

# 由資料採礦技術探討微陣列資料於結腸癌之基因篩選

Yi-Hua Tsai (蔡依樺)

輔仁大學統計資訊學系應用統計碩士班

## 摘要

在台灣結腸癌正以每年新增九千例快速成長，一般認為生活型態的西化和結腸癌患者的快速增加有關，但有一部分的結腸癌患者是遺傳而來，即肇因於 DNA 修補基因突變。故本研究針對 NCBI 微陣列資料庫之結腸癌基因表現量來探討相關基因。

本研究利用美國國家衛生研究中心(NCBI)資料庫所提供的微陣列資料為資料來源，並利用健保資料庫結腸癌資料相呼應，並利用資料採礦技術之方法(如：決策樹、羅吉斯回歸…)，分析各種方法之準確性，進而找出影響結腸癌之顯著基因，由分析結果得知建模方法以隨機森林為最佳，其分析是否為結腸癌較為準確，並找出顯著基因，希望能提供醫學研究方面的參考。

關鍵字：微陣列、資料採礦、結腸癌