

## كشف سرطان الثدي باستخدام القيم النسبية للمؤشرات الحيوية مع الشبكات العصبونية

هبة الله عيسى، أ. د. م. محمد فراس الحناوي

1طالبة ماجستير، قسم الهندسة الطبية، كلية الهندسة الميكانيكية والكهربائية، جامعة دمشق.

2أستاذ في قسم الهندسة الطبية، كلية الهندسة الميكانيكية والكهربائية، جامعة دمشق.

### الملخص

سرطان الثدي هو أكثر أنواع السرطانات شيوعاً التي تصيب النساء في جميع أنحاء العالم. على الرغم من الاستخدام المكثف للتصوير الشعاعي للثدي كمعيار ذهبي للكشف عن السرطان وتحديد موقعه، أصبحت المؤشرات الحيوية القائمة على تحاليل الدم أداة واعدة لتصنيف وتشخيص وتوقع وجود جميع أنواع السرطانات والأورام الخبيثة وخاصة سرطان الثدي. تم تطبيق العديد من خوارزميات تعلم الآلة وتحسينها لاستخدام عينات المؤشرات الحيوية في تشخيص سرطان الثدي واكتشافه. في هذا البحث، تم اقتراح طريقة جديدة للكشف عن سرطان الثدي تعتمد على استبدال القيم الخام للمؤشرات الحيوية والعينات المتوفرة بسمات جديدة نسبية مستخرجة منها. تعبر هذه السمات عن قيم نسبية تشير الى مدى قرب وبعد كل عينة لكل مؤشر الى متوسط القيم السليمة والسرطانية للمؤشر الحيوي نفسه. هذه السمات الجديدة ستستخدم كدخل للشبكة العصبونية ذات التغذية الأمامية. تم استخدام قاعدة بيانات Coimbra بسبب أهميتها واعتمادها في العديد من الأبحاث. نتائج الاختبار أظهرت دقة تصنيف نهائية 91.7% ونسبة حساسية 92.3%. وتظهر الطريقة المقترحة القدرة على اعطاء دقة تصنيف تتجاوز 85% في حال تدريب الشبكة العصبونية على 50% فقط من البيانات.

تاريخ الإيداع: 2022/6/20

تاريخ القبول: 2022/9/22



حقوق النشر: جامعة دمشق - سورية،

يحتفظ المؤلفون بحقوق النشر بموجب

الترخيص CC BY-NC-SA 04

**الكلمات المفتاحية:** سرطان الثدي، الشبكات العصبونية ذات التغذية الأمامية، المؤشرات

الحيوية.

# Detecting Breast Cancer Based on Normalized Features of Biomarkers and Neural Networks

Hiba Allah Essa<sup>1</sup>, Prof. Mhd Firas Alhinawwi<sup>2</sup>

<sup>1</sup>MSc student, Department of Biomedical Engineering, Faculty of Mechanical and Electrical Engineering, Damascus University, Damascus.

<sup>2</sup>Professor, Department of Biomedical Engineering, Faculty of Mechanical and Electrical Engineering, Damascus University.

## Abstract

Received: 20/6/2022

Accepted: 22/9/2022



**Copyright:** Damascus University- Syria, The authors retain the copyright under a CC BY- NC-SA

Breast cancer is the most common type of cancer affecting women worldwide. Despite the extensive use of mammography as the gold standard for cancer detection and locating, biomarkers based on blood analyzes have become a promising tool to classify, diagnose, and predict the presence of all types of cancers and malignancies, especially breast cancer. Lots of machine learning algorithms have been implemented and improved to use bio-samples in breast cancer diagnosis and detection. In this paper, a new method for breast cancer detection is proposed based on replacing the raw values of biomarkers with new normalized features based on the average healthy and cancer values. These features express normalized values that indicate ratio between sample/biomarker and the average healthy and cancerous values of the same biomarker respectively. These new features will be used as input to the feedforward neural network. Coimbra Database has been used because of its importance and reliability in many research papers. The test results show a final classification accuracy of 91.7% and a sensitivity of 92.3%. The proposed method shows the ability to give a classification accuracy of more than 85% if the neural network is trained on 50% of the data only.

