

التنبؤ باحتمالات إصابات مرضى السرطان في محافظات غزة
- باستخدام تقنية الشبكات العصبية الاصطناعية -

Forecasting the probabilities of cancer patients in Gaza provinces
- Using artificial neural network technology methodology -

Prévision des probabilités de patients atteints de cancer dans les
provinces de Gaza - Par l'utilisation de la méthodologie de la technologie
des réseaux neuronaux artificiels -

شريف شحده مصلح* & مولود حشمان**

تاريخ قبول النشر: 2023-12-05

تاريخ استلام المقال: 2022-11-02

Abstract:

This research aims to formulate a time series model using artificial neural networks approach to predict the number of cancer patients in Gaza governorates, based on the monthly data for the period (2009-2018), obtained from the Palestinian ministry of health, Health Information Systems. Unit (2018), Monthly Cancer Patient Incidence Series, Gaza. The results of the data analysis showed that the appropriate model is the autoregression of neural networks NNAR (12.40), according to the root of the mean of the residuals to measure the accuracy of the prediction. The number of cancer patients is predicted monthly, for the next two years. The results, whether statistical or graphical tracing approach, showed high quality of forecasting.

Key words: Artificial Neural Networks, Numbers of cancer patients, Forecasting, Gaza provinces.

* المدرسة الوطنية العليا للإحصاء والاقتصاد التطبيقي (الجزائر) - دكتوراه

(المؤلف المرسل) sh.pal2010@gmail.com

** جامعة الجزائر 3 (الجزائر) - أستاذ

hachemane@hotmail.com

Résumé:

Cette recherche vise à formuler un modèle de série chronologique utilisant l'approche des réseaux neuronaux artificiels pour prédire le nombre de patients atteints de cancer dans les gouvernorats de Gaza, en se basant sur des données mensuelles pour la période (2009-2018) obtenues auprès du ministère palestinien de la Santé, Unité des systèmes d'information sanitaire (2018), Monthly Cancer Patient Incidence Series, Gaza. Les résultats de l'analyse des données ont montré que le modèle approprié est l'autoregression des réseaux neuronaux NNAR(12,40), selon la racine de la moyenne des résidus. Le nombre de patients atteints de cancer est prévu mensuellement, pour les deux prochaines années. Les résultats, qu'il s'agisse d'une approche de traçage statistique ou graphique, ont montré une grande qualité de prévision.

Mots clés: Réseaux de neurones artificiels, Nombre de patients atteints de cancer, Prévisions, Provinces de Gaza.

ملخص:

يهدف هذا البحث إلى صياغة نموذج سلاسل زمنية باستخدام تقنية الشبكات العصبية الاصطناعية للتنبؤ بأعداد المصابين بأمراض السرطان في محافظات غزة، اعتماداً على البيانات الشهرية للفترة (2009-2018)، تم الحصول عليها لدى وزارة الصحة الفلسطينية، وحدة نظم المعلومات الصحية (2018)، سلسلة إصابات مرضى السرطان الشهرية، غزة. أظهرت نتائج تحليل البيانات أن النموذج الملائم هو الانحدار الذاتي للشبكات العصبية NNAR(12,40)، وفقاً لمعيار جذر متوسط مربعات البواقي لقياس دقة التنبؤ. حيث يتم التنبؤ بأعداد المصابين بأمراض السرطان شهرياً، ولسنتين قادمتين. بينت النتائج سواء الإحصائية أو مقارنة اقتفاء الأثر البياني جودة عالية في التنبؤ.

الكلمات المفتاحية: الشبكات العصبية الاصطناعية، أعداد المصابين بأمراض السرطان، التنبؤ، محافظات غزة.

مخطط المقال:

مقدمة

1) الجانب النظري

1-1) تقديم الدراسة

2-1) الشبكات العصبية الاصطناعية

2) الجانب التطبيقي

1-2) النتائج

2-2) مناقشة النتائج

خاتمة

مقدمة:

إن عملية التنبؤ في السلاسل الزمنية تتأثر بشكل مباشر باختيار النموذج المناسب وقد تتأثر دقة التنبؤ بحجم مشاهداتها، وتؤثر هذه الخطوة في دقة التنبؤات المتحصل عليها، حيث أن محافظات غزة تتعرض لكارثة صحية وتحديا كبيرا يتمثل في توفير الإمكانات الصحية المختلفة من مستشفى وكادر طبي متخصص ومعدات طبية متطورة للحد من الازدياد المضطرد في إصابات مرضى السرطان، حيث بلغ عدد الحالات المكتشفة في الفترة ما بين (2009-2018) 13893 حالة، ولبناء استراتيجيات مستقبلية لمواجهة ذلك.

تمحورت مشكلة الدراسة في التساؤل الرئيس التالي: ما النموذج الأمثل والأكثر كفاءة من بين النماذج ANN وبعد المفاضلة بينهما وفقا للمعايير الإحصائية المناسبة للتنبؤ بسلوك واحتمالات الإصابة بمرضى السرطان بمحافظات غزة؟

ومن أجل التوصل إلى إجابات على الأسئلة المطروحة في الإشكالية يستوجب الاستعانة بالفرضيتين:

- السلسلة الزمنية لمرضى السرطان بمحافظات غزة غير مستقرة في المتوسط والتباين؛
- يعتبر نموذج الشبكات العصبية NNAR الذي يحمل وزن وفقا معادلة FARAWAY & CHATFIELD (1998) وأقل فرق مطلق للمعيار الإحصائي RMSE قيم لمعيار RMSE من بين النماذج أكثر كفاءة للتنبؤ باحتمالات الإصابة بمرضى السرطان.

(1) الجانب النظري:

إن التنبؤ من أهم أدوات اتخاذ القرار وأهم عنصر في عملية التخطيط للمستقبل، وهو أحد أهم القضايا الأساسية المستخدمة في عدة مجالات وأهمها المجال الاقتصادي والصحي، ويعتبر تحليل السلاسل الزمنية من الأساليب الإحصائية الشائعة المستخدمة في التنبؤ على نطاق واسع في الكثير من التطبيقات الإحصائية والاقتصادية والصحية وفي مجال علوم الحاسبات والشبكات العصبية الاصطناعية وغيرها، حيث يتم التنبؤ بسلوك المتغير التابع استناداً إلى سلوكه في الماضي. والشبكات العصبية نظام حسابي يتكون من مجموعة من العناصر البسيطة والمرتبطة مع بعضها البعض لتشغيل البيانات على نحو ديناميكي كاستجابة لمداخل خارجية. لذلك، سوف نقوم بدراسة هذه الظاهرة بالاعتماد على بيانات السلسلة الزمنية الشهرية للفترة (2009-2018) باستخدام النموذج الأمثل والأكثر كفاءة من بين نماذج تقنية الشبكات العصبية، للتنبؤ باحتمالات الإصابة بمرض السرطان وفقاً لأقل القيم لمعايير دقة التنبؤ.

