Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 13 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i13.36185

PAPER

Building a Recommender System to Predict the Shape of Bacteria in Urine Cytobacteriological Examination Using Machine Learning

Mohammed Amine Lafraxo[1](✉), Hinde Hami[1], Tarik Merrakchi[2], Ali Azghar[3,4], Ahmed Remaida[1], Mohammed Ouadoud[5], Adil Maleb[3,4], Abdelmajid Soulaymani[1]

[1]Faculty of Science, Ibn Tofail University, Kenitra, Morocco

[2]Hassan II University, Casablanca, Morocco

[3]Mohammed First University, Oujda, Morocco

[4]Mohammed VI University Hospital, Oujda, Morocco

[5]Abdelmalek Essaadi University, Tetouan, Morocco

lafraxo.ma@gmail.com

ABSTRACT

This study aimed to build a recommender system that predicts the shape of bacteria for biological requests of urine cytobacteriological examination (UCBE) using machine learning techniques, to reduce the time taken to identify the shape of bacteria (*Cocci* or *Bacilli*). We used different methods and techniques in the process: Unified Modelling Language (UML) was used for digital design architecture, Rstudio tool with R programming language for system development, and Random Forest (RF) algorithm for the prediction. Experimental results showed that the time needed to identify the shape of bacteria is decreased, and bacilli bacteria are better recognized by the algorithm with an error rate of 3%. In addition to that, the proposed recommender system allows biologists to validate and correct the prediction and improve the accuracy of the classification algorithm used in the future.

KEYWORDS

recommender system, machine learning, artificial intelligence (AI), urine cytobacteriological examination (UCBE), random forest (RF) algorithm, prediction

1 INTRODUCTION

Machine learning is a scientific field and, more specifically, a subset of artificial intelligence (AI) that enables computers to automatically improve through experience [1]. Specifically, it enables algorithms to autonomously learn to perform a task or make predictions from data and improve their performance over time. The extremely rapid evolution of technological development has had a profound impact on several areas of society, including health. And with the emergence of AI, it improved and became a 21st-century innovation [2]. Over the past decades, several scientific studies and research projects have shown the importance of

Lafraxo, M.A., Hami, H., Merrakchi, T., Azghar, A., Remaida, A., Ouadoud, M., Maleb, A., Soulaymani, A. (2023). Building a Recommender System to Predict the Shape of Bacteria in Urine Cytobacteriological Examination Using Machine Learning. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(13), pp. 92–107. https://doi.org/10.3991/ijoe.v19i13.36185

Article submitted 2022-10-18. Revision uploaded 2023-02-10. Final acceptance 2023-02-24.

applying machine learning techniques through recommender systems in a health context, especially in epidemiology, radiology, biology, and medicine.

Recommender systems have been studied in several areas: information retrieval, web technologies, e-commerce, health sciences, educational technologies, and many others. The best-known recommender systems are those used on e-commerce websites. The principle is to use a customer's needs and interests as input to propose a list of recommended products. Many applications rely solely on the products customers buy and explicitly rate them to represent their interests, but such systems can also take into account other attributes, including products viewed and demographic data [3]. The objective of a recommender system is to make it easier for the user to find relevant objects according to his preferences. It considerably reduces the time it takes the user to search for the most interesting objects and to find objects that he might not have noticed. A recommender system is a system capable of providing personalized recommendations or guiding the user to interesting or useful resources within a large data space [4].

In the field of biology, several benefits have been observed while utilizing new information and communication technologies, especially AI [5]. A reduction in the workload, thanks to the automation of the most time-consuming tasks, is one such benefit. It has helped in improving diagnosis accuracy, including for rare diseases, thanks to the cross-referencing of data with thousands, then millions, of already annotated examinations. This, in turn, will inevitably improve patient care through better prioritization of cases, especially in the emergency departments.

This work aims to develop a recommender system that predicts the shape of bacteria for biological requests. A urine cytobacteriological examination (UCBE) is a microbiological examination that diagnoses a urinary tract infection by identifying the causative germ and helps to choose the best treatment. It is the most frequently requested examination in medical practice, and its interpretation is relatively easy, in theory. UCBE is one of the main diagnostic activities in a microbiology laboratory [6]. However, the latest consensus on this condition recommends non-systematic recourse to urine cytobacteriological examination [6]. In the absence of this examination, antibiotic treatment is based on up-to-date bacterial epidemiological data and geographically adapted [7] [8]. Such data can also be used to inform health agencies about the evolution of antibiotic resistance. The main objective is to use machine learning techniques to reduce the time needed to identify the form of bacteria (*Cocci* or *Bacilli*) in this examination.

This paper consists of four main sections:

- In the first section, we have defined the key concepts of machine learning algorithms, and then we present the proposed model used to develop the recommender system.
- In the second section, we present the graphical interfaces of the proposed recommendation system, and how it will work.
- In the third section, we have highlighted related work.
- In the fourth section, we present the conclusion and future work.

2 BASIC DEFINITIONS

Predictive algorithms can be applied in many fields. They can be used in the world of health. By exploiting medical information, we can establish a diagnosis of the state

of health of a patient and anticipate possible risks. In this study, we exploited biological data to predict the shape of bacteria for biological requests for urine cytobacteriological examination using machine learning techniques. The most commonly used algorithms for building predictive models are Logistic regression (LR), Naive Bayes (NB), Decision trees (DT), Random Forests (RF), Neural networks (NN), and K-nearest neighbor (KNN).

- Logistic regression (LR): Logistic regression is a supervised classification machine learning algorithm. It is generally used in binary classification problems [9]. It can be applied to three types of classification tasks: binary classification, multiple classifications, and classification in a hierarchy [10].
- Naive Bayes (NB): Naive Bayes is a machine learning algorithm. It is a Supervised Learning algorithm used for classification. It is particularly useful for text classification. An example of the use of NB is spam email filters [11].
- Decision trees (DT) [12]: Decision trees are among the most widely used classification and prediction techniques. It is a tree structure. DT usually consist of a central node from which several possible data points can be drawn. The nodes lead to other nodes, which in turn point to several other possibilities.
- Random Forests (RF): Random Forest is one of the most robust machine learning algorithms, it consists of an ensemble of individual DT to improve the prediction, it's a supervised machine learning algorithm that can be used for classification and regression [13].
- Neural Networks (NN): In the field of health, NN are increasingly being applied to survival data to make predictions [14], [15]. NN are inspired by the structure of the human brain, with neurons organized in layers that use all the information in the previous layer. These networks are only a very simplified model of our brain, but they have the particularity of being able to use their layers to represent knowledge acquired at different degrees of abstraction [16].
- K-nearest neighbor (KNN): In AI, more precisely in machine learning, the KNN is a supervised learning method. it is a simple and easy-to-implement algorithm that can be used to solve classification and regression problems. However, it is more widely used for classification prediction [17].