**Keywords:** Breast cancer; feedforward neural networks; biomarkers.

## المقدمة

يعد سرطان الثدي (BC) Breast cancer أكثر أنواع السرطانات التي تصيب النساء في جميع أنحاء العالم، وفي عام 2018، أصبح سرطان الثدي (BC) أكثر أنواع السرطانات شهرة لدى النساء مع 15% من الوفيات السنوية المرتبطة بالسرطان (Azamjah et al., 2019, 2015-2018). على الرغم من الاستخدام الشائع للتصوير الشعاعي للثدي كمعيار ذهبي للكشف عن السرطان وتحديد موقعه، فإن استخدام التصوير الشعاعي للثدي، وخاصة للأعمار الصغيرة دون سن الثلاثين، يجعله خطيراً على النساء (Lebovic et al., 2021, 260-261). على الرغم من أن العديد من الأبحاث السابقة استخدمت التصوير الشعاعي ومخرجاته من أجل تشخيص سرطان الثدي، وأظهرت نتائج ممتازة، ولكن بسبب تطور تقنيات التحليل الطيفي وتحليل الدم وأنظمة الذكاء الاصطناعي، انتقل الباحثون للعمل على مفهوم المؤشرات الحيوية والواسمات السرطانية وبعض المركبات الحيوية في جسم الإنسان التي يمكن أن تساعد في تصنيف الحالة والتنبؤ بوجود جميع أنواع السرطانات والأورام الخبيثة بما في ذلك سرطان الثدي (Oloomi et al., 2020, 6).

حددت العديد من الدراسات بعض الجزيئات التي تشير إلى تراكيز غير طبيعية. تم تصنيف هذه التراكيز غير المنتظمة على أنها مؤشرات حيوية، واستخدمت كمسات features تميز بين حالة سرطان الثدي والحالة السليمة. المؤشرات الحيوية المعتمدة على تحاليل الدم هي سمات غير جراحية قابلة للقياس يمكن استخدامها في اكتشاف السرطان والتشخيص المسبق له وتشخيص مرحلة السرطان، مثل أوستيوبونتين (OPN) osteopontin، مستضد السرطان cancer antigen (CA15-3 و CA125). توفر الكمية الغنية من العناصر الجزيئية والخلوية المختلفة في الدم والحالة الصحية الاختلاف الدقيق في التركيز بينها وبالتالي المزيد من السمات والبيانات

لبناء أنظمة آلية فيما يخص التشخيصات غير الجراحية لسرطان الثدي والتنبؤ به (Li et al., 2020, 2767). تشير المؤشرات الحيوية بشكل عام إلى مجموعات من البيانات الكبيرة ولكن مع تباين بسيط وانحراف معياري ضئيل بين الحالة السرطانية والسليمة. تجعل هذه المشكلة أدوات تعلم الآلة Machine learning (ML) حاجة مستمرة للوصول إلى المستوى المطلوب من تشخيص السرطان والتنبؤ به باستخدام المؤشرات الحيوية غير الجراحية المتاحة. مؤخراً، أصبح لخوارزميات تعلم الآلة دوراً رئيسياً في مجموعة واسعة من التطبيقات الطبية، مثل معالجة اللغات الطبيعية، وتصنيف الصور الطبية. ويعتبر اكتشاف السرطانات المستند إلى المؤشرات الحيوية أحد أهم المجالات التي يمكن أن يستخدم معها ML (Rozen et al., 2021, 1774-1783).

