

摘要

肝臟疾病全球高度流行，也是台灣地區最主要的疾病，根據衛生福利部統計疾病及死亡人數結果得知，慢性肝病及肝硬化，排在台灣死亡原因的前十，而肝癌也排在癌症死亡人數第二。肝臟疾病在初期的症狀並不是非常明顯，等到身體有感覺不適的時候，才會發現自身以患有肝臟疾病之症狀，所以，如何建立一個在初期的血液檢查中，就可以判斷出是否患有慢性疾病的機制，減少病患的侵入性檢查，來降低醫療成本以及風險，有其必要性。因此，本研究希望透過資料探勘之技術，採用決策樹以及隨機森林找出其中要特徵值，最後再利用決策樹、隨機森林以及 KNN 演算法，將其資料進行分類，找出最佳 K 值。研究結果得知，利用決策樹及隨機森林所擷取出來的重要特徵因子為 26 個指標中的 AFP 以及 PT，並且利用三種分類方法進行分類，分類結果為 KNN 模型中的 K=3 準確率 73% 為三種分類方法中最高。本研究透過資料探勘的結果，提供給醫療機構，幫助醫療人員減少的看診時間，提高看診效率，達到更完整的醫療服務，也不要耽誤到病患的黃金救援時間。

關鍵字：慢性肝炎、肝硬化、肝癌、資料探勘