(1-1) تقديم الدراسة:

يوجد طريقة حديثة أكثر دقة وكفاءة للتنبؤ يمكنها استخدام المنطق في عملياتها بدلاً من فكرة العلاقة الثابتة بين الرموز وردود الأفعال وهي الشبكات العصبية الاصطناعية Artificial Neural Networks (ANNs) أحد فروع الذكاء الاصطناعي حيث ظهرت في الفترة الأخيرة اهتمامات خاصة لإدخال هذه الطريقة إلى التطبيقات الإحصائية، وبما أن أسلوب (ANNs) لا يتطلب افتراضات عن طبيعة السلسلة الزمنية كونها خطية أو طبيعية أو مستقرة، يعتقد أن استخدام هذا الأسلوب قد يكون مناسباً في تمثيل العلاقات بين المتغيرات بشكل مختلف عن الطرق التقليدية، ويمكنها أن تعالج العديد من البيانات الخطية وغير الخطية بالإضافة إلى البيانات الناقصة Incomplete والمشوشة Noisy، فالشبكات العصبية نظام حسابي يتكون من مجموعة من العناصر البسيطة والمرتبطة مع بعضها البعض لتشغيل البيانات على نحو ديناميكي كاستجابة لمداخل خارجية، لذا تعتبر نظام معالجة للمعلومات، له مميزات أداء معينة بأسلوب يحاكي الشبكات العصبية الحيوية¹.

على ذلك، فإن البحث يقوم على فكرة تصميم نموذج باستخدام تقنية الشبكات العصبية، ونظراً لازدياد إصابات مرض السرطان في الآونة الأخيرة، فقد جاءت هذه الدراسة لبناء نموذج قياسي أمثل لكشف هذه الظاهرة والتي ازدادت بشكل متنام في محافظات غزة التي تأثرت

بالأسلحة المحرمة دولياً، وأهمية التنبؤ باحتمالات الإصابة بمرضى السرطان كأحد أدوات توظيف الدراسات المستقبلية وهذا ما يكون صانع السياسة بحاجة إليه لنجاح تحليل المشكلات، وتشخيص دقيق للإختلالات، وبالتالي القدرة على اتخاذ القرارات الصائبة، ووضع الخطط المستقبلية الملائمة والتي تعود للمعرفة الجيدة من قبل متخذي القرار بسلوك متغير مرضى السرطان أو أي متغيرات أخرى، والقدرة على التنبؤ الدقيق بها لوضع الخطط طويلة وقصيرة الأمد بما يتناسب مع تغيرات وظروف المستقبل، وذلك عندما تكون التنبؤات على درجة عالية من الدقة والتأكد مما يقلل من نسبة الخطأ التي تؤدي لفشل الخطط المعدة.

تتجلى أهمية هذه الدراسة في تعويض نقص الدراسات التي أجريت على مرضى السرطان باستخدام أسلوب تقنية الشبكات العصبية، وتكمن أهمية هذه الدراسة في أنها تتكون من جانبين: تطبيقي وإحصائي، فمن الجانب التطبيقي فإن بناء نموذج قياسي أمثل للتنبؤ يحقق الكثير من المنافع الصحية في محافظات غزة، تتمثل في إمداد أصحاب القرار بمعلومات مستقبلية عن موضوع مهم بسلوك واحتمالات الإصابة بمرضى السرطان، وما يترتب على ذلك من وضع الخطط حسب النتائج المتوقعة لمواجهة المشاكل الصحية المترتبة على التغير في إصابات مرضى السرطان، أما من الجانب الإحصائي؛ فإن استخدام التحليل الإحصائي المعتمد على نماذج وأساليب حديثة ومبتكرة للتنبؤ بالسلسلة الزمنية مثل نماذج الشبكات العصبية يجعل النموذج المستخدم أكثر كفاءة.

- تعي هذه الدراسة إلى تحقيق جملة من الأهداف التي يمكن إيجازها في النقاط التالية:
- بناء نموذج قياسي وصياغته رياضياً لطبيعة مرضى السرطان في محافظات غزة، بحيث يجعل البواقي بين القيم الحقيقية للسلسلة والقيم المقدرة، أقل ما يمكن، ولا يوجد بها أي نوع من الترابط الداخلي فيما بينهم؛
 - بناء النماذج NNAR، والمفاضلة بينها وفقاً لأقل معايير التقييم الإحصائية المناسبة، وذلك لاختيار النموذج الأمثل والأكثر كفاءة من بين النماذج؛
 - التنبؤ باحتمالات الإصابة بمرضى السرطان باستخدام النموذج الأمثل والأكثر كفاءة في التنبؤ من بين النماذج NNAR شهرياً لعينة الاختبار سنة 2018 ولسنتي 2019 و2020؛
 - ودعم جهود متخذي القرار عند قيامهم بالتخطيط للحد من مرضى السرطان.

استناداً لدراسات سابقة، تعتمد الدراسة على المنهج الإحصائي الوصفي والمنهج القياسي. تعتمد على المنهج الإحصائي الوصفي لوصف بيانات السلسلة الزمنية لمرضى السرطان الشهرية في محافظات غزة من خلال عرضها وتلخيصها بطرق إحصائية مثل جدول وشكل بياني وبعض المقاييس الوصفية كمقاييس النزعة المركزية والتشتت من أجل معرفة التفاصيل

المتعلقة بهذه البيانات مثل اتجاهات التزايد أو التناقص أو الثبات في البيانات، ومعرفة طبيعة التغيرات التي تؤثر في السلسلة الزمنية محل الدراسة. أما المنهج الإحصائي القياسي، فيستخدم للتحقق من مدى صحة انطباق الفرضية من عدمها، وذلك لزيادة الدقة في التنبؤ، وهذه الأسلوب متمثل في تقنية الشبكات العصبية.

