

Article

# Détection précoce du charançon rouge du palmier, *Rhynchophorus ferrugineus* (Olivier), Infestation à l'aide du Data Mining

Heba Kurdi <sup>\*1,2</sup> , Amal Al-Aldawsari<sup>1</sup>, Isra Al-Turaiki<sup>3</sup>  et Abdulrahman S. Aldawood <sup>4</sup> 

<sup>1</sup> Département d'informatique, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Arabie saoudite ; 437203624@student.ksu.edu.sa

<sup>2</sup> Département d'ingénierie mécanique, Massachusetts Institute of Technology (MIT), Cambridge, MA 02142-1308, USA

<sup>3</sup> Département des technologies de l'information, Collège des sciences de l'informatique et de l'information, Université du Roi Saud, Riyad 11451, Arabie saoudite ; ialturaiki@ksu.edu.sa

<sup>4</sup> Département de la protection des plantes, Collège des sciences de l'alimentation et de l'agriculture, Université du Roi Saud, Riyad 11451, Arabie saoudite ; aldawood@ksu.edu.sa

\* Correspondance : hkurdi@ksu.edu.sa

[creativecommons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/).



**Citation** : Kurdi, H. ; Al-Aldawsari, A. ; Al-Turaiki, I. ; Aldawood, A.S. Early Detection of Red Palm Weevil, *Rhynchophorus ferrugineus* (Olivier), Infestation Using Data Mining. *Plants* **2021**, *10*, 95. <https://doi.org/10.3390/plants10010095>

Reçu : 23 novembre 2020

Accepté : 31 décembre 2020

Publié : 6 janvier 2021

**Note de l'éditeur** : MDPI reste neutre en ce qui concerne les claims juridictionnels dans les cartes publiées et les affiliations institutionnelles.



**Copyright** : © 2021 par les auteurs. Licensee MDPI, Bâle, Suisse. Cet article est un article en libre accès distribué selon les termes et conditions de la licence Creative Commons Attribution (CC BY) (<https://>

**Résumé :** Au cours des 30 dernières années, le charançon rouge du palmier, *Rhynchophorus ferrugineus* (Olivier), un ravageur très destructeur pour tous les types de palmiers, s'est rapidement répandu dans le monde entier. Cependant, la détection d'une infestation par le charançon est très difficile car les symptômes ne sont pas visibles avant que la mort du palmier ne soit inévitable. En outre, l'utilisation d'outils automatisés d'identification du charançon du palmier pour prédire l'infestation est compliquée par le manque d'ensembles de données sur le charançon du palmier. Dans cette étude, nous avons évalué la capacité de 10 algorithmes de classification de pointe, Naive Bayes (NB), KSTAR, AdaBoost, bagging, PART, arbre de décision J48, perceptron multicouche (MLP), machine à vecteur de support (SVM), forêt aléatoire et régression logistique, à utiliser la taille des plantes et les mesures de température collectées sur des arbres individuels pour prédire l'infestation par le

charançon du palmier à huile dans ses premiers stades, avant que des dommages significatifs ne soient causés à l'arbre. Les performances des algorithmes de classification ont été évaluées en termes d'exactitude, de précision, de rappel et de mesure F à l'aide d'un ensemble de données réelles sur les tordeuses de la pelouse. Les résultats expérimentaux ont montré qu'il est possible de prédire les infestations de TBI avec une précision allant jusqu'à 93 %, une exactitude supérieure à 87 %, un rappel égal à 100 % et une mesure F supérieure à 93 % en utilisant l'exploration de données. En outre, nous avons constaté que la température et la circonférence sont les caractéristiques les plus importantes

---

pour prédire les infestations par les TBI. Cependant, nous demandons instamment de collecter et d'agréger davantage d'ensembles de données sur les térébrants afin de mener davantage d'expériences pour valider ces résultats et fournir des conclusions plus concluantes.

**Mots-clés :** charançon rouge du palmier ; *Rhynchophorus ferrugineus* ; palmier ; infestation ; prédiction ; data mining

## 1. Introduction

Le charançon rouge du palmier (CRP), *Rhynchophorus ferrugineus* (Olivier), est un ravageur très destructeur qui affecte les espèces de palmiers dans le monde entier [1]. Le charançon rouge du palmier menace de nombreux types de palmiers, notamment le cocotier, le palmier à sucre, le palmier à huile, le sagoutier, le palmier palmyre, le palmier royal, le palmier de Washington et le dattier [2]. L'infestation par le TBI a été détectée pour la première fois au milieu des années 1980 dans la région du golfe Arabo-Persique. Depuis, elle s'est rapidement répandue dans le monde entier, atteignant le Moyen-Orient, l'Asie du Sud, l'Afrique du Nord, la Russie, l'Espagne et de nombreuses autres régions [2,3]. Il a été rapporté que jusqu'à 30 % de la production de palmiers dattiers peut être perdue à cause des ravageurs et des maladies [4] ; des millions de dollars sont dépensés pour la lutte contre les ravageurs et l'élimination des arbres infestés. Récemment, les palmiers infestés ont été brûlés afin de minimiser la propagation de la maladie et de sauver d'autres palmiers de l'infestation.

