Communications of the CCISA
Vol. 31 No. 4 Nov. 2025

應用深度學習與注意力機制於物聯網多類別攻擊偵測之研究

張智開[1]、林詠章[2*]
中興大學資訊管理學系
[1] G7112029016@smail.nchu.edu.tw、[2]iclin@nchu.edu.tw

摘要

隨著物聯網 (IoT) 設備普及與 Mirai 殭屍網路原始碼被公開，IoT 環境面臨日益嚴峻的安全威脅。為解決這些挑戰，本研究提出一種基於深度學習與注意力機制的 IoT 多類別攻擊偵測方法。本研究以涵蓋 33 種攻擊類型的 CIC-IoT-2023 資料集為基礎進行實驗。針對 IoT 攻擊的時序特性，本研究改良並實現了三種時序深度學習模型：長短期記憶網路 (LSTM)、GRU (無譯名)，以及時間卷積網路 (TCN)。模型均結合統一注意力機制，以強化關鍵時序特徵的辨識。同時，針對 IoT 攻擊資料普遍存在的嚴重不平衡現象，研究建構多層次平衡策略，包括採用隨機欠取樣機制平衡各類別樣本分布、結合類別權重的 Focal Loss(Class-Weighted Focal Loss) 及加權隨機取樣 (Weighted Random Sampling) 技術，確保訓練批次的多類別均衡性。實驗結果顯示，TCN 結合注意力機制在多項指標皆達 99.5%以上，整體表現最佳；GRU 則具最高計算效率。相較其他研究部分，本研究在應用層攻擊偵測方面表現尤為突出，針對網頁攻擊以及偵察攻擊仍維持高偵測率，有效克服了現有研究在應用層 (Layer 7) 攻擊類型識別上的不足。本研究貢獻在於：(1)系統性比較 LSTM、GRU、TCN 於 IoT 多類別攻擊偵測之效能；(2)結合注意力與多元資料平衡策略，顯著提升應用層及少數類威脅之識別率。

關鍵字：物聯網安全、入侵偵測系統、深度學習、注意力機制、不平衡資料、CIC-IoT-2023

* 通訊作者 (Corresponding author.)

Communications of the CCISA
Vol. 31 No. 4 Nov. 2025

A Study on Multi-Class Attack Detection in IoT Networks Using Deep Learning and Attention Mechanisms

Chang, Chih-Kai[1], Lin, Iuon-Chang[2]
[12] Master of Business Administration in Management Information Systems , National Chung Hsing University
[1] G7112029016@smail.nchu.edu.tw、[2]iclin@nchu.edu.tw

Abstract

With the widespread adoption of Internet of Things (IoT) devices and the public release of the Mirai botnet source code, IoT environments are facing increasingly severe security threats. To address these challenges, this study proposes a deep learning–based multi-class IoT attack detection method incorporating an attention mechanism. Experiments were conducted using the CIC-IoT-2023 dataset, which includes 33 types of attack traffic. Considering the temporal characteristics of IoT attacks, this study implements and customizes three temporal deep learning models—Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Network (TCN). Each model is integrated with a unified attention mechanism to enhance the recognition of critical temporal features. Meanwhile, to mitigate the serious data imbalance commonly found in IoT attack datasets, a multi-level balancing strategy was developed. This strategy combines random undersampling to equalize class distributions, Class-Weighted Focal Loss, and Weighted Random Sampling to ensure balanced batches during training. Experimental results show that the TCN achieved the best overall performance, with all major metrics exceeding 99.5%, while the GRU achieved the highest computational efficiency. Compared with existing studies, the proposed method demonstrates superior performance in application-layer attack detection, maintaining high detection rates for web and reconnaissance attacks, and effectively overcoming current limitations in Layer 7 attack recognition. The contributions of this work are as follows: (1) a systematic performance comparison of LSTM, GRU, and TCN models for multi-class IoT attack detection; and (2) the integration of attention mechanisms with multi-level data balancing strategies, which significantly improve the detection of application-layer and minority-class threats.

Keywords: IoT security, intrusion detection system, deep learning, attention mechanism, imbalanced data, CIC-IoT-2023

一、緒論

隨著物聯網 (Internet of Things, IoT) 設備應用日漸普及，根據 IoT Analytics 的 *State of IoT Summer 2024* 報告，全球 IoT 設備數量預計於 2030 年將達到 400 億台，廣泛應用於智慧城市、醫療保健與工業自動化等領域 [1]。研究指出，物聯網因節點資源受限，難以實施標準安全機制，進而增加遭受攻擊的風險 [2]。特別是自 2016 年 Mirai 惡意軟體原始碼公開後 [3]，針對 IoT 設備的攻擊手法持續演進，從分散式阻斷服務 (Distributed Denial of Service, DDoS) 到多態的惡意軟體，對網路安全與資料隱私構成重大挑戰 [4]。

根據 OWASP Internet of Things 項目的報告，IoT 設備常因薄弱的安全措施成為攻擊目標 [5]，例如 2016 年 Mirai 殭屍網路利用設備弱點發動大規模 DDoS 攻擊，2022 年更超過 1.12 億台 IoT 設備遭受攻擊。常見的漏洞包括跨站腳本攻擊 (XSS) 和 SQL 注入 [6]，這些應用層漏洞對 IoT 設備儼然構成嚴重威脅。Trend Micro 的研究 [7] 進一步指出，攻擊者能夠通過應用層的漏洞竊取用戶認證或推送惡意韌體更新。根據 Cloudflare 2024 年第四季報告，Mirai 攻擊佔所有網路層 DDoS 攻擊的 6%，與上一季相比增長了 131%。在 2024 年 10 月 29 日時，Mirai 變體殭屍網路發起了一次 5.6Tbps 的 UDP DDoS 攻擊，是到當時規模最大的 DDoS 攻擊，來自超過 13,000 個 IoT 裝置 [8]。

為應對 IoT 環境中日益複雜的網路安全威脅，Neto 等人 [4] 指出，大多數現有 IoT 資料集因攻擊類型涵蓋目標不足，難以支持泛用型入侵偵測系統的開發。他們提出的 CIC-IoT-2023 資料集，包含 105 個真實 IoT 設備在 33 種攻擊與正常流量下的封包流量資料，為全面評估入侵偵測系統提供了堅實基礎。

傳統網路入侵偵測系統 (Network Intrusion Detection System, NIDS)，如基於規則的入侵偵測或簡單的特徵匹配方法，在處理 IoT 環境的多樣化威脅時存在諸多限制，包括對未知攻擊的偵測能力有限以及計算資源消耗過高 [9]。近年來，深度學習技術在網路入侵偵測領域展現出顯著優勢，特別是注意力機制 (Attention Mechanism)，能夠動態聚焦於輸入資料的關鍵特徵，並降低計算成本 [10]。

綜上所述，IoT 設備數量日益增長，但因其運作環境異質性與設備資源限制，面臨日益複雜網路安全威脅的情況。基於上述挑戰，本研究旨在開發一種基於深度學習的泛用型入侵偵測系統，針對 IoT 設備環境中多樣化的網路攻擊進行高效偵測。具體而言，本研究將利用提供了多樣化的攻擊場景與真實流量資料的 CIC-IoT-2023 資料集 [4]，使用深度學習結合注意力機制，設計一種能夠適應 IoT 流量時序特性、兼具高效與輕量化的偵測模型，並運用資料平衡策略，進一步提升模型的泛化能力，為 IoT 安全提供可重現構建的解決方案。