3 METHODS

3.1 Method and tools used to develop the recommender system

The Unified Modelling Language (UML) has been used for the design of the recommender system and is the most widely used object modeling standard available today (see Figure 10). The R studio tool was used to develop the recommender system with the R programming language, through the processes of Deming's scientific method [18]: Plan, Do Check, and Act.

3.2 Proposed model

The model studied was based on datasets that exist in a CSV format file, extracted from the UF-1000i machine. The proposed system architecture aims to predict the shape of bacteria in the cytobacteriological examination of urine. The methodology used in this work is consists of five steps. Figure 1 provides an overview

of this methodology. Once the pre-processing has been performed in the initial step of the process, the RF algorithm is used as a classifier to predict whether the bacteria's form is *Bacilli* or *Cocci*. To evaluate the performance and validate the proposed approach, the confusion matrix was used to evaluate sensitivity, specificity, and accuracy for diagnosing the algorithm's progress after each validation or correction.

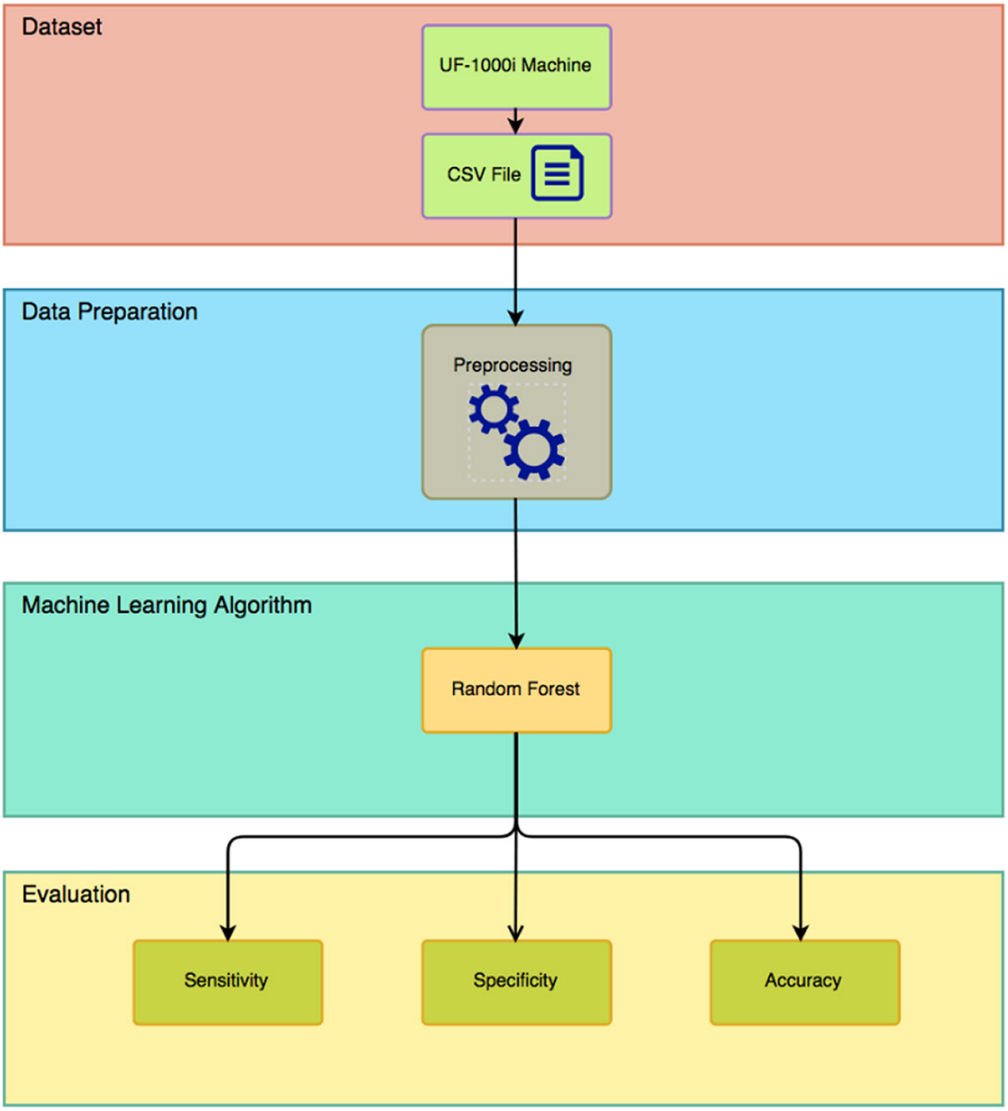

Fig. 1. Flowchart of the proposed prediction model

3.3 Dataset

The data used in this research work is collected from a file in CSV format. This file was extracted from the UF-1000i machine, it is a fully automated urine particle analyzer which processes the ECBU requests thanks to its IPU software installed on the office computer connected to the machine. This file contains several UCBE requests classified by the shape of bacteria *Bacilli*/*Cocci* (see Table 1).

Table 1. Dataset features

Name of the Field	Type	Description
TBAC	Numerical	Number of all particle types
BFSC	Numerical	The intensity of forwarding scattered light (High sensitivity): size/height of the particle
BFLH	Numerical	The intensity of the fluorescent light (high sensitivity): level of staining of the particle nucleus
BFSCW	Numerical	The pulse width of forwarding scattered light (high sensitivity): particle length
SHAPE_BACT	Factor	The shape of the bacteria is the class of the CSV and has 2 distinct values: Bacilli or Cocci

The dataset used consists of 986 rows with five features, it contains various features that are considered important, namely: the count of all particle types, the intensity of forwarding scattered light, the intensity of fluorescent light, the pulse width of forwarding scattered light, and the shape of the bacteria.

3.4 Data preparation

Data preparation is the process of collecting, combining, structuring, and organizing data so that it can be analyzed in data visualization, analysis, and machine learning applications (see Figure 2).

ui.R × server.R ×

```
1  server <- function(input, output, session) {
2      # Dataset import
3      myDataset <- read.csv("data/UCBE.csv", header=TRUE, sep=";")
4      myDataset$ID <- as.character(myDataset$ID)
5      myDataset$TBAC <- as.numeric(myDataset$TBAC)
6      myDataset$BFSC <- as.numeric(myDataset$BFSC)
7      myDataset$BFLH <- as.numeric(myDataset$BFLH)
8      myDataset$BFSCW <- as.numeric(myDataset$BFSCW)
9      myDataset$SHAPE_BACT <- as.factor(myDataset$SHAPE_BACT)
```

Fig. 2. Importing dataset using Rstudio with the R programming language

Data preparation includes preprocessing through data profiling, data cleaning, validation, and data transformation to avoid: omitted values (data not available); duplicate samples; poor annotations; and noise in the data.