Une détection précoce de l'infestation par les tordeuses est essentielle pour un contrôle plus efficace des dommages causés par le ravageur. Cependant, l'ensemble du cycle de vie des larves de tordeuses est dissimulé à l'intérieur du palmier, ce qui empêche la détection d'une infestation à ses premiers stades [5,6]. Les méthodes actuelles de détection précoce comprennent l'inspection visuelle de l'arbre pour observer les symptômes, la détection du son produit par les larves en train de se nourrir, la détection chimique des signatures volatiles produites par les arbres infestés, la détection de la présence d'un ravageur dans l'arbre et la détection de la présence d'un ravageur dans l'arbre.

Les méthodes classiques peuvent être plus coûteuses dans certains cas (par exemple, la détection sonore), mais l'observation des symptômes est la méthode la plus fiable pour détecter les tordeuses de la pomme de terre. Les méthodes classiques peuvent être dans certains cas plus coûteuses (par exemple, la détection sonore) ; cependant, l'observation des symptômes est la méthode la plus fiable pour détecter les tordeuses de la pomme de terre. Cependant, comme les symptômes ne se manifestent qu'à un stade avancé, la détection et la prédiction des infestations de TBI à un stade précoce restent un défi important. C'est pourquoi de nombreux chercheurs de différents domaines ont été motivés pour travailler dans ce domaine. Plutôt que de s'appuyer sur les méthodes de détection traditionnelles, il est nécessaire de développer des méthodes informatiques fiables et durables pour détecter et gérer les populations de tordeuses de la grappe afin de réduire les pertes potentielles de palmiers. Il est également nécessaire d'identifier les caractéristiques clés qui sont pertinentes pour l'identification des palmiers infestés.

L'exploration de données est le processus d'identification de modèles et de relations cachés dans de grandes quantités de données [8]. Le potentiel des techniques d'exploration de données a été démontré dans de nombreux domaines, tels que les soins de santé, les affaires et l'éducation. Les techniques d'exploration de données ont également été appliquées à des problèmes agricoles, avec des résultats montrant un potentiel élevé [9-11]. Cependant, à notre connaissance, aucun des travaux antérieurs n'a étudié le RPW, ce qui peut être attribué à la non-disponibilité d'ensembles de données publiés. Dans cet article, nous avons évalué la capacité de dix algorithmes de classification de pointe, Naive Bayes (NB) [8], KSTAR [12], AdaBoost [13], bagging [14], PART [15], arbre de décision J48 [16], perceptron multicouche (MLP) [17], SVM [18], forêt aléatoire [19], et régression logistique [20], et la régression logistique [20], qui a été construit sur la base de données réelles collectées en Arabie Saoudite [21] dans le cadre d'un vaste projet national visant à réduire les pertes de palmiers en permettant l'administration des traitements et des injections disponibles avant que les pertes ne se produisent. Le reste du présent document est organisé comme suit : la littérature connexe est examinée à la section 2. La section 3 décrit les matériaux et les méthodes utilisés, tandis que la section 4 présente et discute les résultats expérimentaux. Enfin, les conclusions de l'article sont énoncées à la section 5.

## 2. Revue de la littérature

Les techniques d'exploration de données ont été appliquées avec succès dans un large éventail de domaines, tels que les soins de santé [22], les affaires [23] et l'éducation [24]. Dans les domaines liés à l'agriculture, Ref. [25] a présenté l'une des premières études sur l'utilisation des techniques d'exploration de données pour résoudre les problèmes agricoles. L'exploration de données a montré son potentiel lorsqu'elle est appliquée à la prédiction de la fermentation du vin, à l'estimation des paramètres de l'eau du sol, à la reconnaissance des bruits d'animaux et à la détection des farines de viande et d'os. Les auteurs ont mis en évidence l'utilisation de quatre techniques d'exploration de données : les réseaux neuronaux artificiels (ANN), les SVM, les k-voisins les plus proches (KNN) et les k-moyennes.

Pour la prédiction des maladies parasitaires, [26] a présenté un système qui utilise des techniques d'exploration de données et des réseaux de capteurs sans fil. L'objectif était de déterminer les corrélations entre les conditions météorologiques, les parasites et les maladies. L'ensemble de données utilisé dans leur étude représentait des données en temps réel collectées à l'aide du GPRS au cours de quatre saisons. La classification et l'analyse des corrélations ont été effectuées à l'aide de l'algorithme Gaussian Naive Bayes. Les résultats ont révélé que le modèle cumulatif était plus précis que le modèle empirique en raison de sa proximité avec les données au niveau du sol. En particulier, il a permis de protéger les cultures contre les ravageurs et les maladies et d'améliorer les rendements.

Les chercheurs de [27] ont utilisé des algorithmes d'arbres de décision pour prédire la productivité des cultures. Pour l'étude, un ensemble de données sur les cultures a été obtenu auprès de l'Association économique éthiopienne. Trois algorithmes d'arbre de décision ont été utilisés : J48, REPTree et forêt aléatoire. Les auteurs ont conclu que l'attribut "utilisation d'engrais" présentait le pouvoir prédictif le

plus élevé. En outre, REPTree a surpassé les autres classificateurs et a été suivi par la forêt aléatoire.

