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

Prediction of Impact Covid-19 of Physical Activity Based on Ensemble Learning

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Doctor of Philosophy in
Computer Science**

By

Hanan Anas Qasim Ismaeel Aldabagh

Supervised by

Asst. Prof. Dr. Ghayda Abdulaziz Majeed Altalib

2023 A.D. **1445 A.H.**

Abstract

People around world have suffered severe panic as a result of coronavirus Disease (COVID-19). It is necessary to conduct comprehensive assessments in their psychological, physical, and social areas to understand the possible outcomes for patients who have recovered from it, and to ascertain if they have any impact from this virus on their future. Mild COVID-19 infections can take a long time to recover, especially when trying to return to normal life. This leaves patients unsure of how and when they will resume physical activity after COVID-19.

Artificial intelligence played an important role in the battle against COVID-19 virus, as it helped diagnose and predict the spread of the virus, predict deaths resulting from infection, sterilize areas, serve patients, and more. In this thesis, the effects that may occur on recoveries from COVID-19 were known and identified, and an intelligent model was created to predict these effects, in order to warn the infected people and the medical staff of dangers of continuing symptoms. This thesis made use of three groups of COVID-19 dataset: the first and second are follow-up dataset, the first is data collected in this thesis and were divided into three groups according to their follow-up period: the first six months after infection (FSM), and the second six months after infection (SSM), and more than a year after infection (MTY). The second data set is a general data set, the follow-up period is the first six months after infection only. The third data represents infection dataset that was used to ensure the quality of the proposed model.

Multiple pre-processing was performed on the datasets, then resampling was done collected and infection datasets because it is imbalanced, multi-label classification problem have been solved, then using Glowworm Swarm Optimization (GSO) algorithm to select the best features and tune the hyper-parameters, and finally predict the effects that might accrue on people recovering from COVID-19 using Hybrid Ensemble Model (HEM) , which is a hybrid model consisting of multiple Decision Tree that are generated depending on one of these models (Logistic Regression, Decision Tree, Support Vector Machine, Gaussian NB/ Bayesian Ridge, and K-Nearest Neighbor).

The performance results showed that HEM overcame traditional ML algorithms. For the first dataset that was collected, the best prediction was obtained for the impact of the renal system in the FSM group for all effects, which is AUC = 99%, and for predicting the impact of physical activity, the highest value was obtained in the SSM group, represented by AUC = 92%. As for the second dataset, the least error was obtained when predicting the impact of physical functioning, where MSE = 0.1171.

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

التنبؤ بتأثير جائحة كورونا على النشاط البدني اعتمادا على التعليم التجميعي

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

من قبل

حنان أنس قاسم اسماعيل الدباغ

بإشراف

أ.م.د. غيداء عبد العزيز مجيد الطالب

١٤٤٥هـ ٢٠٢٣م

الخلاصة

عانى الناس في جميع أنحاء العالم من ذعر شديد نتيجة انتشار فيروس كورونا (COVID-19). حيث أصبح من الضروري إجراء تقييمات شاملة في المجالات النفسية والبدنية والاجتماعية لفهم النتائج المحتملة للمرضى الذين تعافوا منه والتأكد مما إذا كان لديهم أي تأثير من هذا الفيروس على مستقبل حياتهم. قد تستغرق عدوى COVID-19 الخفيفة وقتًا طويلاً للتعافي خاصة عند محاولة العودة إلى الحياة الطبيعية وهذا يترك المرضى غير متأكدين من كيف ومتى سيستأنفون نشاطهم البدني بعد الإصابة، حيث ربما حاول بعض الأشخاص استئناف نشاطهم الطبيعي لكنهم لم يتمكنوا من ذلك.

لعب الذكاء الاصطناعي دورًا مهمًا في المعركة ضد COVID-19 حيث ساعد في التشخيص والتنبؤ بانتشار الفايروس والتنبؤ بالوفيات الناتجة عن العدوى وفي وسائل التعقيم للمناطق وخدمة المرضى وغيرها .في هذه الاطروحة، تم التعرف على التأثيرات التي قد تحدث على المتعافين من مرض COVID-19 حيث تم إنشاء نموذج ذكي للتنبؤ بهذه التأثيرات من أجل تحذير المصابين والطاقم الطبي من مخاطر استمرار الأعراض. استفادت هذه الأطروحة من ثلاث مجموعات بيانات لكوفيد ١٩: المجموعة الأولى والثانية عبارة عن مجموعة بيانات متابعة، الأولى هي البيانات التي تم جمعها في هذه الأطروحة وتم تقسيمها إلى ثلاث مجموعات وفقًا لفترة متابعتها: الأشهر الستة الأولى بعد العدوى (FSM)، والثاني ستة أشهر بعد الإصابة (SSM)، وأكثر من عام بعد الإصابة (MTY). مجموعة البيانات الثانية هي مجموعة بيانات عامة، وفترة المتابعة هي الأشهر الستة الأولى بعد الإصابة فقط. تمثل البيانات الثالثة مجموعة بيانات الإصابة التي تم استخدامها لضمان جودة النموذج المقترح. تم إجراء معالجات مسبقة على مجموعات البيانات ثم عمل إعادة موازنة (Balancing) لمجموعة البيانات المحلية وذلك لعدم توازنها وتم حل مشكلة التصنيف المتعدد (multi-label) وبعدها تم استخدام خوارزمية GSO لتحديد أفضل الميزات وضبط المعلمات الفائقة (hyper-parameters) وأخيرًا التنبؤ بالتأثيرات التي قد تحدث على الأشخاص الذين تعافو من كوفيد ١٩ باستخدام نموذج المجموعة المختلطة (HEM) ، وهو نموذج هجين يتكون من أشجار قرار متعددة يتم إنشاؤها اعتمادًا على أحد هذه النماذج (Logistic Regression، Decision Tree، Support Vector Machine، Gaussian NB / Bayesian Ridge ، و K-Nearest Neighbour).

أظهرت نتائج أداء النموذج المقترح افضلية على خوارزميات ML التقليدية، بالنسبة لمجموعة البيانات المحلية حيث تم الحصول على أفضل نتيجة تنبؤ عند التنبؤ بمدى تأثر الجهاز الكلوي في مجموعة FSM بالنسبة لجميع التأثيرات والتي كانت (AUC = 99٪)، وبالنسبة للتنبؤ بتأثر النشاط البدني تم الحصول على أعلى قيمة في مجموعة SSM وهي (AUC = 92%)، اما بالنسبة لمجموعة البيانات الثانية تم الحصول على اقل خطأ عند التنبؤ بتأثر الاداء البدني حيث كانت (MSE = 0.1171).