HNSJ Humanitarian & Natural Sciences Journal
ISSN: (e) 2709-0833
www.hnjournal.net
مجلة العلوم الإنسانية والطبيعية
مجلة علمية محكمة (التصنيف: ISI و NSP)
معامل التأثير العربي للعام 2020 = 0.44
معامل تأثير ISI للعام 21/20 = 0.352

عنوان البحث

استخدام نماذج الشبكات العصبية الاصطناعية في التحليل التمييزي بالتطبيق على عينة من المصابين وغير المصابين بمرض البروستات

د. إبراهيم محمد إبراهيم سيد أحمد[1] د. أشرف ادريس سعيد محمد[2] د. محمد أحمد محمد حسن[3]

[1] أستاذ مساعد بقسم الاحصاء التطبيقي كلية العلوم والتقانة /جامعة شندي.
[2] أستاذ مساعد بقسم الاحصاء كلية الاقتصاد /جامعة القضارف.
[3] أستاذ مساعد بقسم الاحصاء التطبيقي كلية العلوم والتقانة /جامعة شندي.

تاريخ النشر: 2021/07/01م تاريخ القبول: 2021/06/11م

المستخلص

تهدف هذه الدراسة إلي بناء نموذج احصائي حديث لتصنيف المصابين بمرض البروستات وغير المصابين بناء على المتغيرات (العمر، الهمقلوبين ،كريات الدم البيضاء،الصفائح الدموية،(P.S.A)،اليوريا ،الكرياتين(S.C) باستخدام نماذج الشبكات العصبية الاصطناعية في التحليل التمييزي حيث تم استخدام المنهج الوصفي والتحليلي باستخدام برنامج spss21 للوصول لنتائج الدراسة وتوصلت الدراسة الى أن الشبكة العصبية الاصطناعية في التحليل التمييزي هي شبكة دالة الاشعاع لبناء نموذج الشبكة العصبية الاصطناعية وتم التوصل الي أن أهم متغير يؤثر علي الاصابة بمرض البروستات هو العمر ثم يليه الهيمقولبين ثم يليه الكراتينين ثم يليه اليوريا ثم يليه (P.S.A) ويليه الصفائح الدموية ، ومن أهم ما توصي به الدراسة استخدام نماذج شبكات العصبية الاصطناعية في التحليل التمييزي لثبات دقتها وقلة الاخطاء بنماذجها المقدرة .

RESEARCH ARTICLE

THE USE OF ARTIFICIAL NEURAL NETWORKS MODELS IN THE DISCRIMINATORY APPLICATION ANALYSIS ON A SAMPLE OF INFECTED AND NON-INFECTED WITH PROSTATE

Ebraheem mohmmed Ebraheem seed[1] Ashraf Edress Saeed[2] Mohammed Ahmed Mohammed Hassan[3]

[1] Assistant Professor, Department of Applied Statistics, College of Science and Technology, Shendi University.
[2] Assistant Professor, Department of Statistics, Faculty of Economics, Gedaref University.
[3] Assistant Professor, Department of Applied Statistics, College of Science and Technology, Shendi University.

Published at 01/07/2021 **Accepted at 11/06/2021**

Abstract

This study aims to construct a modern statistical model for classifying prostate and non-infected people based on variables (age, hummus, white blood cells, platelets, P.S.A.), urea, creatine (S.C) using artificial neural network models In the discriminatory analysis. The researchers used the descriptive analytical method then used spss21 to come out with the results of the study. The study found that the artificial neural network in discriminatory analysis is the Radial Basis Function (RBF) to construct the model of the artificial neural network and it was concluded that the most important variable affects the incidence of prostate disease It is age and then followed by hemoglobin, followed by creatine, followed by urea, followed by P.S.A., followed by platelets, and one of the most important recommended by the study is the use of artificial neural network models in discriminatory analysis to stabilize their accuracy and reduce errors with their estimated models.

