مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (7) العدد(1) 2023
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (7) No. (1) 2023

دراسة تأثير استخلاص الميزات على دقة كشف الاختراقات الأمنية في الشبكات الحاسوبية باستخدام الشبكات العصبونية

د. راغب طعمه *

م. ريم مالك ابراهيم **

(تاريخ الإيداع 2022/11/1 . قُبِل للنشر في 2023/1/17)

□ ملخّص□

في مجال أمن الشبكات، يوجد اشتكشاف مستمر للهجمات الالكترونية التي تؤدي إلى عدم استقرار الشبكة ،حيث أدى زيادة استخدام الانترنت وانتشار الأجهزة الذكية بشكل كبير إلى تصاعد الأنشطة الخبيثة في الشبكة. فكان لابد من انشاء أنظمة قوية لكشف الاختراقات و الحالات الشاذة في الشبكة والعمل على جعل هذا النظام يحارب الوصول غير المصرح به إلى موارد الشبكة وبالتالي تأمين المعلومات.بالرغم من اقتراح العديد من الأساليب لاكتشاف الحالات الشاذة في الشبكة الا ان سعي المهاجمين إلى تغيير سلوكهم باستمرار يجعل بناء نظام آمن تحديا في هذا المجال. تم في هذا البحث دراسة تأثير استخلاص السمات على دقة كشف الإختراقات الأمنية باستخدام تقنية الشبكة العصبونية. إذ يناقش البحث إجراء عدة تجارب تم في كل تجربة استخدام طريقة من طرق استخلاص الميزات ودراسة مدى تأثر دقة الكشف ،كما تم توضيح آلية الحصول على دقة الكشف بشكل مفصل بالاعتماد على مصفوفة الارتباك والشبكة العصبونية اذ تم اعتماد التجريب في اختيار بنية الشبكة العصبونية بما يناسب البحث.

و أظهرت النتائج المنجزة في بيئة الماتلاب افضل طريقة من طرق استخلاص السمات بالنسبة لقاعدة البيانات المستخدمة في بيئة الشبكة العصبونية وهي تدريب الشبكة العصبونية على مجموعة السمات بعد حساب معامل الترابط.

الكلمات المفتاحية: الاختراقات الأمنية، استخلاص السمات، الشبكات العصبونية، الدقة، تخفيض السمات، مصفوفة الدقة.

* مدرس في قسم هندسة تكنولوجيا المعلومات – كليّة هندسة تكنولوجيا المعلومات والاتصالات – جامعة طرطوس–سوريا

** ماجستير في قسم هندسة تكنولوجيا المعلومات – كليّة هندسة تكنولوجيا المعلومات والاتصالات – جامعة طرطوس–سوريا

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (7) العدد(1) 2023
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (7) No. (1) 2023

Studying the effect of feature extraction on the accuracy of security breach in computer networks detection using neural networks

Dr.Ragheb Toemeh *
Eng.Reem Malek Ibrahem **

(Received 1/11/ 2022 . Accepted 17/1/2023)

□ ABSTRACT

In network security,there is a continuous detection of electronic attacks that lead to a destabilised network. Moreover, the increased use of the Internet and the spread of smart devices have resulted in an escalation of malicious activity in the network. Therefore, it was necessary to establish strong systems to detect intrusions and anomalies in the network and work to make this system fight unauthorized access to network resources and thus secure information. Although many methods have been suggested to detect anomalies in the network, the quest of attackers to constantly change their behavior makes building a secure system a challenge in this area. In this research, the effect of feature extraction on the accuracy of detecting security breaches was studied using neural network technology. The research focuses on the conduct of multiple experiments, in each experiment using a characteristic extraction method, and the study of the extent to which detection accuracy is affected. The mechanism of achieving detectio accuracy has been explained in detail from the confusion matrix and the neural network,as the experiment was adopted by choosing the structure of the neural network as a function of the research. The results achieved in the Matlab environment showed the best method of feature extraction for the database used in the neural network environment, which is to train the neural network on the set of features after calculating the correlation coefficient.

Key Words: Security breaches, feature extraction, neural networks, accuracy, attribute reduction, accuracy matrix

* Teacher, Information Technology Engineering Department, Information and communication Technology Engineering, Tartous University, Syria.
** Master Information Technology Engineering Department, Information and communication Technology Engineering, Tartous University, Syria.

1. المقدمة

هجوم الشبكة الحاسوبية هو محاولة الحصول على تصريح غير شرعي للوصول إلى الشبكات الحاسوبية وذلك بهدف سرقة البيانات أو اختراق أمن المعلومات للقيام بأعمال وأنشطة تضر بمصالح المؤسسات، حيث تعد هذه الهجمات من مهددات أمن المعلومات.

تطورت الطرق المتبعة في عملية كشف الاختراقات الأمنية في الشبكات وقواعد البيانات الضخمة مع تطور تقنيات تعلم الآلة، كما أن الزيادة السريعة في تقنيات الانترنت والأجهزة الذكية زاد عدد ونوع الهجومات على البيانات، ولمواكبة هذا التطور في أنواع الهجومات كان لابد من إجراء العديد من الأبحاث مستخدمة العديد من التقنيات التقليدية والمحسنة بهدف حماية هذه البيانات وكشف الاختراقات بدقة وسرعة[1].

تعد أنظمة كشف الاختراقات مسؤولة عن مراقبة الأنظمة واكتشاف الهجمات وتحديد الهجمات التي يمكن أن تأتي للنظام وتسبب له ضرراً، وبالتالي يمنع الوصول غير المصرح به الى الأنظمة عن طريق اصدار تنبيه للمسؤول قبل التسبب بأضرار كبيرة لا يمكن تصحيحها.[2]

بالمقارنة مع طرق تعلم الآلي التقليدية، تعد الشبكات العصبونية من أحدث التقنيات المستخدمة في مجال كشف الاختراقات الأمنية لما تبديه من دقة وسرعة وسهولة في عملية كشف الهجومات، كما أنها مناسبة من أجل عملية كشف الهجمات على الشبكات. في حين أن طرق التعلم الآلي التقليدية تعاني من الدقة المنخفضة وبعض الأخطاء في عملية التصنيف .[3]مازالت الأبحاث جارية في مجال الشبكات العصبونية بهدف الحصول على أفضل بنية للشبكة العصبونية تساهم في عملية تصنيف الهجومات بدقة عالية.

1-1 الدراسات المرجعية:

قامت العديد من الأبحاث باستخدام تقنيات الذكاء الاصطناعي في عملية كشف الاختراقات على قاعدة البيانات الكبيرة KDD-99.

