مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (10) العدد (10) 2026
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (10) No. (10) 2026

# نموذج هجين للكشف عن الاختراقات في شبكات الحوسبة السحابية باستخدام تقنيات الذكاء الصنعي

*** أ. د. م. فادي غصنه**

**** د. م. ناجي محمد**

***** م. رنيم يونس**

**(تاريخ الإيداع 2026/3/31 . قُبِل للنشر في 2026/5/18 )**

**ملخّص**

مع التطور السريع لتقنية الحوسبة السحابية، والتي سمحت بربط العديد من الأجهزة الذكية، تخزين وتبادل البيانات بين الأنظمة بسهولة أكبر، وتمكين الاستفادة من الخدمات بسعر أقل. إلا انه في نفس الوقت زاد عدد الهجمات والتهديدات على شبكات الحوسبة السحابية ولم تعد الحلول التقليدية مجدية في حماية البيانات والكشف عن التهديدات أو منعها في الوقت الحقيقي بما في ذلك أنظمة كشف التسلل. ومن هنا أتت أهمية دخول تقنيات الذكاء الصنعي على هذه الأنظمة حيث تم تطوير طرق متقدمة تهدف إلى زيادة أداء أنظمة كشف التهديدات مثل تقنيات تعلم الآلة أو التعلم العميق.

أجرينا في هذه البحث دراسة إحصائية على قاعدة بيانات CIC-IDS والتي كشفت عن عدة خصائص مهمة لسلوك حركة مرور الشبكة، إضافة إلى خصائص بررت استخدام تقنية التشفير التلقائي ( AE: Auto Encoder) كمرحلة مهمة في عملية تصنيف الهجوم. بعد الدراسة الإحصائية تم اختبار خوارزمية الشبكة العصبونية العميقة (DNN: Deep Neural Network) في كشف الهجمات حيث أظهرت النتائج أنها كانت قادرة فقط على تمييز حركة البيانات الطبيعية ولكنها لم تكشف أنواع الهجوم الموجودة ضمن قاعدة البيانات، ولهذا تم اقتراح نظام هجين يتكون من تقنية التشفير التلقائي (AE: Auto Encoder) كمرحلة أولى والذي يعمل على استخراج الميزات المهمة وتقليل الأبعاد بين البيانات وكشف الحالات الشاذة من خلال تدريبه فقط على البيانات الطبيعية وخرج هذه التقنية يكون دخل لخوارزمية DNN والتي بدورها تستخدم لتصنيف حركة البيانات. أظهرت النتائج أن النموذج الهجين أعطى أداء عال في قدرته على تصنيف أنواع الهجمات والذي بدوره يؤدي إلى تحسين أنظمة الكشف عن التسلل.

**الكلمات المفتاحية:** حوسبة سحابية– أمن – تعلم عميق– شبكات عصبونية عميقة – تشفير تلقائي

*أستاذ في قسم هندسة تكنولوجيا الاتصالات– كلية هندسة تكنولوجيا المعلومات والاتصالات – جامعة طرطوس– سوريا

** أستاذ مساعد في قسم هندسة تكنولوجيا الاتصالات – كلية هندسة تكنولوجيا المعلومات والاتصالات – جامعة طرطوس– سوريا

***طالبة دكتوراه في قسم هندسة تكنولوجيا الاتصالات– كلية هندسة تكنولوجيا المعلومات والاتصالات–جامعة طرطوس–سوريا.

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (10) العدد (10) 2026
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (10) No. (10) 2026

# A hybrid model for detecting intrusion in cloud computing networks using artificial intelligence techniques

**Dr. Fadi Ghosna ***
**Dr. Naji Mohamad ****
**Eng. Ranim Younes *****

**(Received 31/3/2026. Accepted 18/5/2026)**

**ABSTRACT**

With the rapid development of cloud computing technology, which has enabled the connection of numerous smart devices, facilitated easier data storage, exchange between systems, and allowed for more affordable access to services. The number of attacks and threats to cloud computing networks has simultaneously increased. Traditional solutions are no longer effective in protecting data and detecting or preventing threats in real time, including intrusion detection systems (IDS). This is where the importance of integrating artificial intelligence (AI) technologies into these systems lies. Advanced methods have been developed to enhance the performance of threat detection systems, such as machine learning and deep learning.

In this research, we conducted a statistical study on the CIC-IDS database, revealing several important characteristics of network traffic behavior. These characteristics justify the use of Auto Encoder (AE) technology as a crucial step in the attack classification process. Following statistical analysis, the Deep Neural Network (DNN) algorithm was tested for attack detection. The results showed that it was only able to distinguish normal data traffic but could not detect the types of attacks present in the database. Therefore, a hybrid system was proposed, consisting of Auto Encoder technology as a first stage. This technology extracts important features, reduces data spacing, and detects anomalies by training it solely on normal data. The output of this technology serves as the input for the DNN algorithm, which is then used to classify the data traffic. The results demonstrated that the hybrid model performed well in classifying attack types, leading to improved intrusion detection systems.

**Keywords**: Cloud Computing – Security – Deep Learning – DNN – Auto Encoder.

*professor, Communication Technology Engineering Department, Information and communication Technology Engineering, Tartous University, Syria .
**Assiestant professor, Communication Technology Engineering Department, Information and communication Technology Engineering, Tartous University, Syria
***PhD Student- Communication Technology Engineering Department, Information and communication Technology Engineering, Tartous University, Syria

## 1- مقدمــة (Introduction):

قدمت تقنية الحوسبة السحابية العديد من الخدمات المميزة منها البنية التحتية كخدمة ( IaaS: Infrastructure as a Service) والمنصة كخدمة (PaaS: Platform as a Service) وأيضاً قدمت البرامج كخدمة (SaaS: Software as a Service) (Singh et al.,2016). والآن مع التطورات السريعة التي تحدث في الشبكات تزايد عدد مقدمي خدمات الحوسبة السحابية بشكل كبير، فضلاً عن تزايد وتنوع الهجمات الالكترونية التي تستهدف الشركات ومقدمي الخدمات. مما أدى إلى العديد من المشكلات الأمنية التي تهدد بعدم إمكانية الوصول للبيانات أو سلامتها أو سريتها. وبالتالي أصبح من الضروري استخدام نظام لكشف التسلل (IDS: Intrusion Detection System) لحماية البيانات من الاختراق(Das et al., 2019). إلا أن أنظمة IDS التقليدية، غير قادرة على اكتشاف الهجمات الغير معروفة، ولا تستطيع التعامل مع البيانات الغير متوازنة مثلاً عدد الهجومات أقل بكثير من البيانات الطبيعية، إضافة إلى، صعوبة تعامله مع كميات ضخمة من البيانات والتي تولدها الشبكات الحديثة (Gopireddy, 2021). ومن هنا أتت الحاجة إلى دمج تقنيات الذكاء الصنعي مع أنظمة كشف التسلل، لما تقدمه من فوائد مميزة مثل اكتشاف الهجومات وحركة البيانات الشاذة الجديدة والغير معروفة مسبقاً وقدراتها على التعامل مع البيانات الضخمة (Ayyadapu, 2023).