المقدمة

تمهيد

ان التحليل التمييزي هو عبارة عن تقنية احصائية رياضية قوية , تستخدم لتوصيف عناصر المجتمع المدروس والموزعة على مجموعات محددة ومنفصلة ومتكاملة (أكثر من أو يساوي مجموعتين) , وتحديد الحدود الفاصلة بينها واستخلاص قاعدة معينة لتحديد انتماء أي عنصر اليها كما يمكن استخدام التحليل التمييزي للتنبؤ بانتماء أي عنصر جديد الى احدى المجموعات المصنفة ، ويتم استخدام الدالة التمييزية لتصنيف تلك المجموعات بعدة طرق مختلفة ، ولكن ظهرت حديثا مايسمى نماذج الشبكات العصبية الاصطناعية والتي يمكن أن تستخدم في التحليل التمييزي وتصنيف البيانات [1].

(1)مشكلةالبحث

تتمثل مشكلة البحث في التعرف على أهم ندرة النماذج الاحصائية الحديثة التى تقوم بدور التحليل التمييزي بصورة أدق وندرة تطبيقها في مجالات الحياة المختلفة وهي نماذج الشبكات العصبية الاصطناعية في التحليل التمييزي للوصول لنموذج مناسب يتصف بالدقة الاحصائية وقلة خطأ التصنيف .

(2) أهمية البحث

أهمية هذا البحث من جانبين جانب صحي وجانب إحصائي فمن الجانب الصحي فإن بناء نموذج احصائي يمكن تصنيف الأشخاص بأنهم مصابين بمرض البروستات أو غير مصابين به بناء على متغيرات وبيانات تؤخذ منهم شيء مفيد للمؤسسات الصحية حيث, أما من الجانب الإحصائي فإن استخدام التحليل الإحصائي المعتمد علي نماذج حديثة دقيقة مثل نماذج الشبكات العصبية الاصطناعية التي تستخدم في تحليل التحليل التمييزي يعتبر اضافة حقيقية لعلم الاحصاء التطبيقي .

(3)أهداف البحث

يهدف هذا البحث الي الاتي :

1- التعرف على نماذج الشبكات العصبية الاصطناعيةوانواعها وخطواتها .
2- التعرف على كيفية استخدام نماذج الشبكات العصبية الاصطناعية في التحليل التمييزي.
3- استخدام النموذج المقترح لمعرفة لتصنيف بيانات الأشخاص القادمون للمؤسسات الصحية الى مرضى البروستات وغير مرضى.

(4)فرضيات البحث

نموذج الشبكات العصبية الاصطناعية المقدر لتصنيف بيانات الدراسة نموذج مناسب لبيانات الدراسة.

يوجد تأثير معنوي من قبل متغيرات الدراسة على النموذج المقدر .

(5)منهجية البحث

يستخدم في هذا البحث المنهج الإحصائي الوصفي , والمنهج الإحصائي التحليلي باستخدام برنامج spss21.

الجانب النظري

مفهوم الشبكات العصبية الاصطناعية Concept of Artificial Neural Networks

تعرف الشبكة العصبية الاصطناعية بأنها هي محاولة رياضية برمجية لمحاكاة طريقة عمل الدماغ البشري ، وهي عبارة عن مجموعة مترابطة من عصبونات افتراضية تعمل عمل العصبون البايولوجي تستخدم لمعالجة المعلومات بناءً علي الطريقة الاتصالية في الحاسوب[3]

والشبكات العصبية الاصطناعية تتعلم بطريقة تشابه تعلم الانسان من خلال الامثلة والتدريب, والشبكات العصبية تُهيأ وتنظم لتطبيقات محددة مثل نموذج التمييز والإدراك أو تصنيف البيانات من خلال عملية التعلم . والتعلم في النظام البيولوجي يستخدم تكييف نقاط الاشتباك العصبي بين العصبونات وهذه هي الفكرة الجوهرية في عمل الشبكات العصبية[4] .

