Public Health and Informatics
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210196

On the Efficiency of Machine Learning Models in Malaria Prediction

Ousseynou MBAYE[a], Mouhamadou Lamine BA[a,1] and Alassane SY[a]
[a] *Université Alioune Diop de Bambey, Bambey, Senegal*

Abstract. Malaria is still a real public health concern in Sub-Saharan African countries such as Senegal where it represents approximately 35% of the consultation activities in the hospitals. This is mainly due to the lack of appropriate medical care support and often late and error-prone diagnosis of the disease. For instance, largely used tools like Rapid Diagnosis Test are not fully reliable. This study proposes an extensive study of the efficiency of the most popular machine learning models for the task of Malaria occurrence prediction. We have considered patients from Senegal and have evaluated the overall accuracy of each considered algorithm based on sign and symptom information. Our main result is that machine learning algorithms are promising, in particular Naive Bayesian presents a recall very close to that of a rapid diagnostic test while improving highly its precision by 9%.

Keywords. Malaria, prediction, ML, performance, evaluation, Sign, Symptom

1. Introduction

Malaria is a transmissible disease through the bites of infected female Anopheles mosquitoes. It comes with symptoms such as fever, headache, and chills in its early stage and can evolve to more severe health problems (severe anemia, respiratory distress, etc.) often leading to death. In 2019, the number of Malaria cases worldwide has been estimated to be 229 million. The number of deaths caused by Malaria has been approximately estimated to 409,000 in 2019; the African area represents around 94% of the reported Malaria cases and deaths in 2019, thanks to the annual world Malaria report [1]. Over the past years, many efforts have been made by governmental and non-governmental organizations (e.g. WHO) to eradicate Malaria in the world. In the research field, many studies, aiming at understanding the disease from the Plasmodium mosquito point of view or proposing automated detection tools, have been conducted [2,3,4,5]. The Rapid Diagnostic Test (RDT) [5] is one of the most successful and prominent introduced tools to automatically predict whether or not a given patient suffers from Malaria. It relies on the detection of the presence of specific Plasmodium proteins, PfHRP2, pLDH andaldolase in human blood. The RDT is largely used and adopted as a standard in many Sub-Saharan African countries such as Senegal. However, as proved in [5], RDT is not fully reliable: Section 2 shows that the precision of RDT is about 90% for data used in this study. Despite those advanced tools, Malaria is still a real public health issue in Africa because of the lack of appropriate care support or late and error-

---

[1] Corresponding Author, Mouhamadou Lamine BA, LIMA, Université Alioune Diop, BP.3400 Bambey, Senegal; E-mail: mouhamadoulamine.ba@uadb.edu.sn

prone detection of the disease. Artificial intelligence is now recognized as a domain that may help medical actors in their decision-making process [6,7] and to overcome the lack of enough health resources.

This paper proposes an extensive comparative study of the efficiency of the most popular machine learning models for the task of Malaria prediction. The evaluated and compared ML algorithms are Naive Bayes (NB), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN). We conducted experiments on a real-world dataset about patients living in Senegal. This dataset contains clinical data such as signs, symptoms, the final diagnostic of the doctor, as well as the outcome of the RDT. As a main contribution, our tests first show that the precision of all considered machine learning models outperform all RDT in terms of precision. Then, it also proves that NB, SVM and ANN are close to RDT in terms of recall. The rest of the paper is organized as follows. We start by presenting the methods used in this work in section 2. Then, we detail in section 3 the results of our experimental evaluation before concluding this paper in section 4.

2. Methods

For the purposes of studying the efficiency of the most popular machine learning algorithms for the task of Malaria prediction. We rely on real data and some performance evaluation metrics. We detail next the methodology used in this study.

2.1. Data collection and preparation

In order to carry out our experiments in a real setting, we have collected real-world data about patients living in Senegal. Our dataset, referred to as DT, has been collected in 2016 during the "Grand Magal" of Touba, a big religious event in Senegal that gathers several millions of people every year [8]. After the collection step, we have conducted some cleaning, transformation and imputation tasks on the raw data in order to deal with noisy information and missing values. We have then proceeded to feature selection in order to only consider the data attributes (or variables) such as *lack of appetite, tiredness, fever, cephalalgia, nausea, arthralgia, digestive disorders, dizziness, chill, myalgia, diarrhea, and abdominal pain* pertaining for our study; we defer the reader to [9] for the details. For privacy reasons and certain restrictions in the use of the data, we have ignored patient personal data. Table 1 summarizes the main characteristics of the dataset. One can easily observe that the dataset regarding the individuals of each class (i.e. Malaria and Not Malaria) as shown in table 1. To solve our problem of imbalanced dataset, we used the algorithm SMOTE [10], which is a synthetic minority oversampling technique, through its Python implementation in the package imbalanced-learn [11]. SMOTE consists of predicting a sample of synthetic dataset based on the value of the minority class of the targeted class. It randomly chooses the k-nearest neighbors of a given record in order to randomly create new observations. We have applied an over-sampling of the minority class into our patient dataset for having a dataset DT with balanced classes; this enables to avoid over-fitting of a given prediction model.

