

## تطوير نظام تشخيص لسرطان الثدي باستخدام الشبكات العصبونية الاصطناعية

د. م. فادي متوج<sup>1</sup> ود. م. فاتن عجيب<sup>2</sup>

### الملخص

يعدُّ سرطان الثدي أكثر أنواع السرطانات شيوعاً بين النساء، وهو ثاني سبب رئيس لوفيات السرطان في شتى أنحاء العالم. إنَّ الكشف المبكر للسرطان متبوعاً بعلاج مناسب يمكن أن يقلل من خطر الموت. يمكن أن يخطئ الأطباء المختصون في تشخيص المرض، ومن ثمَّ فإنَّ استخدام أساليب الذكاء الاصطناعي مثل الشبكات العصبونية يمكن أن يحسن بشكل كبير من دقة التشخيص. في هذا البحث طُوِّرَ نظام باستخدام الشبكات العصبونية الاصطناعية (ANN) يمكن أن يساعد الأطباء في تشخيص سرطان الثدي بسرعة و بدقة. أظهرت النتائج دقة نظام التشخيص المقترح وموثوقيته، إذ بلغت الحساسية و النوعية و الدقة 96.2%، 100%، 99% على الترتيب.

**الكلمات المفتاحية:** التشخيص الطبي - الذكاء الاصطناعي - الشبكات العصبونية الاصطناعية - سرطان الثدي.

<sup>1</sup>دكتور مدرس في قسم هندسة الميكاترونك في كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، سورية.  
<sup>2</sup>دكتورة مدرسة في قسم الهندسة الطبية في كلية الهندسة الميكانيكية والكهربائية، جامعة دمشق، سورية.

## Developing a Breast Cancer Diagnosis System using Artificial Neural Networks

Dr. Fadi Motawej<sup>1</sup> and Dr. Faten Ajeeb<sup>2</sup>

### Abstract

Breast cancer is the most common type of cancer in women and is the second leading cause of cancer deaths worldwide. Early detection followed by appropriate cancer treatment can reduce the deadly risk. Medical professionals can make mistakes while identifying the disease. The help of artificial intelligence techniques such as neural networks can substantially improve the diagnosis accuracy. In this study, we developed a system for diagnosis of breast cancer using Artificial Neural Networks (ANNs) which can assist the doctors in diagnosis of the disease quickly and accurately. The results showed the accuracy and the reliability of the proposed diagnosis system with sensitivity, specificity and accuracy reached 96.2%, 100% and 99% respectively.

**Keywords:** Medical diagnosis, Artificial intelligence, Artificial neural networks, Breast cancer.

---

<sup>1</sup>Doctor in the Department of Mechatronic Engineering at the Faculty of Mechanical and Electrical Engineering, Tishreen University, Syria.

<sup>2</sup>Engineer in the Department of biomedical Engineering at the Faculty of Mechanical and Electrical Engineering, Damascus University, Syria.

## 1- مقدمة:

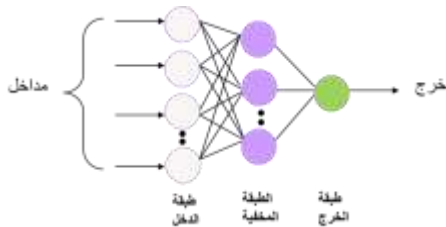
مرضى السكري ([1]، [2])، وتشخيص السل ([3]، [4])، وتصنيف اللوكيميا [5]، وتحليل صور الأشعة أو حتى الأنسجة الحية ([6]، [7]).