تعتمد الدراسة على البيانات الشهرية لمرضى السرطان، حيث تم تقسيم البحث إلى حدود زمنية وحدود مكانية، وبيانات البحث عبارة عن سلسلة زمنية واقعية ذات مدى طويل (120) مشاهدة:

- الحدود الزمنية: حددت فترة الدراسة على مرضى السرطان بمحافظات غزة من 01 يناير 2009 حتى 31 ديسمبر 2018؛
 - والحدود المكانية (الجغرافية): انحصرت الدراسة على محافظات غزة - فلسطين.
- وتم الحصول على بيانات الدراسة من وحدة نظم المعلومات بمركز المعلومات الصحية في وزارة الصحة الفلسطينية².

1-2) الشبكات العصبية الاصطناعية:

إن الشبكات العصبية الاصطناعية إحدى أهم أنواع الذكاء الاصطناعي وأقوى تقنياتها، هذا العلم الحديث الذي تبنى عليه جميع التطبيقات الحديثة والمعقدة من أنظمة التنبؤ والتعرف، ونظم دعم القرار، وأنظمة التحكم الآلي، وهي نظام معالجة معلومات تفرعي ومكونة من عناصر بسيطة ومتوازية التركيب تسمى بالعصبونات أو العقدمتصلة بروابط موزونة، صممت كنماذج رياضية لمحاكاة الشبكة العصبية الحيوية الموجودة في دماغ الإنسان البشري مثل المقدرة على التعلم، الاستنتاج وحل المشاكل المعقدة وكل أنماط السلوك الذي يتميز به الإنسان عن بقية الكائنات الحية ودفع تقنية الذكاء الاصطناعي للأمام.

تتكون الشبكة العصبية الاصطناعية من مجموعة مركبة من عدة عناصر معالجة والتي تسمى خلايا عصبية أو عصبونات (عقد) ويطلق عليها اسم نيورونات (Neurons) والتي تعمل في انسجام لحل مشكلة معينة، وهذه العناصر متصلة فيما بينها في شكل ترابطات تسمى بالشبكة العصبية الاصطناعية وتعمل بشكل مقارب للشبكة العصبية الحيوية ولها القدرة على إجراء العمليات الحسابية من خلال عدة خطوات³:

1. تتم معالجة البيانات في عناصر معالجة بسيطة تسمى نيورونات؛
2. تمر الإشارات بين النيورونات عبر خطوط ربط ناقلة؛
3. يرفق كل خط ربط بوزن معين والذي يضرب مع الإشارات الداخلة إلى النيورون؛

4. ويطبق كل نيورون دالة تنشيط معينة، عادة تكون هذه الدالة غير خطية على إجمالي مدخلاته (مجموع الإشارات الداخلة الموزونة) ليحدد إشارة المخرجات الناجمة منه.

للشبكة العصبية ثلاث مستويات حيث يمثل المستوى الأول طبقة المدخلات والمستوى الثاني يمثل الطبقة المخفية والمستوى الثالث يمثل طبقة المخرجات ويتضمن كل مستوى عدداً من العقد، وبين كل طبقة من هذه الطبقات هناك ضبط الأوزان الخاصة بكل وصلة بينية، وتحتوي الشبكة العصبية على طبقة واحدة فقط من وحدات الإدخال ولكنها قد تحتوي أكثر من طبقة من طبقات المعالجة⁴.

الخلية العصبية الاصطناعية أو العصبون هي مكونة من مكونات الشبكة العصبية الاصطناعية، وهي مستوحاة من الخلية البيولوجية، وهي تعرف على أنها وحدة آلية مبسطة متصلة بوحدات آلية مجاورة، وهي مزودة بحالة داخلية تنشيطية تؤثر بها على وحدات الشبكة وينتقل هذا الأثر عبر الارتباطات المشبكية التي تتمثل في الترميمات الشبكية، وبصفة عامة هي عبارة عن وحدة حساب قوية الارتباط فيما بينها تقوم بمعالجة البيانات في الشبكة الاصطناعية، ويتكون العصبون الاصطناعي من ثلاث أجزاء رئيسية هي⁵:

1. إشارات الدخل Input signs: إشارة الدخل ($x_1, x_2, x_3, \dots, x_n$) تكون على صورة بيانات كمية أو نوعية أو تكون مخرجات لوحدة معالجة أخرى، أو تكون عبارة عن نصوص أو صور أو صوت أو أشكال أو ظواهر أخرى.

2. الروابط الموزونة Weighted connections: قوى الأوزان ($w_{1j}, w_{2j}, w_{3j}, \dots, w_{nj}$) تعبر عن قوة الارتباط بالخلايا أو شدة الترابط بين عنصر قبله وعنصر بعده، ويعرف w_{ij} على أنه قيمة الترجيح الخاص بالارتباط الناتج عن الإشارة الصادرة من الوحدة j إلى الوحدة i بالنسبة لنظام بسيط.

3. عناصر المعالجة (J) Processing Elements: وهي العناصر التي تقوم بعملية معالجة المعلومات في الشبكة العصبية وتتصل هذه العناصر بطرق مختلفة بواسطة الوصلات البينية، ويتألف عنصر المعالج من مكونين أساسيين:

➤ دالة الجمع Summation Function:

تقوم دالة الجمع بحساب صافي إشارة الدخل وهو ناتج مجموع حاصل ضرب المدخلات في وزن الرابط لكل عناصر المدخلات التي تم إدخالها، وبذلك يتم الوصول إلى مجموع مرجح والذي يعطي بالمعادلة الرياضية الآتية⁶:

$$S_j = \sum_{i=1}^n w_{ij}x_i = w_1x_1 + w_2x_2 + \dots + w_nx_n$$

➤ دالة التنشيط الابتدائي Activation Function:

بعد عملية الجمع يأتي عمل دالة التنشيط والتي تسمى أيضًا أو التحويل حيث يتم تحويل ناتج الجمع الموزون في الخطوة الأولى إلى أحد القيم التي يفترض أن تكون ضمن نواتج الشبكة المطلوبة، ويتم ذلك بمقارنة نتيجة الجمع مع قيمة تسمى العتبة Threshold ويرمز لها بالرمز (F) ليتحدد الناتج، ويطبق على المجموع عادة قبل المقارنة تابع تنشيط معين وتتوقف مخرجات الشبكة بصورة أساسية على هذا التوزيع، فإذا كان مجموع أوزان الإشارات يزيد عن عتبة تكون الإشارة المخرجة هي واحد وإذا كان أقل يكون الناتج صفر.