L'exploration de données a également été utilisée pour faciliter les décisions relatives à la protection des cultures de kiwis en Nouvelle-Zélande. Les chercheurs de [28] ont présenté un modèle utilisant les algorithmes d'exploration de données suivants : arbre de décision, Naive Bayes, forêt aléatoire, AdaBoost, SVM et régression logistique. Les classificateurs ont été appliqués à des journaux de pulvérisation et à des ensembles de données de surveillance des parasites. Les performances de ces modèles ont été évaluées en termes de précision et de rappel. Les résultats ont révélé que

que les modèles comportant quelques attributs fournissent des prédictions efficaces. En outre, AdaBoost a surpassé les autres classificateurs et a été suivi par le classificateur Naive Bayes.

Dans [29], un modèle de prévision des maladies du riz a été développé. Le modèle incrémental proposé était basé sur la technique d'optimisation par essaims de particules (PSO) et l'extraction de règles d'association. Il a été conçu pour déterminer un nombre optimal de règles de classification. La précision de la classification a été évaluée sur un ensemble de données de test et comparée à celle d'autres méthodes de classification de pointe, telles que Naive Bayes, KSTAR, AdaBoost, bagging, PART, J48, MLP, SVM, forêt aléatoire, régression logistique et PSO. Le modèle incrémental a obtenu la plus grande précision dans trois des cinq ensembles de données. La méthode proposée a permis de réduire le temps de calcul et a démontré son importance et son efficacité.

Pour la détection précoce des maladies pathogènes des cerises, [30] a proposé un système utilisant quatre techniques d'exploration de données : un modèle discriminant linéaire, un modèle discriminant quadratique, un modèle discriminant pseudo-linéaire et un modèle d'arbre de classification compact (CCT). La précision des modèles a été mesurée et une comparaison entre les modèles a été présentée pour sélectionner le meilleur modèle. L'ensemble de données de test a été créé aléatoirement en dix itérations à partir de l'ensemble de données de formation, et les modèles ont été testés sur ces dix itérations. Les résultats de l'évaluation ont révélé que le modèle CCT présentait une précision de 93,6 %, la plus élevée parmi tous les modèles dans les dix itérations de test.

Des SVM avec différents noyaux et l'extraction de signaux acoustiques ont été utilisés pour détecter les infestations de termites [31]. Les résultats expérimentaux ont révélé que le SVM avec la fonction de noyau polynomiale atteignait une précision de classification élevée.

Récemment, les chercheurs de [32] ont utilisé l'exploration de données pour résoudre le problème de la détection de la maladie du *Fusarium du* poivron. Les réflexions de la lumière sur les feuilles de poivron, mesurées à l'aide d'un spectroradiomètre, ont été utilisées pour classer la plante en quatre catégories : saine, fusariée avec champignon mycorhizien, fusariée avec champignon mycorhizien. Des expériences ont été menées en utilisant des ANN, Naive Bayes et KNN pour la classification. Les résultats avec des valeurs de précision élevées ont démontré l'efficacité des techniques d'exploration de données. Un réseau de capteurs visuels sans fil a été mis en œuvre pour surveiller les mauvaises herbes dans les cultures de riz [33]. Les images ont été collectées et traitées pour éliminer l'arrière-plan du sol, et différentes caractéristiques de forme ont été extraites. Le système met en œuvre des algorithmes de classification de type forêt aléatoire et SVM. Les résultats et les observations obtenus à partir de la configuration expérimentale du système ont révélé que le classificateur de la forêt aléatoire était plus performant que le SVM classificateur.

Malgré les efforts précédents, la recherche sur la détection et la prédiction de l'infestation par les TBI à l'aide de techniques d'exploration de données est encore insuffisante. Nous pensons que cela est dû à la difficulté d'obtenir des ensembles de données représentatifs de l'infestation. Dans [34], un système automatisé basé sur des ANN a été développé pour identifier les taupes brunes afin d'aider à la détection et à la prédiction des taupes brunes. Le modèle ANN proposé a été entraîné et évalué à l'aide d'un ensemble de données d'images contenant 326 images de TBI et 93 images d'autres insectes. Les auteurs ont conclu qu'un ANN à trois couches utilisant un gradient conjugué avec l'algorithme de redémarrage de Powell-Beale pour l'apprentissage supervisé en amont est optimal pour l'identification de l'abeille rouge du palmier. Cependant, le système proposé était très exigeant en termes de calcul. Pour surmonter cette limitation, Ref. [35] a utilisé un plus petit nombre de descripteurs d'images en combinaison avec des modèles ANN. Le nouveau système a permis une meilleure identification des NPR et s'est avéré jusqu'à 14 fois plus rapide lors de l'apprentissage et trois fois plus rapide lors du test des images d'insectes.

Cependant, les deux études précédentes sur le RPW sont basées sur un ensemble de données d'images, ce qui signifie que les palmiers infestés ont déjà atteint un niveau d'infestation sévère à partir duquel les symptômes sont visibles. En revanche, notre étude vise la détection précoce et la prédiction avant que les symptômes ne deviennent visibles.

---

Elle ne repose donc pas sur des images, ce qui rend la mise en œuvre du système plus faisable si l'on tient compte des frais généraux de stockage et de calcul.

