تقييم أنظمة كشف هجمات إنترنت الأشياء المعتمدة على تقنيات التعلم والميزات المحددة على مستوى التدفق والحزمة

***طالب الماجستير : م. مجد المحسن**

****المشرف العلمي: أ.د. إبراهيم الشامي ***المشرف المشارك: أ.د. علي ذياب**

الملخص:

يُمثل الأمن تحديًا حاسمًا لمكونات إنترنت الأشياء حيثُ إنها معرضة بشدةً للهجمات بسبب طبيعتها المحدودة لذلك لا بدا من الكشف عن الهجمات وإطلاق الإنذار في حال ملاحظة أي خطر أو محاولة استغلال . يتعمق هذا العمل في دراسة ومراجعة التطورات الحديثة في أنظمة الكشف عن هجمات إنترنت الأشياء حيث قدمت هذه الدراسة مسحاً للعديد من أنظمة الكشف المعتمدة على تقنيات التعلم الآلة والعميق والتعلم الجماعي مع تسليط الضوء على البيانات المستخدمة وأنواع الهجوم وتقديم دراسة موسعة للهجمات الأكثر خطورة في شبكات إنترنت الأشياء لتحديد وتفسير الميزات التي تعمل على تحسين أنظمة الأمن لكي يتم في النهاية تقييم هذه الأنظمة على مستوى تدفق الشبكة وعلى مستوى الحزمة حيثُ حققت نماذج الكشف بالاعتماد على الميزات المحددة التي بلغ عددها 12 ميزة على مستوى التدفق و 14 ميزة على مستوى الحزمة نتائج على الشكل

***مجد المحسن** طالب **ماجستير ومعيد** في قسم هندسة التحكم الآلي والحواسيب كلية الهندسة الميكانيكية والكهربائية – جامعة البعث .

** دكتور **إبراهيم الشامي** أستاذ في قسم هندسة التحكم الآلي والحواسيب كلية الهندسة الميكانيكية والكهربائية – جامعة البعث

*** دكتور **علي ذياب** أستاذ في قسم هندسة التحكم الآلي والحواسيب كلية الهندسة الميكانيكية والكهربائية – جامعة البعث

التالي: وصلت النماذج المعتمدة على التعلم الجماعي الى أفضل دقة بلغت 99.97% ونسبة FP منخفضة بلغت 0.00018 على مستوى التدفق في حين حققت على مستوى الحزمة دقة 100% ونسبة FP معدومة أي إن الميزات المحددة حققت السرعة والدقة العالية ونسبة FP منخفضة خلال زمن تدريب صغير . كما تمت الاشارة الى التحديات التي تواجهها إنترنت الأشياء في نشر أنظمة الكشف المعتمدة على تقنيات التعلم وتقترح المجالات المحتملة للبحث المستقبلي. سيُرشد هذا الاستطلاع الباحثين والمهتمين في المجال الأمني بمعلومات تمكنهم من بناء نماذج كشف خفيفة الوزن ومناسبة لعملية النشر في أجهزة إنترنت الأشياء .

الكلمات المفتاحية: إنترنت الأشياء، كشف الهجوم ، تقنيات التعلم ، مستوى التدفق، مستوى الحزمة.

Evaluation of IoT Attack Detection Systems Based on Learning Techniques and Specific Features at the Flow and Packet Levels

Prepared by : **Eng. Majd Almohsen**

Supervisor : **Prof. Dr Ibrahim Chami**

Co-Supervisor : **Prof. Dr Ali Diab**

Abstract

Security represents a critical challenge for IoT components, as they are highly vulnerable to attacks due to their limited nature. Therefore, it is necessary to detect attacks and raise an alarm if any threat or attempted exploitation is noticed. This work delves into the study and review of recent developments in Internet of Things attack detection systems where This study presented a survey of many detection systems based on machine, deep, and ensemble learning techniques, Highlighting the data used and attack types and providing an extensive study of the most dangerous attacks in IoT networks to identify and explain features that improve security systems in order to finally evaluate these systems at the network flow level and at the packet level . the detection models based on the specific features, which numbered 12 features at the flow level and 14 features at the packet level, achieved results as follows: Models based on group learning reached the best accuracy. It reached 99.97% and a low FP ratio of 0.00018 at the flow level, while at the packet level it achieved an accuracy of 100% and a FP ratio of zero. The selected features achieved high speed, accuracy, and low FP ratio within a small training time . The challenges faced by the Internet of Things in deploying detection systems based on learning techniques are also pointed out and potential areas for future research are suggested. This survey will guide researchers and those interested in the security field with information that enables them to build lightweight detection models suitable for deployment in IoT devices.

Keywords: IOT , Attacks Detection, Learning Techniques , Netflow-Level , Packet-Level

1- مقدمة

ظهرت إنترنت الأشياء في عام 2002 بعبارة تناولتها مقالة كيفن أشتون (Kevin Ashton) بمجلة فوربس (Forbes) عندما قال: " نحن بحاجة إلى ' الإنترنت من أجل الأشياء ' وهي طريقة موحدة لأجهزة الحاسوب لتُدرك العالم الحقيقي" [1] . إنها لا تزال منصة متزايدة ومتوسعة، ونتيجة هذا التوسع المتزايد تواجه إنترنت الأشياء العديد من التحديات والقضايا التي يجب معالجتها من قبل الدراسات البحثية المختلفة. يُمثل الأمن تحديًا حاسمًا لمكونات إنترنت الأشياء حيث أنها معرضة بشكل كبير للهجمات لعدة أسباب متعلقة بطبيعتها والمتمثلة بمحدودية الموارد وطبيعة النشر ومساحة التخزين الصغيرة وبالتالي لا يمكنها تنفيذ مخططات معقدة تدعم الأمن لحماية الأجهزة والشبكات المتصلة [2] . يعتبر الجدار الناري المستوى الأول من الحماية [3] على سبيل المثال في حال قام مدير الشبكة بإعداد قواعد الجدار الناري لتصفية بعض المنافذ ومنع الوصول اليها عندما يقوم المهاجمون بمحاولة الوصول الى هذه المنافذ سيتم منعهم بواسطة الجدار الناري أما عندما يستثني بعض المنافذ من التصفية وحاول المهاجم الوصول الى هذه المنافذ يحدث عندها اختراق للمخدم دون إطلاق أي تحذير من قبل الجدار الناري ومن جهة أخرى يقوم بحظر بعض طلبات الاتصال الشاذة لرسائل المصادقة لبروتوكول TCP عندما يقوم المهاجم بالتجسس على الشبكة لمعرفة حالة المنافذ والخدمات المتوفرة بهدف القيام بهجوم فعلي لاحقاً [4] . لذلك لا بدا من مكاملة الجدار الناري مع مجموعة من المكونات التي تسهم في تحقيق الحماية الكاملة للشبكة حيثُ يُعتبر نظام كشف الاختراقات الذي يعمل كنظام كشف وتنبيه أحد أهم مكونات الحماية في كشف هجوم التسلل الى الشبكة وذلك بمراقبة الأحداث التي تحدث في الأنظمة أو الشبكات وتحليلها ومن خلال ملاحظة التغيرات في السلوك يمكن الكشف عن وجود هجمات ضمن الشبكة أو باستخدام مطابقة التوقيعات [5] التي تم إنشاء توقيع لها مسبقًا . يتم نشر أنظمة الكشف بأماكن موزعة أو أماكن مركزية من الشبكة [6] حسب مستوى الكشف وأيضاً إن منع وكشف محاولات الهجوم على إنترنت الأشياء يتم توفيره في العديد من الأعمال باستخدام آليات متعددة منها المعتمد على التشفير ونظرًا لأن أجهزة إنترنت الأشياء مقيدة بالموارد، فإن استخدام تقنيات التشفير الحديثة يؤدي إلى

زيادة كبيرة في التكاليف وتأخير عملية وصول البيانات الحساسة أي لا يمكن منع كل محاولات الهجوم على الشبكة وبالتالي لا بدا من أنظمة توفر أمان فعال في اتصالات الشبكة [7] .

يعرض الجدول1 مقارنة بين آليات الكشف [8].

الجدول 1: المقارنة بين أنظمة كشف الهجمات

–	الإيجابيات	السلبيات
الكشف المعتمد على التواقيع (signature)	طريقة دقيقة في الكشف عن الهجمات المعروفة مسبقاً	غير فعالة في حالة الهجمات الجديدة ويمكن للجهات الخبيثة ببساطة تعديل تسلسل هجماتها
الكشف المعتمد على السلوكيات (anomaly)	يعتمد على مراقبة نشاط الشبكة والنظام حيثُ يتم اكتشاف الحالات الشاذة بعدة طرق، وغالباً باستخدام تقنيات الذكاء الصنعي وتعلم الآلة	معدل خطأ عالي . تحتاج الى بيانات دقيقة والى كمية ضخمة من البيانات .
الكشف المعتمد على التشفير (Crypto) [9]	إن دمج ميزات التشفير ومراقبة البنية التحتية يساعد في توفير خدمة أمان موثوقة واكتشاف العقد الضارة في الشبكة عن طريق ضبط قيم الثقة للعقد الضارة.	تحتاج الى تقنيات التشفير الثقيلة لبناء الثقة القوية في الشبكة بالإضافة لذلك هناك مقايضة بين الأمن واستخدام الطاقة .
الكشف المعتمد على تحليل السجلات (Log Analysis)	يندرج ضمنها الكشف المعتمد على السلوكيات واستخدام النماذج الإحصائية	تحتاج الى الاحتفاظ بسجلات مفصلة لفترة زمنية محددة وبعد الضبط والتحديث المنتظم لهذه النماذج أمرًا مهمًا لضمان بقائها دقيقة وفعالة .

وبالتالي تكمن المشكلة الاساسية للبحث بأن شبكات إنترنت الأشياء عرضة للهجمات الالكترونية المختلفة والتي تشكل تهديدًا أمنياً لبيانات إنترنت الأشياء وأيضاً للبروتوكولات المستخدمة في هذا المجال MQTT و COAP ... كهجوم فيضانات حزم النشر (Publish flood) وهجمات إعادة التوجيه (COAP Replay) وهجمات

حجب الخدمة البطيء(SlowITe) وهجمات آخرى مثل المسح والاستكشاف (Recon) وهجمات الشخص في الوسط (MITM) وهجمات برامج الفدية (Ransomware)وغيرها وبالتالي من هذه المشكلة الأساسية نستطيع تحديد مشكلات جزئية في أنظمة كشف الحالات الشاذة (anomaly) المعتمدة على تقنيات التعلم وهي كالتالي : 1- مشكلة عدم التحديد الدقيق لميزات الكشف المعبرة عن سلوك الهجمات في إنترنت الأشياء 2- مشكلة عدم تحديد نموذج التعلم الأفضل والمناسب لمستويات الكشف (مستوى التدفق الشبكة ومستوى الحزمة) وبالتالي إنَ حل هذه المشاكل تساعد في إنتاج أنظمة خفيفة الوزن ومناسبة لعملية النشر ضمن إنترنت الأشياء وتؤدي الى تقليل الأختناق في الكفاءة الحسابية من أجل التنفيذ في الزمن الحقيقي والسرعة في عملية التنبؤ بالهجوم مع تحقيق نسبة إيجابيات خاطئة منخفضة ودقة عالية للوصول الى مستوى عالي من الأمان . **يساهم هذا البحث** : في تقديم دراسة استقسائية (Survey) شاملة حول الاتجاهات الحديثة التي تسعى الى تحسين أنظمة كشف هجمات إنترنت الأشياء المُعتمدة على تقنيات التعلم من خلال الاهتمام بجانبين وهما تقديم ميزات فعالة وذات عدد محدد بالإضافة لذلك الاهتمام بطريقة تصميم نموذج الكشف حيثُ قدمت هذه الدراسة تقييم لمدى أهمية وفعالية الميزات المحددة من خلال التقييم النظري الذي اعتمد على تحليل الهجمات لتفسير ارتباط الميزات بالهجوم ومن ثم التقييم العملي الذي تمَ بدراسة تأثيرها على العديد من أنظمة الكشف ومساهمتها في تدريب النماذج بزمن قصير جدا وتحقيق دقة عالية وقيمة منخفضة من FP .