الدراسة الأولى: Hadoop based parallel binary bat algorithm for network intrusion detection

اقترح الباحثون خوارزمية محسنة لاستخلاص الميزات بالاعتماد على خوارزمية Bat Binary وتم استخدام مصنف Naïve Bayes لكشف الاختراقات، تم تطبيق الخوارزمية على 10% من قاعدة البيانات KDD-99.استخدمت هذه الخوارزمية لزيادة كفاءة عملية اختيار الميزات وتحسين معدل الكشف، وتم تقييم النموذج من خلال: Detection Rate (DR). وتم التوصل إلى أن معدل الكشف في الدراسة المقترحة هو الأعلى إذ أن معدل الكشف تم تحسينه وزمن الكشف تم تقليله.[1]

الدراسة الثانية:A Machine Learning Approach for Intrusion Detection System on NSL-KDD Dataset

في هذه الدراسة قام الباحثون باستخدام عدة تقنيات للتصنيف بهدف تصنيف البيانات الى بيانات طبيعية وبيانات هجوم من هذه التقنيات المستخدمة SVM، ETC، DT، KNN، LR، NB، RF وغيرها من التقنيات. تم تحليل أداء النموذج على أربع مجموعات فرعية مستخرجة من مجموعة البيانات المعتمدة في هذا البحث وهي NSL-dataset. تمت عملية المعالجة المسبقة للبيانات للحصول على السمات الأكثر صلة وتوصل النموذج المقترح إلى أن أداء ETC و RF و DT كان أعلى من التقنيات الأخرى ومن اجل كل الفئات.[4]

الدراسة الثالثة:Research of intrusion detection algorithm based on parallel SVM

اقترح الباحثون خوارزمية تصنيف تفرعيه تعتمد على خوارزمية SVM بالاعتماد على التعلم المتكامل على 10% من قاعدة البيانات PCA.KDD-99 استخدمت لتحليل البيانات واستخلاص الميزات لتقليل الأبعاد بالاعتماد على التعلم المتكامل. هنا يتكون المصنف من عدة مصنفات فرعية يتم اختيار مجموعة التدريب لكل مصنف فرعي بشكل عشوائي من مجموعة التدريب الأصلية، ويسمح باختيار عينات تدريب متكررة. تم بناء النظام لكشف أربع أنواع من الهجمات ومقارنة النتائج من حيث الدقة والزمن على ثلاث طرق: SP-PCA-، Parallel SVM،PCA-SVM SVM.[5]

الدراسة الرابعة:Big Data: controlling fraud by using machine learning libraries on Spark

قام الباحثون باستخدام تقنيات تعلم الآلة لكشف الاحتيال تم استخدام تقنية ال clustering واستخدام خوارزمية ال K-means. الدراسة عبارة عن تطبيق برمجي لكشف أي من حركة الشبكة هو عادي أو يدل على الاحتيال. تم استخدام 400 ألف حركة من حركات الشبكة مأخوذة من قاعدة بيانات ضخمة. في هذا البحث لم يتم استخدام استخلاص الميزات لتحديد الميزات الأكثر صلة بالعمل تم التوصل إلى أن spark أكثر ملائمة من ال Hadoop عند العمل مع الخوارزميات التكرارية مثل K-Means .[6]

الدراسة الخامسة: Intrusion detection model using machine learning algorithm on Big Data environment

قدم الباحثون نظاماً لاكتشاف الاختراقات الهدف منه مراقبة وتحليل البيانات لاكتشاف أي اختراق في النظام اذ تم اقتراح طريقة تصنيف تسمى Spark-Chi-SVM وأثبتت النتائج أن المصنف المقترح يقلل وقت التدريب ومناسب للبيانات الكبيرة[7]

2. أهمية البحث وأهدافه

يقدم البحث مساهمة جديدة في مجال كشف الاختراقات الأمنية في الشبكات الحاسوبية من حيث دراسة تأثير استخلاص السمات على دقة كشف الاختراقات الأمنية وذلك بهدف اكتشاف السلوكيات غير الطبيعية في الشبكات وتحديد نوع هذه السلوكيات، وتكمن أهمية هذا البحث في القيام بعدة تجارب توضح أثر استخلاص الميزات على دقة الكشف من خلال استخدام طرق استخلاص الميزات وشرحها بالتفصيل مع توضيح آلية حساب قيم الدقة بالاعتماد على مصفوفة الارتباك . وبالتالي يهدف هذا البحث إلى تصنيف سجلات قاعدة البيانات إلى سجلات هجوم وسجلات طبيعية ولكن بأعلى دقة ممكنة من خلال القيام بعدة تجارب وحساب الدقة في كل تجربة.

3. طرائق البحث ومواده

تم الاعتماد في هذا البحث على قاعدة البيانات المشهورة KDD99، وهي عبارة عن مجموعة بيانات قياسية تم توليدها عن طريق محاكاة بيئة ضمن شبكة بيانات ضخمة، حيث تضم بيانات طبيعية وغير طبيعية .تم استخدام الشبكات العصبونية في عملية تصنيف البيانات، إذ تعد الشبكات العصبونية من أهم التقنيات المستخدمة في التصنيف نظراً للنتائج الدقيقة التي تحققها في عملية التصنيف وبوقت قصير. تم اختبار طرق استخلاص السمات على قاعدة البيانات المستخدمة ومقارنة دقة كل طريقة باستخدام مصفوفة الارتباك< تم استخدام برنامج الماتلاب في عملية التصنيف.

3-1-مجموعة البيانات Dataset (KDD-99)

تتوفر المجموعات الخاصة بأمن الشبكات بطريقتين،الأولى من برامج مراقبة الحزم مثل Tcpdump و WinDump وغير ذلك،لكن هذه البيانات غير مصنفة وتستغرق وقت في عملية التصنيف وبالتالي غير مناسبة لأغراض النمذجة لكن يمكن أن تخدم غرض التحقق من صلاحية الوقت الذي يضمن متانة النموذج.

الطريقة الثانية هي استخدام مجموعات بيانات مفتوحة المصدر والمتاحة للتنزيل المجاني فهي توفر الوقت في الحصول على البيانات وتزيد من كفاءة البحث لأنها تتطلب تنظيفا أقل كما أنها مصنفة وبالتالي مناسبة لمصمم النماذج،من هذه المجموعات DARPA و KDD-99 و NSL و ADFA وغير ذلك.[8]

في هذا البحث تم استخدام مجموعة البيانات KDD-99 تضم هذه المجموعة 41 سمة بالإضافة إلى سمة ال class التي تعبر عن سجل الاتصال هل هو سجل طبيعي أم سجل هجوم، كما تتدرج هذه السجلات تحت أربع أنواع رئيسة من الهجومات تتدرج تحتها 22 نوع هجوم فرعي .

في الواقع ، يمكن تقسيم حركة مرور الشبكة إلى فئتين (حركة المرور العادية وحركة المرور الضارة). علاوة على ذلك ، يمكن أيضًا تقسيم حركة مرور الشبكة إلى خمس فئات: عادي Normal ، و DoS (هجمات رفض الخدمة) ، و R2L (هجمات الجذر إلى المحلية) ، و U2R (هجوم المستخدم إلى الجذر) ، والتحقيق (هجمات الاستكشاف). وبالتالي يمكن اعتبار اكتشاف الإختراقات على أنها مشكلة تصنيف. من خلال تحسين أداء المصنفات في تحديد حركة المرور الضارة بشكل فعال ، يمكن تحسين دقة اكتشاف التسلل إلى حد كبير.[9]

تتضمن مجموعة البيانات أربعة أنواع رئيسية من الهجومات يندرج تحت كل نوع من هذه الأنواع هجومات فرعية :

1- هجوم حجب الخدمة Denial of service(DOS) : محاولة جعل الجهاز غير متاح لمستخدميه.
2- هجوم مستخدم إلى جذر User to Root(U2R) :استغلال المهاجم للنظام الذي يبدأ بحساب طبيعي يحاول من خلاله الحصول على امتيازات مستخدم رئيسي
3- هجوم بعيد إلى محلي Remote to Local (R2L) :يستغل المهاجم ميزات جهاز محلي من خلال ارسال حزم عبر الانترنت إلى جهاز لا يملك الوصول اليه بهدف استغلال نقاط ضعف الأجهزة .
4- هجوم التحقق Probe :الهدف منه تعريض النظام للخطر من خلال قيام المهاجم بمسح جهاز شبكي لتحديد نقاط الضعف من اجل استغلالها فيما بعد.

