



Study of survival data using artificial neural networks (ANN) and Kaplan-Meyer metho

*دراسة بيانات البقاء باستعمال الشبكات العصبية الاصطناعية (ANN) وطريقة كابلان – مايير

مرتضى فرحان حويل **م.د. شروق عبد الرضا سعيد

Abstract: Through this research the method of artificial neural networks was briefly presented as well as Kaplan-Mayer method in the theoretical side after presenting a general methodology of the research, which is the introduction, objective, problem and methodology of the research 2008 - 2018 from the Center for Genetic Hematology / Dhi Qar Health Department and then Kaplan-Mayer method to determine the impact of each of the variables in the survival period, while the last data analysis is the analysis of data in the manner of neural networks in the first two cases included On all the variables of the study and the second case included the moral variables derived from the Kaplan - Mayer method if the neural network in the second case showed preference over those included in the analysis of all variables, and this proved the importance of using the Kaplan - Mayer method before proceeding with the analysis.

المستخلص : من خلال هذا البحث تم طرح اسلوب الشبكات العصبية الاصطناعية بصورة موجزة فضلا عن طريقة كابلان – مايير في الجانب النظري بعد عرض منهجية عامة للبحث تتمثل بالمقدمة وهدف ومشكلة البحث ومنهجيته , اما الجانب التطبيقي فيمثل ملخص لبيانات البحث المتمثلة ببيانات المرضى المصابين بمرض التلاسيميا الكبرى لفترة التسجيل ٢٠٠٨- ٢٠١٨ والمستحثة من مركز امراض الدم الوراثية / دائرة صحة ذي قار ثم طريقة كابلان – مايير لتحديد مدى تاثير كل متغير من المتغيرات في فترة البقاء , اما اخر عمليات تحليل البيانات فتتمثل تحليل البيانات باسلوب الشبكات العصبية بحالتين الاولى منهما اشتملت على جميع متغيرات الدراسة والحالة الثانية اشتملت على المتغيرات المعنوية المستخلصة عن طريقة كابلان – مايير اذا اظهرت الشبكة العصبية في الحالة الثانية

*بحث مستل
*جامعه كربلاء /كلية الادارة والاقتصاد

افضليتها على تلك المشتملة في تحليلها على جميع المتغيرات , وهذا ما اثبت اهمية استعمال طريقة – كابلان ماير قبل الشروع في عمليات التحليل .

منهجية البحث

المقدمة

يعد هذا البحث اضافة للدراسات السابقة المتعلقة ببيانات البقاء والشبكات العصبية الاصطناعية اذ تشكل بيانات البقاء مجموعة واسعة للبيانات التي تتعلق بظواهر وحالات معينة يُهتم من خلالها بدراسة الوقت السابق لحدوث حدث معين وتعتمد اعتماد كلي في تحليل البيانات الناتجة عن تلك الابحاث على الاساليب الاحصائية التي تهتم بدراسة بيانات البقاء كجداول الحياة وانموذج كوكس (cox) والتوزيع الاسي (Exponential Distribution) وانموذج كومبيرتز (Gompertz Model) وغيرها, ومع التطور الحاصل في اساليب تحليل البيانات امست الحاجة ضرورية لتطوير تلك الاساليب او اضافة اخرى تكاد تكون جديد على الاقل لنوع البيانات المستعمل , وهذا ما دفعنا الى استعمال اسلوب لتحليل البيانات (اسلوب الشبكات العصبية الاصطناعية (ANN)) غير تقليدي لهذا النوع من البيانات (بيانات البقاء) اذ ان بيانات البقاء لها نماذج وطرق تحليل خاصة كما اسلفنا , اما فيما الشبكات العصبية الاصطناعية (ANN) التي تعد احدى نتاجات علم الذكاء الاصطناعي , ذلك الجزء من علم الذكاء الصناعي الذي يحاكي الذكاء الطبيعي (الدماغ البشري) في طريقة عمله باعتماد البرمجة الحاسوبية كي يتم الاعتماد عليه في مهمات معينة بديلاً للذكاء الطبيعي واصبحت من الاساليب المهمة في تمثيل ووصف واعطاء فكرة او تحديد حل مشكلة ما او غير ذلك من المسائل التي تحتاج الى تحليل مفصل للبيانات .

هدف البحث

دراسة بيانات البقاء باستعمال اسلوب الشبكات العصبية الاصطناعية لتحديد اهم العوامل المؤثرة في فترة البقاء , والمفاضلة بين المعماريات الهيكلية المختلفة للشبكة العصبية لاختيار افضلها لتحليل البيانات قيد الدراسة , كذلك معرفة مدى اهمية او افضلية استعمال طريقة كابلان – ماير عند اختبار معنوية العوال المؤثرة في فترة البقاء قبل استعمال الشبكات العصبية الاصطناعية .

مشكلة البحث

تكمن مشكلة البحث في مدى توفيق اسلوب الشبكات العصبية الاصطناعية لدراسة بيانات البقاء باعتبارها اسلوب غير تقليدي لدراسة هكذا نوع من البيانات .

تم العمل وفق المنهج الاستقرائي في هذا البحث وفيه نبدأ بملاحظة المشكلات ثم وضع الفروض الخاصة بها وبعدها يتم اختبارها باستعمال الاساليب الاحصائية .

الجانب النظري

الشبكات العصبية الاصطناعية (ANN): [9],[6]

ان نموذج الشبكة العصبية الاصطناعية ممثل بالادراك الحسي (perceptron) الذي يمثل محاكاة للخلية العصبية الطبيعية في طريقة عملها اقترح من قبل العالم (Frank Rosenblatt) والمتكون من وحدات تمثل مدخلات معينة تشابه الاشارات الداخلة الى الخلية العصبية الطبيعية , ويتم ربط المدخلات بوحدة جمع بواسطة اوزان معينة تماثل في عملها الاشتباك العصبي الحيوي , ثم مخرجات وحدة الجمع التي يتم معالجتها خلال دالة معينة

اذ تعرف الشبكة العصبية الاصطناعية على انها نظام حاسوبي يتكون من عدد محدد من وحدات المعالجة (عصبونات) (Processing Elements) المتشابهة فيما بينها وتتصف بطبيعتها الديناميكية في معالجة البيانات المدخلة اليها .

اذ يمكن تصنيفها الى انواع بحسب عدد الطبقات المكونة لها حيث تقسم الى الشبكة وحيدة الطبقة (Single Perceptions) والشبكة متعددة الطبقات (Mullt Perceptions) , او تصنيفها بحسب اتجاه تدفق المعلومات فيمكن من خلال ذلك تصنيفها الى شبكات التغذية الامامية (Feed forward Net.) وهنا يتم تدفق المعلومات فيها من الخلف الى الامام مروراً بالطبقات دون السماح لها بالعودة العكسية , وشبكات التغذية الخلفية (Feed backward Net.) والتي تسمح بتدفق المعلومات من الخلف الى الامام خلال الطبقات مع السماح للمعلومات بالعودة العكسية اي الى الطبقات او الطبقة السابقة .

مميزات الشبكات العصبية الاصطناعية : [1],[5]

الشبكات العصبية الاصطناعية تتميز بالعديد من المميزات اهمها :

١- تستند الى اسس رياضية قوية .

٢- امكانية عملها على البيانات الكمية والنوعية على حد سواء.

٣- قدرتها على تخزين المعرفة المكتسبة من البيانات السابقة لنفس الظاهرة قيد الدراسة وتكيفها مع البيانات الجديدة .

٤- مرونة تتيح للمستخدمين امكانية استخدامها في كافة المجالات .