أنواع الشبكات العصبية الاصطناعية

1. شبكات أمامية Feed forward :

في هذه الشبكات يتم الاتصال بحيث تتدفق باتجاه واحد ويتم ذلك من خلايا الدخل الي خلايا الخرج .

2. شبكات إرجاعية Recurrent :

في هذه الشبكات يتم الاتصال بحيث تتدفق باتجاهين امامي وخلفي , وتنقسم الشبكات الإرجاعية الي شبكات إرجاعية تامة وإرجاعية جزيئة .

وحدات المعالجة (العصبونات) في الشبكة العصبية الاصطناعية

وحدات المعالجة او العصبونات هي الوحدات التي تقوم بعملية معالجة المعلومات في الشبكة العصبية وهي تشكل المكونات الاساسية التي تتألف منها كل طبقات الشبكة العصبية. وتتصل هذه الوحدات بطرق مختلفة بواسطة الوصلات البينية لتعطي الشكل العام او البنية المعمارية للشبكة العصبية.

تتبع عناصر المعالجة نظام المعالجة المتوازية في اجراء الحسابات المسندة اليها او معالجة البيانات وهي في ذلك تتبع عمل العقل البشري.

وتتألف اي وحدة معالجة او عصبون من المكونات الأساسية التالية:

1- معاملات الأوزان
2- دالة الجمع
3- دالة التحويل
4- دالة الإخراج

أنواع نماذج الشبكات العصبية الاصطناعية التي تستخدم في تحليل الانحدار:

هنالك العديد منها واهمها :

1- البيرسبترون متعدد الطبقات Multilayer perceptron (MLP)
2- شبكات دالة القاعدة الاشعاعية Radial Basis Function (RBF)
3- شبكات الاسترجاع الخلفي Recurrent Networks

4- شبكات Sigma- Pi& Pi sigma

5- شبكات Ridge Polynomial

وهنا سوف نتطرق الي شبكتي MLP & RBF[4]

- **البيرسبترون متعدد الطبقات Multilayer perceptron (MLP)**

شبكة البيرسبترون متعدد الطبقات وتسخدم في عدة نماذج احصائية منها التحليل التمييزي .من اكثر الشبكات استخداماً في تحليل الانحدار وسنخلص خطوات تحليل تصنيف البيانات باستخدام هذه الشبكة بالخطوات التالية :

1- من الأمر analyze نختار الأمر الخاص بهذه الشبكات وهو Neural Networks .

2- نختار الشبكة multilayer perceptron من Neural Networks .

3- نقوم بادخال المتغير التابع في dependent variables والمتغيرات المستقلة في covariates او factors حسب نوعها متغيرات او عوامل مصنفة .

4- تحديد الطريقة للتحليل من طرق الانحدار مثلا standardized من الامر rescalling of covariates

5- بعد ادخال المتغيرات في مكانها الصحيح نذهب للامر partitions وهنا نقوم بعملية توزيع عينة الدراسة على ثلاث اجزاء (التدريب – الاختبار – التحقق) ويفضل تقسيمها بالتساوي (33% لكل جزء مثلا).

6- من الأمر output نختار الامر description لوصف بيانات الشبكة والامر model summery للحصول لمعلمات نموذج الدالة التمييزية المقدرة المقدر بواسطة الشبكة والاهمية النسبية لكل متغير ، والأمر Classification results بينما الامر diagramلرسم الشبكة بيانيا .[13]

- **شبكات دالة القاعدة الاشعاعية Radial Basis Function (RBF)**
- تتكون معمارية هذه الشبكة من طبقتين حيث تحتوي علي طبقة خفية واحده مع دالة تحفيز قاعدة اشعاعية وطبقة مخرجات مع دوال تحفيز خطية ،ويتبع نفس الخطوات التي ذكرت في الشبكة MLP في التحليل التمييزي فقط نقوم في البداية باختيار الشبكة شبكات دالة القاعدة الاشعاعية Radial Basis Function (RBF).[2]

خطوات بناء الشبكة العصبية الاصطناعية

1- تجميع واعداد البيانات :

2- معالجة البيانات :

يتم اجراء بعض العمليات علي البيانات المستخدمة مثل تحديد الاتجاه العام , التركيز علي العلاقة بين المشاهدات ,ايجاد توزيع البيانات .