## 2- الشبكات العصبونية الاصطناعية: الشبكات

العصبونية الاصطناعية هي تمثيل هندسي يهدف إلى محاكاة عمل الشبكات العصبونية في الدماغ البشري. تتألف الشبكات العصبونية الصناعية من عناصر معالجة بسيطة متصلة فيما بينها تدعى عصبونات (neurons). ينجز كل عصبون مهمة حسابية بسيطة لكن السلوك الكلي للشبكة يتحدد من خلال الوصلات بين مختلف هذه العناصر [8]. أهم ما يميز الشبكات العصبونية الاصطناعية هي قدرتها على التعلم والتعميم. لهذا السبب، تُطبَّق على نطاق واسع لأنه من الممكن أن تقوم بنمذجة النظم غير الخطية إذ تكون العلاقة بين المتغيرات غير معروفة أو معقدة جداً

استخدمنا في بحثنا شبكة عصبونية من نوع بيرسيبترون متعدد الطبقات (MLP) إذ تعدّ الأكثر شيوعاً في مجال التشخيص الطبي. تتكون الشبكة العصبونية من نوع بيرسيبترون متعدد الطبقات (MLP) عادةً من ثلاث طبقات موضحة بالشكل (1) وهي:

1. طبقة الدخل Input layer
2. الطبقة المخفية Hidden layer
3. طبقة الخرج Output layer



الشكل (1) بنية الشبكة العصبونية الاصطناعية MLP [8]

طبقة الدخل لا تقوم بأي عملية معالجة، فهي ببساطة مكان تغذية الشبكة بشعاع البيانات، تقوم طبقة الدخل بعد

سرطان الثدي هو من أكثر أنواع السرطانات انتشاراً بين السيدات. وهو السبب الأكثر شيوعاً للوفاة في الفئة العمرية 35-55 سنة عند النساء. إنَّ الاكتشاف المبكر لسرطان الثدي ومن ثمَّ العلاج المبكر هو أفضل الطرائق المتاحة لتقليل عدد الوفيات، وزيادة عدد الناجيات من المصابات بالمرض. كلما اكتُشِفَ المرض مبكراً كلما زادت فرص النجاة وتجنب العلاج الشاق.

سرطان الثدي يعني أنَّ عدداً من خلايا الثدي بدأت تتكاثر تكاثراً غير طبيعي. هذه الخلايا تنقسم بسرعة أكبر من الخلايا السليمة

ويمكن أن تبدأ في الانتشار في أنحاء نسيج الثدي كله، إلى داخل الغدد الليمفاوية، بل إلى أعضاء أخرى في الجسم.

إن اكتساب الأطباء لخبرة التشخيص الضرورية هي عملية بطيئة نسبياً ومكلفة. فضلاً عن ذلك يمكن أن يخطئ الأطباء أحياناً في تشخيص المرض (إعطاء نتيجة سلبية مع وجود المرض، أو إعطاء نتيجة إيجابية في حال غياب المرض) بسبب نقص الخبرة. لذلك كان هناك حاجة للبحث عن طريقة سريعة وموثوق بها تقوم بمساعدة الأطباء والإطار الطبي في إعطاء تشخيص صحيح للمرض. و في هذا الإطار يمكن لتقنيات الذكاء الاصطناعي مثل الشبكات العصبونية الاصطناعية أن تحسن بشكل كبير من دقة التشخيص. هدَفَ بحثنا إلى تطوير نظام تشخيص باستخدام شبكة عصبونية اصطناعية (ANN) لتحديد هل كان المرضى لديهم سرطان الثدي أم لا؟

تستخدم الشبكات العصبونية الاصطناعية (ANNs) على نطاق واسع في التشخيص الطبي. بفضل قدرتها على التعامل مع البيانات المتنوعة، فقد أثبتت الشبكات العصبونية أنَّها مفيدة في تحليل عينات الدم والبول عند

مخبرياً. وتتكون قاعدة البيانات هذه من 699 سجلاً، منها 458 سجلاً لورم حميد، و 241 سجلاً لورم خبيث. وتستخدم تسع سمات لوصف كل سجلاً من قاعدة بيانات ويسكونسن لسرطان الثدي. من بين 699 هناك 16 سجل يحتوي على سمات قيمها غائبة، ولذلك لجأنا إلى حذف هذه السجلات كلها من قاعدة البيانات. تفاصيل السمات الموجودة في مجموعة البيانات المذكورة في الجدول (1).