One of the key goals of data preparation is to make sure that the data is accurate and consistent so that the applications' analysis outputs are reliable. Data is frequently created with blank values, mistakes, or other flaws. Additionally, it is frequently necessary to manage the many formats of data sets that are kept in separate files or databases. A sizable portion of the data preparation process involves fixing errors, checking, and putting data sets together.

In our research, the dataset exported from the CSV file had 4,56% of the data cleaned. The original database contains 986 rows, after the preparation process, the resulting database contains 941 rows (95,44%). Also, the histogram of all features is shown in Figure 3.

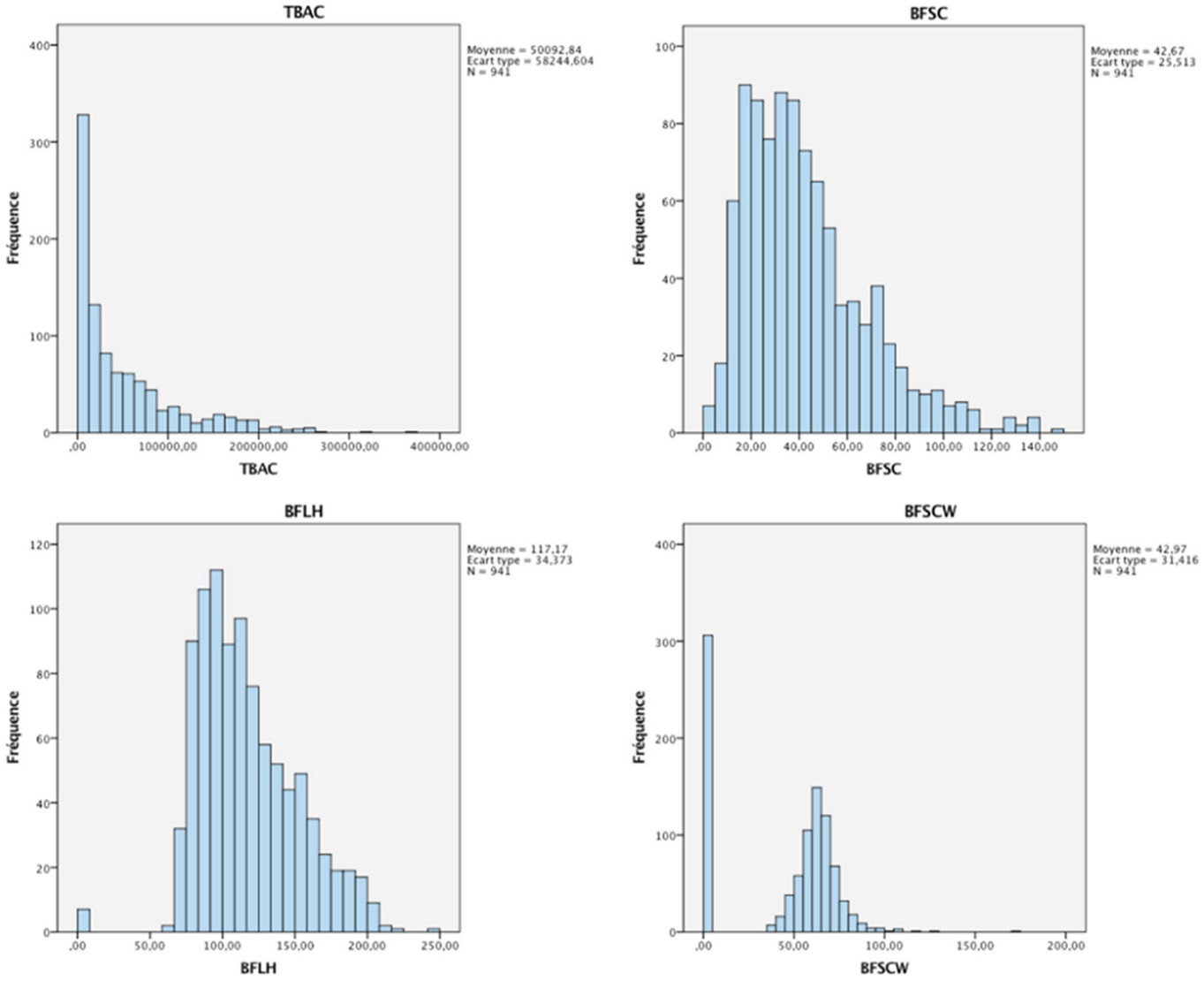

Fig. 3. Histogram of attributes

3.5 Machine learning algorithm

Random Forest algorithm. Random Forest is a machine learning algorithm (see Figure 4). It is a classifier that has been imported into the open-source software R via the RF package. This algorithm is based on three principles that make it very powerful [14].

- Firstly, the CART (Classification And Regression Tree).
- The second principle of the RF algorithm is Tree Bagging.
- Finally, the third principle of the algorithm is "Bootstrap."

The choice of the RF algorithm was made for the following reasons [19]:

- For classification, the RF algorithm will avoid the overlearning problem.
- For both classification and regression tasks, the same RF algorithm can be used.

- Lower chance of variance and overfitting of training data compared to DT since RF takes the average value from the outcomes of its constituent decision trees.
- Accuracy is higher compared to other algorithms such as DT or the naive Bayesian algorithm.
- Performs well on massive data.
- It can provide estimates of important variables or attributes in the classification.

To use the random forest algorithm in the R language, simply use the RF package. RF is a very good technique for fitting a more accurate model by averaging many DT, reducing variance, and avoiding the problem of overlearning trees. DT themselves perform poorly, but when used in conjunction with assembly techniques such as bbagging, and RF, their predictive performance is greatly improved.