3- تقسيم البيانات الي مجاميع :

تقسم البيانات المتوفرة الي المجاميع الاتية :

- مجموعة التدريب Training set : وهي مجموعة تعلم وتحدد نموذج للبيانات .
- مجموعة الاختبار Testing set: والتي يمكن عن طريقها تقدير مهارة الشبكة الافتراضية وامكانية استخدامها بصورة عامة .
- مجموعة التحقيق Validation set: وهي مجموعة لاجراء اختبار نهائي لاداء الشبكة .

4- تحديد تركيبة الشبكة :

- شبكات ذات طبقات خفية .
- شبكات ذات طبقات مزدوجة .

5- اختيار خوارزمية التعلم .

6- تحديد قيم الأوزان الابتدائية .

7- تدريب الشبكة :

يتم تحديد مجموعة الأوزان بين العصبونات ومن ثم تحسين هذه الأوزان نتيجة التدريب والتي تحدد أُقل قيمة لمربع الخطأ للوصول الي أوزان تعطي نتائج دقيقة .

8- الاختبار(معيار التقويم) :

إن المعيار المستخدم في الشبكات العصبية الاصطناعية هو مجموع مربعات الخطأ.

9- التنفيذ:

وهي من أهم الخطوات ,اذ تختبر الشبكة من حيث قدرة التكيف مع حالة التغير في الدورة وامكانية اعادة التدريب والوصول الي أقل مربع خطأ عند تغير البيانات الي الوصول الي حالة الاستقرار .[10]

النتائج والمناقشة

تم استخدام أسلوب الشبكات العصبية الاصطناعية لبيانات الدراسة في تحليل الانحدار باستخدام برنامج SPSS21وكانت النتائج كالتالي :

وصف الشبكة العصبية الاصطناعية لبيانات الدراسة

جدول رقم (1) :وصف نموذج الشبكة العصبية الاصطناعية المقدرة للتحليل التمييزي

اسم الشبكة	خطأ التدريب	خطأ الاختبار	خطأ التحقق	دالة التنشيط
RBF7-5-2	0.077	0.429	000.	Softmax

المصدر : من تحليل بيانات الدراسة باستخدام برنامج spss21

نوع الشبكة :

نلاحظ انه تم استخدام شبكة دالة القاعدة الاشعاعية RBF لبناء نموذج الشبكة العصبية الاصطناعية في التحليل التمييزي لبيانات الدراسة (العمر، الهمقلوبين ،كريات الدم البيضاء،الصفائح الدموية،(P.S.A) ،اليوريا ،الكرياتنين(S.C).

وتعتبر شبكة MLP من أهم انواع الشبكات العصبية الي تستخدم في التحليل التمييزي.

بنية النموذج :

تم تحديد معمارية وبنية الشبكة من خلال تجريب عدد من التراكيب المختلفة والمفاضلة بينها باستخدام معيار أخطاء التدريب واخطاء الاختبار واخطاء التحقق حيث أن هذه العملية تكرر حتي تصل الي بنية معينة اعتمادا علي اقل اخطاء التدريب للبنيات المختلفة حيث يتم التكرار حتي يظهر التدني المستمر في قيمة الخطأ المصاحب لعملية التدريب وعندما نلاحظ أن قيمة الخطأ قد تدني كثيرا, توقف في الانخفاض عنده يتم التوقف في عملية التدريب والاعتماد علي البنية والمعمارية التي توقفت فيها عملية التدريب .

