

行政院及所屬各機關出國報告
(出國類別：開會)

參加「第8屆雲端運算與大數據運算國際會議(2024 8th International Conference on Cloud and Big Data Computing)」出國報告

服務機關：財政部財政資訊中心

姓名職稱：張妤平副組長

莊榮元科長

派赴國家/地區：英國/牛津

出國期間：113年8月14日至113年8月20日

報告日期：113年11月11日

摘要

雲端運算與大數據運算國際會議自2017年開始舉辦，已成為固定召開之重要會議，每年均聚集來自世界各地優秀學者、研究者及專家共同分享研究成果。2024年會議經徵集後，共3篇主題演講(Keynote Speech)及20篇論文發表，發表的論文主題，多數採人工智慧機器學習技術，大都運用類神經網路演算法就各式複雜問題建模，並通過大量資料集對模型進行訓練及調校，以取得最佳研究結果，並廣泛運用在不同領域，包含醫療疾病輔助偵測與判斷、影像自動化判斷與處理、建築物構造力學模擬與預測等。

人工智慧建模除應研究各種人工智慧技術及應用外，更重要的是對問題領域瞭解程度，能夠準確定義問題，提出資料參數及特徵，才能有效進行資料前處理流程，並擇選最適演算法建模、訓練與解釋。因此，除資訊技術外，還需與問題領域專業人士合作，如預測牆體淨有效風壓須與具土木及資訊背景作者合作，建立桌球動作資料庫須與桌球專業人士合作等。

本次會議發表的論文，對於人工智慧模型訓練資料的收集與預處理流程，均適當採用演算法或模型，如預測牆體淨有效風壓使用MIDAS NFX軟體透過CFD模型模擬產生訓練資料；自建桌球發球及接球動作辨識模型自動捕捉大部分發球動作片段作為訓練資料，有效提高訓練資料準備效率，提供與會者未來對於資料收集及前處理新的思考方向。

目錄

壹、 目的.....	1
貳、 過程.....	2
一、 地點.....	2
二、 會議簡介.....	3
三、 論文分享.....	6
參、 心得與建議.....	11
附錄.....	13

壹、目的

隨著新興科技發展，網際網路蓬勃及硬體效能大躍進，大量資料流通與儲存成為可能，如何運用大量儲存的資料，從中分析歸納，找出有用資訊，成為一門顯學。鑑於大數據發展與應用正改變著各行各業，帶來新的契機與挑戰，並普遍應用在工程、醫學、金融、商業與政府等不同領域，如何透過大數據分析與管理，獲得更深入洞察、做出更佳決策，是項重要課題。

財政部財政資訊中心保有龐大及豐富的財政資料，如何妥適運用資料，透過大數據分析管理，獲得更深入洞察，從中發掘有助施政資訊，進而輔助更佳決策，是值得探討的議題。

雲端運算與大數據運算國際會議自2017年開始舉辦，已成為固定召開之重要會議，每年均聚集來自世界各地優秀學者、研究者及專家共同分享研究成果，本次參加會議期能更瞭解巨量資料技術發展趨勢及研究成果，作為精進管理與應用電子發票巨量資料之參考。

貳、過程

一、地點

本次會議於英國牛津布魯克斯大學黑丁頓校區(Headington Campus, Oxford Brookes University)舉行，牛津布魯克斯大學是位於英國牛津郡的公立研究型大學，其校史最早可追溯到1865年成立的牛津藝術學校(Oxford School of Art)，迄今，已成立150年以上，該校1992年由牛津理工學院(Oxford Polytechnic)改制為大學，曾被衛報(The Guardian，英國報紙)形容為最佳新型大學之一。