قدم هذا البحث دراسة واسعة للعديد من نماذج الكشف التي تعتمد على تقنيات التصميم مختلفة مثل E_Serial , E_Stack , RNN , LSTM , DNN , CNN وخوارزميات التعلم الآلة DT, SVM ,RF .

2- الهدف من البحث

يهدف هذا البحث الى دراسة وتقييم أداء أنظمة كشف الهجمات المعتمدة على تقنيات التعلم (الآلة والعميق والجماعي) من حيث دقته في الكشف ونسبة الإيجابيات الكاذبة وبالتالي نستطيع تلخيص هدف البحث في عدة نقاط مكملة لبعضها على الشكل التالي :

- دراسة وتقييم الميزات المحددة القابلة للتفسير على مستوى التدفق الشبكة و الحزمة (طبقة التطبيق) وتأثيرها على أداء أنظمة الكشف المعتمدة على تقنيات التعلم في زيادة الدقة وتقليل معدل الإيجابيات الخاطئة مقارنة بجميع الميزات.
- دراسة وتقييم أنظمة الكشف المصممة بعناية من حيث الدقة والإيجابيات الخاطئة لتحديد النموذج الأفضل والملائم ليعمل كمحرك كشف .

3- أهمية البحث

تأتي أهمية هذا البحث في إيجاد المعايير الأساسية لتصميم نظام كشف مبني على خوارزميات التعلم وذلك بتحديد الميزات التي تعكس سلوك الهجوم وتحديد التقنية المستخدمة في تصميم نموذج الكشف الذي يُحقق سرعة عالية في عملية الكشف وتقليل معدل الإنذارات الكاذبة لتُصبح شبه معدومة فإن نتائج هذا العمل تعكس أهميته بما تقدمه من دراسة ونتائج لمجتمع البحث المهتم بالقضايا الأمنية لإنترنت الأشياء.

4- أدوات وطرائق البحث

تم العمل على بيئة PyCharm لتنفيذ النماذج الكشف في لغة البرمجة Python ذو الإصدار 3.10 وتم استخدام العديد من المكتبات، مثل Pandas و sklearn و Keras وذلك لتنفيذ النماذج المعتمدة على الشبكات العصبونية مثل CNN و DNN و LSTM . تم استخدام مجموعتين من البيانات لتقييم أنظمة كشف الهجمات المعتمدة على تقنيات التعلم على مستوى التدفق الشبكة ومستوى الحزمة : **المجموعة الأولى** : [10] بيانات القياس الخاصة بإنترنت الأشياء " TON-IOT" التي تحتوي بيانات حركة تدفق الشبكة و تتمتع مجموعات البيانات بخصائص:

- مجموعة متنوعة من الأحداث العادية والهجومية .
- مصادر البيانات غير المتجانسة والناتجة عن أجهزة متعددة من إنترنت الأشياء .
- حيث تحتوي إحصائيات مجموعة بيانات ToN-IoT في أنظمة التشغيل Windows 7,10 و Network على العديد من الأنواع الطبيعية والهجومية. حيث تحتوي على 42 ميزة وتسع فئات من الهجوم ضد العناصر الضعيفة موضحة في الجدول 2 .

الجدول 2:أنواع وعدد التدفقات المضمنة في مجموعة البيانات TON-IOT

فئة التدفق	عدد السجلات
Benign	300000
Scanning	20000
MITM	1043
Ransomware	20000
Cross-Site Scripting	20000
DoS	20000
DDoS	20000
Backdoors	20000
Injection	20000
Password	20000

المجموعة الثانية : البيانات الخاصة بالرعاية الصحية "IOT healthcare " على مستوى الحزمة [11] هذه المجموعة تتعامل مع وحدة العناية المشددة للمرضى لذلك تم استخدام أداة إنترنت الأشياء IoT-Flock لتوليدها. تتميز IoT-Flock بأنها قادرة على توليد أحدث الهجمات التي تستغل MQTT وCOAP. وهذا غير مدعوم حتى الآن بواسطة أي أدوات أخرى. تم تنفيذ سيناريوهات الهجوم لإطلاق عدة أنواع من الهجمات ضد بروتوكولات ومخدمات إنترنت الأشياء وتتضمن هجوم MQTT لنشر الفيضانات وهجوم brute force وهجوم SlowITE وهجوم الحزم المعدة بشكل ضار حيث تحتوي على 50 ميزة وهي موضحة في الشكل 1 .

Attack	PCAP Size (bytes)	Number of Packets	Time (mm:ss)
flooding denial of service	49,875,539	130,223	05:00
MQTT Publish flood	8,212,656	613	05:00
SlowITe	972,272	9202	10:00
malformed data	1,038,590	10,924	06:00
brute force authentication	1,397,132	14,501	30:00

الشكل 1:أنواع الهجوم في مجموعة بيانات IOT-healthcare

سيتم تقييم نماذج متنوعة من أنظمة الكشف الواردة في الدراسات المرجعية والمعتمدة على تقنيات التعلم العميق DNN و CNN و LSTM و RNN و GRU ونماذج معتمدة

على التعلم الجماعي والمكونة من عدة متعلمين ضعفاء مثل المكدس Stack والتسلسلي Serial وعدة نماذج معتمدة على خوارزميات تعلم الآلة سيتم ذكر هذه النماذج لاحقاً.

5- الدراسات المرجعية

ناقشت العديد من الدراسات المرجعية أنظمة الكشف المعتمدة على تقنيات التعلم في إنترنت الأشياء حيث إنه تم إجراء مسح شامل في [12] بواسطة Mohammed Ali Al-Garadi وزملائه عام 2018 لتقنيات التعلم الآلة والعميق المستخدمة لتطوير أساليب أنظمة الكشف في إنترنت الأشياء تم عرض التهديدات الأمنية ومناقشة مختلف أسطح الهجوم المحتملة , قدمت هذه الورقة البحثية مقارنة من حيث الإيجابيات والسلبيات لطرق ML/DL المستخدمة فقط دون عرض البيانات المستخدمة ونوعها وعلاقتها بإنترنت الأشياء. في [13] قدمت Nadia Chaabouni ورفاقها عام 2019 تحليلاً شاملاً لأنظمة الكشف الشبكية NIDSs المعتمدة على تقنيات التعلم المختلفة لإنترنت الأشياء. تم تصنيف مخاطر إنترنت الأشياء وطرق الكشف الحديثة المعتمدة على تقنيات التعلم مقارنة بآليات الحماية التقليدية من حيث البنية ومنهجيات الكشف واستراتيجيات التحقق من الصحة والتهديدات الأمنية في حين إنَ جميع البيانات التي تم تصنيفها قديمة وهجماتها محدودة . تم تقديم مراجعة في [14] من قبل Javed Asharf وزملائه لأنظمة الكشف في شبكات إنترنت الأشياء التي تستخدم أساليب الكشف المعتمدة على ML وDL حيثُ تمت مناقشة بنية إنترنت الأشياء والبروتوكولات ونقاط الضعف والهجمات على مستوى بروتوكول إنترنت الأشياء ومقارنة العديد من الدراسات البحثية المهتمة بأنظمة الكشف IDS لإنترنت الأشياء مع التركيز على تقنيات ML وDL المختلفة المتاحة لنظام IoT- IDS المستخدمة كآلية كشف. الدراسة في [15] تُقدم استبياناً حول اكتشاف الحالات الشاذة في إنترنت الأشياء عن طريق التمييز بين السلوكيات الطبيعية والشاذة أثناء تحليل حركة تدفق الشبكة تستعرض هذه الدراسة الأبحاث السابقة التي اعتمدت على تقنيات التعلم العميق واستخدام مجموعتين من البيانات -CSE-CIC

IDS2018 وBot-IoT في الكشف كان التوجه نحو تصنيف الأبحاث والمقارنة بينها من حيث التقنيات المستخدمة بالاعتماد على تدفق الشبكة فقط واستخدام مجموعة بيانات محدودة وقديمة بالنسبة لإنترنت الأشياء.

سلط المؤلفون الضوء في [16] بحثاً استطلاعياً يدور حول الأبحاث التي أهتمت بأنظمة كشف الهجمات لتحقيق الأمان لشبكة إنترنت الأشياء . في هذا الاستطلاع، سلط المؤلفون الضوء على تقديم تحليلًا تفصيليًا لأنظمة IDS التي تم تطويرها باستخدام خوارزميات تعلم الآلة في بيئة إنترنت الأشياء والتي قدمها العديد من الباحثين بالإضافة إلى ذلك تم تقسيم معرفات أنظمة الكشف في إنترنت الأشياء إلى أربع فئات، وهي IDS المعتمدة على اكتشاف الشذوذ والتوقيع والمواصفات والطريقة الهجينة وأجروا تقييم للعديد من أنظمة الكشف . ومع ذلك، فإن أحدث الاتجاهات وأنماط الهجوم المختلفة تتطلب مزيدًا من التحليل، ومعرفة ميزات وخصائص هذه الهجمات لتحسين دقة الكشف ورفع مستوى الأمان في الشبكة غير ذلك إن الميزات التي تم اختيارها محدودة وتم تحديدها باستخدام تقنية دون المقارنة بين الأبحاث من حيث الميزات . قدمت الدراسة الاستطلاعية في [17] مقارنة بين أساليب التعلم العميق الواردة في الأوراق البحثية للكشف عن التهديدات السيبرانية في إنترنت الأشياء بالاعتماد على مقياس الدقة وتقديم القضايا المتعلقة بكل بحث لكنهم لم يهتموا بقضية تحليل الهجمات وتحديد الميزات الهامة التي تعبر عن الهجمات المختلفة التي تؤدي الى انتاج نماذج من التعلم خفيفة الوزن تقلل من استهلاك الطاقة ومناسبة لطبيعة النشر ضمن إنترنت الأشياء .