本研究將利用通過系統性比較三種主流深度學習架構：長短期記憶網路 (Long Short-Term Memory, LSTM)、GRU (無譯名，Gated Recurrent Unit) 與時間卷積網路 (Temporal Convolutional Network, TCN)，分別結合注意力機制，使用針對 IoT 裝置進行攻擊所建立資料集 CIC-IoT-2023 [4]，並考量網路流量中封包的時間順序性對多樣本偵

測任務的表現，評估其精確率、召回率與模型複雜度，並能有效改善資料不平衡現象導致的少數類召回率過低的問題。

二、文獻探討

2.1. IoT安全挑戰與深度學習應用概述

IoT 因運作環境異質性、設備資源限制及大規模連網特性，面臨獨特的安全挑戰 [2]。這些挑戰包括設備間通訊協定的多樣性、運算能力的限制，以及因大規模連網導致的複雜攻擊模式，亦即多類別攻擊的適用性。深度學習 (Deep Learning, DL) 作為機器學習 (Machine Learning, ML) 的子領域，自 LeCun 等人 [11] 提出卷積神經網路 (Convolutional Neural Network, CNN) 與 Hochreiter 等人 [12] 開發 LSTM 以來，在處理高維度、非結構化資料方面展現出顯著優勢；在 IoT 安全中，深度學習被用於異常偵測 (Anomaly Detection) 與網路行為分析 (Network Behavior Analysis)，能有效識別傳統機器學習難以捕捉的細微模式，如：支援向量機 (Support Vector Machine, SVM)、單純貝氏分類器 (Naive Bayes) [13]。此外，注意力機制作為深度學習的關鍵進展，已被應用於網路安全中，以增強對複雜攻擊模式的辨識能力 [14, 15]；針對 IoT 資料集常見的資料不平衡問題，資料平衡策略已被證實能有效提升模型性能，並被廣泛應用於網路入侵偵測系統研究 [16, 17, 18, 19]。透過整合深度學習、注意力機制與資料平衡策略，本研究旨在探討適應 IoT 安全特性並具備多類別攻擊偵測的穩健框架。

2.2. 時序模型與注意力機制的運用

Tseng 等人 [14] 使用深度神經網路 (Deep Neural Network, DNN)、RNN、CNN、CNN+RNN、CNN+LSTM、Transformer 模型處理 CIC-IoT-2023 資料集，發現在多分類中 Transformer 模型的表現最佳。該研究暗示了注意力機制在處理複雜多類別分類任務時的優勢，但未深入探討其作用機制。在過往的研究中，Sun 等人 [20] 利用 CNN 和 LSTM 來提升偵測準確度，同時納入了 payload 作為特徵，並引入了類別權重方法來提高模型的穩健性，在 CIC-IDS-2017 資料集上對每種攻擊類型的準確度均高於 99.50%。但 Akbari 等人 [21] 指出，至 2019 年，超過 90%的網頁瀏覽器流量採用 HTTPS 等對訊息加密的協定，在此情況下將 payload 納入模型沒有意義，只會增加過擬合的可能性。Jony 和 Arnob [22] 使用了 LSTM 應用於 CIC-IoT-2023 資料集對 33 個攻擊與正常流量進行分類，並取得了 98.75%的準確率，驗證了 LSTM 在處理 IoT 流量方面的有效性。Wu 等人提出 RTIDS (Robust Transformer-based IDS) [15]，專注於解決多資料不平衡資料集的處理問題，用自注意力機制防止過擬合，並使用合成少數過取樣技術 (Synthetic Minority Oversampling Technique, SMOTE) 處理資料不平衡問題，在 CIC-IDS-2017 資料

集上達到 98.45%準確率。Nazre 等人 [23] 針對 Edge-IIoTset 資料集的 14 種攻擊類型和正常流量進行多類別分類，相較於 CNN、CNN-GRU、CNN-LSTM、CNN- BiLSTM (雙向長短期記憶網路，Bi-directional Long Short-Term Memory) 和 CNN-LSTM-GRU，發現 TCN 模型表現最佳，TCN 不但能平行計算降低訓練時間，還能克服傳統循環神經網路 (RNN 相關模型) 的梯度消失與高計算複雜度問題。Khan [24] 提出 HCRNNIDS 混合架構，結合 CNN 和循環神經網路 (Recurrent Neural Network, RNN) 的優勢。在 CIC-IDS-2018 資料集上實現 97.75%的偵測準確率，誤報率降至 2.5%。該研究的主要貢獻是將用於空間特徵提取的 CNN 與用於時序特徵學習的 RNN 有效結合，並採用過取樣技術處理資料不平衡問題。Wang 等人 [25]提出 DL-BiLSTM 模型，結合 DNN 與 BiLSTM，針對 CIC-IDS-2017、N-BaIoT 和 CIC-IoT-2023 資料集進行 IoT 入侵偵測。該模型採用增量主成分分析 (Instrumented Principal Component Analysis, IPCA) 降維與動態量化 (Dynamic Quantization) 技術，顯著降低模型大小。

