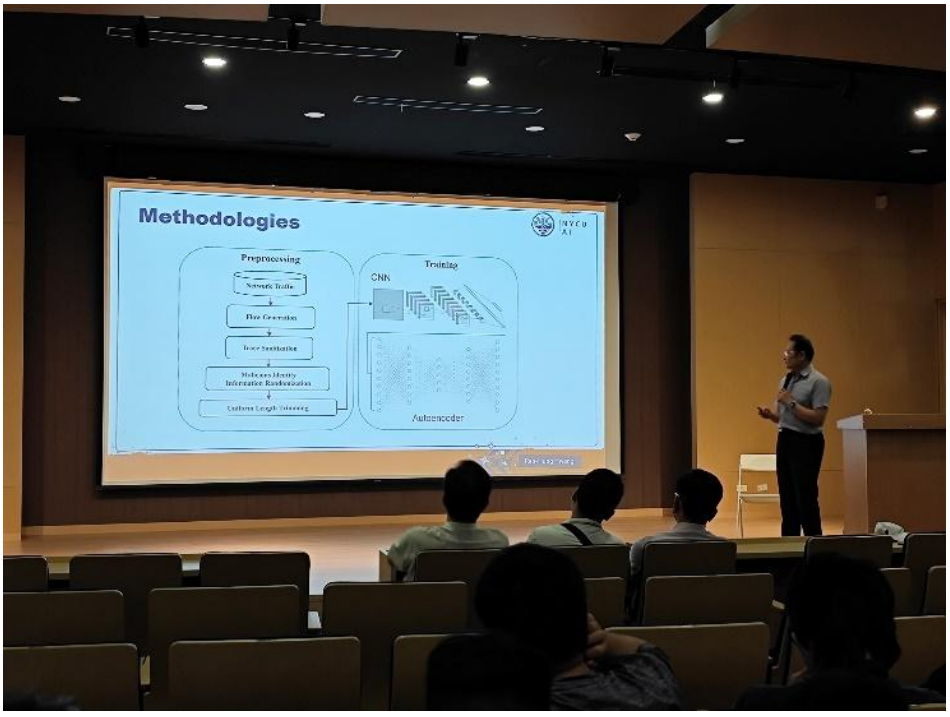
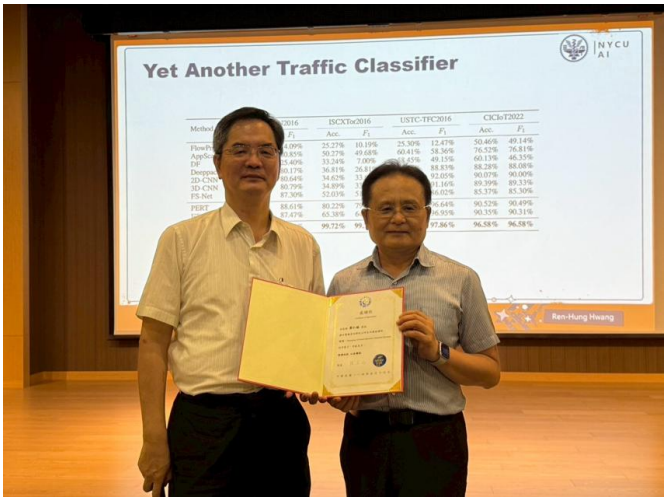
國立臺北大學 114 年度　高教深耕計畫 - 執行成果紀錄