يوضح الجدول 3 مقارنة بين الدراسات المرجعية السابقة وهذا البحث.

الجدول 3: المقارنة بين الدراسات المرجعية (Surveys) وهذا البحث

الدراسات المسحية (Survey) لأنظمة الكشف عن هجمات إنترنت الأشياء المعتمدة على تقنيات التعلم									
العام	Ref	بارامترات المقارنة بين الدراسات المسحية وهذا البحث							
		IDS-IOT	تقنية نظام الكشف	وجود الأختبار والتقييم	طريقة التصميم	استخدام بيانات حديثة خاصة بإنترنت الأشياء	تحليل الهجمات		أنظمة الكشف من حيث نوعية وعدد الميزات
							الاكتفاء بتعريفها فقط	تحليل مُعمق لسلوكها	
2018	[12]	✓	DL/ML	X	✓	X	X	X	X
2019	[13]	✓	DL/ML	X	✓	X	✓	X	X
2020	[14]	✓	DL/ML	X	✓	X	✓	X	X
2021	[15]	✓	DL	X	✓	X	X	X	X
2022	[16]	✓	ML	✓	✓	X	✓	X	✓
2023	[17]	✓	DL	X	✓	✓	✓	X	X
2024	هذه الدراسة	✓	DL/ML/EL	✓	✓	✓	✓	✓	✓

العديد من الدراسات المسحية (Survey) التي تمت دراستها تعاني من عدة سلبيات فبعضها اهتمَ بإحدى تقنيات التعلم دون الأخرى **(تقنية نظام الكشف)** والكثير من الدراسات لم يقدموا اختبار وتقييم واضح بل تمَ الاكتفاء بالمقارنة النظرية فقط **(وجود الأختبار والتقييم)** ومعظم دراستهم كانت تتجه نحو تصنيف ومقارنة نماذج الكشف من حيث تقنية التعلم المستخدمة في تصميم النموذج والتغيير بها من أجل الحصول على دقة عالية ومعدل منخفض من الإيجابيات الخاطئة **(طريقة التصميم)** غير ذلك بعض الدراسات اعتمدت على بيانات قديمة بالنسبة لإنترنت الأشياء وهجماتها محدودة للغاية **(استخدام بيانات حديثة خاصة بإنترنت الأشياء)** ولم يهتموا بدراسة وتحليل الهجمات للخروج بأنظمة كشف تُحقق دقة عالية ومعدل منخفض من الإيجابيات الكاذبة أي دراستهم لم تواكب طبيعة شبكة إنترنت الأشياء والمتمثلة بمحدودية الموارد وحركة التدفق الكبيرة على فترات زمنية قصيرة حيثُ إن هناك أدبيات تتجه نحو مواكبة هذه المشكلة عن طريق تحليل الهجمات ودراسة ارتباطها بالميزات على مستوى تدفق الشبكة والتطبيق لإنترنت الأشياء ومن ثم تحديد الميزات التي تعبر بشكل واضح عن سلوك الهجمة

للخروج بميزات عالمية دقيقة تتعلم منها نماذج الكشف وتساعد في إنتاج أنظمة خفيفة الوزن وذات تعقيد حسابي صغير ومناسبة لعملية النشر ضمن إنترنت الأشياء هذا الجانب لم يتم الاهتمام به في معظم هذه الدراسات المسحية . لذلك لا بدا من معالجة هذه الثغرات وذلك بدراسة العديد من أنظمة الكشف المهتمة بطريقة التصميم وبعضها اهتم بتحديد الميزات والمقارنة بينها ومن ثم تحليل العديد من الهجمات للتحديد الدقيق لميزات الكشف الفعالة وسد الثغرة المتمثلة باختلاف عدد ونوع الميزات المحددة المعتمدة في أنظمة الكشف كما سنرى لاحقاً وإجراء التقييم للعديد من هذه النماذج المصممة بعناية للمفاضلة بينها .

6- خطوات الدراسة

الخطوة الأولى تشمل القسم 7 الذي يتضمن دراسة العديد من أنظمة كشف الهجمات المعتمدة على تقنيات التعلم في إنترنت الأشياء من حيث طريقة تصميم النموذج وميزات الكشف المستخدمة التي تسعى لتحقيق دقة عالية ومعدل منخفض من الإيجابيات الكاذبة والمقارنة بينها

الخطوة الثانية تشمل القسم 8 و9 تتجه نحو تحليل الهجمات الأكثر خطورة على شبكات إنترنت الأشياء مثل هجمات استغلال نقاط الضعف لبروتوكولات MOTT و COAP وهجمات المسح والاستكشاف (Recon) وهجمات برامج الفدية وهجمات XSS وهجمات DDOS وهجمات الشخص في الوسط MITM وذلك من حيث آلية الهجوم ونوعه وتأثيره على الشبكة وتحديد الميزات المهمة لكل هجمة .

الخطوة الثالثة في القسم 10 تعتمد على ربط نتيجة التحليل الناتجة عن الخطوة الثانية بالميزات المحددة في الدراسات بالخطوة الأولى وذلك لتحديد متطلبات البحث من حيث الميزات المحددة.

الخطوة الرابعة في القسم 11 إجراء التقييم للعديد من النماذج الكشف المعتمدة في الدراسات المرجعية مع بيان اهمية الميزات المحددة مقارنة بجميع الميزات وذلك بتأثيرها على أنظمة الكشف . ثم مناقشة النتائج والاستنتاجات.

7- أنظمة كشف الهجمات المعتمدة على تقنيات التعلم في إنترنت الأشياء

اقترح الباحثون العديد من طرق تصميم نماذج كشف الهجوم باستخدام الأساليب القائمة على التعلم والتي تُعد أحدَ أنواع تقنيات الكشف عن الشذوذ في السلوكيات . اقترحوا في [6] نموذج يعتمد على التكديس (Stacking) لتصميم نموذج من التعلم الجماعي يجعل أجهزة إنترنت الأشياء أكثر ذكاءً في اكتشاف السلوك غير المعتاد لبيانات IOT . في[18] تم تصميم إطارًا هجيناً باستخدام الشبكة العصبونية التلافيفية (CNN) لاستخراج التمثيل الدقيق لميزات البيانات وتصنيفها باستخدام نموذج الذاكرة الطويلة قصيرة المدى (LSTM) تم إجراء الاختبار باستخدام مجموعة بيانات IOT_23. في [19] تمت معالجة مشكلة ارتفاع قيمة المعدل الإيجابي الخاطئ باستخدام مخطط تدريبي لنموذج DNN باستخدام أشجار القرار (DTs) والنموذج الرياضي PCA . في [20] قاموا بإضافة مستوى عالٍ من الأمان إلى شبكات الأحداث VANETs باستخدام المصنف التسلسلي XGBoost الذي حقق نتائج مبهرة في مهام التصنيف الثنائية . في [21] عملوا على تحسين نظام الكشف المبني على خوارزميات التعلم الآلة باستخدام تحليل الارتباط الإحصائي لإزالة بعض المعطيات ذات الارتباط الضعيف من مجموعة الميزات الكلية وحققت دقت كشف تبلغ 99.5 % . في [22] تم تقديم نظاماً أمنياً يعتمد على تقنية التعلم التجميعي XGboost تميز نموذجهم بزيادة الدقة إلى الحد الأقصى عن طريق تقليل تابع الخسارة بانتظام. في [23] تم تصميم نموذج هجين من التعلم العميق يعتمد على الوحدة المتكررة ذات البوابات GRU والشبكة العصبية المتكررة RNN . يهدف البحث في [24] الى تحسين نظام كشف الهجمات المعتمد على التعلم العميق عن طريق الاعتماد على ميزات محددة قابلة للتفسير مع وجود هجمة معينة على مستوى الحزمة . قام الباحثين في [25] تم اختيار الميزات الأكثر فعالية من أكثر من مجموعة من بيانات إنترنت الأشياء على مستوى تدفق الشبكة والتي يمكن أن تساعد في إنتاج أنظمة كشف عالية الدقة وقابلة للاستخدام في تطبيقات مختلفة لإنترنت الأشياء. في [26] قدموا طريقة لاختيار الميزات المترابطة باستخدام مخطط الارتباط لبيرسون لكن ميزاتهم كانت تحتوي على الزمن والذي يؤثر سلباً على تدريب النموذج وتم أخذ ميزة خاصة بتشفير SSL والذي قد لا يكون خياراً دائماً لإنترنت الأشياء[27] . في [28] و

[29] تم الاعتماد على عشرة ميزات هامة على مستوى الحزمة لتحسين أنظمة الكشف وتحقيق مستوى عالي من الأمان لشبكة إنترنت الأشياء . يعرض الجدول **4** تلخيص ومقارنة بين العديد من الدراسات المرجعية التي اعتمد بعضها على تحليل ارتباط الهجمات بميزات الكشف وبعضها اعتمد على طريقة تصميم دون تحديد الميزات .

الجدول 4: المقارنة بين الدراسات المعتمدة على ميزات محددة والميزات الكلية لكشف الهجوم

طُرق زيادة الدقة وتقليل معدل الإيجابيات الخاطئة في أنظمة الكشف المعتمدة على تقنيات التعلم				*مستوى الكشف*		–	نظام الكشف	البيانات المستخدمة	–	–
لم يتم التحديد	*ميزات محددة على مستوى الحزمة*	*ميزات محددة على مستوى التدفق*	*تفسير ارتباط الميزات بالهجوم*	*التدفق*	*الحزمة*	الدقة (%)	التقنية	Data	year	Ref
T	F	F	F	T	F	99.5	MLs	IOT-23	2020	[21]
T	F	F	F	T	F	97.06	DL	IOT−23	2021	[18]
T	F	F	F	T	F	98.2	EL	TON-IOT	2021	[20]
F	T	F	F	F	T	99.7	MLs	IoT Healthcare	2021	[29]
F	T	F	F	F	T	90	MLs	MQTTSet	2021	[30]
T	F	F	F	T	F	86	EL	TON-IOT	2022	[6]
T	F	F	F	T	F	88.6	DL	NSL-KDD	2022	[19]
F	T	F	T	F	T	**99.97**	DL	MQTTset	2022	[24]
F	F	T	T	T	F	**99.62**	ML	TON-IOT IOT-23 IOT-ID	2022	[25]
T	F	F	F	F	T	99.64	EL	IoT Healthcare	2023	[22]
T	F	F	F	T	F	99	DL	TON-IOT	2023	[23]
F	F	T	F	T	F	99.9	EL	TON-IOT	2023	[26]
F	T	F	F	F	T	99.99	EL	IoT Healthcare	2023	[28]
F	T	F	F	F	T	95	ML	MQTTSet	2023	[31]

1.7- المقارنة بين الدراسات المرجعية المعتمدة على ميزات محددة

يتم تلخيص العلاقة بين الدراسات التي اهتمت بتحديد ميزات الكشف وذلك في الجدول **5** حيثُ هدفت الى جعل النموذج يتدرب على الميزات المهمة في مرحلة التدريب وبالتالي تحقيق دقة عالية .