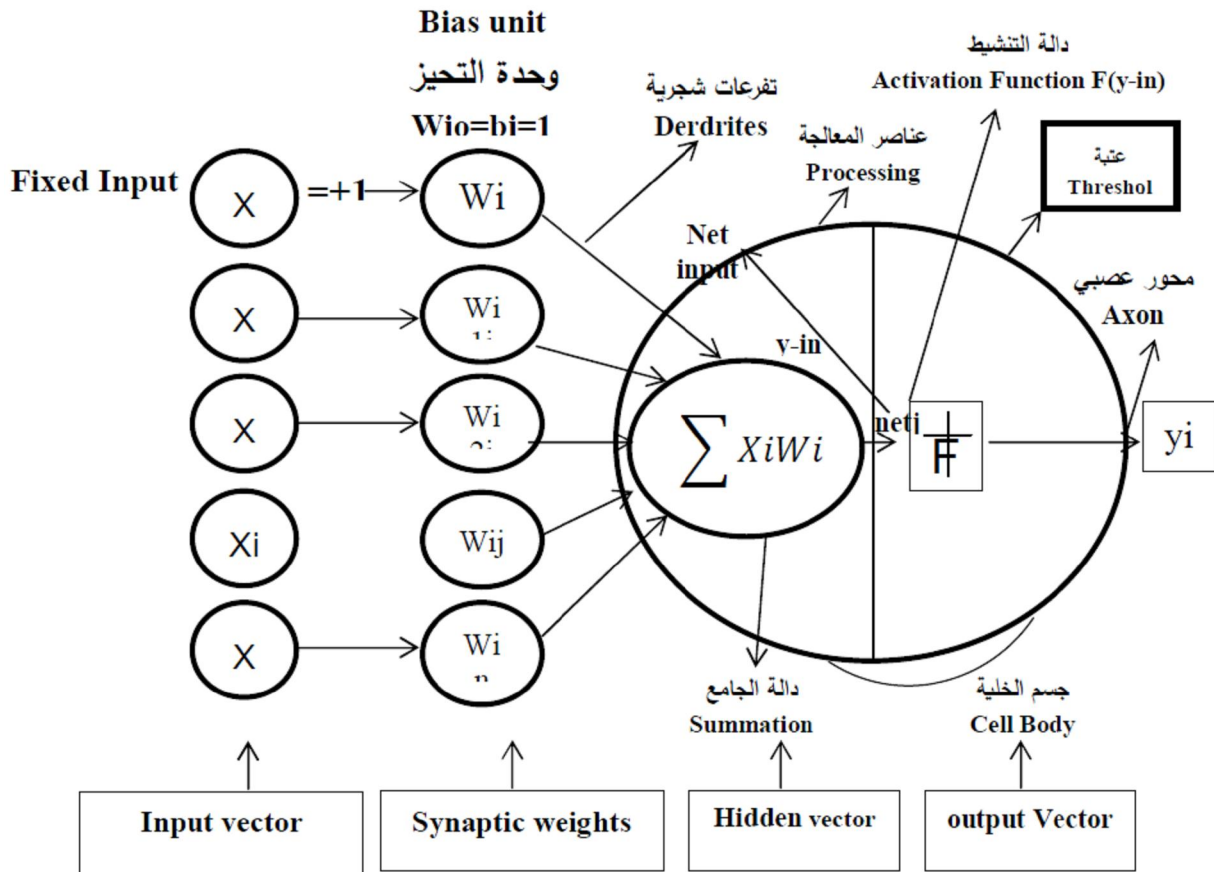
ودوال التنشيط تحد من خرج العصبون، لذا يسمى بتابع التخميد Squashing حيث يجعل المخرجات محصورة ضمن المجال [0,1] أو ضمن المجال [-1,1]⁸، ويتم التحكم بمخرجات العصبون (J) باستخدام دالة التنشيط F(.) من خلال المعادلة الآتية⁹:

$$Y_i = F \left[\sum_{j=1}^n W_{ij} X_j + b_i \right]$$

ومن أهم الدوال، دوال السيغمويد Sigmoid function، دالة الظل القطعي الزائدي Hyperbolic tangent function، دالة الخطوة Step function، الدالة الخطية بعتبة Linear function by threshold، الدالة الخطية Linear function، دالة الإشارة Sign function.

يوضح الشكل 01 الهيكل الرياضي للعصبون الاصطناعي كالتالي¹⁰:

الشكل 01 - الهيكل الرياضي للعصبون الاصطناعي



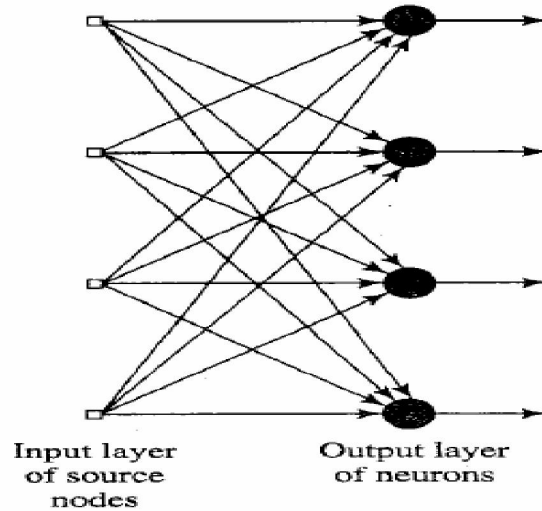
4. الإخراج (Yi) Out put: الإخراج عبارة عن الناتج أو الحل للمشكلة محل الدراسة.

تم بناء هذه النماذج واختيار وتحديد النموذج الأمثل من خلال عدة مراحل، بداية من استخدام أسلوب شبكة البيرسبترون متعددة الطبقات (Multi-layer Perceptron (MLP لبناء نماذج الشبكات العصبية متعددة الطبقات ذات التغذية الأمامية لبيانات السلسلة محل الدراسة، وصولاً إلى عملية بناء الشبكات العصبية. وتعتبر شبكة البيرسبترون متعددة الطبقات من أكثر نماذج الشبكات العصبية استخداماً خاصة في مجال التنبؤ بالسلاسل الزمنية، وتقوم فكرة هذه الشبكة على استخدام القيم السابقة للسلسلة الزمنية y_{t-1} كمداخلات الشبكة، ويتم تجميع الأوزان في الطبقة الخفية بالنسبة للمدخلات، وترتبط هذه الطبقات فيما بينها بوصلات