نلاحظ من الجدول اعلاه أنه قد تم اختيار المعمارية RBF 7-5-2 لبيانات الدراسة اعتمادا علي اخطاء التدريب التي توقف الانخفاض فيها وهي1.59 .وتعني هذه البنية 7-5-2 انه توجد سبعة طبقات ادخال وهي متغيرات الدراسة التي, وتوجد بالبنية 5 طبقا خفية طبقةواحدة وتوجد طبقتين اخراج (مخرجات النموذج مصابين وغيرمصابين).

دوال التحفيز :

تم استخدام دوال تحفيز او دوال تنشيط حيث تم استخدام دالة التنشيط Softmax

وتم استخدام خوارزمية الانتشار السريع للتدريب .

مدخلات الشبكة :

تم تغذية الشبكة العصبية الاصطناعيةمتغيرات الدراسة وهي (العمر، الهمقلوبين ،كريات الدم البيضاء،الصفائح الدموية،(P.S.A) ،اليوريا ،الكرياتنين(S.C)) لمجموعة من المصابين ومجموعة أخرى من غير المصابين.

, وتم تقسيم البيانات المدخلة الي الشبكة بطريقة الثلث تلافياً لحدوث اي مشاكل في التدريب حيث :

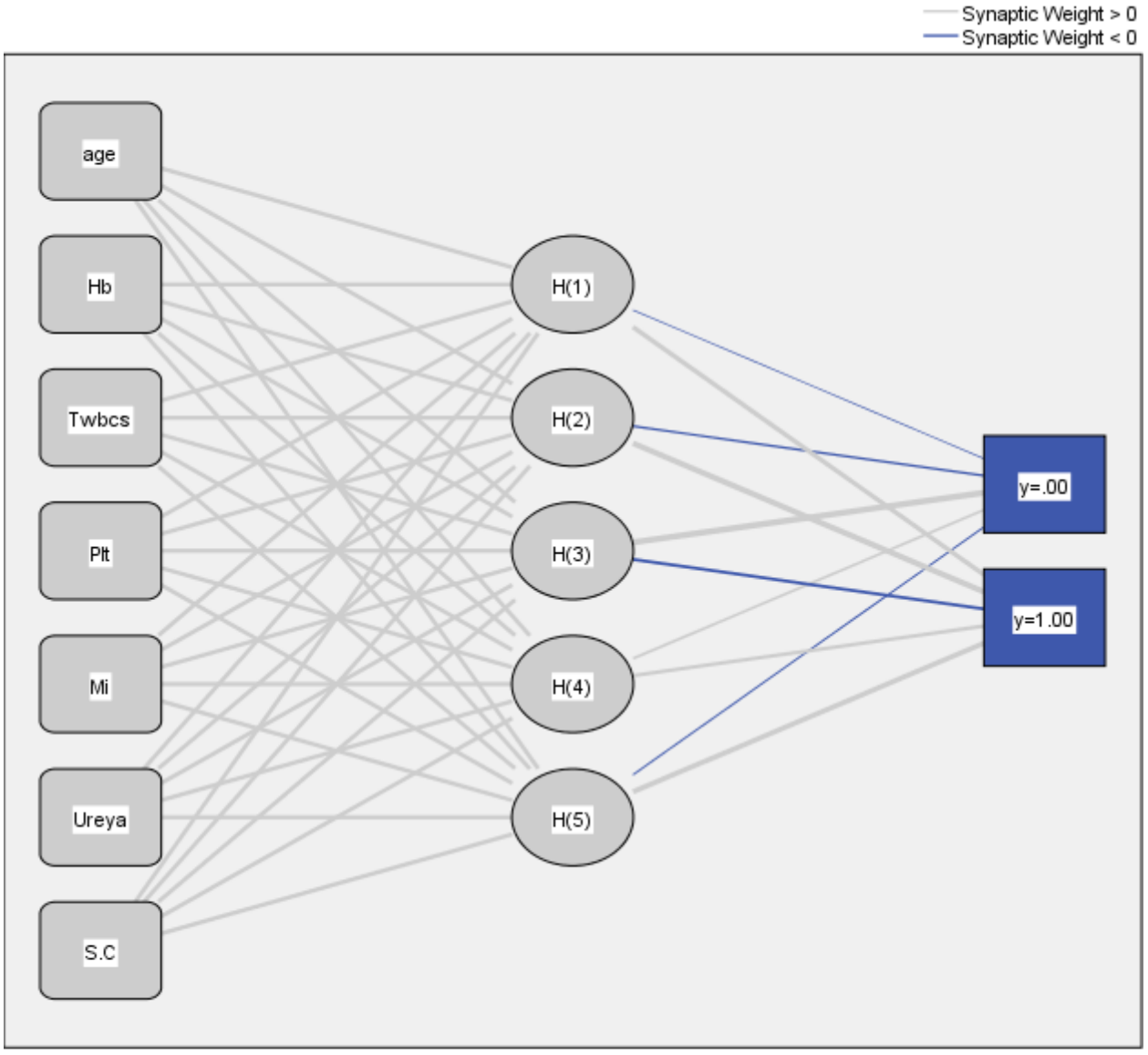
34% من البيانات لاجراء التدريب .

