الجمهورية العربية السورية

جامعة تشرين

كلية الهندسة المعلوماتية

قسم النظم والشبكات الحاسوبية

A Deep Learning-Based Intrusion Detection and Preventation System for Detecting and Preventing Denial-of-Service Attacks

إعداد: نوح عبود

إشراف :د أحمد أحمد

2024/2023

Contents

:ABSTRACT-1

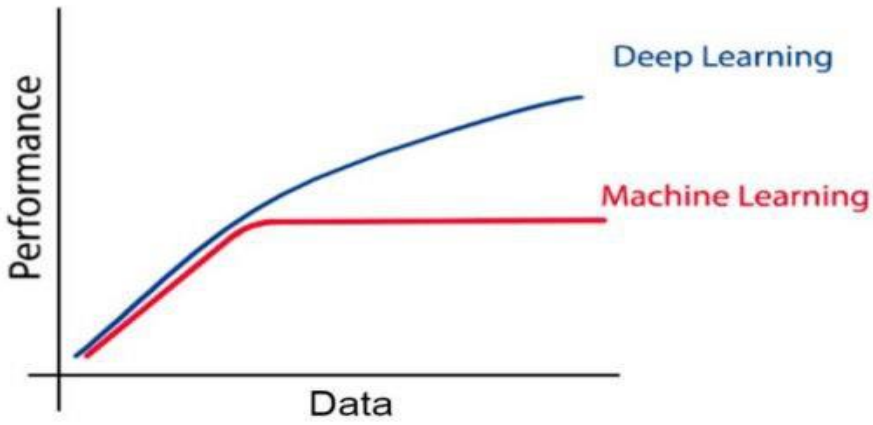
يوضح هذا التقرير كيفية تصنيف واختيار وتدريب خوارزمية التعلم العميق(multi-layered Deep Feed Forward neural network) لإنشاءIDS/IPS (نظام منع/كشف التسلل) يسمى Dique(نسبة الى سد Dique Potrerillos الموجود في الارجنتين) والذي يمكنه اكتشاف ومنع هجمات رفض الخدمة (DoS). باستخدام نموذج التعلم العميق المقترح تنقسم الحزم الواردة إلى خادم الشبكة إلى فئتين: الحميدة (وهي حزم مرور عادية) والخبيثة (والتي يعتبرها النظام أنها تحتوي على هجمات حجب الخدمة المحتملة). لدى Dique واجهة مستخدم رسومية (GUI) حيث يمكننا "في الوقت الفعلي" عرض المعلومات الرسومية والنصية التي تم التقاطها وتصنيف الحزم، ويسمح لنا بالتبديل بين وضع IDS ووضع IPS لتشغيل النظام. يستخدم النموذج المقترح شبكة عصبية متعددة الطبقات Deep Feed Forward. تم استخدام مجموعة بيانات CICDDoS2019 للتدريب وتم تحقيق دقة قدرها 0.994. تم تنفيذ سبعة أنواع مختلفة من هجمات DoS (خمسة موجودة في مجموعة بيانات التدريب واثنتان غير موجودتين في مجموعة البيانات المذكورة) والتي يمكن للمستخدمين إطلاقها بشكل انتقائي ضد خادم الشبكة.

: INTRODUCTION-2

نظرًا لحجم الإنترنت المتزايد باستمرار وعلى الرغم من الجهود المبذولة لحماية أنظمة الكمبيوتر، تستمر الجرائم الإلكترونية في النمو بشكل كبير. وفي ضوء هذا الوضع، شعر الخبراء بالحاجة إلى تطوير أدوات أكثر تطوراً للكشف عن المتسللين والتصرف كرد فعل سريع و بشكل استباقي.ونظرا لذلك اتجهت معظم الدراسات نحوا إمكانية دمج ميزات التعلم الألي مع أنظمة الحماية للوصل الى أساليب حماية أفضل.

على الرغم من أن التعلم الآلي يبدو أسلوبًا واعدًا لمعالجة العديد من مشكلات الأمن السيبراني، إلا أنه يحتوي على العديد من العيوب. أحدها هو أنه يمكن أن يخطئ في تصنيف حزمة بيانات ضارة على أنها حميدة ويقبلها للتعلم، مما قد يؤدي إلى إفساد الخوارزمية تمامًا، بالإضافة إلى التعامل مع الحزم الشاذة على أنها حزم حميدة. والعيب الثاني هو أن العدد الكبير من الحزم التي تنتقل عبر الشبكة يتطلب مستوى عال من المعالجة، مما يجعل من الصعب تحليل الحزم في الوقت الحقيقي وقد يؤثر على أداء نظام الكمبيوتر. نظرًا للحاجة إلى تحسين

تقنيات التعلم الآلي لاكتشاف التسلل، يقوم خبراء الذكاء الاصطناعي (AI) بالتحقيق في استخدام أساليب التعلم العميق (DL) - وهو مجال فرعي من التعلم الآلي - لحل هذه المشكلات. كما هو موضح في الشكل 1، يتناسب أداء خوارزميات التعلم العميق مع كمية البيانات المعالجة، في حين يميل أداء خوارزميات التعلم الآلي إلى الاستقرار بمرور الوقت



الشكل 1 Machine learning vs. deep learning.

3- Machine learning vs deep learning :

تقدم DL العديد من الفوائد لاكتشاف الهجمات عند مقارنتها بـ ML :

	Machine learning	deep learning
Data size	تعمل خوارزميات ML بشكل أفضل مع مجموعات البيانات الصغيرة.	تعمل خوارزميات DL بشكل أفضل بكثير مع كميات كبيرة من البيانات
Time	تتطلب وقت أقل	خوارزميات DL تتطلب مزيدًا من الوقت للتدريب، إلا أنه يتم تعويض هذا الوقت

		الإضافي أثناء مرحلة التشغيل في الوقت الفعلي.
Dedication	تتطلب خوارزميات التعلم الآلي (ML) تحديد الميزات والتسميات (المخرجات) مسبقًا حتى يتمكن النظام من التعلم وإجراء التنبؤات بدقة.	في خوارزميات التعلم العميق (DL)، يقوم النظام تلقائيًا بتحديد الميزات (المدخلات) لتحليل البيانات

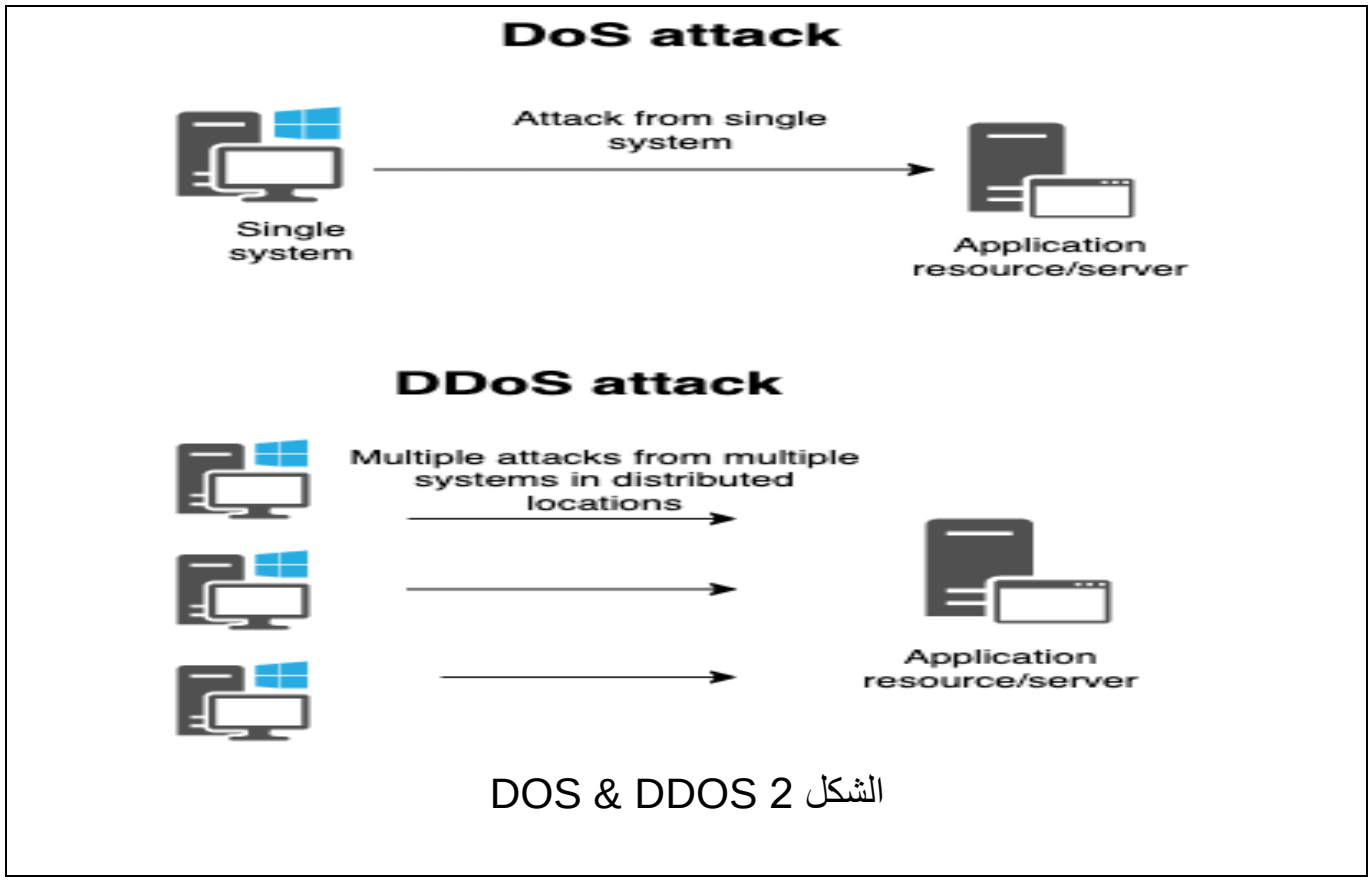
4- DOS & DDOS :