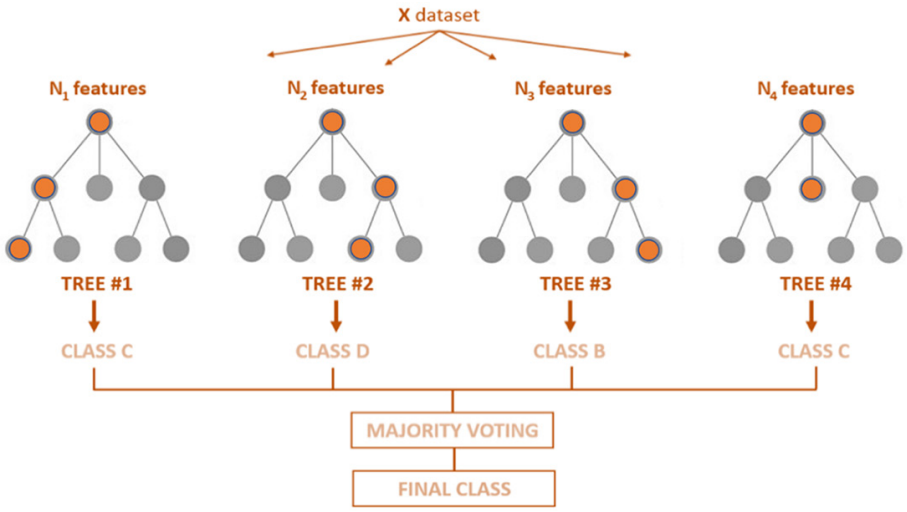

Fig. 4. Random Forests algorithm

Package "RandomForest". After reading the Comma-Seperated Value (CSV) file, the RF package was used to explain the "SHAPE_BACT" variable according to the other variables (see Table 1). The question to be answered is: which variables in our model best discriminate between the *Bacilli* form of bacteria and the *Cocci* form of bacteria?

The RF algorithm produces a considerable number of classification trees on a random fraction of the data to get the smallest possible out of bag (OOB). To minimize this value, two parameters were set: the number of trees constructed by the algorithm (ntree) and the number of variables tested at each division (mtry).

Choice of ntree and mtry. The method adopted is to choose the number of trees by keeping the default value of mtry and test several values by evaluating them, for example, with the following command, which displays a graph showing how to reduce the OOB according to the number of trees generated (see Figure 5).

```
plot(dataset_RandomForest$err.rate[, 1], type = "l", xlab = "Number of trees",
ylab = "OOB error")
```

The number of trees is then chosen when the value stabilizes at a minimum, the OOB value stagnates from 4500 trees, which is the case with a forest of 5000 trees.

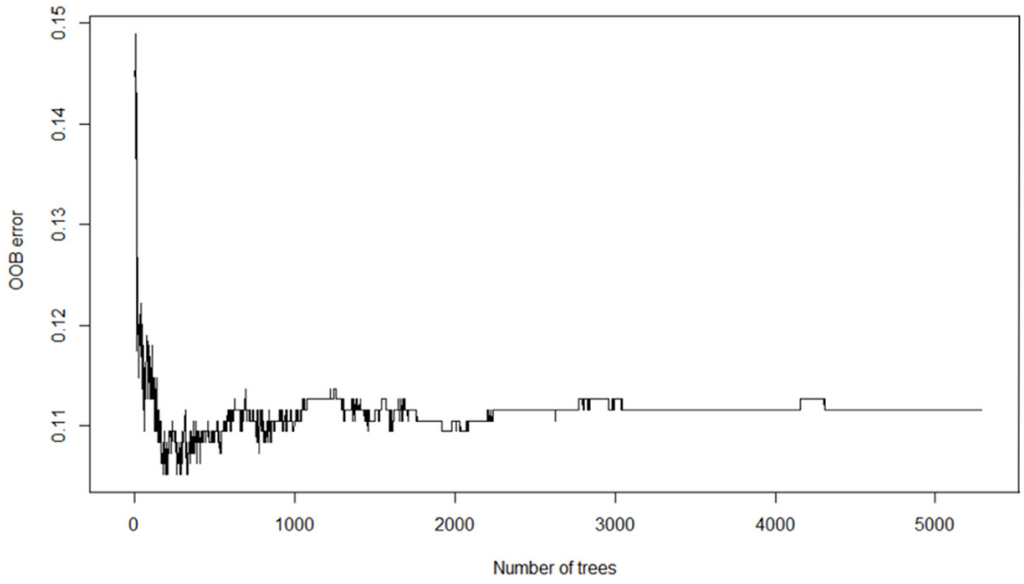

Fig. 5. The plot of OOB according to the number of trees generated

There are four explanatory variables in our model, we repeat the same method used for ntree while changing the value of mtry between 1 and 4. According to the results achieved, we have selected the value of 4 for the mtry variable since it allows us to have a stable OOB.

Out-of-bag estimate error. The content of the dataset_RandomForest object obtained is described in Figure 6:

```
Call:
 randomForest(formula = SHAPE_BACT ~ TBAC + BFSC + BFLH + BFSCW,
 data = myDataset, ntree = 4500, mtry = 4, na.action = na.roughfix)
               Type of random forest: classification
                     Number of trees: 4500
No. of variables tried at each split: 4
```

Fig. 6. Presentation of the content dataset_RandomForest

In the proposed prediction model, the confusion matrix of a two-class classifier, *Cocci*, and *Bacilli* (see Figure 7).

```
matrice_confiusion = dataset_RandomForest$confusion
colnames(matrice_confiusion) <- c("Bacilli", "Cocci", "Error (%)")
TP <- matrice_confiusion[1,1]
FN <- matrice_confiusion[1,2]
FP <- matrice_confiusion[2,1]
TN <- matrice_confiusion[2,2]
```

Fig. 7. Confusion matrix of a two-class classifier, *Cocci*, and *Bacilli*

All these measures can be calculated based on the confusion matrix (see Table 2). The command that can display the confusion matrix on Rstudio is dataset_RandomForest $confusion, it is a table showing the observed data in terms of rows and the data predicted by the algorithm in terms of columns. The OOB error rate of this model is 11.11% and *Bacilli* bacteria is better recognized by the algorithm with an error rate of 3%.

Table 2. Confusion matrix for Random Forest algorithm

	Predicted		
	Negative	**Positive**	
Negative	True Negative: Predicted No, and the target was Negative.	False Positive: Predicted Yes, and the target was Negative.	Specificity = 0.30 (1)
Positive	False Negative: Predicted Negative, and the target was Positive.	True Positive: Predicted Positive, and the target was Positive.	Sensitivity = 0.96 (2)
	Negative Predictive value	Precision = 0.91 (3)	Accuracy = 0.88 (4)

$$Specificity = \frac{TN}{TN + FP} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Classification of explanatory variables. Now we have a model that has a 3% error for the Bacilli shape. Next, we reveal the most discriminating variables that appear in our model for the recognition of the Bacilli and Cocci shapes. The following graph shows the importance of the explanatory variables for distinguishing the two shapes of bacteria (see Figure 8).