تعتبر مجموعة بيانات Coimbra الخاصة بسرطان الثدي من موقع UCI لتعلم الآلة واحدة من أهم مجموعات البيانات الأكثر استخداماً للتحقيق في تشخيص واكتشاف BC باستخدام تعلم الآلة والذكاء الاصطناعي وتحتوي على العديد من المؤشرات الحيوية القائمة على تحليل عينات الدم مثل الجلوكوز Glucose والريزيسيتين resistin واللبتين leptin وستة مؤشرات أخرى (Patrício et al., 2018, 1-8).

النماذج التنبؤية الأكثر شهرة لكشف وتشخيص سرطان الثدي بالاعتماد على السمات المتوفرة هي متجه الآلة الداعم (SVM) support vector machines والغابة العشوائية (RF) random forests، والانحدار اللوجستي (LR) logistic regression، والجيران الأقرب (KNN) K-nearest neighbors. يستخدم الباحثون عادة التحليل متعدد المتغيرات multivariate analysis لتحديد أهمية كلاً من المؤشرات على حدة في تحسين كشف السرطان. ثم يتم تقييم مخرجات التصنيف باستخدام الحساسية sensitivity والنوعية specificity والدقة accuracy

للشبكة العصبونية ذات التغذية الأمامية وتدريب الشبكة على البيانات ودراسة كفاءة اختبار وتشخيص عينات غير مدربة مسبقاً.

### الطرائق والمواد

#### 1. بيانات المؤشرات الحيوية

في هذه الورقة البحثية، تم دراسة المؤشرات الحيوية الخاصة بقاعدة البيانات Coimbra Dataset من موقع UCI (جامعة كاليفورنيا ، إيرفين) لتحسين اكتشاف سرطان الثدي باستخدام تقنيات التعلم العميق. تحتوي مجموعة البيانات هذه على 116 عينة (52 عينة لأشخاص أصحاء و 64 عينة لمرضى سرطان الثدي) (Patricio et al., 2018, 1-8). تتكون البيانات من تسع سمات (مؤشرات حيوية) وفئتين مصنفتين على أنهما سليم وسرطان الثدي. السمات التسع لمجموعة البيانات هي: العمر (سنة)، الجلوكوز Glucose (mg/dL)، الأنسولين (µU/mL) Insulin، نموذج تقييم الاتزان homeostatic model assessment (HOMA)، اللبتين Leptin (ng/mL)، سيتوكين الدهني Adiponectin (µg/mL)، ريزيستين (ng/mL) Resistin، و MCP-1 (pg/dL) كما هو موضح في الجدول (1).

من خلال حساب متوسط والانحراف المعياري لقيم كل سمة للعينات السليمة والسرطانية الجدول (1) يتبين لدينا وجود تباين مقبول بين العينات السليمة والسرطانية وخاصة المتوسط الحسابي، بالتالي يمكن الاعتماد عليه في زيادة كفاءة السمات في يخص كشف سرطان الثدي باستخدام الشبكات العصبونية ذات التغذية الأمامية.

(Dlamini et al., 2020, 2300-2302). فيما يتعلق بمجموعة بيانات Coimbra، تُظهر نماذج SVM التي تستخدم Resistin ومؤشر كتلة الجسم، والجلوكوز، والعمر أداءً أفضل حيث تصل النوعية إلى (85% إلى 90%)، والحساسية (82% إلى 88%) والمنطقة الواقعة تحت المنحنى (AUC) كانت (0.91, 0.87) (Patricio et al., 2018, 1-8). اختبر Xixuan Chen و Yixuan Li عام 2018 خمس خوارزميات مختلفة لتعلم الآلة مثل SVM، RF والشبكات العصبونية و LR، ووجدوا أن RF القائمة على بارامترات التقييم هي أفضل نموذج لاستخدامه مع مجموعة بيانات Coimbra (Li et al., 2018, 212-214). استخدم كلاً من Silva وزملاءه عام 2019 طريقة الشبكات العصبونية الضبابية fuzzy neural networks لكشف سرطان الثدي بالاعتماد على Coimbra وتوصلوا إلى دقة نهائية 62% وحساسية كشف سرطان 78.3% (Silva et al., 2019, 479-481). لاحقاً تم استخدام شجرة القرار الضبابية fuzzy decision tree (FDT) لتصنيف بيانات Coimbra وتوصل الباحثون إلى دقة نهائية 70.68% وحساسية 78.3% (Idris et al., 2021, 11-13).