في السنوات الأخيرة، عمل الباحثون في هذا المجال بتجربة أساليب جديدة باستخدام العديد من خوارزميات الذكاء الصنعي، والتي تهدف إلى التقليل من أي جانب سلبي لأنظمة كشف التسلل. في (Xu et al., 2020) عمل الباحثون على توظيف خوارزميات الذكاء الصنعي لتحديد الهوية، حيث تم استخدام خوارزمية أقرب جار (KNN) لاكتشاف الحالات الشاذة في الشبكة. في (Qazi et al., 2022) عمل الباحثون على استخدام مشفر تلقائي عميق (DAE) غير متماثل بهدف كشف الاختراقات ضمن الشبكات حيث تمت التجارب باستخدام مجموعة بيانات KDD CUP'99. في (Guijuan et al., 2020) تم استخدام تقنية المشفر التلقائي (AE) حيث تم توضيح ميزات هذه التقنية وتفوقها على الخوارزميات المعقدة الأخرى من خلال عمله على اكتشاف تمثيل البيانات وتجاهل إشارات الضجيج وتم اعتباره كأكثر النهج فاعلية في مجال الأمن السيبراني. في (Gamage & Samarabandu, 2020) استخدم الباحثون أربع خوارزميات في تصنيف أنواع الهجومات وهي خوارزمية الشبكات العصبونية العميقة (DNN) والتشفير التلقائي (AE) وشبكات الذاكرة طويلة الأمد (LSTM). أظهرت النتائج أن نموذج DNN أعطى دقة أعلى من باقي النماذج.

تم في هذه البحث العمل على نظام هجين لكشف اختراقات الشبكة يعتمد على خصائص إحصائية قائمة على تدفق البيانات باستخدام قاعدة البيانات المعيارية (CICIDS)، حيث يهدف هذا النظام إلى الكشف الدقيق عن هجمات حجب الخدمة DOS وتصنيفها بدقة عالية إلى أنواع مختلفة باستخدام منهجية تصنيف متعددة الفئات. حيث أن النظام الهجين المقترح يتكون من نموذجين من التعلم العميق: الأول هو المشفر التلقائي (AE) والثاني شبكة عصبونية عميقة (DNN). حيث في المرحلة الأولى تم تدريب خوارزمية المشفر التلقائي على حركة مرور طبيعية بهدف تعلم أنماط حركة البيانات الطبيعية واستخلاص تمثيلات الميزات ذات الصلة، بعدها يتم استخدام هذه التمثيلات المستخرجة من قبل خوارزمية DNN لتحسين كشف وتصنيف الهجمات.

### 1-1 هدف البحث (Research Objective):

يهدف هذا البحث إلى تحسين أداء أنظمة كشف التسلل في شبكات الحوسبة السحابية من خلال اقتراح نظام هجين من تقنيات التعلم العميق تعمل على استخلاص وتشفير الميزات المهمة بالاعتماد على خوارزمية التشفير التلقائي ومن ثم إدخال الميزات المشفرة إلى خوارزمية التعلم العميق بهدف اكتشاف حركة مرور البيانات وتصنيف أنواع الهجوم.

## 2- مواد وطرق البحث (Research materials and methods):

### 2-1 نظام كشف التسلل (IDS: Intrusion Detection System):

نظام كشف التسلل (IDS) هو عبارة عن نظام برمجي مهمته الأساسية هي التمييز بين الأحداث الطبيعية والغير طبيعية من خلال مراقبة حركة مرور البيانات ضمن الشبكة. يتم تصنيف نظام كشف التسلل إلى ثلاث أصناف رئيسية يختلف كل صنف عن الاخر في أسلوب تحليل البيانات. حيث يتم اختيار الأسلوب المناسب اعتماداً على عدة عوامل، مثل نوع الشذوذ، ومستوى الأمان المطلوب، والتكلفة (Kaur et al., 2017).

1- **أسلوب الكشف القائم على التوقيعات:** يتم اختيار هذا النوع من الكشف في الحالات التي يطلب بها كشف الهجمات المعروفة. يتميز بأنه سهل التنفيذ في الوقت الحقيقي، منخفض التكلفة. أما عن سلبياته فهو غير فعال ضد هجماتzero-day ، إضافة إلى انه يحتاج تحديث قاعدة البيانات بشكل دوري بهدف الكشف عن الهجمات الجديدة.

2- **أسلوب الكشف القائم على الشذوذ** يتم اختيار هذا النوع من الكشف في الحالات التي يطلب بها كشف الهجمات الغير معروفة بمعنى أنها غير موجودة ضمن قاعدة البيانات لديه. إلا أن له قيود تكمن في صعوبة التنفيذ أثناء التشغيل إضافة إلى ان دقة الكشف تتأثر بعدد الميزات الموجودة في قاعدة البيانات.

3- **الكشف الهجين:** يعمل الكشف الهجين على الدمج بين تقنيات الكشف القائم على الشذوذ والقائم على التوقيعات. يتميز بأنه يعطي معدل دقة عالٍ في الكشف إلا ان له تكلفة مرتفعة (Alzahrani & Hong, 2017).

أما بالنسبة لأهم التحديات التي تواجه أنظمة كشف التسلل نتيجة التطورات المستمرة في الأنظمة والشبكات فهي أولاً تزايد كمية البيانات المخزنة والمرسلة عبر الشبكات(De la Hoz et al., 2015). بمعنى أنه يجب أن يكون نظام كشف التسلل قادرًا على تحليل كمية كبيرة من البيانات في زمن قليل مع ضمان إعطاء دقة عالية في الكشف. التحدي الثاني يكمن في التنوع الكبير في عدد البرتوكولات الموجودة في الشبكات الحديثة مما يؤدي الى صعوبة التمييز بين حركة البيانات الطبيعية والشاذة. إضافة إلى مشكلة وجود هجمات منخفضة التردد والتي تنجح في اختراق أنظمة كشف التسلل التقليدية(Koc et al., 2012).

### 2-2 التعلم العميق (DL: Deep Learning):

التعلم العميق هو فرع من فروع التعلم الآلي. يعمل على نمذجة العلاقات بالاعتماد على طبقات تمثيل متعددة ويحتاج إلى كمية كبيرة من البيانات ليتمكن من التعرف عليها. لعبت خوارزميات التعلم العميق دوراً هاماً في تطوير أنظمة كشف التسلل في الشبكات (Avci et al., 2021). لذا، في هذا البحث، استُخدمت طريقتان من التعلم العميق، وهما التشفير التقائي والشبكات العصبونية العميقة، وتم دمجهما للحصول على نتائج أفضل ودقة أعلى في الكشف.

### 2-2-1 خوارزمية الشبكات العصبونية العميقة (DNN: Deep Neural Network):

تعرف الشبكة العصبونية الصناعية بأنها نموذج حسابي مستوحاة من آلية عمل الدماغ البشري وقدرته على التعلم واكتساب المعرفة من خلال تحديد العلاقات بين البيانات عبر التجربة. يعمل هذا النموذج على تمرير البيانات

في الشبكة لتحديد إشارة الدخل المناسبة التي سيتم نقلها إلى الطبقة التالية. ويسمى النموذج خطي بسيط إذا كان لا يحتوي على عصبونات داخلية، اما إذا احتوى على عصبونات سيعمل عندها على تحديد العصبونات التي يريد تنشيطها. وفي حال احتوت الشبكة العصبونية على طبقات مخفية متعددة عندها يطلق عليها مصطلح شبكات عصبونية عميقة(DNN)( Sana et al., 2021 ).