تعد هجمات رفض الخدمة (DoS) وهجمات رفض الخدمة الموزعة (DDoS) من الهجمات الإلكترونية المعروفة التي يتم فيها محاولة استهلاك موارد الحوسبة الخاصة بالمضيف أو الشبكة، مما يجعل الوصول إليها غير ممكن للمستخدمين الشرعيين أو يؤثر بشكل كبير على تشغيل نظام الكمبيوتر الخاص بهم. يمكن تصنيف هجمات DoS إلى عمليات استغلال البرامج وهجمات الفيضانات. في عمليات استغلال البرمجيات، يستغل المهاجم نقاط الضعف في خادم الضحية لتعطيل خدماته أو تقليل أدائه بشكل كبير. في هجمات الفيضان، يقوم المهاجم باستنزاف موارد النظام عن طريق إرسال عدد كبير من الطلبات الكاذبة.

يمكن تصنيف الأساليب المستخدمة للكشف عن هجمات DoS إلى ثلاث فئات:

- ✓ **signature-based detection :** الاكتشاف المعتمد على التوقيع هو طريقة تستخدم في أنظمة كشف التسلل لتحديد المتسللين بناءً على الأنماط أو التوقيعات المعروفة للهجمات السابقة. يتضمن هذا الأسلوب مقارنة حركة مرور الشبكة الواردة بقاعدة بيانات لتوقيعات الهجوم المحددة مسبقًا لاكتشاف الأنشطة الضارة وحظرها. يعتبر الاكتشاف المعتمد على التوقيع فعالاً في التعرف على تهديدات محددة ولكنه قد يواجه صعوبة في اكتشاف هجمات جديدة أو غير معروفة لا تتطابق مع التوقيعات الموجودة مسبقا.

- ✓ **anomaly-based detection:**على عكس الكشف القائم على التوقيع، والذي يعتمد على أنماط معروفة من الهجمات، يركز الكشف القائم على الشذوذ على اكتشاف الانحرافات عن سلوك الشبكة العادي.
- ✓ **hybrid detection technique :** طريقة تجمع بين تقنيات الكشف القائمة على الشذوذ والتوقيع على أساس لتعزيز دقة وفعالية تحديد التهديدات السيبرانية ومنعها. ومن خلال دمج نقاط القوة في كلا النهجين، يمكن لأنظمة الكشف المختلطة اكتشاف الهجمات المعروفة وغير المعروفة، مما يوفر آلية دفاع أكثر شمولاً ضد أنواع مختلفة من التهديدات السيبرانية في الوقت الفعلي.



الشكل 2 DOS & DDOS

5- IPS & IDS :

IDS (Intrusion Detection system)هو إما جهاز أو برنامج يقوم بتحليل حركة مرور الشبكة الواردة بحثًا عن أنشطة ضارة أو انتهاكات للسياسة (تحليل سلوك الشبكة) ويصدر تنبيهات

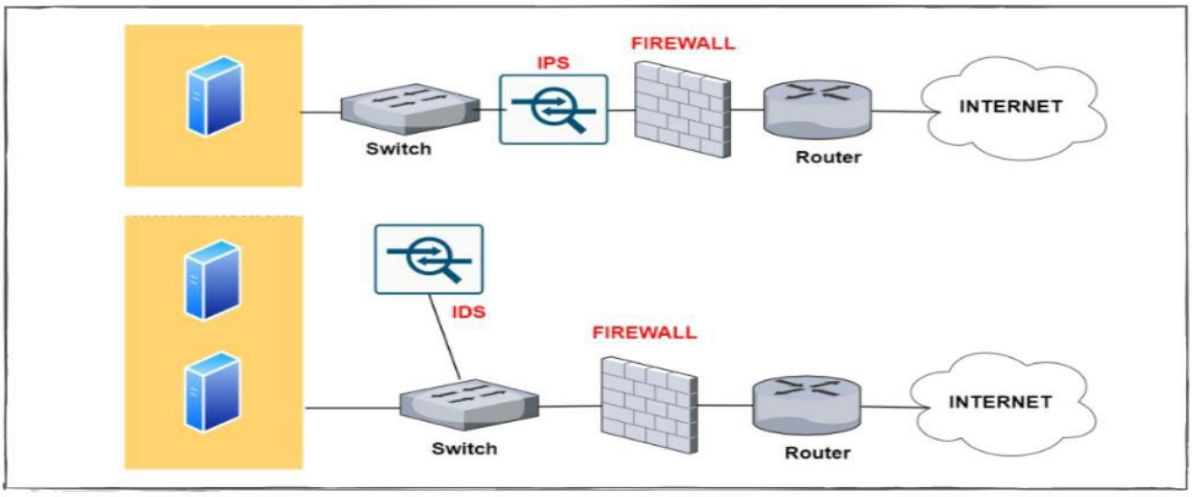
عند اكتشافها. فهو يكتشف حركة المرور في الوقت الفعلي ويبحث عن توقيعات الهجوم أو أنماط حركة المرور، ثم يرسل الإنذارات. على عكس IPS، فإن نظام كشف التسلل إلى الشبكة لا يتماشى مع مسار البيانات، لذلك يمكنه فقط التنبيه والإنذار عند اكتشاف الحالات الشاذة.

- يطبق IDS طريقتين للكشف عن الحالات الشاذة في الحزمة الموجودة في الشبكة. هم:

- Signature-based detection
- Anomaly-based detection

IPS- (Intrusion preventation System) تعني نظام كشف التسلل والوقاية منه. وكما يوحي الاسم، فهو يكتشف الحزم الضارة ويحظر الحزمة. على عكس IDS، الذي يقوم فقط باكتشاف الحزمة والإبلاغ عنها، يحاول IPS حظر الحزم أيضًا. وبالتالي، فإن IPS متقدم بعض الشيء وأكثر فعالية من IDS.

- يتم توصيل IPS بشكل مباشر بتدفق الحزمة. كما هو موضح من هيكل الشبكة (الشكل 3) بالاسفل (جدار الحماية مع IPS)، عادةً ما يكون جهاز IPS متصلاً خلف جدار الحماية ولكنه ضمن مسار الاتصال الذي ينقل الحزم من/إلى الشبكة الداخلية حتى يتمكن جهاز IPS من حظر حركة المرور الضارة على الفور قبل الوصول إلى الخوادم الداخلية



الشكل 3 IPS & IDS

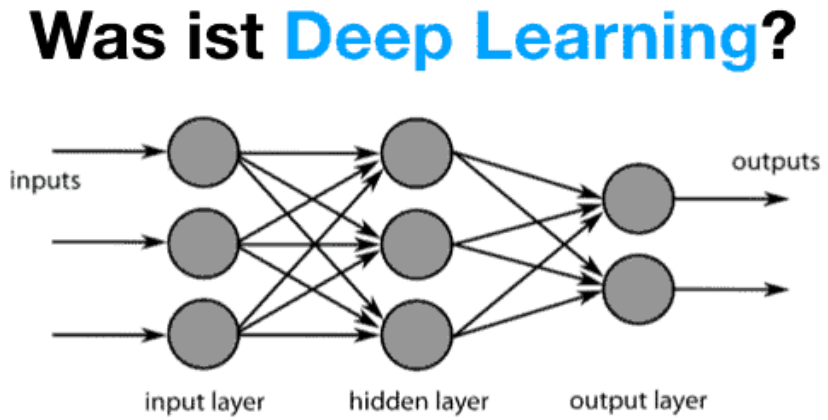
-- تطبق IPS ثلاث طرق للكشف عن الحالات الشاذة وحظر الحزمة في الشبكة. هي:

- Signature-based detection:مطابق لما هو علية في IDS ولكن هنا يتم رفع التنبيه واسقاط الحزمة.
- Anomaly-based detection:أيضا مطابق لما هو علية في IDS ولكن هنا يتم رفع التنبيه واسقاط الحزمة.
- Stateful protocol analysis detection: يعتمد الكشف هنا على اختلاف البروتوكول. تتم مقارنة الحزم الواردة مع ملف التعريفات المقبولة وبالتالي يتم إسقاط الحزمة أو السماح بها

6-Deep Learning :

التعلم العميق هو فرع من فروع التعلم الآلي (Machine Learning) يعتمد على الشبكات العصبية الاصطناعية (Artificial Neural Networks) لمحاكاة عمليات التعلم في الدماغ البشري. يتميز التعلم العميق بقدرته على معالجة البيانات الكبيرة والمعقدة مثل الصور والفيديوهات والنصوص، واستخلاص الأنماط والمعلومات منها.
الشبكات العصبية الاصطناعية هي النواة الأساسية للتعلم العميق. تتكون الشبكة العصبية من طبقات من الخلايا العصبية الاصطناعية(Neurons) ، وتنقسم إلى ثلاثة أنواع رئيسية من الطبقات (الشكل 4):

1. **الطبقة المدخلة** :**(Input Layer)** تستقبل البيانات الخام (مثل الصور أو النصوص).
2. **الطبقات المخفية** :**(Hidden Layers)** تعالج البيانات وتستخلص الأنماط منها. يمكن أن تكون هناك عدة طبقات مخفية.
3. **الطبقة الخارجة** :**(Output Layer)** تنتج النتيجة النهائية (مثل تصنيف الصور أو التنبؤ بالنصوص).