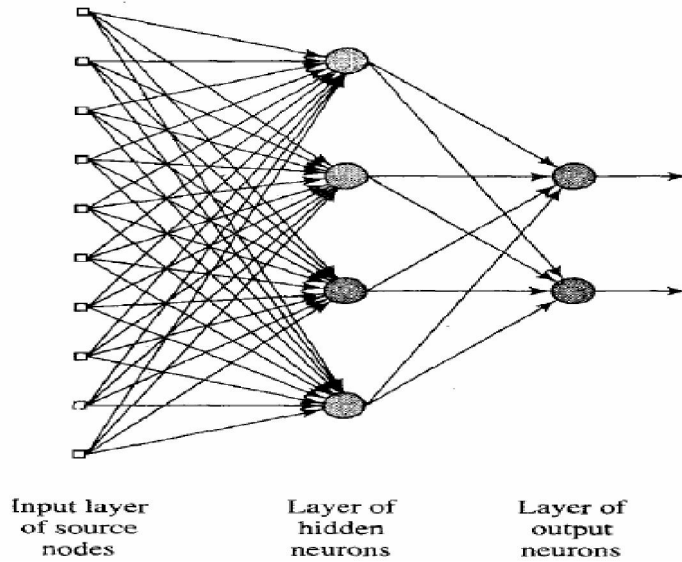
بينية تحمل الأوزان، وقد تم استخدام دالة التنشيط السيغمويدي في كل من الطبقة الخفية وطبقة المخرجات.

تتحرك المعلومات في هذه الشبكة في اتجاه واحد فقط، إنها تتحرك إلى الأمام من خلال إدخال وإخراج الطبقات المخفية (إن وجدت)، فإذا كانت الشبكة أمامية التغذية تحتوي على طبقة مدخلات وطبقة مخرجات فقط بدون طبقات مخفية يطلق عليها شبكات عصبية أمامية التغذية وحيدة الطبقة كما في الشكل 02، أما إذا احتوت الشبكة الأمامية على الأقل على طبقة مخفية واحدة فإنها تصبح شبكة أمامية متعددة الطبقات كما في الشكل 03^{1f}:

الشكل 02 - شبكة أمامية ذات طبقة وحيدة



الشكل 03 - شبكة أمامية متعددة الطبقات



من أهم الاعتبارات في الشبكة العصبية استخدام خوارزمية التعلم المناسبة، وتوجد العديد من خوارزميات التعلم ويتوقف اختيار الخوارزمية المناسبة على حسب نوع الشبكة، ومن أكثر الخوارزميات استخدامًا: خوارزمية الانتشار العكسي Back Propagation Algorithm، وخوارزمية التغذية الأمامية Feed Forward Algorithm، وخوارزمية التعلم الهيبان Habbian Learning Algorithm، وخوارزمية التعلم التنافسي Competitive Learning (AlgorithmAmpazis, 1998).

تهدف خوارزميات التعلم في الشبكات العصبية إلى تحديد قيمة الأوزان المرجحة من خلال تدريب الشبكة بإحدى طرق التعلم المختلفة والتي يمكن تقسيمها إلى ثلاثة طرق رئيسية وهي التعلم المراقب Supervised Learning، التعلم غير المراقب Unsupervised Learning، والتعلم بإعادة التدعيم Reinforcement Learning¹²:

(2) الجانب التطبيقي:

تتألف هذه السلسلة الشهرية من 120 مشاهدة، وتعود إلى الفترة من يناير عام 2009 إلى ديسمبر عام 2018، وأن هذه البيانات محل الدراسة تمثل الأعداد المسجلة لمرضى السرطان بمحافظة غزة، وتم الحصول عليها من سجلات وزارة الصحة الفلسطينية "وحدة نظم المعلومات" في مركز المعلومات الصحية، وكما موضحة في الجدول التالي:

(1-2) النتائج:

الجدول 01 - أعداد مرضى السرطان بمحافظة غزة *

Year/ Month	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
January	89	86	98	88	96	126	160	115	144	144
February	76	74	90	73	100	137	130	120	134	135
March	80	78	95	82	103	105	138	123	135	136
April	81	79	100	93	107	135	142	145	148	148
May	86	84	118	99	110	110	120	149	147	147
June	75	73	98	102	120	116	122	153	144	145
July	70	69	90	105	123	112	128	157	143	144
August	81	79	69	112	125	114	133	165	147	148
September	76	76	87	115	128	115	140	165	151	152
October	84	83	73	118	130	162	139	160	159	160
November	73	74	72	121	132	142	134	140	149	149
December	74	75	47	123	140	128	142	110	143	142
Sum	945	940	1037	1231	1414	1502	1628	1702	1744	1750

المصدر: وزارة الصحة الفلسطينية، وحدة نظم المعلومات الصحية (2018)، «سلسلة إصابات مرضى السرطان الشهرية في الفترة (2009-2018)»، غزة.

(2-2) مناقشة النتائج:

تقسيم ومعالجة البيانات:

- تم تقسيم البيانات المدخلة إلى مجموعتين، هما:
1. بيانات التدريب train-data والتي تحتوي على 108 مشاهدة متسلسلة من البيانات الموجودة تحت تصرفنا من 1-1-2009 إلى 31-12-2017 لتدريب وتجهيز النماذج؛
 2. بيانات الاختبار test- data والتي تحتوي على البيانات الـ 12 المتبقية للتحقق من صلاحية النماذج، ومعرفة كفاءة هذه التقنية في تحليل السلاسل الزمنية واختبار الشبكة. من الممارسات الجيدة جعل البيانات تأخذ نمط معين، بإجراء عملية Normalization، بهدف التخلص من وحدات البيانات الملوثة، مما يتيح لنا مقارنة البيانات بسهولة من مواقع مختلفة، وهذا إجراء مهم للغاية في بناء شبكة عصبية، لأنه يتجنب النتائج غير الضرورية أو عمليات التدريب الصعبة للغاية التي تؤدي إلى مشاكل الخوارزمية، وفي دراستنا استخدمنا طريقة Mini Max Method Normalization (تسمى عادة تحجيم المعالم) للحصول على جميع البيانات المقاسة في النطاق الثنائي (0,1) بهدف أن تتناسب دالة التنشيط السيمغويد المستخدمة في الطبقة المخفية، ومنع النموذج من أن تؤثر عليه المتغيرات ذات القيم المتطرفة، والصيغة الرياضية للطريقة ذكرناها سابقا وهي كالتالي:

$$X_{new} = z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

تحديد الهندسة المعمارية للشبكة العصبية:

- تم استخدام برنامج التحليل الإحصائي R لتركيب نموذج الشبكة العصبية وتتكون معمارية الشبكة العصبية المختارة من ثلاث طبقات (طبقة المدخلات - الطبقة الخفية - طبقة المخرجات)، وقد توزعت عناصر المعالجة في طبقات الشبكة كما يلي:
- طبقة المدخلات: إن المدخلات هنا تمثل بيانات شهرية لمرضى السرطان بمحافظة غزة من 1-1-2009 إلى 31-12-2018 وبذلك فإن طبقة المدخلات تتكون من وحدة معالجة (عصبون)؛
 - الطبقة المخفية: سنقوم بإنشاء شبكة عصبية مع طبقة مخفية واحدة فقط وتحتوي على عدد من وحدات المعالجة والتي يتم تحديدها من خلال منهجية التجريب والخطأ وذلك

- بتغيير عدد عصبونات الطبقة المخفية وتدريب الشبكة بهدف الحصول على أقل قيم لمعياري RMSE؛
- طبقة المخرجات: تحتوي هذه الطبقة على الأعداد الشهرية لمرضى السرطان المتتبا بها وعليه فإنها تحتوي على وحدة معالجة واحدة.