**2-2-2 خوارزمية المشفر التلقائي:**

تعتبر المُشفِّرات التلقائية (AutoEncoders – AEs) واحدة من نماذج التعلم العميق غير الخاضعة للإشراف، حيث يتم استخدامها في عمليات الكشف عن الحالات الغير طبيعية من خلال استخدام طرق تعلم التمثيل الكامن للبيانات. يتم تدريب هذه النماذج على تمثيل الأنماط الطبيعية بهدف إعادة بناء السلوك الطبيعي للبيانات، بعدها يتم اختيار المدخلات غير الطبيعية من خلال ملاحظة ارتفاع خطأ إعادة البناء (Li et al., 2023; Li et al., 2025). وانطلاقًا من هذه الأفكار، اعتمد هذا البحث في المرحلة الأولى على مُشفِّر تلقائي أساسي كنموذج فعال لإزالة الضجيج وتقليل الأبعاد واستخراج السمات من مجموعة بيانات، حيث يعمل على ضغط البيانات الأصلية عالية الأبعاد ويُقلل من البيانات المكررة ويحد من الضجيج، وهذا الذي سيتم ملاحظته في التحليل الإحصائي الذي تم إجراؤه على قاعدة البيانات المستخدمة. وخرج هذا المشفر استُخدم لاحقًا كمدخل لشبكة عصبونية عميقة (DNN)، وبالتالي الحصول على تحسين كفاءة الكشف عن الشذوذ.

**2-3 قاعدة البيانات والتحليل الإحصائي (Dataset & Statistical Analysis):**

تم في هذا البحث استخدام مجموعة جزئية من بيانات CIC-IDS المحملة من موقع Kaggle بهدف تقييم إطار عمل كشف الهجمات على شبكات الحوسبة السحابية المقترح. تعد مجموعة البيانات هذه مرجعية ومتاحة ومحدثة، تتضمن حركة مرور بيانات ضمن شبكة واقعية وتوضح أيضاً أنواع الهجمات على هذه الشبكات. تم استخراج قاعدة البيانات هذه باستخدام أداة CICFlowMeter حيث تتضمن البيانات خصائص إحصائية كاملة معتمدة على التدفق.

تم إجراء فحص أولي لقاعدة البيانات التي تم استخدامها في هذا البحث. حيث احتوت على ما يقارب 1.04 مليون سجل تدفق. تم رصد عدد قليل من القيم المفقودة (4921 مدخلاً)، وهو ما يمثل أقل من 0.5% من مجموعة البيانات. وقد أُزيلت هذه القيم المفقودة أثناء المعالجة المسبقة لضمان موثوقية التحليل الإحصائي وتدريب النموذج. وعدد الميزات (Features) كبير 79 ميزة أي أن العلاقات بينها معقدة وغير خطية.

**2-3-1 توزيع الفئات ضمن قاعدة البيانات:**

نلاحظ من الجدول (1) أن مجموعة البيانات غير متوازنة للغاية، حيث تمثل حركة المرور الطبيعية (Benign) غالبية التدفقات حوالي 95%، بينما تمثل حركة مرور الهجوم DOS حوالي 5% من مجموعة البيانات. وهذا طبيعي جداً في بيانات الشبكات الحقيقية إلا أنه يجعل كشف الهجمات أصعب.

**الجدول(1): توزيع الفئات ضمن قاعدة البيانات**

| الفئة | العدد | النسبة |
|---|---|---|
| حركة مرور طبيعية بدون هجوم (Benign) | 988050 | 95% |
| هجوم DOS نوع GoldenEye | 41508 | 5% |
| هجوم DOS نوع SlowLoris | 10990 | |

لدينا نوعين من هجمات DOS: أولاً GoldenEye وهو هجوم عالي السرعة يزيد الحمل على السيرفرات عن طريق إرسال كثيف للطلبات. ثانياً SlowLoris وهو هجوم يفتح عدداً كبيراً من الاتصالات ويُبقيها مفتوحة لفترة طويلة لاستهلاك موارد الخادم. أي أنه لدينا أكثر من نوع هجوم، وكل نوع له نمط حركة مختلف وخصائص مختلفة وهذا يزيد تعقيد المشكلة. إنّ وجود سلوكيات هجوم متعددة ومساحة ميزات عالية الأبعاد يجعل النماذج التقليدية السطحية غير كافية لالتقاط خصائص حركة مرور الشبكة المعقدة واكتشاف الهجمات. لذلك، تم اعتماد تقنيات التعلم العميق مثل المشفرات التلقائية والشبكات العصبونية العميقة لتعلم التمثيلات الكامنة وتحسين أداء الكشف.

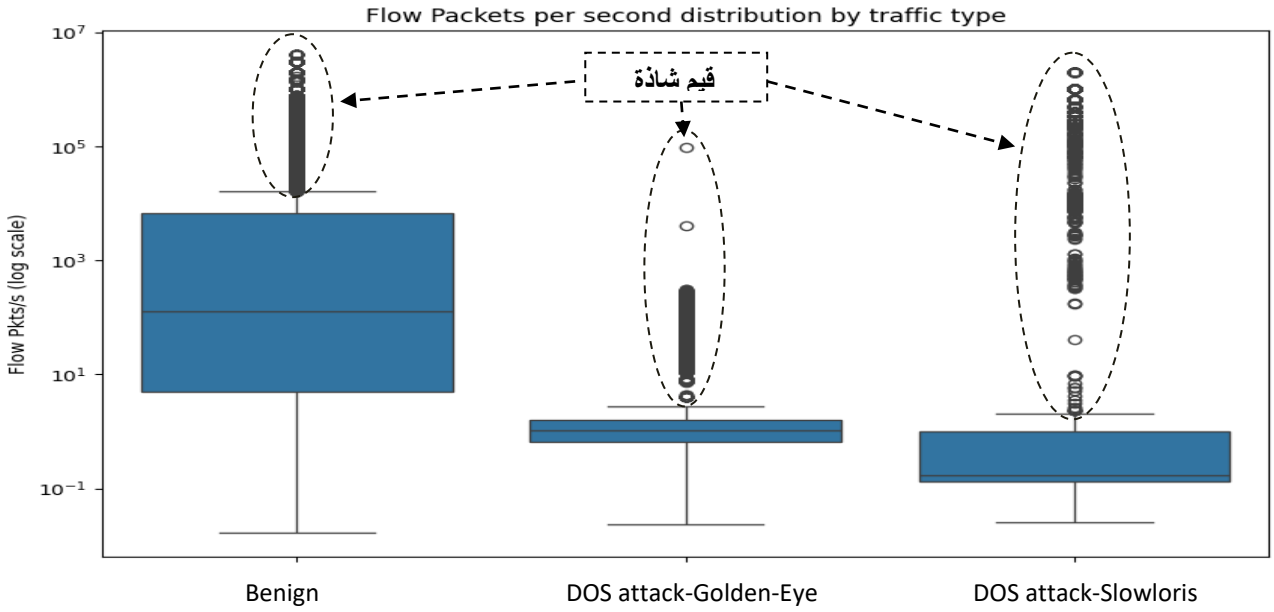
تكشف المقارنة الإحصائية في الجدول (2) بين حركة البيانات السليمة وحركة البيانات الهجومية عن اختلافات سلوكية واضحة في العديد من الخصائص، بما في ذلك مدة التدفق، والفترات الزمنية بين وصول الحزم، وعدد الحزم. تميل تدفقات البيانات الهجومية إلى إظهار مدة أطول، وفترات خمول أكبر، وخصائص نقل حزم مختلفة مقارنةً بحركة البيانات السليمة. تتوزع هذه الاختلافات على خصائص مترابطة متعددة بدلاً من سمة تمييزية واحدة، مما يُبرز تعقيد مهمة التصنيف. لذلك، تُستخدم أساليب التعلم العميق، مثل تمثيل الخصائص القائم على المشفر التلقائي والشبكات العصبونية العميقة، لالتقاط العلاقات غير الخطية ضمن بيانات حركة البيانات عالية الأبعاد.