الشكل 4 Artificial Neural Networks

يتم تدريب الشبكة العصبية وفقل للخطوات الأتيه:

1. **إعداد البيانات** :تجميع وتنسيق البيانات وتحديد المدخلات والمخرجات.
2. **تهيئة الشبكة العصبية** :تحديد عدد الطبقات العصبية وعدد الخلايا العصبية في كل طبقة.
3. **التغذية الأمامية (Forward Propagation)**: تمرير البيانات عبر الشبكة العصبية للحصول على النتيجة.
4. **حساب الخطأ (Error Calculation)**: مقارنة النتيجة المتوقعة بالنتيجة الفعلية وحساب الخطأ.
5. **التغذية الخلفية (Backpropagation)**: تحديث الأوزان في الشبكة العصبية لتقليل الخطأ باستخدام خوارزميات مثل.Gradient Descent
6. **التكرار (Iteration)**: تكرار العمليات السابقة لعدة مرات حتى تصل الشبكة العصبية إلى مستوى مقبول من الدقة

-أنواع الشبكات العصبية:

هناك عدة أنواع من الشبكات العصبية والتقنيات المستخدمة في التعلم العميق، كل منها مناسب لمهام معينة:

- الشبكات العصبية الالتفافية(Convolutional Neural Networks - CNNs)
تستخدم بشكل رئيسي في معالجة الصور والفيديوهات. تتكون من طبقات تلافيفية (Convolutional Layers)التي تكتشف الأنماط في الصور مثل الحواف والأشكال.
- .2الشبكات العصبية التكرارية(Recurrent Neural Networks - RNNs)
تستخدم لمعالجة البيانات التسلسلية مثل النصوص والصوتيات. تتميز بقدرتها على الاحتفاظ بالمعلومات من الخطوات السابقة في التسلسل.
- شبكات الذاكرة الطويلة القصيرة الأمد(Long Short-Term Memory - LSTM)
هي نوع من RNNs مصممة للتغلب على مشكلة النسيان في الشبكات العصبية التكرارية العادية. تستخدم بشكل رئيسي في معالجة النصوص والترجمة الآلية.
- الشبكات العصبية المتقدمة(Advanced Networks)
مثل شبكات المحولات (Transformers) التي تستخدم في مهام معالجة اللغة الطبيعية(NLP) مثل الترجمة الآلية وتوليد النصوص.
- Deep Feedforward Neural Networksالشبكات العصبية العميقة الموجهة (DFNN) تمثل نموذجًا قويًا لمعالجة البيانات المعقدة واستخلاص الأنماط منها. تعتمد على طبقات متعددة من الخلايا العصبية التي تتعلم الميزات بشكل تدريجي من المدخلات الخام إلى المخرجات النهائية. عملية التدريب تشمل التغذية الأمامية لحساب المخرجات، وحساب الخطأ، والتغذية الخلفية لتحديث الأوزان، مما يمكن النموذج من التحسن مع مرور الوقت وزيادة دقته في التنبؤات.

: LITERATURE REVIEW-7

يوضح الجدول الأتي(الشكل 5) العلاقة بين خوارزميات التعلم العميق ومجموعات البيانات المختلفة لاكتشاف هجمات رفض الخدمة (DoS) والتي تم استخدامها في دراسات سابقة . ويعرض معدلات الدقة التي حققتها نماذج مختلفة عند تدريبها واختبارها على مجموعات بيانات محددة مثل NSL-KDD و99 KDD. وتسلط النتائج الضوء على فعالية أساليب التعلم العميق، مثل GRU وLSTM وDeep Feedforward وRandom Forest ,(CNN) وLSTM، في اكتشاف الهجمات السيبرانية ومنعها، مع الحاجة إلى مزيد من التحسين لتطبيقات العالم الحقيقي.

Algorithms \ Dataset	MawiLab 2017	ISCX 2012	NSLKDD	UNSW NB15	CICIDS2017	CNTC 2017	CSIC 2010	DARPA 1998	Netflow	KDD CUP 99	Darpa 2009 DDoS	Yahoo S5 Web	Dataset propio SSE
Recurrent Neural Network	100 [17]		98.8 [1]							99.49 [13]			
Convolutional Neural Network	100 [17]					94,11 [7]		99,36 [7]	99.41 [1]				
Stacked Recurrent Neural Network	100 [17]												
Long short Term Memory		98,88 [4]	99,51 [14]			95,12 [7]	96,13 [7]	99,98 [7]		99.8 [1]			
Deep Radial Intelligence			99,69 [8]	96,15 [8]									
Genetic Algorithm(Simulated Annealing Algorithm (SAA)					99.92 [9]								
VCNN/FNN			99.2 [20]										
Deep Belief Network			97.5 [1]						97.6 [1]	93.49 [1]			
Gradient Recurrent Unit									84.15 [1]				
Deep Feed Forward			93.78 [19]								99.63 [24]		
(Bidirectional Gradient Recurrent Unit)BGRU + MLP			99.24 [15]							99.84 [15]			
Restricted Boltzmann Machine			73.23 [16]							92,12 [10]			
CNN + LSTM												98.6 [12]	
AutoEncoders										91,86 [10]			98,99 [10]
Deep Defense(CNN, RNN, LSTM, GRU)		97,60 [10]											

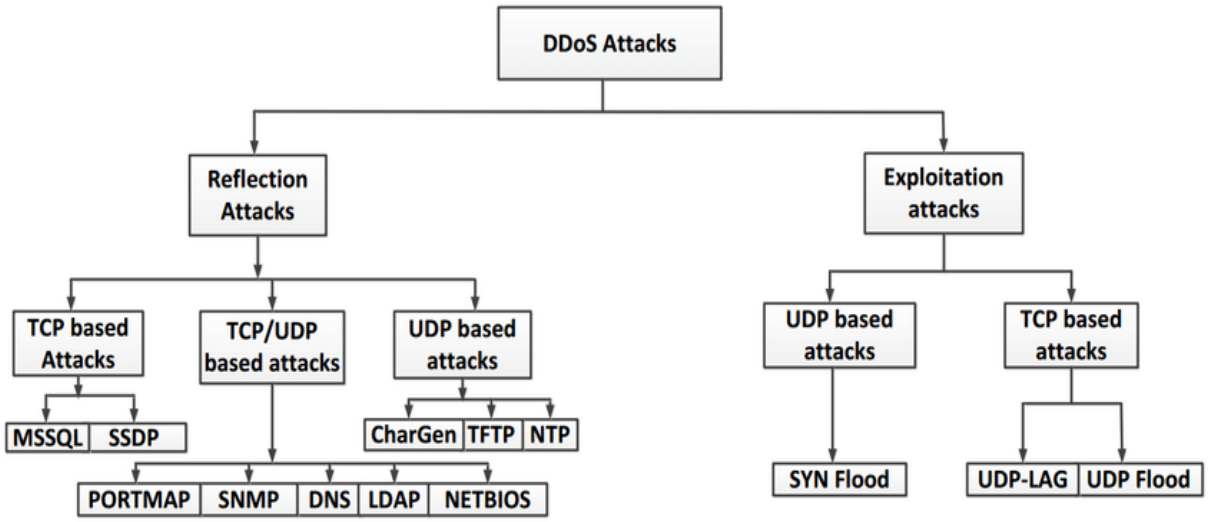
الشكل 5 Relationship between algorithms vs dataset of DL for DoS attacks

DATASET-8 :

كما يوحي الاسم، هي مجموعات من البيانات ذات الصلة المستخدمة لتدريب نموذج DL.
إن الخطوة الأكثر تحديًا في تقييم أداء أنظمة دفاع DoS هي العثور على البيانات المناسبة.
إحدى الطرق هي مراقبة وجمع المعلومات من الشبكة. ومع ذلك، نظرًا لأن جمع المعلومات من الشبكة أمر مكلف، فيمكن استخدام مجموعات البيانات المتاحة على الإنترنت.