分項計畫	1-3.3-1『深化教學研國際合作』
活動名稱	電資院大師系列講座 主題：**Emerging AI-based Intrusion Detection Systems** 講師：黃仁竑院長(國立陽明交通大學智慧科學暨綠能學院教授兼院長)
活動日期	114 年 05 月 16 日星期五
活動時間	13:30-15:00
活動地點	電資院 B1 萬榮講堂
活動內容	今天的演講，我們非常榮幸邀請到陽明交通大學的黃仁竑院長蒞臨分享。黃院長首先介紹了自己早期在「網路流量異常偵測（Network Traffic Anomaly Detection）」領域的研究經歷，主要聚焦於基於機器學習的入侵偵測系統（IDS）技術，期望能將 AI 模型應用於即時預測任務中。 他提到，當時多數學者仍以傳統機器學習方法處理異常偵測問題，因此他希望引入深度學習技術，打造能自動學習資料特徵，並具備快速偵測惡意流量能力的系統。為此，他嘗試以卷積神經網路（CNN）處理一維的流量數據，並透過模型僅觀察前幾個封包來進行判斷，這也是他長期累積的網路領域經驗所帶來的洞見。 為了進一步偵測未知攻擊（zero-day attacks），他採用了自編碼器（autoencoder）來學習正常流量的分布，並以此識別異常流量。為了讓模型能自動擷取有效特徵，他也嘗試引入 CNN 與 Transformer 架構。演講中，黃院長詳細介紹了其模型架構，其中一項關鍵策略是：在訓練模型時，刻意移除某些明顯指出惡意流量的特徵，避免模型學會「偷看答案」，而是透過真正具代表性的特徵進行判斷。 在具體實作上，他以監督式學習訓練 CNN 模型，再將 CNN 提取的特徵送入 autoencoder，學習正常流量的行為。autoencoder 的目的是讓輸出儘可能接近輸入，因此當模型遇到從未見過的異常流量時，其重建誤差會明顯變大，藉此作為惡意偵測的依據。令人印象深刻的是，儘管當時黃院長並無 GPU 資源、僅能使用 CPU 進行實驗，但憑藉簡潔的模型設計，仍達成了極為快速的偵測效果。 後來，因應 CNN 在研究上的逐漸退潮，黃院長也嘗試將 Transformer 架構應用於惡意流量偵測。他參考了 ViT（Vision Transformer）的設計，將原本用於處理二維圖像的架構調整為適合處理一維流量資料，成功替換原先 CNN 結構，並取得相當不錯的結果。 在演講後半段，黃院長拋出了一個值得深思的問題：目前大多數研究方法僅在少數幾個資料集上訓練與測試，那麼這些模型是否具備跨資料集的泛化能力呢？實驗結果顯示，單一資料集訓練出來的模型往往難以套用到其他資料集。為了解決這個問題，他進一步嘗試將多個資料集合併訓練，發現能大幅提升模型在多個資料集間的整體表現，因此萌生將「持續學習（continual learning）」應用於惡意流量偵測領域的構想。 最後，黃院長也分享了其他研究者如何將他所提出的技術延伸應用到「加密流量」的分析上，顯示其方法具備相當的靈活性與延展性。

活動成效	本次活動透過邀請黃仁竑院長進行專題演講，成功引導與會師生深入了解機器學習與深度學習技術於網路異常偵測的實務應用。黃院長以自身研究經驗出發，讓與會者認識 AI 技術如何實際應用於網路安全防護上，尤其是在即時偵測與未知攻擊（zero-day attack）上的策略與挑戰。演講深入探討了使用 CNN 與 Transformer 處理一維網路流量資料的模型架構與設計理念，有助於參與者建立實作深度學習模型時的關鍵觀念。黃院長所提出的跨資料集模型泛化能力問題，以及後續將「Continual Learning」導入惡意流量偵測的想法，引導與會者思考目前研究方法的侷限，並啟發對於更具彈性與長期學習能力模型的想像與探索。最後院長也補充其技術在加密流量分析上的延伸應用，展現所提出方法的廣泛適用性，讓與會者更具備將資安技術拓展至其他相關領域的視野與靈感。
活動照片（至少 2 張並檢附說明）	 黃院長分享自己為何想將深度學習應用到 IDS 研究的動機，並分享這項領域會使用到的 benchmark。
	 此圖展示了黃院長設計的模型架構，闡述其如何處理並學習流量資料的特徵。



活動文宣

活動紀錄：陳靚 68892