**الجدول(2): مقارنة الإحصائيات بين حركة مرور البيانات الطبيعية والهجوم (Benign Vs DOS)**

| الميزة | طبيعي Benign | الهجوم Attack | التفسير |
|---|---|---|---|
| مدة التدفق Flow Duration ($\mu s$) | 13.9 | 23.9 | نلاحظ أن تدفقات هجمات DoS تستمر لفترة أطول من الاتصالات الطبيعية. وذلك بسبب أن الهجوم يحاول إبقاء الاتصال نشطًا لفترة طويلة لإشغال موارد الخادم. |
| معدل البيانات Flow Bytes/s | 425.606 | 66.509 | معدل البيانات في الهجوم أقل لأن هجمات مثل Slowloris ترسل بيانات ببطء شديد. وهذا أسلوب معروف لإبقاء الاتصال مفتوحًا دون استهلاك عرض نطاق كبير. |
| معدل الحزم Flow Pkts/s | 42.439 | 13.675 | الاتصالات الطبيعية قد تحتوي تبادل بيانات أكثر، بينما بعض هجمات DoS تعتمد على إبطاء الاتصال. |
| الزمن بين الحزم (IAT) ($\mu s$) | 2.8 | 7.0 | الزمن في الهجوم بين الحزم يكون أكبر، وهذا يتوافق مع سلوك هجمات SlowLoris التي ترسل الحزم ببطء. |
| فترات النشاط ($\mu s$) | 84 | 650 | في الهجوم فترات النشاط وفترات الخمول تكون أطول وهذا يدل على سلوك غير طبيعي للتدفقات. |
| فترات الخمول ($\mu s$) | 3.8 | 13.7 | |

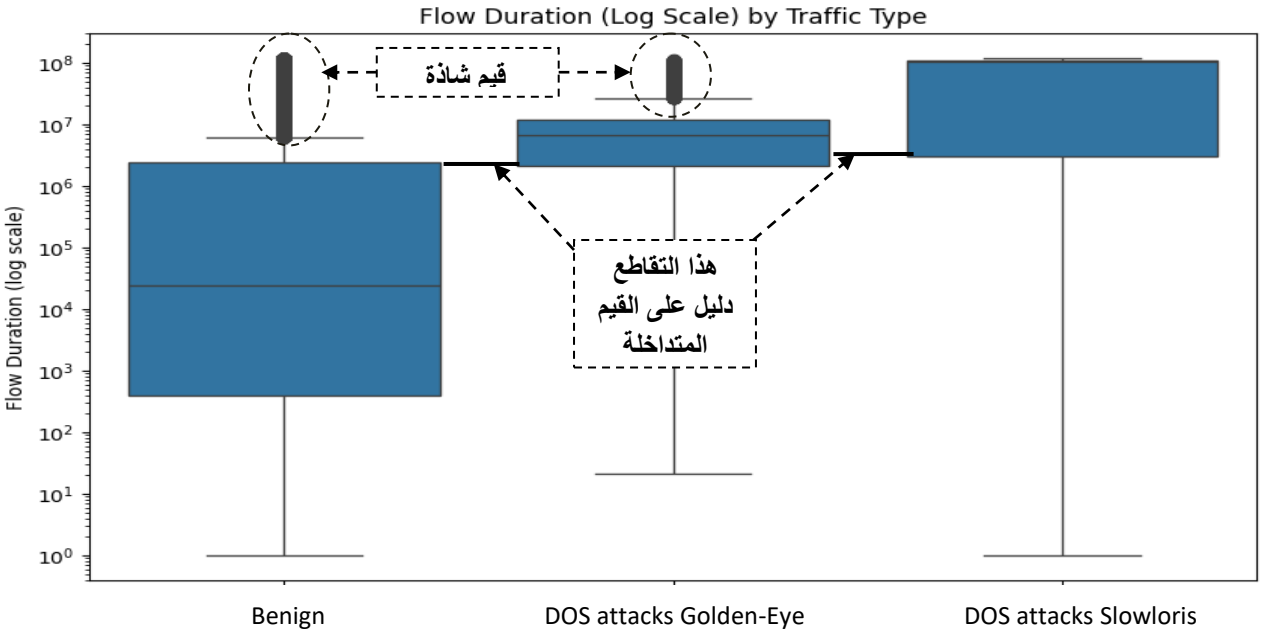
يوضح الشكل (1) توزيع حزم البيانات في الثانية الواحدة باستخدام مقياس لوغاريتمي. تُظهر حركة البيانات العادية (Benign) أن القيم تمتد لمستويات عالية بمعنى أن الاتصالات الطبيعية أحياناً تحتوي معدل حزم عالي. أما في هجوم DOS GoldenEye يظهر أن معظم القيم صغيرة أي أن هذا الهجوم لا يعتمد دائماً على معدل حزم عالي بل على نمط إرسال معين. وفي هجوم Dos Slowloris يظهر أن القيم أقرب للصفر بمعنى أن هذا الهجوم يرسل الحزم ببطء شديد.

النتيجة من هذا الشكل هي الفرق الواضح بين الهجوم والحركة الطبيعية وحتى أن كل هجوم له سلوك مختلف. ويشير أيضاً وجود قيم متداخلة وقيم شاذة إلا أن هذه الخاصية وحدها غير كافية للتصنيف الدقيق، مما يحفز استخدام نماذج التعلم العميق لرصد لأنماط المعقدة.



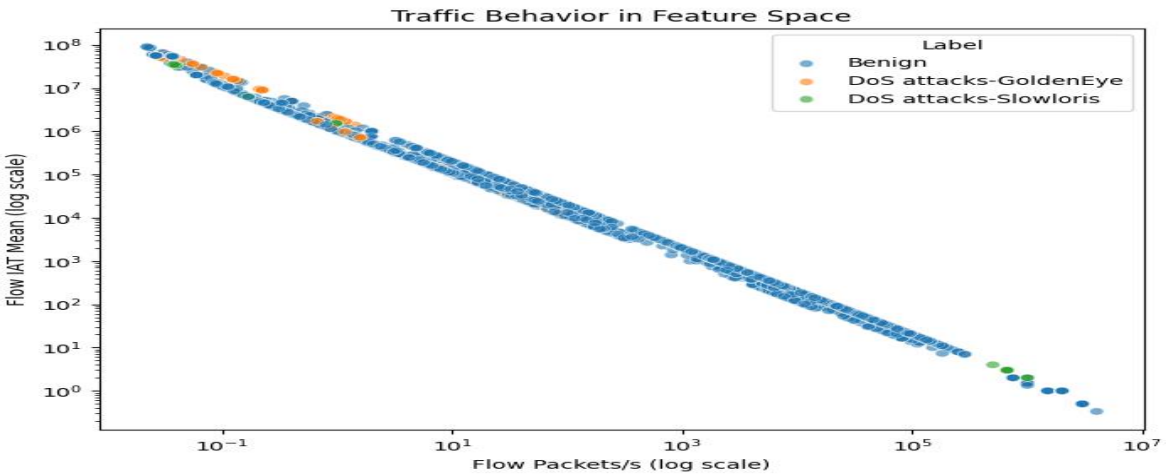

**الشكل (1): يوضح توزيع حزم البيانات في الثانية الواحدة باستخدام مقياس لوغاريتمي**

يوضح الشكل (2) مخطط صندوقي لتوزيع مدة التدفق عبر أنواع حركة المرور المختلفة باستخدام مقياس لوغاريتمي. يُلاحظ أن حركة المرور الطبيعية تكون فيها القيم موزعة على نطاق واسع والوسيط منخفض نسبياً بمعنى أن هذه الاتصالات غالباً قصيرة، بينما تحافظ هجمات حجب الخدمة، وخاصةً هجوم Slowloris، على اتصالات أطول بكثير. يشير وجود العديد من القيم الشاذة إلى تباين كبير وتوزيع غير طبيعي للبيانات. تؤكد هذه الملاحظات أن مدة التدفق سمة مميزة للتمييز بين حركة المرور الطبيعية وحركة المرور الهجومية ولكنها غير كافية لوحدها لتتم عملية التصنيف.