مجموعة البيانات المستخدمة في هذه الدراسة هي CICDDoS2019 (مجموعة بيانات تقييم DDoS) .تحتوي مجموعة البيانات هذه على 11 نوعًا من هجمات حجب الخدمة (الشكل 6)، والتي تم إنشاؤها في بيئة خاضعة للرقابة، و88 ميزة مستخرجة باستخدام CICFlowMeter، وهي أداة تستخرج ميزات ملفات pcap. بالإضافة إلى ذلك، قام

منشئو مجموعة البيانات هذه (خبراء من المعهد الكندي للأمن السيبراني) بتطوير عملية لتحديد الميزات الأكثر صلة بكل هجوم DoS وحصلوا على 22 ميزة



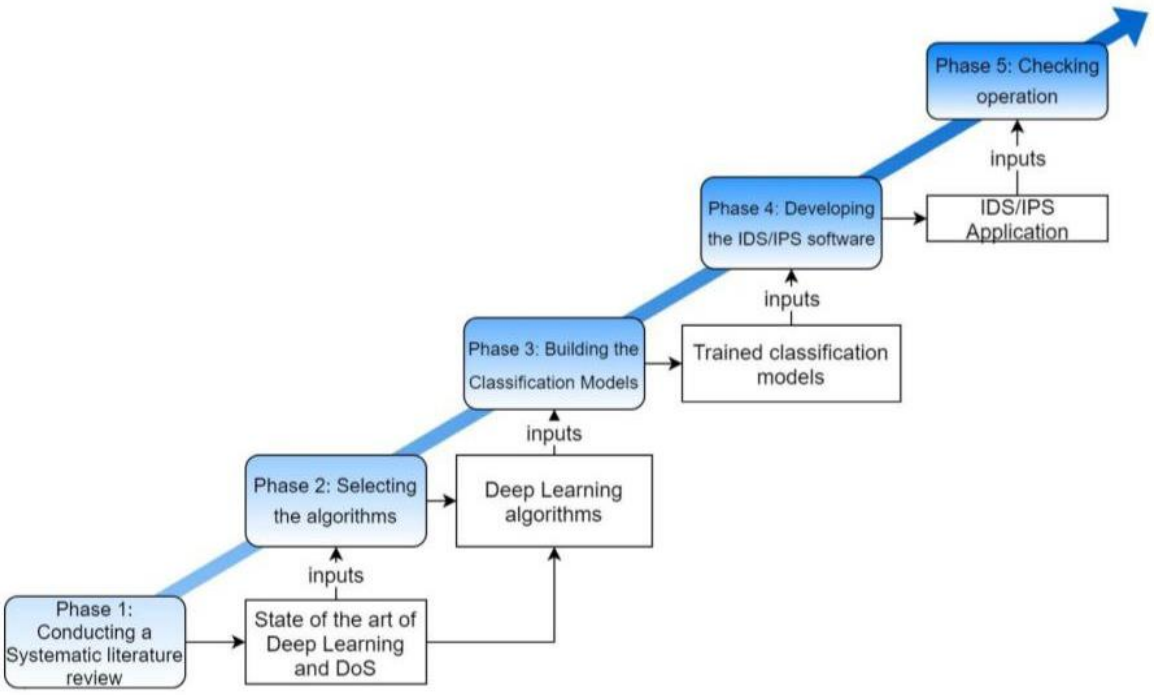
الشكل 6 Types of DDoS attacks on the CICDoS2019 dataset

وفيما يلي مقارنة بين النوعين الأساسيين لهجمات ال DDOS

المعايير	هجمات الانعكاس (Reflection Attacks)	هجمات الاستغلال (Exploitation Attacks)
طريقة العمل	تعتمد على استغلال خوادم مفتوحة على الإنترنت تقوم بالرد على طلبات غير مشروعة مرسلة من قبل المهاجم، حيث يتم تزوير عنوان المصدر ليكون عنوان الهدف. نتيجة لذلك، تستجيب الخوادم المفتوحة بطلبات مكثفة إلى الهدف، مما يؤدي إلى إغراقه.	تعتمد على استغلال نقاط الضعف في البرمجيات أو البروتوكولات لتحقيق انقطاع الخدمة أو تعطيلها. يمكن أن تكون هذه الهجمات موجهة نحو التطبيقات، الأنظمة، أو الشبكات.
التعقيد	متوسط إلى عالي	متوسط إلى عالي
الأهداف	الشبكات والخوادم	التطبيقات والأنظمة والشبكات
المثال الشائع	DNS Reflection, NTP Reflection	Slowloris, HTTP Flood

9- METHODOLOGY:

يوضح الشكل الأتي (الشكل 7)الخطوات المتبعة لانشاء نظام كشف التسلل (Dique) المعتمد على التعلم العميق:



الشكل 7 phase of the methodology implemented in this study

وهذه المراحل هي :

a) Conducting a systematic literature review:

تضمنت المرحلة الأولى من الدراسة إجراء مراجعة منهجية SYSTEMATIC LITERATURE REVIEW (SLR) لفهم الخوارزميات ومجموعات البيانات المستخدمة في البحث حول أساليب التعلم العميق (DL) وهجمات رفض الخدمة (DoS) بين عامي 2009 و2020. وتضمنت هذه المرحلة :

1. تحديد أسئلة البحث و تحديد مصطلحات البحث التي سيتم إدخالها في مربع البحث بقاعدة البيانات(التعلم العميق، رفض الخدمة، نظام كشف التسلل.....)
2. اختيار قواعد البيانات
3. تقييم جودة الدراسة
4. تحليل الدراسات ذات الصلة لتوصيف خوارزميات DL للكشف عن هجوم DoS ومجموعات البيانات المستخدمة لنماذج التدريب حيث يقدم الجدول السابق (الشكل 5)تلخيصا لاهم الخوارزميات وقواعد البيانات المستخدمة.

b) Selecting the algorithm:في هذه المرحلة تم العمل على ايجاد الخوارزمية الافضل من بين الخوارزميات التي تم جمع المعلومات حولها في الخطوة الاولى وذلك وفقا لعدة معايير (الشكل 8)

No.	Selection criteria
1	The algorithm is available in the Java Deep Learning library.
2	The model's training configuration used with the DL algorithm is available.
3	The dataset employed with the DL algorithm to train the model contains DoS attack packets.
4	The accuracy of the model trained with the DL algorithm is above 90%.

الشكل 8Selection criteria .

وبناءا على هذه المعيار تم اختبار عدة خوارزميات لتحديد مدى تطابقها مع المعاير وهي موضحة في الجدول الاتي (الشكل 9)

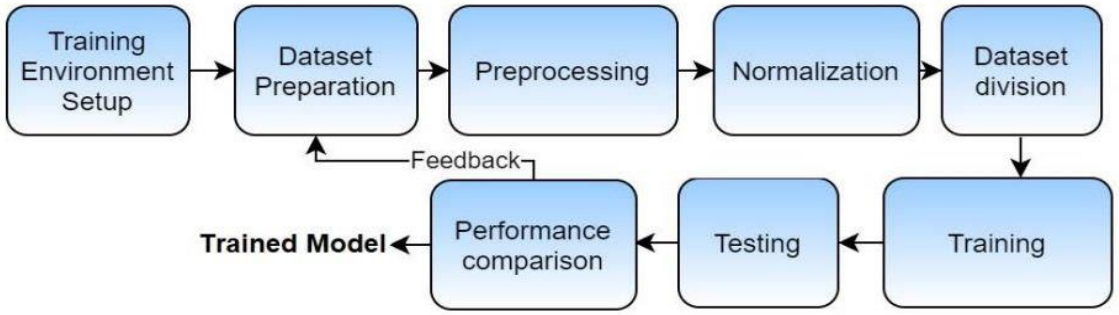
Algorithms \ Criteria	1	2	3	4	Adequate
Recurrent Neural Network (RNN)	✓	✓	✓	✓	✓
Convolutional Neural Network (CNN)	✓	X	✓	✓	
Stacked RNN	X	X	X	X	
Long Short-Term Memory (LSTM)	✓	✓	✓	✓	✓
Deep Radial Intelligence	X	X	✓	✓	
Improved Genetic Algorithm (IGA) and Simulated Annealing Algorithm (SAA)	X	X	✓	✓	
VCNN/FNN	X	X	✓	✓	
Deep Belief Network	✓	X	✓	✓	
Gated Recurrent Unit (GRU)	X	X	X	X	
Deep Feedforward	✓	✓	✓	✓	✓
Bidirectional Gated Recurrent Unit (BGRU) + Multi-Layer Perceptron (MLP)	X	X	✓	✓	
Restricted Boltzman Machine	X	X	✓	X	
CNN + LSTM	X	X	X	✓	
Autoencoders	X	X	✓	✓	
Deep Defense (CNN, RNN, LSTM, GRU)	X	X	✓	✓	

الشكل 9Algorithms vs. selection criteria

ومن خلال الجدول السابق (الشكل 9) نجد أن الخوارزمية التي حققت أعلى نسبة توافق مع المعايير الموضوعه(الشكل 8) هي Deep Feedforward

c) Building the classification models:

يصف الشكل الاتي (الشكل 10) كيفية بناء نماذج التصنيف باستخدام DL. سيتم شرح كل خطوة متبعة لبناء نماذج التصنيف أدناه



الشكل 10 Steps of building the classification models

وفيما يلي شرح الخطوات

1. Training environment setup :

تم اختيار نظام التشغيل Debian لتدريب النموذج بسبب استهلاكه المنخفض للموارد. يتم تفصيل مواصفات الخادم المستخدمة للتدريب النموذجي في الجدول التالي(الشكل 11)، مما يوفر معلومات حول إعداد الأجهزة والبرامج المستخدمة في بيئة التدريب. يعد هذا الاختيار لنظام التشغيل ومواصفات الخادم أمرًا بالغ الأهمية لتدريب نظام كشف التسلل ومنعه القائم على التعلم العميق والذي تمت مناقشته في البحث بكفاءة.

TABLE 4. Training server characteristics.

