1 مزايا وعيوب شجرة القرار:

تتميز شجرة القرار عن باقي أدوات دعم القرار بالعديد من المزايا:

- سهولة الفهم والتفسير، حيث يمكن للأفراد العاديين فهم نماذج شجرة القرار بعد شرح مختصر.

• لها قيمة حتى بوجود القليل من البيانات، فمن الممكن تشكيل مهمة بناءً على توصيف الخبراء للوضع (والوضعيات البديلة والتكاليف) وتفاصيل الخرج.

• كما أنها تسمح بإضافة سيناريوهات جديدة ممكنة.

• تساهم في تحديد أسوأ وأفضل قيمة متوقعة للسيناريوهات المختلفة.

• من الممكن دمجها مع باقي تقنيات القرار [10].

عيوب شجرة القرار :

• عند استخدام بيانات تتضمن متغيرات فئوية بأعداد مختلفة من المستويات، فإن المعلومات الناتجة من

شجرة القرار تكون منحازة للصفات التي تحتوي على مستويات أكثر.

• يمكن للحسابات أن تكون معقدة جداً خصوصاً إذا كانت العديد من القيم غير أكيدة [10].

3-2 خطوات رسم شجرة القرارات:

إن رسم شجرة القرارات لا يتم بشكل اعتباطي بل وفق قواعد وخطوات محددة وواضحة في ضوء البيانات المتوفرة عن المشكلة، وكلما كان الشكل البياني معبراً بشكل كامل وصحيح عن أصل المشكلة وتفرعاتها، كلما كان ذلك عاملاً مساعداً وأساسياً في التوصل إلى حلها، وبشكل عام توجد خطوات متسلسلة تستخدم في عملية رسم وتحليل شجرة القرار [11].

يتم رسم شجرة القرار بداية من يسار الصفحة والاتجاه نحو اليمين:

1- رسم شجرة القرار باستخدام المربعات للتعبير عن القرارات واستخدام الدوائر للتعبير عن الحالات الطبيعية.

2- تقييم شجرة القرار بغرض التأكد من احتوائها على كل العوائد المحتملة.

3- حساب قيم الشجرة ابتداءً من اليمين والاتجاه نحو اليسار.

4- حساب القيم المتوقعة للبدائل من خلال ضرب قيم العوائد في احتمالاتها [11].

هناك صيغ معروفة لبناء شجرة القرار المستخدمة من قبل خوارزمية (التصنيف والانحدار)، لأنه يقوم على مبدأ

الاحتمال المربع لكل هدف في العقدة ويصل إلى الحد الأدنى لها (أي الصفر) عند كل الحالات [12].

يعتمد بناء شجرة القرار على التوزيع احتمالي لمجموعة البيانات (S) عندئذ فإن المعلومات المحتملة بهذا التوزيع

تسمى $Entropy(S)$ ، حيث تُعطى وفق المعادلة التالية:

$$H(S) = Entropy(S) = \sum_{c \in C} -P(c) * \log_2(P(c)) \dots \dots \dots (1)$$

حيث:

S : هي مجموعة البيانات الحالية المراد حساب الانتروبي لها (والتي تتغير عند كل مرور في خوارزمية ID3).

C : هي مجموعة الأصناف وفي نظامنا المقترح هي {Yes, No}.

P(c) : نسبة عدد العناصر في الفئة (c) إلى عدد العناصر الكلية في المجموعة (S).

عندما تكون $H(S) = 0$ يتم تصنيف المجموعة (أي أن جميع العناصر في S من نفس الفئة).

في خوارزمية ID3 يتم حساب $Entropy$ لكل سمة متبقية، يتم استخدام السمة التي تحتوي على أصغر قيمة

$Entropy$ لتقسيم المجموعة S على هذا التكرار، كلما زادت قيمة $Entropy$ زادت إمكانية تحسين التصنيف هنا.

لتحديد وظيفة اختيار الاختبار الذي يجب أن يعنون العقدة الحالية، يمكننا استخدام تابع الريح أو تابع كسب

المعلومات (Information gain IG(A)) وهو مقياس الاختلاف في الانتروبي قبل وبعد تقسيم المجموعة S على

السمة A، بمعنى آخر هي انخفاض مقدار عدم اليقين بعد تقسيم المجموعة S على السمة A. ويستخدم أيضاً لترتيب السمات وبناء شجرة القرار حيث يوجد في كل عقدة سمة ذات أعلى كسب للمعلومات [12].
تعطى علاقة الربح لكل عقدة بالمعادلة التالية :

$$IG(A, S) = H(S) - \sum_{t \in T} (P(t) * H(t)) \dots \dots \dots (2)$$

H(S): الانتروبي للمجموعة S.