من المفاهيم الأساسية للشبكات العصبية الاصطناعية ان الاشارة الخارجة من اي خلية عصبية تعتمد على المدخلات الى تلك الخلية وان مفهوم التدريب المعتمد في الشبكات معتمداً في الاساس على تقليل مقدار الخطأ , حيث ان الشبكة تستمر في تحديث اوزانها لحين الوصول الى اقل مقدار للخطأ الذي يحقق افضل هيكلية معمارية ممكن , هذا و تجري عملية التعلم تلك باستعمال خوارزميات مختصة تسمى خوارزميات التعلم والتي تتولى عملية التدريب اذ تعمل على تعديل اوزان الشبكة للوصول الى اقل حد خطأ ممكن وبالتالي تحسين اداء الشبكة ككل , ان الهدف الاساسي من خوارزميات التعلم هو تحديد الاوزان المثلى , وهناك طريقتين للتدريب او التعلم هي :

١-التعلم المراقب :

Controlled Learning:

ويتم هذا الاسلوب للتعلم بإدخال بيانات التدريب للشبكة العصبية بشكل نمطين هما نمط المدخلات ونمط المخرجات وكل منها يرتبط بمتجه المخرجات المستهدفة , فتعتمد الشبكة في هذه الحالة من التعلم الفرق بين مخرجات الشبكة والمخرجات الحقيقية (نمط المخرجات) لتحديد اوزان الشبكة .

٢- التعلم غير المراقب:

Unattended Learning:

ان اسلوب التعلم غير المراقب يتم بإدخال بيانات التدريب للشبكة العصبية بشكل نمط واحد فقط نمط المدخلات , فتعتمد هذه الطريقة (الاسلوب) على اساس قدرة الشبكة في اكتشاف ملامح نمط المخرجات التي يتم تعميمها على باقي المخرجات , اي بدون سابق معرفة بقيمة المخرجات الحقيقية وهذا عكس الاسلوب السابق (التعلم المراقب) اذ يستند هذا الاسلوب من التعلم على الارتباطات بين قيم الادخال .

تشغيل البيانات في الشبكة العصبية الاصطناعية : [6],[10]

ان تشغيل المعلومات في الشبكة العصبية الاصطناعية يتطلب التعرف على المفاهيم الاتية :

- الطبقات (Layers) : وتمثل عدد الطبقات المكونة للشبكة العصبية , اي طبقة الادخال (Input Layer) وطبقة الاخراج (Output Layer) وعدد الطبقات المخفية فيما بينهما.
- طبقة المدخلات (Input Layer): وتمثل المتغيرات المستقلة وهو المستوى الاول للشبكة العصبية .
- الطبقة المخفية (Hidden Layer): وتأتي بعد طبقة المدخلات من حيث الترتيب داخل الشبكة وقد تكون بمستوى واحد او اكثر حسب نوع الشبكة , والتي تربط بين طبقة المدخلات من جهة وطبقة المخرجات من جهة اخرى.
- طبقة المخرجات (Output Layer): وتمثل المستوى الاخير للشبكة حيث مخرجات الشبكة النهائية .

• الاوزان (Weights) : يمكن من خلالها تحديد مدى قوة ترابط بين الطبقات , ان كل عقدة ترتبط مع الطبقة السابقة لها بوزن ومن جهة اخرى ترتبط مع الطبقة اللاحقة بوزن معين اخر , وفي العموم فان الشبكات العصبية لها ثلاثة مستويات من الاوزان المستوى الاول مستوى الاوزان بين طبقة الادخال والطبقة المخفية (Input to Hidden Weight) والمستوى الثاني للاوزان بين الطبقات المخفية (Weight layer between hidden levels) (في حالة اكثر من طبقة مخفية واحدة) , والمستوى الثالث مستوى الاوزان بين الطبقة المخفية وطبقة الاخراج (Hidden to output weight).

• دالة التحويل (Transformation Function): ان دالة التحويل تعتمد في الاساس على الدالة التجميعية $Y = \sum_{i=1}^n X_i W_{ij}$ والتي تمثل دالة التحفيز الداخلية ويطلق عليها دالة التنشيط , وتعتبر دالة التحويل اللوجستية الدالة الاكثر استخداماً في معظم الدارسات والابحاث وتأخذ الدالة الصورة الاتية:

$$Y_x = \frac{1}{1 + e^{-y}} \dots \dots \dots (1)$$

Y_x : تمثل قيمة اشارات الدخل الموزونة للطبقة X .

X_i : تمثل قيم المدخلات للشبكة .

W_{ij} : تمثل الاوزان .

٢-١-٤- خوارزمية الانتشار العكسي للخطأ:^[2]

ان خطوة الاساسية المعتمدة وفق هذه الخوارزمية تتم بتحديد اوزان (طبقة المدخلات – الطبقة المخفية) ثم تحديد اوزان (الطبقة المخفية – طبقة المخرجات) , بعدها يتم تعديل الاوزان اعتماداً على مقدار الخطأ للوصول الى الاوزان المثلى للشبكة , اما قيم الاوزان الاولية في بداية عملية التدريب تؤخذ بشكل ارقام تولد عشوائياً من احد التوزيعات الاحصائية ان استعمال خوارزمية الانتشار العكسي لتدريب الشبكة العصبية يتضمن ثلاث مراحل هي :

- مرحلة الانتشار الامامي للخطأ .
- مرحلة الانتشار الخلفي للخطأ .
- مرحلة توليف اوزان الشبكة .

ويمكن ايجاز هذه المنهجية او الخوارزمية بالخطوات التالية :

١- توليد قيم اوزان اولية للشبكة من احد التوزيعات الاحصائية .

٢- استقبال كل عقدة من عقد طبقة المدخلات اشارة دخلها ويتم ارسال هذه الاشارة الى كل عقدة من عقد الطبقة المخفية , حيث تقوم كل عقدة من عقد الطبقة المخفية بتجميع الاشارات الموزونة الداخلة لها وفق المعادلة :

$$O_j = \frac{1}{1 + e^{-\sum_{i=1}^n X_i W_{ij}}} \dots \dots \dots (2)$$

٣- تقدير قيم مخرجات الطبقة المخفية من خلال تطبيق تابع التنشيط , وترسل بعد ذلك الى عقد طبقة المخرجات .

٤- تجمع اشارات الدخل الموزونة عند كل عقدة من عقد المخرجات وفق المعادلة :

$$I_k = \frac{2}{1 + e^{-\sum_{j=1}^n O_j W_{jk}}} \dots \dots \dots (3)$$

٥- حساب مقدار الخطأ لعقدة الاخراج يتم ذلك بحساب الفرق بين قيمة مخرجات العقدة (I_k) والقيمة الحقيقية للعقدة (r_k) وفق المعادلة :

$$E_k = r_k - I_k \dots \dots \dots (4)$$

يتم بعد ذلك مقارنة مخرجات الشبكة العصبية النهائية مع القيم الحقيقية لتقدير الخطأ وفق المعادلة :

$$E_k = (r_k - I_k)f(t) \dots \dots \dots (5)$$

$f(t)$: تمثل دالة لوجستية في حالة عقد المخرجات غير خطية وتساوي واحد في حالة عقد المخرجات خطية .