33 % من البيانات لاجراء الاختبار .

33% من البيانات لاجراء التحقق .

رسم بياني للشبكة العصبية الاصطناعية لبيانات الدراسة :

شكل (1) يوضح رسم الشبكة العصبية الاصطناعية لتصنيف بيانات الدراسة



المصدر : من تحليل بيانات الدراسة باستخدام برنامج spss21

معلمات نموذج الدالة التمييزية المقدرة باستخدام الشبكات العصبية الاصطناعية :

جدول رقم (3)معلمات نموذج الدالة التمييزية المقدرة بواسطة الشبكة العصبية الاصطناعية

المتغير	القيمة المقدرة
الثابت	0.634
العمر	0.968-
الهمقلوبين	0.758
كريات الدم البيضاء	0.323
الصفائح الدموية	0.496
(P.S.A)	0.347
اليوريا	0.580
الكرياتنين(S.C)	0.761-

المصدر : من تحليل بيانات الدراسة باستخدام برنامج spss21

من الجدول (3) يتضح لنا أن النموذج المقدر للدالة التمييزية باستخدام نموذج الشبكات العصبية الاصطناعية هو

$$\hat{y} = .634 - .968x_1 + .758x_2 + .323x_3 + .496x_4 + .347x_5 + .580x_6 - .761x_7$$

جدول التصنيف :

جدول رقم (4) جدول التصنيف :

Sample	Observed	Predicted		
		NO	YES	Percent Correct
	NO	11	3	78.6%
	YES	0	31	100.0%
	Overall Percent	24.4%	75.6%	93.3%

المصدر : من تحليل بيانات الدراسة باستخدام برنامج spss21

من الجدول (4) نلاحظ أن نموذج الشبكة العصبية الاصطناعية المقدر لتصنيف بيانات الدراسة أن نسبة التصنيف الصحيح هو93.3% وهي نسبة كبيرة تدل على قوة النموذج في التصنيف ، بينما نسبة التصنيف الخاطئ هي 6.7%.