圖1 牛津布魯克斯大學校園

牛津郡位於英國倫敦西北方，為牛津大學(University of Oxford)英語國家中歷史最久的大學發源地，城市處處可見古老建築物散落其中，讓牛津成為許多旅客至倫敦時，會順道遊歷的地點。本次參訪期間，主要使用巴士作為代步工具，牛津的地鐵及巴士均可使用信用卡感應支付(若為交通卡並綁定手機錢包，手機無需解鎖即可感應支付，超便捷)，每日收費設上限，一旦搭乘金額達上限，同一天後續搭乘均不再收費，對外國旅客相當友善，另商店、市集亦支援信用卡小額支付，尤其是無人超商不提供現金交易機臺，完全實現無現金生活環境。此次參訪後，建議欲前往英國旅客無需換匯攜帶現金。

二、會議簡介

雲端運算與大數據運算國際會議自2017年開始舉辦，迄今(2024)年來第8屆，本會議與計算機協會(Association for Computing Machinery, ACM)合作，會議論文集由ACM在國際會議論文系列出版，並存檔於 ACM數位圖書館，本次會議於2024年8月15至17日在英國牛津布魯克斯大學舉行，會議目標是建立國際論壇，讓工程師及科學家有園地，分享他們在雲計算及大數據計算領域的想法與研究經驗，會議論文徵求主題包含，但不限於下列項目：

- ◆ 大數據算法、應用及服務
- ◆ 大數據挖掘及分析
- ◆ 大數據處理及查詢
- ◆ 雲計算及語義網技術
- ◆ 雲計算技術在大數據分析的應用
- ◆ 雲優化及自動化
- ◆ 物聯網(IoT)雲及大數據
- ◆ 雲計算及大數據軟件工程
- ◆ 軟體、硬體及演算法協同設計與高效能計算
- ◆ 大數據管理軟體及工具
- ◆ 雲計算解決方案及平臺
- ◆ 雲及大數據隱私與安全
- ◆ 大數據可視化
- ◆ 數據驅動創新、計算建模及數據整合
- ◆ 數據密集型計算理論及技術
- ◆ 虛擬化技術、雲數據管理與存儲、雲資源管理及服務質量(QoS)



圖2 ICCADC 2024主視覺

SCHEDULE AT A GLANCE

August 15th (Thursday)

Entrance, John Henry Brookes Building	
10:00-12:00 13:30-17:00	Participants Registration & Conference Kits Collection

August 16th (Friday)

Floor 2, JHB201, John Henry Brookes Building		
Host	Dr. Yanlong Zhang, Manchester Metropolitan University, UK	
9:30-9:35	Opening Remark	Prof. Hong Zhu, Oxford Brookes University, UK
9:35-9:40	Welcome Address	Prof. Huseyin Seker, Birmingham City University, UK
9:40-10:25	Keynote	Prof. Giancarlo Fortino, University of Calabria, Italy <i>Topic: Integrating Machine Learning and Multi-Agent Systems for Fully Enabling Device-Edge-Cloud Continuum in Complex IoT Worlds</i>
10:25-10:45	Coffee Break and Group Photo	
10:45-11:30	Keynote	Prof. Hong Zhu, Oxford Brookes University, UK <i>Topic: User Centric Evaluation of Large Language Models as Program Code Generation Tools</i>
11:30-12:15	Keynote	Prof. Huseyin Seker, Birmingham City University, UK <i>Topic: The Power of Data and The Things It Empowers</i>
12:15-13:30	Lunch	
13:30-15:30	Session 1 - Data Modeling and Machine Learning	
15:30-15:45	Coffee Break	
15:45-17:30	Session 2 - Image Processing and Information Security	
17:45-18:30	Fork Buffet	

August 17th (Saturday)

Activity	
10:30-12:00	Oxford Walking Tour for Voluntary Participants

圖3 會議議程

本次會議徵集後，共3篇主題演講(Keynote Speech)及20篇論文發表，並依發表論文主題區分為資料建模與機器學習(Data Modeling and Machine Learning)及影像處理與資通安全(Image Processing and Information Security)等2大議題。



圖4 大合照

三、論文分享

本次會議各論文發表者可簡報15分鐘，挑選2篇令人印象深刻的論文，擇重點分享如下：

(一)利用人工神經網路及赤池信息準則預測牆體淨有效風壓

(Prediction of Net Effective Wind Pressure in Walls using Artificial Neural Network and Akaike Information Criterion)