٦- يتم حساب التغير في مقدار الخطأ وفق المعادلة :

$$\Delta U_{jk} = \alpha O_j E_k \dots \dots \dots (6)$$

وتجمع كل عقدة من عقد الطبقة المخفية اشارات المدخلات الموزونة الى (E_k) وفق المعادلة :

$$\Delta_j = \sum_{k=1}^n E_k U_{jk} \dots \dots \dots (7)$$

ثم يتم حساب (E_j) بضرب قيمة المعادلة (7) بتابع التنشيط , ليتم بعدها حساب التغير في حجم الخطأ وفق المعادلة :

$$\Delta U_{ij} = \alpha E_j X_i \dots \dots \dots (8)$$

٧- يتم تحديد الازان الجديدة لكل عقدة من عقد طبقة المخرجات وفق المعادلة :

$$W_{ik}(\text{new}) = W_{ik}(\text{old}) + \Delta U_{ik} \dots \dots \dots (9)$$

كذلك يتم تحديد الازان الجديدة لكل عقدة من عقد الطبقة المخفية وفق المعادلة :

$$W_{ij}(\text{new}) = W_{ij}(\text{old}) + \Delta U_{ij} \dots \dots \dots (10)$$

وتستمر الشبكة في هذا النهج اي تحديث الازان خلال عملية التدريب لحين الوصول الى الازان المثلى بالتالي الوصول الى قيم مخرجات الشبكة العصبية المثلى (افضل توفيق للأنموذج قيد البحث) .

Kaplan & Meier : طريقة كابلان – مايير : [8],[4]

Method

تم اقتراح هذه الطريقة من قبل العالمان (Kaplan & Meier) عام (١٩٥٨م) وتدعى ايضا بتقدير كابلان – مايير او مقدر الضرب الحدي , وتعرف على انها احتمال البقاء للمفردة لمدة معينة من الزمن بمراعاة كون الزمن بفترات متعددة , اذ تستعمل لتحليل بيانات البقاء بصورة فردية لكل متغير من متغيرات الدراسة مع المتغير التابع

والذي يمثل الزمن في حال كون البيانات تحوي اكثر من متغير لتعطينا فهم اوسع عن مدى الاختلاف بين دوال البقاء لمجموعتين او اكثر للمتغير الحدي ويساهم في تقديرها جميع مفردات الدراسة سواء أكانت مراقبة ام غير مراقبة وهذا ما يميزها عن طرق تحليل بيانات البقاء (Survival Data) الاخرى كجداول الحياة وانموذج انحدار كوكس (Cox) التي تستعمل المتغيرات المستقلة مجتمعة لتحليل البيانات ويستعمل في هذه الطريقة الاختبارات اللامعلمية كاختبار ولكوكسن واختبار لوغاريتم الرتبة (Log-Rank) .

٣- الجانب التطبيقي :

٣-١- البيانات :

تمثل البيانات قيد الدراسة عينة عشوائية بحجم (١٨٦) للمصابين بمرض الثلاثيميا الكبرى للفترة (٢٠٠٨-٢٠١٨)، وتم جمع البيانات اعتماداً على السجلات المرضية والطبقات لكل حالة مرضية في مركز امراض الدم الوراثية / دائرة صحة ذي قار، واشتملت البيانات الخام على ١٦ متغير ، ١٤ متغيراً منها مستقلاً ومتغيراً تابعاً واحداً ومتغير الحالة (Case)، والجدول التالي يمثل الوصف المفصل للبيانات :

جدول رقم (١) بيانات الدراسة .

المتغير	المتغير	تغير	
	المعتمد		
الفترة الزمنية ما بين تاريخ تشخيص الحالة المرضية وتاريخ الوفاة او المشاهدة مقاسة بعدد الايام .	زمن بقاء المريض على قيد الحياة		
المتغيرات المستقلة			
(A+) = 1 (B+) = 2 (AB+) = 3 (O+) = 4 (A-) = 5 (B-) = 6 (AB-) = 7 (O-) = 8	فصيلة الدم	X ₁	
-١ ذكر -٢ انثى		X ₂	
الفترة الزمنية ما بين تاريخ الولادة وتاريخ تشخيص الحالة المرضية مقاساً بعدد الايام.	سند تشخيص الحالة	X ₃	
=١ الحضر =٢ القرى والارياف	سكن	X ₄	

1 = توجد اصابات سابقة في العائلة 2 = لا توجد اصابات سابقة في العائلة	عائلة الطبي	X ₅	
1 = الالاء اقرباء 2 = الالاء غير اقرباء	ربى للأوبى	X ₆	
1 = اعزب 2 = متزوج / متزوجة 3 = غير ذلك	لاجتماعية	X ₇	
1 = موظف , 3 = متقاعد 2 = كاسب , 4 = عاطل عن العمل , ربة بيت 5 = طالب		X ₈	
1 = مصاب 2 = غير مصاب	بأمراض الطحال	X ₉	
1 = مصاب 2 = غير مصاب	بداء السكري	X ₁₀	
1 = مصاب 2 = غير مصاب	بأمراض القلب	X ₁₁	
1 = مصاب 2 = غير مصاب	بمرض	X ₁₂	
1 = مصاب 2 = غير مصاب	بوهن العظم	X ₁₃	
1 = مصاب 2 = غير مصاب	بالتهاب الكبد	X ₁₄	
0 = فقدان المشاهدة (عدم معرفة الحالة النهائية للمشاهدة) 1 = عدم فقدان المشاهدة (حدوث حدث الموت للمشاهدة)	نهائية للمشاهدة		

٢-٣- نتائج طريقة كابلان-مايبر:

تم استعمال طريقة كابلان-مايبر لتحليل البيانات بصورة فردية لكل متغير من المتغيرات المذكورة في الجدول (1) مع المتغير التابع الذي يمثل زمن البقاء على قيد الحياة , لمعرفة اي من المتغيرات المستقلة له تأثير على زمن

البقاء قبل الاسهاب في طريقة التحليل الاساسية للدراسة والمتمثلة بالشبكات العصبية الاصطناعية والجدول التالي يبين مدى معنوية كل متغير من متغيرات الدراسة .

جدول (٢) نتائج تحليل طريقة كابلان – ماير لجميع المتغيرات .

P-Value	Chi-Square	النسبة	لمشاهدات المراقبة	مشاهدات الحدث	المتغير	
٠.١١٧	١١.٥٤١	%38	٢٦	٤٤	A+	فصيلة الدم
		%16	٧	٢٣	B+	
		%7	٥	٨	AB+	
		%31	١٢	٤٥	O+	
		%4	١	٦	A-	
		%2	١	٢	B-	
		%1	٠	١	AB-	
		%3	٢	٣	O-	
		%38	٢٦	٤٤	A+	
0.364	0.٨٢٤	%٥٤	٣٢	٦٩	ذكر	جنس المريض
		%٤٦	٢٢	٦٣	انثى	
0.462	0.541	%39	15	57	الحضر	محل السكن
		%61	39	75	الريف	
0.000	57.604	%58	26	82	العائلة لديها اصابات سابقة	تاريخ العائلة الطبي
		%42	28	50	العائلة ليس لديها اصابات سابقة	
٠.٣٩٥	٠.٧٢٤	%76	٤٣	٩٩	الابوين اقرباء	صلة القرى للابوين
		%24	١١	٣٣	لابوين غير اقرباء	
٠.٢٦٨	٢.٦٣٤	%87	٥٣	١٠٩	اعزب	الحالة الاجتماعية
		%11	١	١٩	متزوج / متزوجة	
		%2	٠	٤	غير ذلك	
0.757	٠.٠٩٦	%٢	0	4	موظف	العمل
		%٩٨	54	128	طل عن العمل/ربة	
٠.٠٠٠	٦٢.٠٩١	%51	٣٥	٦٠	مصاب	الاصابة بأمراض الطحال
		%49	١٩	٧٢	غير مصاب	
٠.٠٠٠	٧٣.٥٨٣	%52	٤١	٥٦	مصاب	الاصابة بأمراض السكري
		%48	١٣	٧٦	غير مصاب	
٠.٥٨٣	٠.٣٠٢	%16	٥	٢٤	مصاب	الاصابة بأمراض القلب
		%84	٤٩	١٠٨	غير مصاب	