الجدول (1) سمات البيانات المستخدمة في البحث [9].

Attribute	Domain
Sample code number	id number
Clump Thickness	1 - 10
Uniformity of Cell Size	1 - 10
Uniformity of Cell Shape	1 - 10
Marginal Adhesion	1 - 10
Single Epithelial Cell Size	1 - 10
Bare Nuclei	1 - 10
Bland Chromatin	1 - 10
Normal Nucleoli	1 - 10
Mitoses	1 - 10
Class	2 for benign, 4 for malignant

- Clump thickness: تميل الخلايا الحميدة إلى أن تكون مجمعة في طبقات وحيدة، في حين غالباً ما تتجمع الخلايا السرطانية في طبقات متعددة.
- Uniformity of cell size/shape: تميل الخلايا السرطانية إلى أن تختلف في الحجم والشكل. وهذا هو السبب في أن هذه المعايير هي قيمة في تحديد هل كانت الخلايا سرطانية أم لا؟
- Marginal adhesion: تميل الخلايا الطبيعية إلى التمسك معاً، أمّا الخلايا السرطانية فتميل إلى فقدان هذه القدرة لذلك فإن فقدان الالتصاق هو علامة على الأورام الخبيثة.
- Single epithelial cell size: الخلايا الظهارية التي تكون متضخمة بشكل كبير قد تكون خلية خبيثة.
- Bare nuclei: النوى العارية هو مصطلح يستخدم للنوى التي لا تحيط بها السيتوبلازما (بقية الخلية). وعادة ما ينظر إليها في الأورام الحميدة.

ذلك بتغذية الطبقة المخفية بالمعلومات، ومن ثمّ تقوم الطبقة المخفية بتغذية طبقة الخرج. المعالجة الفعلية للبيانات تتم في الطبقة المخفية وطبقة الخرج [8].

عندما يكون هناك عدد كافٍ من العصبونات في كل طبقة، تكون الشبكة قادرة على التدريب (training) للقيام بأشياء مفيدة بالاستعانة بخوارزميات التدريب (training algorithms).

خلال عملية التدريب تقوم الشبكة بتعديل مستمر للأوزان (weights) المرتبطة بكل عصبون، إذ يشكل تعديل الأوزان محور عملية التعلم للشبكة العصبونية. بعد كل تعديل للأوزان يقترب خرج الشبكة أكثر فأكثر من قيمة الخرج الحقيقية. تتعلم الشبكة عن طريق إعطائها مجموعة من الأمثلة، التي يجب أن تكون مختارة بعناية، لأن ذلك سيسهم في سرعة تعلم الشبكة. ومجموعة الأمثلة هذه تسمى مجموعة معطيات التدريب. تستمر عملية التدريب على المعطيات المتوافرة كلها وتعديل الأوزان المرتبطة بكل عصبون حتى يصبح خرج الشبكة أقرب ما يمكن للواقع.

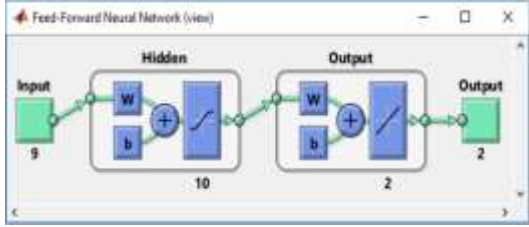
### 3-خطوات إجراء البحث:

- تتضمن خطوات البحث المراحل الآتية:
- تجميع المعطيات.
- بناء الشبكة العصبونية الاصطناعية المناسبة.
- تدريب الشبكة العصبونية على معطيات المرضى التي جُمعت.
- اختبار مدى كفاءة الشبكة المدربة على عينات جديدة لم تمرر للشبكة خلال عملية التدريب.
- في حال اجتازت الشبكة الاختبار بنجاح تصبح جاهزة للاستخدام العملي في التشخيص.