T : المجموعات الفرعية التي تم إنشاؤها من تقسيم المجموعة S حسب السمة A.

P(t): نسبة عدد العناصر في t إلى عدد العناصر في المجموعة S.

H(t): الانتروبي للمجموعة الفرعية T.

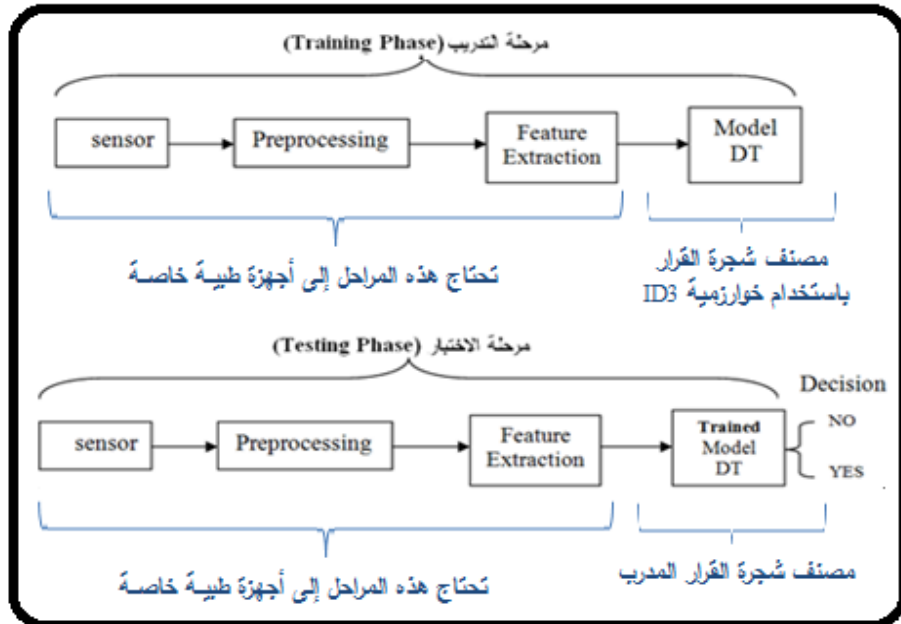
في ID3 يمكن حساب كسب المعلومات (بدلاً من الانتروبيا) لكل سمة متبقية، حيث يتم استخدام السمة ذات الكسب الأكبر للمعلومات لتقسيم المجموعة S على هذا التكرار.

4- النموذج المقترح للتعرف على مرضى سرطان الثدي:

يمكن توصيف مشكلة التعرف على مرضى سرطان الثدي من خلال البيانات المستخلصة من التحاليل المأخوذة للمرضى من قبل الأطباء والمخزنة في قاعدة بيانات عالمية.

4-1 بنية النظام المقترح:

إن النظام المقترح للتعرف على مرضى سرطان الثدي يتضمن مرحلتين أساسيتين: مرحلة التدريب ومرحلة الاختبار (التعرف)، يبين الشكل (2) بنية النظام المقترح.



الشكل (2): بنية النظام المقترح للتعرف على مرضى سرطان الثدي.

يبين الشكل (2) مرحلة التدريب بمراحلها الأربعة حيث تحتاج المراحل الثلاث التالية (sensor, Preprocessing, Feature Extraction) إلى أجهزة طبية ومختصين للحصول عليها لذلك تم الاعتماد على تحميل قاعدة بيانات جاهزة خاصة بمرضى سرطان الثدي تم استخراج شعاع الميزات فيها لتدريب النموذج المقترح عليها، في المرحلة الرابعة تم الحصول على الموديل المدرب باستخدام شجرة القرار، وضمن مرحلة الاختبار سنمر في نفس المراحل السابقة ولكن الموديل المدرب سيعطي قراراً فيما إذا كانت العينات المدخلة لشخص سليم أو مريض.