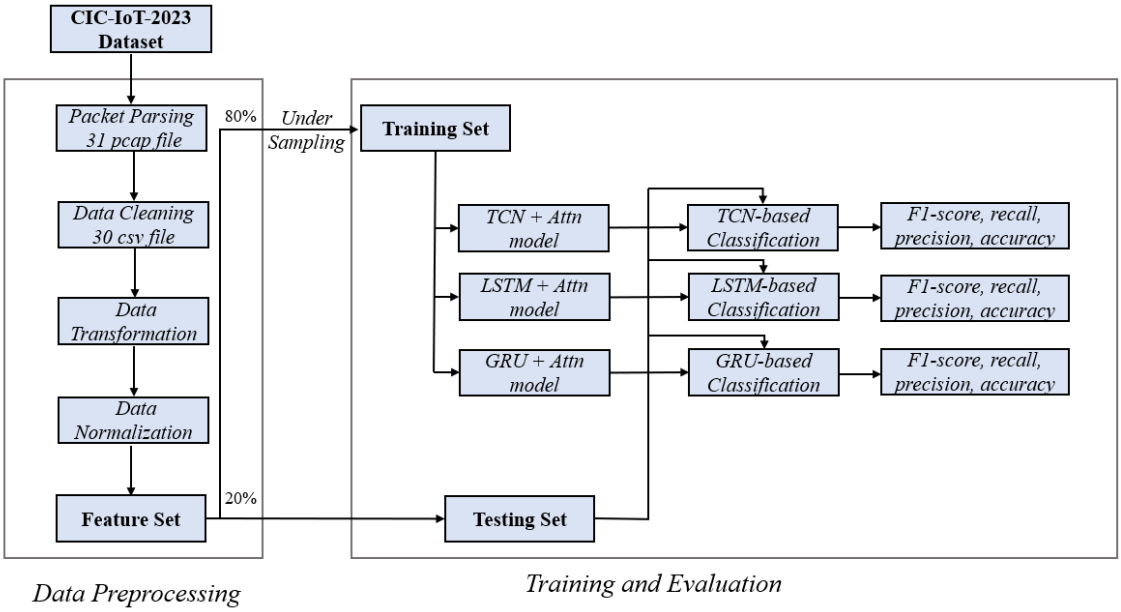
2.3. 資料不平衡問題的解決策略

IoT 攻擊樣態多元，各項研究所產出的資料集種類眾多，部分攻擊資料種類樣本數龐大，如 DoS/DDoS 類別，部分攻擊資料種類樣本數稀少；如應用層的攻擊類型，但這些攻擊類型是相當重要且不可忽略的。資料不平衡是 NIDS 資料集普遍存在的問題，現有研究已經有了多種解決策略，但多為單一技術的應用。Behera 等人 [16] 設計了軟體定義 IoT (SDIoT) 環境的 DDoS 偵測方法，整合 SMOTE-ENN 資料平衡策略、SFMI 特徵選擇、PCA 降維，並使用決策樹 (Decision Tree)、K-近鄰演算法 (K-Nearest Neighbor)、高斯單純貝氏 (Gaussian Naive Bayes)、RF 和多層感知器 (Multilayer Perceptron, MLP) 進行分類，證實了 RF 的效能最好，且可在僅使用 10 個特徵的情況下進行分類。Gheni 和 Al-Yaseen [17] 提出結合 GSK(Gaining-Sharing Knowledge) 演算法的聚類方法來處理資料不平衡問題與特徵降維 (Dimension Reduction)，配合 MLP 和自編碼器 (Autoencoder, AE)，並在 MLP 上達到了 97.46%的準確率。Doost 等人 [18] 在 UNSW-NB15 資料集上進行實驗，使用 CNN 進行特徵降維與 RF 的分類方法，並針對少數攻擊類型過取樣 (Over-Sampling) 和對多數類別欠取樣 (Under-Sampling)，使精確率超過 98%。Mulyanto 等人 [19] 提出了一種基於 Focal Loss 的神經網路方法，稱為 FL-NIDS，展示了 Focal Loss 在 DNN 和 CNN 架構中，對於少數類分類結果的顯著改進。

三、研究架構設計

本研究以 CIC-IoT-2023 資料集為基礎，設計一套系統化的資料處理流程，將原始網路封包 (pcap 檔案) 轉換為適合網路入侵偵測系統的結構化特徵資料集，結合三種資料平衡策略，採用欠取樣技術、結合類別權重的 Focal Loss (Class-Weighted Focal Loss) 以

及加權隨機取樣 (Weighted Random Sampling) 方法，應用於三種基於深度學習的異常偵測模型：LSTM、GRU 與 TCN，每種模型均結合注意力機制。圖一展示了本研究的整體方法流程，包括資料處理、模型訓練與評估的各個階段。



圖一：整體架構圖

本研究採用由加拿大網路安全研究所 (Canadian Institute for Cybersecurity, CIC) 發布的 CIC-IoT-2023 資料集 [4]，利用該資料集的多樣化攻擊場景與高品質特徵進行訓練和測試資料。該資料集中的正常網路流量類型包含音響設備、攝影設備、照明設備、電源插座、感測器、智慧家電等裝置產生之資料；而惡意流量包含 33 種攻擊，分屬 7 大類別：DDoS, DoS, Recon, Web-based, brute force, spoofing 和 Mirai。是目前公開發表中規模最大、最全面的 IoT 攻擊資料集，最少樣本的類別樣本量也超過了 1000 筆，設備規模涵蓋 105 個真實 IoT 設備(詳見表一)，模擬真實智慧家庭環境的大規模 IoT 部署。

表一：CIC-IoT-2023 資料集設備分類統計

設備類別	數量	代表設備	主要功能
音響設備	8	Amazon Alexa Echo Dot, Google Nest Mini	語音控制、智慧助理
攝影設備	25	AMCREST WiFi Camera, Nest Indoor Camera	安全監控、影像串流
照明設備	8	Philips Hue Bridge, LIFX Lightbulb	智慧照明控制
電源插座	15	Teckin Plug, Wemo Smart Plug	電源管理、遠端控制
感測器	23	Fibaro Motion Sensor, Ring Contact Sensor	環境監測、安全偵測
智慧家電	12	LG Smart TV, iRobot Roomba	智慧型電視、清潔機器人
攻擊設備	7	Raspberry Pi 4	攻擊執行平台
中樞設備	7	SmartThings Hub, Ring Base Station	設備管理、協定轉換

在資料集取樣部分，為整合網路流量順序並適應本研究設定的情境，本研究依據 TCP/UDP Stream ID 劃分流量，排除僅涉及 Layer 2 與 Layer 3 的封包，因此將影響較為明顯的類別排除：ARP Spoofing 的封包都僅使用了第二層的協議，因此無法藉由第四層的協議來劃分；DDoS ICMP Fragmentation 和 DDoS ICMP Flood 大多封包都可被轉換成流量，但有大量僅涉及到第三層的 ICMP 跟 IGMP 的封包被排除，僅包含第三層協定的封包因無法明確劃分來源與目的，無法轉換為時間序列，因此本研究框架不將其納入處理與預測範圍。再來由於 GREIP Flood 與 GREETH Flood 的差異主要在 GRE 協議下的特徵，因此本研究將特徵相近的 GREIP Flood 與 GREETH Flood 合併為 GRE Flood。考慮到資料量充足，本研究挑選各攻擊類別的第一個 pcap 檔案(最大 2GB)作為訓練與測試資料，以確保資料代表性與計算效率。

為確保模型訓練與評估的公正性，本研究將 CIC-IoT-2023 資料集轉換成時序資料後，再參考 [26] 使用分層取樣按 8：2 的比例分割為訓練集與測試集，避免隨機分割破壞時序資料的連續性，並確保劃分後的資料集比例相等。為明確特徵資料在網路架構中的作用，本研究分析了各特徵所屬的網路層、對應協定及其在 IoT 安全偵測中的用途，詳見表二。在資料連結層、網路層、傳輸層、應用層有不同共 **16 種**特徵。

表二：網路流量特徵

特徵類別	OSI 層	對應協定	說明
frame.len	資料連結層	Ethernet	描述封包總長度，識別異常流量模式
ip.proto	網路層	IP	標識傳輸層協議類型，區分 TCP/UDP 等流量
ip.ttl	網路層	IP	表示封包生存時間，偵測異常路由行為
icmp.type	網路層	ICMP	表示 ICMP 訊息類型
tcp.window_size	傳輸層	TCP	控制 TCP 流量窗口，識別流量擁塞或攻擊
dstport	傳輸層	TCP/UDP	標識目標端口，分析服務類型與攻擊目標
srcport	傳輸層	TCP/UDP	標識來源端口，追蹤流量來源與攻擊模式
transport.layer_len	傳輸層	TCP/UDP	描述傳輸層資料長度，偵測異常流量大小
stream	傳輸層	TCP/UDP	標識流量會話，支援流量的分組
tcp.flags	傳輸層	TCP	標識 TCP 控制位，分析協議狀態與攻擊行為
direction	傳輸層	TCP/UDP	標識封包在該流量中是 inbound/outbound
http.request.method	應用層	HTTP	表示 HTTP 請求類型，偵測應用層攻擊
dns.qry.name	應用層	DNS	表示 DNS 查詢名稱，偵測域名相關攻擊
quic.version	應用層	QUIC	表示 QUIC 協議使用情況，偵測新興協議攻擊
timestamp	-	-	提供時間資訊，用於時序分析與異常偵測
label	-	-	分類，標示攻擊類別與正常流量