### **3. Matériels et méthodes**

Dans cette recherche, nous avons évalué les performances de dix algorithmes de classification pour prédire l'infestation des palmiers avec le RPW. Tous les algorithmes de classification ont été

Les modèles sont mis en œuvre à l'aide du Waikato Environment for Knowledge Analysis (WEKA) et sont exécutés sur un MacBook Pro 9.2 avec un système d'exploitation OS X 10.9.5 et un Intel Core i7 à 2,9 GHz avec 8 Go de mémoire. Tous les modèles sont évalués à l'aide d'une validation croisée 10 fois. L'ensemble de données utilisé dans cet article est un ensemble de données réelles de palmiers en Arabie Saoudite, collectées après la coupe des arbres pour détecter l'infestation et l'application d'un traitement [21].

### 3.1. Ensemble de données

L'ensemble de données représente une expérience qui a été menée sur des palmiers dattiers dans la région de Kharj à Riyad, en Arabie saoudite, en avril 2017, dans le but de tester une méthode de lutte utilisant un traitement au nématode (Koppert's "Palmanem" (nématode entomopathogène, *Steinernema carpocapsae*)) (Koppert Biological Systems, Rodenrijs, Pays-Bas).

Dans cette expérience, les palmiers dattiers ont été sélectionnés selon qu'ils étaient infestés ou sains. Tous les attributs de l'arbre (15 attributs) ont été mesurés. Ensuite, les arbres ont été disséqués, puis la présence ou l'absence d'infestation a été vérifiée. L'ensemble de données RPW comporte 36 caractéristiques et 1200 enregistrements. L'ensemble de données comporte quelques valeurs manquantes qui ont été remplacées par le mode de chaque classe. L'ensemble de données est déséquilibré : 384 enregistrements appartiennent à des arbres infestés, tandis que 816 enregistrements appartiennent à des arbres sains. Nous avons rééchantillonné l'ensemble de données afin d'obtenir un nombre égal d'enregistrements, 600, pour chaque classe. Les attributs ayant une seule valeur pour tous les enregistrements ont été supprimés. L'ensemble de données final se compose de 15 attributs, comme le montre le tableau 1 : traitement, schéma d'injection, durée d'exposition, réplication, date de traitement, hauteur de la plante-1, hauteur de la plante-2, diamètre, circonférence, température au nord, température à l'est, température au sud, température à l'ouest, coordonnées (au nord) et coordonnées (à l'est).

**Tableau 1.** Description des attributs de l'ensemble de données.

Attribut	Définition
Traitement	Dans l'expérience, chaque traitement représente une tactique de contrôle (avec un nématode insecte (Palmanem de Koppert (nématode entomopathogène) ou tactique de l'eau). Les tactiques sont administrées dans le palmier dattier par le biais d'injections. Le schéma d'injection révèle la position des
Modèle d'injection	injection dans le palmier dattier, c'est-à-dire injection à la base de l'arbre, injection à un mètre de hauteur, injections en spirale pour couvrir l'ensemble du tronc (nord, sud, est et ouest).
Durée d'exposition expérimentales. Répétition	Il s'agit d'une période pendant laquelle un traitement est appliqué aux unités Répétition des traitements expérimentaux.
Date du traitement	Date d'application du traitement.
Hauteur de la plante-1	Longueur du palmier dattier de la base à la zone située sous la
Hauteur de la plante-2	Longueur totale du palmier dattier.
Diamètre dattier. Circonférence	La largeur totale du tronc du palmier La distance totale autour du tronc du
Température au nord	Température interne du tronc du palmier dattier en direction du
Température Est	Température interne du tronc du palmier dattier en direction de
Température sud	Température interne du tronc du palmier dattier en direction du
Température ouest	Température interne du tronc du palmier dattier en direction de
Coordonnées (nord)	Coordonnées associées aux positions dans la direction nord.
Coordonnées (est)	Coordonnées associées aux positions dans la direction est.

### 3.2. Sélection des caractéristiques

Pour déterminer les caractéristiques clés qui contribuent à la classification d'un palmier comme infesté ou sain, nous avons calculé le coefficient de corrélation de Pearson entre les attributs et la classe. En outre, nous avons calculé le gain d'information et le ratio de gain de chaque attribut par rapport aux classes [36]. Dans l'ensemble de données RPW, les attributs ayant une valeur nulle pour l'une des mesures calculées ont

### 3.3. Algorithmes de classification

Dans cette étude, nous avons considéré 10 algorithmes de classification de pointe, Naive Bayes (NB) [8], KSTAR [12], AdaBoost [13], bagging [14], PART [15], l'arbre de décision J48 [16], perceptron multicouche (MLP) [17], SVM [18], forêt aléatoire [19] et régression logistique [20],

en raison de leur popularité et de leur disponibilité dans de nombreuses boîtes à outils d'exploration de données. La figure 1 montre les étapes suivies dans cette étude.