٠.٠٠٠	٥٢.٠٠٤	%52	٣٩	57	مصاب	الاصابة بتاخر النمو
		%48	١٥	٧٥	غير مصاب	
٠.٠٠٠	١٩.٤٧٣	%38	٣٠	٤١	مصاب	الاصابة بنخر او وهن العظم
		%62	٢٤	٩١	غير مصاب	
٠.٠٠٠	٦٣.٧٣٠	%47	٣٠	٥٨	مصاب	الاصابة بالتهاب الكبد الفيروسي
		%53	٢٤	٧٤	غير مصاب	

من خلال النتائج المثبتة في الجدول السابق يتضح ان المتغيرات (فصيلة الدم , جنس المريض , محل السكن , صلة القربى للابوين , الحالة الاجتماعية , العمل , الاصابة بأمراض القلب) ليس لها تاثير في فترة بقاء المرضى على قيد الحياة , اما المتغيرات (العمر عند اكتشاف الاصابة , تاريخ العائلة الطبي , الاصابة بأمراض الطحال , الاصابة بأمراض السكري , الاصابة بتاخر النمو , الاصابة بنخر او وهن العظم , الاصابة بالتهاب الكبد الفيروسي) فقد تبين ان لها تاثير معنوي في فترة بقاء المريض على قيد الحياة .

٣-٣- الشبكات العصبية الاصطناعية : Artificial Neural Netorks

٣-٣-١- نتائج تطبيق الشبكات العصبية الاصطناعية لجميع متغيرات الدراسة :-

ان تحليل البيانات باستعمال الشبكات العصبية الاصطناعية في هذه الفقرة من الدراسة سيتم باستعمال جميع المتغيرات المستقلة الوارد ذكرها في جدول (١) والبالغ عددها اربعة عشر متغيراً مستقلاً والذي يكون مساوياً لعدد وحدات طبقة المدخلات اضافة الى وحدة التحيز واحدة , وثبات عدد وحدات طبقة المخرجات بوحدتين واختلاف عدد وحدات الطبقة المخفية ونسب مجموعتي التدريب والاختبار وبعده تشكيلات معمارية لكل نسبة من تلك النسب في كل مرة , باعتماد خوارزمية الانتشار العكسي للخطأ ودالة تنشيط اللوجستية ثنائية القطب عند الطبقة المخفية ودالة تنشيط العتبة عند طبقة المخرجات , للتوصل الى افضل معمارية ممكنة للشبكة العصبية باقل خطأ ممكن واقل نسبة تنبؤ زائف لمجموعة الاختبار , والجدول التالي يبين ذلك:

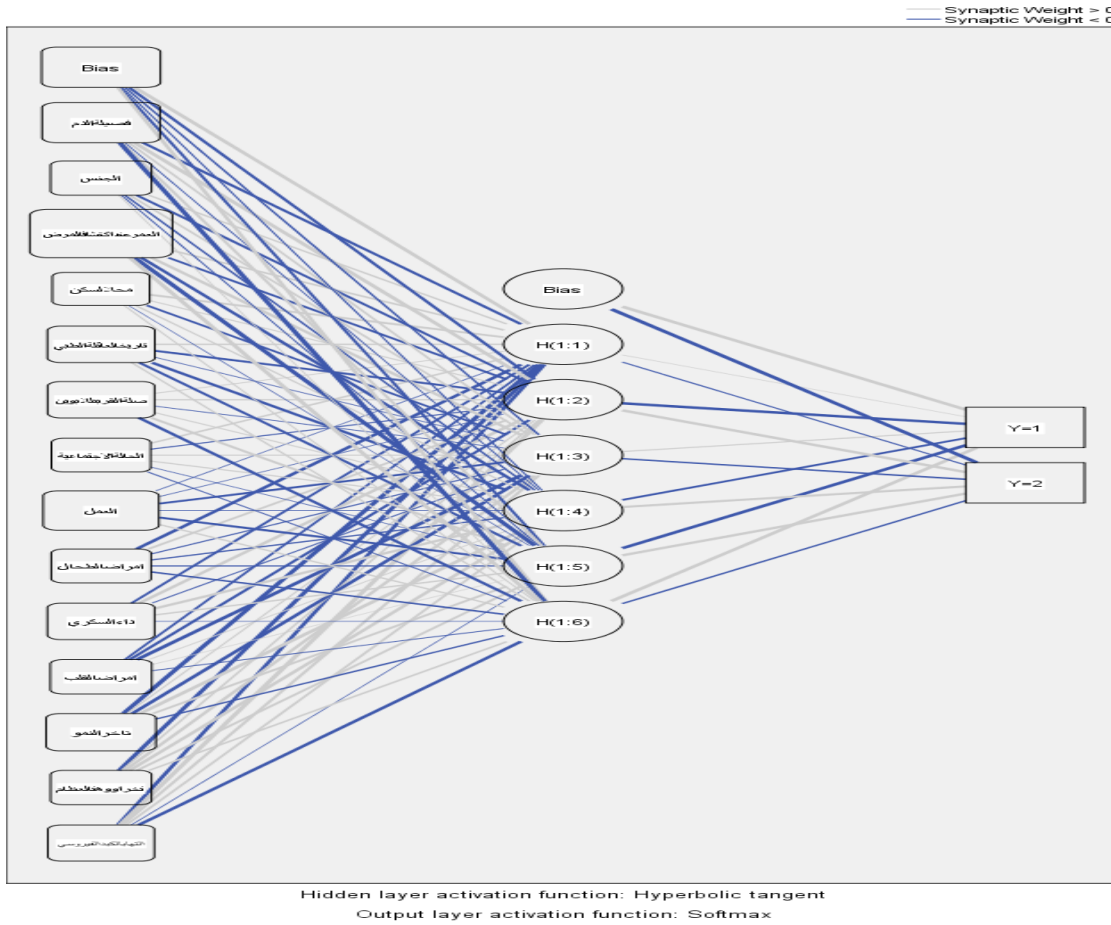
جدول (3) نتائج الشبكات العصبية الاصطناعية .

زمن التدريب	مجموعة الاختبار			مجموعة التدريب		معمارية الشبكة خالات , مخفية , مخرجات	ت
	التنبؤ الزائف	الخطأ	النسبة	الخطأ	النسبة		
ثانيتان	%10.00	6.928	%١٠	7.399	%٩٠	(14,2,2)	١
٨ ثوان	%5.90	5.142	%١٠	5.509	%٩٠	(14,3,2)	٢
٣ ثوان	%11.80	6.043	%١٠	23.29	%٩٠	(14,4,2)	٣
٥ ثوان	%7.70	7.485	%١٠	10.63	%٩٠	(14,5,2)	٤
ثانيتان	%9.10	8.769	%١٠	9.868	%٩٠	(14,6,2)	٥
ثانيتان	%11.60	6.114	%١٠	7.144	%٩٠	(14,7,2)	٦