為將 CIC-IoT-2023 資料集轉換為適於深度學習模型的時間序列格式，本研究從原始 pcap 檔案中提取特徵，而非使用資料集提供的預處理 csv 檔案，以確保特徵的時序一致性，充分利用時序資訊，並提升方法的通用性，即適用於其他採用不同特徵的資料集。本研究之資料預處理流程包含四個步驟：封包解析 (Packet Parsing)、資料清理 (Data

Cleaning)、資料轉換 (Data Transformation) 及資料正規化 (Data Normalization)。

第一步是 Packet Parsing，從原始 pcap 檔中提取網路封包的特徵，本研究採用 tshark 工具進行封包解析，tshark 是一種廣泛使用的網路協議分析工具，能夠高效提取封包中的特徵。

第二步是 Data Cleaning，旨在去除無效或不相關的資料。首先排除 ip.proto 為空的封包，因為這些封包通常表示不完整的記錄或解析錯誤，無法提供有意義的特徵。接著對資料進行隨機欠取樣，通過隨機移除多數類樣本，減少資料集中的資料不平衡問題。然後將封包依 TCP/UDP Stream ID 重新組織為時間序列，單一時間序列長度上限為 20，組成後移除 TCP/UDP Stream ID，同時添加特徵「direction」，用以表示封包的傳輸方向。

第三步是 Data Transformation，將原始特徵轉換為適合機器學習模型的格式，同時提取與入侵行為相關的衍生特徵。利用 timestamp 來計算相對時間 (Relative Time, rt)，封包間隔時間 (Inter-Arrival Time, iat)，rt 表示封包相對於第一個封包的時間偏移，iat 表示相鄰封包之間的時間差；對 http.request.method 和 ip.proto 分別使用 One Hot Encoding，將字串轉換為數值以適應機器學習模型的輸入需求；對 tcp.flags 進行位元拆解 (bitwise decomposition)，並轉換為多個獨立的二元特徵；對 quic.version, icmp.type 和 dns.qry.name 進行二元特徵轉換。

第四步是 Data Normalization，由於前面已經對字串型態的資料進行轉換，可以對所有特徵進行正規化處理，減少模型受到極值影響，提高模型訓練的穩定性。其中x為原始資料，μ為該特徵之平均值，σ為該特徵之標準差，x'為正規化後的資料，x_{min}為該特徵值的下限，x_{max}為該特徵值的上限。

對所有類別的 rt 與 iat 使用標準化 (Standardization)，將其轉換為均值為 0、標準差為 1 的分布，以消除不同時間尺度的影響。方程式如下：

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

對其餘數值型特徵利用封包特徵大小受規範限制的特性，進行最小值最大值歸一化 (Min-max scaling)，方程式如下：

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

在網路流量分類任務中，資料集的類別分布通常不平衡。例如，在 CIC-IoT-2023 資料集中，DDoS 相關攻擊佔據大部分樣本，應用層的攻擊類型樣本數較少，顯示出了不平衡問題的挑戰。設資料集包含C個類別，總樣本數為 N，每個類別 i 的樣本數為N_i，類別分布的方程式可表示為：

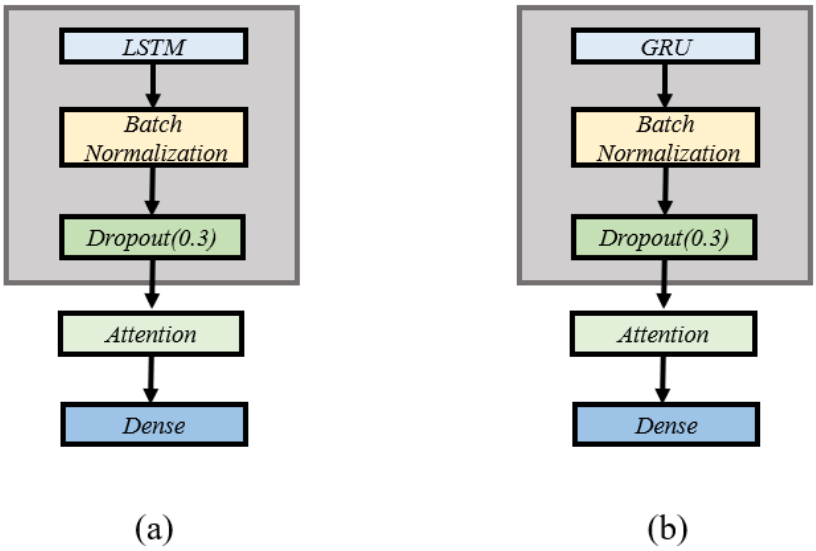
$$N = \sum_{i=1}^{C} N_i \quad (3)$$

不平衡資料集中，某些N_i(多數類)遠大於N_j(少數類)，導致模型偏向多數類，少數類的召回率和 F1 分數較低，亦即在少數類攻擊的偵測能力上有所缺陷。本研究通過以下三種方法應對這一問題：

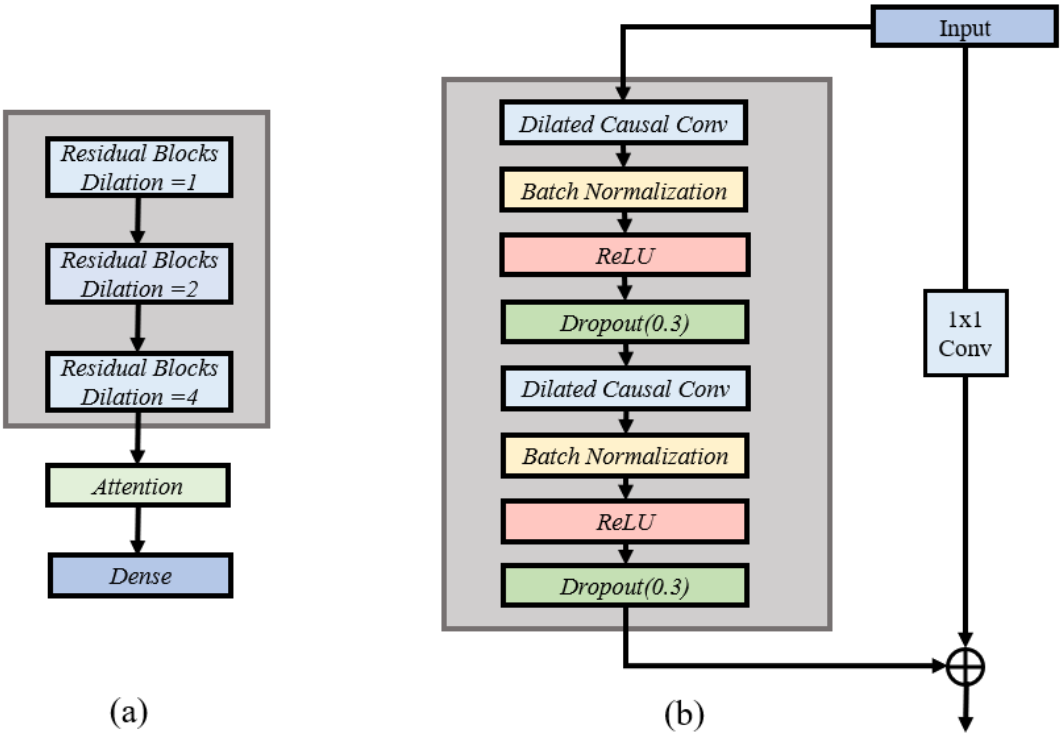
- 隨機欠取樣：減少多數類樣本數。
- 結合類別權重的 Focal Loss：在損失函數 Focal Loss 中，為少數類分配更高權重。
- 加權隨機取樣：通過加權隨機取樣確保批次級別的類別平衡。