Table 1. Main characteristics of our real-world dataset about patients living in Senegal

Variables	Observation	Variable Types		Classes	
		Numeric	**Boolean**	**Malaria**	**Not Malaria**
16	21083	2	14	614	20469

2.2. Machine learning algorithms

We have considered and compared the six most popular machine learning approaches [12,13] which are Decision Tree (DT) [14], Random Forest (RF) [15], Naive Bayes (NB) [16], Logistic regression (LR) [17], Support Vector Machine (SVM) [18], Artificial Neural Network (ANN) [19]. All of them are supervised learning algorithms, i.e., require a training phase with labelled data.

2.3. Experimentation setting

We have trained and validated each algorithm of our dataset using *stratified-5-fold cross-validation* [2] in the same experimentation environment. We have relied on ML implementation of the algorithms available with the Scikit-Learn Python library. To evaluate the efficiency of each algorithm, we finally measured its *precision*, *recall*, *F1-score*, *AUC*[3], and *specificity*. For RDT, we deduced its performance measures from the dataset.

3. Results and Discussions

We start by comparing the performances of our six tested machine learning models on our real-world dataset. Table 2 details the results of the experiments with the different algorithms by giving the precision, the recall, the F1-score, the AUC, and the specificity of each algorithm tested. On the other hand, Figure 1 provides a comparative view of the performance of these models with respect to each metric. Given results in Table 2 and Figure 1, we first observe that all the ML models present a very high precision but different values for the other metrics. More specifically, we can secondly remark that the classifiers LR, NB and ANN present better overall accuracy values regarding all the metrics. Finally, we can conclude that NB offers the best trade-off between precision, recall, AUC, and specificity compared to the others; it presents better true positive and true negative prediction rates.

Table 2. Performance measures of the tested Machine Learning algorithms on our real-world dataset

ML algorithms	Precision	Recall	F1-score	AUC	Specificity
Decision Tree	0.99	0.84	0.91	0.76	0.58
Random Forest	0.99	0.84	0.91	0.76	0.60
Logistic Regression	0.90	0.78	0.88	0.84	0.75
Naive Bayesian	0.99	0.82	0.90	0.84	0.71
Support Vector Machine	0.99	0.86	0.92	0.80	0.62
Artificial Neural Networks	0.99	0.84	0.91	0.79	0.65

---

[2] https://scikit-learn.org/stable/modules/cross_validation.html

[3] https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=fr

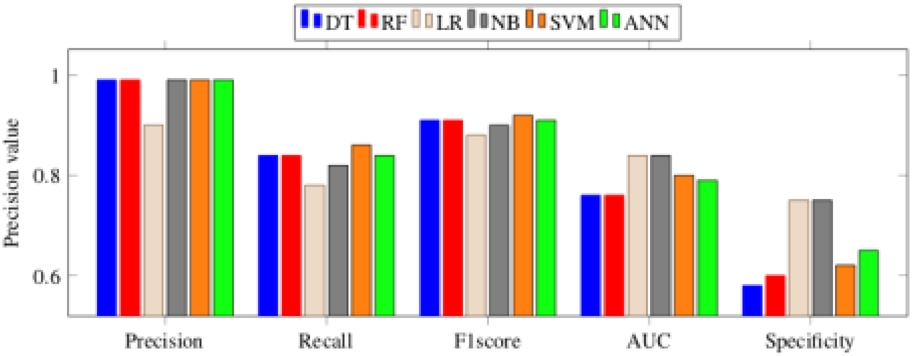

Figure 1. Comparison of the performances of the six tested ML models on DT

Secondly, we compare NB, which offers the best trade-off amongst all the tested ML models, with the widely used rapid test diagnostic model. Table 3 contains the precision, recall, F1-score, and specificity of the RDT on the patient dataset used in this study: the outcome of the RDT was available for each patient in the data so that we were able to deduce its overall performances. Figure 2 depicts a pairwise comparison of the precision, recall and specificity of Naive Bayesian and Rapid Diagnostic Test. We observe that NB highly improves the precision of RDT by 9% proving its efficiency in predicting the patients that really suffer from Malaria. However, even if NB and RDT are close in terms of sensitivity or recall, the latter present a better specificity. In other words, BN is less accurate in predicting the true negative within each class.