3-2 استخلاص الميزات:

تلعب المعالجة المسبقة دورًا مهمًا يتم من خلاله إزالة الميزات غير الرقمية / الفئوية أو استبدالها باستخدام تقنيات معينة لاستخراج / الميزات. تقلل المعالجة المسبقة من التعقيد الحسابي وتزيد من معدل تصنيف النموذج.[10]

هو عملية تقليل عدد متغيرات الدخل عند تطوير نموذج تنبؤي، توفر لنا طرق استخلاص الميزات طريقة لتقليل وقت الحساب، وتحسين أداء التنبؤ، وفهم أفضل للبيانات في التعلم الآلي أو التعرف على الأنماط[11] الميزة هي خاصية فردية قابلة للقياس للعملية التي تتم ملاحظتها. باستخدام مجموعة من الميزات يمكن لأي خوارزمية للتعلم الآلي إجراء التصنيف[7].الهدف من اختيار الميزات في التعلم الآلي هو العثور على أفضل مجموعة من الميزات التي تسمح للشخص ببناء نماذج مفيدة للظواهر المدروسة. يتم اختيار طرق استخلاص الميزات بالاعتماد على نوع التعلم

المستخدم سواء كان تعلم رقابي أم تعلم غير رقابي، لكل منها تقنيات خاصة به بالإضافة لذلك يتم اخذ نوع البيانات بعين الاعتبار في حال كانت رقمية ام فئوية.

يمكن تصنيف تقنيات اختيار الميزات في التعلم الآلي على نطاق واسع إلى الفئات التالية:

1- طرق التصفية Filter Methods: تستخدم في حال كانت المصنفات من نوع التعلم الرقابي تعد هذه الطرق أسرع وأقل تكلفة من الناحية الحسابية مقارنة مع الطرق الأخرى، وهي مناسبة عند التعامل مع بيانات ذات عدد سمات كبير.

يندرج تحت هذا النوع من الطرق عدة تقنيات يتم اختيار التقنية المناسبة حسب نوع البيانات التي يتم التعامل معها. يمكن أن نميز نوعين من البيانات: بيانات فئوية categorical تضم الأحرف والرموز والأرقام مع أحرف(1st) بالإضافة إلى القيم البوليانية(true, false) ، النوع الثاني من البيانات هو البيانات الرقمية Numeric تضم الأرقام الصحيحة والحقيقية.

يمثل الشكل التالي آلية عمل هذا النوع من التقنيات عند تطبيقها من اجل أغراض التصنيف، إذ يتم أخذ مجموعة جزئية من السمات هي الأفضل من بين مجموعة السمات الكلية ثم إخضاعها إلى خوارزمية تعلم ثم تقييم الآداء.

الشكل(1) :مراحل عمل طرق التصفية

تقنيات استخلاص الميزات من النوع Filter methods :

a) Chi-square Test: مناسب من اجل البيانات الفئوية يتم حسابه بين كل سمة والهدف ونختار السمات ذات افضل نتيجة لقيمة chi-square. يعطى بالعلاقة:

$$X = \sum \frac{(opserved\ value - expected\ value)}{expected\ value} \quad \ldots\ldots\ldots (1)$$

b) معامل الارتباط هو مقياس للعلاقة الخطية بين اثنين أو اكثر من المتغيرات يمكن من خلاله التنبؤ بواحد من السمات من خلال السمة الأخرى.

عند ارتباط متغيرين يمكن توقع أحدهما من الآخر وبالتالي إذا كانت السمتان مترابطتان يحتاج النموذج واحدة منهم فقط لأن السمة الثانية لا تضيف أي ميزات أو معلومات إضافية عن السمة الأولى. العلاقة الرياضية التي تعبر عن معامل الترابط:

$$Corr(x,y) = \frac{Cov(x,y)}{\sigma x.\sigma y} \quad (2)$$

حيث : σx:الانحراف المعياري ل x ، σy:الانحراف المعياري ل y ، cov(x,y) :التباين

حيث أن الانحراف المعياري هو الجذر التربيعي للتباين.

c) Variance التباين هو أحد مقاييس مقياس التشتت الإحصائي بين القيم لعينة ما ،يقيس مقدار تشتت القيم عن الوسط الحسابي وعن بعضها البعض. إذا كانت قيمة التباين كبيرة يعني ذلك ان القيم متباعدة عن بعضها البعض وعن الوسط الحسابي وفي المقابل إذا كانت قيمته صغيرة هذا يعني أن القيم متقاربة من بعضها البعض ومن الوسط الحسابي أما إذا كانت قيمته صفر يعني ذلك أن القيم متماثلة. العلاقة الرياضية التي تعبر عن التباين:

$$Cov(x) = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{xi} - \mathrm{X}) \qquad (3)$$

حيث: X :المتوسط الحسابي لقيم x_i ، n: عدد العينات

d) Main Absolute Difference (MAD): متوسط الانحراف المطلق هو مقياس لمتوسط المسافة المطلقة بين كل قيمة بيانات ومتوسط مجموعة البيانات< يعطى بالعلاقة:

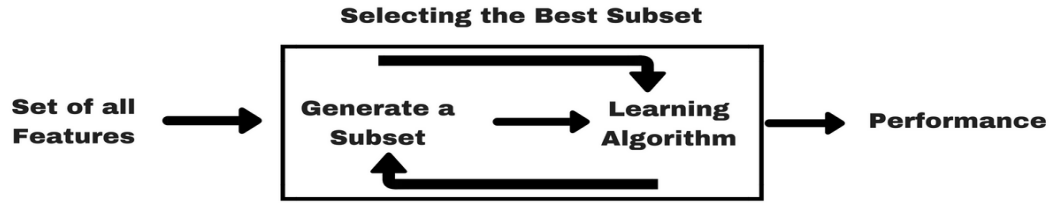
$$MAD = \frac{\sum |xi - X|}{n} \qquad (4)$$

حيث n عدد العينات

e) Dispersion Ratio: نسبة التشتت تدل على نسبة المتوسط الحسابي AM لعينة ما على المتوسط الهندسي GM لنفس العينة. المتوسط الحسابي هو مجموع قيم العينة على عددها في حين أن المتوسط الهندسي هو الجذر من المرتبة n لجداء قيم العينة. تتراوح نسبة التشتت من 1 إلى اللانهاية، وتدل نسبة التشتت الأعلى على أن الميزة أكثر صلة وأكثر ارتباطا.

2- طرق التغليف Wrapper Methods: تناسب هذه الطرق خوارزميات التعلم غير الرقابي، في طرق التجميع، يتم استخدام مجموعة فرعية من الميزات وتدريب النموذج على استخدامها. استنادًا إلى الاستدلالات التي تم استخلاصها من النموذج السابق، تقرر إضافة الميزات أو إزالتها من المجموعة الفرعية. يتم تقليل المشكلة بشكل أساسي إلى مشكلة بحث. عادة ما تكون هذه الطرق مكلفة للغاية من الناحية الحسابية.

