**TOWARDS AN EFFICIENT PREDICTION MODEL OF MALARIA CASES IN SENEGAL**

Ousseynou Mbaye, Mouhamadou lamine Ba, Gaoussou Camara, Alassane Sy

UADB

{ousseynou.mbaye, mouhamadoulamine.ba, gaoussou.camara, alassane.sy}@uadb.edu.sn

**Résumé**

Le paludisme est l’une des maladies les plus mortelles dans le monde plus particulièrement en Afrique subsaharien. La situation est critique dans des pays comme le Senegal à cause du manque de services de santé de qualité et de personnel médical qualifie et capable de faire des diagnostics précis des maladies dont souffrent les patients.

Ainsi la nécessité de trouver des outils automatisées pour aider les acteurs médicaux dans le processus de prise de décision après avoir réalisé le diagnostic.

Dans ce papier nous proposons les premières étapes vers la réalisation d’un algorithme de diagnostic du paludisme basé sur les signes et symptômes du patient en plus du TDR (Test de Diagnostic Rapide). Notre modèle de prédiction est basée sur la régression logistique.

Les résultats des premiers tests sur un jeu de données réel et semi-synthétique de patients se sont très prometteurs concernant l’efficacité de l’approche proposée.

Mots clés : Diagnostic, Paludisme, Modèle de Prédiction, Imputation des données

1. **Contexte**

Le paludisme est une maladie potentiellement mortelle due à des parasites transmis à l’homme par des piqûres de moustiques femelles infectés. Selon le dernier Rapport sur le paludisme dans le monde, publié en novembre 2017, il y a eu 216 millions de cas de paludisme en 2016, contre 211 millions en 2015. On estime à 445 000 le nombre de décès dus au paludisme en 2016, un chiffre similaire à celui de l’année précédente (446 000). La Région OMS de l'Afrique supporte une part disproportionnée de la charge mondiale du paludisme. En 2016, 90% des cas de paludisme et 91% des décès dus à cette maladie sont survenus dans cette région. 80% de la charge de morbidité due au paludisme pesaient sur une quinzaine de pays – tous situés en Afrique subsaharienne, sauf l’Inde. Malgré les efforts consentis par les Etats et organismes non gouvernementaux, le paludisme demeure toujours un facteur majeur de mortalité à travers le monde.

Cependant, la plus part de ces décès sont dus à un diagnostic incomplet qui ne révèle pas le type exact de paludisme.

1. **Problématique**

Le diagnostic tardif d’un paludisme grave conduit très souvent à la mort. Cependant, les techniques de diagnostic du paludisme existant ne donnent pas une idée très précise du niveau de gravité de la maladie. En effet, le diagnostic basé sur la recherche des plasmodies (par microscopie) ou le test de diagnostic rapide (TDR) ne permet pas de faire une classification rigoureuse de l’état d’avancement de la maladie. Ainsi, il est nécessaire de trouver de nouvelles méthodes qui permettront d’avoir un diagnostic complet et fiable.

1. **Objectifs**

**2.1 Objectif Général**

Le but de cette étude est de proposer un algorithme de diagnostic du paludisme basé sur les signes et symptômes du patient en plus du TDR (Test de Diagnostic Rapide).

2.2 **Objectifs spécifiques**

Les objectifs spécifiques de cette étude sont :

* Traitement données, sélection des attributs du paludisme et formalisation des données pour obtenir un jeu données correcte pour la prédiction
* Imputation des données manquantes et la création d’un jeu de données synthétiques pour rééquilibrer les classes des attributs
* Produire un algorithme de prédiction du paludisme en utilisant la régression logistique avec le logiciel python
* Tester l’algorithme avec les données réelles de patient atteint de paludisme

1. **Résultats Préliminaires :**

Dans cette étude l’objectif est de prédire si un patient est atteint de paludisme ou non connaissant ces données de diagnostics et nominales. Etant donné que notre problème est un problème de classification binaire avec les classes Paludisme et Non-Paludisme nous utilisons la régression logistique qui est un algorithme de machine learning fortement utilisé dans le domaine médical.

Nous avons appliqué cet algorithme à deux jeux de données : un jeu de données réel de patients consultés lors du grand Magal de Touba(DT1) et un jeu de données semi-synthétique obtenu à partir du premier jeu de donné(DT2).

Pour chaque jeu de données donné, nous avons effectué deux types de tests avec notre modèle de prédiction: un test sans inclure le résultat du test de diagnostic rapide (TDR) et un autre avec le résultat du test de diagnostic rapide (TDR) dans les attributs que nous mettons en entré du modèle.

Les expériences prouvent que notre modèle de prévision basé sur la régression logistique fonctionne bien en général et en particulier lorsque le résultat du test de diagnostic rapide est considéré comme une caractéristique du processus d’apprentissage, plus précisément, la précision de nos prédictions est supérieure à 90% pour les jeux de données DT1 et DT2, atteignant 100% pour DT1. Pour le cas spécifique du jeu de données TD1 contenant des données manquantes à l'aide d'un algorithme d'imputation, cette précision ne diminue pas, même si le TDR n'est pas pris en compte pendant le processus de prévision, et seules les caractéristiques du paludisme, c'est-à-dire les signes et les symptômes, sont prises en compte.

En conséquence, nous pensons qu'il est possible de construire un modèle de prévision efficace et robuste du paludisme sans qu'il soit nécessaire d'effectuer le test de diagnostic rapide ; c’est-à-dire prédire si un patient donné est atteint ou non de paludisme

1. **Perspectives :**

Pour améliorer les résultats de ce modèle nous prévoyons :

* d’introduire un facteur de risque en prenant en compte le taux de morbidité et le taux de létalité par région
* d’utiliser d’autres modèles de classification binaires tels que Support Vector Machine (ou SVM) et comparer leurs résultats à ceux obtenus avec le modèle basé sur la régression logistique.

1. **Références**

1. OpenRefine. http://openrefine.org/, online; accessed 30 Octobery 2018

2. Python implementation of missforest. https://pypi.org/project/predictive imputer/, online; accessed 31 October 2018

3. Adimi, F., Soebiyanto, R.P., Safi, N., Kiang, R.: Towards malaria risk prediction in afghanistan using remote sensing. Malaria Journal 9(1), 125 (May 2010)

4. Aminot I, D.M.: The use of logistic regression in the analysis of data concerning good medical

practice. Rev Med Ass Maladie 33(2), 157–143 (2002)

5. AS, A., AM, V., SH., K.: Malaria parasite development in the mosquito and infection of the mammalian host pp. 195–221 (2009)

6. Bbosa, F.,Wesonga, R., Jehopio, P.: Clinical malaria diagnosis: rule-based classification statistical prototype. SpringerPlus 5(1), 939 (Jun 2016)

7. WHO: World malaria report in 2017 (2017)

Annexes :