作者：Dante L. Silva, Kevin Lawrence M. De Jesus, Benjamin D. Flores, Russell L. Diona, Ralph Alwin M. De Jesus, Orlando P. Lopez

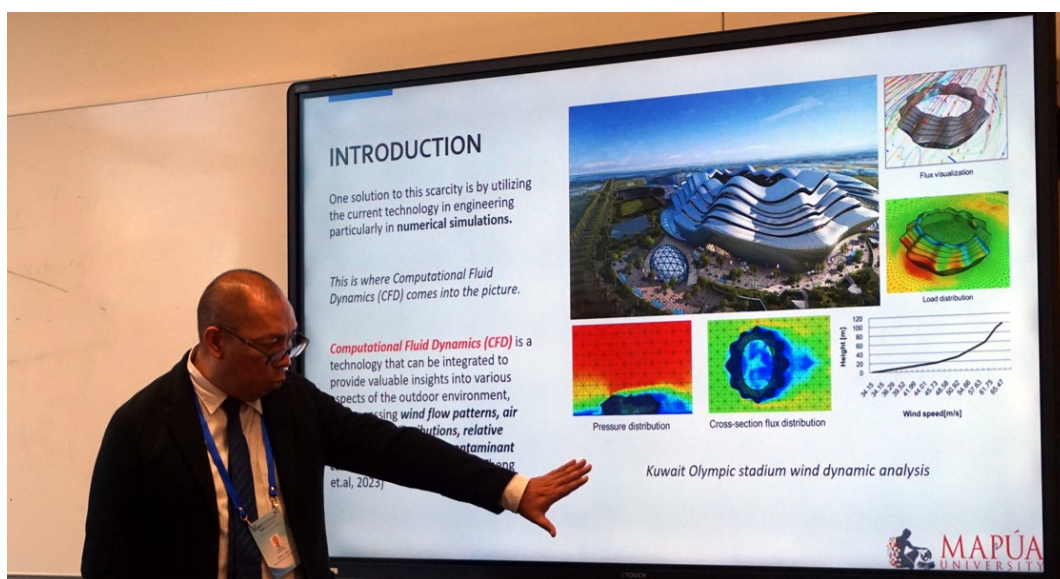


圖5 菲律賓作者分享

作者來自菲律賓，其發表考量菲律賓每年均遭受許多颱風侵襲，該研究利用計算流體動力學(CFD)及人工神經網路(ANN)機器學習技術，預測結構牆體有效淨風壓(ENWP)，以減少風力對建築物的潛在破壞。考量建築物未能承受的風力可能造成重大損壞衝擊，對經常遭遇強風的地區，準確計算有效淨風壓是設計安全結構的重要關鍵因素之一。

作者使用MIDAS NFX軟體(韓國MIDAS IT公司開發，為結構及流體分析常用之專業模擬工具)透過CFD模型(以Reynold's Averaged Navier-Stokes, RANS創建)，模擬產生ANN預測建模機器學習訓練所需資料庫。CFD模擬共生成900個資料集，其參數包括屋脊距離、圍護結構類型、表面類型、地面高度、風向、基本風速、牆面/表面開口存在及有效淨風壓。