يوضح الشكل آلية عمل طرق التغليف، يتم توليد مجموعة حزئية من مجموعة السمات الكلية وإخضاعها الى خوارزمية تعلم و إعادة هذه المرحلة عدة مرات للحصول على أفضل السمات ثم تقييم النموذج.



الشكل (2):آلية عمل طرق التغليف

بعض طرق التغليف المستخدمة في استخلاص الميزات:

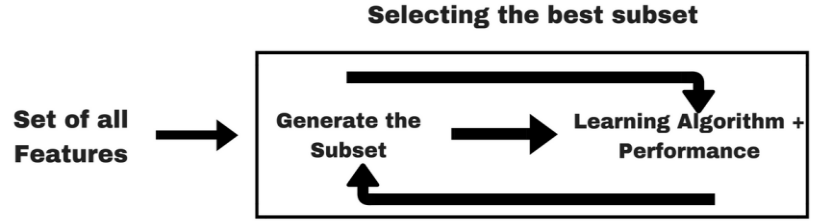
a. Forward Selection: هو طريقة تكرارية يتم البدء بها بدون وجود أي ميزة في النموذج. في كل تكرار، نستمر في إضافة الميزة التي تعمل على تحسين نموذجنا بشكل أفضل حتى لا تؤدي إضافة متغير جديد إلى تحسين أداء النموذج.

b. Backward Elimination: تبدأ بجميع الميزات وإزالة الميزة الأقل أهمية في كل تكرار مما يحسن أداء النموذج. يتم تكرار ذلك حتى يتم ملاحظة عدم تحسن في إزالة الميزات.

c. Recursive Feature elimination: تهدف إلى العثور على أفضل مجموعة فرعية من الميزات أداءً. يقوم بإنشاء نماذج بشكل متكرر ويتجاهل الميزة الأفضل أو الأسوأ أداءً في كل تكرار. يقوم ببناء النموذج التالي حتى يتم استنفاد جميع الميزات. ثم يقوم بترتيب الميزات بناءً على ترتيب إزالتها.

3- الطرق المضمنة Embedded Methods: تجمع الطرق المضمنة بين صفات طرق التصفية والغلاف. يتم تنفيذه من خلال الخوارزميات التي لها طرق اختيار الميزات المضمنة الخاصة بها.

يمثل الشكل آلية عمل الطرق المضمنة يتم فيها توليد مجموعة جزئية من مجموعة السمات الكاملة ثم تقييم أداء خوارزمية التعلم وتكرار هذه الخطوة حتى نحصل على مجموعة السمات الأفضل والأكثر ارتباطا.

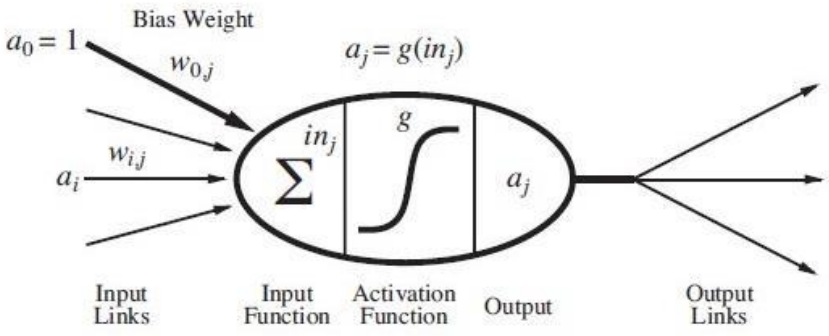


الشكل(3) :آلية عمل الطرق المضمنة

3-4 الشبكات العصبونية

تُعرَّف الشبكة العصبونية الصنعية بأنها نظام لمعالجة البيانات بشكل يحاكي ويشابه الطريقة التي تتبعها الشبكات العصبية الطبيعية عند الإنسان. تحتوي الشبكة العصبونية على عدد كبير من العناصر الصغيرة لمعالجة المعلومات تسمى الخلية العصبية أو العصبون، لها المقدرة على الاستجابة لإشارة الدخل والتعلم لتتلاءم مع الوسط المحيط وتعطي الخرج المناسب،وتصنف الشبكات العصبونية حسب نموذج الوصل بين العصبونات إلى شبكات الطبقة المفردة وشبكات متعددة الطبقات وهذا يسمى البنية، أو حسب طريقة تعيين الأوزان المرافقة للوصلات إلى التعليم بوجود مشرف والتعليم مع عدم وجود مشرف وهذا يسمى التدريب أو التعليم. تعتبر عملية التصنيف والترميز أبرز نشاط تقوم به الشبكات العصبونية. [12]

أثبتت الشبكات العصبونية قدرتها على حل كثير من المشاكل ضمن مسائل عديدة وحقول متنوعة.



الشكل(4) نموذج الشبكة العصبونية

3-5 الخوارزمية المقترحة:

الخطوات المتبعة:

1- تحميل مجموعة البيانات على برنامج الإكسل وفحصها بالتأكد من عدم وجود خلايا فارغة أو سجلات مكررة ومعالجة النقص في حال وجوده.

2- تحميل مجموعة البيانات على برنامج الماتلاب .

3- اختيار الشبكة العصبونية المناسبة بالاعتماد على التجريب بتغيير عدد الطبقات والعصبونات بكل طبقة ومعامل التعلم بالإضافة إلى تابع النقل للحصول في النهاية على بنية بمعامل خطأ صغير نسبيا.

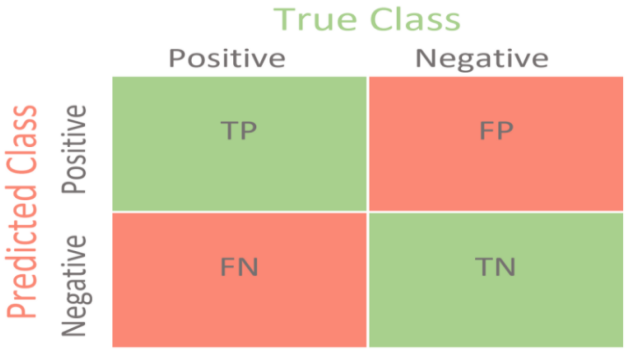
4- تطبيق طرق استخلاص السمات وحساب الدقة في كل طريقة.

5- مقارنة نتيجة طرق استخلاص الميزات ومناقشة النتائج.

