Cardiovascular Diseases Classification Using Support Vector Machine Algorithm with the Flow Direction Algorithm Optimization

Prepared by

Fuad AbdUlhadi Yahia

Supervisor by

Dr. Hossam Mustafa

Abstract

One of the most important factors in the success of treatments and recovery from diseases is early treatment. As a result, the relevance of related early treatment research rises, and one of these fields of research is heart disease. Cardiovascular diseases are one of the main causes of death worldwide, killing 17.9 million people each year, according to the World Health Organization. In this thesis, we propose a classification model based on an optimization algorithm to improve results in predicting whether persons of all genders and ages would develop cardiovascular disease. In previous studies, researchers have been interested in finding the best ways to classify diseases by selecting disease-related features through optimization algorithms and optimization models, and in this thesis we present the Flow Direction Algorithm algorithm (FDA)،which will select the features most

relevant to the disease in order to raise the classification accuracy rate in order to educate patients before contracting the disease, which leads to reducing the number of deaths when infection is avoided. In the prediction stage, the support vector machine (SVM) is combined with the flow direction optimization algorithm (FDA) and it is also combined with the latest version of the optimization algorithm called the Opposition Flow Directional Algorithm (OFDA). To ensure more convenient results, we have investigated both the (FDA) and (OFDA) optimization algorithms with the K-Nearest Neighbors (KNN) algorithm. The proposed algorithm's performance is assessed using accuracy, recall, and precision performance metrics. The proposed algorithm based on (SVM) and (KNN) is compared with the results of previous research that used the same three datasets, first dataset is Heart Failure Clinical Dataset (HFCD).The second dataset is called Heart Dataset (HD), the third dataset is the Heart Disease Prediction Dataset (HDPD), both datasets are obtained from the UCI machine learning repository. Which were used to predict heart disease, the results showed the best accuracy at 90.75 % when using a KNN-FDA algorithm on heart failure clinical records dataset, The result was almost equal when using the SVM-FDA algorithm, on heart disease prediction dataset, where accuracy is 90.74%.

تصنيف أمراض القلب بأستخدام خوارزمية آلة المتجهات الفائقة (SVM) المدمجة مع خوارزمية اتجاه التدفق(FDA)

إعداد

فؤاد عبد الهادي يحيى

إشراف

الدكتور حسام مصطفى

الملخص

يعتبر العلاج المبكر من أهم عوامل نجاح العلاج والشفاء من الأمراض. نتيجة لذلك، تزداد أهمية أبحاث العلاج المبكر ذات الصلة، وأحد مجالات البحث هذه هو أمراض القلب. تعد أمراض القلب والأوعية الدموية أحد الأسباب الرئيسية للوفاة في جميع أنحاء العالم، حيث تقتل 17.9 مليون شخص كل عام وفقًا لمنظمة الصحة العالمية. في هذه الرسالة، نقترح نموذج تصنيف بالاعتماد على خوارزمية التحسين، لتحسين النتائج في التنبؤ بما إذا كان الأشخاص من جميع الأجناس والأعمار سيصابون بأمراض القلب والأوعية الدموية. في الدراسات السابقة، كان الباحثون مهتمين بإيجاد أفضل الطرق لتصنيف الأمراض من خلال اختيار الميزات المتعلقة بالأمراض من خلال خوارزميات التحسين ونماذج التحسين، وفي هذه الأطروحة نقدم خوارزمية Flow Direction Algorthim (FDA)، والتي ستحدد الميزات الأكثر ذات الصلة بالمرض من أجل رفع معدل دقة التصنيف من أجل توعية المرضى قبل الإصابة بالمرض مما يؤدي إلى تقليل عدد الوفيات عند تجنب الإصابة. في مرحلة التنبؤ، يتم دمج آلة دعم المتجهات SVM مع خوارزمية

تحسين اتجاه التدفق FDA ويتم دمجها أيضًا مع إصدار اخر من خوارزمية التحسين FDA والتي تسمى خوارزمية اتجاه التدفق المعاكس OFDA لضمان نتائج أكثر ملاءمة ، قمنا بالتحقيق في كل من خوارزميات التحسين FDA و OFDA باستخدام خوارزمية K-Nearest Neighbors (KNN) .يتم تقييم أداء الخوارزمية المقترحة باستخدام مقاييس أداء الدقة والعرض والاسترجاع. تمت استعراض نتائج الخوارزمية المقترحة بناءً على خوارزميات التصنيف SVM و KNN وجراء المقارنة بينها واستخدم مجموعات البيانات الثلاثة ، الأولى هي مجموعة البيانات السريرية لفشل القلب HFCD، وتسمى مجموعة البيانات الثانية Heart Dataset (HD)، مجموعة البيانات الثالثة هي مجموعة بيانات التنبؤ بأمراض القلب HDPD ، ويتم الحصول على مجموعتي البيانات من مستودع التعلم الآلي لـ UCI والتي تم استخدامها للتنبؤ بأمراض القلب ، أظهرت النتائج أفضل دقة عند 90.75٪ عند استخدام خوارزمية KNN-FDA اما دقة عند استخدام خوارزمية SVM-FDA على مجموعة بيانات السجلات السريرية لفشل القلب بلغت الدقة 90.74٪ وكانت متساوية تقريبًا.