**الشكل (2): مخطط صندوقي لتوزيع مدة التدفق عبر أنواع حركة المرور المختلفة باستخدام مقياس لوغاريتمي**

يُظهر مخطط التشتت في الشكل (3) علاقة عكسية قوية بين عدد حزم البيانات في الثانية والفاصل الزمني بين وصولها، مما يُشكل نمطاً خطياً واضحاً على المقياس اللوغاريتمي. مع ذلك، تقع جميع أنواع حركة البيانات، بما في ذلك التدفقات الآمنة والهجومية، على نفس خط الاتجاه، مما يُشير إلى تداخل كبير بينها. مما يشير إلى أن الفئات لا يمكن فصلها خطياً وأن تمثيلات الميزات منخفضة الأبعاد غير كافية للتصنيف الدقيق، ويُبرز الحاجة إلى نماذج متقدمة لرصد الأنماط الخفية في البيانات عالية الأبعاد مثل الشبكات العصبونية العميقة.




**الشكل (3) مخطط التشتت لتدفق حزم البيانات مع الفاصل الزمني بين الحزم**

### 2-3-2 تبرير استخدام خوارزمية المشفر التلقائي كمرحلة أولى في النموذج الهجين:

من خلال هذا التحليل الإحصائي الذي أُجري على مجموعة بيانات CIC-IDS والذي كشف عن عدة خصائص مهمة لسلوك حركة مرور الشبكة. تُبرز النتائج أن السمات الفردية، أو حتى مجموعة بسيطة من السمات، لا تكفي لتحقيق تصنيف دقيق. تشير التوزيعات المتداخلة والعلاقات غير الخطية إلى أن المشكلة معقدة بطبيعتها، ولا

يمكن معالجتها بفعالية باستخدام الطرق التقليدية أو الطرق الخطية. لذا، يلزم استخدام تقنيات متقدمة للذكاء الصنعي لاستخلاص ميزات ذات دلالة من البيانات تلقائياً. في هذا السياق، تُستخدم المشفرات التلقائية تحديداً (AE) لأنها تعمل على استخلاص ميزات أقوى من الميزات الأصلية وتقليل الأبعاد والتقاط الأنماط الغير خطية وتقليل الضجيج والقيم الشاذة، مما يسمح للنموذج بالتقاط الأنماط الخفية والبنى الكامنة في البيانات. ثم تدخل هذه التمثيلات المستخلصة إلى شبكة عصبونية عميقة (DNN) لإجراء تصنيف دقيق. يُمكّن هذا المزيج النموذج من التغلب على قيود الأساليب التقليدية وتحسين دقة الكشف عن هجمات حجب الخدمة (DOS) بشكل ملحوظ.

**2-4 المعالجة المسبقة للبيانات:**

في مرحلة المعالجة المسبقة للبيانات الأولية، تم تنظيف مجموعة البيانات بإزالة القيم المفقودة والقيم الغير محدودة. بعد ذلك تم حذف الخصائص الغير مفيدة، مثل الطوابع الزمنية (Timestamp) حيث مثل هذه السمات يتم استخدامها لتتبع الاتصالات الضارة وقت حدوثها ولا تقدم سوى فائدة قليلة جدا اثناء تدريب الخوارزمية لذلك تم حذفها. وتم تدريب النماذج على السمات الرقمية فقط.

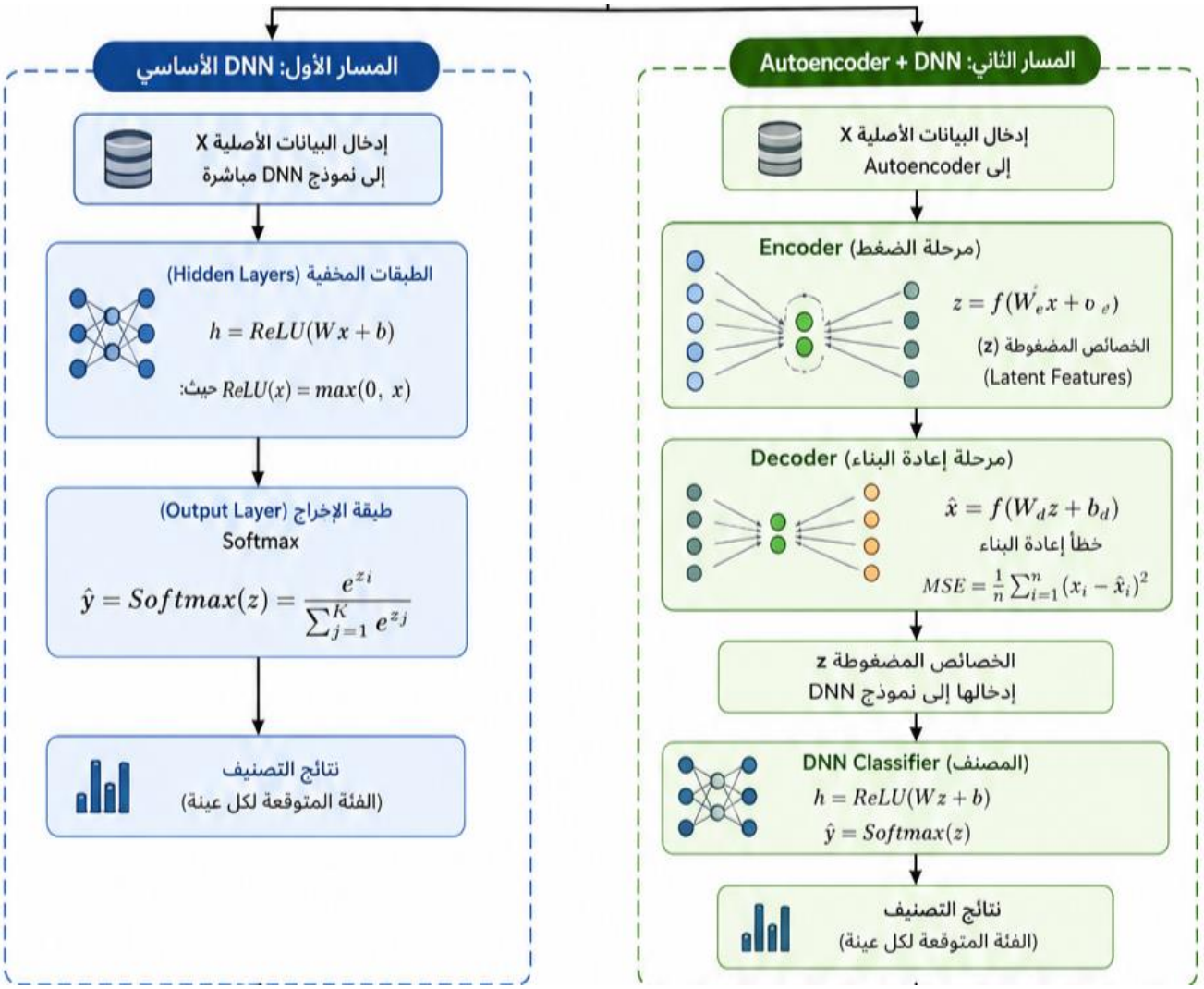
تم الاعتماد على حزم Pandas، NumPy، learn-Scikit ضمن لغة Python لتنفيذ عمليات المعالجة المسبقة للبيانات. وأيضاً تم استخدام ترميز التصنيفات من أجل إعطاء مجموعة البيانات. بعد ذلك تم تطبيق تحجيم السمات باستخدام تسوية MinMax بهدف تحويل السمات المدخلة إلى نطاقين بين 0 و1، والذي يؤدي إلى تحسين أداء خوارزميات التعلم العميق. بعد الانتهاء من مرحلة المعالجة المسبقة، تم تقسيم مجموعة البيانات إلى مجموعتين 70% من البيانات خاصة بالتدريب و30% من أجل الاختبار.

