International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

December 2022 | Issue 53 | Volume 14 | Number 4 | Pages 25-34

INTELLIGENT SYSTEM FOR REMOTE DISEASES DIAGNOSIS

K. Slime [1] A. Maizate [1] O. Bouaissi [1] A. Hilmani [1] C. Mahmoudi [1] N. Mouine [2]

1. RITM-ESTC / CED-ENSEME, University of Hassan II, Casablanca, Morocco, khaoulaslime@gmail.com, maizate@hotmail.com, oumaima.bouaissi97@gmail.com, adilhilmani@gmail.com, charif.mahmoudi.us@gmail.com
2. Cardiology B Department, Ibn Sina Hospital Center, Mohamed V University, Rabat, Morocco, mouine2@yahoo.fr

Abstract- The restrictions introduced as mitigation for the novel coronavirus (COVID-19) are required measures that hurts the global economy. The need of social distancing and the quarantines, as two of the major measures, could be drastically reduced by the adoption of novel approaches that foster the technology to implement alternative that will ease the burden on the world's economy. In this paper we argue that the targeted instrumentation of high-risk citizen with monitoring system can provide reasonably accurate real-time results to let the low-risk population to resume its activity. We present CovidTrack, an intelligent system of deep learning that delivers a high accuracy (95%) and enables a practitioner-in-the-loop capability by leveraging mobile technologies. We describe the proposed system and expose the results of the study where we focused on the persons with heart-related precondition.

Keywords: COVID-19, Prediction, Deep Learning, Machine Learning.

1. INTRODUCTION

Since December 2019 coronavirus 2 has not stopped spreading all over the world, and it is affecting 213 countries and territories around the world. More than 922,212 deaths and more than 29,185,779 confirmed cases [1] and surely, they are more infected people. A lot of countries suffer from the large outbreaking of the coronavirus and healthcare systems are enabled to cope with this situation. The big challenge with this virus is how to stop outbreaking and limit infection. Early detection seems to be the best solution for this big dare. To ease the charge on the healthcare system, IT and especially AI (Artificial Intelligence) intervention is needed. Multiple studies have been done to diagnose and give predictive models so that relevant COVID-19 research results can be shared quickly and openly to apprize and guide the public health committee answer and help save many lives [2]. In fact, noticed that the allocation of healthcare resources in the context of the pandemic is limited, various prediction models that merge several variables or features to estimate the risk that a person will be infected have been put in place to help medical staff in the triage of patients. [3].

Testing is a crucial element of the fight against COVID-19. The availability of testing in high-risk areas according to Peto, et al. [4] is a key for the exit from the lockdown restrictions. Whereas most of the population in high-risk zone are asymptomatic, the transmission rate still high to the individuals with pre-condition significantly high [5]. The early diagnosis showed a significant reduction on the lethality in patients with precondition [6]. In this paper, we focus in patients with cardiovascular-related preconditions, especially as diagnosis techniques are reasonably mature and have been studied closely by the National Institutes of Health (NIH) [7] and the National Health Commission of China [8].

Towards the early direction, Song et Al. develop a COVID-19 precocious alert score (COVID-19 EWS) as a multiparameter screening tool to pick out heavily suspected people [9] based on multivariate logistic regression. Mbida et Al. use artificial intelligence to develop smart system for quick analysis and detection [10]. A new sort tool of Artificial Intelligence diagnosis system for Suspected COVID-19 in fever clinics is proposed by Feng et Al [11]. Deep learning is also used to diagnose and identify the virus SARS-CoV-2 from the viral genome sequence [12-13]. The hospital stay of COVID-19 patients can be predicted by analyzing CT images with machine learning algorithms [14] and also using chest radiographs [15]. Some choose to pursue the progression of COVID-19 (sever or critical) which helps to determine the treatment priority [16] and many others estimate the risk of mortality in patients suspected or confirmed with COVID-19 [17-18-19-20-21-22-23].

Our approach aims to provide a comprehensive detection solution composed by an interactive mobile application backed by an intelligent backend system. Odeh et AL. also present a work using mobile application with artificial intelligence to increase the efficiency of healthcare services [24].

The mobile application has three main features: enables data collection from the patients and notifies the patient when if the infection prediction leans towards this conclusion; enables the practitioners to input information following a consultation; and interact with the intelligent backend. The intelligent backend system is composed from a series of machine learning models to predict the potential risk of being infected with COVID-19.

We focused on the major symptoms as diagnostic criteria of the disease to calculate risk score with mobile e application and then predict with IA models the probability of being positive patient. The label of our data is 'corona result' and it can take three values: High risk, medium risk, or low risk. The random forest has the highest accuracy (95%) followed by the Decision tree, KNN and naive Bayes (85%) and the last one is Logistic regression (75%). Moreover, deep learning algorithms give us an accuracy of 95%. To achieve these results, we followed the CRISP-DM data analysis process which is a description of common approaches of a standard process model often used by data mining experts [25]. It defines a set of instructions and recommendations that aid to set out, plan, and implement the data analysis or data mining projects, it is well explained by Figure 1 diagram.

