Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Sciences
and Mathematics
Department of Computer Sciences

HEART ATTACK DETECTION USING INTELLIGENT TECHNIQUES

A Dissertation Submitted to the Council of the College of Computer Sciences and Mathematics
University of Mosul
As a partial Fulfillment for the Requirements
for the Degree of Doctor of Philosophy
in Computer Science

By

Omar Shakir Hasan Sofi Ali

Supervised by

Prof. Dr. Ibrahim Ahmed Saleh

2022 A.D. **1444 A.H.**

ABSTRACT

Heart attacks pose a severe health threat to human society in the world. Intelligent techniques can detect the precision of the heart attack risk to identify patients at risk and risk factors effectively, take the appropriate intervention measures in time, and help medical institutions improve resource customization. The hospitals contain a large amount of data associated with diagnosing heart attacks. These data include different features of the patient, including what is related to heart attacks and what is not related. Modeling this type of data requires finding relationships, and selecting the features related to heart attacks. Data sets for heart attacks are often excessively unbalanced. When this type of data is modeled, the machine learning models are biased to the majority category in the training stage and finding hyperparameters for machine learning models to find the target group (heart attacks) is considered one of the problems facing developers when modeling.

This dissertation deals with the detection of heart attacks by studying the dataset, analyzing the importance of features, the relationships between them. The data modeling process passes through several stages. In the stage of selecting the features with influence on the results of the models and reducing bias, in addition to increasing the efficiency of the models, the swarm algorithms (PSO, BAT, BCS) have been applied. And method called a merge swarm features selection (MSFS) was proposed to select the features of influence in the three algorithms and compare the proposed method, the second stage is the stage of Detection. A set of machine learning models has been configured (Naive Bayes, Decision TREE, XGBoost), and the proposed of using a Catboost algorithm. To improve the accuracy of the model, the Hyperopt technique was proposed to find and apply hyper parameters, in the third stage, the precision scale was used for the minority category as a targeted value of the fitness function for data balance and increased accuracy of the model to Detection ict patients at risk of a heart attack which is the category with the ultimate goal of the entire study. After building and testing models using relevant performance measures of study (accuracy, F1 degree, retrieval, accuracy, confusion matrix), the comparison is made, determining the importance of proposed methods and the possibility of their application.

The study found that the proposed method of merging results of the chosen swarm in selecting features is a promising solution in the field of selecting features

and increases the accuracy of the system and that the traditional machine learning models are biased in the case of unbalanced data set and that selecting features of importance in accordance with the target category has an impact on performing models in addition to the determination of hyper parameters reduces the bias of the chosen model. The CatBoost model with Hyperopt and merge swarms for feature selecting, achieved 98% accuracy and 80% precision for the minority category.

وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم الحاسوب

كشف النوبة القلبية بأستخدام التقنيات الذكائية

أطروحة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة دكتوراه فلسفة في
علوم الحاسوب

من قبل

عمر شاكر حسن صوفي علي

بإشراف
أ.د. ابراهيم احمد صالح

١٤٤٤هـ ٢٠٢٢م

الملخص

تشكل النوبة القلبية تهديدًا صحيًا خطيرًا للمجتمع البشري في العالم. يمكن للتقنيات الذكية الكشف عن دقة مخاطر النوبة القلبية لتحديد المرضى المعرضين للخطر وعوامل الخطر بشكل فعال ، واتخاذ تدابير التدخل المناسبة في الوقت المناسب ، ومساعدة المؤسسات الطبية على تحسين تخصيص الموارد ، وهو أمر له أهمية عملية مهمة. تحتوي دور الرعاية الصحية والمستشفيات على كم هائل من البيانات المرتبطة بتشخيص النوبات القلبية تتضمن هذه البيانات سمات مختلفة للمريض منها ماهو مرتبط بالنوبات القلبية ومنها ماهو غير مرتبط. يتطلب نمذجة هذا النوع من البيانات إيجاد علاقات واختيار الميزات المتعلقة بالنوبات القلبية .ان مجموعات البيانات الخاصة بالنوبات القلبية على الغالب غير متوازنة بشكل مفرط وعند نمذجة هذا النوع من البيانات تنحاز نماذج تعلم الالة الى فئة الاغلبية في مرحلة التدريب كما ان ايجاد المعلمات الفائقة لنماذج تعلم الالة لايجاد الفئة المستهدفة (النوبات القلبية) تعتبر من المشاكل التي تواجه المطوريين عند النمذجة.

تتناول هذه الاطروحة اكتشاف النوبة القلبية من خلال دراسة مجموعات البيانات وتحليل اهمية الميزات والترابط مابينها. تمر عملية نمذجة البيانات بعدة مراحل ففي مرحلة تحديد السمات ذات التاثير على نتائج النماذج وتقليل الانحياز بالاضافة زيادة كفاءة النماذج تم تطبيق خوارزميات الاسراب الثنائية (PSO, BAT, BCS) وتم اقتراح طريقة تدعى دمج الاسراب لتحديد السمات (HSFS) لاختيار السمات ذات التاثير في الخوارزميات الثلاثة ومقارنه الطريقة المقترحة. والمرحلة الثانيه هي مرحلة الاكتشاف , تم تكوين مجموعة من نماذج تعلم الالة (naive Bayes, Decision Trees, XGboost) واقتراح استخدام خوارزمية (CatBoost)، ولتحسين دقة النموذج تم اقتراح استخدام تقنية Hyperopt للعثور على المعلمات الفائقة وتطبيقها ، اما المرحلة الثالثة تم استخدام مقياس (precision) لفئة الأقلية كقيمة مستهدفة لوظيفة الملاءمة لموازنة البيانات وزيادة دقة النموذج للتنبؤ بالمرضى المعرضين لخطر الإصابة بنوبة قلبية لأنها الفئة ذات الهدف النهائي للدراسة بأكملها. بعد بناء النماذج واختبارها باستخدام مقاييس الأداء ذات الصلة بهدف الدراسة (الدقة ، درجة F1 ، الاسترجاع ، الدقة ، مصفوفة الارباك) ، يتم إجراء المقارنة ، وتحديد أهمية الطرق المقترحة وإمكانية تطبيقها.

توصلت الدراسة الى ان الطريقة المقترحة في دمج نتائج خورازميات الاسراب المختارة في تحديد الميزات يعد حلا واعدا في مجال تحديد الميزات ويزيد من دقة النظام وان نماذج تعلم الالة التقليدية تنحاز في حالة مجموعات البيانات الغير متوازنة و ان تحديد السمات ذات الاهمية وفقا للفئة المستهدفة ذات تاثير على اداء النماذج بالاضافة الى ان تحديد المعلمات الفائقة يقلل من

انحياز النموذج المختار. حقق نموذج CatBoost مع Hyperopt ومع دمج ألاسراب لاختيار الميزات ، دقة 98٪ و 80٪ لفئة الأقلية.

.