في المرحلة الأولى تم تدريب نموذج المشفر التلقائي AE على حركة بيانات طبيعية فقط من أجل تعلم سلوك الشبكة الطبيعي واستخراج تمثيلات الميزات. خرج هذه المرحلة تم استخدامه كمدخلات لتدريب خوارزمية الشبكة العصبونية العميقة (DNN). أخيراً تم تقييم أداء النموذج من خلال مجموعة التحقق، بهدف تحسين قدرته عند دخول بيانات حديثة على النموذج.

**2-5 المنهجية المقترحة:**

يوضح الشكل (4) المنهجية المتبعة في هذا البحث للنظام المقترح بعد عملية تحميل البيانات وإجراء عملية المعالجة المسبقة للبيانات مع توضيح المعادلات الرياضية في كل مرحلة. حيث أن البارامترات تمثل مايلي:

X: البيانات الأصلية
W: مصفوفة الأوزان
b: الانحياز
Z: الخصائص المضغوطة
$\hat{x}$ : البيانات المعاد بناؤها
$\hat{y}$ : الفئة المتوقعة
ReLU: دالة التنشيط
Softmax: دالة الاحتمال




**الشكل(4): المخطط الانسيابي لمراحل مقارنة نموذجي DNN و AE+DNN**

### 6-2 مقاييس الأداء:

تم الاعتماد على مصفوفة الارتباك لأنها تعمل على تحديد التصنيف الحقيقي والمتوقع. حيث تقسم النتيجة إلى قسمين، طبيعي وغير طبيعي. مصفوفة الارتباك تتضمن أربع حالات أساسية يجب قياسها.

الإيجابي الحقيقي (TP): أي أن النموذج دقيق ويتنبأ بنتائج إيجابية. بمعنى أن النموذج تنبأ هجوم وكان بالفعل هجوم.

النتيجة السلبية الخاطئة (FN): وهذا يعني أن النموذج غير دقيق. فهو يصنف حركة البيانات الخبيثة أنها طبيعية.

النتيجة الإيجابية الخاطئة (FP): وهذا يعني أن النموذج غير دقيق. يتنبأ النموذج بحركة المرور الطبيعية على أنها هجوم. لذلك يسمى بالإنذار الكاذب.

السلبي الحقيقي (TN): وهذا يعني أن النموذج دقيق ويتنبأ بنتائج صحيحة. بمعنى ان النموذج تنبأ أن البيانات طبيعية وكانت بالفعل طبيعية.

تم تقييم أداء النظام المقترح باستخدام معايير الاداء التالية الدقة، والضبط، والاستدعاء، ودرجة F1:

$$\text{الدقة} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{الاستدعاء (Recall)} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{درجة F1} = \frac{T2\times Precision\times Recall}{Precision + Recall} \quad (3)$$

## 3- النتائج والمناقشة:

### 3-1 نتائج البحث:

تم في هذا القسم تقديم النتائج التي تم الحصول عليها بعد تطبيق نموذجين من تقنيات التعلم العميق الشبكات العصبونية العميقة والتشفير التلقائي. أجريت التجارب باستخدام مجموعة بيانات CIC-IDS، وتحديداً سيناريو لنوعين من هجوم حجب الخدمة الموزع (DOS).

حيث تم في المرحلة الأولى تدريب نموذج الشبكة العصبونية العميقة (DNN) واختباره باستخدام ميزات التدفق الإحصائية المستخرجة، بعد ذلك تم تقييم أداء النموذج بالاعتماد على مصفوفة الارتباك ( confusion matrix) وتقرير التصنيف.

نستنتج من الجدول (3) أن النموذج أعطى نتائج عالية في اكتشاف حركة مرور البيانات الطبيعية، حيث حققت هذه الفئة دقة 0.95، واستدعاء 1.00، ودرجة F1 = 0.97. بينما فشل هذا النموذج في الكشف عن هجمات حجب الخدمة بنوعيها GoldenEye و Slowloris، حيث نلاحظ من الجدول أن قيم الدقة والاستدعاء ودرجة F1 كلها تساوي الصفر لنوعي الهجوم.

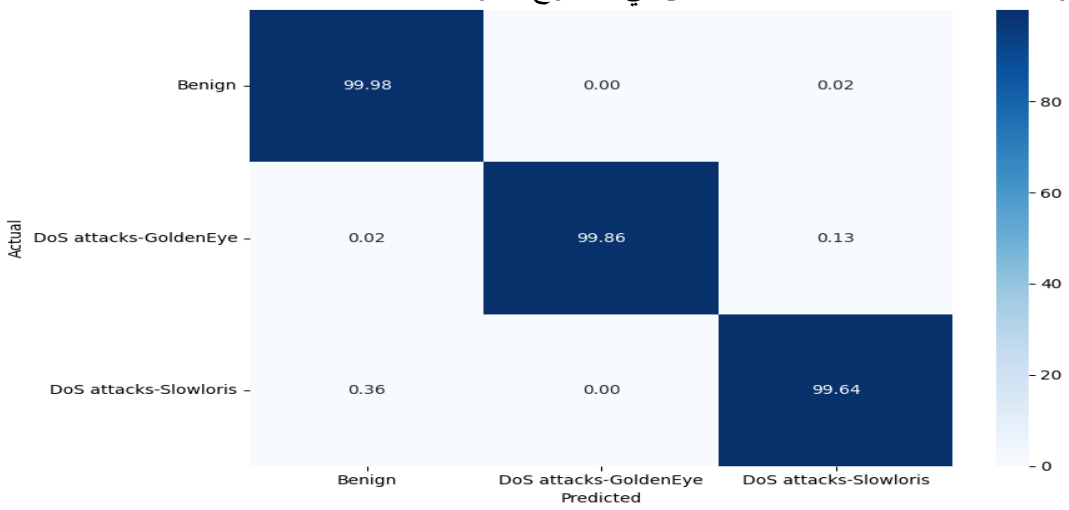
تُفسّر النتائج أن النموذج على الرغم من الدقة الإجمالية العالية التي حققها ولكن هذا لا يعني أنه ناجح في اكتشاف حركة مرور البيانات الطبيعية بل يعني أنه يميل نحو الفئة الأكثر ضمن قاعدة البيانات، متجاهلاً تماماً فئات الهجمات الأقل شيوعاً وهذا بسبب مشكلة عدم توازن الفئات ضمن مجموعة البيانات.