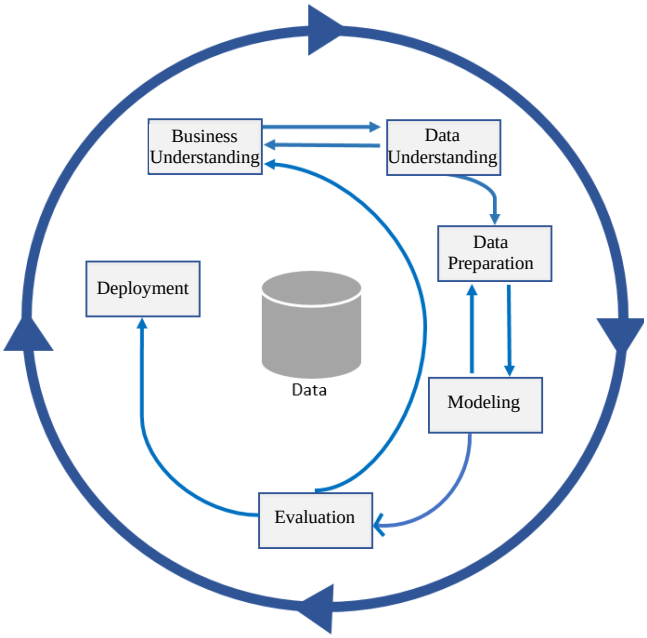

Figure 1. Cross Industry Standard Process for Data Mining [26]

In this work we propose the following contributions:
- An innovative approach for early diagnosis of the COVID-19 in patients with cardiovascular preconditions.
- An intelligent system that leverages deep learning techniques alongside sensing and mobile technologies to enable an accurate prediction of being infected SARS-CoV-2 virus. The reminder of this paper is structured in two sections.

2. METHODOLOGY

The objective of this section is describing the methodology that we practiced to build the intelligent backend, our methodology is based on the CRISP-DM (Cross Industry Standard Process for Data Mining) process where the first step is the business understanding. We define the problematic as:

How to strengthen the diagnosis predictions to precede the observable individual symptoms?

Patient symptomatology can be a simple cold or a lung infection responsible for acute respiratory distress. Cardiac patients are more threatened with this virus and as much as fast it's detected the probability of saving their life increase.

That's why we think about the development of an e-health system: CovidTrack, which provide is telemedicine solution based on AI methods which help to pursue and predict the risk of having COVID-19.

2.1. Data Understanding

The data understanding process, as the second step of CRISP-DM process is crucial to the enhancement of the accuracy of the predictions. This sub-section describes our refinement process where we first identify the initial set of data. Secondly, we analyse the structure of the collected dataset. Last, we check the soundness of the dataset according important properties.

2.1.1. Collection of Initial Data

Our search criteria were centered on the size and the openness of the data source. As our attempt to obtain local data were significantly delayed by the administrative process, we were able to obtain dataset containing information about patient and the different COVID-19 symptoms. We searched for this in different data science platform such as: google dataset and Kaggle.

We begin our search with the global and general term 'COVID-19' we get about 1140 results. This number is less than what we pretend to get after six months of the first apparition of the coronavirus. The majority of the datasets founded are about statistics such as the number of deaths, number of infected peoples and also about the propagation of the COVID-19 on different countries [27]. After this we reduce our search with 'COVID-19 prediction' terms which give us 39 results but we didn't find among them significant data for our study. For example, one dataset gives only general data about the COVID-19. For more precision we continue our search with 'COVID-19 symptoms' & 'COVID-19 symptoms' and we get 59 datasets which seems to be very promising as results. However, it was not the case. The majority of these datasets didn't represent our research terms and many others datasets are destined for different aims and didn't contain what we looked for which captures the date of symptom onset for a sample of confirmed US coronavirus cases, which contain data on hospital bed capacity. It's a collection of 306 U.S. hospital markets, as well as data for nine different models of COVID-19 infection scenarios. Only 6 datasets from the 59 results are suitable for our research. After preprocessing we discover that the first and second one and the third one has information about a few symptoms and in our study, we need re parameters for our model. The fourth one describes each symptom as one line of the dataset s we fund a lot of duplicated lines with the same patient information.

The fifth dataset has only positives COVID-19 cases and didn't treat negative ones. That's why we finally work with the last one witch even if it has only 127 lines but it's the most adapted to our needs. In fact, it contains all the symptoms of coronavirus and significant data.

The dataset used in this paper is taken from Kaggle and it contains exactly what we need as parameters for our AI platform to predict the risk of COVID-19.