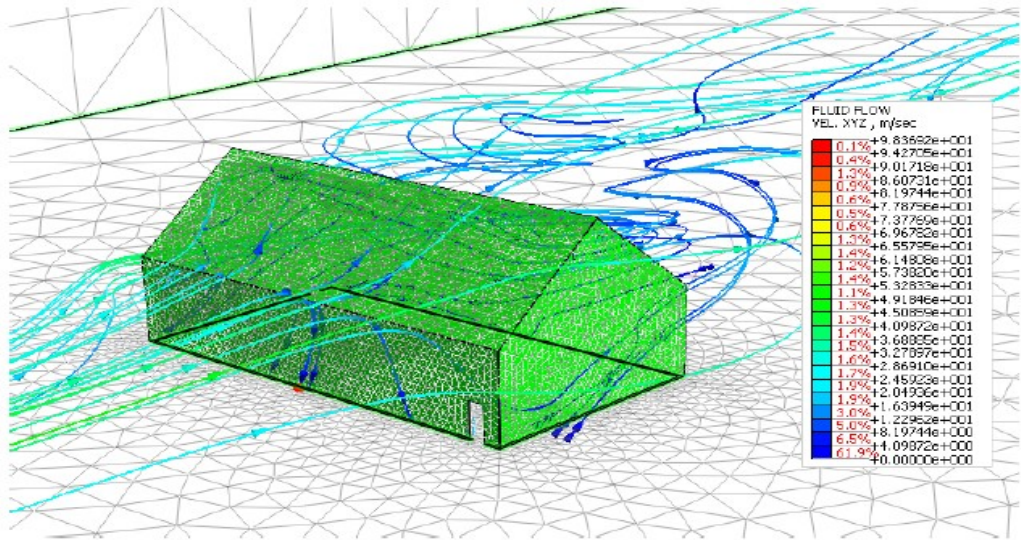


圖6 使用MIDAS NFX透過CFD模型產生的風流路徑

ANN模型開發，使用反向傳播人工神經網路(BP-ANN)，搭配Levenberg-Marquardt演算法(LMA)及雙曲正切激活函數(HTSF)，調校網路結構。經多次試驗，確定7-20-1網路結構為最佳模型，其性能指標，決定係數(R值)達0.99868；均方誤差(MSE)為0.000749；平均絕對百分比誤差(MAPE)為5.036%，顯示預測有高度精確率。

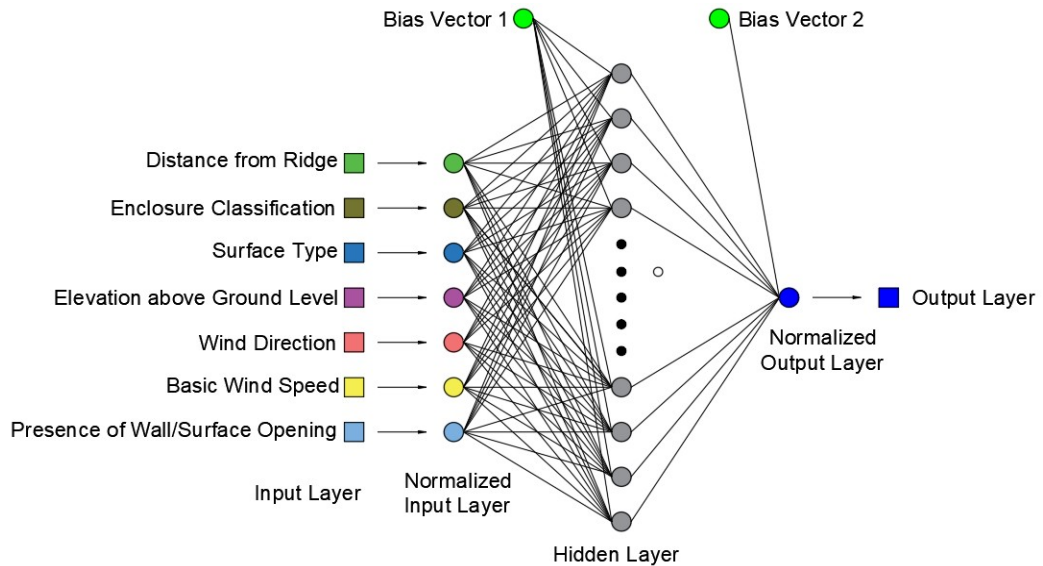


圖7 有效淨風壓模型網路架構

除基本性能評估，還使用赤池信息準則(AIC)來評估模型有效性，結果顯示7-20-1結構達最低AIC值，進一步驗證所建模型在本研究的最適性。此外，為深入分析各個輸入參數對ENWP的影響，應用Garson演算法(GA)進行敏感性分析，評估參