**الجدول (3): نتائج التقييم لنموذج DNN**

| الصنف (Class) | الدقة (Precision) | الاستدعاء (Recall) | F1 Score |
|---|---|---|---|
| البيانات الطبيعية (Benign) | 0.95 | 1.00 | 0.97 |
| هجوم DOS-GoldenEye | 0.00 | 0.00 | 0.00 |
| هجوم DOS-Slowloris | 0.00 | 0.00 | 0.00 |

انطلاقاً من الدراسة الإحصائية التي أجريت وانطلاقاً من هذه النتائج وبهدف تحسين دقة اكتشاف الهجمات تم الانتقال إلى فكرة النموذج الهجين حيث تم دمج خوارزمية التشفير التلقائي مع الشبكة العصبونية العميقة (AE+DNN) واستخراج النتائج ومقارنتها مع نتائج المرحلة السابقة. حيث تم تدريب نموذج AE على حركة مرور طبيعية بهدف تعلم تمثيل السلوك الطبيعي للبيانات وتشفير الميزات. بعد ذلك تم استخدام الميزات المشفرة كدخل لخوارزمية DNN. تم تقييم أداء هذا النموذج الهجين بنفس المقاييس التي تم اعتمادها في الفقرة السابقة. حيث وكما يظهر في الشكل (5) مصفوفة الارتباك حيث تُظهر تحسن كبير في اكتشاف الهجمات بنوعيها مقارنة مع نموذج DNN لوحده. تبين أنه تم تصنيف حركة المرور الطبيعية بنسبة 99.98% بشكل صحيح واكتشاف الهجمات من نوع DOS–GoldenEye بنسبة 99.86% والهجمات من نوع DOS–

Slowloris بنسبة 99.64%. وهذا ما يفسر أن النموذج دقيق ومستقر وقادر على التمييز بين الأصناف وأن خوارزمية AutoEncoder عملت بشكل دقيق في استخراج الميزات.




**الشكل(5): مصفوفة الارتباك للنموذج الهجين المقترح AE+DNN**

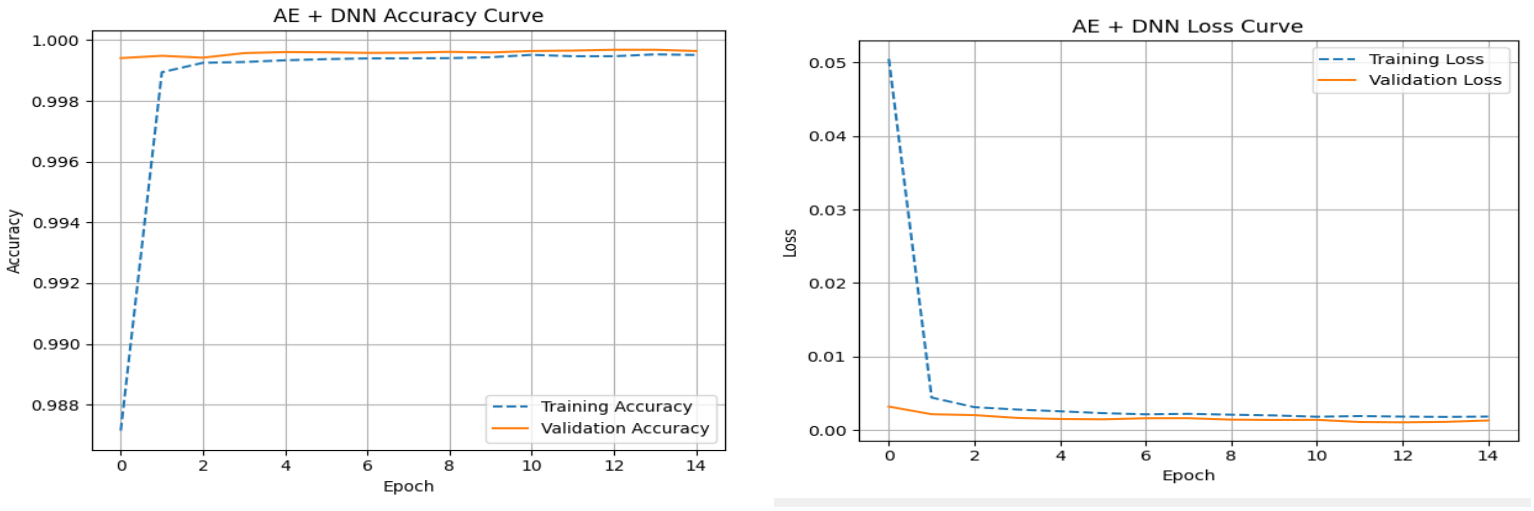
ويبين الجدول (4) نتائج النموذج الهجين، حيث تبين النتائج أن هذا النموذج قادر على اكتشاف جميع الهجمات بنوعيها كون قيم Recall=1 في الحالتين. ونلاحظ أن قيم الدقة عالية وهذا يعني أن النموذج لا يعطي إنذارات كاذبة.

**الجدول (4): نتائج التقييم للنموذج الهجين المقترح AE+DNN**

| الصنف (Class) | الدقة (Precision) | الاستدعاء (Recall) | F1 Score |
|---|---|---|---|
| **بيانات طبيعية (Benign)** | 1.00 | 1.00 | 1.00 |
| **هجوم DOS-GoldenEye** | 1.00 | 1.00 | 1.00 |
| **هجوم DOS-Slowloris** | 0.98 | 1.00 | 0.99 |

يوضح الشكل (6) منحنيات التدريب والخسارة للنموذج الهجين المقترح AE+DNN. نلاحظ من المنحني (6-a) أن دقة التدريب تزداد بسرعة في البداية وتستقر عند حوالي 99.9%، مما يدل على أن النموذج تعلم بسرعة ووصل إلى أداء عالي ومستقر ولا يوجد Overfitting واضح.

أما بالنسبة لمنحني الخسارة (6-b) فنلاحظ انخفاض سريع في المرحلة الأولى ويتقارب إلى قيم منخفضة جداً. وهذا يعني أن النموذج يقلل من أخطاء التصنيف بشكل كبير. وهذه النتائج تؤكد أن النموذج الهجين المقترح قادراً على كشف الهجوم والاختراقات بشكل كبير.




(a) (b)

**الشكل(6): منحنيات الدقة (Accuracy) والخسارة (Loss) للتدريب والتحقق لنموذج AE+DNN المقترح**

## 3-2 مقارنة النتائج مع الدراسات السابقة:

يوضح الجدول (5) مقارنة أداء النموذج المقترح مع دراستين في مجال كشف التسلل. أظهرت الدراسة الأولى المعتمدة على Sparse Autoencoder دقة بلغت 94%، بينما حققت الدراسة الثانية التي اعتمدت على نموذج هجين CNN+RNN دقة مقدارها 98.9%. أما النموذج المقترح في هذه الدراسة، والذي يعتمد على دمج Autoencoder مع DNN، فقد حقق أعلى دقة تصنيف، مما يؤكد فعالية الدمج بين تقليل الأبعاد والتصنيف العميق في تحسين أداء أنظمة كشف التسلل.

**الجدول(5): مقارنة نتائج النموذج المقترح مع بعض الدراسات السابقة**

| المرجع | المنهجية المستخدمة | الدقة | ملاحظات |
|---|---|---|---|
| (Aminanto & Kim, 2017) | Sparse Autoencoder | 94.81% | استخدم المشفر التلقائي لاستخراج الخصائص وتصنيف الهجمات اللاسلكية. |
| (Qazi et al., 2023) | CNN+RNN | 98.9% | نموذج هجين للتعرف على هجمات الشبكات باستخدام التعلم العميق. |
| النموذج المقترح في هذا البحث | AE+DNN | 99.9% | دمج AE مع DNN أدى إلى تحسين الأداء بشكل واضح. |

## 4- الاستنتاجات والتوصيات:

تم في هذا العمل إجراء تحليل إحصائي على مجموعة بيانات والذي كشف عن عدة خصائص مهمة لسلوك حركة مرور الشبكة.

