Centre d'Etudes Doctorales : Sciences et Techniques et Sciences Médicales

Avis de Soutenance

THESE DE DOCTORAT

Présentée par

Madame NOURA QARMICHE

Discipline : Informatique
Spécialité : Informatique

Sujet de la thèse

Approches optimisées de machine Learning pour la prédiction de la susceptibilité au cancer colorectal au Maroc

Formation Doctorale " Sciences de l'Ingénieur, Sciences Physiques, Mathématiques et Informatique"

Thèse présentée et soutenue **le vendredi 16 juin 2023 à 15h** à l'Ecole Nationale des Sciences Appliquées de Fès, devant le jury composé de :

Nom et Prénom	Titre	Etablissement	
El Habib NFAOUI	PES	Faculté des Sciences Dhar El Mehraz de Fès	Président
Moulay Youssef HADI	PES	Ecole Supérieure de Technologie de Kenitra	Rapporteur
Mohamed BERRADA	PES	Ecole Nationale des Sciences Appliquées de Fès	Rapporteur
Karima AISSAOUI	PH	Faculté des Sciences d'Oujda	Rapporteur
Karima EL RHAZI	PES	Faculté de Médecine et de Pharmacie et Médecine Dentaire de Fès	Examinateur
Nada OTMANI	Pr Agrégée	Faculté de Médecine et de Pharmacie et Médecine Dentaire de Fès	Examinateur
Younes CHIHAB	PH	Ecole Supérieure de Technologie de Kenitra	Examinateur
Nour El Houda CHAOUI	PES	Ecole Nationale des Sciences Appliquées de Fès	Directeur de thèse

Laboratoire de recherche : Laboratoire Intelligence Artificielle & Sciences des Données & Systèmes Émergents
Etablissement : Ecole Nationale des Sciences Appliquées de Fès

Centre d'Etudes Doctorales : Sciences et Techniques et Sciences Médicales

Résumé de la thèse

Dans les pays en voie de développement, le cancer colorectal (CCR) est en train de devenir un problème de santé publique majeur, avec une augmentation significative de l'incidence et de la mortalité. Les chiffres sont alarmants et nécessitent des politiques de prévention efficaces pour lutter contre cette maladie. Les facteurs les plus importants contribuant au CCR sont la nutrition et le style de vie, qui jouent un rôle clé dans les différentes étapes de la cancérogenèse colorectale. Étant donné que ces pays sont en transition épidémiologique, démographique, économique et sociale, leurs habitudes alimentaires et modes de vie sont en train de subir d'importants changements qui les orientent vers un modèle plus occidental. L'étude CCR Nutrition, qui est une étude cas témoins, multicentrique Marocaine a montré que les facteurs nutritionnels et le mode de vie sont étroitement liés à l'incidence du cancer colorectal (CCR) au Maroc. En effet, l'adoption d'un mode de vie sain peut potentiellement réduire le risque de CCR à long terme. Identifier les personnes à risque de développer un CCR permet de les orienter vers un dépistage et de les motiver à changer de comportement.

Le but principal de cette thèse était de créer et de valider un modèle prédictif de risque de cancer colorectal (CCR) basé sur l'âge et les éléments de l'indice de mode de vie sain étendu (EHLI) dans la population marocaine, en utilisant la base de données de CCR Nutrition. Pour atteindre cet objectif, treize modèles d'apprentissage automatique, à la fois simples et combinés, ont été développés pour déterminer le meilleur. Une application web a également été développée pour une utilisation plus pratique en milieu médical. Les tests ont été réalisés à l'aide d'une validation croisée à cinq échantillons. Le modèle Soft-Voting, basé sur les techniques CatBoost, LightGBM et Gradient-Boosting, a fourni des résultats améliorés, avec une précision moyenne de 0,6583 ±0,054. Grâce à cette prédiction de risque de CCR basée sur l'âge et les éléments de l'EHLI par Soft-Voting, les personnes à risque de cancer colorectal pourraient être identifiées dans la population marocaine, contribuant ainsi à améliorer la prévention en encourageant l'adoption d'un mode de vie sain.

Malgré ses avantages, le modèle Soft-Voting présente certaines limites, notamment l'utilisation de l'échelle de mesure très détaillée FFQ pour recueillir les caractéristiques alimentaires, ce qui la rend difficile à utiliser par un clinicien ou un simple utilisateur. C'est pourquoi il est important de chercher des méthodes simplifiées pour calculer les scores du FFQ. Dans cette optique, nous avons exploré les techniques d'apprentissage automatique non supervisé pour réduire la dimensionnalité du FFQ et regrouper les profils alimentaires de la population marocaine associés au CCR.

Les aliments ont été classés en 30 groupes alimentaires réduits sur six dimensions grâce à une analyse en composantes principales (ACP). En appliquant la méthode de K-means dans le sous-espace ACP, deux modèles ont été identifiés : le "modèle prudent" qui implique une consommation modérée de presque tous les aliments avec une légère augmentation des fruits et légumes, et le "modèle dangereux" qui comprend l'huile végétale, le gâteau, le chocolat, le fromage, la viande rouge, le sucre et le beurre, avec une faible variation entre les composantes et les groupes. Le test de Student a révélé une relation significative entre les groupes et toutes les consommations alimentaires, à l'exception de la volaille. En utilisant le test de régression logistique simple, nous avons constaté que les personnes appartenant au "modèle dangereux" ont un risque plus élevé de développer un CCR, avec un OR de 1,59 et un intervalle de confiance à 95 % de [1,37 ; 1,38]. Les résultats de cette étude pourraient contribuer à l'élaboration de recommandations pour une alimentation préventive du CCR dans la population marocaine.

En termes de perspectives de recherche, nous envisageons tout d'abord de développer un nouvel indice de prédiction basé, entre autres, sur le profil alimentaire plutôt que sur les détails nutritionnels. Ensuite, nous prévoyons de créer un Framework qui permettra aux utilisateurs de connaître leur profil alimentaire ainsi que leur probabilité de développer un cancer colorectal et de recevoir des recommandations personnalisées.

Mots clés : cancer colorectal, EHLI, machine learning, ensemble learning, réduction de dimensions, analyse de cluster, profil alimentaire, facteurs de risque, Maroc