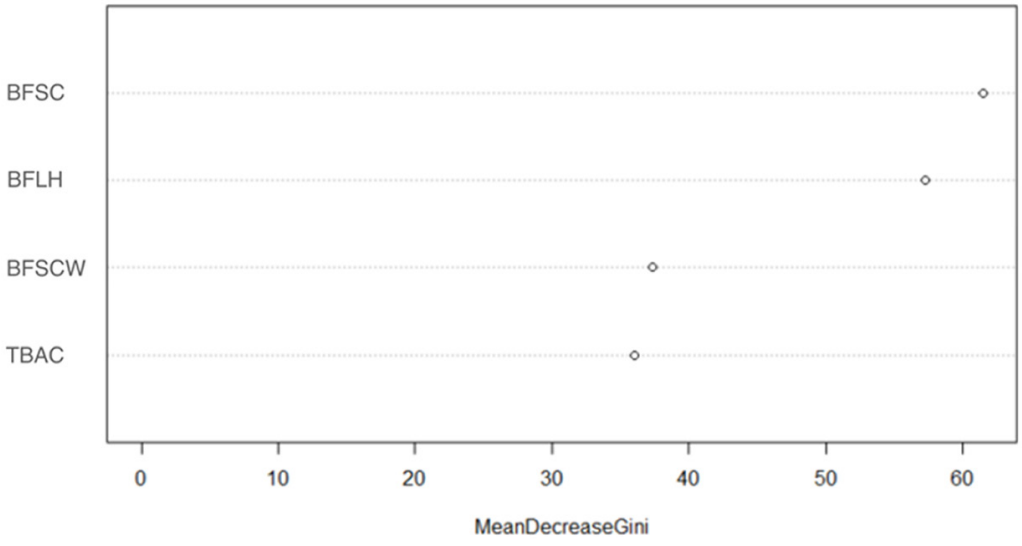

Fig. 8. The plot of the importance of explanatory variables

In this study, the two most important criteria for distinguishing between *Bacilli* and *Cocci* are the size/height of the particle (BFSC) and the level of staining of the particle nucleus (BFLH).

4 PROPOSED RECOMMENDER SYSTEM

4.1 Objective

The recommender system proposed in this work is articulated in two services integrated into a web-based information and communication platform, as shown in Figure 11:

- The first is the prediction service, whose main objective is to predict the shape of bacteria for biological requests of urine cytobacteriological examination.
- The second is the confirmation and correction service, its main objective is to validate or correct the bacterial form of a new case, from the urine cytobacteriological examination.

Therefore, the proposed recommender system is a powerful program for the prediction of the shape of bacteria in the UCBE examination (*Cocci* or *Bacilli*, see Figure 9) by offering the user validation and correction of the prediction, to improve the accuracy of the classification algorithm used and make it optimal in the future.

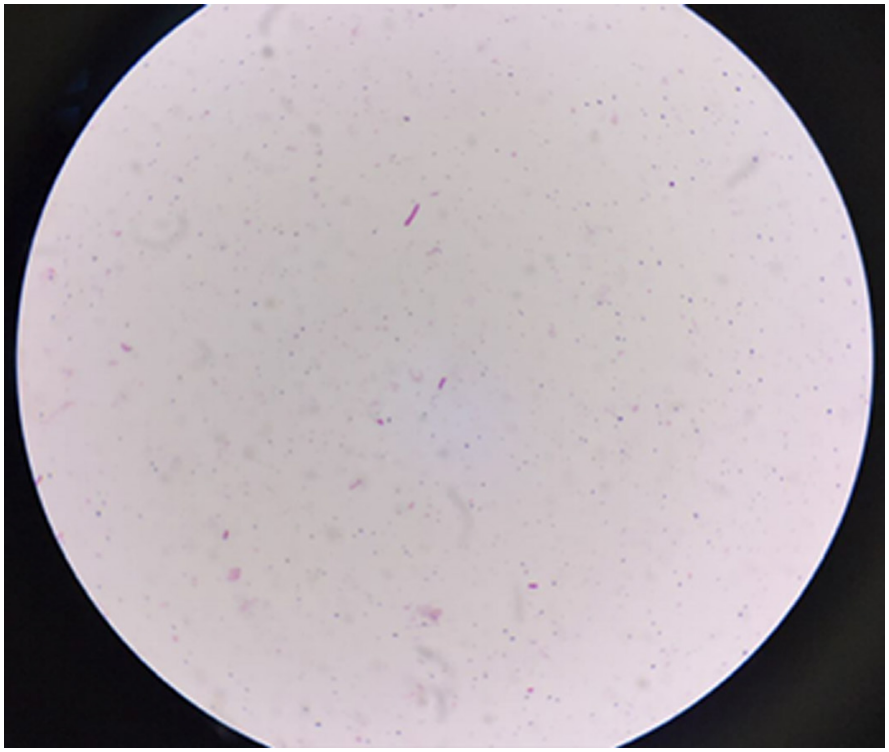

Fig. 9. The shape of bacteria in UCBE examination

4.2 Performance

The user can use a computer or a smartphone to access the system and consult its services in a dynamic and interactive web environment. Figure 10 shows the sequence diagram of the proposed recommender system. It presents the action of

predicting the shape of bacteria for the cytobacteriological examination of urine. Firstly, the user can access the system with a login and password on the authentication page, then he can enter a new case, and the system displays the result of the prediction. If the prediction is correct, the user will validate it otherwise, the user will correct the prediction, and the system will add the case to the database in a regular way.