本研究引用的三種基於注意力機制的深度學習框架，包含 LSTM、GRU 以及 TCN，以高效捕捉 PCAP 轉化後的時序特徵並實現異常流量偵測。該框架結合批次標準化 (BatchNorm1d, BN)、Dropout、注意力機制與全連接 (Fully Connected, Dense) 層，模型能夠有效捕捉時序資料中的複雜依賴關係並聚焦關鍵特徵，充分應對資料不平衡與多維度特徵的挑戰。以下分節介紹模型的架構、注意力機制、全連接層和分類方法的由來、設計及實現方式。

本研究提出的 LSTM + Attention 架構如圖二 (a) 所示，包含 3 層 LSTM 層、BN 層和 Dropout 層。另一個模型是 GRU + Attention 架構，如圖二 (b) 所示，由主要由 3 層 GRU 層、BN 層、Dropout 層、注意力層和全連接層組成。GRU 是 LSTM 的精簡版，專注於記憶效率優化，在確保效能的同時不影響其準確性，用於處理序列資料，通過更新閥和重置閥捕捉時序依賴、確保因果性。為了減少過擬合，在訓練過程中對每個批次進行標準化，並隨機丟棄部分神經元，防止對特定特徵的過度依賴。第三個模型是 TCN + Attention 架構，如圖三 (a) 所示，由三個殘差塊構成。殘差塊的架構如圖三 (b) 所示，由 kernel size=5 的因果卷積 (Causal Convolutions) 層、BN 層和 Dropout 層組成。TCN 使用因果卷積確保時間序列的因果性，利用殘差學習 (Residual Connection) 增強特徵傳遞能力，即使其他部分在訓練過程中出現了梯度消失或梯度爆炸也能順利反向傳播。



圖二：(a) LSTM 架構，(b) GRU 架構



圖三：(a) TCN 模型架構，(b) 殘差塊架構

四、實驗結果

為提升深度學習模型的訓練與推論效率，本研究採用 NVIDIA CUDA 12.6.0 與 PyTorch 2.6.0 構建 GPU 加速環境，相較於純 CPU 計算，可大幅縮減訓練時間。
為確保三種時序模型在 IoT 多類別攻擊偵測任務中的公平比較，本研究對所有模型採用統一的超參數設定。於表三中介紹各超參數的值。

表三：Hyperparameter of models

Batch Size	128
Epochs	50
Optimizer	Adam
Learning Rate	0.001
Dropout	0.3
Hidden Layers	3
Neurons(LSTM, GRU)	256
Tunnel(TCN)	64, 128, 256
Loss Function	Focal Loss(gamma=2.0, label smoothing =0.1)

GRU + Attention、LSTM + Attention 和 TCN + Attention 的評估結果如表四所示。

表四：本研究三個不同模型的效能評估

	$\boldsymbol{Accuracy_{micro}}$	$\boldsymbol{PRC_{macro}}$	$\boldsymbol{Recall_{macro}}$	$\boldsymbol{F1_{macro}}$
GRU + Attention	0.9936	0.9620	0.9790	0.9698
LSTM + Attention	0.9942	0.9648	0.9879	0.9756
TCN + Attention	**0.9954**	**0.9711**	**0.9885**	**0.9793**

可發現三者在整體準確率表現、類別平衡偵測能力、攻擊識別能力都很高，其中 TCN + Attention 的擁有最好的性能，因此接下來與其他論文的對比都是以 TCN + Attention 為主。每個模型的隱藏神經元數量及對應的參數量如表五所示，可發現 TCN + Attention 所需的記憶體相對較高，不過 TCN 的結構設計使其在處理序列資料時可能具有更高的並行性和計算效率。

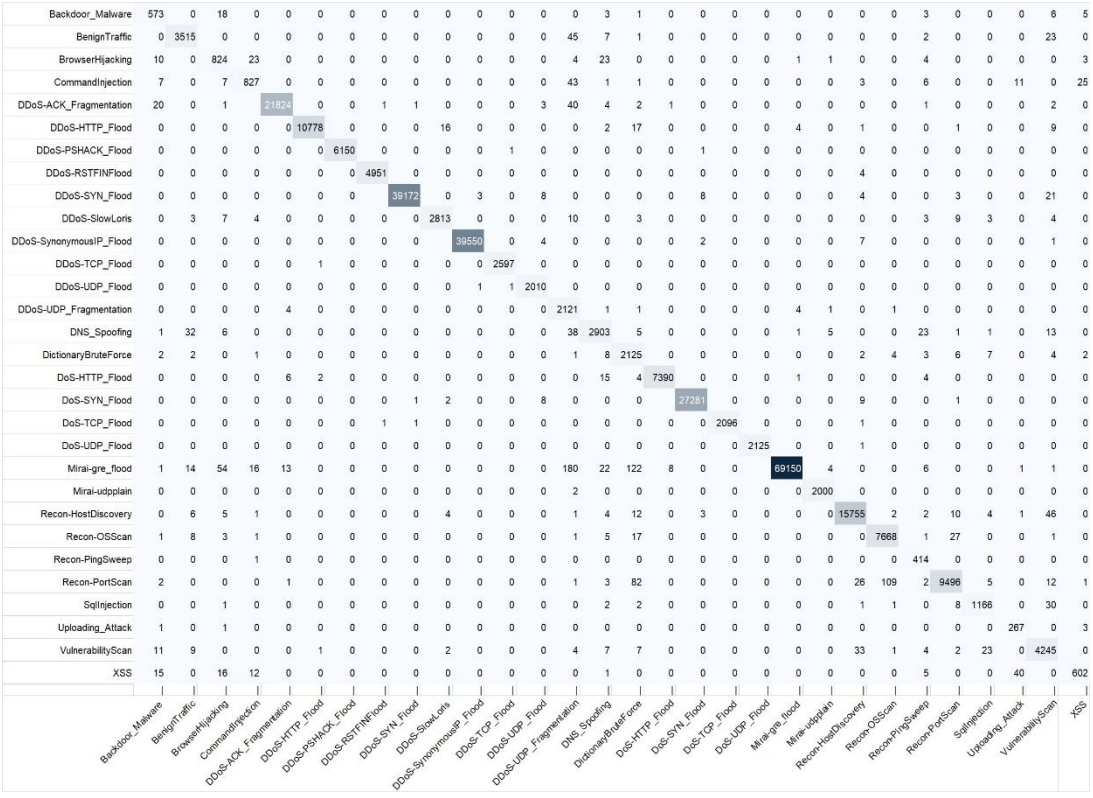
TCN 模型因其並行卷積結構能夠有效捕捉時序特徵，結合注意力機制後進一步提升了對關鍵時序模式的識別能力。同時，殘差學習設計有助於緩解梯度消失問題，使模型能夠學習更深層的特徵表示。

表五：本研究三個不同模型的隱藏神經元數量和參數量

Label	隱藏神經元數量	參數量
GRU + Attention	3 層×256=768	1,115,168
LSTM + Attention	3 層×256=768	1,451,296
TCN + Attention	每層 2 組×(64+128+256)=986	1,502,912