數相對貢獻。結果證明，基本風速對有效淨風壓的影響最為顯著，顯示其相對重要性為23.068%。

研究結果不僅強調基本風速在結構設計的關鍵作用，還開發建立可靠的預測模型，能幫助工程師在設計過程考慮風力的影響。這模型成功開發預示著，未來可進一步整合機器學習技術來提高結構設計的安全性與效率。

總結來說，本研究展示結合CFD及ANN創新方法，建立有效牆體風壓預測模型，不僅能夠在學術研究推進相關領域的發展，也對實際工程設計提供了重要的參考依據。

(二) 桌球比賽視頻基於球員動作自動分段以增強數據獲取

(Automatic Segmentation of Table Tennis Match Video Clips Based on Player Actions for Enhanced Data Acquisition)

作者：Zhong-Kai Wei, Jieh-Ren Chang



圖8 臺灣作者發表

作者來自臺灣宜蘭大學研究者，其發表利用計算機視覺及機器學習技術，實現桌球比賽影片自動剪輯任務，提高資料收集效率，減少人工觀看及手動編輯影片的人力資源需求，加快資料生成速度，確保高準確度的視頻片段辨識。本研究有效自動化勞動密集的資料準備過程，解決獲取機器學習模型訓練所需大量帶標記資料集的問題。此外，本自動剪輯框架可應用於其他動作辨識任務，提高各個領域半監督式學習模型的準確性。

本研究使用訓練數據來自自建的桌球發球及接球動作資料庫，命名為桌球動作20(TTA 20)。TTA 20資料庫係支援辨識發球及接球動作模型訓練所需創建，與

桌球專業人士合作，數據從線上平台收集，這些平台播放全球排名前10男性選手比賽完整影片。透過手動剪輯這些完整比賽視頻，提取發球及接球動作片段，每個片段根據其動作類別進行標記，作為後續動作辨識模型的訓練資料。

作者利用計算機視覺及類神經網路，訓練開發的桌球發球及接球動作辨識模型(TTPSRM)，結合人類姿態估計技術(OpenPose，開源姿態估計庫，用於人體姿勢、手勢及面部表情實時檢測與追蹤)及門控循環單元(GRU，屬循環神經網路RNN，適合處理順序資料)，準確辨識球員動作狀態，以此為基礎，找出影片所有發球及相應接球時機。

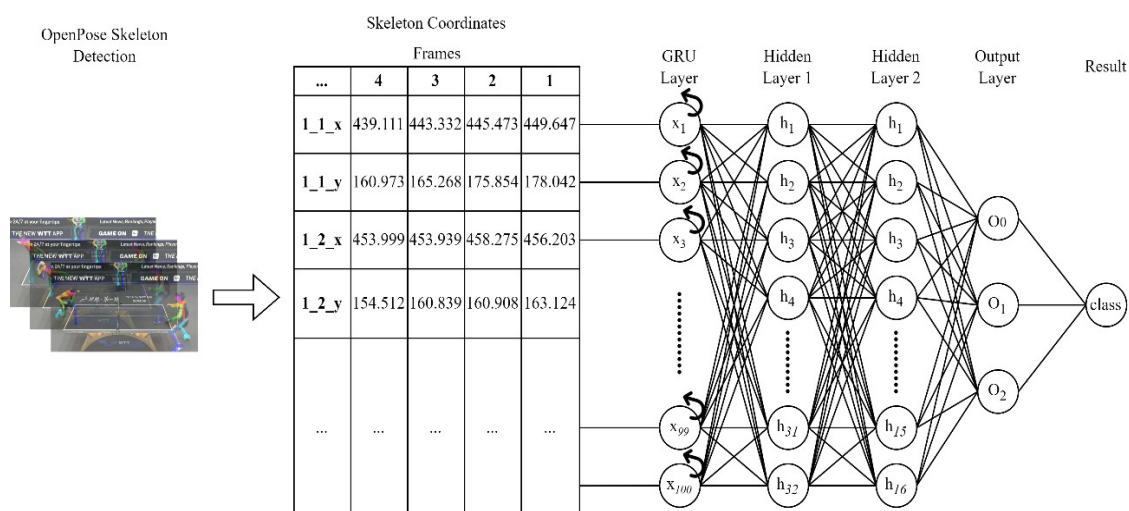


圖9 桌球比賽狀態辨識模型(TTPSRM)