الجدول 5: الميزات المحددة التي حددتها الأدبيات بتحليل ارتباطها مع الهجمات

المراجع \ الميزات المحددة	المرجع [25]	المرجع[1] [26]	المرجع[28] و [29]	المرجع [24]	المرجع [31]	المرجع [30]
Tcp-len	-	-	-	■	■	■
Time_stream	-	-	■	-	■	■
mqtt- Mes- Ty	-	-	■	■	■	■
Keep Alive	-	-	-	■	■	-
Topic-len mqtt-len	-	-	-	-	■	■
QOS level[2]	-	-	■	-	■	■
Con ACK	-	-	-	■	■	■
Tcp-flags	-	-	■	-	■	■
Mqtt_retain	-	-	■	-	■	-
Mqtt_protocol	-	-	-	-	■	-
Mqtt_hdrflags	-	-	■	■	■	■
Sourc_Port	-	■	-	■	-	-
Dest_Port	■	■	-	-	-	-
conn-stat[3]	■	-	-	-	-	-
proto	■	■	-	-	-	-
src- & Des-pkts pkts	■	-	-	-	-	-
src_ip_byte dst_ip_byte	■	-	-	-	-	-
رسائل DNS	-	■	-	-	-	-
Datasets	IoT-و TON_IoT IoT-ID و23	TON_IOT	IoT Healthcare Security	MQTTset	MQTTset	MQTTset
مستوى الكشف	التدفق		الحزمة			

.....

أخيراً نلاحظ إن بعض الدراسات ركزوا على طريقة تصميم النموذج بحيث يحقق دقة عالية من دون أخذ نوع وعدد الميزات في الاعتبار وبعض الدراسات الأخرى اعتمدوا على تحديد ميزات هامة تُمكن من إنتاج أنظمة كشف قابلة للعمل في عدة تطبيقات لإنترنت الأشياء لكن

[1] في [19] بعض الميزات لم يتم وضعها في جدول لأن ميزات SSL وHTTP غير مدعومة لأنه قد لا يبدو SSL/TLS خيارًا دائمًا [27].

[2] QoS level : مستوى جودة الخدمة في رسائل بروتوكول MQTT تضمن وصول الرسائل للوجهة .

[3] Conn-stat : حالة الاتصال بين المرسل والمستقبل .

بعضها اعتمدَ على تقنية اختيار الميزة وبعضها اعتمد على تحليل الهجمات في اختيار الميزة ونلاحظ من الجدول **4** الدقة العالية في المرجع [24] و [25]. ومن ملخص الدراسات التي اعتمدت على ميزات محددة كما هو موضح في الجدول **5** يتضح إن هناك تفاوت بين عدد الميزات ونوعها في كل دراسة لذلك لا بدا من إجراء تحليل لهذه الهجمات على مستوى التدفق الشبكة وعلى مستوى الحزمة لتبيان أهمية كل ميزة وتحديد معايير التقييم.

8- الهجمات على إنترنت الأشياء

في الأقسام التالية يتم دراسة العديد من الهجمات على شبكة إنترنت الأشياء.

1.8- مجموعة الهجمات التي تستغل بروتوكولات إنترنت الاشياء MQTT و COAP

فيما يلي مجموعة من الهجمات التي يعتمد بها المهاجم على استغلال النقاط الضعف الموجودة في بروتوكولات إنترنت الأشياء MQTT و COAP .

1.1.8-هجوم فيضانات MQTT (MQTT Flood)

وفقًا للمرجع رقم [32] فإن نشر رسائل MQTT بمعدل مرتفع يمكن أن يتسبب في هجوم رفض الخدمة. فيما يلي بعض أنواع هجوم الفيضان :1- هجوم فيضانات الحمولة الأخيرة (Last will) : يستغل هذا الهجوم ميزة الوصية الأخيرة المتوفرة في بروتوكول MQTT بحجم حمولة كبير لتأثير على عرض النطاق الترددي حيثُ يقوم المهاجم بالاتصال بالوسيط وقطع الاتصال به بشكل متكرر[33] في هذا الهجوم من المهم جداً مراقبة طول رسالة الوصية (mqtt.willmsg_len) وحالة الاتصال على مستوى طبقة التطبيق(mqtt-conflags) .2- هجوم فيضانات حزم النشر (Publish flood) : لا يستطيع مخدم MQTT التعامل مع حزم النشر التي يتجاوز حجمها 10000 بايت [29] . في هذا النوع من الهجوم من المهم مراقبة التغيرات في طول الحزم على مستوى طبقة التطبيق والنقل مثل (mqtt-len && tcp-len) . رسائل الاتصال ذو جودة خدمة من المستوى 2 قد تسبب فترات تأخير كبيرة [34] لهذا السبب فأن مراقبة تغيرات مستوى جودة الخدمة (mqtt-header-flags) تعتبر هامة في كشف هذا النوع من الهجمات .

2.1.8- هجوم حجب الخدمة البطيء (MQTT-Slowlte)

يستغل هجوم حجب الخدمة البطيء (Slowlte) بروتوكولMQTT المعتمد على بروتوكول

النقل TCP .هذا الهجوم يستخدم الحد الأدنى من النطاق الترددي ويُحاول المهاجم السيطرة على جميع الاتصالات المتاحة للخادم وإبقاء مخدم MQTT مشغولاً لأطول فترة عن طريق تعديل قيمة البارامتر Keep-Alive على قيمة عشوائية T في حزمة الاتصال [27] . وبالتالي نستطيع من خلال مراقبة هذا البارامتر كشف أمرين الأول إنقاذ وسيط MQTT من هجمة DOS والثانية الكشف عن مهاجم ينتحل عنوان معين .

3.1.8- هجمة الحزم المصممة لتعطيل وسيط MQTT

في الثغرة الأمنية [35] يقوم المهاجم بإجراء اتصال مع وسيط MQTT على طبقة النقل ويبدء بإرسال رسائل النشر Publish بدلاً من ارسال طلب اتصال الى وسيط MQTT وتكون هذه الحزم المُرسلة ذات قيمة خاطئة في طول الموضوع وبالتالي الوصول إلى حالة رفض الخدمة (DoS) باستخدام الحد الأدنى من النطاق الترددي. في هجوم malformed يُرسل العميل طلب اشتراك إلى الوسيط ثم يبدء العميل بإرسال حزم نشر الى الوسيط ، ومع ذلك يتلقى الوسيط حزمة النشر من العميل مما يؤدي إلى حدوث إرباك للوسيط [36] . يجب مراقبة نوع الرسالة (mqtt-type) اضافة لذلك لا بدا من مراقبة التغيرات في طول الموضوع وطول حزمة MQTT (mqtt-len && topic-len) ومراقبة حالة تأكيد الاتصال (con-ack) عندما بقوم المهاجم بإرسال رسائل غير صحيحة .

4.1.8- هجوم إعادة توجيه حزم بروتوكول إنترنت الأشياء COAP

كما في بروتوكول MQTT هناك إمكانية لتواجد المهاجم بين العميل والخادم . في هذا الهجوم يقوم المهاجم بتعديل الحمولة عن طريق استبدالها ببيانات غير دقيقة وإرسالها إلى خادم COAP باستخدام تقنية خادعة تحاكي أجهزة الاستشعار [29] تتمثل عملية الكشف بمراقبة السلوك المنحرف في قيمة البيانات والتي تكون مُعدلة وخارجة عن طبيعتها .

2.8- هجمات برامج الفدية (Ransomware)

تستغل هجمات برامج الفدية الثغرات الأمن في بروتوكول SMB عن طريق الاستغلال EternalBlue الذي يستفيد من ثلاثة أخطاء يرتكبها بروتوكول SMB لكتابة وتنفيذ كود الأوامر البرمجية والتمكن من السيطرة على النظام [37] . إنَ سلوك برامج الفدية التشفيرية التي يتم ملاحظتها اثناء تحليل حركة تدفق رسائل بروتوكول SMB يتم تخليصها في[38] . يجب مراقبة رسائل القراءة والكتابة والحذف التي تتم بسرعة عالية/نسبة عالية مقارنة بالحالة الطبيعية/. يقوم المهاجم بإجراء مسح للشبكة للحصول على معلومات عن حالة المنافذ التي تستخدم خدمة SMB والتي رقمها 445 وبالتالي يجب مراقبة أي حالة اتصال شاذة تُستخدم لمعرفة حالة المنافذ الشهيرة وخاصة التي تعمل على بروتوكول TCP .

3.8- هجمات الشخص في الوسط MITM

يتم تنفيذ هذا الهجوم بعدة طرق منها : إعادة توجيه رسالة ICMP [39] . نظرًا لعدم مصادقة رسائل إعادة توجيه ICMP يمكن للمهاجمين انتحال هذه الرسائل لكن شرط تحقيق هذا الهجوم: إن يكون عنوان وجهة الحزمة المستقبليَّة يقع في فضاء عناوين الشبكة الَّتي عُنونت بها الحزمة السابِقة (التهديد داخلي) [40] . ويمكن أن يتم هجوم MITM عن طريق تسمم ذاكرة التخزين المؤقت لبروتوكول ARP باستغلال نقاط الضعف في رسائل بروتوكول ARP لخداع جهاز الضحية وجهاز التوجيه في [41] تم وصف هذا الهجوم. يمكن كشفها باستخدام أنظمة الكشف المعتمدة على السلوكيات عن طريق مراقبة التغيرات في اطوال الحزم المرسلة والمستقبلة (src_ip_bytes && dst_ip_bytes) وعدد الحزم (src_pkts && dst_pkts) وحجم الحمولة (dst_bytes && src_bytes) على مستوى تدفق الشبكة ومراقبة الحركة الشاذة لرسائل DNS وللكشف الدقيق عن MITM يمكن دراسة بعض الميزات مثل حساب نسبة التباين بين الحزم المرسلة والمستقبلة (Variance-packets).

4.8- هجمات المسح (Scanning Attack)

إن هجوم المسح هو إحدى الخطوات الأولية للهجوم السيبراني قبل إطلاق الهجوم الفعلي. تتكون تقنيات مسح المنافذ (port scan) من إرسال رسالة إلى الضحية والاستماع للحصول على إجابة. عادةً مايتم فحص المنافذ على منافذ TCP أي المنافذ التي تُرجع ردود فعل جيدة للمهاجم ولها عدة أنواع منها أعلام التخفي وهي SYN و FINوNULL [42] . وهجمات مسح TCP مثل TCP Connect وSYN وFIN وACK وXMAS [4]. ويحدث المسح أيضًا على منافذ UDP ولكنها لا توفر المعلومات بسهولة للمهاجمين [43]. يعرض الجدول 6 تقنيات المسح.