Table 3. Performance measures of the RDT on our real-world dataset

Precision	**Recall**	**F1-score**	**Specificity**
0.90	0.85	0.87	0.90

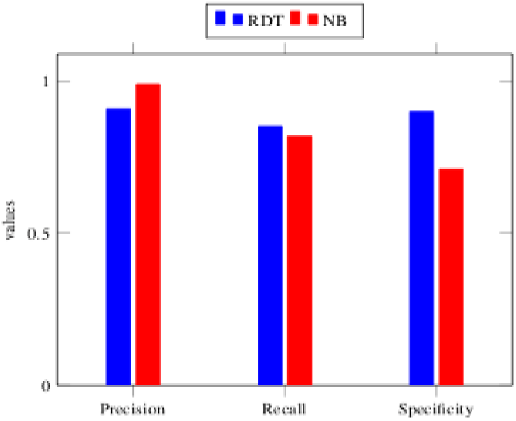

Figure 2. Pairwise comparison of the performances of the RDT and Naive Bayesian

4. Conclusion

In this study, six classifiers using a wide variety of operating procedures have been extensively tested and compared over a real-world health dataset from Senegal in order to evaluate their performance for the task of predicting the occurrence or not of Malaria in a patient knowing his signs and symptoms. The results obtained show that ML models achieve very high precision compared to the baseline Rapid Diagnostic Test Model, but lower specificity scores. Amongst these ML models, we showed that NB offers a better trade-off when focusing on precision, recall, AUC, and sensitivity. In addition, we observed that NB is close to RDT in terms of recall. As future works, we plan to enhance NB so that it can outperform RDT in terms of sensitivity and specificity. Another research direction is to investigate an assembling approach based on the classifiers offering the best performances in our present study. But also, to compare these performances with the ensemble methods for their validation.

References

[1] World Health Organization. 2019 World Malaria Report; 2019. Available at: https://www.who.int/publications/i/item/9789241565721. Accessed Jan 2020.

[2] Garrido-Cardenas JA, Cebrián-Carmona J, González-Cerón L, Manzano-Agugliaro F, Mesa-Valle C. Analysis of Global Research on Malaria and Plasmodium vivax. Int J Environ Res Public Health. 2019 May 31;16(11):1928.

[3] Lepes T. Review of Research on Malaria. Bulletin of the World Health Organization. 1974;50(3-4):151 – 57.

[4] Ermert V, Fink AH, Jones AE, Morse AP. Development of a new version of the Liverpool Malaria Model. I. Refining the parameter settings and mathematical formulation of basic processes based on a literature review. Malar J. 2011 Feb 11;10:35.

[5] Houzé S. Les tests de diagnostic rapide pour le paludisme [Rapid diagnostic test for malaria]. Bull Soc Pathol Exot. 2017 Feb;110(1):49-54.

[6] Mitchell TM, et al. Machine learning. McGraw-Hill; 1997.

[7] Yadav A, Arpitha MS, Mithun KA. Better Healthcare using Machine Learning. International Journal of Advanced Research in Computer Science. 2010;9(1).

[8] Sokhna C, Mboup BM, Sow PG, Camara G, Dieng M, Sylla M, et al. Communicable and non-communicable disease risks at the Grand Magal of Touba: The largest mass gathering in Senegal. Travel Medicine and Infectious Disease. 2017;19:56 – 60.

[9] Mbaye O, Ba ML, Camara G, Sy A, Mboup BM, Diallo A. Towards an Efficient Prediction Model of Malaria Cases in Senegal. International Conference on Innovations and Interdisciplinary Solutions for Underserved Areas. Springer; 2019. p. 173–88.

[10] Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. The Journal of Machine Learning Research. 2017;18(1):559-63.

[11] Wang J, Xu M, Wang H, Zhang J. Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding. IEEE Proceedings of 8th international Conference on Signal Processing. 2006 vol. 3.

[12] De Oliveira H, Prodel M, Augusto V. Binary Classification on French Hospital Data: Benchmark of 7 Machine Learning Algorithms. Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2018. p. 1743-48.

[13] Tomar D, Agarwal S. A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology. 2013;5(5):241-66.

[14] Rokach L, Maimon O. Decision Trees. Springer; 2006;6:65-92.

[15] Breiman L. Random Forests. Machine Learning. 2001;45(1):5–32.

[16] Kaviani P, Dhotre S. Short Survey on Naive Bayes Algorithms International Journal of Advance Engineering and Research Development. 2017;4(11), 607-11.

[17] Morgan SP, Teachman JD. Logistic Regression: Description, Examples, and Comparisons. Journal of Marriage and Family. 1988;50(4):929-36.

[18] Evgeniou T, Pontil M. Support Vector Machines: Theory and Applications. In: Studies in Fuzziness and Soft Computing. 2001;2049:249-57.

[19] Mehlig B. Artificial Neural Networks. arXiv. Available at https://arxiv.org/pdf/1901.05639.pdf. Accessed Jan 2020.