基於混淆矩陣的深入分析，本研究發現了不同攻擊類型的偵測特徵模式。在最佳模型 (TCN + Attention) 中，DDoS 系列攻擊 (包括 TCP Flood、UDP Flood、SYN Flood 等) 均達到 95%以上的偵測準確率，證實了這類攻擊的網路流量特徵明顯且易於識別。相較之下，應用層攻擊的偵測準確率相對較低，反映了這些攻擊在網路流量層面的隱蔽能力，也再次證實應用層攻擊偵測在 IoT 安全領域的挑戰性。

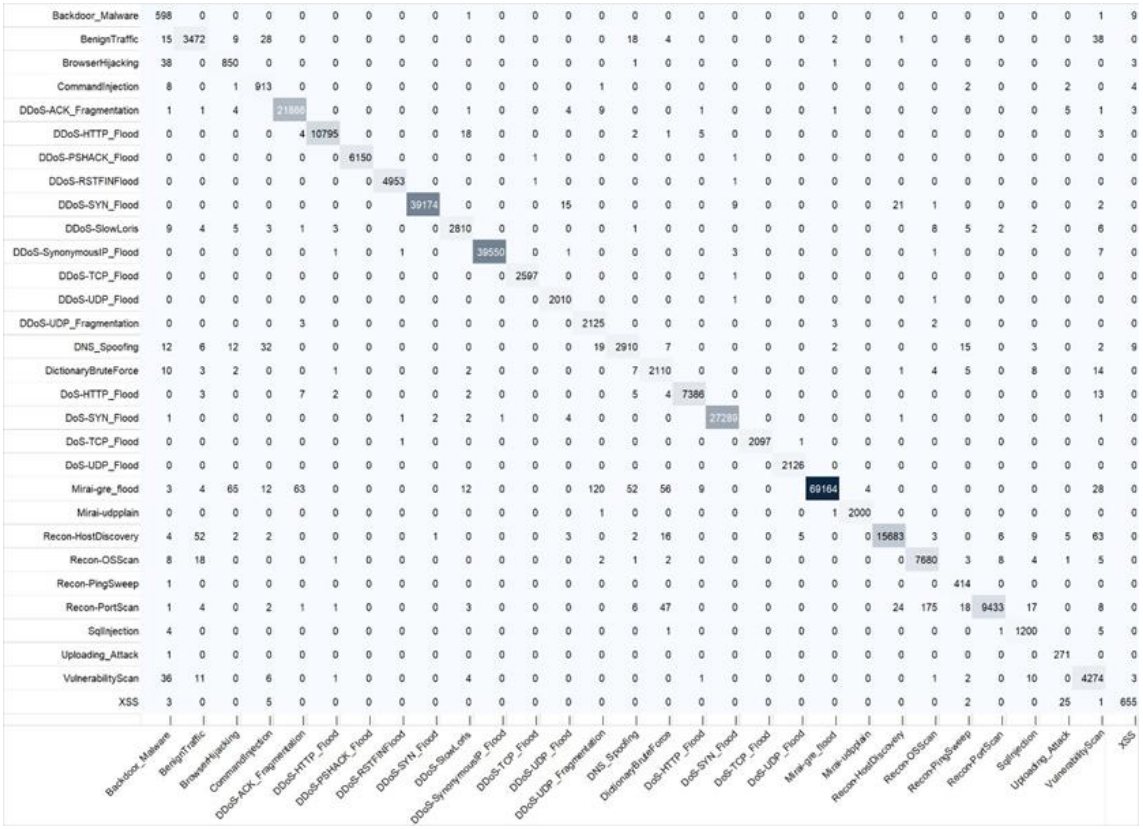
GRU + Attention 模型實驗結果的混淆矩陣如圖四所示，有六個類別的精確率不足 90%，分別為 Backdoor Malware (88.98%)、Browser Hijacking (87.38%)、DDoS UDP Fragmentation (85.15%)、Dictionary Brute Force (88.47%)、Recon Ping Sweep (85.71%)、SQL Injection (83.44%)。此外，有兩個類別的召回率不足 90%，分別為 Command Injection (88.83%) 和 Vulnerability Scan (87.12%)，這兩種類別樣本數也偏少。



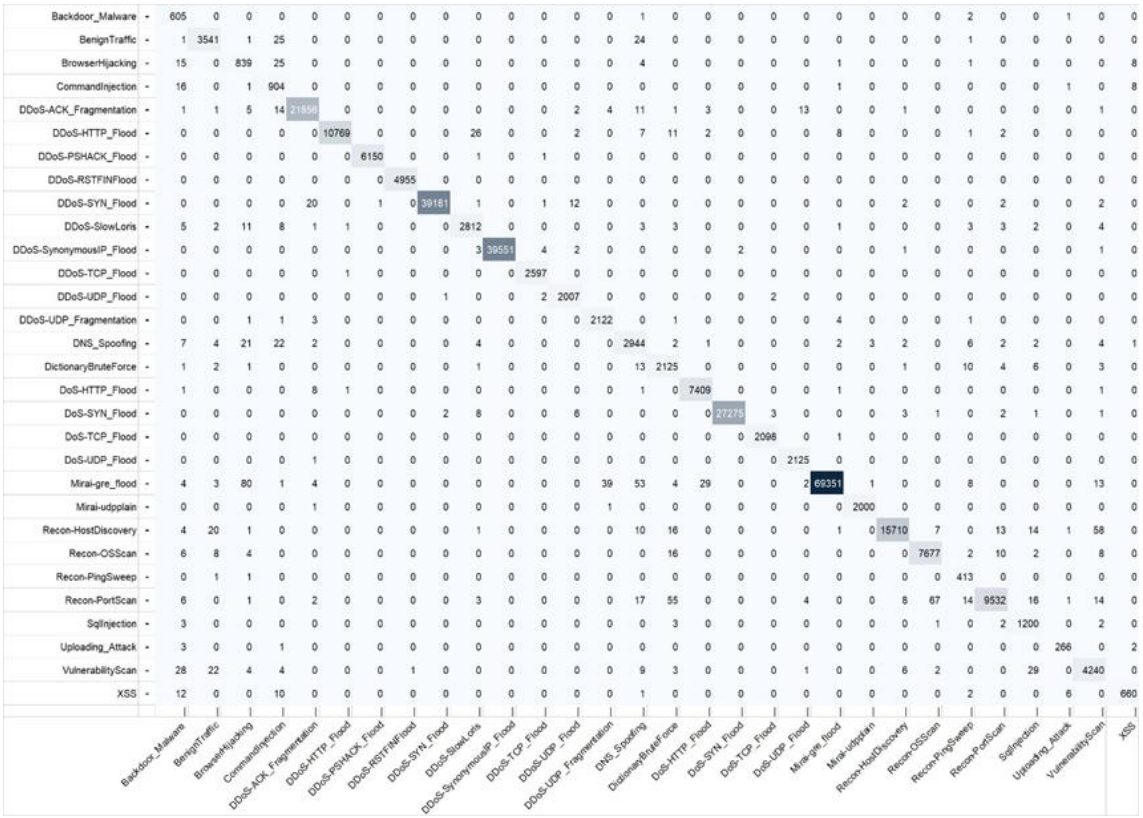
圖四：GRU + Attention 模型的混淆矩陣

LSTM + Attention 模型實驗結果的混淆矩陣如圖五所示，有五個類別的精確率不足90%，分別為 Backdoor Malware (79.42%)、Browser Hijacking (89.47%)、Command Injection (89.60%)、Recon Ping Sweep (87.71%) 和 Uploading Attack (87.70%)。召回率方面，LSTM 模型未有類別低於 90%，表現相對穩定。