الجدول :6 معلومات عن هجمات مسح المنافذ والاستجابات والمنع على مستوى جدار الحماية

هجوم المسح	البروتوكول المستخدم	استجابة المنفذ المفتوح	استجابة المنفذ المغلق	الكشف على مستوى جدار الحماية
TCP Connect()	TCP	ACK	RST	Yes
SYN Scan	TCP	ACK	RST	Yes
SYN|ACK Scan	TCP	RST	RST	Yes
FIN Scan	TCP	No	RST	No
ACK Scan	TCP	No	RST	No
Null Scan	TCP	No	RST	No
XMAS Scan	TCP	No	RST	No
UDP Scan	UDP	No	Port unreachable	No
FTP Scan	FTP	No	No	No

بعد اكتشاف منافذ TCP/UDP باستخدام إحدى طرق المسح يقوم المهاجم باستجواب هذه المنافذ لتحديد المزيد حول ما هو قيد التشغيل مثل اكتشاف أنظمة . غالباً تكون الحزم ذو حمولة فارغة أو حمولة صغيرة جداً فمن المهم مراقبة تغيرات حجم الحمولة المرسلة والمستقبلة (src_bytes && dst_bytes) على مستوى تدفق الشبكة غير ذلك يجب في كل تدفق مراقبة حزم الاتصال (fin && rst & sack && syn) . يقوم بعض المهاجمين الأذكياء بإرسال حزم TCP مُعدة خصيصاً لاستكشاف نظام التشغيل المستخدم وأصداره عن طريق إعداد حجم النافذة TCP (Window-size) ومن المهم

مراقبة منافذ الهدف (Dest-Port) وحالات الاتصال الشاذة (conn-stat). وللكشف الدقيق عن الهجوم يمكن ادخال مفهوم الفجوة الزمنية لمعرفة التكرار في الاتصالات المشبوهة (num-conn).

5.8 – هجمات حقن النصوص البرمجية الضارة في المواقع XSS

يتمكن المهاجم في هذا الهجوم من تنفيذ برنامج نصي ضار في متصفح مستخدم آخر [44] . مثال بسيط عن بنية إنترنت الأشياء حيث يتم إرسال درجة حرارة غرفة معينة بانتظام إلى التطبيق ويعرضها التطبيق على واجهة مستخدم الويب . بافتراض أن المهاجم لديه حق الوصول إلى شبكة MQTT أو COAP ويمكنه النشر لنفس الموضوع عندها يستطيع نشر حمولة XSS الضارة[45] . يتم الكشف بتحليل الحمولة (Payload) لكن في نظم كشف السلوكيات الشاذة يمكن مراقبة حجم الحمولة التي تكون ذو قيمة معينة في

هذا الهجوم (src_bytes & dst_bytes) ومراقبة حجم الحزم على مستوى IP

6.8– هجمات DDOS كفيضانات من TCP/UDP/ICMP

يقوم هجوم DDOS باستهداف الضحية (victim) لاغراقها بالرسائل مثل هجوم فيضانات UDP أو فيضانات ICMP وهجوم الفيضانات TCP-SYN الخاصة بالمصافحة الثلاثية لبروتوكول TCP وبالتالي يكون لدى الضحية الكثير من الاتصالات نصف المفتوحة ، مما يؤدي إلى استهلاك موارد نظام الضحية [46]. يتم الكشف بمراقبة المعدل الطبيعي لوجود رسائل ICMP ورسائل UDP وTCP ومن جهة أخرى مراقبة اطوال الحزم على طبقة النقل وزمن تدفق حزم TCP (tcp.time_delta) ومراقبة معدل الإرسال والاستقبال.

9- ملخص الدراسة التحليلية لهجمات إنترنت الأشياء وعلاقتها بميزات الكشف

يلخص الجدول 7 الهجمات الأكثر خطورة في شبكات إنترنت الأشياء مع توضيح نوع الهجوم ومستوى وجوده في الشبكة وبعض الميزات التي تكشف عن سلوك كل هجمة .

الجدول 7: ملخص هجمات إنترنت الأشياء الأكثر خطورة

الهجوم	أنواع الهجوم	المستوى	الميزات
SlowITe Attack	هجوم DOS بطيء	التطبيق	KeepAlive
MQTT Flood Attack	فيضانات الحمولة الأخيرة فيضانات ACK-PSH فيضانات حزم النشر .	التطبيق– النقل	Tcp-len/ Mqtt-len Mqtt-willmsg-len / Tcp-flags Mqtt-Qos / Mqtt-hdrflags
MQTT packet crafting attack	هجمة DOS هجوم الحركة المشوهة	التطبيق	Topic-len / Mqtt-len Mqtt-msgtype / Mqtt-flagack
COAP Replay Attack	هجوم انتحال يتلاعب بقيمة الحمولة	التطبيق– الشبكة	سلوك MITM
Recon Attack	OS scan Port scan Vulnerability scan	الشبكة – النقل	Window-size / Dest-port tcp.connection / prot-type conn-state / num-conn
MITM Attack	ARP- Spoofing ICMP Redirect	الشبكة – المستوى الفيزيائي	dst_ip_bytes/src_ip_bytes src_pkts / dst_pkts src_bytes / dst_bytes Variance-packets
XSS Attack	هجوم على مواقع الويب	التطبيق	src_bytes / dst_bytes
DDOS	فيضانات من رسائل TCP/UDP/ICMP	النقل	tcp.time_delta / tcp-len rate-(tcp/udp/icmp)
Ransomware Attack	crypto- ransomware	التطبيق– الشبكة – النقل	Dest-port/ Conn-state Tcp-flags/ Rate-sent receve -read -write

الجزء العملي :

10- متطلبات البحث لاختبار نماذج الكشف .

1.10- الميزات المحددة في الاختبار والتقييم على مستوى التدفق والحزمة .

من الجدول 7 نلاحظ إنَ هناك ميزات مشتركة بين كل هجمة وبعض الهجمات اِنفردت بميزات خاصة بها أما الميزات المحددة في الجدول 5 على مستوى التدفق فقد اعتمدَ الباحثون في [25] على تقديم ميزات عالمية بدراسة عدة تطبيقات لإنترنت الأشياء

والمتمثلة في مجموعات البيانات TON-IOT و IOT-23 و IOT-ID للخروج بميزات تستطيع عن طريقها أنظمة الكشف العمل بمختلف التطبيقات ومن خلال المقارنة بين الجدولين 5 و 7 سنعتمد على الميزات الموضحة في الجدول 8 وذلك على مستوى تدفق الشبكة . أما بالنسبة للميزات المحددة على مستوى الحزمة قام كل من [24] [30], [31] , بتحديد الميزات الهامة في شبكة إنترنت الأشياء المطبقة في المنازل الذكية ومن الجدول 5 نلاحظ إن الاختلاف بينهم قليل جداً. وفي[28] و [29] اعتمدوا على ميزات محددة للخروج بأنظمة كشف دقة عالية . وبالمقارنة مع الجدول 7 سنعتمد على الميزات المحددة على مستوى الحزمة والموضحة في الجدول 9 .

الجدول 8: الميزات المحددة على مستوى التدفق

Dst_port	int
proto	object
Scr_bytes	int
Dst_bytes	int
Conn_state	object
Src_pkts	int
Dst_pkts	int
Src_ip_bytes	int
Dst_ip_bytes	int
Dns_qtype	int
Dns_AA	object
Dns_RD	object

الجدول 9: الميزات المحددة على مستوى الحزمة

Tcp_time_delta	float
Tcp_flags_ack	int
Tcp_flags_push	int
Tcp_flags_reset	int
Tcp_flags	int
Tcp_len	int
Mqtt.retain	int
Mqtt.hdrflags	float
Mqtt.conack.flags	int
Mqtt.kalive	int
Mqtt.len	int
Mqtt.msgtype	int
Mqtt.qos	int
Mqtt.topic_len	int

2.10- المقاييس المستخدمة لتقييم نماذج الكشف

بافتراض إن الهجمة تأخذ القيمة "1" والبيانات الطبيعية تأخذ القيمة "0" في الخرج فإن:

الجدول 10: بارامترات التقييم بافتراض التعبير التالي :الهجمة="1" والبيانات الطبيعية="0"

السجلات الطبيعية		السجلات الهجوم	
سجلات طبيعية صحيحة	سجلات هجوم خاطئة	سجلات طبيعية خاطئة	سجلات هجوم صحيحة
True Negative TN	False Negative FN	False Positive FP	True Positive TP

اعتماداً على البارامترات السابقة يتم استخلاص أربع مقاييس مستخدمة في عملية التقييم:

1- الدقة Accuracy (ACC) وتعطى بالعلاقة التالية :

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

2- Precision (PRE) وتعطى بالصيغة التالية :

$$PRE = \frac{Tp}{Tp + Fp} \quad (2)$$

3- المعدل الإيجابي الحقيقي (TPR) او Recall وتعطى بالصيغة التالية :

$$REC = \frac{Tp}{Tp + FN} \quad (3)$$

4- عامل F1-Score :عامل مهم وخصوصا في البيانات الضخمة ويعطى بالصيغة التالية

$$F1 = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \quad (4)$$

3.10- المعالجة المسبقة للبيانات المستخدمة

اعتماداً على الميزات المحددة في الجدولين 8 و 9 سنعتمد على مجموعتين من البيانات :

المجموعة الأولى : **"TON-IOT"** [10] **.** والمجموعة الثانية : **"IoT Healthcare Security"** [11]. حيث إن هناك العديد من التحديات في مجموعات البيانات هذه مثل القيم المفقودة، والفئات التصنيف غير المتوازنة والميزات الوصفية والميزات غير الضرورية التي تؤثر على أداء النماذج سلباً . لقد حددنا العينات ذات القيم المفقودة، مثل NaN، وتم التخلص منها باستبدالها بمتوسط قيمة الميزة. وتم تحويل الميزات الوصفية إلى رقمية حيث تحتوي مجموعة بيانات ToN-IoT على العديد من الميزات الوصفية مثل حالة الاتصال ونوع بروتوكول طبقة النقل والخدمة المتوفرة .... ولتحويلها الى قيم رقمية تم استخدام الترميز one-hot encoding. وتم حل مشكلة عدم توازن فئات التصنيف والتي تؤثر على أداء النماذج باستخدام تقنية SMOTE لموازنة فئة الهجوم مع الفئة الطبيعية . أما بالنسبة للميزات التي تؤثر سلباً على أداء النموذج مثل الزمن فقد تم حذفه وبعض الميزات الأخرى مثل عناوين IP فقد تم حذفها لان عناوين IP متغيرة أيضاً ليكون النموذج أكثر عمومية وشامل وأيضاً بسبب الأرقام الكبيرة في الزمن وعدد عناوين IP الكبير التي تؤثر سلباً على النموذج . يمكن أن يؤدي وجود قيم تقع خارج النطاق إلى نتائج غير صحيحة لذلك تم استخدام تقنية التقييس المعتمدة على الحد الأدنى والحد الأعلى Min-Max لتتراوح قيم الميزات بين [0:1] وفق العلاقة التالية:

$$Z = \frac{x - x_{min}}{x_{max} - x_{min}}$$

حيث x هي قيمة الميزة، وZ هي القيمة بعد التقييس، وxmax وxmin هما القيمتان الكبرى والصغرى للميزة . تم الاعتماد على تقسيم البيانات لتكون % 80 للتدريب النماذج و % 20 لاختبار النماذج وتقييمها .