١١ ثوان	%13.50	7.341	%١٠	24.62	%٩٠	(14,8,2)	٧
ثانيتان	%9.40	4.996	%١٠	9.129	%٩٠	(14,9,2)	٨
٣ ثوان	%20.00	6.737	%١٠	10.19	%٩٠	(14,10,2)	٩
ثانيتان	%18.20	8.437	%١٠	8.68	%٩٠	(14,11,2)	١٠
٣ ثوان	%9.10	4.834	%١٠	6.91	%٩٠	(14,12,2)	١١
ثانيتان	%11.20	5.807	%٢٠	6.345	%٨٠	(14,2,2)	١٢
ثانيتان	%6.30	5.125	%٢٠	10.2	%٨٠	(14,3,2)	١٣
٥ ثوان	%9.70	4.655	%٢٠	61.82	%٨٠	(14,4,2)	١٤
٦ ثوان	%3.80	5.457	%٢٠	27.38	%٨٠	(14,5,2)	١٥
٦ ثوان	%4.50	11.6	%٢٠	6.12	%٨٠	(14,6,2)	١٦
ثانيتان	%7.00	9.919	%٢٠	7.765	%٨٠	(14,7,2)	١٧
ثانيتان	%13.30	8.292	%٢٠	4.513	%٨٠	(14,8,2)	١٨
٣ ثوان	%4.30	9.236	%٢٠	3.76	%٨٠	(14,9,2)	١٩
ثانيتان	%10.00	8.195	%٢٠	8.734	%٨٠	(14,10,2)	٢٠
٣ ثوان	%10.30	8.962	%٢٠	8.523	%٨٠	(14,11,2)	٢١
ثانيتان	%11.40	6.938	%٢٠	7.809	%٨٠	(14,12,2)	٢٢
٣ ثوان	%4.10	5.854	%٣٠	8.828	%٧٠	(14,2,2)	٢٣
٤ ثوان	%3.40	7.473	%٣٠	7.544	%٧٠	(14,3,2)	٢٤
ثانيتان	%6.60	7.283	%٣٠	6.667	%٧٠	(14,4,2)	٢٥
٦ ثوان	%6.60	12.84	%٣٠	11.15	%٧٠	(14,5,2)	٢٦
ثانيتان	%2.20	4.594	%٣٠	9.237	%٧٠	(14,6,2)	٢٧
ثانيتان	%5.80	9.538	%٣٠	4.591	%٧٠	(14,7,2)	٢٨
٥ ثوان	%3.30	4.665	%٣٠	16.6	%٧٠	(14,8,2)	٢٩
ثانيتان	%2.90	7.322	%٣٠	4.63	%٧٠	(14,9,2)	٣٠
ثانيتان	%6.10	4.95	%٣٠	7.59	%٧٠	(14,10,2)	٣١
٣ ثوان	%6.40	5.335	%٣٠	6.823	%٧٠	(14,11,2)	٣٢
ثانيتان	%7.40	7.613	%٣٠	8.266	%٧٠	(14,12,2)	٣٣
٤ ثوان	%2.70	6.237	%٤٠	6.893	%٦٠	(14,2,2)	٣٤
٣ ثوان	%10.10	25.26	%٤٠	7.236	%٦٠	(14,3,2)	٣٥
ثانيتان	%4.80	9.128	%٤٠	6.22	%٦٠	(14,4,2)	٣٦
٣ ثوان	%8.30	15.92	%٤٠	14.03	%٦٠	(14,5,2)	٣٧
ثانيتان	%7.40	12.88	%٤٠	4.73	%٦٠	(14,6,2)	٣٨
ثانيتان	%2.40	7.831	%٤٠	6.284	%٦٠	(14,7,2)	٣٩
٣ ثوان	%7.30	5.883	%٤٠	5.91	%٦٠	(14,8,2)	٤٠
٣ ثوان	%4.30	4.748	%٤٠	7.358	%٦٠	(14,9,2)	٤١
ثانيتان	%10.30	17.43	%٤٠	6.133	%٦٠	(14,10,2)	٤٢

ثانيتان	5.50%	8.045	40%	5.242	60%	(14,11,2)	43
3 ثوان	5.50%	10.08	40%	6.643	60%	(14,12,2)	44

ان افضل معمارية للشبكة العصبية المستخدمة هي (2-6-14) بنسبة 70% لمجموعة التدريب ومقدار خطأ 9.237. ونسبة 30% من العينة الكلية لمجموعة الاختبار ومقدار خطأ 4.094. ونسبة التنبؤ الزائف كانت بنسبة 2.2%, اذ تم اختيار المعمارية الهيكلية للشبكة الحالية باعتماد اقل خطأ ممكن لمجموعة الاختبار واقل نسبة تنبؤ زائف في نفس الوقت.



شكل (1) معمارية الشبكة العصبية الاصطناعية المختارة .

من شكل (1) تتضح معمارية الشبكة العصبية الاصطناعية المتكونة من ثلاث طبقات الاولى من اليسار طبقة المدخلات بأربع عشرة وحدة اضافة الى وحدة التحيز (Bias) والثانية الطبقة المخفية بست وحدات ووحدة تحيز (Bias) واحدة والثالثة طبقة المخرجات بوحدتين , الوحدة الاولى (NEWY=1) في حالة كون قيم فترة البقاء المريض على قيد الحياة المنتبأ بها اقل من متوسط فترة البقاء (Y) والوحدة الثانية (NEWY=2) عندما تكون قيم فترة بقاء المريض على قيد الحياة المنتبأ بها اكبر من متوسط فترة البقاء (Y) , ولمعرفة معلومات الشبكة العصبية الاصطناعية المختارة ندرج الجدول (4) .

جدول (٤) معلومات الشبكة العصبية الاصطناعية .

Network Information		
Input Layer	Covariates	1 فصيلة الدم
		2 جنس المريض
		3 العمر عند اكتشاف الإصابة
		4 محل السكن
		5 تاريخ العائلة الطبي
		6 صلة القرى للأبوين
		7 الحالة الاجتماعية
		8 العمل
		9 الإصابة بأمراض الطحال
		10 الإصابة بداء السكري
		11 الإصابة بأمراض القلب
		12 الإصابة بتأخر النمو
		13 الإصابة بوهن او نخر العظم
		14 الإصابة بالتهاب الكبد الفيروسي
Number of Units ^a		14
Scaling Method for Covariate		Standardized
Number of Hidden Layers		1
Number of Units in Hidden Layer		6
Activation Function		Hyperbolic tangent
Output Layer	Dependent Variable	1 Y*
	Number of Units	2
	Activation Function	Softmax
	Error Function	Cross-entropy

a. Excluding the bias unit

ولغرض معرفة ملخص لنتائج الشبكة العصبية المختارة نلحظ الجدول التالي :

جدول (٥) ملخص الشبكة العصبية الاصطناعية .

Model Summary		
Training	Cross Entropy Error	9.237
	Percent Incorrect Prediction	2.1%
	Stopping Rule Used	secutive step(s) decrease in error
	Training Time	0:00:00.02
Testing	Cross Entropy Error	4.594
	Percent Incorrect Prediction	2.2%

Dependent Variable: Y*

Error computations are based on the testing sample

تتضح اهمية المعمارية الحالية للشبكة العصبية من خلال جدول (5) اذا ان مقدار الخطأ بلغ ٩.٢٣٧ ونسبة التنبؤ الزائف ٢.١% لمجموعة التدريب ومقدار خطأ ٤.٥٩٤ ونسبة التنبؤ الزائف ٢.٢% لمجموعة الاختبار مما جعلها الافضل بين المعماريات الاخرى المدرجة في جدول (3). اما نتائج التصنيف (التنبؤ الصحيح) لمجموعتي الاختبار والتدريب فتتضح من الجدول التالي :

جدول (٦) التنبؤ الصحيح (التصنيف) لمجموعتي التدريب والاختبار.