5- المناقشة و نتائج التجارب: (Discussion and Experimental Results)

لتقييم النموذج المقترح، تم تنفيذ بعض التجارب باستخدام الحزمة البرمجية Matlab R2018b على حاسوب مكتبي بالمواصفات التالية:

المعالج: Intel® Core™ i3-7100 CPU @ 3.90 GHz

الذاكرة المثبتة (RAM): 4GB

نظام التشغيل: 64Bit

نفذت مجموعة من الاختبارات على قاعدة البيانات الخاصة بمرضى سرطان الثدي، بغية التصنيف باستخدام شجرة القرار.

إن قاعدة البيانات الخاصة بمرضى سرطان الثدي المستخدمة في هذا البحث تحتوي أشعة الميزات المستتبهة من مرحلة استخراج السمات ويمكن تضمينها في Matlab R2018b من خلال التعليمة.

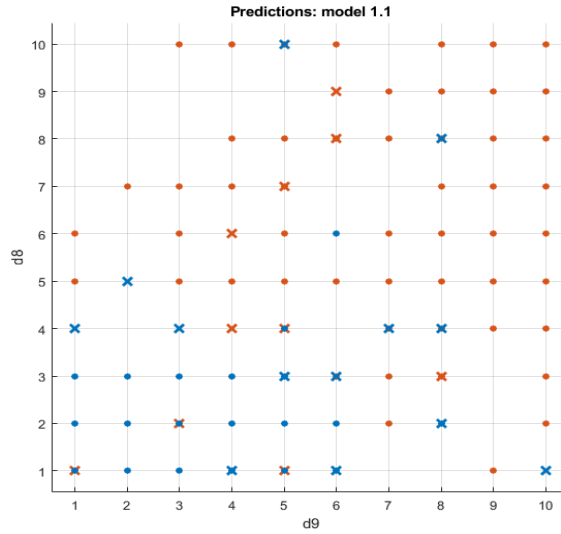
[d,g] = breast_cancer_dataset.

يمثل كل عنصر في المتجه d عينات الدم لمريض واحد من أصل 569 مريض، يحتوي المتجه g على 569 عنصر كل عنصر منها يمثل بقيمة [0] أو [1]، القيمة [0] تشير إلى مريض بسرطان الثدي، القيمة [1] تشير إلى مريض عادي.

توضح التجربة كيفية إنشاء ومقارنة أشجار التصنيف المختلفة باستخدام معلم التصنيف (Classification Learner)، نقوم بتحميل مجموعة البيانات المراد تطبيق الخوارزميات المطلوبة لاستخدامها في التصنيف.

في علامة التبويب التطبيقات (Apps)، في مجموعة التعلم الآلي (Machine Learning group)، انقر فوق تصنيف المتعلم (Classification Learner)، ثم انقر فوق جلسة جديدة (New Session).

لإنشاء نموذج شجرة التصنيف، في علامة التبويب Classification Learner، في قسم نوع النموذج نختار AllTree نختار "تدريب"، يظهر لدينا الشكل (3) التالي:



الشكل (3): طريقة توزيع البيانات بعد التدريب باستخدام شجرة القرار.

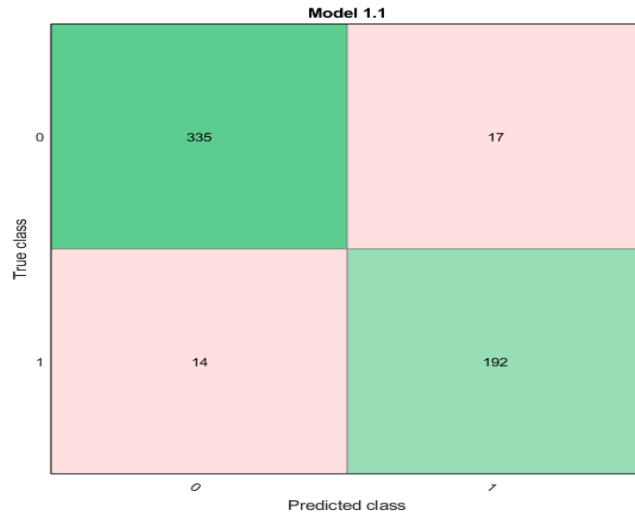
يوضح الشكل (3) طريقة توزيع البيانات بعد التدريب باستخدام شجرة القرار، حيث يشير الرمز X إلى النقاط التي تم تصنيفها بشكل خاطئ. وباقي النقاط تم تصنيفها بشكل صحيح .
يوضح الجدول (1) نتائج تدريب ثلاثة نماذج من أشكال شجرة القرار وصحة النتائج والزمن المستغرق لكل منها.