為實現影片自動化剪輯任務，作者設計基於滑動視窗技術的連續動作影片辨識方法，透過將比賽視頻分割為固定大小的視窗，逐幀輸入TTPSRM進行動作分類，從而精確捕捉發球及接球時機，再以視頻各個發球時機依序找到最近接球時機，並將找到的發球與接球時間間隔門檻設定小於4秒，確保找到同一回合的發球與接發，據此實現自動化影片分割方法。

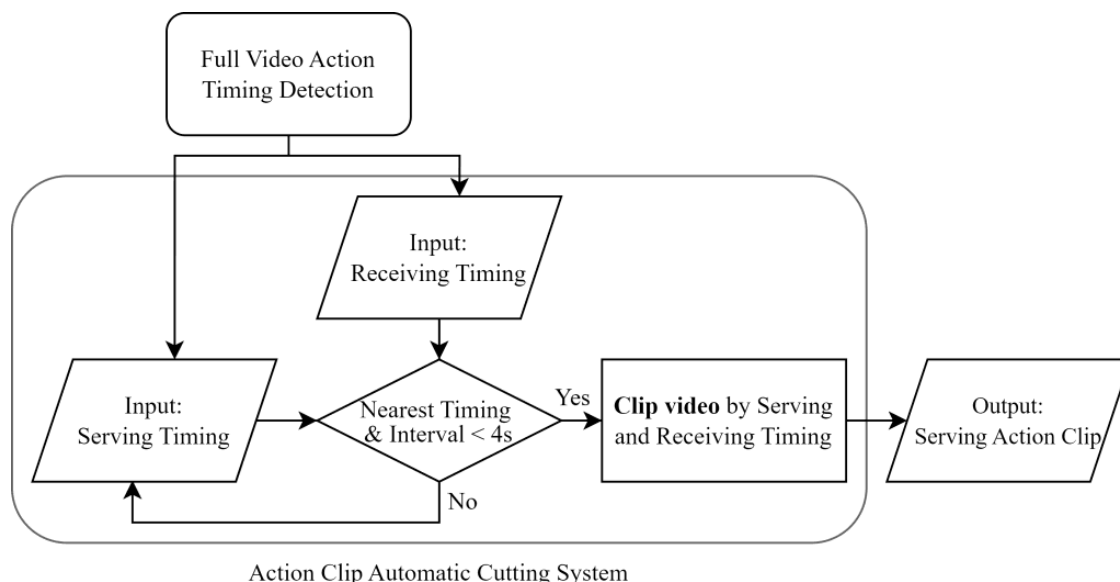


圖10 動作片段自動分割系統

桌球發球及接球動作辨識模型(TTPSRM)實證結果，在準確辨識選手動作方面有顯著效果，發球辨識整體準確率約98.4%；接球辨識準確率約96.6%，顯示TTPSRM能準確辨識比賽影片中選手狀態，提供後續系統有價值的數據。

在自動分割系統性能評估方面，使用2022年ITTF比賽影片測試，可自動剪輯70%發球動作片段，平均起始點誤差為0.547秒；結束點誤差為0.287秒，存在輕微誤差，但準確性仍然很高，與手動剪輯結果相近。雖有些發球動作片段未被分割，但系統已自動捕捉大部分發球動作片段，適合資料前處理工作流程。

本研究可快速從桌球影片蒐集大量數據，提高訓練資料準備的效率，克服目前手動方法效率低且依賴專業知識的限制，實現更快、更準確的數據收集，從而支援桌球領域更先進、更有效的研究與開發。