Component	Specification
Processor	Intel Xeon 1.90 GHz 64-bit, 8 cores
RAM	32 GB DDR4 2666
Hard disk	Magnetic, 100 GB
Operating system	Debian 10
IDE	Eclipse
Libraries	Java 1.8, DL4J

الشكل 11 Training server characteristics

2. **Dataset preparation** :تم اختيار مجموعة بيانات CICDDoS2019 لتدريب النموذج نظرًا لحداثتها وحجمها الكبير، حيث تحتوي على أكثر من 48 مليون سجل لهجمات رفض الخدمة (DoS) حصريًا. تم تقسيم مجموعة البيانات هذه إلى 11 ملفًا (الشكل 12)، يحتوي كل منها على معلومات مفصلة عن حزم الشبكة وبيانات التدفق، وهي ضرورية لتدريب نظام كشف التسلل ومنعه باستخدام تقنيات التعلم العميق.

TABLE 5. Files from the CICDDoS2019 dataset.

Name	Size	Number of records
DrDoS_DNSFile	2.0 GiB	5074414
DrDoS_LDAPFile	874.8 MiB	218153
DrDoS_MSSQLFile	1.8 GiB	4524499
DrDoS_NetBIOSFile	1.6 GiB	4094987
DrDoS_NTPFile	615.1 MiB	1217008
DrDoS_SNMPFile	2.0 GiB	5161378
DrDoS_SSDPFile	1.2 GiB	2611375
DrDoS_UDPFile	1.4 GiB	3136803
SynFile	607.8 MiB	1582682
TFTPFile	8.7 GiB	20107828
UDPLagFile	150.7 MiB	370606

الشكل 12 Files from the CICDDoS2019 dataset

يمثل كل سجل معلومات في حزمة الشبكة مكونة من 88 متغيرًا تحتوي على بيانات مثل عنوان IP المصدر وعنوان IP الوجهة ومنفذ الوجهة ومنفذ المصدر والبروتوكول. تتضمن هذه الحزمة أيضًا معلومات التدفق التي تم استخراجها باستخدام أداة CICFlowMeter المستخدمة لاستخراج الميزات. يعرض العمود الأخير في كل ملف متغير LABEL، الذي يشير إلى نوع هجوم DoS الذي ينتمي إليه كل سجل. إذا لم يكن السجل جزءًا من هجوم DoS، فسيتم إدراجه على أنهBENIGN .

وفقًا للدراسة التي أجراها المعهد الكندي للأمن السيبراني، هناك 22 متغيرًا أكثر صلة، من بين 88 متغيرًا موجودة في مجموعة البيانات، من حيث أهمية كل خاصية لكل فئة. على سبيل

المثال، وفقًا لدراستهم، يعد متغير Packet_lenght_std واحدًا من أكثر المتغيرات ذات الصلة بحزم BENIGN. يسرد الجدول التالي (الشكل 13) هذه الميزات الـ 22 ويصفها، والتي سيتم استخدامها لتدريب النماذج

Feature	Description
Destination Port	Destination port
Protocol	Protocol number in the IP header
Flow Duration	Flow duration in microseconds
Fwd Packet Length Max	Maximum packet size in source-to-destination direction
Fwd Packet Length Min	Minimum packet size in source-to-destination direction
Fwd Packet Length Std	Standard deviation of packet size in source-to-destination direction
Flow IAT Mean	Mean time between two packets sent in the flow
Flow IAT Max	Maximum time between two packets sent in the flow
Fwd IAT Mean	Mean time between two packets sent to the source-to-destination direction
Fwd IAT Max	Maximum time between two packets sent to the source-to-destination direction
Fwd IAT Min	Minimum time between two packets sent source-to-destination direction
Fwd Header Length	Total bytes used by headers at the source-to-destination direction
Fwd Packets/s	Number of packets sent per second from source to destination
Min Packet Length	Minimum length of a packet
Max Packet Length	Maximum length of a packet
Packet Length Std	Standard deviation of the length of a packet
ACK Flag Count	Number of packets with ACK
Average Packet Size	Average packet size
Fwd Header Length.1	Packet header length in source-to-destination direction
Subflow Fwd Bytes	Average number of bytes in a subflow in the source-to-destination direction
Init_Win_bytes_forward	Total bytes sent in an initial window at the source-to-destination direction
min_seg_size_forward	Minimum size of the segment observed in the source-to-destination direction

الشكل 13 List of features.

3. DATASET PREPROCESSING :

في المعالجة المسبقة لمجموعة البيانات لنظام كشف التسلل ومنعه، يجب أن تكون قيم التسمية (output labels) رقمية حتى تعمل خوارزمية التدريب بفعالية. ولتحقيق ذلك، تم تحويل هذه القيم من تنسيقها الاصلي (سلاسل) إلى قيم رقمية، حيث يمثل "1" بيانات ضارة و"0" يمثل بيانات

حميدة. يسمح هذا التحويل للخوارزمية بمعالجة البيانات والتعلم منها بشكل أكثر كفاءة أثناء التدريب.

4. DATASET NORMALIZATION:

تتضمن تسوية مجموعة البيانات تسوية نطاق بيانات الإدخال إلى نطاق معين، عادة بين -1 و1. تعد هذه العملية ضرورية للشبكات العصبية لأنها تساعد على تحسين أداء التدريب من خلال ضمان أن تقع البيانات ضمن النطاق الأمثل لوظائف التنشيط المستخدمة في الشبكة. يؤدي تطبيع الميزات الـ 22 في مجموعة البيانات إلى تعزيز عملية التدريب من خلال جعل البيانات أكثر ملاءمة لتدريب الشبكات العصبية، مما يؤدي في النهاية إلى أداء تصنيف أفضل.

5. DATASET DIVISION :

تتضمن عملية تقسيم مجموعة البيانات تقسيم البيانات إلى مجموعات تدريب واختبار لتقييم أداء النموذج. في حين أن الممارسة القياسية هي تقسيم مجموعة البيانات إلى 70% للتدريب و30% للاختبار، في هذا البحث، تم اختيار توزيع مختلف مع تخصيص 90% من البيانات للتدريب و10% للاختبار، مع إعطاء الأولوية لمرحلة التدريب. تم تنفيذ هذا التقسيم باستخدام برنامج تم إعداده بلغة Java.

6. Training:

خلال مرحلة التدريب على نظام كشف التسلل ومنعه القائم على التعلم العميق، تم اختبار تكوينات الشبكة العصبية المختلفة للعثور على النموذج الذي حقق أفضل أداء في تصنيف حزم الشبكة على أنها ضارة أو حميدة. تم تطوير إجمالي 20 نموذج تصنيف، ويعرض الجدول التالي (الشكل 14)تفاصيل أفضل نموذج تم الوصول اليه وتهدف هذه العملية إلى تحسين أداء النموذج في تحديد وتصنيف حركة مرور الشبكة بدقة لأغراض كشف التسلل.

Proposed model			
Seed		7	
Epochs		200.000	
Data size		44.000	
Batch size		44.000	
Activation function		tan	
Initial weight values		Xavier	
Updater		Adam 0.001	
Features		73	
Layer 1	InputLayer	Input	73
		Output	73
Layer 2	DenseLayer	Input	73
		Output	128
Layer 3	DropoutLayer	0.5	
Layer 4	DenseLayer	Input	128
		Output	256
Layer 5	DenseLayer	Input	256
		Output	512
Layer 6	DenseLayer	Input	512
		Output	256
Layer 7	DenseLayer	Input	256
		Output	128
Layer 8	DropoutLayer	0.5	
Layer 9	OutputLayer	Input	128
		Output	2

الشكل 14 Final configuration for training

7. Testing :

بعد تدريب النموذج، تم حفظه في ملف ثنائي (.bin). أثناء مرحلة الاختبار، تم تحميل النموذج وتقييمه باستخدام 10% من مجموعة البيانات المخصصة للاختبار. تم استخدام مقاييس الأداء مثل الدقة والضبط والاستدعاء ودرجة F1 لتقييم فعالية نموذج التعلم العميق المُدرب في تصنيف حزم الشبكة على أنها ضارة أو حميدة.

وفيما يلي شرح لمقاييس الاداء:

- **True Positives (TP)** عدد العينات التي تم التنبؤ بها بشكل صحيح كإيجابية.
- **True Negatives (TN)** عدد العينات التي تم التنبؤ بها بشكل صحيح كسلبية.
- **False Positives (FP)** عدد العينات التي تم التنبؤ بها بشكل خاطئ كإيجابية (أي أنها سلبية في الحقيقة).
- **False Negatives (FN)**: عدد العينات التي تم التنبؤ بها بشكل خاطئ كسلبية (أي أنها إيجابية في الحقيقة).