4.10 – الخطوات والمعايير المُستخدمة في تقييم نماذج الكشف الهجمات

سيتم التقييم وفق المعايير التالية :

- تقييم الأنظمة على مستوى تدفق الشبكة وعلى مستوى الحزمة .
- سيتم تقييم نماذج متنوعة من أنظمة الكشف المصممة بعناية والتي تمت دراستها ومقارنتها في الجدول 4 من القسم 7:

نماذج المعتمدة على التعلم الجماعي

[6] **Naila Naz** : نموذج يعتمد على التكديس (Stacking) لتصميم نموذج من التعلم الجماعي استخدمَ هذا النموذج مجموعة من خوارزميات التعلم الآلة الكلاسيكية كأعضاء في النموذج الجماعي المقترح وهي LR و SVM و NB و RF و LDA .

[22] **Bharathi (XGBoost):** يعتمد على تقنية التعلم التجميعي التسلسلي المتدرج Ensemble crossover (EC) XG boost .

[26] **Ayoob Almotairi :** نظام يعتمد على التكديس (stacking) ومكون من RF و NB و SVM و KNN .

نماذج المعتمدة على التعلم العميق

[18] **Amiya Kumar** : تم تصميم النموذج باستخدام الشبكة العصبونية التلافيفية (CNN) المؤلفة من أربع طبقات التفافية لاستخراج التمثيل الدقيق لميزات البيانات وتصنيفها باستخدام نموذج الذاكرة الطويلة قصيرة المدى (LSTM) المكونة من طبقتين امامية وخلفية.

[19] **Shoayee Dlaim Alotaibi** : هذا النموذج مكون من مخطط تدريبي يتم استخدام أشجار القرار (DTs) لتعمل كواجهة لفلترة العينات الخاصة بجميع الميزات المتاحة ومن تمريرها الى نموذج رياضي PCA يعمل على تقليل أبعاد المعطيات الى 11 ميزة وأخيراً تقديمها الى محرك الكشف خفيف الوزن DNN . جيث يتكون DNN من طبقتين 140 خلية عصبونية في الطبقة الأولى و70 خلية عصبونية في الطبقة الثانية .

[23] **Noor Wali Khan** : صمموا نموذج هجين من التعلم العميق مؤلف من ثلاثة طبقات وهي الوحدة المتكررة ذات البوابات GRU وطبقة Dense وطبقة من الشبكة العصبونية المتكررة RNN والتي يمكنها تصنيف الهجمات عبر الطبقات متعددة من الشبكة

[24] **Sunoh Choi** :تألفَ نموذجهم من ثلاثة طبقات حيث إن حجم طبقة التضمين Embedding وعدد عقد LSTM و طبقة Dense كل منها تم ضبطه على 128 ومعدل التطبيع 0.2 .

نماذج المعتمدة على التعلم الآلة

هذه النماذج مثل الغابات العشوائية واشجار القرار وخوارزمية SVM

Abdallah R.GAD (SVM) , Maheshi (DT) [30] **, Nicolas (RF)** [21]

[20]

لتحقيق هدف البحث سنعتمد على الخطوات التالية في التقييم:

* الخطوة الاولى: تطبيق جميع الميزات على نماذج الكشف وملاحظة مقاييس التقييم
* الخطوة الثانية:تطبيق الميزات المحددة على نماذج الكشف وملاحظة مقاييس التقييم
* الخطوة الثالثة: المقارنة بين الخطوتين الاولى والثانية من حيث الدقة والإيجابيات الخاطئة وزمن التدريب.

إعادة الخطوات الثلاثة للاختبار على مستوى الحزمة .

11- التقييم من حيث النتائج والمقارنة

في هذه القسم سيتم تقييم العديد من نماذج الكشف المصممة بعناية لاكتشاف الاختراقات الأمنية في شبكة إنترنت الأشياء لكي يتم الإجابة عن سؤالين في نفس الوقت :

1- ما هو تأثير الميزات السلوكية المحددة ذو المستويين (مستوى التدفق والحزمة) على أنظمة الكشف وهل تُحقق قيمة منخفضة من الإيجابيات الخاطئة مقارنة بجميع الميزات .

2- ما هو النموذج الكشف الأفضل (طريقة التصميم) الذي يعطي إيجابيات خاطئة منخفضة على المستويين من الشبكة.

1.11- التقييم على مستوى تدفق الشبكة

1.1.11-نتائج تقييم نماذج الكشف بالاعتماد على الميزات المحددة على مستوى التدفق

في هذا القسم تم تقييم نماذج الكشف على الميزات المحددة القابلة للتفسير على مستوى تدفق الشبكة والتي بلغ عددها 12 ميزة (موضحة في الجدول 8) من اصل 42 ميزة مختلفة باستخدام TON-IOT يتم عرض النتائج في الجدول 11.

الجدول 11: نتائج تقييم نماذج الكشف على الميزات المحددة على مستوى تدفق الشبكة

False positive	F1-Score	Recall	PRE	ACC	اسم النموذج
0.00025	0.99963	0.9996	0.9997	0.9996	***Naila Naz*** [6]
0.00018	0.9991	0.9991	0.9993	0.9997	***Bharathi (XGBoost)*** [22]
0.0013	0.9982	0.9986	0.998	0.9986	***Ayoob Almotairi*** [26]
0.02	0.98157	0.9841	0.9790	0.982	***Shoayee Dlaim Alotaibi*** [19]
0.0031	0.99643	0.996	0.9969	0.996	***Amiya Kumar*** [18]
0.024	0.9889	0.983	0.995	0.977	***Noor Wali Khan*** [23]
0.036	0.987	0.99	0.985	0.983	***Sunoh Choi*** [24]
0.0168	0.89702	0.9316	0.984	0.95	***Abdallah R.GAD (SVM)*** [20]
0.00024	0.9989	0.99912	0.9987	0.9989	***Maheshi (DT)*** [30]
0.00062	0.99930	0.9992	0.9994	0.9993	***Nicolas (RF)*** [21]

2.1.11- نتائج تقييم نماذج الكشف بالاعتماد على جميع الميزات على مستوى التدفق

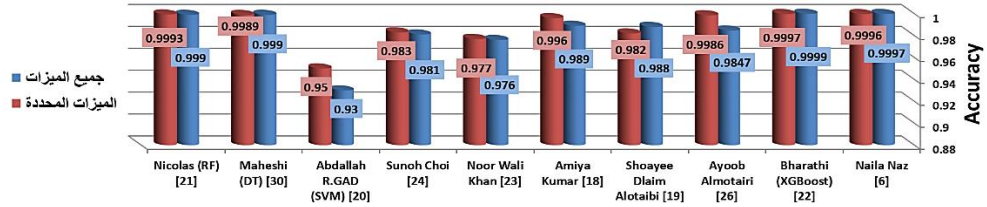
هنا تم إجراء تقييم النماذج على جميع الميزات المستخرجة على مستوى التدفق وذلك لملاحظة اهمية الميزات المحددة في تعليم النموذج يتم عرض النتائج في الجدول 12 .

الجدول 12: نتائج تقييم نماذج الكشف على جميع الميزات على مستوى تدفق الشبكة

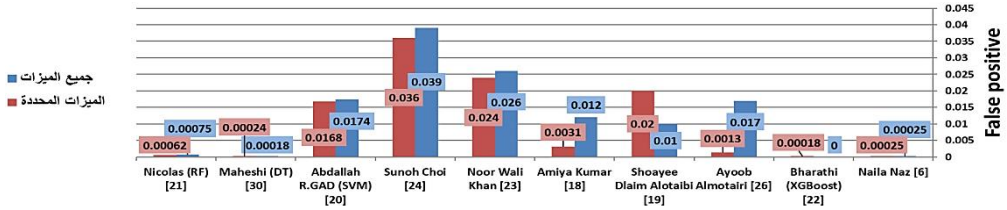
FP	F1-Score	Recall	PRE	ACC	اسم النموذج
0.00025	0.9997	0.9998	0.99976	0.9997	***Naila Naz*** [6]
0	0.9993	0.9996	0.9991	0.9999	***Bharathi (XGBoost)*** [22]
0.017	0.984	0.987	0.982	0.9847	***Ayoob Almotairi*** [26]
0.01	0.987	0.990	0.985	0.988	***Shoayee Dlaim Alotaibi*** [19]
0.012	0.9897	0.9918	0.9876	0.989	***Amiya Kumar*** [18]
0.026	0.988	0.993	0.985	0.976	***Noor Wali Khan*** [23]
0.039	0.992	0.996	0.988	0.981	***Sunoh Choi*** [24]
0.0174	0.9396	0.8973	0.9861	0.93	***Abdallah R.GAD (SVM)*** [20]
0.00018	0.9994	0.9992	0.9997	0.999	***Maheshi (DT)*** [30]
0.00075	0.9995	0.9996	0.9992	0.999	***Nicolas (RF)*** [21]

3.1.11- المقارنة بين تأثير الميزات المحددة والميزات الكلية على أنظمة الكشف على مستوى التدفق الشبكة

يوضح الشكل 2 و3 المقارنة من حيث الدقة والإيجابيات الخاطئة على التوالي :

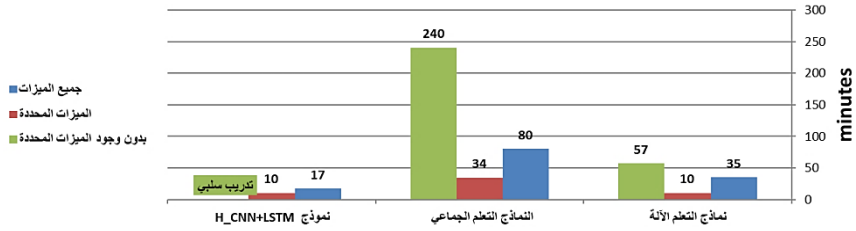


الشكل 2: المقارنة بين الميزات المحددة والميزات الكلية على مستوى التدفق من حيث دقة الكشف (Accuracy)



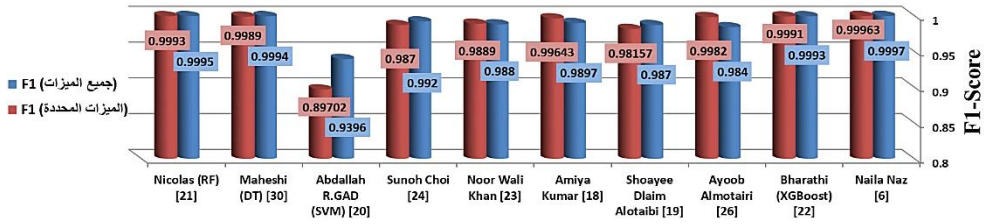
الشكل 3: المقارنة بين الميزات المحددة والميزات الكلية على مستوى التدفق الشبكة من حيث الإيجابيات الخاطئة

تمت إضافة مقياس آخر وهو زمن تدريب النماذج لمعرفة وتحديد مدى قوة الميزات المحددة في فعالية وسرعة تعليمها لنموذج الكشف وهذا المقياس موضح في الشكل 4 .