### 4-تجميع المعطيات

في هذا البحث، تم استُخدمت قاعدة بيانات ويسكونسن لسرطان الثدي [9]. وقد تم أنشئت قاعدة البيانات هذه في مستشفيات جامعة ويسكونسن ماديسون، وهي عبارة عن بيانات خزر أُخذت من أناس مرضى وسليمين، وحُلّت

أداء. يبين الشكل (3) بنية الشبكة العصبونية المستخدمة في بحثنا، إذ استخدمنا برنامج الماتلاب كأداة برمجية [10].



الشكل (3) بنية الشبكة العصبونية المستخدمة كما تظهر في الماتلاب.

### 6- تدريب الشبكة العصبونية الاصطناعية

الآن الشبكة على استعداد للتدريب. وتنقسم العينات تلقائياً إلى مجموعة للتدريب، ومجموعة للتدقيق، وأخرى للاختبار.

- عدد العينات الكلي 683 عينة، تمثل عدد المرضى الإجمالي.
  - مجموعة عينات التدريب 479 عينة وهو ما نسبته 70% من عدد العينات الكلي.
  - مجموعة عينات التدقيق 102 عينة، وهو ما نسبته 15% من عدد العينات الكلي.
  - مجموعة عينات الاختبار 102 عينة، وهو ما نسبته 15% من عدد العينات الكلي.
- تظهر البارامترات المستخدمة في الشبكة العصبونية في نافذة التدريب (الشكل 4).

استُخدمت تابع التدريب Levenberg-Marquardt لتدريب الشبكة العصبونية، إذ أظهر نتائج جيدة من حيث دقة التصنيف، وزمن التدريب مقارنة بتوابع التدريب الأخرى التي جُرِّبَتْ. كذلك استُخدمت طريقة التوقيف المبكر (Early Stopping) لتجنب حصول مشكلة التدريب المفرط (Overfitting). وتعتمد هذه الطريقة على بيانات التدقيق في عملها إذ تقوم بإيقاف عملية التدريب في اللحظة التي يبدأ فيها منحنى تابع الخطأ المقيس على هذه

• Bland Chromatin: في الخلايا الحميدة يصف الكروماتين نسيجاً موحداً للنواة. في الخلايا السرطانية يميل إلى الكروماتين إلى أن يكون خشناً.

• Normal nucleoli: النويات العادية هي جسيمات صغيرة ترى في النواة. في الخلايا الطبيعية عادة ما تكون النويات صغيرة جداً، أو حتى تكاد تُرى. في الخلايا السرطانية تصبح النويات أكثر وضوحاً، وأحياناً هناك العديد منها.

• Mitoses: هو الانقسام النووي، وهي العملية التي تنقسم فيها الخلية وتكرر. يقوم مختصو علم الأمراض بتحديد درجة السرطان عن طريق عد عدد هذه الانقسامات.

ويوضّح الشكل (2) عينة صغيرة من قاعدة بيانات ويسكونسن لسرطان الثدي التي استُخدمت في بحثنا.

```
1002945,5,4,4,5,7,10,3,2,1,2
1015425,3,1,1,1,2,2,3,1,1,2
1016277,6,8,8,1,3,4,3,7,1,2
1017023,4,1,1,3,2,1,3,1,1,2
1017122,8,10,10,8,7,10,9,7,1,4
1018099,2,1,1,2,2,10,3,1,1,2
1018561,2,1,2,1,2,1,3,1,1,2
1033078,2,1,1,1,2,1,1,1,5,2
1033078,4,2,1,1,2,2,1,2,1,2
1035283,1,1,1,1,1,1,3,1,1,2
1036172,2,1,1,2,2,1,2,1,1,2
1041801,5,3,3,2,2,3,4,4,1,4
1043999,1,1,1,2,2,3,3,1,1,2
1044572,8,7,5,10,7,9,5,5,4,4
1047630,7,4,6,4,6,1,4,3,1,4
1048672,4,1,1,1,2,1,2,1,1,2
```

الشكل (2) عينة صغيرة من قاعدة بيانات ويسكونسن لسرطان الثدي.