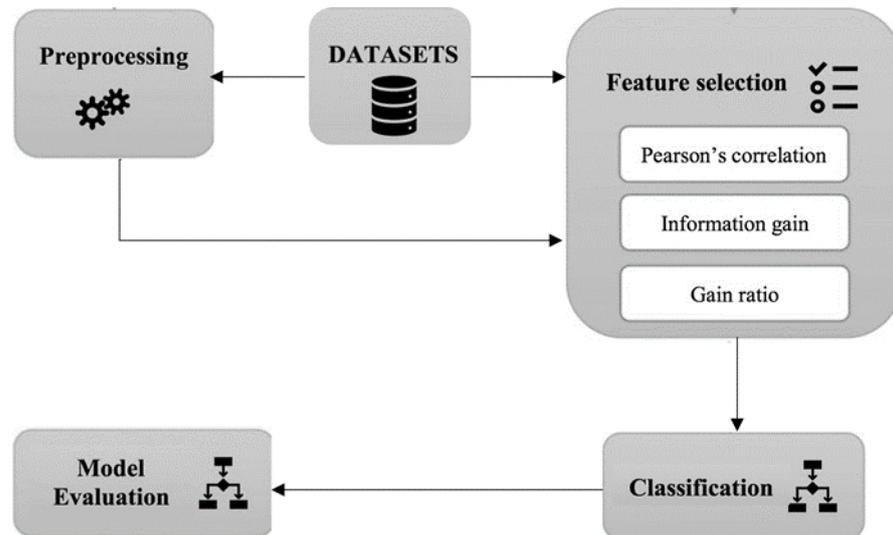


Figure 1. Déroulement de l'étude.

### 3.4. Mesures de performance

Nous avons évalué les performances de prédiction des algorithmes de classification sélectionnés en fonction des mesures suivantes :

- Précision : pourcentage d'échantillons correctement classés, calculé sur la base de (1) :

$$\text{Précision} = (TP + TN) / (TP + TN + FP + FN). \quad (1)$$

- Rappel : fraction des échantillons correctement classés comme infestés par rapport au nombre total d'échantillons infestés, calculée sur la base de (2) :

$$\text{Rappel} = (TP) / (TP + FN). \quad (2)$$

- Précision : fraction des échantillons correctement classés comme infestés sur l'ensemble des échantillons prédits comme infestés, calculée sur la base de (3) :

$$\text{Précision} = (TP) / (TP + FP). \quad (3)$$

- Mesure F : moyenne harmonique de la précision et du rappel, calculée sur la base de (4) :

$$F - \text{Mesure} = \frac{2 \times \text{précision} \times \text{rappel}}{\text{précision} + \text{rappel}} \quad (4)$$

où TP (vrais positifs) représente le nombre total d'échantillons correctement classés comme infestés, FP (faux positifs) représente le nombre total d'échantillons incorrectement classés comme infestés, FN (faux négatifs) représente le nombre total d'échantillons incorrectement classés comme sains, et TN (vrais négatifs) représente le nombre total d'échantillons correctement classés comme sains.

## 4. Résultats et discussion

Les valeurs obtenues pour le coefficient de corrélation de Pearson ont révélé que les attributs de température ouest, est, nord et sud sont fortement corrélés à l'étiquette de la classe. Les valeurs de corrélation obtenues sont respectivement de 0,63937, 0,51306, 0,49284 et 0,25765. La hauteur de la plante-1 a obtenu une valeur de corrélation très faible de 0,00316, ce qui indique qu'elle peut être moins révélatrice d'une infestation. Les valeurs de corrélation globales sont comprises entre 0,00316 et 0,63937.

Les attributs de température présentent un gain d'information élevé. Les valeurs sont de 0,71068, 0,71068 et 0,709856 pour les attributs nord, est et ouest, respectivement. La circonférence semble également importante pour la classification, puisqu'elle présente un gain d'information de 0,707054. Parmi tous les attributs, la durée d'exposition, qui décrit la durée (en mois) de l'exposition au traitement, présentait un gain d'information nul ; elle a donc été déplacée. En ce qui concerne le ratio de gain, les valeurs globales étaient légèrement inférieures à celles des deux mesures précédentes. Elles se situent dans une fourchette comprise entre zéro et 0,401156. Les attributs les mieux classés sont liés à la température et à la circonférence. En outre, la durée d'exposition semble à nouveau ne pas être pertinente pour la classification, puisqu'elle a de nouveau obtenu la valeur zéro, ce qui explique qu'elle ait été supprimée. Par conséquent, nous avons inclus les 15 attributs pour la méthode d'évaluation des attributs de corrélation et nous avons supprimé l'attribut "temps d'exposition" pour les deux autres méthodes. Nous disposons donc de 14 attributs pour les méthodes d'évaluation des attributs de gain d'information et de rapport de gain.