الشكل 4: متوسط زمن تدريب نماذج الكشف على مستوى تدفق الشبكة

يوضح الشكل 5 المقارنة باستخدام المقياس F1-Score الذي يعطي دقة و كفاءة النموذج و يعتبر أكثر شمولاً من Precision و Recall .



الشكل 5: المقارنة بين الميزات المحددة والميزات الكلية على مستوى التدفق من حيث F1-

مُناقشة النتائج : اعتماداً على الجداول والمخططات البيانية السابقة نلاحظ إن 90% من النماذج حَققت باستخدام الميزات المحددة دقة أكبر أو تساوي دقتها بالنسبة لجميع الميزات وفي **الشكل 3** نلاحظ أن التقييم والاختبار على الميزات المحددة قللَ نسبة الإيجابيات الخاطئة في 8 نماذج (Naila Naz [6] و Bharathi (XGBoost) [22] و Ayoob Almotairi [26] و Amiya Kumar [18] و Noor Wali Khan [23] و Sunoh Choi [24] و Abdallah R.GAD (SVM) [20]) مقارنة بوجود جميع الميزات، وكانت نسبة الإيجابيات الخاطئة قريبة من بعضها ومنخفضة جداً في النماذج (Naila Naz [6] و Bharathi (XGBoost) [22] و Maheshi (DT) [30] و Nicolas (RF) [21]) ونموذج واحد (Shoayee Dlaim Alotaibi [19]) لم يحقق قيمة منخفضة ومن الشكل 4 نلاحظ إنهُ تم الوصول الى نسبة منخفضة من الإيجابيات الخاطئة في الكثير من النماذج وذلك بتدريبها على الميزات المحددة التي حققت متوسط زمن تدريب منخفض جداً مقارنة مع جميع الميزات أو بدون وجود الميزات وبالنظر الى الشكل 5 نلاحظ إن اختبار النماذج على الميزات المحددة حققت نسبة عالية في المقياس F1-Score الذي يعكس دقة وكفاءة النماذج.

2.11- التقييم على مستوى الحزمة

1.2.11-نتائج تقييم نماذج الكشف بالاعتماد على الميزات المحددة على مستوى الحزمة

في هذا القسم تم تقييم نماذج الكشف على الميزات المحددة القابلة للتفسير على مستوى الحزمة والتي بلغ عددها 14 ميزة (موضحة في الجدول 9) من اصل 50 ميزة مختلفة باستخدام IoT Healthcare Security . يبين الجدول 13 نتائج التقييم :

الجدول 13: نتائج تقييم نماذج الكشف على الميزات المحددة على مستوى الحزمة

FP	F1-Score	Recall	PRE	ACC	اسم النموذج
0	1	1	1	1	***Naila Naz*** [6]
0.00004	0.99993	0.99991	0.99995	0.9999	***Bharathi (XGBoost)*** [22]
0	1	1	1	1	***Ayoob Almotairi*** [26]
0	1	1	1	1	***Shoayee Dlaim Alotaibi*** [19]
0.0003	0.9958	0.9929	0.9988	0.998	***Amiya Kumar*** [18]
0.0192	0.9887	0.997	0.980	0.9888	***Noor Wali Khan*** [23]
0.013	0.9874	0.985	0.99	0.989	***Sunoh Choi*** [24]
0.0012	0.9971	0.9976	0.9965	0.998	***Abdallah R.GAD (SVM)*** [20]
0	1	1	1	1	***Maheshi (DT)*** [30]
0	1	1	1	1	***Nicolas (RF)*** [21]

2.2.11-نتائج تقييم نماذج الكشف بالاعتماد على جميع الميزات على مستوى الحزمة

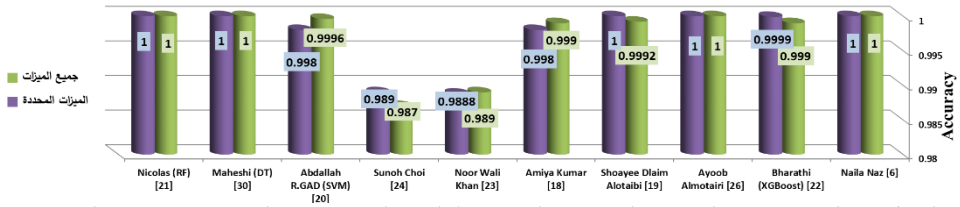
هنا تم إجراء تقييم النماذج على جميع الميزات على مستوى الحزمة وذلك لملاحظة اهمية الميزات المحددة يتم عرض نتائج التقييم في الجدول 14 .

الجدول 14:نتائج تقييم نماذج الكشف على جميع الميزات على مستوى الحزمة

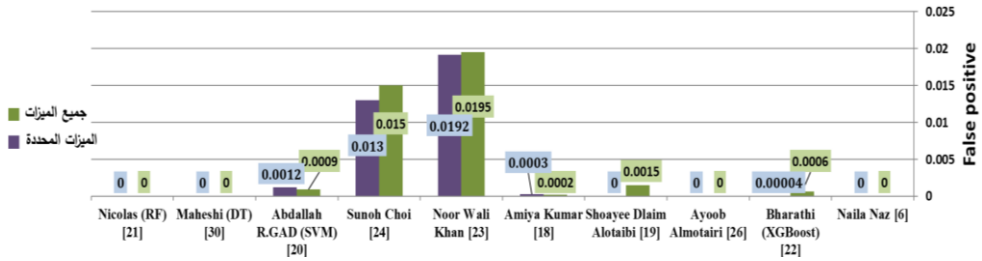
FP	F1-Score	Recall	PRE	ACC	اسم النموذج
0	1	1	1	1	***Naila Naz*** [6]
0.0006	0.997	0.997	0.998	0.999	***Bharathi (XGBoost)*** [22]
0	1	1	1	1	***Ayoob Almotairi*** [26]
0.0015	0.99921	1	0.9984	0.9992	***Shoayee Dlaim Alotaibi*** [19]
0.0002	0.99987	1	0.99975	0.999	***Amiya Kumar*** [18]
0.0195	0.9828	0.996	0.970	0.989	***Noor Wali Khan*** [23]
0.015	0.9894	0.988	0.991	0.987	***Sunoh Choi*** [24]
0.0009	0.9996	0.999510	0.9997	0.9996	***Abdallah R.GAD (SVM)*** [20]
0	1	1	1	1	***Maheshi (DT)*** [30]
0	1	1	1	1	***Nicolas (RF)*** [21]

3.2.11-المقارنة بين تأثير الميزات المحددة والميزات الكلية على أنظمة الكشف (مستوى الحزمة)

يوضح الشكل 6 و 7 المقارنة من حيث الدقة والإيجابيات الخاطئة بين الميزات المحددة والميزات الكلية على التوالي .

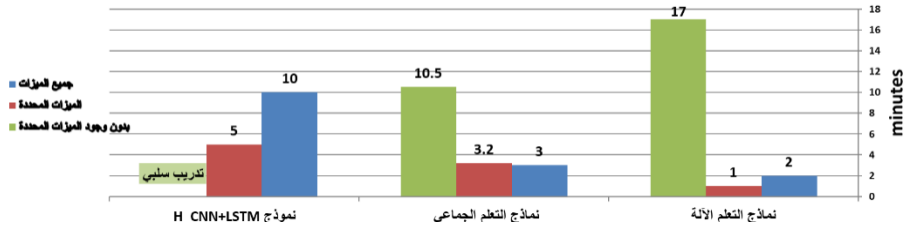


الشكل 6:المقارنة بين الميزات المحددة والميزات الكلية على مستوى الحزمة من حيث الدقة



الشكل 7: المقارنة بين الميزات المحددة والميزات الكلية على مستوى الحزمة من حيث الإيجابيات الخاطئة

ولتحديد مدى فعالية الميزات المحددة في تعليم النماذج وسرعة الكشف عن الهجمات مقارنة بالميزات الكلية تم أخذ مقياس وهو زمن تدريب النماذج الموضح في الشكل 8 .



الشكل 8: متوسط زمن تدريب نماذج الكشف على مستوى الحزمة

مُناقشة النتائج : من المخططات البيانية السابقة والموضحة في الأشكال 6 و 7 و 8 ، يمكن ملاحظة وصول النماذج المُدربة على الميزات المحددة الى دقة عالية وتحقيق نسبة إيجابيات خاطئة منخفضة مقارنة بالنماذج المدربة على جميع الميزات على

مستوى الحزمة وكانت النتائج على الشكل التالي: وصلت دقة 7 نماذج الى دقة أكبر أو تساوي دقة النماذج المدربة على جميع الميزات والتي هي(Naila Naz [6] و Bharathi (XGBoost) [22] و Ayoob Almotairi [26] و Shoayee Dlaim Alotaibi [19] و Sunoh Choi Choi [24] و [30] **)** Maheshi (DT) و Nicolas [21] (RF)) . في حين حققت بعض النماذج (Naila Naz [6] و Ayoob Almotairi [26] و Shoayee Dlaim Alotaibi [19] و Maheshi (DT) [30] Nicolas (RF) [21]) دقة مثالية وخالية تماماً من الإيجابيات الخاطئة وهي موضحة في الشكل 7 .

12- دراسة نتائج البحث

تمت مناقشة النتائج في الاقسام 3.1.11 و 3.2.11 بالاعتماد على التقييم والمقارنة ونتيجة هذه المناقشات نستخلص ما يلي :

1- الميزات المحددة على مستوى التدفق الشبكة والحزمة في طبقة التطبيق لشبكة إنترنت

الأشياء ساهمت في تحسين أنظمة الكشف من حيث تقليل التعقيد في العملية الحسابية للنماذج من أجل الكشف في الزمن الحقيقي وساعدت النموذج على التركيز على الميزات المهمة في مرحلة التدريب وبالتالي حققت السرعة في عملية التنبؤ بالهجوم مع تحقيق نسبة إيجابيات خاطئة منخفضة ودقة عالية .

2- نماذج الكشف المعتمدة على التعلم التجميعي التسلسلي المتدرج XGBoost ونموذج CNN&LSTM حققت قيمة منخفضة من حيث الإيجابيات الخاطئة مقارنة بالنماذج الأخرى على مستوى التدفق والحزمة .

3- إن عملية الكشف على مستوى الحزمة ذو دقة اعلى من الكشف على مستوى التدفق.

4- الميزات المحددة تُمكن من تصميم نماذج غير مقيدة ببيئة معينة وإنما تصميم نظام موسع يعمل على عدة تطبيقات من إنترنت الأشياء IOT .