تدريب الشبكة:

في هذه المرحلة يتم تحديد خوارزمية التعلم التي تتناسب مع تركيبة الشبكة، حيث تم الاعتماد على دالة التنشيط السيغمويد (افتراضي) كدالة تحفيز في الطبقة الخفية، وكذلك في طبقة المخرجات، وتم استخدام خوارزمية الانتشار العكسي Backpropagation Algorithm لتدريب الشبكة، وتم استخدام التدريب المراقب بإشراف.

تقدير نموذج الشبكة:

في هذه المرحلة تم تقدير عدد من النماذج التي تم دراستها للوصول إلى أفضل نموذج وقد تم استخدام دالة Nnetar في برنامج التحليل الإحصائي R لتقدير تلك النماذج، حيث أن النموذج المولد من هذه الدالة يسمى (NNAR) Neural Networks Auto Regressive. وللحصول على هذه النتائج نبدأ بخمس عقد مخفية في الطبقة المخفية ونزيد تدريجيا 5 عقد في كل مرة حتى الوصول إلى 55 عقدة كالتالي:

- (5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55) مع فجوات زمنية متسلسلة من 1 إلى 15 عقدة في طبقة المدخلات

- بعد تدريب الشبكة وتطبيق خوارزمية الانتشار العكسي ذات التغذية الامامية، نحصل على نتائج RMSE لمائة خمسة وستون نموذجا، كما هو موضح في الجدول 02 وللمفاضلة بين هذه النماذج والحصول على النموذج الأفضل، يتم إتباع الخطوات التالية:
- ✓ يتم البحث عن أفضل عدد لوحات المعالجة في الطبقة المخفية لكل فجوة زمنية، وذلك من خلال البحث عن أقل فرق مطلق للمعيار الإحصائي RMSE المقابل لكل وحدة معالجة في الطبقة المخفية بحيث نتوقف عن البحث عندما لا يتجاوز هذا الفرق للنموذجين المراد المفاضلة بينهما عن 0.0001

$$|RMSE_i - RMSE_{i+1}| = 0.0001$$

- ✓ يتم البحث عن أفضل فجوة زمنية، وذلك من خلال البحث عن أقل فرق مطلق للمعيار الإحصائي RMSE المقابل لوحات المعالجة التي تم اختيارها في الخطوة الأولى، بحيث نتوقف عن البحث عندما لا يتجاوز هذا الفرق للنموذجين المراد المفاضلة بينهما عن 0.0001.

الجدول 02 - نتائج RMSE لنماذج ANNs المرشحة لبيانات سلسلة مرضى السرطان

			Numbers of nodes in hidden layer								
Lags p	5	10	15	20	25	30	35	40	45	50	55
1	0.094865	0.092121	0.0917689	0.091255	0.0917558	0.090907	0.0913742	0.0912922	0.091367	0.091997	0.091733
2	0.081627	0.076895	0.0747315	0.073301	0.0711347	0.0726064	0.0730654	0.0714812	0.07168	0.071232	0.072169
3	0.073207	0.060625	0.0565078	0.05507	0.0517072	0.0521079	0.0492395	0.04887	0.048479	0.050874	0.049175
4	0.059551	0.040744	0.0330504	0.028809	0.0277387	0.0246508	0.02602	0.0230497	0.024365	0.022764	0.02326
5	0.058412	0.033855	0.0273202	0.020488	0.0197492	0.0159032	0.0157553	0.0175922	0.017409	0.018637	0.017213
6	0.047142	0.022637	0.0144317	0.012487	0.0101558	0.0085364	0.0087996	0.0078859	0.007801	0.008403	0.007953
7	0.042437	0.016859	0.0111247	0.007522	0.0068519	0.006897	0.0065752	0.0066396	0.006932	0.0061	0.006548
8	0.03607	0.010988	0.00639	0.006234	0.0060166	0.0062462	0.0057491	0.0057419	0.005817	0.005916	0.005981
9	0.031333	0.009837	0.0063548	0.006814	0.006063	0.0062148	0.0057282	0.0058886	0.005624	0.006135	0.005695
10	0.028587	0.007693	0.0053546	0.005675	0.0054526	0.0052946	0.0058317	0.005928	0.005252	0.006005	0.005472
11	0.024527	0.007023	0.0059792	0.005584	0.0047081	0.0050536	0.0045293	0.0046344	0.004822	0.004739	0.004734
12	0.021565	0.005017	0.0043499	0.004983	0.0043594	0.0035467	0.0040322	0.003864	0.003912	0.003553	0.003255
13	0.01824	0.000668	0.0002781	0.000385	0.0001068	0.0001114	0.0001191	0.0001461	0.000128	0.000106	0.000115
14	0.014814	0.000787	5.93E-05	0.001447	7.04E-05	7.56E-05	8.51E-05	7.58E-05	6.74E-05	6.77E-05	7.09E-05
15	0.016596	0.000413	8.31E-05	0.002052	3.68E-05	8.25E-05	5.40E-05	5.59E-05	5.78E-05	5.02E-05	7.95E-05

المصدر: بالاعتماد على مخرجات برنامج التحليل الاحصائي (R).

يتكون النموذج الأفضل من أربعون وحدة معالجة (عقدة مخفية) في الطبقة المخفية، بالإضافة إلى إثني عشرة فجوة زمنية حيث كانت قيمة معيار جذر متوسطات مربعات الخطأ للتدريب Train RMSE تساوي 0.003864، وبالتالي فإن النموذج الأمثل هو NNAR(12,40)، وعليه فإن طبقة المدخلات تحتوي على 1 عصبون، أما الطبقة المخفية تحتوي على 8 عصبون، أما طبقة المخرجات فإنها تحتوي على واحد عصبون.

