

以資料探勘技術分析癌症篩檢的相關因素-以子宮頸癌、乳癌為例

蔡興國 吳家鈞 褚祐貝 鄭乃華 余慧妤

中國醫藥大學醫務管理學系

根據國際癌症研究機構統計，世界新的癌症病例數男女統計約一千兩百萬人。衛生署亦公佈 101 年國人十大主要死因，惡性腫瘤連續 31 年蟬連首位。美國國家癌症研究院報告，早期診斷延長了癌症病患的存活率，透過適當篩檢工具早期發現早期治療降低癌症發生率及死亡率。本研究探討影響癌症篩檢利用的相關因素，並了解這些因素之間彼此影響的情形。本研究使用國家衛生研究院「2009 年國民健康訪問暨藥物濫用調查」的資料庫，從中擷取相關樣本，以資料探勘技術的決策樹進行分析。採用可能影響癌症篩檢的相關因素包括年齡、教育程度、健康行為、健康狀態、社會人口學等變項。研究工具以資料探勘軟體 WEKA 的演算法 J48 (CART) 決策樹來進行分類預測。結果發現，影響子宮頸癌篩檢利用的重要變項依次為「年齡」、「婚姻狀況」、「有無做政府提供免費的健康檢查」；對於「年齡」介於 30 到 68.2 歲的族群，以「婚姻狀況」及「有無吸菸」為重要影響因素；68.2 歲到 78.1 歲的族群的重要影響因素，則為「有無做免費健康檢查」。對於影響乳癌篩檢利用的重要變項依次為「免費健檢」、「年齡」、「自費健檢」、「教育程度」；「年齡」介於 40 歲到 49.3 歲、有做「免費健檢」與「自費健檢」的族群中，其重要影響因素為「近期內有無看中醫門診」與「教育程度」。因此，對於子宮頸癌與乳癌篩檢的利用，應該分別從上述重要影響因素來聚焦宣導，較能提升成效。

**關鍵字：**資料探勘、決策樹、CART、癌症篩檢、子宮頸癌、乳癌