تهدف هذه الورقة البحثية إلى اقتراح طريقة جديدة للكشف عن سرطان الثدي القائم على العلامات الحيوية بالاعتماد على منهجية جديدة تعتمد على استبدال القيم الخام للسمات والعينات المتوفرة بسمات منسوبة normalized مستخرجة منها حيث تعبر هذه السمات عن قيم نسبية تشير إلى مدى قرب وبعد كل عينة لكل سمة إلى متوسط القيم السليمة والسرطانية للسمة نفسها، ليتم استخدام هذه السمات الجديدة كمدخلات

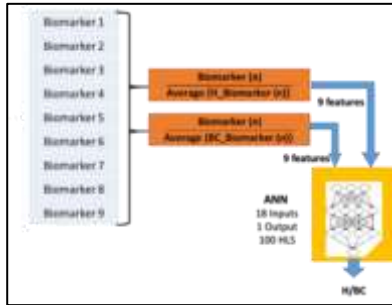
الجدول (1) المؤشرات الحيوية المتوفرة في قاعدة بيانات Coimbra والعينات المتوفرة

على شكل متوسط حسابي وانحراف معياري.

المؤشر الحيوي	العينات السليمة (متوسط ± انحراف معياري)	العينات السرطانية (متوسط ± انحراف معياري)
Age (years)	58.07 ± 18.77	56.67 ± 13.38
BMI (kg/m2)	28.31 ± 5.37	26.98 ± 4.58
Glucose (mg/dL)	88.23 ± 10.09	105.56 ± 26.34
Insulin (µU/mL)	6.93 ± 4.81	12.51 ± 12.22

المؤشر الحيوي	العينات السليمة (متوسط $\pm$ انحراف معياري)	العينات السرطانية (متوسط $\pm$ انحراف معياري)
HOMA	1.55 $\pm$ 1.20	3.62 $\pm$ 4.55
Leptin (ng/mL)	26.63 $\pm$ 19.14	26.59 $\pm$ 19.06
Adiponectin ( $\mu$ g/mL)	10.32 $\pm$ 7.55	10.06 $\pm$ 6.14
Resistin (ng/mL)	11.61 $\pm$ 11.33	17.25 $\pm$ 12.53
(pg/dL) MCP-1	499.73 $\pm$ 289.41	563.01 $\pm$ 380.98

والحالات السرطانية " Average {BC\_Biomarker (n)}  
(n) "، سيتم إعادة صياغة مجموعة البيانات الى 18  
سمة، حيث كل عينة من سمة محددة سيتم تنسيبها  
الى متوسط الحالات السليمة ومتوسط الحالات  
السرطانية، بالتالي يصبح لدينا مؤشر نسبي حول مدى  
قرب كل عينة من السليم ومن السرطان كما هو  
موضح في الشكل (1).



الشكل (1) مخطط المنهجية المتبعة في استخراج

الخصائص النسبية من القيم العددية للمؤشرات الحيوية  
وإستخدامها كمدخلات للشبكة العصبونية للحصول على تصنيف  
الحالة.