TCN + Attention 模型實驗結果的混淆矩陣如圖六所示，有四個類別的精確率不足90%，分別為 Backdoor Malware (84.26%)、Browser Hijacking (86.40%)、Command Injection (89.06%)、Recon Ping Sweep (89.01%)。與 GRU 和 LSTM 相比，TCN 在精確率低於 90% 的類別數量較少，顯示其分類穩定性較高。召回率方面，TCN 模型同樣未有類別低於 90%。



圖五：LSTM + Attention 模型的混淆矩陣



圖六：TCN + Attention 模型的混淆矩陣

表五彙總精確率 (Precision) 和召回率 (Recall) 低於 90%的類別，與基於 GRU 的模型和基於 LSTM 的模型相比，基於 TCN 的模型在精確率低於 90%的類別數量最少，顯示其分類穩定性較高；召回率方面，基於 TCN 的模型與基於 LSTM 的模型未有類別低於 90%。

表五：各模型對應評估指標低於 90%的類別

模型	類別		評估指標	(%)
GRU + Attention	Web	Backdoor Malware	精確率	88.98
	Web	Browser Hijacking		87.38
	DDoS	UDP Fragmentation		85.15
	Dictionary Brute Force			88.47
	Recon	Ping Sweep		85.71
	Web	SQL Injection		83.44
	Web	Command Injection	召回率	88.83
	Recon	Vulnerability Scan		87.12
LSTM + Attention	Web	Backdoor Malware	精確率	79.42
	Web	Browser Hijacking		89.47
	Web	Command Injection		89.60
	Recon	Ping Sweep		87.71
	Web	Uploading Attack		87.70
TCN + Attention	Web	Backdoor Malware	精確率	84.26
	Web	Browser Hijacking		86.40
	Web	Command Injection		89.06
	Recon	Ping Sweep		89.01

本研究以 CIC-IoT-2023 資料集為基準，比較 Macro-Average 精確率、召回率及 Micro-Average 準確率。表六為與其他文獻比較的結果，其中僅顯示該文獻中最優秀模型，數值是從參考文獻中取得，或根據其提供的混淆矩陣計算得出。粗體表示採用該資料集的 33 種攻擊與正常流量(共 34 類)進行 Fine-Grained 分類訓練，否則採用 7 大類與正常流量(共 8 類)進行 Coarse-Grained 分類訓練。現有研究在分類策略上呈現明顯差異，[4]、[14]、[17]、[16]均採用 Fine-Grained 分類，[4]、[14]、[17]都是採用 8 類分類，[16] 則是排除正常流量採用 7 類分類；而本研究與 [22]、[25]都是採用 Fine-Grained 分類，[22] 採用 34 類，[25] 採用 8 類分類，本研究則是採用 30 類分類。

其中，[4] 作為資料集的原始發布研究，採用羅吉斯迴歸 (Logistic Regression)、感知器 (Perceptron)、Adaboost、RF 和 DNN 進行 34 類攻擊分類，但根據其文字敘述與 8 類分類混淆矩陣數值比對，以及從 20%測試集的資料量進行計算比對，可發現他提供的 Accuracy 數值正確，但 Recall 與 Precision 較可能是誤植，因此，本研究表 12 是以該文獻中具有混淆矩陣數值的 8 類分類來進行比較。在 Wang 等人 [25] 的研究中，僅針對屬於偵察 (Recon) 攻擊與 Mirai 攻擊這兩個 IoT 流量中較新穎的攻擊類型的類別以及正常流量進行 Fine-Grained 分類。

不同研究文獻在處理 CIC-IoT-2023 資料集的資料不平衡問題時採用不同策略。[4]、[14]、[22]、[25] 未採用專門平衡技術，[17] 使用 Gaining-Sharing Knowledge，[16] 採用 SMOTE-ENN 技術。本研究整合欠取樣、類別權重和加權隨機取樣多種資料平衡策略，有效解決資料不平衡問題，使 Macro-Average 召回率達到 98.85%，在目前可見文獻中表現最佳。

在技術創新與實用性方面，本研究展現多重優勢。實驗結果呈現 30 類細分攻擊偵測中達到 97.11%的精確率，相較於現有 Fine-Grained 分類研究 [22] 提升了 21.14%，證明了多種資料平衡策略的關鍵作用。其次，98.85%的召回率表現顯著降低了漏檢率，對於入侵偵測系統的實用性具有重要意義。實驗結果顯示，與現有研究相比，本研究在相同 CIC-IoT-2023 資料集上展現出顯著優勢，這些資料證實了本研究提出的技術組合在

表六：與使用 CIC-IoT-2023 建模的 NIDS 文獻進行比較

研究	模型	分類數量	資料平衡策略	**Precision**$_{macro}$ (%)	**Recall**$_{macro}$ (%)	**Accuracy**$_{micro}$ (%)
[4]	RF	8	-	91.74	70.57	99.44
[14]	Transformers	8	-	91.89	74.47	99.40
[16]	SMOTE-ENN +SFMI +PCA+ RF	7	SMOTE-ENN	-	85.64	-
[17]	Clustering+MLP	8	Gaining-Sharing Knowledge	62.87	60.76	97.46
[22]	LSTM	34	-	75.97	67.46	98.75
[25]	DL-BiLSTM	8		61.34	76.23	91.13
本研究	TCN + Attn	30	1.欠取樣 2.類別權重 3.加權隨機取樣	97.11	98.85	99.54

IoT 多類別攻擊偵測上的有效性。最後，99.54%的整體準確率結合 30 類細分識別能力，證明本研究所提方案具備多類別攻擊偵測的穩健框架，為 IoT 攻擊偵測提供正確的解決方向。

陸、結論與未來展望

本研究成功實現了三個主要研究目標，並在 CIC-IoT-2023 資料集上取得顯著成果。通過構建並比較三種整合注意力機制的深度學習模型 (GRU、LSTM、TCN)，本研究在 30 類 IoT 攻擊偵測任務上展現了優異效能，在多類別的召回率有顯著提升，且各項指標都超越其他文獻。統一的實驗框架系統性地比較了三種深度學習架構在 IoT 攻擊偵測上的表現。實驗結果揭示了各模型的性能特徵：TCN 模型在整體準確率上表現最好；GRU 模型以最少參數量達到相近的準確率，展現極佳的計算效率，其輕量化的特性較適合進行部屬測試，甚至實際應用於邊緣運算；本研究設計的統一注意力機制成功整合到三種不同的模型架構中，實驗結果證實了其有效性，所有配備注意力機制的模型均達到 99% 以上的 Micro-average 召回率，顯著優於未使用注意力機制的基線模型。注意力機制透過自適應權重分配，有效提升了模型對關鍵時序特徵的學習能力，特別是在處理複雜攻擊模式時表現突出。

基於本研究的實驗結果，未來工作可從以下這幾個方面進行優化。首先，面對部分類別精確率仍有待改進的問題，可得知有些樣本較多的類別並未被完全習得特徵，導致被誤判成少數類。因此可探索更有效的取樣方法，使其特徵被完全習得；其次，未來工作應擴展到多個資料集的交叉驗證，進行跨資料集的模型遷移和性能評估；同時，考慮在真實 IoT 部署環境中進行模型驗證，評估實驗室環境和實際應用環境間的性能差異，並據此調整模型參數和偵測策略，以推動 IoT 安全技術的實用化。

參考文獻

[1] S Sinha, “State of IoT 2024: Number of connected IoT devices growing 13% to 18.8 billion globally,” IoT Analytics, https://iot-analytics.com/number-connected-iot-devices/, Sept. 2024 (accessed Jul. 2025).