الدقة(Accuracy) :

هي نسبة التنبؤات الصحيحة (كل من الإيجابيات الحقيقية والسلبيات الحقيقية) إلى مجموع جميع التنبؤات. تعطي فكرة عامة عن عدد التنبؤات الصحيحة من إجمالي التنبؤات. مناسبة عندما تكون الفئات متوازنة. .يتم حسابها باستخدام العلاقة الرياضية التالية:

$$Accuracy = TP + TN \ / \ TP + TN + FP + FN$$

الضبط(Precision) :

هو نسبة الإيجابيات الحقيقية إلى مجموع الإيجابيات الحقيقية والإيجابيات الخاطئة. مهم عندما تكون تكلفة الإيجابيات الخاطئة عالية. يتم حسابه باستخدام العلاقة الرياضية التالية:

$$Precision = TP \ / \ TP + FP$$

الاستدعاء(Recall) :

هو نسبة الإيجابيات الحقيقية إلى مجموع الإيجابيات الحقيقية والسلبيات الخاطئة. مهم عندما تكون تكلفة السلبيات الخاطئة عالية. يتم حسابه باستخدام العلاقة الرياضية التالية

$$Recall = TP \ / \ TP + FN$$

درجة(F1 Score)

هي مقياس يجمع بين الضبط والاستدعاء في قيمة واحدة. توازن بين الضبط والاستدعاء، وتكون مفيدة عندما نحتاج إلى توازن بينهما .يتم حسابه باستخدام المتوسط التوافقي للضبط والاستدعاء:

$$F1 = 2 * ((Precision * Recall) / (Precision + Recall))$$

8. MODEL PERFORMANCE COMPARISON :

تضمنت خطوة "مقارنة أداء النماذج" تقييم النماذج المدربة بناء على متغيرات محددة لتحديد كفاءتها. ومن خلال مقارنة نتائج كل نموذج، تمكن الباحثون من تحديد النموذج الأفضل أداءً وإجراء أي تعديلات مطلوبة لتحقيق الأداء الأمثل. تم عرض نتائج هذه المقارنة في الجدول الاتي(الشكل 15) ، مع عرض أداء كل نموذج مدرب لمزيد من التحليل والتحسين

Model No.	Features	Labels	Accuracy	Precision	Recall	F1-Score
1	22	2	0.9582	0.9711	0.9835	0.9772
2	22	2	0.96	1	0.9321	0.9649
3	22	2	0.9785	0.9889	0.9765	0.9836
4	22	2	0.9282	0.9997	0.9215	0.959
5	22	2	0.9266	0.9831	0.9348	0.9583
6	73	2	0.9823	0.9997	0.9807	0.9901
7	73	2	0.9847	0.9996	0.9841	0.9918
8	73	2	0.9064	0.9064	1	0.9509
9	73	2	0.9818	0.9982	0.9817	0.9899
10	73	2	0.9859	0.9972	0.9872	0.9922
11	73	2	0.9093	0.9944	0.9038	0.947
12	73	2	0.9889	0.9913	0.9965	0.9939
13	73	2	0.9814	0.9982	0.9812	0.9896
14	73	2	0.9842	0.9966	0.9859	0.9912
15	73	2	0.9722	0.9893	0.9798	0.9845
16	73	2	0.9022	0.9022	1	0.9486
17	73	2	0.9952	0.998	0.9914	0.9947
18	73	2	0.997	0.9975	0.996	0.9967
19	73	2	0.998	0.998	0.997	0.9975
20	73	2	0.9994	0.9995	0.999	0.9993

```
=========================Evaluation Metrics=========================
 # of classes:     2
 Accuracy:         0.9994
 Precision:        0.9995
 Recall:           0.9990
 F1 Score:         0.9993
Precision, recall & F1: reported for positive class (class 1 - "1") only
```

الشكل 15 Performance of the proposed models

d) Developing the IPS/IDS software :

تتضمن المرحلة الرابعة انشاء برنامج لنظام كشف التسلل (IDS) ونظام منع التسلل (IPS) لاكتشاف ومنع هجمات رفض الخدمة (DoS). يدمج هذا البرنامج نموذج التعلم العميق (DL) الذي تم تدريبه بأفضل أداء لتعزيز قدرة النظام على تحديد التهديدات المحتملة وتخفيفها في الوقت الفعلي على الشبكة. تم تصميم البرنامج لالتقاط حزم الشبكة، وتصنيفها على أنها

ضارة أو حميدة، وعرض معلومات الحزم ذات الصلة، وتخزين بيانات حركة مرور الشبكة لمزيد من التحليل والتحقق من الصحة.

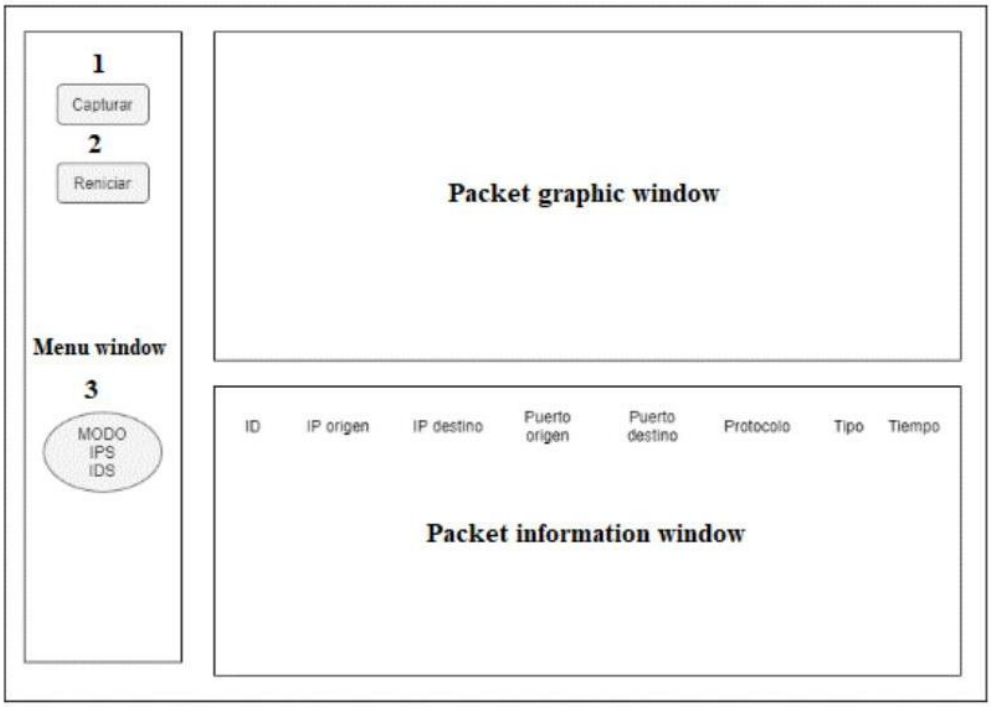
تقسم هذه المرحلة الى عدة مراحل:

a) **REQUIREMENT ANALYSIS** : فيما يلي المتطلبات الوظيفية الأكثر صلة المقترحة لتطوير Dique :

1) يجب أن يلتقط النظام الحزم من واجهة الشبكة
2) يجب أن يكون النظام عبارة عن تطبيق ويب حتى يمكن مشاهدته من أي جهاز
3) يجب أن يكون لدى النظام ألية تسجيل دخول حتى يتمكن المستخدمون المصرح لهم فقط من الوصول إليه
4) يجب على النظام تصنيف حزم الشبكة إلى نوعين: ضارة وحميدة
5) يجب أن يقوم النظام بتصنيف الحزم في الوقت الفعلي
6) يجب أن يعرض النظام الوقت وعنوان IP المصدر وعنوان IP الوجهة ومنفذ المصدر ومنفذ الوجهة وبروتوكول الحزم الملتقطة
7) يجب أن يشتمل النظام على زر لبدء وإيقاف التقاط الحزم
8) يجب أن يكون لدى النظام زر لإعادة تشغيل التقاط الحزم (إعادة التشغيل)
9) يجب أن يقوم النظام بتخزين كل حركة مرور الشبكة في ملف pcap لمزيد من التحليل
10) يجب إنشاء قاعدة بيانات لتخزين البيانات الخاصة بالمستخدمين المعتمدين

b) SYSTEM DESIGN :

يعرض الشكل 16 واجهة المستخدم الرسومية المصممة للنظام الوقائي المقترح (Dique)



الشكل 16 Scheme of the graphical user interface designed for the proposed preventive system

تشتمل واجهة المستخدم الرسومية المصممة على ثلاثة مكونات:

- Menu window : ويحتوي على الأزرار الثلاثة التالية التي تتيح للمستخدمين التحكم في التطبيق: (1) زر الالتقاط الذي يبدأ عملية التحليل أو التصنيف. (2) زر إعادة التشغيل، وله وظيفتان: الإيقاف أو إعادة التشغيل (تتغير تسميته حسب حالته الحالية). و(3) زر وضع IDS/IPS، الذي يقوم بتبديل وضع التنفيذ بين الكشف والمنع.

- Packet graphic window: ويحتوي على رسم بياني ثنائي الأبعاد يوضح في الوقت الفعلي عدد الحزم التي يصنفها النظام. يشير المحور الأفقي إلى الوقت بالثواني؛ ويشير المحور الرأسي إلى عدد الحزم المصنفة.

- Packet information window : ويحتوي على جدول يعرض معلومات حول الحزم مثل عنوان IP المصدر، وعنوان IP الوجهة، ومنفذ المصدر، ومنفذ الوجهة، والبروتوكول، والنوع (ضار/حميد)، والوقت الذي تم فيه التقاطها.

C) ENCODING :

يشير "الترميز" إلى عملية ترجمة المتطلبات المحددة وبنية النظام والتصميم إلى كود فعلي باستخدام لغة برمجة محددة، وهنا تم استخدام لغة Java. استخدم المطورون بيئة التطوير المتكاملة Eclipse (IDE) لكتابة وتنظيم التعليمات البرمجية للنظام الوقائي.