Nos résultats concernant l'importance de la température pour prédire les infestations par les TPP sont cohérents avec des études antérieures [37,38], qui montrent que la température interne du tronc du palmier-dattier est un attribut important, car l'infestation par les TPP entraîne un changement significatif de la température interne du tronc du palmier-dattier par rapport à celle de l'environnement. En conséquence, ils ont suggéré que le changement de température peut être utilisé pour prédire la présence de TPP à l'intérieur du tronc du palmier-dattier. En plus de la température, certains chercheurs ont également signalé que la population de TBI à l'intérieur du tronc de palmier-dattier est positivement corrélée à l'épaisseur (diamètre/circonférence) du tronc [39]. Les performances des algorithmes de pointe sur l'ensemble de données équilibré de RPW étaient très similaires pour les trois méthodes de sélection des caractéristiques, comme l'illustre le tableau 2. Tous les algorithmes (à l'exception de Naive Bayes) ont obtenu des résultats très similaires pour les quatre mesures de performance : exactitude (93,08 %), précision (87,80 %), rappel (100 %) et mesure f (93,50 %). Naive Bayes a obtenu des valeurs inférieures à celles des autres algorithmes en termes de précision (82,58 %), de rappel (72,80 %) et de f-mesure (80,70 %), ce qui peut être attribué au fait qu'il suppose implicitement que tous les attributs sont mutuellement indépendants, ce qui n'est pas réaliste dans les ensembles de données réels.

**Tableau 2.** Performances des 10 classificateurs sur l'ensemble de données du charançon rouge du palmier (RPW).

Algorithme	Précision	Précision	Rappel	Mesure F
Naive Bayes	82.58%	90.50%	72.80%	80.70%
Logistique	93.08%	87.80%	100%	93.50%
MLP	93.08%	87.80%	100%	93.50%
SVM	93.08%	87.80%	100%	93.50%
KSTAR	93.08%	87.80%	100%	93.50%
AdaBoost	93.08%	87.80%	100%	93.50%
Ensachage	93.08%	87.80%	100%	93.50%
PARTIE	93.08%	87.80%	100%	93.50%
J48	93.08%	87.80%	100%	93.50%
Forêt aléatoire	93.08%	87.80%	100%	93.50%

## 5. Conclusions

Au cours des dernières décennies, les palmiers ont été de plus en plus infestés par le ravageur RPW. Pour minimiser la perte de palmiers et contrôler ce ravageur, nous devons détecter l'infestation sans avoir à couper les arbres. L'objectif principal de cette étude est donc d'évaluer la capacité de différents algorithmes de classification par exploration de données à prédire avec précision l'infestation par le TBEP sans perdre de palmiers. Nous avons appliqué 10 techniques de classification de pointe qui ont démontré leur efficacité dans différents domaines. Ces algorithmes ont été exécutés en utilisant une validation croisée de 10 fois sur un ensemble de données réel d'infestation par les tordeuses de la pelouse. Nous avons comparé la capacité de ces algorithmes de

---

classification à prédire l'infestation par les térébrants. Les performances globales des algorithmes de classification étaient identiques, avec une précision de prédiction de 93 %, Naïve Bayes affichant la précision la plus faible (82 %). Cela contraste avec [29], où les mêmes 10 algorithmes de classification ont été appliqués à la prédiction des maladies du riz

et Naïve Bayes ont montré la plus grande précision, 88%, tandis qu'AdaBoosr a eu la plus faible précision, 37%, et que les autres classificateurs ont légèrement varié dans la gamme (81-87).

En ce qui concerne les caractéristiques clés pour la classification des TBI, les résultats expérimentaux indiquent que la circonférence et la température sont des caractéristiques importantes pour prédire l'infestation. Ces résultats sont conformes à ceux d'études antérieures [37,38], qui ont montré que les infestations de tordeuses de la grappe provoquent une augmentation significative de la température interne du tronc du palmier-dattier par rapport au palmier-dattier sain et à la température atmosphérique ambiante. En outre, dans [39], il a été montré que la population de térébrants à l'intérieur du tronc du palmier-dattier présentait une corrélation positive avec la circonférence du tronc du palmier-dattier. Néanmoins, le classement des attributs pour les mesures de sélection des caractéristiques révèle la difficulté de déterminer les caractéristiques spécifiques de l'infestation par les térébrants.

Toutefois, l'ensemble de données utilisé doit être élargi et amélioré pour que les résultats soient plus concluants. Nous demandons instamment de collecter et d'agréger davantage d'ensembles de données RPW afin d'effectuer plus d'expériences pour valider ces résultats. En outre, la précision de la prédiction de l'IPR peut être améliorée en exploitant le potentiel de techniques d'exploration de données encore inexploitées.

**Contributions des auteurs :** Conceptualisation, A.A.-A. et I.A.-T. ; Conservation des données, A.S.A. ; Acquisition de fonds, H.K. ; Investigation, H.K., I.A.-T. et A.S.A. ; Méthodologie, I.A.-T. ; Logiciel, A.A.-A. ; Supervision, H.K. et I.A.-T. ; Rédaction - version originale, A.A.-A. ; Révision de la rédaction et édition, I.A.-T. Tous les auteurs ont lu et approuvé la version publiée du manuscrit.

**Financement :** Cette recherche n'a bénéficié d'aucun

financement externe. **Déclaration du comité d'examen**

**institutionnel :** Sans objet. **Déclaration de consentement**

**éclairé :** Sans objet.

**Déclaration de disponibilité des données :** Non applicable.

**Remerciements :** Cette recherche a bénéficié d'une subvention de l'Unité de soutien aux chercheurs, numéro de projet (RSP- 2020/204), Université du Roi Saud, Riyad, Arabie saoudite.