### 3. الشبكة العصبونية

بعد الانتهاء من تحديد بنية البيانات المتوفرة  
الخاصة بالمؤشرات الحيوية للحالات السليمة  
والسرطانية، واستخراج السمات الجديدة (18 سمة) التي  
تعتمد على تنسيب القيم العددية الى المتوسط الحسابي  
لكل مؤشر حيوي لكل من السليم والسرطان على حده.  
في هذا البحث، سيتم العمل على تصنيف السمات  
الجديدة الى سليم أو سرطان باستخدام الشبكة  
العصبونية ذات تغذية أمامية (FNN) feedforward  
neural network. شبكة FNN هي نوع من الشبكات

## 2. معالجة البيانات واستخراج الخصائص

### النسبية

في معظم الأدبيات المتعلقة باستخدام المؤشرات  
الحيوية المعتمدة على تحاليل الدم في كشف سرطان  
الثدي وبالاعتماد على الذكاء الاصطناعي، يتم  
استخدام القيم الخام للسمات كما هي، ولكن بسبب  
تقارب القيم العددية للسمات بين الحالات السليمة  
والسرطانية (Rasool et al., 2022, 3211-3213)،  
فإن العديد من خوارزميات الذكاء الاصطناعي وتعلم  
الألة لا تستطيع الحصول على دقة تصنيف عالية،  
وذلك بسبب نسبة الارتياب والتقارب العالية بين السليم  
والمصاب، وخاصة في حالة قلة السمات المستخدمة  
بالإضافة الى قلة العينات.

من خلال دراسة مطوّلة وموسعة حول سلوك تغير  
القيم العددية للسمات بين السليم والحالة السرطانية،  
تبين لدينا أنه يمكن استخدام النسبة بين قيمة كل عينة  
الى متوسط كلاً من القيم السليمة والسرطانية للسمة  
نفسها، كسمات جديدة ونسبية معتمدة على بعد القيم  
العددية الخام عن كل من متوسط السليم والسرطان  
لكل سمة والتي توفر لدينا دخل ذو تباين عالي للذكاء  
الاصطناعي وخاصة الشبكات العصبونية.

في الطريقة المقترحة، لدينا قاعدة بيانات  
Coimbra التي تحتوي 9 مؤشرات حيوية، بعد حساب  
المتوسط الحسابي لقيم كل سمة (n) Biomarker  
(حيث n هو رقم المؤشر الحيوي في قاعدة البيانات)  
للحالات السليمة " Average {H\_Biomarker (n)} "

True Positive (TP): عدد البيانات المصنفة بشكل صحيح وهي تدل على وجود السرطان (BC).  
False Positive (FP): عدد البيانات المصنفة بشكل خاطئ على أنها سرطان (BC) وهي سليمة (H).  
True Negative (TN): عدد البيانات المصنفة على أنها سليمة (H) وهي تدل على عدم وجود السرطان.  
False Negative (FN): عدد البيانات المصنفة على أنها سليمة (H) وهي سرطانية (BC).

### النتائج والمناقشة

في هذا القسم، سنقوم بعرض نتائج تنفيذ المنهجية المقترحة لكشف سرطان الثدي باستخدام قاعدة المؤشرات الحيوية Coimbra، حيث تتلخص المنهجية بالخطوات التالية: (1) قراءة البيانات، (2) استخراج المتوسط الحسابي لجميع العينات لكل سمة على حده وبشكل منفصل بين السليم والمصاب بالسرطان، (3) تتسبب قيم العينات المتوفرة لكل سمة الى متوسط القيم السليمة والسرطانية للسمة نفسها وبالتالي العينة السمة سينتج عنها سمتين جديدتين، لذلك سيتحول حجم الدخل المستخدم مع شبكة FNN من 9 سمات الى 18 سمة، (4) خطوة التصنيف تتلخص باستخدام الشبكات العصبونية ذات التغذية الأمامية (FNNs) مع السمات المتوفرة وتدريبها وفق عينات Coimbra الى السليم والمصاب، بالإضافة الى تقييم أداء الشبكة على عدد من العينات غير المدربة مسبقاً.