3-6 مصفوفة الارتباك (أو مصفوفة الدقة)

❖ مصفوفة الارتباك (confusion matrix): تعد مصفوفة الارتباك من أهم الوسائل المستخدمة في تقييم المصنفات، إذ يتم تقييم المصنف من خلال قدرته على التصنيف الصحيح أي القدرة على تحديد نوع الصنف الذي ينتمي إليه سجل الاتصال أكان طبيعياً أم هجوماً، وعند مقارنة نتيجة التصنيف مع الواقع الفعلي نجد أربع حالات مختلفة [2016]:

- الإيجابيات الصحيحة True Positive (TP): الحدث إيجابي وتم التنبؤ أن الحدث إيجابي. (صحيح)
- الإيجابيات الخاطئة False Positive (FP): الحدث سلبي وتم التنبؤ أن الحدث إيجابي. (خطأ)
- السلبيات الخاطئة False Negative (FN): الحدث إيجابي وتم التنبؤ أن الحدث سلبي. (خطأ)
- السلبيات الصحيحة True Negative (TN): الحدث سلبي وتم التنبؤ أن الحدث سلبي. (الشواذ)



الشكل(5) :حالات مصفوفة الارتباك

من خلال هذه الحالات يتم حساب القيم التالية:

الدقة Accuracy: المقياس الأكثر شيوعا لتقييم المصنف، يقيم كامل الخوارزمية ،يعطى بالعلاقة التالية[2015]:

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad \ldots\ldots\ldots\ldots\ldots\ldots(5)$$

معدل الخطأ Error Rate:

$$Error\ Rate = 1 - \text{Accuracy} \ldots\ldots (6)$$

4.النتائج والمناقشة

تم استخدام مجموعة البيانات KDD-99، هي قاعدة بيانات كبيرة مكونة من 42 سمة وحوالي اربع ملايين سجل بيانات بصيغة txt، تم استخدام 25% من هذه القاعدة في هذا البحث، عدد السجلات المستخدمة 1100623 سجل مندرجة تحت اربع هجومات أساسية بشكل عام و23 هجوم فرعي.

تم التعبير عن كل سمة برقم حسب ترتيبها في مجموعة البيانات ،وفيما يلي شرح بعض سمات مجموعة البيانات :

السمة 1 : Duration عدد الثواني في الاتصال، السمة 2 : Protocol_type نوع البروتوكول ...tcp,udp، السمة 3 : Service خدمة الشبكة مثل ...http,telnet ، السمة 4 : Src_bytes عدد بايتات البيانات من المصدر إلى الوجهة ، السمة 5 : Dst_bytes عدد بايتات البيانات من الوجهة إلى المصدر ، السمة 6 : Flag حالة الاتصال طبيعي أم خاطئ.

في الشكل جزء من بيانات مجموعة البيانات كل سطر يدل على بيانات اتصال مع تحديد نوعها هل هي هجوم ام سجل طبيعي.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP
1	0	2	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	229	10	0	0	1	1	0.04	0.06	0	255	10	0.04	0.06	0	0	0	0	1	1	neptune
2	0	2	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	136	1	0	0	1	1	0.01	0.06	0	255	1	0	0.06	0	0	0	0	1	1	neptune
3	2	2	7	8	12983	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	134	86	0.61	0.04	0.61	0.02	0	0	0	0	normal
4	0	1	11	8	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	65	0	0	0	0	1	0	1	3	57	1	0	1	0.28	0	0	0	0	saint
5	1	2	1	2	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	8	0	0.12	1	0.5	1	0	0.75	29	86	0.31	0.17	0.03	0.02	0	0	0.83	0.71	guess_passwd
6	0	2	2	8	267	14515	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	4	4	0	0	0	0	1	0	0	155	255	1	0	0.01	0.03	0.01	0	0	0	normal
7	0	2	3	8	1022	387	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	1	255	28	0.11	0.72	0	0	0	0	0.72	0.04	normal
8	0	2	1	8	129	174	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	255	255	1	0	0	0	0.01	0.01	0.02	0.02	guess_passwd
9	0	2	2	8	327	467	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	33	47	0	0	0	0	1	0	0.04	151	255	1	0	0.01	0.03	0	0	0	0	normal
10	0	2	4	8	26	157	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	52	26	0.5	0.08	0.02	0	0	0	0	0	guess_passwd
11	0	2	1	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	255	128	0.5	0.01	0	0	0	0	0.66	0.32	guess_passwd
12	0	2	3	8	616	330	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	1	255	129	0.51	0.03	0	0	0	0	0.33	0	normal
13	0	2	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	111	2	0	0	1	1	0.02	0.07	0	255	2	0.01	0.07	0	0	0	0	1	1	neptune
14	0	2	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	120	120	1	1	0	0	1	0	0	235	171	0.73	0.07	0	0	0.69	0.95	0.02	0	neptune
15	37	2	1	8	773	364200	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	38	73	0.16	0.05	0.03	0.04	0	0.77	0	0.07	normal
16	0	2	2	8	350	3610	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	8	8	0	0	0	0	1	0	0	71	255	1	0	0.01	0.04	0	0	0	0	normal
17	0	2	2	8	213	659	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	24	24	0	0	0	0	1	0	0	255	255	1	0	0	0	0	0	0	0	normal
18	0	2	2	8	246	2090	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	16	16	0	0	0	0	1	0	0	35	255	1	0	0.03	0.05	0	0	0	0	normal
19	0	3	5	8	45	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	505	505	0	0	0	0	1	0	0	255	255	1	0	1	0	0	0	0	0	normal
20	0	2	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	204	18	0	0	1	1	0.09	0.07	0	255	18	0.07	0.07	0	0	0	0	1	1	neptune
21	0	2	13	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	118	19	0	0	1	1	0.16	0.05	0	255	19	0.07	0.05	0	0	0	0	1	1	neptune

الشكل(6) :جزء من مجموعة البيانات KDD

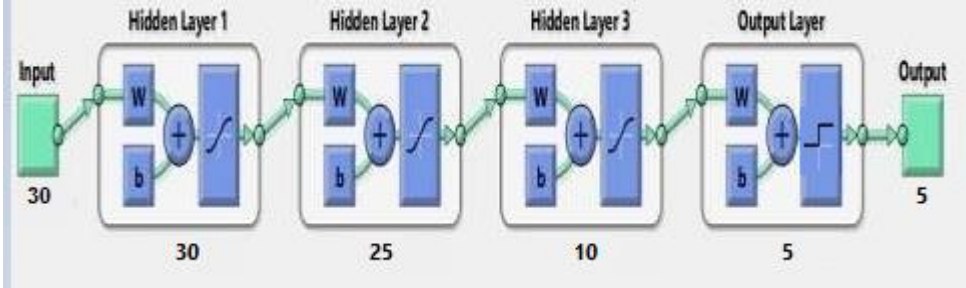
تتم نمذجة الشبكة العصبونية الصنعية في بيئة ماتلاب بإدخال بيانات التدريب، تمثل هذه البيانات مدخلات الشبكة المتمثلة بشعاع السمات الخاص بسجل الاتصال (30 سمة) والقيم المرغوبة للخرج المتمثلة بنتيجة التصنيف لسجل الاتصال. لاختيار عدد الطبقات المناسب وعدد العصبونات المناسبة تم إجراء عدة تجارب على الشبكة العصبونية ومن أجل قاعدة البيانات السابقة.تم في كل تجربة اختيار عدد طبقات محدد وعدد عصبونات محددة، وفي كل مرة تم حساب معامل الخطأ، تم البدء بطبقة خفية واحدة مع تغيير عدد العصبونات ضمنها ثم حساب معامل الخطأ ثم طبقتين وهكذا ، خطأ كما تم أخذ معامل التعلم بعين الاعتبار وتم تغيير قيمته أثناء التجريب.

بناء على ما سبق وباختيار عدة سيناريوهات لتدريب الشبكة العصبونية باستخدام تابع التدريب الممثل لخوارزمية الانحدار التدريجي للخطأ ذات معامل معدل التعلم المتغير القيمة، نجد أن هيكلية الشبكة العصبونية المختارة في البحث هي الموافقة للسيناريو التالي والتي تمتلك القيم الأقل لمعامل الخطأ في مرحلة التدريب والاختبار كما هو مبين في التركيبة التالية:

- طبقة الدخل: مكونة من شعاع السمات الخاص بسجل الاتصال 30 سمة، ثلاث طبقات خفية، عدد عصبونات الطبقة الخفية الأولى 30، أما عدد عصبونات الطبقة الخفية الثانية 25، 10 عصبون في الطبقة الخفية الثالثة، طبقة الخرج نتيجة التصنيف لسجل الاتصال تضم 5 عصبونات.

إن الاستمرار بزيادة عدد الطبقات الخفية وزيادة عدد العصبونات فيها سيؤدي إلى تحقيق قيمة أصغرية لمتوسط مربع الخطأ ولكنه يزيد من حجم وتعقيد الشبكة، لذلك تم اعتبار ان السيناريو السابق يحقق اختيار

الهيكلية الأفضل للشبكة العصبونية المختارة في البحث. يبين الشكل هيكلية الشبكة العصبونية النهائية المستخدمة في البحث.



الشكل (7) الشبكة العصبونية المستخدمة

باعتماد السيناريو السابق الذي تم التوصل إليه تم إجراء عدة تجارب على الشبكة العصبونية توضح تأثير استخلاص الميزات على عملية التصنيف ومدى تأثر دقة الكشف بذلك بهدف اختيار التقنية المناسبة لاستخلاص السمات. إذ تم الاعتماد على التعليمات البرمجية في بيئة الماتلاب لحساب القيم المذكورة . تفيد عملية استخلاص الميزات في الحصول على السمات المفيدة والأكثر ترابطا أو الأكثر تأثيرا بالنظام،وبالتالي تقليل مساحة التخزين ووقت المعالجة من جهة بالإضافة إلى الحصول على أفضل دقة من جهة أخرى.

التجربة الأولى: تم تدريب مصنف الشبكة العصبونية على مجموعة السمات كاملة (41 سمة) وعرض نتائج دقة المصنف بهدف المقارنة لاحقا من اجل كل عملية تغيير في عدد السمات.

التجربة الثانية: تدريب مصنف الشبكة العصبونية على مجموعة السمات بعد عملية حساب التباين للسمات (39) سمة. تم التوصل إلى أن السمتين 20 و 21 لها تباين صفري أي لاتتغير قيمتها من أجل كل السجلات في قاعدة البيانات وبالتالي يمكن الاستغناء عنهما مما يقلل من حجم قاعدة البيانات دون التأثير على دقة التصنيف، وأصبح عدد السمات بعد هذه المرحلة 39 سمة. كلا السمتين قيمتهما صفر على طول قاعدة البيانات ومن اجل كل السجلات وبالتالي حذفهما لا يؤثر على عملية التصنيف.

التجربة الثالثة: تدريب مصنف الشبكة العصبونية على مجموعة السمات بعد عملية حساب معامل الترابط بين السمات (30 سمة).

وجدنا عدة سمات مرتبطة ببعضها البعض وبالتالي يمكن الأستغناء عن إحداها والإبقاء على الأخرى مثلا السمة 25 هي النسبة المئوية لعدد مرات إعادة الخطأ مرتبطة بشكل كبير مع السمة 26 التي تعبر عن النسبة المئوية لعدد الاتصالات لنفس الخدمة ونفس المضيف أي وجود قيمة لأحد هذه السمتان يغني عن وجود السمة الأخرى بمعرفة عدد مرات حصول فشل بالاتصال يمكننا من معرفة عدد مرات الاتصال الناجح، ونفس الامر ينطبق على كل السمات المترابطة.

التجربة الرابعة :تدريب مصنف الشبكة العصبونية على مجموعة السمات بعد عملية حساب متوسط الانحراف المطلق MAD (29 سمة).

التجربة الخامسة : تدريب مصنف الشبكة العصبونية على مجموعة السمات بعد عملية حساب نسبة التشتت(30 سمة)

كانت النتائج كما هو موضح بالجدول التالي:

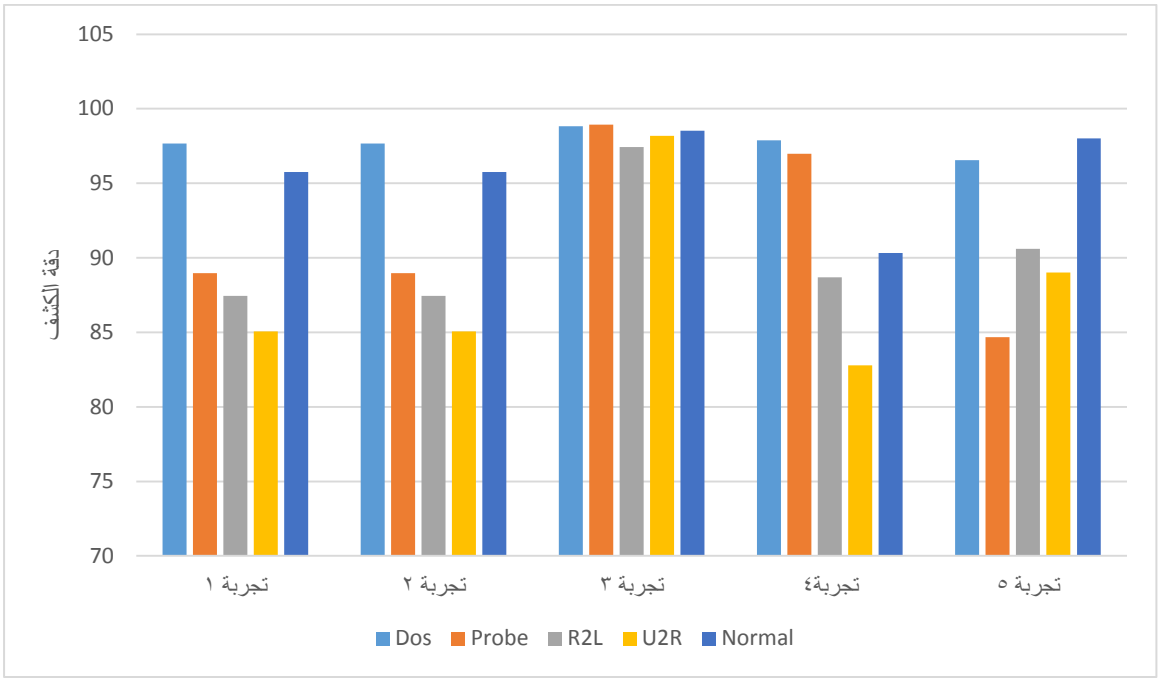
جدول (1) نتائج دقة الكشف في التجارب الخمسة

	عدد السمات	Dos	Probe	R2L	U2R	Normal
تجربة 1	41	97.66%	88.97%	87.45%	85.06%	95.76%
تجربة 2	39	97.66%	88.97%	87.45%	85.06%	95.76%
تجربة 3	30	98.82%	98.93%	97.43%	98.19%	98.53%
تجربة 4	29	97.89%	96.98%	88.69%	82.78%	90.33%
تجربة 5	30	96.55%	84.67%	90.60%	89.01%	98.02%

يوضح الجدول (1) نتائج دقة الكشف التي تم الحصول عليها بعد تطبيق إحدى تقنيات استخلاص الميزات ومن أجل الهجومات الأربع الأساسية بالإضافة إلى دقة كشف السجلات الطبيعية normal . هذه النتائج تم الحصول عليها من مصفوفة الارتباك الناتجة عن محاكاة الشبكة العصبونية ،إذ تم تحليلها وإجراء الحسابات اللازمة وتنظيمها في الجدول السابق. نلاحظ ان نتائج التجربتين 1و2 متماثلة وهذا يؤكد عدم مساهمة السمتين 20 و21 في عملية التصنيف حيث أعطت قيم صفرية للتباين.السمة (20) هي Num_outbound_cmds : عدد الأوامر الصادرة في جلسة بروتوكول ftp والسمة (21) هي Is_hot_login: تأخذ قيمة 1 اذا تم تسجيل الدخول الى hot list و0 فيما عدا ذلك .

كلا السمتين قيمتهما صفر على طول قاعدة البيانات ومن اجل كل السجلات وبالتالي حذفهما لا يؤثر على عملية التصنيف.

أما في التجربة 3 بعد حساب قيم معامل الترابط للسمات والإبقاء على السمات ذات الترابطات الأقوى نلاحظ تحسن دقة التصنيف بشكل واضح مقارنة مع التجربة 1 حيث تم استخدام كامل السمات اما في التجربة 4 بعد حساب قيمة متوسط الانحراف المطلق نلاحظ ان السمات اقل من كل التجارب السابقة واعطت دقة عالية لكن تجربة 3 دقتها اعلى ، اما التجربة 5 تم تقييم المصنف على 30 سمة كما التجربة 3 لكن الدقة في التجربة 3 اعلى منها من التجربة 5، *وبالتالي تكون دقة المصنف في التجربة 3 هي الأفضل وهو المصنف المعتمد لحساب وتقييم النتائج. فيما يلي مقارنة بين التجارب السابقة من حيث الدقة:*

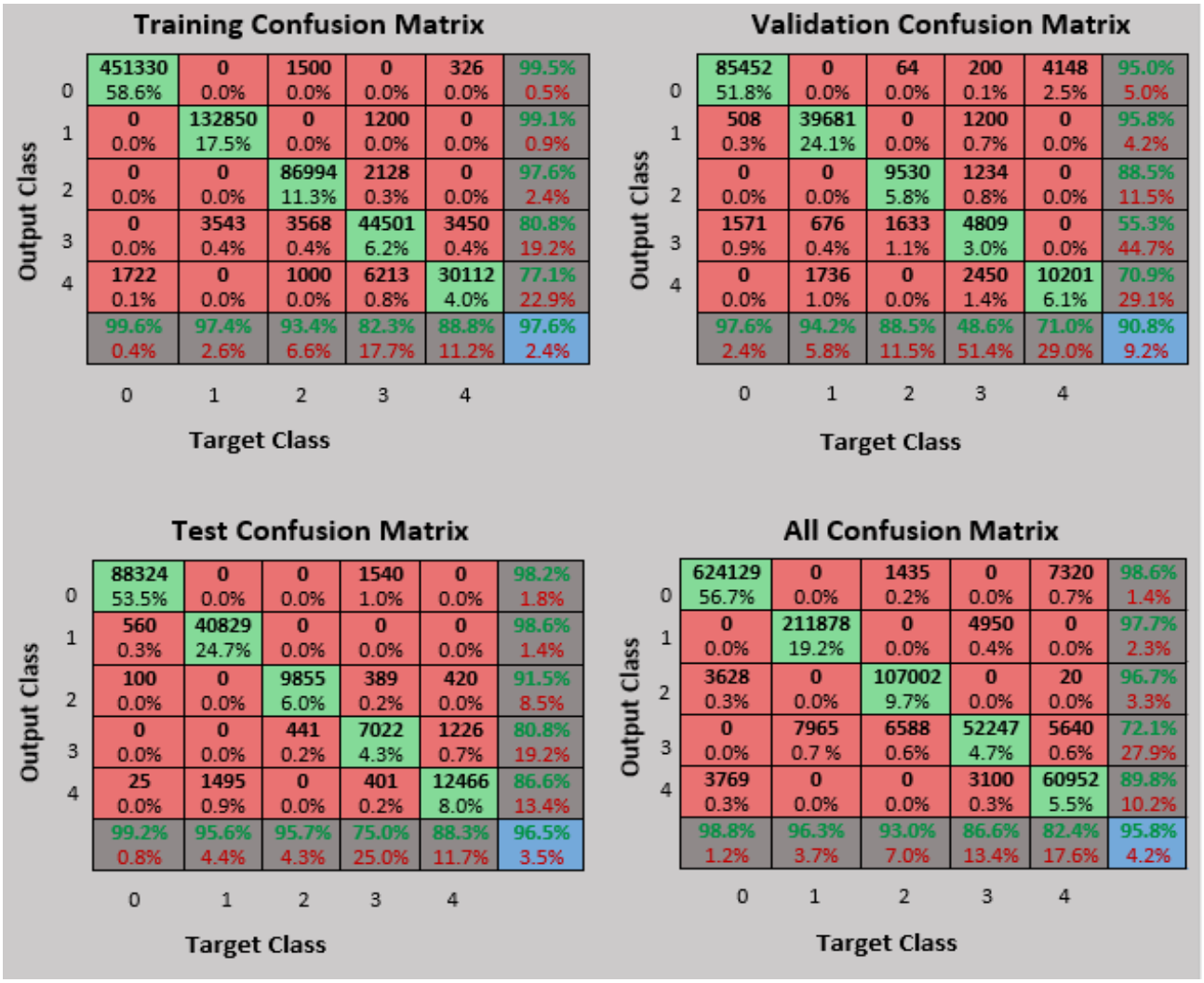


الشكل (8): مخطط بياني يوضح مقارنة بين التجارب الخمسة

فيما يلي تفصيل آلية الحصول على قيم الدقة باعتماد السيناريو السابق الذي تم التوصل إليه كمصنف للبيانات ضمن هذا البحث، تم إجراء عملية التصنيف على قاعدة البيانات 1,100,623 سجل، وحساب دقة الكشف تم بالاعتماد على مصفوفة الارتباك، تم تقسيم البيانات إلى 3 مجموعات وفق التالي:

مجموعة التدريب 70% تتضمن 770623 سجلاً ، مجموعة التحقق 15%تتضمن 165093 سجلاً ، مجموعة الاختبار 15% تتضمن 165093سجلاً .

مصفوفة الدقة الناتجة من أجل الشبكة العصبونية السابقة ومن أجل 1100623 سجلاً حيث في هذه المرحلة تم تصنيف السجلات للهجومات الرئيسية الأربعة، حيث يمثل 0 السجل الطبيعي ويمثل 1سجل هجوم من النوع Dos ويمثل 2سجل هجوم من النوع Probe ويمثل3 سجل هجوم من النوع R2L ويمثل4 سجل هجوم من النوع U2R



الشكل (9) نتائج مصفوفة الدقة من اجل خمسة أصناف

من مصفوفة الارتباك الناتجة كانت النتائج كما هو موضح بالجدول التالي:

الجدول(2) نتائج مصفوفة الدقة

Confusion Matrix			Actual Class	
			yes	No
Predicted Class	Normal	yes	624129 (56.7%)	7397 (0.7%)
		no	8755 (0.8%)	460342 (41.8%)
	Dos	yes	211878 (19.2%)	7965 (0.7%)
		no	4950 (0.4%)	875830 (79.5%)
	Probe	yes	107002 (9.7%)	8023 (0.7%)
		no	3648 (0.3)	981950 (89.2%)
	R2L	yes	52247 (4.7%)	8050 (0.7%)
		no	20193 (1.8%)	1020133 (92.6%)
	U2R	yes	60952 (5.5%)	12960 (1.0%)
		no	6869 (0.6%)	1019822 (92.6%)

حساب عامل الدقة وفق العلاقة:

$$Accuracy(Normal) = \frac{TN + TP}{TN + TP + FN + FP} = \frac{624129 + 460342}{624129 + 460342 + 8755 + 7397}$$
$$= 0.9853 = \mathbf{98.53}\%$$

$$Accuracy(Dos) = \frac{211878 + 875830}{211878 + 875830 + 4950 + 7965} = 0.9882 = \mathbf{98.82}\%$$

$$Accuracy(Probe) = \frac{107002 + 981950}{107002 + 981950 + 3648 + 8023} = 0.9893 = \mathbf{98.93}\%$$

$$Accuracy(R2L) = \frac{52247 + 1020133}{52247 + 1020133 + 20193 + 8050} = 0.9743 = \mathbf{97.43}\%$$

$$Accuracy(U2R) = \frac{60952 + 1019822}{60952 + 1019822 + 6869 + 12980} = 0.9819 = \mathbf{98.19}\%$$

من قيم الدقة يمكن حساب معامل الخطأ:

$$Error\ Rate = 1 - \text{Accuracy}$$
$$Error\ Rate(Normal) = 1 - 0.9853 = 0.0147$$
$$Error\ Rate(Dos) = 1 - 0.9882 = 0.0118$$
$$Error\ Rate(Probe) = 1 - 0.9893 = 0.0107$$
$$Error\ Rate(R2L) = 1 - 0.9743 = 0.0257$$
$$Error\ Rate(U2R) = 1 - 0.9819 = 0.0181$$

من النتائج السابقة نجد أن دقة التصنيف عالية كما أن معدلات الخطأ منخفضة جداً مما يدل على أن الشبكة العصبونية قامت بالتصنيف بشكل صحيح لأغلب السجلات.

4-الاستنتاجات والتوصيات:

تم في هذا البحث دراسة تأثير استخلاص السمات على دقة كشف الاختراقات الأمنية باستخدام الشبكات العصبونية، من خلال هذه الدراسة تم استنتاج مايلي:

- استخدم البحث الشبكات العصبونية في عملية التصنيف ، اذ تم بناء شبكة عصبونية يدويا بالاعتماد على التجريب وتم تجريب أكثر من شبكة اعتماداً على عدد العصبونات وعلى عدد الطبقات الخفية، وبالنهاية تم اعتماد الشبكة العصبونية الأفضل تبعاً لمعامل الخطأ الأقل.
- تمَّ دراسة تأثير استخلاص السمات على عمل ودقة الشبكة العصبونية، اذ تم إجراء خمس تجارب في كل تجربة تم استخدام طريقة من طرق استخلاص السمات وتم التوصل إلى أن التجربة الثالثة أفضل من حيث الدقة من باقي التجارب.
- توصلت الشبكة العصبونية المقترحة إلى معدل كشف 98.53% من أجل السجلات الطبيعية، ومعدل 98.82% بالنسبة لهجوم Dos،و 98.93% لهجوم probe، ومعدل 97.43% لهجوم R2L ومعدل 98.19% لهجوم U2L

ومن المستحسن: تطوير بنية الشبكة المستخدمة بحيث يتم زيادة عدد العصبونات الخفية وعدد الطبقات الخفية في الشبكة بهدف الحصول على أفضل دقة، أو تطويرها بجعلها تتمتع بخاصية التعميم اذ تصبح قادرة على كشف أي هجوم تتعرض له الشبكة وغير موجود ضمن مجموعة البيانات التدريبية المستخدمة، بالإضافة إلى استخدام طرق استخلاص ميزات مختلفة عن الطرق المستخدمة قد تساهم في تقليل زمن التصنيف وزيادة دقة الكشف

المراجع

[1] Natesan P, et al. 2017, *Hadoop based parallel binary bat algorithm for network intrusion detection*. Int J Parallel Program, Vol. 45, No.5,1194–213.

[2] Bilal.M; Ekhlas.K. G. 2021, *"Intrusion Detection System for NSL-KDD dataset based on deep learning and recursive feature elimination,"* Engineering and Technology Journal, Vol. 39, No. 07, pp. 1069-1079.

[3] Daradkeh, M؛.Abualigah, L.; Atalla, S.; Mansoor, W. 2022,*Scientometric Analysis and Classification of Research Using Convolutional Neural Networks: A Case Study in Data Science and Analytics*. Electronics, 11, 2066. https://doi.org/10.3390/ electronics11132066.

[4] Abrar.I;Ayub.Z;Masoodi.F,2020. *A Machine Learning Approach for Intrusion Detection System on NSL-KDD Dataset.* Conference on Smart Electronics and Communication (ICOSEC 2020) IEEE Xplore Part Number: CFP20V90-ART; ISBN: 978-1-7281-5461-9.

[5] Wang H, Xiao Y, Long Y. 2017, *Research of intrusion detection algorithm based on parallel SVM*. conference on electronics information and emergency communication (ICEIEC),. Piscataway, p. 153–156.(IEEE)

[6] Ferhat K, Sevcan A. Big Data: controlling fraud by using machine learning libraries on Spark. Int J Appl Math Electron Comput. 2018;6(1):1–5.

[7] Suad O , Fadel B,Nabeel A, Amal A . 2018. Intrusion detection model using machine learning algorithm on Big Data environment. Creative Commons Attribution 4.0 International License.http://creat iveco mmons .org/licen ses/by/4.0/.

[8] Rawat.Sh;Srinivasan.A;Ravi.V;Ghosh.U.2020. Intrusion detection systems using classical machine

learning techniquesA vs integrated unsupervised feature learning and deep neural network. wileyonlinelibrary.com/journal/itl2.

[9] SU.T;SUN.H;ZHU.J;WANG.SH;LI.A.2020. *BAT: Deep Learning Methods on Network Intrusion Detection Using NSL-KDD Dataset*.(IEEE)

[10] Masoodi.F;Bamhdi.A;Teli.T.2021.Machine learning for Classification analysis of Intrusion Detection on NSL-KDD Dataset.Turkish Journal of Computer and Mathematics Education. Vol.12,No.10,2286-2293.

[11] OZGUR, A.; et al. , April 14 2016, *A Review of KDD99 Dataset Usage in Intrusion Detection and Machine Learning between 2010 and 2015*. PeerJ Preprints.

[12] BEKKAR, M.; et al. 2013, *Evaluation Measures for Models Assessment over Imbalanced Data Sets*.Information Engineering and Applications, Vol.3 No.10,27-39.

[13] KDD Cup 1999 Data. https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
13/May/2021.