يبين الجدول (1) أن شجرة القرار أعطت نتائج مشجعة في الأنواع الثلاث السابقة الذكر حيث كانت النتائج أفضل ما يمكن في نموذج (Medium Tree) من حيث صحة النتائج والزمن المستغرق للتدريب.

المصنف (Classifier)	الدقة (accuracy)	الزمن (Time)
Fine Tree (maximum number of splits is 100).	94.4%	11.583sec
Medium Tree (maximum number of splits is 20)	94.4%	10.116sec
CoarseTree (maximum number of splits is 4)	93.7%	9.978Sec

الجدول (1): نتائج تدريب ثلاثة نماذج من أشكال شجرة القرار.

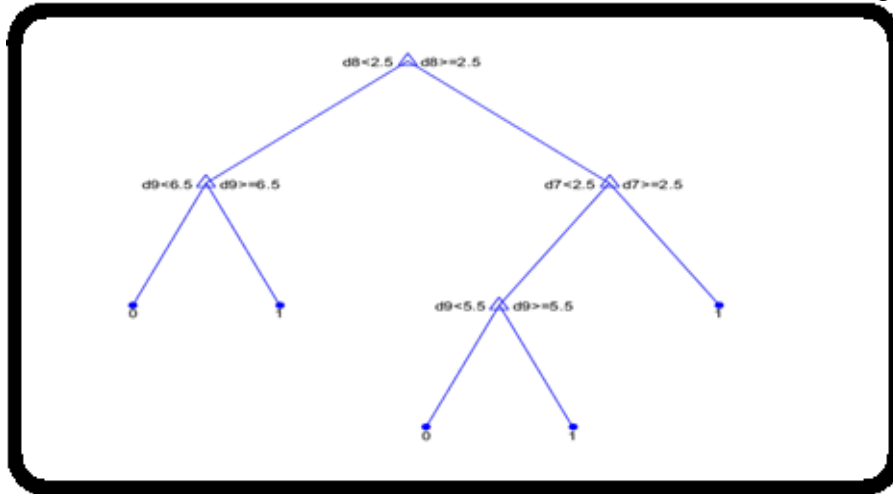
يبين الشكل (4) مصفوفة النسب المئوية للتصنيفات الصحيحة وغير الصحيحة (Confusion Matrix).



الشكل(4): مصفوفة النسب المئوية للتصنيفات باستخدام شجرة القرار.

لعرض نموذج شجرة القرار الذي تم تدريبه نقوم بتصدير (Export) أفضل نموذج تم تدريبه إلى مساحة العمل، ونكتب التعليمة التالية في Matlab فيظهر الشكل(5) التالي:

`view(trainedModel.ClassificationTree,'Mode','graph')`



الشكل(5): أفضل نموذج لشجرة القرار الذي تم تدريبه.

6- دراسة مقارنة: (Comparative Study)

أجريت باستخدام الحزمة البرمجية Matlab R2018b مجموعة من التجارب على قاعدة البيانات السابقة الخاصة بمرضى سرطان الثدي بغية تقييم النموذج المقترح وذلك بتغيير نوع المصنف حيث تم مقارنة استخدام ثلاثة أنواع من المصنفات وهي:

Back Propagation Neural Network(BPNN), K-Nearest Neighbors(KNN), Support Vector Machine(SVM) and Decision Trees

يبين الجدول (2) أن المصنفات الثلاث السابقة الذكر أعطت نتائج مشجعة بشكل عام، لكن الشبكة العصبونية من نوع 'Trainscg' (BPNN) أعطت نتائج أفضل من حيث الدقة والزمن المستغرق للتدريب.