參、心得與建議

本次赴英國牛津參加會議，此行體驗了英國成熟信用卡支付環境，從各種想得到的交通、超市、便利商店及連鎖餐廳，到較想不到的市集、路邊小店及教堂捐款等小額支付，都可使用信用卡感應付款，實現無現金零接觸環境，本次出國行程可說完全無需使用現金，讓人開了眼界。本次會議地點在牛津郡，該郡外表充滿歷史城市建築下，內在則有現代化技術支撐，體現了英國這個工業革命起源地的底蘊。另就本次會議內容提出下列心得與建議：

一、跨領域合作交流

本次會議發表的論文主題，多數採人工智慧機器學習技術，大都運用類神經網路演算法就各式複雜問題建模，並通過大量資料集對模型進行訓練及調校，以取得最佳研究結果，並廣泛運用在不同領域，包含醫療疾病輔助偵測與判斷、影像自動化判斷與處理、建築物構造力學模擬與預測等。

建立模型除應研究各種人工智慧技術及應用外，更重要的是對問題領域瞭解程度，能夠準確定義問題，提出資料參數及特徵，才能有效進行資料前處理流程，並擇選最適演算法建模、訓練與解釋。因此，除資訊技術外，還需與問題領域專業人士合作，如牆體淨有效風壓須與具土木及資訊背景作者合作，建立桌球動作資料庫須與桌球專業人士合作。

財政部從電腦選案輔助查核，至近幾年資料科學工作坊，透過稅務專家與資訊同仁的緊密合作，對於提升查核效能已具相當成效，期許未來面對人工智慧浪潮，除賦稅領域外，亦能在跨域交流下激盪出更多加值應用與成果。

二、資料收集與運用

本次報告所提及2篇論文，對於人工智慧模型訓練資料的收集與預處理流程，均適當採用演算法或模型，如預測牆體淨有效風壓使用MIDAS NFX軟體透過CFD模型模擬產生訓練資料；自建桌球發球及接球動作辨識模型自動捕捉大部分發球動作片段作為訓練資料，有效提高訓練資料準備效率，提供未來對於資料收集及前處理新的思考方向。

人工智慧模型訓練資料的收集與預處理流程，對訓練及產出結果至關重要，財

政部財政資訊中心保有龐大及豐富的財政資料，致許多學研機構希望取得資料以為研究，然財政資料多涉機敏，基於促進政府資料透明及數位治理，以最小化原則提供其他政府機關應用外，對於民間應用處於較封閉狀態。近年透過財稅抽樣資料庫，對資料去識別化，提供外部正規申請管道，是項很好的嘗試，未來可視對外開放使用成果，研究調整提供方向及資料內容，促進公民參與，增進財稅資料運用價值。

附錄

論文列表

項次	題目 作者
1	具有RID規則的雙服務器排隊系統最佳化分析 Optimality Analysis of a Two-Server Queuing System with the RID Discipline <i>Xian Liu, Changcheng Huang</i>
2	使用預測分類優化Hadoop分佈式文件系統複製策略 Optimizing Hadoop Distributed File System Replication Policies with Predictive Categorization <i>Nada A. Zayed, Yasmine N. M. Saleh, Ahmed A. Aboelfarag, Mohamed A. Shaheen</i>
3	通過遷移多任務學習評估預訓練模型在黑色素瘤分類的表現 Assessing Pretrained Model Through Transfer Multi-Task Learn For Melanoma Classification <i>Hongyuan Xie, Yanlong Zhang</i>
4	RECAST：用於特定項目捕捉實際生產過程的開源平台 RECAST: An Open Source Platform for Item-Specific Capturing of Real Production Processes <i>Tim Köhler, Thomas Fraunholz, Dennis Rall, Lars Larsen, Dominik Görick, Alfons Schuster</i>
5	純度：衡量數據集中化質量新維度 Purity: a New Dimension for Measuring Data Centralization Quality <i>Lander Bonilla, María José López Osa, Josu Diaz-de-Arcaya, Ana I. Torre-Bastida, Aitor Almeida</i>
6	通過傳輸層安全確保雲基礎設施軟體定義網絡安全 Securing software defined networking in cloud infrastructure through transport layer security <i>Majid Jalali, Martin J Reed</i>
7	用於預測及理解兒童肥胖可解釋RNN：來自英國千禧年隊列研究的案例 Interpretable RNN for Prediction & Understanding of Childhood Obesity: A Scenario from the UK Millennium Cohort Study <i>Hissam Tawfik, Balbir Singh, Tarek Khater</i>
8	基於區塊鏈醫學影像數據安全存儲與跨域共享機制