تم حساب عدد الأوزان **Weights number** لنموذج الشبكة العصبية الأمثل NNAR(12,40)، كما موضح في الجدول 02 السابق، حيث نلاحظ أن عدد أوزان الشبكة المفضلة NNAR(12,40) يساوي 25 وهذا ما يتطابق مع معادلة FARAWAY & CHATFIELD التالية¹³:

$$P = (X_i + 2) X_u + 1$$

حيث أن: X_i تمثل عدد متغيرات الإدخال

X_u تمثل عدد العصبونات أو عدد الوحدات المعالجة في الطبقة المخفية

P تمثل عدد الأوزان في الشبكة العصبية.

مرحلة التنبؤ

تم تمثيل السلسلة الزمنية لتنبؤات عينة الاختبار سنة 2018 ولسنتي 2019 و 2020 في الجدول التالي، والتي أظهرت نتائج تقاربية مع سلوك السلسلة الأصلية.

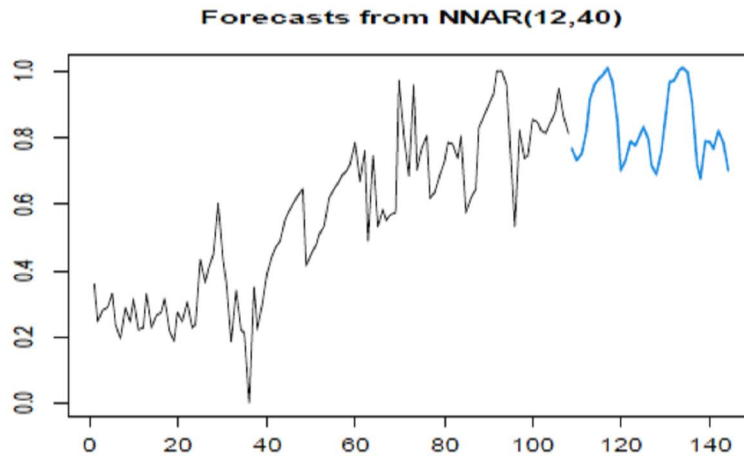
الجدول 03 - تنبؤات NNAR(12,40) لسنة الاختبار 2018 ولسنتي 2019 و 2020

القيمة المتنبأ بها prededction	التاريخ Year-Month	القيمة المتنبأ بها prededction	التاريخ Year-Month	القيمة المتنبأ بها prededction	التاريخ -Month Year
150.54255	Jan/2020	153.20628	Jan/2019	136.03281	Jan/2018
154.98234	Feb/2020	163.70733	Feb/2019	136.77444	Feb/2018
135.44995	Mar/2020	140.83519	Mar/2019	124.03264	Mar/2018
157.21379	Apr/2020	185.88302	Apr/2019	132.38852	Apr/2018
155.83428	May/2020	164.54537	May/2019	135.87622	May/2018
154.82585	Jun/2020	170.78826	Jun/2019	131.97694	Jun/2018
150.82540	Jul/2020	162.00504	Jul/2019	139.64955	Jul/2018
127.27854	Aug/2020	157.05552	Aug/2019	133.83963	Aug/2018
135.48976	Sep/2020	159.69590	Sep/2019	144.49696	Sep/2018
141.71399	Oct/2020	164.55897	Oct/2019	159.60846	Oct/2018
134.18633	Nov/2020	151.50154	Nov/2019	148.12030	Nov/2018
122.80931	Dec/2020	129.47946	Dec/2019	157.51616	Dec/2018

المصدر: بالاعتماد على مخرجات برنامج التحليل الإحصائي (R).

كما تم تمثيل السلسلة الزمنية لتنبؤات عينة الاختبار سنة 2018 والتي لدينا معلومات مسبقة عنها، وللسنتين 2019 و 2020 والموضحة في الشكل 04:

**الشكل 04- منحني القيم الحقيقية مقابل القيم التنبؤية لنموذج NNAR(12,40)
لسلسلة عينة الاختبار 2018 ولسنتي 2019 و2020**



قياس دقة التنبؤ

تعتبر هذه المرحلة ذات أهمية في تقييم النموذج للأغراض التنبؤية، ولتحديد مقدار الخطأ في التنبؤ بين القيم الحقيقية والقيم المتنبأ بها باستخدام نموذج NNAR(12,40) تم استخدام مقاييس دقة التنبؤ MAE, MSE, RMSE, MAPE لمجموعة الاختبار للبيانات خارج العينة ومن أهم المعايير المستخدمة في المفاضلة بين مجموعة النماذج وكانت النتائج كما في الجدول التالي:

الجدول 04 - معايير دقة التنبؤ لمجموعة الاختبار لنموذج NNAR (12,40)

النماذج	MAE	MSE	RMSE	MAPE
NNAR(12,40)	10.2853	127.2559	11.28077	6.622802

المصدر: بالاعتماد على مخرجات برنامج التحليل الإحصائي (R).

خاتمة:

بعد الانتهاء من كافة المجرىات العملية لتحليل بيانات السلسلة الزمنية الشهرية لمرضى السرطان بمحافظة غزة، يتناول هذا الفصل أهم الاستنتاجات التي توصلت إليها الدراسة وأهم التوصيات والآفاق المستقبلية لها.

أثارت الدراسة مجموعة من التساؤلات ومجموعة من الفرضيات وتوصلت إلى الاستنتاجات:

- تم بناء نموذج الشبكات العصبية، باستخدام شبكة البيرسبترون متعدد الطبقات (MLP)، وتبين أن نموذج NNAR (12,40) هو الأمثل، وذلك لامتلاكه أقل فرق مطلق للمعيار الاحصائي (RMSE) وقيمته 0.003864؛
- تم استخدام النموذج NNAR (12,40) للتنبؤ بعينة الاختبار لـ 12 قيمة مستقبلية وللسنتين 2019 و 2020، وأظهرت القيم التنبؤية لعينة الاختبار بأنها متقاربة من القيم الأصلية للسلسلة وأن هناك تزايداً مضطرباً في إصابات مرضى السرطان؛
- النموذج الأمثل هو NNAR(12,40)، وأن طبقة المدخلات تحتوي على 1 عصبون، الطبقة المخفية على 8 عصبون، وطبقة المخرجات على 1 عصبون؛
- ونلاحظ أن عدد أوزان الشبكة المفضلة NNAR(12,40) يساوي 25، وهذا ما يتطابق مع معادلة وفقاً لمعادلة FARAWAY & CHATFIELD (1998)، مما يؤكد لنا بأن النموذج NNAR(12,40) هو الأمثل.