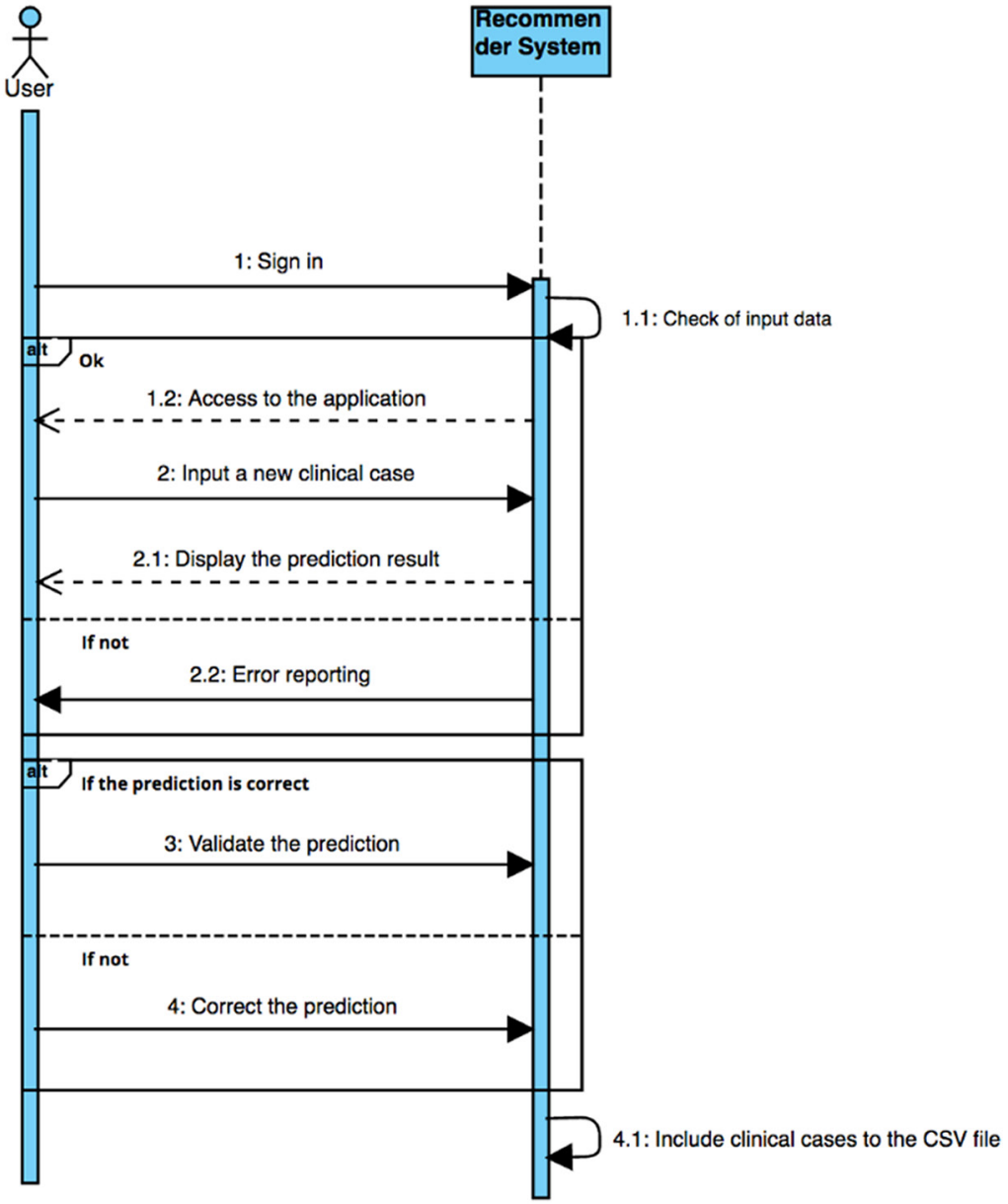

Fig. 10. Sequence diagram of the proposed recommender system

4.3 Interfaces

The following figure shows the graphical interface of the proposed recommender system (see Figure 11):

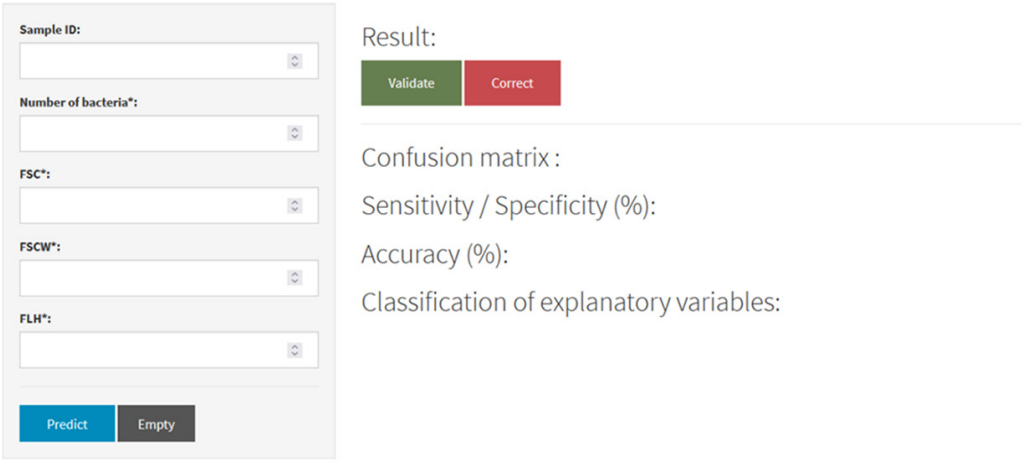

Fig. 11. Principal interface

As shown in Figure 11 the web application contains three main buttons:

- Button "Predict" allows the user to predict the bacterial form of a new case, by entering the form fields. The result of the prediction is displayed on the right below the title "Result" (see Figure 12).
- Button "Validate" allows the user to validate the prediction by adding the new case to the dataset.
- Button "Correct" allows the user to correct the prediction if it is erroneous by adding the new case with the inverse of the predicted class to the dataset.

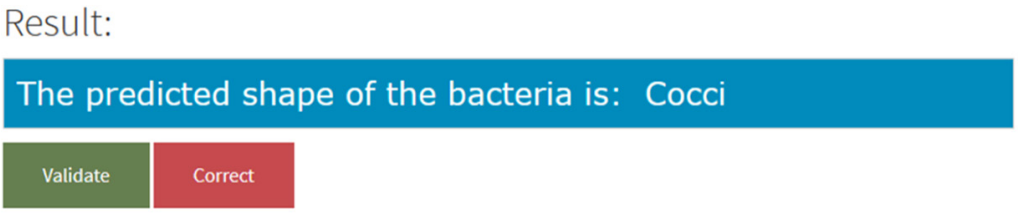

Fig. 12. Prediction result

Experimental results showed that the time required to identify the shape of bacteria was lessser, and *Bacilli* bacteria are better recognized by the algorithm with an error rate of 3%.

By adding each new record to the above-mentioned dataset (CSV file), the "Validate" and "Correct" buttons will improve the accuracy of the random forest algorithm developed for, among other things, predicting the shape of bacteria in urine cytobacteriological examination.

The rest of the interface contains the confusion matrix, sensitivity/specificity, precision, and importance of the explanatory variables to diagnose the state of progress of the algorithm after each validation or correction.

5 RELATED WORKS

In today's world, AI and recommendation systems play an important role in simplifying our personal and professional lives. Due to the rapidly growing interest in

the health field, several studies have been carried out by applying machine learning techniques for prediction, using both supervised and unsupervised machine learning algorithms.

The objective of this work was to reduce the time taken to identify the shape of bacteria (*Cocci* or *Bacilli*) in urine cytobacteriological examination using machine learning techniques and, more precisely, by applying the RF algorithm. Through this approach, a study was conducted to examine indoor environments in homes and schools for health outcomes using machine learning and logistic regression methods [20]. Using the RF approach, an exploratory study was conducted to predict optimal treatment regimens for cancer, which highlighted the role of machine learning in providing recommendations to specialists for selecting appropriate treatments that improve outcomes for breast cancer patients [21].

Since the declaration by the World Health Organization that COVID-19 is a global pandemic, several studies have been published to address this pandemic, presenting effective solutions for prevention and diagnosis using machine learning techniques [22] [23]. G. Grekousis et al. [24] worked on a geographical RF approach in which they demonstrated the importance of demographic, socioeconomic, and underlying health ranking factors on deaths due to COVID-19 in the United States. M. Adhikari and A. Munusamy [25] proposed iCovidCare, which is an intelligent health monitoring and prediction framework for COVID-19 using ensemble RF in edge networks. Interactive health services (eHealth and m-Health) play a very important role in facilitating communication between different actors in the health field [26] through the use of new information and communication technologies (see Figure 13).