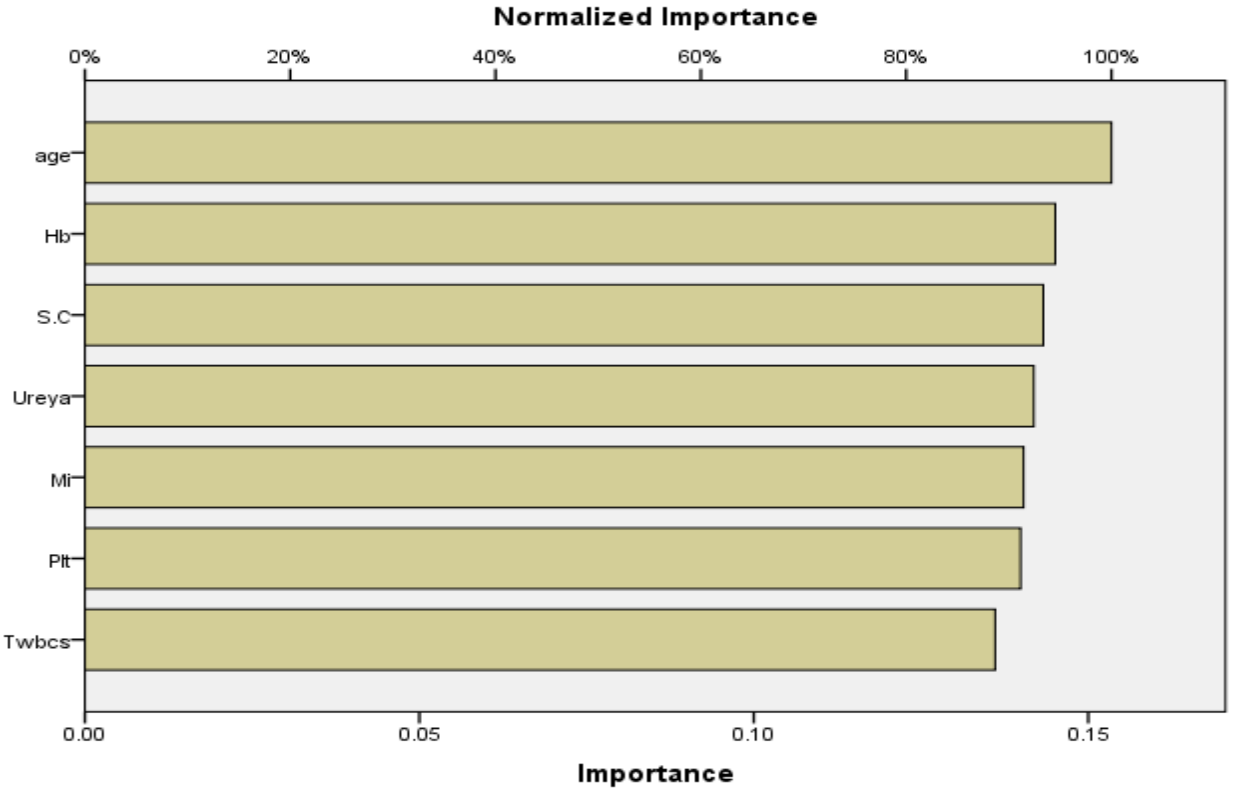
الأهمية النسبية لمتغيرات الدراسة

جدول (5) الأهمية النسبية لمتغيرات الدارسة

	Importance	Normalized Importance
Age	.153	100.0%
Hb(g/dl)	.145	94.5%
TWBCs(cmm)	.136	88.7%
PLT(Ml)	.140	91.2%
(Ml)	.140	91.4%
UREYA(mg/dl)	.142	92.4%
S.C(mg/dl)	.143	93.4%

من الجدول (5) نلاحظ أن أهم متغير يؤثر علي الاصابة بمرض البروستات هو العمر ثم يليه الهيمقولبين ثم يليه الكراتينين ثم يليه اليوريا ثم يليه (P.S.A) ويليه الصفائح الدموية.

شكل (2): الأهمية النسبية لمتغيرات الدارسة



الاستنتاجات :.—

1- تم استخدام شبكة دالة القاعدة الاشعاعية MLP لبناء نموذج الشبكة العصبية الاصطناعية في التحليل التمييزي لبيانات الدراسة (العمر، الهمقلوبين ،كريات الدم البيضاء،الصفائح الدموية،(P.S.A) ،اليوريا ،الكرياتنين(S.C) وكانت معمارية تركيب الشبكةهي 2-5-7 RBF لبيانات الدراسة اعتمادا علي اخطاء التدريب التي توقف الانخفاض فيها وهي1.59 .وتعني هذه البنية 2-5-7 انه توجد سبعة طبقات ادخال وهي متغيرات الدراسة التي, وتوجد بالبنية 5 طبقا خفية طبقةواحدة وتوجد طبقتين اخراج (مخرجات النموذج مصابين وغيرمصابين) وكانت دالة التنشيط هي الدالة Softmax.