1) أولًا، أظهر توزيع السمات الرئيسية، مثل مدة التدفق وعدد حزم البيانات في الثانية، اختلافات كبيرة بين حركة المرور العادية وحركة المرور الهجومية. وعلى وجه الخصوص، تُظهر هجمات حجب الخدمة(DOS)، ولا سيما هجمات Slowloris، مدد تدفق أطول ومعدلات حزم بيانات أقل مقارنةً بحركة المرور العادية.

2) ثانيًا، تتسم البيانات بانحراف كبير وتحتوي على العديد من القيم الشاذة، كما هو موضح في مخططات الصندوق. وهذا يُشير إلى أن توزيعات السمات غير طبيعية (غير غاوسية) وتُظهر تبايناً عالياً.

3) ثالثاً، يُظهر تحليل مخطط التشتت وجود علاقة عكسية قوية بين عدد حزم البيانات في الثانية والفاصل الزمني بين وصول التدفقات. ومع ذلك، وعلى الرغم من هذه العلاقة الواضحة، يوجد تداخل كبير بين حركة المرور العادية وحركة المرور الهجومية في فضاء السمات.

4) أبرزت هذه الملاحظات تبرير استخدام خوارزمية المشفر التلقائي والذي يتميز في قدرته على استخلاص الميزات الأكثر أهمية وتقليل الأبعاد، وحل مشكلة التداخل وتقليل الضجيج.

5) بعد النتائج الإحصائية وقبل إدخال خوارزمية التشفير التلقائي والاعتماد على النموذج الهجين، تم الاعتماد على خوارزمية الشبكات العصبونية العميقة (DNN) لوحدها في عملية تصنيف البيانات بالاعتماد على الخصائص الإحصائية لتدفقات الشبكة أي على الميزات الأصلية. حيث أظهر هذا النموذج نسبة نجاح عالية في كشف حركة البيانات الطبيعية إلا أنه لم يتعرف أبداً على أنواع هجومات DOS الموجودة، وذلك بسبب عدم توازن البيانات.

6) ولحل هذه المشاكل تم اقتراح النموذج الهجين والذي يدمج تقنية التشفير التلقائي (AE) مع خوارزمية الشبكات العصبونية العميقة (DNN)، حيث تم تدريب تقنية التشفير التلقائي على حركة مرور بيانات طبيعية لتعلم سلوك هذه البيانات، ثم استخدام خرج هذه المرحلة كمدخلات لخوارزمية DNN.

7) أظهرت النتائج أخيراً تحسن كبير في الأداء حيث أصبح النموذج الهجين قادر على التمييز بين أنواع الهجمات، إضافة إلى نسبة أخطاء قليلة في التصنيف. وهذا بدوره يؤكد أهمية إضافة تقنية التشفير التلقائي والاعتماد على النماذج الهجينة.

**وبناء على النتائج يمكننا اقتراح بعض التوصيات:**

a) تطبيق النموذج الهجين المقترح (AE+DNN) على كامل قاعدة البيانات بهدف تقييم الأداء على مجموعة كبيرة من الهجمات.

b) مقارنة النموذج المقترح مع نماذج تعلم عميق أخرى على نفس قاعدة البيانات لتقييم الأداء بشكل دقيق.

c) إضافة خوارزميات تحسينية على تقنية DNN.

## 5- المراجع:

**[1]** ALZAHRANI, S., HONG, L. 2017, *A survey of cloud computing detection techniques against DDoS attacks, Journal of Information Security*, vol. 9, no. 1, pp. 45-69.

**[2]** AMINANTO, M.E., KIM, K. 2017, *Improving detection of Wi-Fi impersonation by fully unsupervised deep learning, In: International Workshop on Information Security Applications*, Springer, Berlin/Heidelberg, Germany, pp. 212-223.

**[3]** AYYADAPU, A.K.R. 2023, *Enhancing cloud security with AI-driven big data analytics, International Neurourology Journal*, vol. 27, no. 4, pp. 1591-1597.

**[4]** AVCI, O., ABDELJABER, O., KIRANYAZ, S., HUSSEIN, M., GABBOUJ, M. , INMAN, D.J. 2021, *A review of vibration-based damage detection in civil structures: from traditional methods to machine learning and deep learning applications, Mechanical Systems and Signal Processing*, vol. 147, p. 107077.

**[5]** Das, S., MAHFOUZ, A.M., VENUGOPAL, D., SHIVA, S. 2019, *DDoS intrusion detection through machine learning ensemble, Proceedings of the 19th IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C 2019*), pp. 471-477.

**[6]** DE LA HOZ, E., DE LA HOZ, E., ORTIZ, A., ORTEGA, J., PRIETO, B. 2015, *PCA filtering and probabilistic SOM for network intrusion detection, Neurocomputing*, vol. 164, pp. 71-81.

**[7]** GAMAGE, S. AND SAMARABANDU, J. 2020, *Deep learning methods in network intrusion detection: a survey and an objective comparison, Journal of Network and Computer Applications*, vol. 169, p. 102767.

**[8]** GOPIREDDY, R.R. 2021, *AI-powered security in cloud environments: enhancing data protection and threat detection, International Journal of Science and Research*.

**[9]** GUIJUAN, Z., YANG, L. AND XIAONING, J. 2020, *A survey of autoencoder-based recommender systems, Frontiers of Computer Science*, vol. 14, pp. 430-450.

**[10]** KAUR, P., KUMAR, M. AND BHANDARI, A. 2017, *A review of detection approaches for distributed denial of service attacks, Systems Science & Control Engineering*, vol. 5, no. 1, pp. 301-320.

**[11]** KOC, L., MAZZUCHI, T.A., SARKANI, S. 2012, *A network intrusion detection system based on a hidden naïve Bayes multiclass classifier, Expert Systems with Applications*, vol. 39, pp. 13492-13500.

**[12]** LI, J., ZHENG, K., GAO, L., HAN, Z., LI, Z. , CHANUSSOT, J. 2025, *Enhanced deep image prior for unsupervised hyperspectral image super-resolution, IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, p. 5504218.

**[13]** LI, J., ZHENG, K., LI, Z., GAO, L., JIA, X. 2023, *X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution, IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, p. 5518317.

**[14]** QAZI, E.U.H., IMRAN, M., HAIDER, N., SHOAIB, M., RAZZAK, I. 2022, *An intelligent and efficient network intrusion detection system using deep learning, Computers and Electrical Engineering*, vol. 99, p. 107764.

**[15]** QAZI, E.U.H., FAHEEM, M.H. AND ZIA, T.2023, HDLNIDS: *Hybrid deep-learning-based network intrusion detection system, Applied Sciences*, vol.13, p.4921.

**[16]** SANA, M.U., LI, Z., JAVAID, F., LIAQAT, H.B., ALI, M.U. 2021, *Enhanced security in cloud computing using neural network and encryption*, IEEE Access, vol. 9, pp. 145785-145799.

**[17]** SINGH, S., JEONG, Y.S., PARK, J.H. 2016, *A survey on cloud computing security: issues, threats, and solutions, Journal of Network and Computer Applications*, vol. 75, pp. 200-222.

**[18]** XU, H., PRZYSTUPA, K., FANG, C., MARCINIAK, A., KOCHAN, O., BESHLEY, M. 2020, *A combination strategy of feature selection based on an integrated optimization algorithm and weighted K-nearest neighbor to improve the performance of network intrusion detection, Electronics*, vol. 9, p. 1206.

**[19]** ZHANG, Z., ZHANG, K., KHELIFI, A. 2018, *Multivariate time series analysis in climate and environmental research*, Springer, Berlin/Heidelberg, Germany.