بعد أن أصبح لدينا 18 سمة نسبية، سيتم تدريب الشبكة العصبونية على 92 عينة (80% من البيانات) تتوزع بشكل متكافئ بين الحالة السليمة والمصابة، وسيتم تخصيص 24 عينة (20% من البيانات) كبيانات اختبار. بعد تنفيذ عملية تدريب شبكة (FNN) وفق البارامترات المذكورة سابقاً في الفقرة (3) من قسم

العصبونية الاصطناعية. غالباً ما يشار إليها على أنها شبكة متعددة الطبقات من الخلايا العصبية. تتكون شبكة FNN من طبقة إدخال، وعدد من الطبقات المخفية، وطبقة إخراج (-1, 2015, Laudani *et al.*). تم بناء الشبكة العصبونية FNN باستخدام MATLAB وفق المواصفات التالية: طبقة الدخل تحتوي 18 عقدة، طبقة الخرج تحتوي عقدة واحدة (سليم/مصاب)، حجم الطبقة الخفية hidden layer (HLS) مساوي الى 6 عصبونات، حيث تم حساب عدد العصبونات N للطبقة المخفية من المعادلة (1) حسب (Huang *et al.*, 865–873):

$$N = \frac{0.51}{\sqrt{0.43 \times m + 0.1n^2 + 2.54m + 0.77n + 0.35}} + 0.51 \quad (1)$$

حيث m هو عدد المدخل، و n عدد المخرج. عملية التدريب اعتمدت على قيمة الأداء Performance النهائي المرغوب ( $10^{-8}$ )، عدد التكرارات 2000 iterations، وتابع التدريب باستخدام الانتشار الخلفي من نوع التدرج المرافق المنسوب Scaled conjugate gradient الذي يقوم بتحديث قيم الأوزان والانحياز وفقاً لطريقة التدرج المرافق المنسوب. تم استخدام نسب متغيرة بين عينات التدريب الى الاختبار حيث تم توضيحها في قسم النتائج.

### 4. مقاييس الأداء

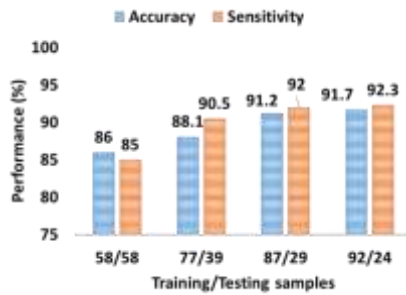
تم تحديد دقة وكفاءة الخرج من خلال دقة التشخيص (AC) accuracy، الحساسية sensitivity (SE)، وتم حسابهم باستخدام القوانين:

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

حيث:

المدرية (الاختبار) الى العينات المدربة. يظهر الشكل (4) قدرة الشبكة العصبونية على اعطاء دقة وحساسية تصنيف عالية تتجاوز نسبة 85% في حال استخدام 50% فقط من العينات للتدريب وأظهرت نسبة العينات 77/39 دقة وحساسية تجاوزت 88% كما هو موضح في الشكل (4).



الشكل (4) دقة وحساسية التصنيف أثناء اختبار شبكة FNN باستخدام نسب مختلفة من عينات التدريب والاختبار. من خلال عرض نتائج الأبحاث السابقة ذات الصلة حول تعزيز كشف سرطان الثدي باستخدام المؤشرات الحيوية وأدوات تعلم الآلة الجدول (2)، يمكن بوضوح ملاحظة الى أي مدى تزيد الطريقة المقترحة من الدقة والحساسية حيث حصل (Idris et al., 2021, 11-13) على دقة نهائية بلغت 70.69%، في المقابل باستخدام المنهجية المقترحة في بحثنا وصلت الدقة الى 91.7% والحساسية الى 92.3% باستخدام 80% من البيانات في عملية التدريب و 20% في عملية الاختبار. ومن أجل التأكيد على كفاءة استخدام السمات المنسوبة، يظهر الجدول (2) فرق الدقة والحساسية في حال تصنيف بيانات Coimbra باستخدام شبكة FNN الحالية مع البيانات الخام التي تحتوي 9 سمات فقط Raw features، حيث أظهرت دقة تساوي الى 79.3% وحساسية تساوي الى 75%.

