Enhancing Cancer Case Screening Efficiency through Machine Learning for Accurate Search and Intelligent Allocation

Yuan-Ting Hunag[1], Yi-Wen Shen[1*], Chia-Hung Li[2], Shu-Jung Yu[1], Yeh-Hsia Liu[1], Chi-Hua Li[1], Chih-Jen Huang[1,3]

[1]Department of Cancer Center, [2]Department of Medical Information, Kaohsiung Medical University Chung-Ho Memorial Hospital, [3]School of Medicine, Kaohsiung Medical University

*Corresponding Author: Yi-Wen Shen　　　　　　Email: iwen69228@hotmail.com

ORCID: iD http://orcid.org/0009-0005-8656-0395

Abstract

Objective: Cancer registration registries serve as the empirical foundation for improving the quality of cancer care. Unlike current methods, which rely on manual review and screening and yield only a 50.4% reporting eligibility, this study leverages machine learning and natural language processing to extract key medical record information, thus enhancing the precision in selecting cases for reporting and in classifying cancer types.

Materials and Methods: The study utilized 3,000 categorized cases from 2017 and 2018, accompanied by 21,994 medical records, imaging reports, and pathology reports from a medical center in southern Taiwan, for machine learning training. A multiclass classification model, ML.NET Multiclass Classification SDCA Maximum Entropy, was employed, and keywords were annotated for 30 types of cancer to construct a smart prediction module.

Results: The screening results were categorized into three groups: "to be reported", "not to be reported", and "suspected cases." The intelligent system achieved an average accuracy rate of 89.7% in case reporting and 89.5% in cancer-type classification.

Conclusion: This smart predictive system enhances the efficiency of cancer case screening, allowing registry staff to focus on the completeness and accuracy of data extraction. Future iterations could incorporate image and text recognition to strengthen the predictive capabilities of the system, thereby providing higher analytical value to clinical teams.

Keywords: Cancer, cancer registry, screening prediction, machine learning

利用機器學習精準搜尋並智慧分案
以提升癌登個案篩選效能

黃圓婷[1]、沈怡妏[1]*、李佳鴻[2]、游淑蓉[1]、劉曄霞[1]、李季樺[1]、黃志仁[1,3]
[1]高雄醫學大學附設中和紀念醫院癌症中心、[2]資訊室、[3]高雄醫學大學醫學系

*通訊作者：沈怡妏　　　　　　　　　　所屬單位：高雄醫學大學附設中和紀念醫院癌症中心
電子信箱：iwen69228@hotmail.com　　　聯絡地址：80756高雄市三民區自由一路100號

ORCID：iD http://orcid.org/0009-0005-8656-0395

摘要

目的：癌症登記資料庫是癌症醫療品質改善的實證根本，目前依賴著人工逐筆檢視篩選，但符合申報條件僅佔50.4%。希冀透過機器學習自然語言處理擷取病歷資訊等關鍵字，能更精準地篩選出需申報的癌症個案並同時正確分類癌別。

材料與方法：利用南部某醫學中心2017年及2018年已分類的3,000筆個案含21,994份病歷資料、影像報告及病理報告進行機器訓練學習。利用多元分類模型(ML.NET Multiclass Classification SDCA Maximum Entropy)，並依30癌別進行關鍵字標註，建立智慧系統預測模組。

結果：篩選結果分為「需申報」、「不需申報」、「疑似個案」三組。智慧系統預測個案申報平均正確率為89.7%及癌別分類平均正確率為89.5%。

結論：智慧預測系統協助癌登個案篩選以提升篩選效能，讓癌症登記師專注於摘錄資料的完整性及正確性，未來期可導入圖文辨識，強化預測系統判讀能力，提供各臨床團隊更高的分析價值。

關鍵詞：癌症、癌症登記、篩選預測、機器學習