項次	<p style="text-align: center;">題目 作者</p>
	<p>Blockchain-based Secure Storage and Cross-domain Sharing Mechanism for Medical Image Data <i>Chuanjia Yao, Rong Jiang, Leijin Long, Jianhang Dong, Chenguang Wang</i></p>
9	<p>利用工作負載預測進行多租戶併行DBMS查詢優化 Leveraging Workload Prediction for Query Optimization in Multi-Tenant Parallel DBMSs <i>Mira El Danaoui, Shaoyi Yin, Abdelkader Hameurlain, Franck Morvan</i></p>
10	<p>使用機器學習進行癡呆症惡化預測 Dementia Deterioration Prediction Using Machine Learning <i>Layla Dawood Alwardoud, Hissam Mouayad Tawfik, Sohaib Majzoub</i></p>
11	<p>Google Cloud部署 BigQuery ML 及 Vertex AI 模型可用性、性能及成本效益分析 Analyzing the Usability, Performance, and Cost-Efficiency of Deploying ML Models on BigQuery ML and Vertex AI in Google Cloud <i>Hongyu Wang, Jeong Yang, Gongbo Liang, Young Lee, Zechun Cao</i></p>
12	<p>利用人工神經網路及赤池信息準則預測牆體淨有效風壓 Prediction of Net Effective Wind Pressure in Walls using Artificial Neural Network and Akaike Information Criterion <i>Dante L. Silva, Kevin Lawrence M. De Jesus, Benjamin D. Flores, Russell L. Diona, Ralph Alwin M. De Jesus, Orlando P. Lopez</i></p>
13	<p>結合聯邦學習及差異隱私進行智能電網的安全異常檢測 Integrating Federated Learning and Differential Privacy for Secure Anomaly Detection in Smart Grids <i>Mohammadreza Mohammadi, Rakesh Shrestha, Sima Sinaei</i></p>
14	<p>使用FinBERT-LSTM及新聞情緒分析預測股票價格 Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis <i>Wenjun Gu, Yihao Zhong, Shizun Li, Changsong Wei, Liting Dong, Zhuoyue Wang, Chao Yan</i></p>
15	<p>肥胖及糖尿病前期對sEMG信號分類的影響 Obesity and pre-diabetic influence on the sEMG signal classification <i>Mohamed Rashed Al-Mulla</i></p>
16	<p>基於瓦瑟斯坦距離的非局部能量分割模型 Non-local energy segmentation model based on Wasserstein distance</p>

項次	<p style="text-align: center;">題目 作者</p>
	<p style="text-align: center;"><i>Qiang Hou, Shangcan Liu, Mingling Ou</i></p>
17	<p>學習自動指導參考提取及挖掘研究 Research on automatic guidelines reference extraction and mining based on representation learning <i>Huijie Han, Shiqi Guo, Xiaocui Gong and Xinying An</i></p>
18	<p>客制化RISC-V安全且節能的邊緣計算平台 Secure and Energy-Efficient Edge Computing Platform with Customized RISC-V <i>Cuong Pham-Quoc, Nguyen The Binh</i></p>
19	<p>eKYC過程應用異常檢測及Meijering濾波器檢測再拍攝身分文件調查 An investigation into the application of Anomaly Detection and the Meijering filter in the eKYC process to detect recaptured identity documents <i>John Magee, Stephen Sheridan, Christina Thorpe</i></p>
20	<p>桌球比賽視頻基於球員動作自動分段以增強數據獲取 Automatic Segmentation of Table Tennis Match Video Clips Based on Player Actions for Enhanced Data Acquisition <i>Zhong-Kai Wei, Jieh-Ren Chang</i></p>