الطرائق والمواد، تم عرض نتائج التدريب وفق مصفوفة الارتباب Confusion matrix حيث توضح التصنيف الصحيح (Actual class) والتصنيف الناتج عن FNN (Predicted class) كما هو موضح في الشكل (2) حيث يعبر رمز H أو (0) عن السليم و BC أو (1) تعبر عن الحالة السرطانية. أظهرت نتائج التدريب دقة تصنيف نهائية 93.5% ونسبة حساسية 96.1%.

Predicted	H (0)	37 40.2%	2 2.2%	94.9% 5.1%
	BC (1)	4 4.3%	49 53.3%	92.5% 7.5%
		90.2% 9.8%	96.1% 3.9%	93.5% 6.5%
		H (0)	BC (1)	
		Actual		

الشكل (2) مصفوفة الارتباب للعينات المصنفة خلال عملية تدريب شبكة FNN.

من خلال تصنيف عينات الاختبار، تبين لدينا أن دقة تصنيف الشبكة العصبونية لهذه العينات تساوي الى 91.7% وحساسية لكشف السرطان تساوي الى 92.3% كما هو موضح في الشكل (3).

Predicted	H (0)	10 41.7%	1 4.2%	90.9% 9.1%
	BC (1)	1 4.2%	12 50.0%	92.3% 7.7%
		90.9% 9.1%	92.3% 7.7%	91.7% 8.3%
		H (0)	BC (1)	
		Actual		

الشكل (3) مصفوفة الارتباب الناتجة عن اختبار شبكة FNN لعينات الاختبار غير المدربة مسبقاً.

من أجل تقييم كفاءة المنهجية في حال لدينا عدد قليل من العينات والتي تعتبر تحدي كبير في مجال تصنيف المؤشرات الحيوية، تم تكرار عملية التدريب والاختبار للشبكة العصبونية بزيادة عدد العينات غير

الجدول (2) مقارنة مقاييس الأداء بين الطريقة المقترحة وبعض الأدبيات المتعلقة بمجموعة بيانات Coimbra.

المرجع	الطريقة	أفضل دقة	أفضل حساسية
(Poorani et al., 2019, 2106-2109)	SVM	85.72%	46.54 %
(Silva et al., 2019, 479-481)	fuzzy neural networks	62%	78%
(Hasdyna et al., 2020, 5)	KNN	85.72%	46.54 %
(Idris et al., 2021, 11-13)	fuzzy decision tree (FDT)	70.69 %	78.3%
البحث الحالي	Raw features and FNN	79.3%	75%
البحث الحالي	Normalized features and FNN	91.7%	92.3%

### الخاتمة

أظهرت دقة تصنيف نهائية 91.7% وقيمة حساسية تساوي الى 92.3%. في حين أن السمات الطبيعية الموجودة في قاعدة Coimbra بعد استخدامها المباشر مع شبكة FNN لم تتجاوز دقتها وحساسيتها نسبة 80%. وبالمقارنة مع أحدث الأبحاث المنجزة عالمياً فيما يخص تحسين كشف سرطان الثدي باعتماد قاعدة بيانات Coimbra وأدوات الذكاء الاصطناعي المحسنة، أظهرت الطريقة المقترحة فرق محسن وكبير في دقة الكشف، وذلك يعود الى اعتماد سمات مفيدة وذات قدرة تمييزية أعلى تعتمد على متوسط قيم العينات السليمة والسرطانية في آن معاً.