### 5- بناء الشبكة العصبونية الاصطناعية

أنشأنا شبكة عصبونية ذات تغذية أمامية بطبقة مخفية واحدة تحوي 10 عصبونات مخفية. هذه الشبكة تتألف من 9 مداخل كل مدخل يمثل سمة استُخلصت من الخزعة المأخوذة من المريض والمحددة في الجدول (1) في حين تحوي الشبكة خرجين يأخذان القيم [1؛0] في حالة ورم حميد و القيم [0؛1] في حالة ورم خبيث. اختير عدد العصبونات في الطبقة المخفية من خلال مراقبة الخطأ التربيعي المتوسط (mse) لعدد مختلف من العصبونات إذ وجدنا أن الشبكة بـ 10 عصبونات مخفية أعطت أفضل

الإريك. (Confusion Matrix) وتظهر مصفوفة الإريك النسب المئوية للتصنيفات (التشخيصات) الصحيحة وغير الصحيحة وينبع اسم هذه المصفوفة من حقيقة أنها تجعل من السهل رؤية هل كان النظام المصمم مرتكباً بين التصنيفين (أي يخطئ بين أحد التصنيفين على أنه الآخر). يظهر الشكل (5) مصفوفة الإريك الأرقام على طول المحور الرئيس للمصفوفة تمثل القرارات الصائبة المتخذة. أمّا الأرقام على طول المحور الثانوي فتمثل القرارات المغلوطة فيها بين الأصناف المختلفة.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP الإيجابي الحقيقي	FP الإيجابي الكاذب
	Negative	FN السلبي الكاذب	TN السلبي الحقيقي

الشكل (5) مصفوفة الإريك.

يبين الشكل (6) مصفوفة الإريك الخاصة بنظامنا:

		Actual Values	
		Positive	Negative
Predicted Values	Positive	<b>26</b> 25.5%	<b>0</b> 0.0%
	Negative	<b>1</b> 1.0%	<b>75</b> 73.5%

الشكل (6) مصفوفة الإريك الخاصة بنظامنا.

وفيما يأتي النسب المئوية الإجمالية للتشخيص الصحيح وغير الصحيح:

Percentage Correct Classification : 99.01%

Percentage Incorrect Classification : 0.98%

وكما هو واضح فمن أصل 102 مرضى نجحت الشبكة بتشخيص حالة 101 مريض، وأخفقت في حالة

البيانات بالتزايد إذ يكون أداء الشبكة أفضل ما يمكن في هذه الحالة.



الشكل (4) نافذة التدريب كما تظهر في برنامج الماتلاب.

## 7- اختبار الشبكة العصبونية الاصطناعية

يمكننا الآن اختبار الشبكة العصبونية المدربة مع عينات الاختبار التي قسمناها من مجموعة البيانات الرئيسية. لم نستخدم بيانات الاختبار في التدريب بأي شكل من الأشكال، ومن ثم فهي توفر مجموعة بيانات مهمة لاختبار الشبكة. وهذا سوف يعطينا تصوراً كيف سيكون أداء الشبكة عند اختبارها مع بيانات من العالم الحقيقي.

## 8- مناقشة النتائج

ولأجل تقييم أداء النظام المقترح حسب بعض المقاييس (المعايير) الشائعة اعتماداً على مصفوفة

تجدر الملاحظة إلى أنّ النظام المقترح يقوم بتصنيف الورم إمّا خبيثاً أو حميداً دون تحديد المرحلة التي وصل إليها الورم (مبكرة، متقدمة، ...).

### 9-الخلاصة:

تمثل ال ANNs أداة فعالة لمساعدة الأطباء على إجراء التشخيص في الحقل الطبي. وقد أثبتت ANNs نجاعتها في تشخيص سرطان الثدي، إذ إنّ استخدامها يجعل التشخيص أكثر دقة و موثوقية، ومن ثمّ يزيد من رضا المرضى. ومع ذلك، لا بدّ من عدّها فقط كأداة لدعم القرار النهائي للطبيب، فهو المسؤول في نهاية المطاف عن التقييم النهائي.

