الملخص

في هذه الرسالة نتقصى ثلاث تقنيات لتنقيب البيانات في تصنيف داء السكري في قواعد طبية مختلفة: (1) خوارزمية التوليد الخلفي للشبكات العصبية (2) شجرة القرار المستند إلى خوارزمية3) C4.5) خوارزمية الجار الأقربKNN , ولتحسين خوارزميةKNN اقترحنا خوارزمية تصنيف جديدة تدعىKNNB1R التي تستند إلى أعطاء وزن لكل خاصية. من الناحية الأخرى لانجاز خوارزميات أكثر كفاءة مع دقة عالية استخدمنا عمليات أعداد البيانات قبل المعالجة مثل: عمليات التنظيف والتطبيع للبيانات, فضلا عما ذكر أعلاه تم معالجة القيم المفقودة أيضا في القاعدتين: (1) قاعدة عالميه (بيانات السكري الهندية), (2) قاعدة محليه (بيانات السكري العراقية) لتحسين النتائج المستحصلة. هذه الأطروحة تقدم أيضا أداة تطبيقية فعّالة لاكتشاف داء السكري عند المرضى وإيجاد نسبة احتمالية الإصابة به.

Abstract

In this thesis, we investigate three efficient data mining techniques for diabetes classification using: Back Propagation Neural Networks, Decision tree based on C4.5 algorithm, and K-Nearest Neighbor methods in medical databases. However, to improve the accuracy of KNN algorithm, we propose a new classification algorithm called K-NN based on One-Rule (KNNBIR) which is based on weight change. On the other hand, to achieve more efficient algorithms with high predictive accuracy, data preprocessing like data cleaning and data normalization, are implemented. Moreover, the missing values are also handled in two data sets to improve the obtained results: (i) International (Indian Diabetes) and (ii) Local (Iraqi Diabetes) datasets. this thesis also introduces an effective application tool for diabetes detection at the patients. The probability of diabetes infection is discovered by using such tool.