Classification				
Sample	Observed	Predicted		
		1	2	Percent Correct
Training	1	80	2	97.6%
	2	1	58	98.3%
	Overall Percent	57.4%	42.6%	97.9%
Testing	1	24	1	96.0%
	2	0	20	100.0%
	Overall Percent	53.3%	46.7%	97.8%

Dependent Variable: Y*

ان نسبة التنبؤ الصحيح لمجموعتي التدريب (Training) والاختبار (Testing) بلغت 97.9% و ٩٧.٨% على التوالي وهي نسبة جيدة مقارنة مع الشبكات الاخرى , ولمعرفة الاهمية النسبية لمتغيرات الدراسة ندرج الجدول التالي :

جدول (٧) مقدار ونسب الاهمية للمتغيرات المستقلة .

Independent Variable Importance

	Importance	Normalized Importance
فصيلة الدم	.125	48.1%
جنس المريض	.031	12.1%
لعمر عند اكتشاف الاصابة	.260	100.0%
محل السكن	.013	4.9%
تاريخ العائلة الطبي	.067	25.8%
صلة القربى للأبوين	.065	25.0%
الحالة الاجتماعية	.023	8.7%
العمل	.113	43.5%
الاصابة بأمراض الطحال	.026	10.0%
الاصابة بداء السكري	.066	25.5%
الاصابة بأمراض القلب	.034	13.2%
الاصابة بتأخر النمو	.081	31.0%
وهن او نخر العظم	.033	12.5%
بالتهاب الكبد الفيروسي	.062	23.9%

ان متغير العمر عند اكتشاف الاصابة يعد الاعم من بين المتغيرات يليه من حيث الاهمية متغير فصيلة الدم والتحصيل الدراسي وهذا ما يتضح من خلال نسب الاهمية ومقدارها الموضحة في الجدول اعلاه .

٣-٢-٣- نتائج تطبيق الشبكات العصبية الاصطناعية للمتغيرات المعنوية فقط :-

في هذه الحالة سيتم استبعاد المتغيرات غير المعنوية والتي ليس لها تأثير معنوي في فترة البقاء وعندها سوف يتم اختبار عدة انواع من الشبكات العصبية بغية التوصل الى افضل نموذج شبكة لتحليل البيانات , باستعمال خوارزمية الانتشار العكسي للخطأ, ودالة التنشيط اللوجستية ثنائية القطب عند طبقة المخفية ودالة التنشيط العتبة لطبقة المخرجات ولعدة تشكيلات معمارية للشبكة بثبات عدد وحدات طبقة المدخلات والذي يكون مساوياً لعدد المتغيرات المستقلة التي اظهرت معنويتها في فترة بقاء المريض على قيد الحياة باستعمال طريقة كابلان – مايبير والتي تم تحديدها مسبقاً , و ثبات عدد وحدات طبقة المخرجات وبالباغة وحدتين , مع تغير في عدد وحدات الطبقة

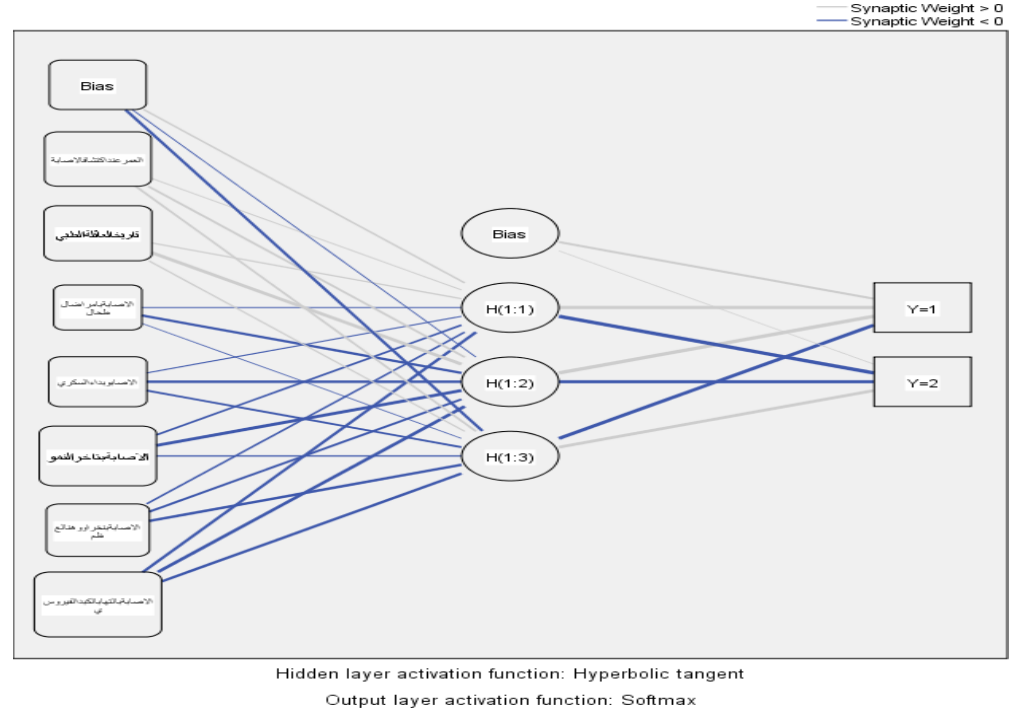
المخفية ونسبة التدريب والاختبار من العدد الكلي لعينة الدراسة في كل مرة للوصول الى افضل هيكلية باقل خطأ ممكن واعلى نسبة تنبؤ صحيح لمجموعة الاختبار المختارة , والجدول التالي يبين اهم النتائج التي تم التوصل اليها :

جدول (٨) نتائج الشبكات العصبية الاصطناعية.

من التدريب	مجموعة الاختبار			مجموعة التدريب		معمارية الشبكة مدخلات , مخفية (مخرجات)	ت
	التنبؤ الزائف	الخطأ	النسبة	الخطأ	النسبة		
٧ ثوان	%3.60	3.977	%١٠	15.39	%٩٠	(٧,٢,٢)	١
٣ ثوان	%4.80	4.499	%١٠	26.02	%٩٠	(٧,٣,٢)	٢
٣ ثوان	%4.80	4.71	%١٠	14.33	%٩٠	(٧,٤,٢)	٣
٣ ثوان	%5.60	2.382	%١٠	24.31	%٩٠	(٧,٥,٢)	٤
٣ ثوان	%5.40	9.8	%١٠	23.38	%٩٠	(٧,٦,٢)	٥
٤ ثوان	%7.60	6.719	%١٠	16.26	%٩٠	(٧,٧,٢)	٦
ثانيتان	%8.90	5.182	%٢٠	22.47	%٨٠	(٧,٢,٢)	٨
ثانيتان	%5.40	4.76	%٢٠	25.45	%٨٠	(٧,٣,٢)	٩
ثانيتان	%5.90	3.048	%٢٠	13.7	%٨٠	(٧,٤,٢)	١٠
٤ ثوان	%8.30	7.234	%٢٠	19.39	%٨٠	(٧,٥,٢)	١١
٤ ثوان	%5.30	5.225	%٢٠	26.3	%٨٠	(٧,٦,٢)	١٢
٤ ثوان	%8.90	5.991	%٢٠	20.39	%٨٠	(٧,٧,٢)	١٣
٣ ثوان	%4.60	8.164	%٣٠	12.06	%٧٠	(٧,٢,٢)	١٥
3 ثوان	%١.٨0	٣.873	%٣٠	٤.٧59	%٧٠	(٧,٣,٢)	١٦
٥ ثوان	%5.90	7.999	%٣٠	11.01	%٧٠	(٧,٤,٢)	١٧
ثانيتان	%5.40	5.283	%٣٠	15.38	%٧٠	(٧,٥,٢)	١٨
٤ ثوان	%5.80	8.933	%٣٠	13.99	%٧٠	(٧,٦,٢)	١٩
٣ ثوان	%7.20	11.24	%٣٠	10.2	%٧٠	(٧,٧,٢)	٢٠
١ ثوان	%5.30	10.56	%٤٠	9.327	%٦٠	(٧,٢,٢)	٢٢
١ ثوان	%5.10	10.48	%٤٠	6.296	%٦٠	(٧,٣,٢)	٢٣
٣ ثوان	%4.10	6.828	%٤٠	14.92	%٦٠	(٧,٤,٢)	٢٤
٣ ثوان	%4.30	11.56	%٤٠	12.31	%٦٠	(٧,٥,٢)	٢٥

