舌下靜脈嚴重程度的機器學習分類

Published article (2022), PMC9663216 | 兩頁中文導讀

研究任務與意義

本文希望用手機舌下靜脈照片，協助將外觀分為輕度與重度，並比較不同特徵擷取方式與機器學習分類器。研究背景是中醫的血瘀視覺判讀，但模型標籤來自醫師對影像的評分；因此它學習的是影像分級，不是獨立診斷癌症、心血管疾病或其他內在病因。其貢獻在於建立可比較的計算流程，探索降維是否能兼顧分類表現與訓練速度。

資料與標註

100位患者來自臺北慈濟醫院2017年4月至2022年3月的資料，納入年齡20–80歲，女性87位、男性13位，主要包含癌症與乾燥症患者。三位資深醫師各按V1到V4給1–4分：V1隱約可見、V2明顯、V3分支、V4隆起或串珠狀；平均≤2為mild，平均>2為severe，各50張。兩兩Cohen κ為0.845、0.930、0.879，Fleiss κ為0.885，顯示評分一致性高，但不等於生物學真值或其他疾病診斷已驗證。

影像前處理與特徵維度

手機使用強制閃光並保存RGB照片，由醫師人工裁出ROI、移除牙齒與鬍鬚等干擾，再置中、縮放128×128、min–max正規化到[0,1]。展平後有49,152個像素特徵，遠多於樣本數。人工ROI是流程的一部分，不能把結果描述為任意手機照片直接輸入就能達同樣表現。原文的裁切與色彩比較是研究中的前處理觀察，尚非外部族群的性能保證。

三種表示方式

Original直接使用像素；PCA＋SIR先以不看標籤的PCA壓縮，再用標籤分組的條件平均估計有效投影方向；CNN則由卷積、pooling與全連接層擷取1,000維中間特徵，交給Ridge或SVM等分類器。PCA保留大變異方向，主成分不相關但未必獨立；SIR估計線性投影方向，可用來描述非線性反應關係。Ridge-CNN是混合模型，不是單純以CNN最後輸出完成分類。

模型與交叉驗證

比較linear與RBF SVM、linear model、Ridge、KNN（k=5）、decision tree及random forest，使用當時scikit-learn預設參數。五折驗證每回80張訓練、20張測試；CNN再把80張分為64訓練、16驗證，五折程序重複四次。原文說CNN訓練40次迭代。正式重現時，PCA、SIR、CNN等擬合步驟應限制於訓練fold，尤其SIR不能看到測試標籤；原文未完整展示全部pipeline，故不假稱已逐項驗證或已證實資訊洩漏。

舌下靜脈嚴重程度的機器學習分類

Published article (2022), PMC9663216 | 兩頁中文導讀

分類結果與特徵的作用

原始像素下，linear model、Ridge、linear SVM的平均accuracy約84.5%–84.8%，高於KNN76.2%、random forest75.5%及decision tree67.7%。PCA＋SIR使各方法集中於約83.5%–84.5%；CNN特徵配Ridge達87.5%，配linear SVM為86.2%。CNN並非改善所有模型，例如linear model降至83.5%。這些是原文報告值，沒有在本導讀重新訓練，也不證明任何模型在所有資料或調參下最優。

Ridge-CNN的取捨

Ridge-CNN平均accuracy87.5%（95% CI83.8–91.2）、sensitivity90.5%、specificity84.5%、PPV86.5%、NPV91.3%。相較Ridge-original，敏感度提升11.0個百分點，但特異度下降5.5個百分點，accuracy提升2.8個百分點。模型較少漏判重度，也較多把輕度判重度。這是重複驗證平均，不應由百分比捏造單一100人的整數混淆矩陣；臨床用途還需衡量兩種錯誤的代價。

與醫師比較及訓練時間

另兩位比較醫師不同於三位標註者，平均accuracy90.0%；與Ridge-CNN比較的原文t-test p=0.26110。未達顯著差異不是等效或非劣性證明，且重複fold與共用影像的相依性需要適當處理。Table2報告Ridge-original訓練1.1秒、PCA＋SIR0.004秒、CNN0.027秒，後者約41倍加速；這是原文量測，並非手機診斷延遲。硬體與CNN計時範圍未充分交代，不能保證任意設備重現相同比例。

外推與研究限制

資料只有單一中心100例，女性占87%，疾病組成也集中，並依賴人工ROI與專家參照標籤。50/50的平衡資料下PPV不等於一般族群PPV：若假設敏感度0.905、特異度0.845不變，重度比例只有10%，Bayes公式給PPV約39.3%。此為教學假設，非新實測。研究尚未提供多中心前瞻驗證、四級自動分級或改善患者結果的證據，不能把原型當作已驗證的醫師替代工具。

後續方向與來源

可進一步以不同醫院、照明、疾病與性別分層進行獨立測試，公開完整模型與切分流程，並評估自動ROI、臨床決策影響及事先設定界值的醫師比較。來源：Ping-Hsun Lu、Chih-Chi Chiang、Wei-Hsuan Yu、Min-Chien Yu、Feng-Nan Hwang，Machine Learning-Based Technique for the Severity Classification of Sublingual Varices according to Traditional Chinese Medicine，Computational and Mathematical Methods in Medicine（2022），3545712，DOI:10.1155/2022/3545712，PMC9663216。