2- نموذج الشبكة العصبية الاصطناعية المقدر لتصنيف بيانات الدراسة أن نسبة التصنيف الصحيح هو93.3% وهي نسبة كبيرة تدل على قوة النموذج في التصنيف ، بينما نسبة التصنيف الخاطئ هي 6.7%.

3- أن أهم متغير يؤثر علي الاصابة بمرض البروستات هو العمر ثم يليه الهيمقولبين ثم يليه الكراتينين ثم يليه اليوريا ثم يليه (P.S.A) ويليه الصفائح الدموية.

التوصيات:

1- استخدام نماذج شبكات العصبية الاصطناعية في التحليل التمييزي لثبات دقتها وقلة الاخطاء بنماذجها المقدرة .

2- عمل المزيد من الدراسات حول نماذج الشبكات العصبية الاصطناعية في التحليل التمييزي مثلا دراسة مقارنة بين نموذج الدالة التمييزية بطرقها الكلاسيكية المختلفة ونموذج الشبكات العصبية الاصطناعية وغيرها .

المراجع المصادر :

1- بريان ف.ج. مانلي – ترجمة عبدالرحمن محمد أبوعمه الأساس والطرق الإحصائية المتعدد المتغيرات – دار النشر العلمي والمطابع – جامعة الملك سعود 1422هـ -2001 م .

2- ريتشارد جونسون، دينوشرن- تعريب عبدالرحمن حامد عزام – التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية – دار المريخ للنشر – 1418 هـ – 1998 م.

3- باسل يونس الخياط , عزه حازم ذكي " استخدام الشبكات العصبية في التكهن بالسلاسل الزمنية "المجلة العراقية للعلوم الاحصائية , الطبعة الثامنة2005 م.

4- ديفيد م سكا بورا " بناء الشبكات العصبية " ترجمة فهد بن عبدالله التركي _ الرياض جامعة الملك سعود 2002م .

5- سولاف صبري ,جواد كاظم (2005) "التنبؤ بمبيعات البنك المركزي من العملة الأجنبية " , رسالة منشورة , المجلة العراقية للعلوم الإحصائية.

6- ظافر مطر , عبدالكريم حسين (2006) " النماذج الحركية لدالة التحويل وتعدد المدخلات " , رسالة منشورة ,

المجلة العراقية للعلوم الإحصائية .

7- عدنان ماجد (2002) , "طرق التنبؤ الإحصائي ", الجزء الأول , جامعة الملك سعود .

8- فاندال والتر (1992) "السلاسل الزمنية من الوجهة التطبيقية ونماذج بوكس –جنكيز ", تعريب عبد المرضي حامد عزام ,دار المريخ للنشر ,الرياض , المملكة العربية السعودية .

9- ندوى خزعل رشاد , عزة حازم رشاد (2010م) استخدام الشبكات العصبية في تحليل الانحدار المتعدد

10- ريم على الجراح , مناهل عبد الكريم عزوزو ,على داؤد النعيمي (2008م) مقارنة بين التحليل التمييزي والشبكات العصبية في التشخيص الطبي لمرضى سرطان الفم ، ورقة علمية منشورة لمجلة تنمية الرافدين العدد 92

مجلد 30 لسنة 2008م.

11- H.K.Cigizoglu,(2003)" Incorporation of ARMA Models into flow forecasting by artificial neural networks"

12- Lukepoli, Helmut, (2005)" New Introduction to Multiple Time Series Analysis", ن Springer Berlin Heidelberg, New York.

13- Spss21