**التمويل:** هذا البحث ممول من جامعة دمشق وفق رقم التمويل (501100020595).

أصبح دعم اتخاذ القرار الطبي باستخدام أدوات تعلم الآلة لتشخيص السرطان المستند إلى المؤشرات الحيوية مجالاً مهماً وواعداً في الدراسات السريرية. في هذا البحث، تم اقتراح طريقة جديدة للكشف عن سرطان الثدي تعتمد على استبدال القيم الخام للمؤشرات الحيوية والعينات المتوفرة بسمات جديدة مستخرجة منها حيث تعبر هذه السمات عن قيم نسبية تشير الى مدى قرب وبعد كل عينة لكل مؤشر الى متوسط القيم السليمة والسرطانية للمؤشر الحيوي نفسه، ليتم استخدام هذه السمات الجديدة كمدخلات للشبكة العصبونية ذات التغذية الأمامية وتدريبها ودراسة كفاءة اختبار وتشخيص عينات غير مدربة مسبقاً، حيث تم استخدام قاعدة بيانات Coimbra. نتائج الاختبار

### المراجع

1. Azamjah N., Soltan-Zadeh Y., Zayeri, F. (2019). **Global trend of breast cancer mortality rate: a 25-year study**. Asian Pacific journal of cancer prevention: APJCP, 20(7), pp.2015-2018.
2. Lebovic, G.S., Hollingsworth, A. and Feig, S.A., 2010. **Risk assessment, screening and prevention of breast cancer: A look at cost-effectiveness**. The breast, 19(4), pp.260-261.
3. Oloomi M., Moazzezy N., Bouzari S. (2020). **Comparing blood versus tissue-based biomarkers expression in breast cancer patients**. Heliyon, 6(4), pp.6.
4. Li J., Guan X., Fan Z., Ching L.M., Li Y., Wang X., Cao W.M., Liu D.X. (2020). **Non-invasive biomarkers for early detection of breast cancer**. Cancers, 12(10), pp.2767.
5. Rozen R., Weihs D. (2021). **Machine-Learning Provides Patient-Specific Prediction of Metastatic Risk Based on Innovative, Mechanobiology Assay**. Annals of Biomedical Engineering, 49(7), pp.1774-1783.
6. Patrício M., Pereira J., Crisóstomo J., Matafome P., Gomes M., Seica R., Caramelo F. (2018). **Using Resistin, glucose, age and BMI to predict the presence of breast cancer**. BMC cancer, 18(1), pp.1-8.
7. Dlamini Z., Francies F.Z., Hull R. Marima R. (2020). **Artificial intelligence (AI) and big data in cancer and precision oncology**. Computational and Structural Biotechnology Journal, 18, pp.2300-2302.



8. Li Y., Chen Z. (2018). **Performance evaluation of machine learning methods for breast cancer prediction.** Appl Comput Math, 7(4), pp.212-214.
9. Silva Araújo V.J., Guimarães A.J., de Campos Souza P.V., Rezende T.S., Araújo V.S. (2019). **Using resistin, glucose, age and BMI and pruning fuzzy neural network for the construction of expert systems in the prediction of breast cancer.** Machine Learning and Knowledge Extraction, 1(1), pp.479-481.
10. Idris N.F., Ismail M.A. (2021). **Breast cancer disease classification using fuzzy-ID3 algorithm with FUZZYDBD method: automatic fuzzy database definition.** PeerJ Computer Science, 7, pp. 11-13.
11. Rasool A., Bunterngchit C., Tiejian L., Islam M.R., Qu Q., Jiang Q. (2022). **Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis.** International journal of environmental research and public health, 19(6), pp.3211.
12. Laudani A., Lozito G.M., Riganti Fulginei F., Salvini A. (2015). **On training efficiency and computational costs of a feed forward neural network: a review.** Computational intelligence and neuroscience, 2015(13), pp.1-4.
13. Huang M.L., Hung Y.H., Chen W.Y. (2010). **Neural network classifier with entropy based feature selection on breast cancer diagnosis.** J Med Syst, 34:pp. 865–73.
14. Poorani S., Balasubramanie P. (2019). **Deep neural network classifier in breast cancer prediction.** International Journal of Engineering and Advanced Technology, 9(1), pp.2106-2109.
15. Hasdyna N., Sianipar B., Zamzami E.M. (2020). **Improving the performance of K-nearest neighbor algorithm by reducing the attributes of dataset using gain ratio.** Journal of Physics: Conference Series, 1566(1), pp. 5.