ثانيتان	10.80%	16.99	40%	5.668	60%	(7,6,2)	26
٤ ثوان	6.60%	14.25	40%	10.84	60%	(7,7,2)	27

من خلال النتائج المثبتة في الجدول (٨) يتضح ان افضل معمارية للشبكة العصبية (٧,٣,٢) اي بعدد ثلاث وحدات للطبقة المخفية بنسبة تدريب ٣٠% ونسبة اختبار ٧٠% بمقدار خطأ ٤.٧٥٩ في عينة التدريب و ٣.٨٧٣ في عينة الاختبار و ١.٨% تنبؤ زائف , كما في الجداول والاشكال الموضحة تالياً .



شكل (٢) معمارية الشبكة العصبية الاصطناعية المختارة .

ويتضح من شكل (٢) معمارية الشبكة العصبية المختارة والتي تتكون من ثلاث طبقات هي طبقة المدخلات بسبع وحدات اضافة الى وحدة التحيز (Bias) , والثانية في الوسط الطبقة المخفية وتتكون من ثلاث وحدات اضافة الى وحدة التحيز (Bias) , والثالثة وهي طبقة المخرجات وتكون بوحدتين , و لمعرفة معلومات الشبكة العصبية الاساسية والملخصة في الجدول التالي :

جدول (9) معلومات الشبكة العصبية الاصطناعية .

Network Information			
Input Layer	Covariates	1	العمر عند اكتشاف الإصابة
		2	تاريخ العائلة الطبي
		3	الإصابة بأمراض الطحال
		4	الإصابة بداء السكري
		5	الإصابة بتأخر النمو
		6	الإصابة بوهن أو نخر العظم
		7	الإصابة بالتهاب الكبد الفيروسي
Number of Units ^a		7	
Rescaling Method for Covariates		Standardized	
Number of Hidden Layers		1	
Hidden Layer(s)	Number of Units in Hidden Layer 1 ^a	3	
Activation Function		Hyperbolic tangent	
Output Layer	Dependent Variables	1	Y*
	Number of Units	2	
	Activation Function	Softmax	
	Error Function	Cross-entropy	

a. Excluding the bias unit

ففي القسم الاول فيه والخاص بطبقة المدخلات (Input Layer) والتي تتكون من سبع وحدات وبسبعة متغيرات مستقلة (Covariates) , اما في القسم الثاني من الجدول والخاص بالطبقة المخفية (Hidden Layer) والمتكونة من ثلاث وحدات وبدالة تنشيط (Hyperbolic tangent) , اما في القسم الاخير من الجدول والذي يوضح معلومات طبقة المخرجات (Output Layer) والمتكون من متغير معتمد (Dependent Variables) واحد (Y*) بعدد وحدتين (Number of Units) ودالة تنشيط (Softmax) وتعرف أيضا بدالة السغمويد (Sigmoid) , اما النتائج متمثلة بمقدار الخطأ ونسبة التنبؤ الزائف لكل من مجموعتي الاختبار والتدريب والملخصة في الجدول التالي:

جدول (10) ملخص نتائج الشبكة العصبية .

Model Summary		
Training	Cross Entropy Error	4.759
	Percent Incorrect Prediction	3.9%
	Stopping Rule Used	Consecutive steps with no decrease in error
	Training Time	0:00:00.03
Testing	Cross Entropy Error	3.873
	Percent Incorrect Prediction	1.8%

Dependent Variable: Y*

اذ نلاحظ في قسمه الاول النتائج المتعلقة بمجموعة التدريب (Training) والتي كانت بقيمة خطأ 4.759 ونسبة تنبؤ زائف 3.9% ومدة تدريب 3 ثوان , اما القسم الاخر للجدول والخاص بنتائج عينة الاختبار (Tasting) والتي كانت بخطأ مقداره 3.873 ونسبة تنبؤ زائف 1.8% , ومن خلال النتائج انفة الذكر تم اختيار النموذج والمعمارية الحالية للشبكة العصبية المستخدمة في تحليل البيانات كونها تحقق افضل النتائج بين نماذج الشبكات العصبية والمبينة معلوماتها في جدول (8) .

جدول (11) اوزان الشبكة العصبية .

Parameter Estimates

Predictor	Predicted				
	Hidden Layer 1			Output Layer	
	H(1:1)	H(1:2)	H(1:3)	[Y=1]	[Y=2]
(Bias)	.332	-.094-	-.545-		
العمر عند اكتشاف الاصابة	.172	.388	.354		
تاريخ العائلة الطبي	.318	1.124	.211		
الاصابة بأمراض الطحال	-.184-	-.592-	-.052-		
الاصابة بداء السكري	-.203-	-.908-	-.397-		
الاصابة بتأخر النمو	-.365-	-1.141-	-.321-		
الاصابة بنخر اووهن العظم	-.349-	-.479-	-.776-		
التهاب الكبد الفيروسي	-.559-	-1.015-	-.713-		
(Bias)				.537	.021
H(1:1)				1.438	-2.005-
H(1:2)				1.819	-1.241-
H(1:3)				-1.221-	1.117

اما اوزان (Weights) الشبكة العصبية والميمنة قيمها في جدول (١١) , في قسمه الاول قيم الاوزان ما بين طبقة المدخلات (Input Layer) والطبقة المخفية (Hidden Layer) , وفي قسمه الاخر قيم الاوزان ما بين الطبقة المخفية (Hidden Layer) وطبقة المخرجات (Output Layer) , كذلك فان نسبة التنبؤ الصحيح لكل من مجموعتي التدريب والاختبار والمدرجة في الجدول التالي :

جدول (١٢) التنبؤ الصحيح (التصنيف) لمجموعي التدريب والاختبار .

Classification				
Sample	Observed	Predicted		
		1	2	Percent Correct
Training	1	٧١	٣	95.9%
	2	٢	5٤	96.4%
	Overall Percent	٥٦.٢%	٤٣.٨%	9٦.١%
Testing	1	31	1	٩٦.٩%
	2	0	24	100.0%
	Overall Percent	55.4%	٤٤.٦%	98.2%

Dependent Variable: Y*

اذ يتضح ان نسبة التنبؤ الصحيح في عينة التدريب (Training) بلغت 96.1% وفي عينة الاختبار (Testing) بلغت 98.2% , ومن تلك النسب المرتفعة نستدل على ان الشبكة العصبية المستخدمة جيدة في تحليل هكذا نوع من البيانات والمعمارية الحالية للشبكة هي الافضل بين الشبكات الاخرى المدرجة في جدول (٨) , اما فيما يخص نسب ومقدار الاهمية لكل متغير فيمكن ملاحظتها من الجدول التالي :

جدول (١٣) مقدار ونسبة الاهمية للمتغيرات المستقلة .