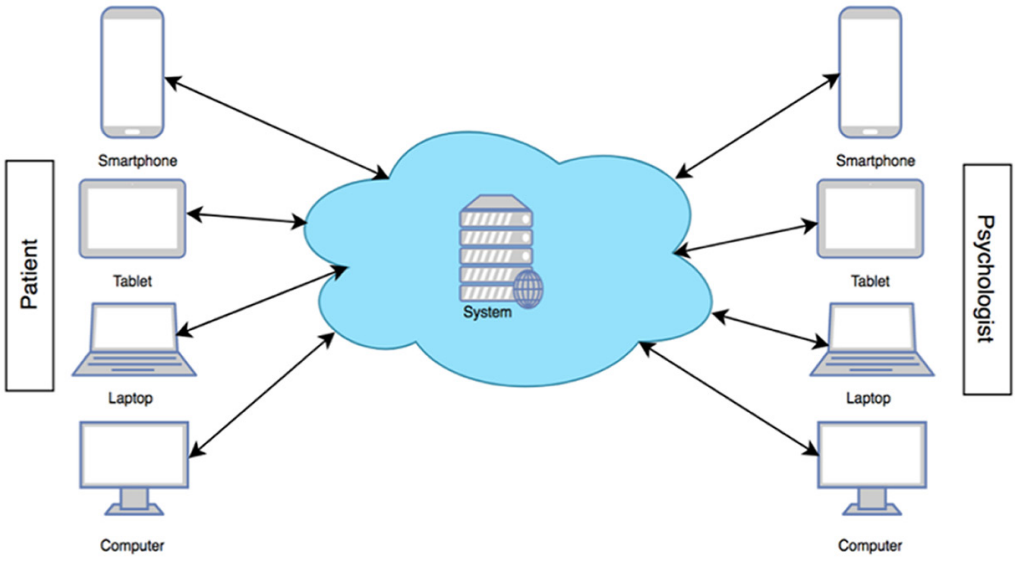

Fig. 13. eHealth platform architecture [27]

A recent study proposes the implementation of an interactive tool for dealing with burnout using Scratch programming code. It is a mobile application that offers several services: training, assessment, and coaching [28]. The third service contains a smart option that allows sending recommendations (adjustment strategies to cope with stress) to the user. However, the results of a recently published study [29] showed that the combination of the two prediction algorithms, RF and AdaBoost, achieved high prediction accuracy for heart disease. This means that the use of

recommender systems has a very important role in the prevention and improvement of the physical and psychological health of the individual.

6 CONCLUSION AND FUTURE WORK

This work aimed to develop a recommender system that predicts the shape of bacteria for biological requests, such as urine cytobacteriological examination, using machine learning techniques to reduce the time taken to identify the shape of bacteria (*Cocci* or *Bacilli*). Experimental results showed that the time needed to identify the shape of bacteria is decreased, and bacilli bacteria are better recognized by the algorithm with an error rate of 3%. In future work, we will work on improving the prediction results of this algorithm and generalizing the solution to other biological examinations.

7 REFERENCES

[1] I. E. Guabassi, Z. Bousalem, R. Marah, and A. Qazdar, "A recommender system for predicting students' admission to a graduate program using machine learning algorithms," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 2, pp. 135–147, 2021. https://doi.org/10.3991/ijoe.v17i02.20049

[2] P. Simon and T. Moulin, "Chapitre 12 – Intelligence artificielle médicale," *Télémédecine et Télésoin*, Éd. Paris: Elsevier Masson, pp. 65–75, 2021. https://doi.org/10.1016/B978-2-294-77544-4.00012-4

[3] E. Negre, "Systèmes de recommandation: Introduction," ISTE Group, 2015.

[4] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Model User-Adap Inter.*, vol. 12, no. 4, pp. 331–370, 2002. https://doi.org/10.1023/A:1021240730564

[5] D. Callegarin and P. Callier, "Enjeux du déploiement de l'intelligence artificielle en santé," *Actualités Pharmaceutiques*, vol. 60, no. 611, pp. 21–24, 2021. https://doi.org/10.1016/j.actpha.2021.10.006

[6] B. Collet, G. Nebout, G. Liabeuf, B. Fantino, D. Ogier-Peronnet, and J. C. Mabriez, "Infection urinaire en ville : enquête sur le diagnostic et le traitement. 2ème partie : examen cytobactériologique des urines en pratique courante," *Medecine et Maladies Infectieuses*, vol. 10, no. 26, pp. 817–821, 1996. https://doi.org/10.1016/S0399-077X(96)80061-8

[7] J. K. Bajaj, R. P. Karyakarte, J. D. Kulkarni, and A. B. Deshmukh, "Changing aetiology of urinary tract infections and emergence of drug resistance as a major problem," *J Commun Dis.*, vol. 31, no. 3, pp. 181–184, 1999.

[8] M. J. Piñero Acín, M. R. Martínez Matías, M. J. Córdoba Gutiérrez, T. Sánchez Rojas, D. López Juárez, and F. J. Rodríguez Alcalá. "Do urocultures change our therapy approach?" *Aten Primaria*, vol. 26, no. 7, pp. 459–463, 2000. https://doi.org/10.1016/S0212-6567(00)78703-8

[9] X. Zou, Y. Hu, Z. Tian, and K. Shen, "Logistic regression model optimization and case analysis," In *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, Dalian, China, pp. 135–139, 2019. https://doi.org/10.1109/ICCSNT47585.2019.8962457

[10] J. R. Brzezinski and G. J. Knafl, "Logistic regression modeling for context-based classification," In *Tenth International Workshop on Database and Expert Systems Applications. DEXA 99*, pp. 755–759, 1999. https://doi.org/10.1109/DEXA.1999.795279

[11] H. Zhang and D. Li, "Naïve bayes text classifier," In *2007 IEEE International Conference on Granular Computing (GRC 2007)*, pp. 708–708, 2007. https://doi.org/10.1109/GrC.2007.40

[12] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991. https://doi.org/10.1109/21.97458

[13] A. Liaw and M. Wiener, "Classification and regression by randomForset," *R News*, vol. 2/3, pp. 18–22, 2002.