**Conflits d'intérêts :** Les auteurs ne déclarent aucun conflit d'intérêts. Les bailleurs de fonds n'ont joué aucun rôle dans la conception de l'étude, dans la collecte, l'analyse ou l'interprétation des données, dans la rédaction du manuscrit ou dans la décision de publier les résultats.

## Références

1. Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO). 2019. Disponible en ligne : <http://www.fao.org/home/en/> (consulté le 5 janvier 2021).
2. Soroker, V. ; Colazza, S. *Handbook of Major Palm Pests : Biology and Management* ; John Wiley & Sons : Chichester, Royaume-Uni, 2017.
3. Al-Kindi, K.M. ; Kwan, P. ; Andrew, N.R. ; Welch, M. Impacts of Human-related Practices on *Ommatissus Lybicus* Infestations of Date Palm in Oman. *PLoS ONE* **2017**, *12*, e0171103. [CrossRef] [PubMed]
4. El-Juhany, L.I. Dégradation des palmiers dattiers et de la production de dattes dans les pays arabes : Causes et réhabilitation potentielle. *Aust. J. Basic Appl. Sci.* **2010**, *4*, 83998-84010.
5. Hoddle, M.S. ; Hoddle, C.D. ; Faleiro, J.R. ; El-Shafie, H.A.F. ; Jeske, D.R. ; Sallam, A.A. Jusqu'où peut voler le charançon rouge du palmier (Coleoptera : Curculionidae) : Computerized Flight Mill Studies with Field Captured Weevils. *J. Econ. Entomol.* **2015**, *108*, 2599-2609. [CrossRef] [PubMed]
6. Giovino, A. ; Bertolini, E. ; Fileccia, V. ; Al Hassan, M. ; Labra, M. ; Martinelli, F. Analyse du transcriptome de *Phoenix canariensis* Chabaud en réponse aux attaques de *Rhynchophorus ferrugineus* Olivier. *Front. Plant Sci.* **2015**, *6*, 817. [CrossRef] [PubMed]
7. Al-Dosary, N.M. ; Al-Dobai, S. ; Faleiro, J.R. Examen de la gestion du charançon rouge du palmier *Rhynchophorus ferrugineus* Olivier dans les pays de l'Union européenne. *Palmier dattier Phoenix Dactylifera*, L. *Emir. J. Food Agric.* **2016**, *28*, 34-44. [CrossRef]
8. Han, J. ; Kamber, M. *Data Mining : Concepts et techniques*, 3e édition ; The Morgan Kaufmann Series in Data Management Systems ; Morgan Kaufmann : Burlington, MA, USA, 2011.
9. Rumpf, T. ; Mahlein, A.K. ; Steiner, U. ; Oerke, E.C. ; Dehne, H.W. ; Plmer, L. Early Detection and Classification of Plant Diseases with Support Vector Machines based on Hyperspectral Reflectance. *Comput. Electron. Agric.* **2010**, *74*, 91-99. [CrossRef]
10. Babbar, R. ; Babbar, S. Predicting River Water Quality Index using Data Mining Techniques. *Environ. Earth Sci.* **2017**, *76*,

- 
504. [\[CrossRef\]](#)
11. Chlingaryan, A. ; Sukkarieh, S. ; Whelan, B. Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture : A review. *Comput. Electron. Agric.* **2018**, *151*, 61-69. [\[CrossRef\]](#)
  12. Cleary, J.G. ; Trigg, L.E. K\* : An Instance-based Learner using an Entropic Distance Measure. In Proceedings of the 12th International Conference on Machine Learning, Tahoe City, CA, USA, 9-12 juillet 1995 ; pp. 108-114.

13. Freund, Y. ; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119-139. [[CrossRef](#)].
14. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123-140. [[CrossRef](#)].
15. Frank, E. ; Witten, I.H. Generating Accurate Rule Sets without Global Optimization. In Proceedings of the 15th International Conference on Machine Learning, Madison, WI, USA, 24-27 July 1998 ; Shavlik, J., Ed. ; Morgan Kaufmann : San Mateo, CA, USA, 1998 ; pp. 144-151.
16. Quinlan, R. *C4.5 : Programs for Machine Learning* ; Morgan Kaufmann Publishers : San Mateo, CA, USA, 1993.
17. Haykin, S. *Neural Networks : A Comprehensive Foundation*, 2e édition ; Prentice Hall : Upper Saddle River, NJ, USA, 1998.
18. Cortes, C. ; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273-297. [[CrossRef](#)].
19. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *4*, 5-32. [[CrossRef](#)]
20. Han, J. ; Kamber, M. ; Pei, J. *Data Mining*, 3e édition ; Elsevier : Burlington, MA, USA, 2012 ; pp. 5-8.
21. Aldawood, A.S. ; Alsagan, F.H. ; Altuwariqi, H. ; Almuteri, A. ; Rasool, K. Red Palm Weevil Chemical Treatments on Date Palms in Saudi Arabia : Results of Extensive Experimentations. In Proceedings of the Colloque Méditerranéen sur les Ravageurs des Palmiers, Nice, France, 16-18 janvier 2013.
22. Georga, E.I. ; Tachos, N.S. ; Sakellarios, A.I. ; Kigka, V.I. ; Exarchos, T.P. ; Pelosi, G. ; Parodi, O. ; Michalis, L.K. ; Fotiadis, D.I. Artificial Intelligence and Data Mining Methods for Cardiovascular Risk Prediction. In *Cardiovascular Computing Methodologies and Clinical Applications* ; Series in BioEngineering ; Springer : Singapore, 2019 ; pp. 279-301.
23. Saghir, M. ; Bibi, Z. ; Bashir, S. ; Khan, F.H. Churn Prediction using Neural Network based Individual and Ensemble Models. In Proceedings of the 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 8-12 janvier 2019 ; pp. 634-639.
24. Altujjar, Y. ; Altamimi, W. ; Al-Turaiki, I. ; Al-Razgan, M. Predicting Critical Courses Affecting Students Performance : Une étude de cas. *Procedia Comput. Sci.* **2016**, *82*, 65-71. [[CrossRef](#)]
25. Mucherino, A. ; Papajorgji, P. ; Pardalos, P.M. Enquête sur les techniques d'exploration de données appliquées à l'agriculture. *Oper. Res.* **2009**, *9*, 121-140. [[CrossRef](#)]
26. Tripathy, A.K. ; Adinarayana, J. ; Sudharsan, D. ; Merchant, S.N. ; Desai, U.B. ; Vijayalakshmi, K. ; Reddy, D.R. ; Sreenivas, G. ; Ninomiya, S. ; Hirafuji, M. ; et al. Data Mining and Wireless Sensor Network for Agriculture Pest/disease Predictions. Dans les actes du Congrès mondial sur les technologies de l'information et de la communication, Mumbai, Inde, 11-14 décembre 2011 ; pp. 1229-1234.
27. Diriba, Z. ; Borena, B. Application of Data Mining Techniques for Crop Productivity Prediction. *HiLCoE J. Comput. Sci. Technol.* **2013**, *1*, 152-155.
28. Hill, M.G. ; Connolly, P.G. ; Reutemann, P. ; Fletcher, D. The Use of Data Mining to Assist Crop Protection Decisions on Kiwifruit in New Zealand. *Comput. Electron. Agric.* **2014**, *108*, 250-257. [[CrossRef](#)]
29. Sengupta, S. ; Das, A.K. Particle Swarm Optimization based Incremental Classifier Design for Rice Disease Prediction. *Comput. Electron. Agric.* **2017**, *140*, 443-451. [[CrossRef](#)]
30. Ilic, M. ; Ilic, S. ; Jovic, S. ; Panic, S. Early Cherry Fruit Pathogen Disease Detection based on Data Mining Prediction. *Comput. Electron. Agric.* **2018**, *150*, 418-425. [[CrossRef](#)]
31. Nanda, M.A. ; Seminar, K.B. ; Nandika, D. ; Maddu, A. Étude comparative des fonctions du noyau dans la machine à vecteur de support et son application pour la détection des termites. *Information* **2018**, *9*, 5 [[CrossRef](#)].
32. Karadagç, K. ; Tenekeci, M.E. ; Tas, altin, R. ; Bilgili, A. Detection of Pepper Fusarium Disease using Machine Learning Algorithms based on Spectral Reflectance. *Sustain. Comput. Inform. Syst.* **2020**, *28*, 100299. [[CrossRef](#)]
33. Kamath, R. ; Balachandra, M. ; Prabhu, S. Raspberry Pi as Visual Sensor Nodes in Precision Agriculture : Une étude. *IEEE Access* **2019**, *7*, 45110-45122. [[CrossRef](#)].
34. Al-Saqer, S.M. ; Hassan, G.M. Système de reconnaissance du charançon rouge du palmier (*Rhynchophorus ferrugineus*, Olivier) basé sur des réseaux neuronaux artificiels. *Am. J. Agric. Biol. Sci.* **2011**, *6*, 356-364. [[CrossRef](#)]
35. Al-Saqer, S.M. Un système d'identification fiable pour le charançon rouge du palmier. *Am. J. Appl. Sci.* **2012**, *9*, 1150-1157.
36. Karegowda, A. ; Manjunath, A. ; Jayaram, M. Comparative study of attribute selection using gain ratio and correlation based feature selection. *Int. J. Inf. Technol. Knowl. Manag.* **2010**, *2*, 271-277.
37. Mozib, M.E.F. ; El-Shafie, H.A. Effet de l'infestation par le charançon rouge du palmier, *Rhynchophorus ferrugineus* (Olivier) sur les profils de température du palmier dattier. *J. Entomol. Nematol.* **2013**, *5*, 77-83. [[CrossRef](#)]
38. El-Faki, M.S. ; El-Shafie, H.A.F. ; Al-Hajhoj, M.B.R. Possibilités de détection précoce des palmiers dattiers (Arecaceae) infestés par le charançon rouge du palmier (Coleoptera : *Curculionidae*) à l'aide des différentiels de température. *Can. Entomol.* **2016**, *148*, 239-245. [[CrossRef](#)]
39. Aldawood, A.S. ; Rasool, K.G. Optimisation de l'élevage du charançon rouge du palmier : *Rhynchophorus ferrugineus* (Coleoptera : *Curculionidae*) sur le palmier dattier : Phoenix dactylifera. *Fla. Entomol.* **2011**, *94*, 756-760. [[CrossRef](#)]