Independent Variable Importance		
	Importance	Normalized Importance
العمر عند اكتشاف الإصابة	.513	100.0%
تاريخ العائلة الطبي	.175	34.0%
الإصابة بأمراض الطحال	.061	11.8%
الإصابة بداء السكري	.036	7.0%
الإصابة بتأخر النمو	.071	13.9%
الإصابة بنخر او وهن		
التهاب الكبد الف		%

من الجدول اعلاه يمكننا التعرف على نسب ومقدار الاهمية لكل متغير من المتغيرات المستقلة (Independent Variable Importance) اذ كان متغير العمر عند اكتشاف الاصابة بنسبة اهمية ١٠٠% , يليه متغير تاريخ العائلة الطبي بنسبة اهمية ٣٤.٠% , ثم متغير الاصابة بالتهاب الكبد الفيروسي اذ بلغت ١٩.٧% , فيما كان متغير تأخر النمو بنسبة اهمية ١٣.٩% ومتغير الاصابة بأمراض الطحال كان نسبة الاهمية ١١.٨% يليه متغير الاصابة بوهن العظم بنسبة اهمية ٨.٤% واخيراً متغير الاصابة بداء السكري باقل نسبة اهمية ٧.٠% .

٤-٣- مناقشة النتائج :-

في هذه الفقرة نحن بصدد مقارنة النتائج بين الحالتين , الاولى منهما وهي حالة تحليل البيانات باستعمال الشبكة العصبية باعتماد كافة متغيرات الدراسة وكانت نتائج تلك الحالة شبكة عصبية بمعمارية (٢-٦-١٤) والحالة الثانية المتمثلة بتحليل البيانات باستخدام الشبكة العصبية باعتماد المتغيرات المعنوية فقط وكانت افضل معمارية في هذه الحالة (٢-٣-٧), اذ تم مقارنة الحالتين باعتماد مقدار الخطأ لمجموعتي الاختبار والتدريب وكذلك نسبة التنبؤ الزائف والتنبؤ الصحيح (التصنيف) للمفاضلة واختيار الشبكة العصبية الافضل منهما والنتائج موضحة في الجدول التالي :

جدول (١٤) نتائج تحليل البيانات في الحالتين.

الحالة الاولى	الحالة الثانية	
(١٤-٦-٢)	(٢-٣-٧)	معمارية الشبكة
مجموعة التدريب		
٧٠%	٧٠%	نسبة المجموعة
٢.١%	٣.٩%	التنبؤ الزائف
٩.٢٣٧	٤.٧٥٩	مقدار الخطأ
٩٧.٩%	٩٦.١%	التنبؤ الصحيح (التصنيف)
مجموعة الاختبار		
٣٠%	٣٠%	نسبة المجموعة
٢.٢%	١.٨%	التنبؤ الزائف
٤.٥٩٤	٣.٨٧٣	مقدار الخطأ
٩٧.٨%	٩٨.٢%	التنبؤ الصحيح (التصنيف)

من جدول (١٤) والذي يبين ملخص نتائج تحليل البيانات باستخدام الشبكات العصبية في الحالتين وتتضح من خلالها ان الافضلية عند مقارنة تلك النتائج للشبكة العصبية بمعمارية (٢-٣-٧) والتي تمثل الحالة الثانية للتحليل , اي

باستعمال المتغيرات المعنوية فقط اذ كانت نسبة التنبؤ الصحيح ٩٨.٢% وهي اعلى من الحالة الاولى التي كانت نسبة التنبؤ الصحيح بها ٩٧.٨% وكذلك الحال لبقية المؤشرات الاخرى المثبتة في الجدول اعلاه.

٤- الاستنتاجات والتوصيات :

٤-١- الاستنتاجات :

١- ان للشبكات العصبية الاصطناعية (ANN) القدرة والامكانية لدراسة وتحليل انواع مختلفة من البيانات بغض النظر عن نوعها سواء كانت بيانات كمية ام وصفية اذ تعد من الاساليب التكيفية لتحليل البيانات وهذا ما اتضح من خلال نسبة التنبؤ الصحيح .

٢- افضلية الهيكلية المعمارية (٧-٣-٢) على الهيكليات الاخرى للشبكة العصبية المستعملة لتحليل بيانات البقاء الخاصة بمرضى الثلاثيميا الكبرى .

٣- ان متغير العمر عند اكتشاف الإصابة يعد اهم المتغيرات المستقلة وله اكبر تاثير في فترة بقاء المريض على قيد الحياة .

٤- ان لطريقة كابلان – ماير اهمية كبير في تحديد مدى معنوية المتغيرات في فترة البقاء قبل تحليل البيانات باستعمال الشبكات العصبية الاصطناعية.

٤-١- التوصيات :

١- اعتماد اسلوب الشبكات العصبية الاصطناعية لتحليل وتصنيف بيانات مرضى الثلاثيميا والتنبؤ بفترة بقاء المريض على قيد الحياة في المستقبل من قبل المؤسسات والباحثين .

٢- نوصي بدراسات مقارنة فيما بين الشبكات العصبية الاصطناعية والاساليب الاحصائية التقليدية كنماذج الانحدار المختلفة كنموذج كوكس (COX) وانحدار توبت للمفاضلة بينها .

٣- دراسة بيانات البقاء لظواهر وحالات اخرى باستعمال الشبكات العصبية الاصطناعية والخروج بنتائج فيما اذا كان لها القدرة على التكيف مع تلك الانواع من البيانات لتعميم نتائجها والاستفادة منها .

٤- استخدام طريقة كابلان – ماير قبل الخوض في غمار عمليات التحليل الاخرى فيما يخص بيانات البقاء .

المصادر :

١- العتيبي - فوزي, (٢٠٠٣م), "استخدام السلاسل الزمنية والشبكات العصبية في التنبؤ بالارقام القياسية" رسالة

ماجستير في الاحصاء التطبيقي ,جامعة قناة السويس ,مصر.

٢- محمود - أسيل سمير محمد ، (٢٠٠٥م) ، "مقارنة بين طرائق تحليل وتنبؤ السلاسل الزمنية وتطبيقها على

مبيعات الشركة العامة لتوزيع كهرباء بغداد" ، أطروحة مقدمة الى قسم الإحصاء , كلية الإدارة والأقتصاد , جامعة

بغداد .

Ƴ-Andereescu,(2004) , " Forecasting Corporate Earning: A Data Mining Approach ", MSc, Thesis in Accounting the Swedish School of Economics and Business administration .

ε- Collett, D. (2003) ,"Modelling Survival Data in Medical" , Research ,Chapman and Hall London .

ϕ-Hannu Pitkänen,(1998),"Multivariate Statistical Methods in Systems Engineering", Email: Hannu.Pitkänen@hut.fi

ϖ-Haykin, S. (1999) , " Neural Networks A Comprehensive Foundation " second edition.

ϗ-Kneale , p.; See, L. and Smith, A. (2000). "Towards Defining Evaluation Measures for Neural Network Forecasting Models", School of Geography, University of Leeds , UK.

8-Mantel,N.,(1966)," Evaluation of survival data and two new rank order statistics arising in its consideration", Cancer Chemotherapy Report.

9-Mitchell, T. M. ,(1997) , "Machin learning Boston", WCB/ Mcgraw-hill,new york.

10-Rajasck, S. & Vijaylak sh mi, G. A. ,(2003) ,"neural networks Algorithms syanthesis and Application", prentice- Hall of India, newdelhi.

11-Tkacz Greg. (2001). "Neural network forecasting of Canadian GDP growth", International Journal of Forecasting 17 (2001) 57–69, Ottawa , Ontario , Canadak .