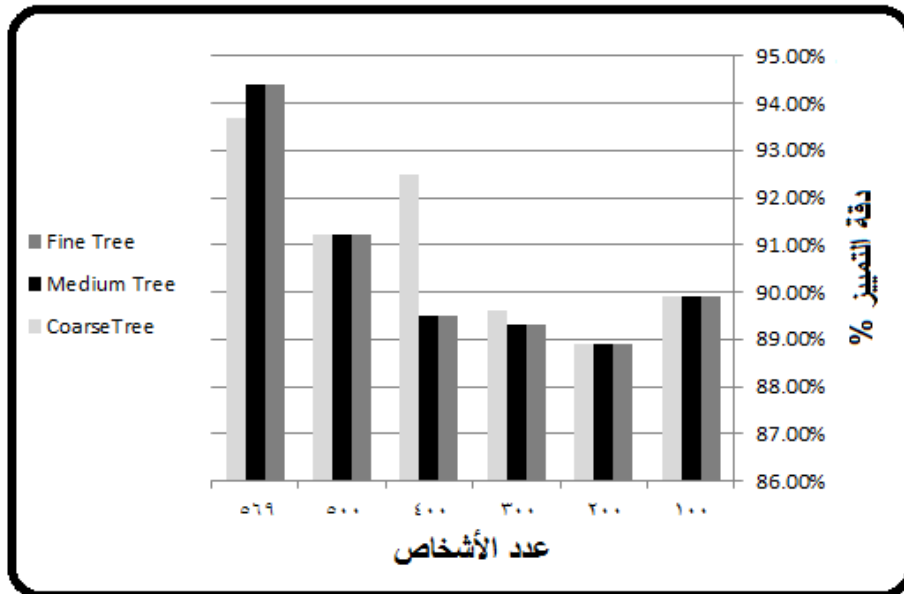
من الملاحظ أن استخدام شجرة القرار في النظام المقترح لا يعطي أفضل دقة أو زمن مقارنةً مع المصنفات الثلاثة الأخرى، إلا أنه تم اختياره في النموذج المقترح بناءً على سلوكه المبين في التجربة التالية.

الجدول (2): مقارنة الأنواع الثلاث من المصنفات.

الزمن (Time)	الدقة accuracy	المصنف (Classifier) 500 شخص
1.59Sec	%93.0	Fine Knn
0.51Sec	%94.2	Medium Knn
0.88Sec	%94.2	Coarse Knn
0.82Sec	%91.2	Fine Tree
0.61 sec	%91.2	Medium Tree
0.54Sec	%91.2	CoarseTree
0.59Sec	%91.0	Fine Guassian SVM
0.51Sec	%96.5	Medium Guassian SVM
0.88Sec	%93.5	Coarse Guassian SVM
1 Sec	% 94.0	'Trainlm' (BPNN)
1 Sec	%94.6	'Trainscg'(BPNN)
1 Sec	%94.0	'Trainbr'(BPNN)
1 Sec	%92.2	'Trainrp'(BPNN)

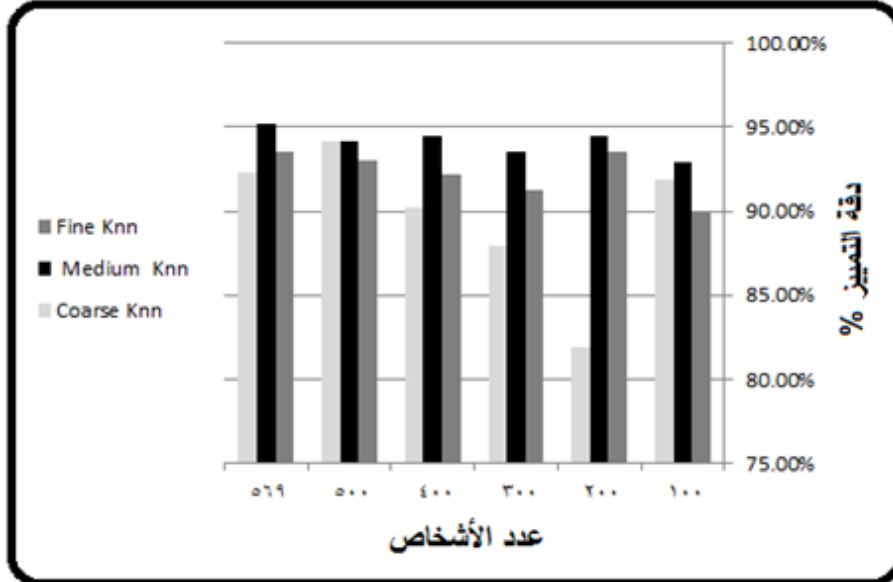
أجريت باستخدام الحزمة البرمجية Matlab R2018b مجموعة من التجارب على قاعدة البيانات السابقة الخاصة بمرضى سرطان الثدي بغية تقييم استقرار النموذج المقترح مقارنةً مع باقي المصنفات السابقة وذلك مع تغيير عدد الأشخاص.