10) RESULTS :

- **SELECTING THE ALGORITHMS :** وضحنا سابقا من خلال الشكل 8 المعايير التي على أساسها اخترنا الخوارزمية المناسبه وكذلك وضحنا من خلال الشكل 9 مدى توافق مجموعة من الخوارزميات مع هذه المعايير ومن خلال ذلك تبين أن الخوارزميات الأكثر تطابقا مع المعايير هي Deep Feedforward وRNN-LSTM الا أنه لم نتمكن من استخدام خوارزمية RNN أو LSTM وذلك لان مكتبة DL4J تطلبت أن تكون مجموعة البيانات منظمة بالفعل ولا تحتاج إلى معالجة مسبقة على الرغم من أنها حققت دقة عالية (الشكل 17)

Algorithm	Accuracy	Precision	Recall	F1-Score
Deep Feedforward	0.9	0.9353	0.9579	0.9464
RNN and LSTM	0.9555	0.9824	0.9222	0.9513

الشكل 17 Algorithms' performance

➢ BUILDING THE CLASSIFICATION MODELS :

وضح الشكل 15 قيم الأداء التي حصلت عليها النماذج المختلفة التي تم تدريبها باستخدام الخوارزمية المحددة DFNN ومن خلال هذا الشكل نجد أن النموذج الذي تم تدريبه بأفضل دقة هو النموذج رقم 20 مع علامتين تصنيفيتين (خبيثة وحميدة). لقد تم البدء في تدريب النماذج باستخدام 22 ميزة ذات صلة، ثم استخدمنا 73 ميزة لتحسين أدائها.

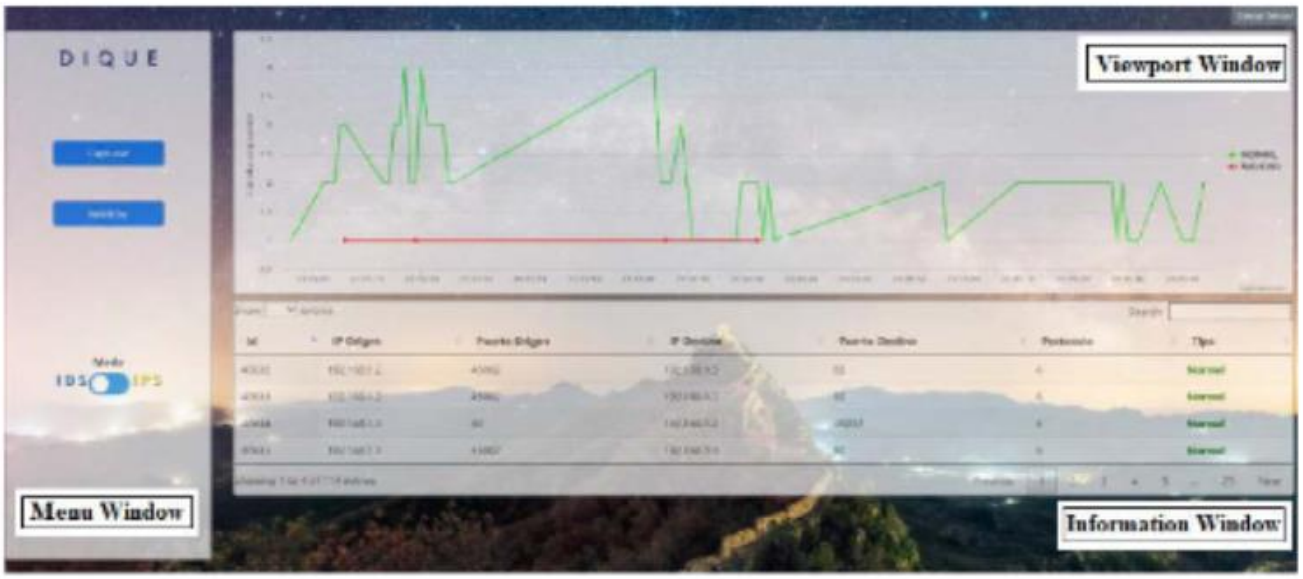
بعد التدريب، تم حفظ كل نموذج في ملف .zip والذي يتضمن ملفين .bin يحتويان على تكوين النموذج وأوزانه. بالإضافة إلى ملف json يعرض تفاصيل تكوين الشبكة العصبية. يعد الملف .zip، المكتوب بلغة Java البرمجية، هو المتطلب الوحيد للنظام الوقائي لاستخدام النموذج المُدرب بشكل فعال ويمكن استيراد هذه الملفات أو تحميلها في النظام الوقائي لاستخدامها.

➢ DEVELOPING THE IDS/IPS (INTRUSION PREVENTION/DETECTION SYSTEM) SOFTWARE :

LOGIN PAGE-1 : تحتوي على نموذج تسجيل دخول يتم من خلاله الوصول إلى التطبيق. يتم تخزين المستخدمين المسجلين في قاعدة بيانات، إلى جانب كلمات المرور المشفرة الخاصة بهم.(الشكل 18)

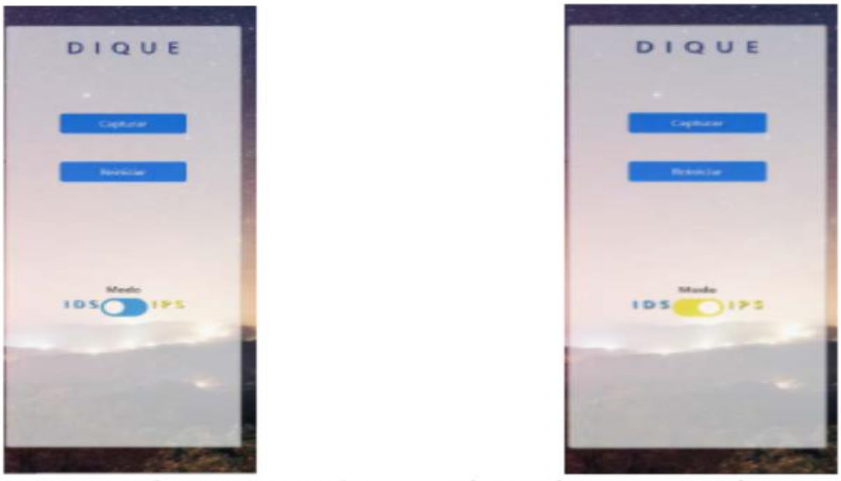
الشكل 18 Login page of Dique

: HOME PAGE-2



HOME PAGE 19 الشكل

: MENU WINDOW-3



A) IDS (Detection) active mode B) IPS (Prevention) active mode

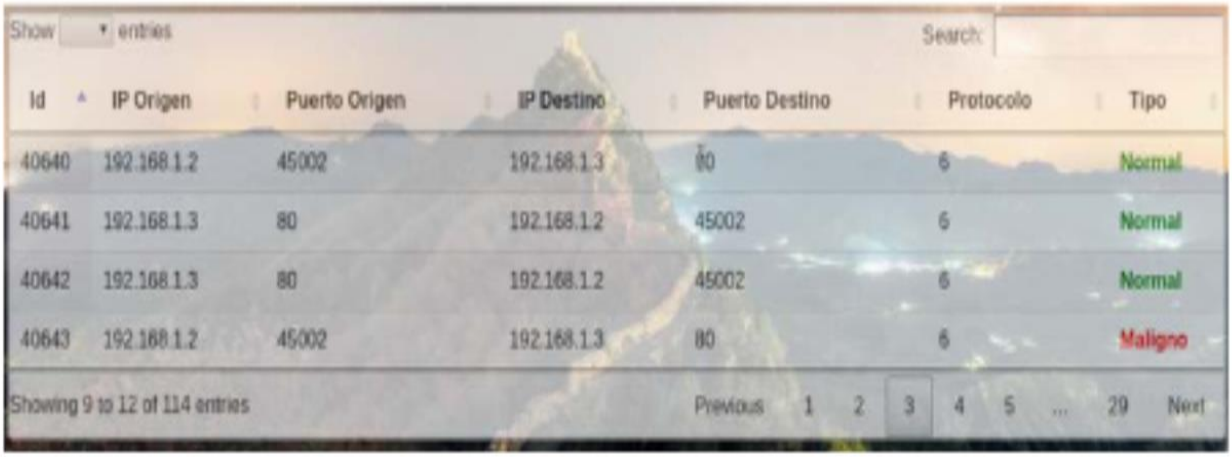
MENU WINDOW 20 الشكل

: VIEWPORT WINDOW-4

اللون الأحمر: الحزم الخبيثة
اللون الأخضر: الحزم الحميدة
المحور الافقي: الوقت بالثواني
المحور العمودي: عدد الحزم

الشكل 21 Packet graphical section in Dique

: INFORMATION WINDOW-5



Id	IP Origen	Puerto Origen	IP Destino	Puerto Destino	Protocolo	Tipo
40640	192.168.1.2	45002	192.168.1.3	80	6	Normal
40641	192.168.1.3	80	192.168.1.2	45002	6	Normal
40642	192.168.1.3	80	192.168.1.2	45002	6	Normal
40643	192.168.1.2	45002	192.168.1.3	80	6	Maligno

Show entries — Search:
Showing 9 to 12 of 114 entries — Previous 1 2 3 4 5 ... 29 Next