### المصطلحات

المصطلحات باللغة العربية	المصطلحات باللغة الانكليزية
الشبكة العصبونية الاصطناعية	Artificial Neural Network
مصفوفة الإرباك	Confusion Matrix
واجهة مستخدم رسومية	GUI
خوارزميات التدريب	Training Algorithms
البيبرسيبترون متعدد الطبقات	Multilayer perceptron (MLP)
التدريب المفرط	Overfitting

واحدة؛ وهي نسبة مرتفعة تعبر عن دقة نظام التشخيص المقترح في بحثنا وموثوقيته. وفيما يأتي بعض المعايير الشائعة التي يمكن حسابها من مصفوفة الإرباك:

• الحساسية Sensitivity

$$Sensitivity = \frac{TP}{TP + FN} \times 100 = 96.2\%$$

• النوعية Specificity

$$Specificity = \frac{TN}{TN + FP} \times 100 = 100\%$$

• الدقة Accuracy

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100 = 99\%$$

ومن أجل الإفادة القصوى من نتائج بحثنا من قبل الأطباء و الإطار الطبي الممارس في المستشفيات والعيادات صمّمنا واجهة رسومية تفاعلية GUI باستخدام برنامج الماتلاب تسهل على الطبيب الممارس استخدام نظام التصنيف الأوتوماتيكي، إذ يقوم الطبيب بإدخال بيانات الخزعة المأخوذة من المرضى في الحقول المخصصة لها، ثم يقوم بالنقر على زر النتيجة لتظهر نتيجة التشخيص بكل سهولة و يسر كما يوضّح الشكل (7).

الشكل (7) واجهة رسومية تفاعلية GUI للطبيب

## References

- [1] Catalogna M, Cohen E, Fishman S, Halpern Z, Nevo U & Ben-Jacob E. (2012). “Artificial neural networks based controller for glucose monitoring during clamp test”. *PloS One*. 7: e44587.
- [2] Fernandez de Canete J, Gonzalez-Perez S & Ramos-Diaz JC. (2012). “Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes”. *Comput Meth Progr Biomed*. 106: 55–66.
- [3] Elveren E & Yumuşak N. (2011). “Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm”. *J Med Syst*. 35: 329–332.
- [4] Er O, Temurtas F & Tanrikulu. (2008). “A Tuberculosis Disease Diagnosis Using Artificial Neural Networks”. *J Med Syst*. 34: 299–302.
- [5] Dey P, Lamba A, Kumari S & Marwaha N. “Application of an artificial neural network in the prognosis of chronic myeloid leukemia”. (2012). *Anal Quant Cytol Histol*. 33: 335–339.
- [6] Barbosa D, Roupar D, Ramos J, Tavares A & Lima C. (2012). “Automatic small bowel tumor diagnosis by using multi-scale wavelet-based analysis in wireless capsule endoscopy images”. *Biomed Eng Online*. 11: 3.
- [7] Saghiri M, Asgar K, Boukani K, Lotfi M, Aghili H, Delvarani A, Karamifar K, Saghiri A, Mehrvarzfar P & Garcia-Godoy F. (2012). “A new approach for locating the minor apical foramen using an artificial neural network”. *Int Endod J*. 45: 257–265.
- [8] Filippo Amato, Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš Hampl & Josef Havel. (2013). “Artificial neural networks in medical diagnosis”. *Appl Biomed*. 11: 47–58.
- [9] [https://archive.ics.uci.edu/ml/datasets/breast+caner+wisconsin+\(original\)](https://archive.ics.uci.edu/ml/datasets/breast+caner+wisconsin+(original)).
- [10] <https://www.mathworks.com>

Received	1/4/2019	إيداع البحث
Accepted for Publ.	24/6/2019	قبول البحث للنشر

[11]