

CED : « Sciences et Techniques de l'Ingénieur »

AVIS DE SOUTENANCE

Sokaina Elkhamlichi

Présentera ses travaux de recherche en vue de l'obtention du
Doctorat en Sciences et Techniques

Intitulé de la thèse :

«Contributions à l'analyse des données biomédicales : cas de cancers et de COVID-19»

<u>Date :</u>	18 marsr 2023
<u>Heure :</u>	10 Heures
<u>Lieu :</u>	Salle Othmane FILALI MEKNASSI ENSA de Tanger

Devant le jury :

<i>Membres de jury</i>		
Pr. Mohammed Reda BRITEL	ENSAT Tanger	Président
Pr. Jaouad EL HARTI	Faculté de Médecine de Rabat	Rapporteur
Pr. Mariam TANANA	ENSAT Tanger	Rapporteur
Pr. Mohcine BENNANI MECHITA	FST Tanger	Rapporteur
Pr Rajaa Chahboune	FMP de Tanger	Examinatrice
Pr. Abdelfettah SEDQUI	ENSAT Tanger	Directeur de Thèse
Pr. Amal MAURADY	FST Tanger	Co-encadrante

RESUME

Historiquement, le secteur de la santé a produit d'énormes quantités de données, motivées par l'amélioration de la qualité des soins offertes aux patients, la tenue de dossiers ainsi que la conformité et les exigences réglementaires. Bien que la majorité des données soient conservées sous format papier, la tendance récente est la numérisation rapide de ces vastes quantités de données. Poussées par les normes réglementaires et la capacité d'améliorer la qualité de la prestation des soins de santé tout en réduisant les coûts, ces énormes quantités de données ont le potentiel de prendre en charge un large éventail de tâches médicales et de soins de santé, y compris, entre autres, la surveillance des maladies, l'aide à la décision clinique et la gestion de la santé des populations. Ces quantités substantielles de données sont avantageuses pour la science des données et stimulent une croissance rapide de l'analyse des données en médecine. Cette thèse vise à analyser les données biomédicales. Premièrement, en menant une analyse descriptive explicative pour évaluer et comparer la situation du COVID-19 entre les pays à revenu intermédiaire inférieur de la région de la Méditerranée orientale. Ensuite, en appliquant une analyse prédictive, en particulier, les approches d'apprentissage automatique pour prédire l'apparition du cancer du col de l'utérus. La première étude réalisée dans cette thèse consiste à comparer et à évaluer la situation du Covid 19 entre les pays à revenu intermédiaire inférieur de la région de la Méditerranée orientale, à savoir le Maroc, la Tunisie, l'Égypte, Djibouti, le Pakistan, le Soudan et la Palestine. La technique de l'analyse de la variance (ANOVA) a été réalisée pour comparer le taux d'infection, de létalité, de guérison et de positivité dans ces pays. Les données rapportées quotidiennement des cas confirmés, des guérisons, des décès et des tests ont été utilisées dans cette étude. Les résultats illustrent les différences significatives dans l'impact du COVID-19 dans les pays susmentionnés en fonction de plusieurs facteurs, notamment les aspects démographiques et les conditions de la comorbidité. La deuxième contribution de cette thèse porte sur la construction d'un modèle d'apprentissage automatique pour prédire le cancer du col de l'utérus afin d'aider les Médecins dans le diagnostic de cette maladie à un stade précoce et de suivre le traitement approprié aux patients. L'étude compare cinq algorithmes d'apprentissage automatique, à savoir, K-plus proches voisins, Gaussien Naïve Bayes, la régression logistique, Forêts aléatoires et l'arbre de décision. L'arbre de décision dépasse tous les autres algorithmes. Ensuite, l'étude s'est poursuivie avec l'amélioration des résultats de l'arbre de décision en équilibrant les données, en sélectionnant les caractéristiques clés ainsi qu'en ajustant les hyperparamètres. Ainsi, le classifieur de l'arbre de décision combiné avec SMOTE tout en réglant les hyperparamètres avec la technique Grid Search présentent la meilleure performance. En pratique, nos contributions pourraient être très utiles dans les milieux médicaux et pourraient sauver des vies et réduire le fardeau du système de soins de santé. En fait, notre première étude énonce un certain nombre de recommandations, qui pourraient être très bénéfiques pour aider les gouvernements à prendre des décisions adéquates et à élaborer des stratégies de confinement plus efficaces et moins invasives. D'autre part, la deuxième contribution de notre thèse fournit un modèle d'apprentissage automatique précis pour prédire le cancer du col de l'utérus sur la base de données cliniques. Ce modèle peut aider les médecins à diagnostiquer précocement le cancer du col de l'utérus et à fournir des interventions appropriées et potentiellement à sauver des vies.

Mots clés: COVID-19; Nouvelle pandémie de coronavirus; Facteurs de risque; Analyse de la variance (ANOVA); Épidémiologie; pays à revenu intermédiaire de la tranche inférieure; Région de la Méditerranée orientale; Apprentissage automatique supervisé; Classification du cancer du col de l'utérus; Sélection des caractéristiques; Technique de suréchantillonnage minoritaire synthétique; Élimination de fonctionnalité récurrente; Arbre de décision.