13- التوصيات والمقترحات

على الرغم من النتائج الإيجابية، لا يزال هذا البحث يعاني من بعض القيود كالعدد القليل من العينات وفئات الهجوم المحدودة ضمن إنترنت الأشياء . ستتضمن الدراسات المستقبلية إعداد أنظمة الكشف والتحقق من قدرتها على اكتشاف الهجمات في الوقت الحقيقي ودراسة وتجربة مجموعات البيانات الفعلية الأخرى على نموذج التعلم المتميز وخاصة مع عصر البيانات الضخمة (Big data) ليتم دراسة وتفسير العديد من الميزات من خلال تحديد ارتباطها بالهجوم مثل Mqtt-willmsg-len و Window-size و num-conn و Variance-packets و rate-(tcp/udp/icmp) و Rate-sent و Rate-receve و Rate-read و Rate-write والتي سيتم الأخذ بها في الدراسة المستقبلية لتحسين أنظمة الكشف عن السلوك الشاذ للهجمات بشكل واضح وخاصة هجمات المسح والاستكشاف التي تحاول سرقة المعلومات الخاصة بالشبكة والجهاز لـتُمكن المهاجمين من تنفيذ هجماتهم لاحقاً بشكلٍ أكثر احكاماً وخطورة وأن كشف هذا النوع من الهجمات يساعد في التصدي والمنع المبكر للعديد من الهجمات اللاحقة قبل تأثيرها السلبي على إنترنت الأشياء.

المراجع

[1] Q. I. Sarhan, "Internet of things: a survey of challenges and issues," *International Journal of Internet of Things and Cyber-Assurance*, vol. 1, pp. 40-75, January 2018.

[2] M. M. Ogonji, G. Okeyo, and J. M. Wafula, "A survey on privacy and security of Internet of Things," *Computer Science Review*, vol. 38, September 2020.

[3] "Network security polic," Wikipedia, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Network_security_policy. [Accessed 1 November 2023].

[4] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Surveying Port Scans and Their Detection Methodologies," *The Computer Journal Advance*, vol. 54, pp. 1565 - 1581, April 2011.

[5] S. O. Amin, M. S. Siddiqui, C. S. Hong, and J. Choe, "A Novel Coding Scheme to Implement Signature based IDS in IP based Sensor Networks," *IFIP/IEEE International Symposium on Integrated Network Management-Workshops*, pp. 269-274, June 2009.

[6] N. Naz et al., "Ensemble learning-based IDS for sensors telemetry data in IoT networks," *Mathematical Biosciences and Engineering*, vol. 19(10), p. 10550–10580, July 2022.

[7] A. S. Jaradat, M. M. Barhoush, and R. B. Easa, "network intrusion detection system: machine learning approac," *Indonesian Journal of Electrical Engineering and Computer Science* , vol. 25, p. 1151~1158, February 2022.

[8] "Anomaly-based intrusion detection system," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Anomaly-based_intrusion_detection_system. [Accessed 20 November 2023].

[9] S. Dhurandher, A. Kumar, M. Obaidat, "Cryptography-Based Misbehavior Detection and Trust Control Mechanism for Opportunistic

Network System," *IEEE SYSTEMS JOURNAL,* vol. 12, pp. 3191 - 3202, 2017.

[10] "The TON_IoT Datasets," From UNSW, [Online]. Available: https://research.unsw.edu.au/projects/toniot-datasets. [Accessed 1 December 2023].

[11] " IoT Healthcare Security Dataset," From kaggle, [Online]. Available: https://www.kaggle.com/datasets/faisalmalik/iot-healthcare-security-dataset . [Accessed 1 December 2023].

[12] M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali, and M. Guizani, "A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security," *IEEE Communications Surveys and Tutorials,* vol. 22, pp. 1646 - 1685, 2018.

[13] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, "Network Intrusion Detection for IoT Security based on Learning Techniques," *IEEE COMMUNICATIONS SURVEYS AND TUTORIALS,* vol. 00, 2019.

[14] J. Asharf, N. Moustafa, H. Khurshid, E. Debie, W. Haider, and A. Wahab, "A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Direction," *Electronics,* vol. 9(7), p. 1177, July 2020.

[15] B. I. Farhan and A. D. Jasim,"Survey of Intrusion Detection Using Deep Learning in the Internet of Things," *Iraqi Journal for Computer Science and Mathematics,* vol. 3, pp. 83 - 93, November 2021.

[16] S. V. N. Santhosh Kumar, M. Selvi, and A. Kannan, "A Comprehensive Survey on Machine Learning-Based Intrusion Detection Systems for Secure Communication in Internet of Thing," *Computational Intelligence and Neuroscience,* p. 24, December 2022.

[17] A. Aldhaheri, F. Alwahedi, M. A. Ferrag, and A. Battah, "Deep learning for cyber threat detection in IoT networks: A review," *Internet of Things and Cyber-Physical Systems,* vol. 4 , p. 110 – 128, September

2023.

[18] A. K. Sahu, S. Sharma, M. Tanveer, and R. Raja, "Internet of Things attack detection using hybrid Deep Learning Model," *Computer Communications,* vol. 176, p. 146–154, August 2021.

[19] S. D. Alotaibi et al., "Deep Neural Network-Based Intrusion Detection System through PCA," *Mathematical Problems in Engineering,* p. 9, April 2022.

[20] A. R. Gad, A. A. Nashat, and T. M. Barkat, "Intrusion Detection System Using Machine Learning for Vehicular Ad Hoc Networks Based on ToN-IoT Datase," *IEEE Access,* vol. 9, pp. 142206 - 142217, October 2021.

[21] N. A. Stoian, "Machine Learning for Anomaly Detection in IoT networks: Malware analysis on the IoT-23 Data set," *UNIVERSITY OF TWENTE STUDENT THESES,* p. 10, 2020.

[22] V. Bharathi and C. N. S. V. Kumar, "Vulnerability Detection in Cyber-Physical System Using Machine Learning," *Scalable Computing: Practice and Experience,* vol. 25, p. 577–591, 2023.

[23] Y. Al Sawafi, A. Touzene, and R. Hedjam, "A hybrid deep learning-based intrusion detection system for IoT networks," *Mathematical Biosciences and Engineering,* vol. 20(8), p. 13491–13520, May 2023.

[24] S. Choi and J. Cho, "Novel Feature Extraction Method for Detecting Malicious MQTT Traffic Using Seq2Seq," *Applied Sciences,* vol. 12, p. 12306, November 2022.

[25] M. M. Alani and A. Miri, "Towards an Explainable Universal Feature Set for IoT Intrusion Detection," *Sensors,* vol. 22(15), p. 5690, July 2022.

[26] A. Almotairi, S. Atawneh, O. A. Khashan, and N. M. Khafajah, "Enhancing intrusion detection in IoT networks using machine learning-based feature selection and ensemble model," *Systems Science & Control Engineering,* vol. 12, p. 2321381 , November 2023.

[27] I. Vaccari, M. Aiello, and E. Cambiaso, "SlowITe, a Novel Denial of Service Attack Affecting MQTT," *Sensors,* vol. 20(10), p. 2932, May 2020.

[28] N. Savanović *et al.*, "Intrusion Detection in Healthcare 4.0 Internet of Things Systems via Metaheuristics Optimized Machine Learnin," *Sustainability,* vol. 15(16), p. 12563, August 2023.

[29] F. Hussain *et al.*, "A Framework for Malicious Traffic Detection in IoT Healthcare Environment," *Sensors,* vol. 21(9), p. 3025, April 2021.

[30] M. B. Dissanayake, "Feature Engineering for Cyber-attack detection in Internet of Things," *I.J. Wireless and Microwave Technologies,* vol. 6, pp. 46-54, December 2021.

[31] Imran, M. F. A. Zuhairi, S. M. Ali, Z. Shahid, M. M. Alam, and M. M. Su'ud, "Improving Reliability for Detecting Anomalies in the MQTT Network by Applying Correlation Analysis for Feature Selection Using Machine Learning Techniques," *applied Sciences,* vol. 13(11), p. 6753, May 2023.

[32] "CVE-2018-1684 Detail," [Online]. Available: https://nvd.nist.gov/vuln/detail/CVE-20181684. [Accessed 20 December 2023].

[33] M. Hadded, G. Lauras, J. Letailleur, Y. Petiot, and A. Dubois,"An Assessment Platform of Cybersecurity Attacks against the MQTT Protocol using SIEM," *HAL,* p. 8, Oct 2022.

[34] J. Aveleira-Mata, A. Ibán-Sánchez, M. T. García-Ordás, I. García-Rodríguez, and H. Alaiz-Moreton, "Review and Replication of CoAP and MQTT Attacks for Dataset Generatio," *Open Access by IOS Press,* pp. 207 - 216, 2020 .

[35] "CVE-2016-10523 Detail," [Online]. Available: https://nvd.nist.gov/vuln/detail/CVE2016-10523 . [Accessed 20 December 2023].

[36] " CVE-2016-9877," [Online]. Available:

https://nvd.nist.gov/vuln/detail/CVE-2016-9877. [Accessed 20 December 2023].

[37] " EternalBlue Exploit: What It Is And How It Works," Sentinelone, May 2019 . [Online]. Available: https://www.sentinelone.com/blog/eternalblue-nsa-developed-exploit-just-wont-die. [Accessed 25 December 2023].

[38] D. Morato, E. Berrueta, E. Magaña, and M. Izal, "Ransomware early detection by the analysis of file sharing traffic," *Journal of Network and Computer Applications*, vol. 124, pp. 14-32, 2018.

[39] F. A. Barbhuiya, S. Roopa, R. Ratti, S. Biswas, and S. Nandi, "An Active Detection Mechanism for Detecting ICMP Based Attacks," *IEEE Computer Society*, pp. 51 - 58, 2012.

[40] S. Waichal and B. B. Meshram, "Router Attacks-Detection And Defense Mechanisms," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 2, pp. 145 - 149 , 2013.

[41] G. Al Sukkar, R. Saifan, S. Khwaldeh, M. Maqableh, and I. Jafar,"Address Resolution Protocol (ARP): Spoofing Attack and Proposed Defense," *Communications and Network*, vol. 8, pp. 118-130, August 2016.

[42] M. U. Nisa and K. Kifayat, "Detection of Slow Port Scanning Attacks," *International Conference on Cyber Warfare and Security (ICCWS)*, October 2020.

[43] "UDP Scan (-sU)," Nmap, [Online]. Available: https://nmap.org/book/scan-methodsudp-scan.html . [Accessed 5 February 2024].

[44] S. Goswami, N. Hoque, D. K. Bhattacharyya, and J. Kalita,"An Unsupervised Method for Detection of XSS Attack," *International Journal of Network Security*, vol. 19(5), pp. 761- 775, Apr 2016.

[45] Abdulrahman, "Lesson 2 : WebAppPen Cross Site Scripting XSS (Code Review)," Medium, 2021. [Online]. Available:

https://ph33r.medium.com/lesson-2-webapppen-cross-site-scripting-xsscode-review-66d04e8cdf99 . [Accessed 6 February 2024].

[46] K. Lee, J. Kim, K. H. Kwon, Y. Han, and S. Kim, "DDoS attack detection method using cluster analysis," *Expert Systems with Applications,* vol. 8, p. 1659–1665, 2008.