الشكل 22 Packet information section in Dique

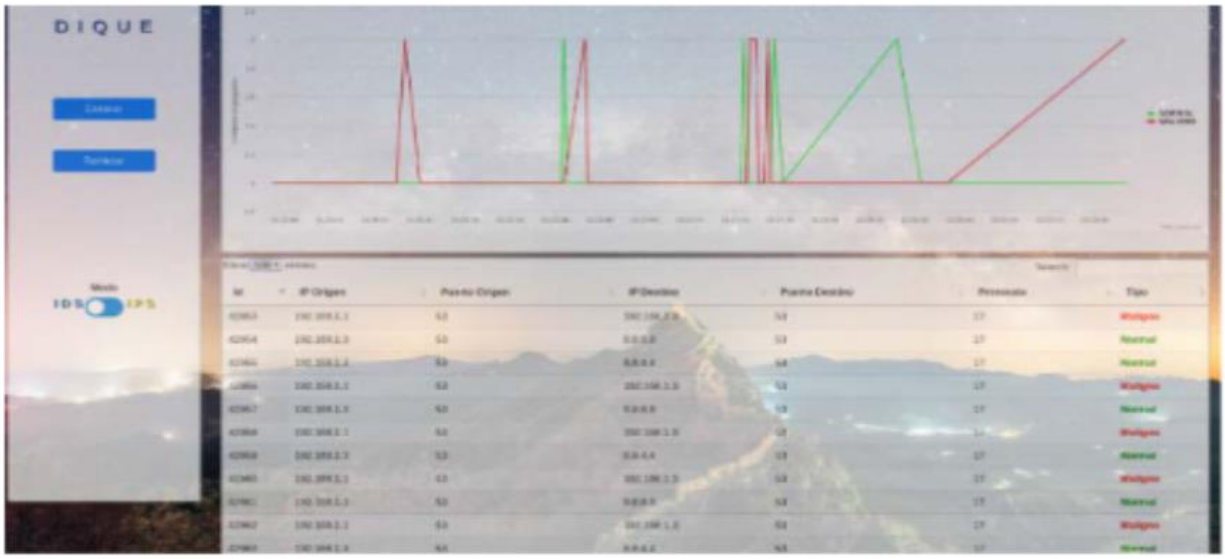
CHECKING DIQUE's OPERATION-6 :

تم اختبار النظام وإخضاعه لعدة أنواع من الهجمات وفيما يلي بعض هذه الهجمات وكيفية تعامل النظام مهعا

- ENIGN HTTP REQUEST : تم إرسال طلب حميد إلى خادم الويب، قام Dique بتصنيف جميع الحزم على أنها عادية، كما هو موضح في الشكل

الشكل 23 View from Dique of a normal HTTP request sent to the web server.

- DNS REFLECTION ATTACK :



الشكل 24 View from Dique of a DNS reflection attack

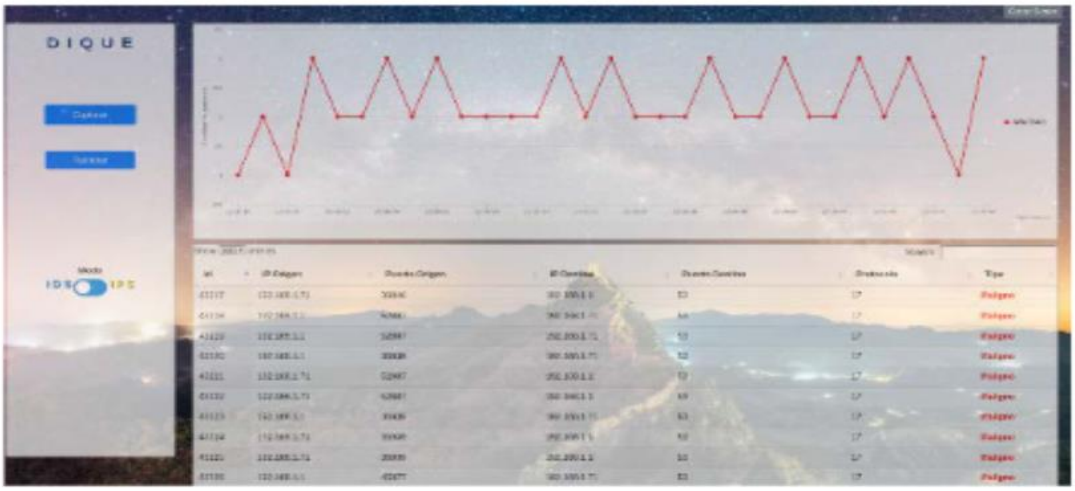
- : NTP REFLECTION ATTACK

View from Dique of a NTP reflection attack 25 الشكل

- : SYN FLOOD ATTACK

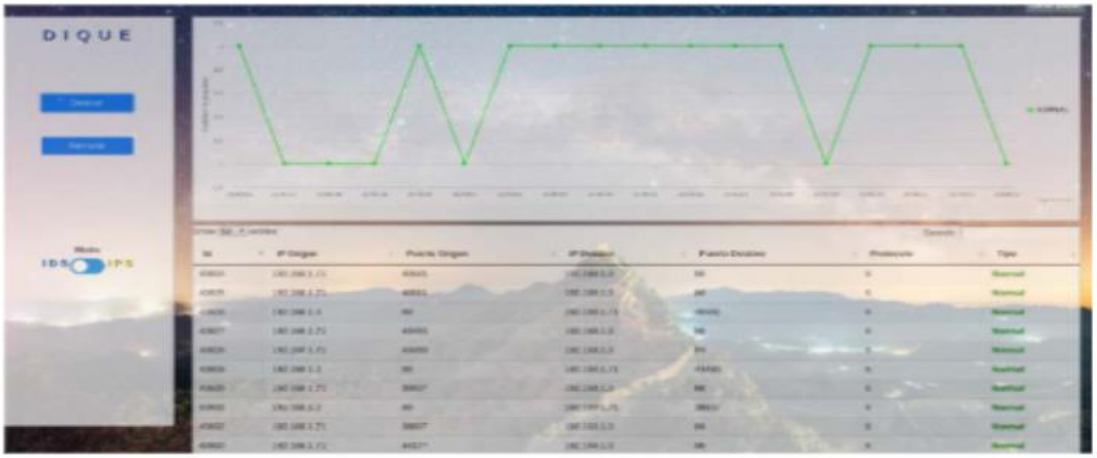
.View from Dique of a SYN flood attack 26 الشكل

- : UDP FLOOD ATTACK



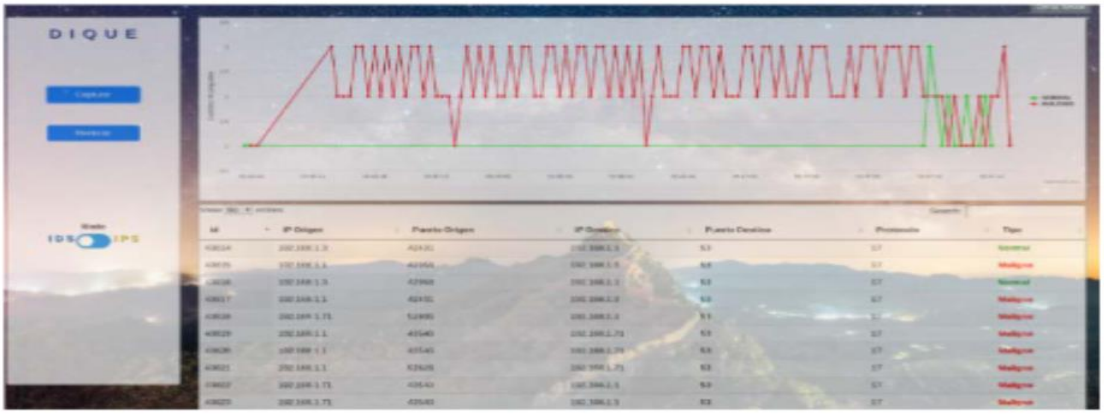
View from Dique of a UDP flood attack 27 الشكل

- : TCP FLOOD ATTACK



View from Dique of a TCP flood attack 28 الشكل

- ICMP FLOOD ATTACK :



الشكل 29 . View from Dique of a ICMP flood attack.

- SLOW HTTP ATTACK :

الشكل 30 View from Dique of a Slow HTTP attack

11) CONCLUSION AND FUTURE WORK :

على الرغم من أن النماذج المدربة المقترحة قد حققت دقة عالية في الأداء، إلا أن هذا لا يضمن من الناحية العملية أنها تصنف على النحو الأمثل الحزم غير المعروفة التي تدخل إلى خادم الويب، نظرًا لأن احتمال التعرف على الحزم مرتفع، ولكن لديه درجة من الخطأ كما أوضحت عملية اختبار النظام على عدة انواع من الهجمات. من الناحية النظرية من المتوقع الحصول على نسبة تصنيف أعلى، إذا تم تدريب النموذج بمزيد من البيانات.

كعمل مستقبلي، يوصى بما يلي:

- ✓ تطوير خوارزمية DL جديدة لهجمات DoS. استنادًا إلى DFNN أو تحسين أداء الأنظمة الحالية.
- ✓ تحسين التدريب على نماذج التصنيف التي تم الحصول عليها في هذه الدراسة، بحيث يمكن للنظام الوقائي المقترح التصنيف في فئات أكثر (على سبيل المثال، التنبؤ بنوع هجوم حجب الخدمة) مع احتمالية أكبر. •
- ✓ إضافة المزيد من البيانات إلى مجموعة بيانات التدريب

: REFERENCES(12

1. Canola Garcia, & Blandon. (2022). A deep learning-based intrusion detection and Preventation system for detecting and preventing denial-of-service attacks. *IEEE Access*, *10*, 83043-83060.