[14] E. Roblin, P. Cournède, and S. Michiels, "Prédiction de survie avec réseaux de neurones et mesures d'incertitude," *Revue d'Épidémiologie et de Santé Publique*, vol. 70, no. 2, pp. 84–85, 2022. https://doi.org/10.1016/j.respe.2022.03.116

[15] E. Roblin, P. Cournède, and S. Michiels, "Développements de réseaux de neurones pour données censurées et prédiction de survie," *Revue d'Épidémiologie et de Santé Publique*, vol. 69, no. 1, pp. 9–10, 2021. https://doi.org/10.1016/j.respe.2021.04.011

[16] C. Laurent, "Intelligence artificielle : une mine d'or pour les entreprises," *Gestion*, vol. 42, no. 1, pp. 76–79, 2017. https://doi.org/10.3917/riges.421.0076

[17] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pp. 1255–1260, 2019. https://doi.org/10.1109/ICCS45141.2019.9065747

[18] Y. Wolmark, "Évaluer la qualité," *Gerontologie et societe*, vol. 24, no. 99, pp. 131–146, 2001. https://doi.org/10.3917/gs.099.0131

[19] S. Uddin, A. Khan, H. md Ekramul, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Medical Informatics and Decision Making*, vol. 19, p. 281, 2019. https://doi.org/10.1186/s12911-019-1004-8

[20] Ziqiang Lin, Shao Lin, Iulia A. Neamtiu, Bo Ye, Eva Csobod, Emese Fazakas, and Eugen Gurzau, "Predicting environmental risk factors in relation to health outcomes among school children from Romania using random forest model – An analysis of data from the SINPHONIE project," *Science of The Total Environment*, vol. 784, 2021. https://doi.org/10.1016/j.scitotenv.2021.147145

[21] Z. Cui, Z. Kadziola, D. E. Faries, I. Lipkovich, B. Ratitch, X. Li, K. Sheffield, and G. Cuyun Carter, "MS3 predicting optimal treatment regimens for HR+/HER2-breast cancer based on electronic health records using random forest," *Value in Health*, vol. 23, pp. 8–9, 2020. https://doi.org/10.1016/j.jval.2020.04.046

[22] Kanika Goel, Renuka Sindhgatta, Sumit Kalra, Rohan Goel, and Preeti Mutreja, "The effect of machine learning explanations on user trust for automated diagnosis of COVID-19," *Computers in Biology and Medicine*, vol. 146, 2022. https://doi.org/10.1016/j.compbiomed.2022.105587

[23] S. Lafraxo and M. el Ansari, "CoviNet: Automated COVID-19 detection from X-rays using deep learning techniques," In *6th IEEE Congress on Information Science and Technology (CiSt)*, pp. 489–494, 2021. https://doi.org/10.1109/CiSt49399.2021.9357250

[24] George Grekousis, Zhixin Feng, Ioannis Marakakis, Yi Lu, and Ruoyu Wang, "Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: A geographical random forest approach," *Health & Place*, vol. 74, 2022. https://doi.org/10.1016/j.healthplace.2022.102744

[25] M. Adhikari and A. Munusamy, "iCovidCare: Intelligent health monitoring framework for COVID-19 using ensemble random forest in edge networks," *Internet of Things*, vol. 14, 2021. https://doi.org/10.1016/j.iot.2021.100385

[26] L. Gossec, A. Cantagrel, M. Soubrier, J-M. Berthelot, J-M. Joubert, B. Combe, W. Czarlewski, D. Wendling, E. Dernis, L. Grange, C. Beauvais, A. Perdriger, H. Nataf, M. Dougados, and H. Servy, "Sanoïa®, plateforme e-santé interactive d'auto-évaluation dans la polyarthrite rhumatoïde. Essai comparatif randomisé de 12 mois sur 320 patients," *Revue du Rhumatisme*, vol. 86, no. 1, 2019. https://doi.org/10.1016/j.rhum.2018.08.002

[27] M. A. Lafraxo, M. Ouadoud, Y. El Madhi, and A. Soulaymani, "Burnout syndrome among nursing staff: Performing data analysis using the SPSS statistic," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 7, no. 4, pp. 11, 2021. https://doi.org/10.3991/ijoe.v17i04.20979

[28] M. A. Lafraxo, M. Ouadoud, Y. El Madhi, M. Rehali, and A. Soulaymani, "Burnout syndrome prevention measures among nursing staff: Implementing a mobile application based on MIT's app inventor tool using the scratch programming code," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 7, no. 4, pp. 81–95, 2021. https://doi.org/10.3991/ijoe.v17i04.20393

[29] H. El Hamdaoui, S. Boujraf, N. El H. Chaoui, B. Alami, and M. Maaroufi, "Improving heart disease prediction using random forest and AdaBoost algorithms," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 11, pp. 60–75, 2021. https://doi.org/10.3991/ijoe.v17i11.24781

8 AUTHORS

Mohammed Amine Lafraxo is a Ph.D. Research Scholar in Health Sciences, Software Engineering, E-learning and Medical Informatics. He is currently working at the Laboratory of Biology and Health, the Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco.

Hinde Hami is a Ph.D. Research Scholar in Human Population Genetics. She is currently working at the Laboratory of Biology and Health, the Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco.

Tarik Merrakchi is a Data Engineer at Bpifrance. He has more than 7 years of professional experience in Big Data and Information System projects in different sectors. In 2017, he got engineering degree in information systems at SupMTI Oujda, Morocco. He has a Master's degree in Big Data & Cloud Computing from the Faculty of Sciences Aïn Chock of Casablanca, Morocco.

Ali Azghar is a Ph.D. student at the Faculty of Sciences, Mohammed First University, Oujda, Morocco. He is a graduate of the Faculty of Medicine and Pharmacy, Morocco, Oujda.

Ahmed Remaida is a Ph.D. student at the Department of Informatics, Logistics, and Mathematics at the National School of Applied Sciences at Ibn Tofail University in Kenitra. He has a Master's degree in Educational Technology from the "École Normale Supérieure, ENS" of Martil, Morocco. His research interests include the use of machine learning in an educational context.

Mohammed Ouadoud is a Ph.D. Research Scholar in Computer Sciences, at the Laboratory of Information System and Software Engineering (SIGL) at the National School of Applied Sciences, Abdelmalek Essaâdi University, Tetouan, Morocco. In 2018, he completed his Ph.D. thesis in computer science at the Faculty of Science in Tetouan, Morocco. His dissertation research focuses on modeling and prototyping a learning management system based on the IMD-LD, the NoSQL, and the hybridization of learning theories. He received a Master's degree in Instructional Design and Multimedia Engineering from the École Normale Supérieure of Martil, Morocco, in 2013. His current research focuses on IT and software engineering. He is a reviewer for several international journals.

Adil Maleb is a Research Scholar with the Mohammed First University, Faculty of Medicine and Pharmacy, Morocco, Oujda.

Abdelmajid Soulaymani is a Research Scholar, and is working at the Laboratory of Biology and Health at the Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco.