

人工智能能看到什麼？ 人工智慧於胸部X光的應用進展

三軍總醫院 家庭暨社區醫學部 廖浩鈞 方文輝

學術專論：人工智能能看到什麼？人工智慧於胸部X光的應用進展

胸部X光是臨床上最重要的工具之一，不論是常規性的健康檢查，乃至具有高盛行率的肺部疾病（如肺炎、結核、肺癌等），皆須要使用成本低且輻射暴露少的胸部X光。然而，由於解剖結構在投影方向的重疊，加之解析度上的限制，解讀胸部X光片可能具有挑戰性。這種影響使得在特定位置檢測細微異常變得困難¹。因此，隨著近期深度學習技術逐漸成熟，人們希冀人工智慧能協助人類對胸部X光進行影像分析，尤其在COVID-19疫情爆發後，這一需求更加迫切，由於疫情失控和病例數的增加，人們試圖找到一種準確且簡單的方法來診斷COVID-19，胸部X光檢查因其低傷害性和相對快速而頗受青睞。然而，因為COVID-19本身較難以藉由單純的胸部X光去跟一般肺炎做區分及放射科醫生需要時間來判讀胸部放射線影像，更需要人工智慧的輔助去避免判斷出錯²。本文將簡單介紹人工智慧及目前其於數個重要疾病的研究進展及臨床應用。

深度學習

隨著計算處理能力的提升，人工智慧有能力在影像辨識中應用了深度學習技術，架構出卷積神經網路(Convolutional neural network)，這是一種專為影像數據設計的人工神經網絡，突破了人類處理非結構化資料的限制，通過自主學習的方式達成無監督學習³。舉例來說，傳統的人工智慧系統就像一位醫學生，需在考試前先聽老師講解如何解讀胸部X光，再依據老師的思路進行判讀，其能力受限於老師的授課水準。然而，擁有深度學習能力的人工智慧

無需教導，只需大量影像資料和對應的診斷結果作為“考古題”，便能建構出其獨特的判讀思路。由於電腦擅長處理龐大的資料量，人工智慧並不會忽略胸部X光的細微改變，深入地分析大量影像數據，甚至揭示出人類尚未察覺的潛在細節，進而達成更好的結果。此外，人工智慧在進行重複性任務時能保持穩定一致的標準，上述特性使其在胸部X光影像判讀中表現尤為顯著。

人工智慧在胸部X光影像應用的現況

在2020年新冠疫情期間，人工智慧的應用在COVID-19預測中得到了廣泛的研究關注，並展現了其強大的診斷輔助潛力。例如，2021年，Des等人利用多個深度學習模型對COVID-19的影像進行診斷，其研究結果顯示模型的敏感度大致在95%左右，而特異度則達到約88%，顯示出深度學習在COVID-19診斷上的顯著成效⁴。後續的研究也揭示了人工智慧判斷胸部X光的視角，展現其在判斷病灶上的準確性⁵。

除了針對新冠肺炎的應用外，其他研究也證實了人工智慧在胸部X光分析中的廣泛潛力。一項來自韓國的隨機交叉研究中，研究者邀請了六位放射科醫師參與，針對肺結節、肺實質病變、間質混濁、胸腔積液和氣胸等常見胸部病變進行判讀。在使用人工智慧輔助後，醫師的判讀效能以AUC（Area Under the Curve，表示分類模型的整體性能）從原先的0.93提升至0.98，並且顯著縮短了判讀時間，顯示人工智能輔助在提升診斷準確性與效率上



圖一 研究者使用可視化神經網路(Grad-CAM)將人工智慧權重較高的部位標示出⁵

的潛力⁶。

此外，其他研究亦針對胸部X光中常見的診斷詞條進行預測模型的開發與驗證，如肺炎、結核病(TB)、氣胸、心臟肥大、肺積水、肺腫瘤、性別等。這些研究顯示，人工智慧在多種病理狀態的自動識別上均展現了高水平的準確性，進一步佐證了其在醫學影像分析中可能的廣泛應用潛力⁷。

在上述的研究中，最被關注的便是肺部癌症，相較於電腦斷層已經被多重運用在治療上（如被用於非侵入性地搜尋某些可能影響治療選擇的突變），擁有較高使用率的胸部X光通常被專注使用在篩檢出潛在的案例⁸。近期，韓國一項針對健康檢查人群的隨機對照試驗，評估了人工智能輔助技術在檢測Lung-RADS（Lung Imaging Reporting and Data System，肺部影像報告與數據系統）4類肺結節的效果。該研究納入了10,476名受試者，隨機分配至人工智能輔助組與非人工智能組，結果顯示人工智能輔助組中放射科醫師檢出的Lung-RADS 4類肺結節達到0.59%，比例顯著高於

非人工智能組的0.25%，相對風險為2.4($P = 0.008$)。此外，人工智能輔助組在檢出惡性肺結節方面的表現也優於非人工智能組，達0.15%，而非人工智能組則未檢出任何惡性肺結節($P = 0.008$)。此研究結果顯示，人工智能輔助技術在低風險健康人群篩檢中具有提高檢出率的效果，尤其在早期肺結節的偵測上表現突出。當然，該研究有提到因Lung-RADS 4類結節的低盛行率(3.8%)及樣本數不足，因此該研究結果不算嚴謹。⁹⁻¹⁰縱然如此，這些結果進一步證實了人工智能輔助技術在提高早期肺結節檢出率及減少漏診風險方面的價值，顯示其在肺部癌症篩查中具備潛在應用前景。

除了這些常見的肺部疾病，令人驚訝的是，其在心血管疾病方面也有一定程度的運用，在新英格蘭醫學期刊2024年的評論專欄討論中，一項研究探討了單次胸部X光檢查是否可以與傳統的風險評估分數一樣，有效地預測10年內發生主要不良心血管事件(major adverse cardiovascular events, MACE)的風險。研究結果顯示，單次胸部X光篩檢可顯著提升

MACE的預測能力，10年內發生 MACE 的風險增高（相對風險為1.73，95%信賴區間為1.47至2.03）。在另外一組包含2,132名已知動脈粥狀硬化性心血管疾病風險(ASCVD)的門診患者中，胸部X光心血管疾病風險模型對MACE的預測效能優於傳統的ASCVD 風險評估模型（相對風險為1.88，信賴區間為1.24至2.85）¹¹。

此外，有研究發現，利用胸部X光進行年齡預測時，有時會出現預測年齡高於實際年齡的情況。進一步分析顯示，當人工智慧預測的“生理年齡”高於實際年齡時，這些患者的心臟衰竭等心血管臨床結局較差。經過對人工智慧模型的影像關注區域進行回顧，發現模型將注意力集中於縱隔腔、心臟輪廓和主動脈弓等區域，暗示胸部X光預測出的生理年齡可能透露了潛在的心血管預後信息²。

進一步而言，一項基於三軍總醫院資料集的研究發現，生理年齡預測殘差較高的患者在全因死亡率、心血管相關死亡率、新發生心衰竭、新發生慢性腎臟病、新發生急性心肌梗塞和新發生中風等長期結局方面均表現出顯著的統計相關性。研究結果表明，當預測年齡超過實際年齡10歲以上時，這些患者在心因性死亡方面的風險顯著增高。內部驗證組顯示此類患者的心因性死亡率高出近八倍，而外部驗證組也有近兩倍的風險上升¹³。考慮到僅有約30%的門診患者擁有足夠的信息來計算ASCVD風險分數，利用簡單且非侵入性的胸部X光檢查來篩檢心血管高風險患者，無疑是一種具有可行性的策略¹⁴。

此外，近期另一項研究更進一步探索了人工智慧於胸部X光影像中檢測左心室收縮功能障礙方面的應用，並顯示出其對預測心血管風險的潛力。該研究開發了一個深度學習模型去預測LVEF $\leq 35\%$ 的可能性，結果顯示模型的ROC曲線下面積在內部和外部驗證組中分別達到0.888和0.867，CXR-EF $\leq 35\%$ 檢測出異常的患者在死亡率、心血管相關死亡以及新發生的心房顫動方面的風險亦顯著升高，分別顯示全因死亡率在內部驗證組的風險比為1.40(95% CI 1.15-1.71)，心血管死亡率的風險比則達到3.02(95% CI 1.84-4.98)。有趣的地方是，該研究進一步分析了初始LVEF $> 50\%$ 但被CXR-EF $\leq 35\%$ 檢測出異常的患者，發現此群患者在長期內發展成LVEF $\leq 35\%$ 的風險顯著增加。內部驗證組的風險比為3.91(95% CI 2.98-5.14)，而外部驗證組則為2.49(95% CI 1.89-3.27)，暗示著CXR-EF或許具有能力抓出部分高危險群，顯示胸部X光影像在檢測心血管風險方面的潛在應用價值¹⁵。

另外，毫無疑問的，胸部X光含有大量骨骼信息，因此在骨骼相關疾病的預測上具有一定的應用潛力。最新的研究進一步探索了人工智慧於胸部X光影像中識別骨質疏鬆症及其預測死亡風險的潛能。該研究開發的模型在內部與外部驗證組中的ROC曲線下面積分別達到0.930與0.892，顯示出其在骨質疏鬆症識別上的高度準確性。在死亡率部分，內部驗證組中，CXR-OP分組患者的全因死亡風險比為2.59(95% CI：1.83–3.67)，外部驗證組中的

HR為1.67(95% CI: 1.61–1.72)。此外，該研究將AI輔助的預測結果與傳統DXA分類進行比較，結果顯示，人工智慧模型即使在無DXA檢測的患者中，也能提供同等高效的風險預測，證明其作為一種早期骨質疏鬆症篩檢工具的潛在價值，或許能成為醫生判斷骨質疏鬆高風險人群的輔助工具。¹⁶

限制性

儘管人工智慧在胸部X光影像分析的應用中展現出顯著潛力，目前的研究仍然面臨多重挑戰，可能影響其在臨床實務中的全面推廣。首先，數據品質和偏倚問題成為一大挑戰。人工智慧模型的有效性高度依賴於高質量且多樣化的訓練數據，但現階段許多模型的數據來源局限於單一機構或特定人群，導致其泛化能力不足¹⁷。這一限制使得模型在不同醫療機構或人群中應用時可能無法保持穩定的準確性，並容易引發診斷偏倚¹⁸。此外，許多公共數據集缺乏亞洲人群的資料樣本，對於台灣或亞洲地區的應用效果仍需進一步驗證。因此，如何確保人工智慧模型在不同種族和地區的普遍性成為亟待解決的問題。

此外，由於人工智慧模型對於訓練數據的需求巨大，許多針對胸部X光應用的人工智能研究仍處於試驗階段。除了針對肺部結節檢測的研究相對較多以外，其他應用方向如心血管或其他胸部病變的人工智能輔助診斷仍缺乏大規模隨機對照試驗的支持。顯然，缺乏大規模隨機對照試驗的驗證將進一步影響臨床醫師和政府相關機構對人工智能輔助診斷技術的信

任，延緩其在實際臨床中的普及應用¹⁹。

另外，模型解釋性問題也是當前的一大挑戰。現今多數人工智慧模型屬於所謂的「黑盒子」系統，即其判讀過程對使用者不可見或難以理解。雖然一些技術，如可視化熱圖(Grad-CAM)，已被用於部分減輕此問題，使模型的判斷依據更加透明，但仍無法完全消除臨床醫師對模型透明度和可靠性的顧慮²⁰。由於模型解釋性的不足，醫師對於人工智能輔助診斷結果的信任度可能受到影響，而這也在一定程度上限制了人工智慧在胸部X光應用中的實用性。整體而言，目前由於研究證據尚不完備，人工智慧在胸部X光影像診斷中的應用仍需更多實證研究的支撐，以強化其臨床可行性。

總結

儘管人工智慧在胸部X光影像應用中面臨多項限制，諸如數據來源偏倚、模型泛化能力不足、缺乏大規模隨機對照試驗以及解釋性問題等，這些挑戰並未掩蓋其在臨床應用中的潛在價值。不可否認，人工智能在胸部X光分析中的出色表現，尤其是在回顧性研究中展示出的高準確率和診斷輔助效果，凸顯了其應用前景。

胸部X光配合人工智能輔助診斷能夠涵蓋多項臨床需求，包括偵測肺部感染如新冠肺炎、篩查肺結節、鑑別肺實質病變、間質混濁、胸腔積液、氣胸，甚至在良惡性結節的判斷及心血管疾病風險的預測中也展現了重要作用。胸部X光作為醫學影像檢查中普及率高、操作便捷的工具，結合人工智能的強大分析能

力，可能在未來對臨床診斷及治療策略產生深遠的影響，甚至有望帶來革命性的改變。

參考文獻

1. Quekel LG, Kessels AG, Goei R, et al: Detection of lung cancer on the chest radiograph: a study on observer performance. *Eur J Radiol* 2001; 39(2): 111-6.
2. Asif S, Zhao M, Tang F, et al: A deep learning-based framework for detecting COVID-19 patients using chest X-rays. *Multimed Syst* 2022; 28(4): 1495-513.
3. LeCun Y, Bengio Y, Hinton G: Deep learning. *Nature* 2015; 521(7553): 436-44.
4. Das AK, Ghosh S, Thunder S, et al: Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network. *Pattern Analysis and Applications* 2021; 24: 1111-24.
5. Ukwuoma CC, Cai D, Heyat MBB, et al: Deep learning framework for rapid and accurate respiratory COVID-19 prediction using chest X-ray images. *J King Saud Univ Comput Inf Sci* 2023; 35(7): 101596.
6. Sung J, Park S, Lee SM, et al: Added value of deep learning-based detection system for multiple major findings on chest radiographs: a randomized crossover study. *Radiology* 2021; 299(2): 450-9.
7. Çallı E, Sogancioglu E, Ginneken B, et al: Deep learning for chest X-ray analysis: A survey. *Med Image Anal* 2021; 72: 102125.
8. Chassagnon G, Margerie-Mellon CD, Vakalopoulou M, et al: Artificial intelligence in lung cancer: current applications and perspectives. *Jpn J Radiol* 2023; 41(3): 235-44.
9. Nam JG, Hwang EJ, Kim J, et al: AI Improves nodule detection on chest radiographs in a health screening population: a randomized controlled trial. *Radiology* 2023; 307(2): e221894.
10. Nam JG, Kim HJ, Lee EH, et al: Value of a deep learning-based algorithm for detecting Lung-RADS category 4 nodules on chest radiographs in a health checkup population: estimation of the sample size for a randomized controlled trial. *Eur Radiol* 2022; 32(1): 213-22.
11. Weiss J, Raghu VK, Paruchuri K, et al: Deep learning to estimate cardiovascular risk from chest radiographs : a risk prediction study. *Ann Intern Med* 2024; 177(4): 409-17.
12. Ieki H, Ito K, Saji M, et al: Deep learning-based age estimation from chest X-rays indicates cardiovascular prognosis. *Commun Med (Lond)* 2022; 2(1): 159.
13. Liao HC, Lin C, WangCH, et al: The deep learning algorithm estimates chest radiograph-based sex and age as independent risk factors

- for future cardiovascular outcomes. *Digit Health* 2023; 9: 20552076231191055.
14. Poplin R, Varadarajan AV, Blumer K, et al: Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018; 2(3): 158-64.
 15. Hsiang CW, Lin C, Liu WC, et al: Detection of left ventricular systolic dysfunction using an artificial intelligence-enabled chest X-ray. *Can J Cardiol* 2022; 38(6): 763-73.
 16. Tsai DJ, Lin C, Lin CS, et al: Artificial intelligence-enabled chest X-ray classifies osteoporosis and identifies mortality risk. *J Med Syst* 2024; 48(1): 12.
 17. Kelly CJ, Karthikesalingam A, Suleyman M, et al: Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; 17(1): 195.
 18. Beam AL, Kohane IS: Big data and machine learning in health care. *JAMA* 2018; 319(13): 1317-8.
 19. Topol EJ: High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; 25(1): 44-56.
 20. Samek W, Wiegand T, Müller KR: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv* 2017; 1708.08296. 🌐