بناءً على نتائج التحليل، توصلنا إلى مجموعة من التوصيات:

- ضرورة إنشاء مستشفى متخصص في الأمراض السرطانية بمحافظة غزة وتجهيزه بأفضل المعدات الحديثة للكشف المبكر عن هذه الأمراض، والمستلزمات العلاجية له، مع توفير كادر طبي بشري ذو كفاءة عالية للحد من هذه الظاهرة؛
- مطالبة السلطة الوطنية الفلسطينية بتقديم لائحة اتهام ضد الاحتلال الصهيوني، وكذلك لمنظمة الصحة العالمية ومحكمة الجرائم الدولية على جرائم الاحتلال الصهيوني المرتكبة بحق الشعب الفلسطيني بعد استخدامها الأسلحة المحرمة دولياً في الحروب التي شنتها على محافظات غزة في 2008 و 2009 و 2012 و 2014 و 2021؛
- وتشكيل لجنة رقابية وطنية على القطاع الزراعي بسبب عدم الاستخدام الأمثل للمبيدات الزراعية من قبل المزارعين؛
- وضرورة استخدام التكنولوجيا الحديثة والأمن في شبكات الاتصالات الثابتة والخلوية، وتجنب تركيب محطات الهاتف النقال وأبراج تقوية الإرسال داخل التجمعات السكنية.

الهوامش والمراجع:

- 1 عيسى علام (2000)، «الشبكات العصبية الاصطناعية - البنية الهندسية - الخوارزميات - التطبيقات»، شعاع للنشر والعلوم، حلب.
- 2 وزارة الصحة الفلسطينية، (2018)، «سلسلة إصابات مرضى السرطان الشهرية في الفترة (2009-2018)»، وحدة نظم المعلومات الصحية، وزارة الصحة الفلسطينية، غزة.
- 3 ارجع إلى:
- طاقية البيومي & الوصيفي الشيماء (2012)، «التنبؤ باستخدام الدمج بين الشبكات العصبية الاصطناعية ونماذج بوكس وجينكنز - دراسة تطبيقية»، جامعة المنصورة، الدقهلية، ص ص 1-22؛
- العباسي عبد الحميد محمد (2013)، «مقدمة في الشبكات العصبية الاصطناعية وتطبيقاتها في العلوم الاجتماعية باستخدام SPSS»، معهد الدراسات والبحوث الإحصائية، جامعة القاهرة، ص ص 1-39.
- 4 B. J. A. KROSE & Patrick Van DER SMAGT (1996), «An Introduction to neural networks», Journal of Computer Science, pp. 1-135,
https://www.researchgate.net/profile/Patrick-Van-Der-Smagt/publication/272832321_An_introduction_to_neural_networks/links/5808737a08ae63c48fec7f2f/An-introduction-to-neural-networks.pdf
- 5 ارجع إلى:
- عمر صابر قاس & إسراء رستم محمد (2012)، «دراسة رياضية تحليلية لخوارزميات الشبكات العصبية الاصطناعية في ملائمة نموذج للتشخيص الطبي»، مجلة الراافدين لعلوم الحاسبات والرياضيات، كلية علوم الحاسبات والرياضيات، جامعة الموصل، الموصل، المجلد 10، العدد 01، ص ص 183-192؛
- على بشار الشريف & عبد الغنى على جمعة (2012)، «تطبيقات على الشبكات العصبية الاصطناعية»، جامعة تشرين، اللاذقية، ص ص 6-13؛
- عائدة يونس محمد المراد (2011)، «مقارنة بين الانحدار الكلاسيكي والشبكات العصبية الاصطناعية في التنبؤ بمستويات نتائج بحوث طلبة كلية التربية الرياضية»، المجلة العراقية للعلوم الإحصائية، كلية علوم الحاسبات والرياضيات، جامعة الموصل، الموصل، العدد 21، ص ص 286-303؛
- يوسف صوار، دياب زقاي & طاوش قندوسي (2012)، «تقنية الشبكات العصبية الاصطناعية كأحد أساليب ذكاء الأعمال لتسيير مخاطر القروض»، المؤتمر العلمي السنوي الحادي عشر حول «ذكاء الأعمال واقتصاد المعرفة»، جامعة الزيتونية الأردنية، عمان، أيام 23-26 أبريل 2012 [كتاب المؤتمر، ص ص 88-107]؛
- العباسي عبد الحميد محمد، مرجع سبق ذكره؛
- Simon HAYKIN (1994), «Neural networks: A comprehensive foundation», Macmillan, New York city, USA.

6 Simon HAYKIN, op. cit.

7 ارجع إلى:

- رياض قادري (2017)، «مقارنة رياضية وقياسية للتنبؤ بالمبيعات - شركة سونلغاز من الكهرباء المنخفض التوتر لسنة 2016 باستخدام نماذج الشبكات العصبية»، أطروحة دكتوراه في العلوم التجارية، تخصص تسويق، كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير، جامعة تلمسان، ص 151؛
- على بشار الشريف & عبد الغنى على جمعة، مرجع سبق ذكره.

8 ارجع إلى:

- على بشار الشريف & عبد الغنى على جمعة، مرجع سبق ذكره؛
- Andries ENGELBRECHT (2002), «Computational intelligence: An introduction», Wiley & Sons, USA.

9 Simon HAYKIN, op. cit.

10 ارجع إلى:

- عمر صابر قاس & إسراء رستم محمد، مرجع سبق ذكره؛
- رياض قادري، مرجع سبق ذكره؛
- عائدة يونس محمد المراد، مرجع سبق ذكره؛
- العباسي عبد الحميد محمد، مرجع سبق ذكره؛
- Sandhya SAMARASINGHA (2007), «Neural networks for applied sciences and engineering», Tayler & Francis Group, LLC., NY., United Kingdom;
- B. J. A. KROSE & Patrick Van DER SMAGT, op. cit.

11 Simon HAYKIN, op. cit.

12 ارجع إلى:

- يوسف صوار، دياب زقاي & طاوش قندوسي، مرجع سبق ذكره؛
- العباسي عبد الحميد محمد، مرجع سبق ذكره؛
- Simon HAYKIN, op. cit.;
- Andries ENGELBRECHT, op. cit.

13 Julian FARAWAY & Chris CHATFIELD (1998), «Time series forecasting with neural networks: A comparative study using the airline data», Journal of the Royal Statistical Society, United Kingdom, Appl. Ststist. n° 47, Part2, pp. 231-250, <https://people.bath.ac.uk/jjf23/papers/neural/nnts.pdf>