تبين الأشكال التالية الشكل (6)(7)(8)(9)، تغيرات دقة التعرف (استقراره نموذج التعرف) مع تغير عدد الأشخاص.

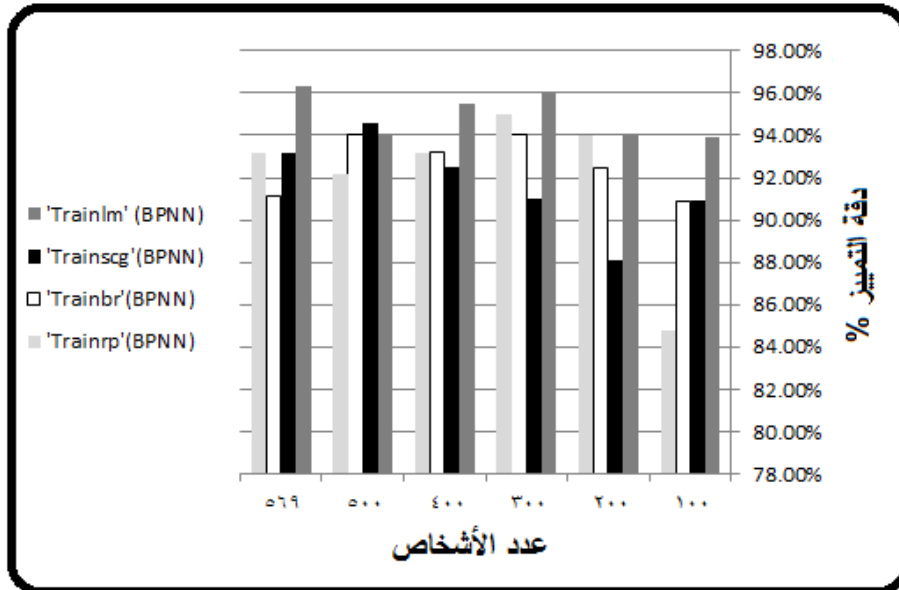


الشكل(6): تغيرات دقة تمييز نموذج التعرف على مرضى سرطان الثدي مع تغير عدد الأشخاص وباستخدام مصنف DT

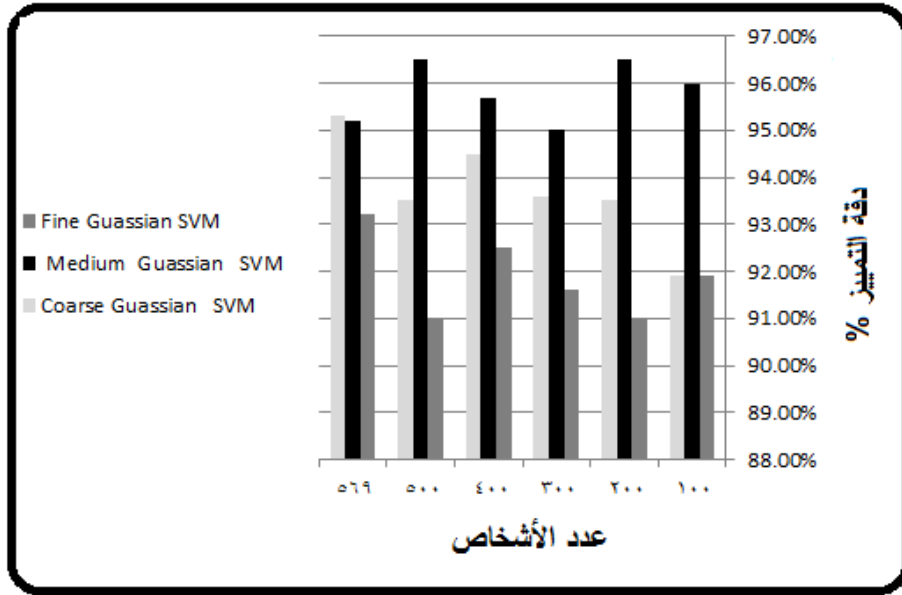
يبين الشكل (6) سلوكاً فريداً لمصنف شجرة القرار (وهو ما أدى إلى اختياره في هذا البحث)، حيث أن دقة التعرف تزداد مع زيادة عدد الأشخاص، على خلاف باقي المصنفات، والتي بينت التجارب العملية عليها استقراراً مقبولاً في دقة التعرف مع زيادة عدد الأشخاص، كما هو موضح في الأشكال (7)(8)(9).



الشكل(7): تغيرات دقة تمييز نموذج التعرف على مرضى سرطان الثدي مع تغير عدد الأشخاص وباستخدام مصنف KNN



الشكل(8): تغيرات دقة تمييز نموذج التعرف على مرضى سرطان الثدي مع تغير عدد الأشخاص وباستخدام مصنف BPNN



الشكل(9): تغيرات دقة تمييز نموذج التعرف على مرضى سرطان الثدي مع تغير عدد الأشخاص وباستخدام مصنف SVM

7- الخلاصة (Conclusion):

تم في هذا البحث دراسة مشكلة التعرف على مرضى سرطان الثدي باستخدام شجرة القرار، حيث تم تنفيذ الدراسة المقترحة مع ثلاثة أنواع من تصنيف شجرة القرار وهي (Fine Tree, Medium Tree, CoarseTree) على قاعدة بيانات خاصة بمرضى سرطان الثدي. أكدت النتائج التجريبية أن نتائج التعرف على مرضى سرطان الثدي كانت أفضل ما يمكن عند استخدام المصنف (Medium Tree)، كما أجريت مجموعة من التجارب المقارنة بين النموذج المقترح مع الشبكات العصبونية (BPNN) (Back propagation Neural Network)، ومع نموذج خوارزمية الدعم الآلي (Support Vector Machine SVM) ومع مصنف الجار الأقرب (K Nearest Neighbors(KNN)) باستخدام الحزمة البرمجية Matlab R2018b، حيث أكدت التجارب فعالية النظام عند استخدام الشبكات العصبونية 'BPNN Trainscg'. كما أجريت باستخدام الحزمة البرمجية Matlab R2018b مجموعة من التجارب على قاعدة البيانات السابقة الخاصة بمرضى سرطان الثدي بغية تقييم استقرار النموذج المقترح مقارنةً مع باقي المصنفات السابقة وذلك مع تغيير عدد الأشخاص، حيث تبين أن مصنف شجرة القرار المستخدم في النظام المقترح يسلك سلوكاً فريداً (وهو ما أدى إلى اختياره في هذا البحث)، حيث أن دقة التعرف تزداد مع زيادة عدد الأشخاص، على خلاف باقي المصنفات.

8- المراجع: (References)

- [1] (2019)World health organization, Available: <http://www.who.int/en/newsroom/factsheets/detail/cancer>.
- [2] David, A.; (2019). *Machine Learning Classification Techniques of Breast Cancer*.
- [3] مصعب عمر محمد عبد الخالق، (2020). *تشخيص مرض سرطان الثدي باستخدام خوارزمية*.
جامعة إفريقيا العالمية.
- [4] Kharya, S., (2014). *Naive Bayes Classifiers: A Probabilistic Detection Model for Breast Cancer*. International Journal of Computer Applications.
- [5] Astri, A., Mousannif H., (2016). *Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis*. Procedia Computer Science.
- [6] Verma, G., Verma, H., (2019). *Predicting Breast Cancer using Linear Kernel Support Vector Machine*. SSRN Electronic Journal
- [7] Kaur, P., Pruthi, Y., (2019). *Empirical Analysis of Cervical and Breast Cancer Prediction Systems using Classification*. Education and Management Engineering
- [8] Osama, R., (2012). *Breast Mass Detection in Mammograms using Modified K-means Clustering*. Faculty of Engineering, Cairo University.
- [9] جلال ابراهيم العبد، (2007). *استخدام الأساليب الكمية في اتخاذ القرارات الإدارية*. دار الجامعة الجديدة للنشر.
- [10] عطية العربي، بن زخرفة بوعلام، (2017). *دور شجرة القرار في تحسين القرارات التسويقية*.
جامعة قاصدي، مجلة العلوم الاقتصادية والتسيير والعلوم التجارية.
- [11] مؤيد الفضل، (2004). *الأساليب الكمية في الإدارة*. دار اليازوري العلمية.
- [12] Sanjeevi, M., (2017). *Decision Trees Algorithms*, Available: <https://www.linkedin.com/in/madhusanjeeviai>