[2] Z. A. Khan and P. Herrmann, “Recent Advancements in Intrusion Detection Systems for the Internet of Things,” *Security and Communication Networks*, vol. 2019, pp. 1-19, Jul. 2019, Art. no. 4301409, doi: 10.1155/2019/4301409.

[3] S. G. Abbas, F. Hashmat, G. A. Shah, and K. Zafar, “Generic signature development for IoT botnet families,” *Forensic Science International: Digital Investigation*, vol. 38, Sept. 2021, Art. no. 301224, doi: 10.1016/j.fsidi.2021.301224.

[4] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," *Sensors*, vol. 23, no. 13, Jun. 2023, Art. no. 5941, doi: 10.3390/s23135941.

[5] OWASP Foundation, "OWASP Internet of Things Project," Open Web Application Security Project, https://owasp.org/www-project-internet-of-things/, 2018(accessed Jul. 2025).

[6] D. Mirza, "Top 10 IoT device vulnerabilities to enhance IoT security," Host Duplex, https://www.hostduplex.com/blog/top-iot-device-vulnerabilities/, Feb. 2024 (accessed Jul. 2025).

[7] Z. Chang, and S. Li, "The IoT attack surface: Threats and security solutions," Trend Micro, https://www.trendmicro.com/vinfo/us/security/news/internet-of-things/the-iot-attack-surface-threats-and-security-solutions, May. 2019 (accessed Jul. 2025).

[8] O. Yoachimik, and J. Pacheco, "Record-breaking 5.6 Tbps DDoS attack and global DDoS trends for 2024 Q4," Cloudflare, https://blog.cloudflare.com/ddos-threat-report-for-2024-q4/, 2024(accessed Jul. 2025).

[9] P. Williams, I. Kaylan Dutta, H. Daoud, M. Bayoumi, "A survey on security in internet of things with a focus on the impact of emerging technologies," *Internet of Things*, vol. 19, Aug. 2022, Art. no. 100564, doi: 10.1016/j.iot.2022.100564.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, Long Beach, California, USA, Dec. 2017, pp. 6000–6010.

[11] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791

[12] S. Hochreiter, and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735

[13] M. M. Rahman, S. A. Shakil, M. R. Mustakim, "A survey on intrusion detection system in IoT networks," *Cyber Security and Applications*, vol. 3, Dec. 2025, Art. no. 100082, doi: 10.1016/j.csa.2024.100082.

[14] S. M. Tseng, Y. Q. Wang, and Y. C. Wang, "Multi-class intrusion detection based on transformer for IoT networks using CIC-IoT-2023 dataset." *Future Internet*, vol. 16, no. 8, Aug. 2024, Art. no. 284, doi: 10.3390/fi16080284.

[15] Z. Wu, H. Zhang, P. Wang and Z. Sun, "RTIDS: A Robust Transformer-Based Approach for Intrusion Detection System," *IEEE Access*, vol. 10, pp. 64375-64387, Jun. 2022, doi: 10.1109/ACCESS.2022.3182333.

[16] A. Behera, K. S. Sagar, T K. Mishra, A. Nayyar, M. Bilal, "Enhancing DDoS detection in SDIoT through effective feature selection with SMOTE-ENN," *PLoS One*, vol. 19, no. 10, Oct. 2024, doi: 10.1371/journal.pone.0309682.

[17] H. Q. Gheni, and W. L. Al-Yaseen, "Two-step data clustering for improved intrusion detection system using CIC-IoT-2023 dataset," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 9, Sept 2024, Art. no. 100673, doi: 10.1016/j.prime.2024.100673.

[18] P. A. Doost, S. S. Moghadam, E. Khezri, A. Basem, and M. Trik, "A new intrusion detection method using ensemble classification and feature selection," *Sci Rep,* vol. 15, Apr. 2025, Art. no. 13642, doi: 10.1038/s41598-025-98604-w.

[19] M. Mulyanto, M. Faisal, S. W.Prakosa, and J.S. Leu, "Effectiveness of Focal Loss for Minority Classification in Network Intrusion Detection Systems," *Symmetry*, vol. 13, no. 1, Dec. 2020, Art. no. 4, doi: 10.3390/sym13010004.

[20] P. Sun , P. Liu, Li, L. Qi , C. X. Lu, R. Hao, J. Chen, "DL-IDS: Extracting Features Using CNN-LSTM Hybrid Network for Intrusion Detection System," *Security and Communication Networks*, vol. 2020, Art. no. 8890306, Aug. 2020, doi: 10.1155/2020/8890306.

[21] I. Akbari, M. A. Salahuddin, L. Aniva, N. Limam, R. Boutaba, B. Mathieu, S. Moteau, and S. Tuffin, "Traffic classification in an increasingly encrypted web," *Communications of the ACM*, vol. 65, no. 10, pp. 75–83, Sept. 2022, doi: 10.1145/3559439.

[22] A. I. Jony, and A. K. B. Arnob, "A long short-term memory based approach for detecting cyber attacks in IoT using CIC-IoT2023 dataset," *Journal of Edge Computing*, vol. 3, no. 1, pp. 28–42, May. 2024, doi: 10.55056/jec.648.

[23] R. Nazre, R. Budke, O. Oak, S. Sawant, and A. Joshi, "A temporal convolutional network-based approach for network intrusion detection," in *Proc. of 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)*, pp. 1–6, Nov. 2024, Kalaburagi, India, doi: 10.1109/ICIICS63763.2024.10860234.

[24] M. A. Khan, "HCRNNIDS: Hybrid Convolutional Recurrent Neural Network-Based Network Intrusion Detection System," Processes, vol. 9, no. 5, May. 2021, Art. no. 834, doi: 10.3390/pr9050834.

[25] Z. Wang, H. Chen, S. Yang, X. Luo, D. Li, and J. Wang, "A lightweight intrusion detection method for IoT based on deep learning and dynamic quantization," *PeerJ Computer Science*, vol. 9, Sept. 2023, Art. no. e1569, doi: 10.7717/peerj-cs.1569.

[26] M. L. Ali, K. Thakur, S. Schmeelk, J. Debello, and D. Dragos, "Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study," *Applied Sciences*, vol. 15, no. 4, Feb. 2025, Art. no. 1903, doi: 10.3390/app15041903.

[27] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary," *J. Artif. Int. Res*, vol. 61, no. 1, pp. 863–905, Jan. 2018, doi: 10.5555/3241691.3241712.

[28] T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, Feb. 2020, pp. 318-327, doi: 10.1109/TPAMI.2018.2858826.

[29] K. Cho, B. V. Merriënboer, D. Bahdanau, and Y. Bengio, "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches," In *Proc. of Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation(SSST-8)*, Doha, Qatar, Oct. 2014, pp. 103–111, doi:10.3115/v1/W14-4012.

[30] S. Bai, J. Z. Kolter, and V. Koltun, ''An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,'' 2018, arXiv:1803.01271.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778.

[32] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in *Proc. of Advances in Neural Information Processing Systems Workshop(NIPS)*, Dec. 2015, pp. 1–9.