2.1.2. Data Analysis

The initial dataset allows to classify the risk of COVID-19 in a person according to a set of symptoms. It contains 127 rows and 21 columns. The line represent information about patients and columns represent ticket of each line and contains general information such as : patient id, age and gender and COVID-19 symptoms and information: Body temperature, Dry Cough, Sour throat, weakness, Breathing problem, Drowsiness, Pain in chest, Travel history to infected countries, Diabetes, Heart disease, Lung disease, Stroke or reduced immunity, Symptoms progressed, High blood pressure, Kidney disease, Change in appetite, Loss of sense of smell, Corona result. After being analyzed by the project team doctor it was clear that the big advantage of this dataset is has it contains the major considered symptoms of the coronavirus which are described at the second section (Table 1).

In the dataset, for the Symptoms two values are possible:
0 → Not present (this patient didn't have this symptoms)
1 → Present (this patient has this symptom)

For the risk Column three values are possible:
0 → Low risk to be a positive case
1 → Medium Risk to be a positive case
2 → High Risk to be a positive case

In this part we examinate our dataset (its columns and their potential values) we conclude that our dataset contains the most frequent symptoms and the majority of its columns have 0 or 1 as value so we can normalize it with this interval.

Table 1. Dataset columns

Column	Description
Sno	Patient id
Age	Age
Gender	Gender (male or female)
Body temperature	Temperature in Fahrenheit
Dry Cough	Symptom can have 0 or 1 as value
Sour throat	Symptom can have 0 or 1 as value
weakness	Symptom can have 0 or 1 as value
Breathing problem	Symptom can have 0 or 1 as value
Drowsiness	Symptom can have 0 or 1 as value
Pain in chest	Symptom can have 0 or 1 as value
Travel history to infected countries	Indicate if patient have been in contact with infected peoples or no can have 0 or 1 as value
Diabetes	Symptom can have 0 or 1 as value
Heart disease	Indicate if patient had heart problem. Can have 0 or 1 as value
Lung disease	Symptom can have 0 or 1 as value
Stroke or reduced immunity	Indicate if patient have immunity problems. Can have 0 or 1 as value.
Symptoms progressed	Give history of symptoms evolution
High blood pressure	Symptom can have 0 or 1 as value
Kidney disease	Indicate if patient have kidney diseases. Can have 0 or 1 as value.
Change in appetite	Symptom can have 0 or 1 as value
Loss of sense of smell	Symptom can have 0 or 1 as value
Corona result	Present the risk of being a positive case. Can have three values: low risk, medium risk and high risk.

2.1.3. Data Quality Verification

In this part we have followed the global verification steps:

- Check the data volume and examine its raw properties.
- Understand the meaning of each attribute and its value in business terms.
- For each attribute, calculate basic statistics (for example, distribution, mean, max, min, standard deviation, variance, mode, skewness).

Table 2. Sample of data

Sno	1	2	3	4	5
Age	20	19	55	40	33
Gender	1	1	0	0	1
Body temperature	98.6	99.0	102.0	100.0	99.2
Dry Cough	0	1	1	0	0
Sour throat	0	0	1	0	1
weakness	0	0	1	0	0
Breathing problem	0	0	1	0	1
Drowsiness	0	0	1	1	0
Pain in chest	0	0	1	1	0
Diabetes	0	0	0	0	0
Heart disease	0	0	0	0	0
Lung disease	0	0	0	0	1
Stroke or reduced immunity	0	0	0	0	1
Symptoms progressed	0	0	0	1	1
High blood pressure	0	0	0	1	0
Kidney disease	0	0	0	0	0
Change in appetite	0	0	1	1	0
Loss of sense of smell	0	0	0	0	1
Corona result	0	0	1	2	2

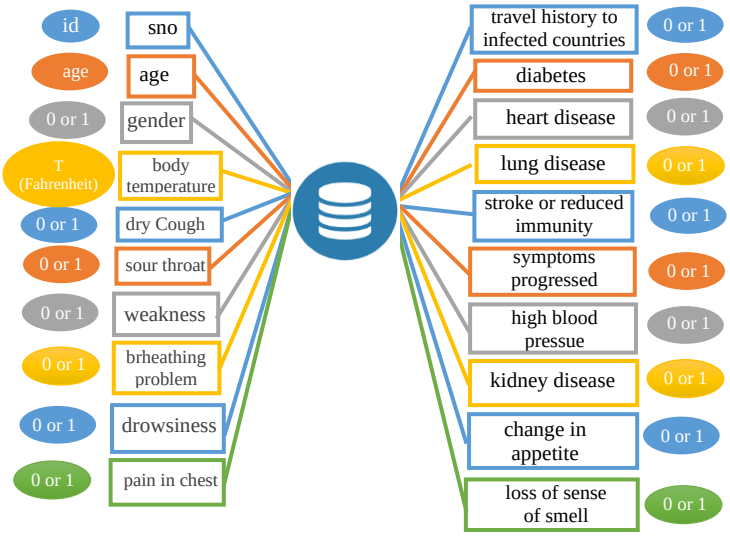

Figure 2. Data comprehension

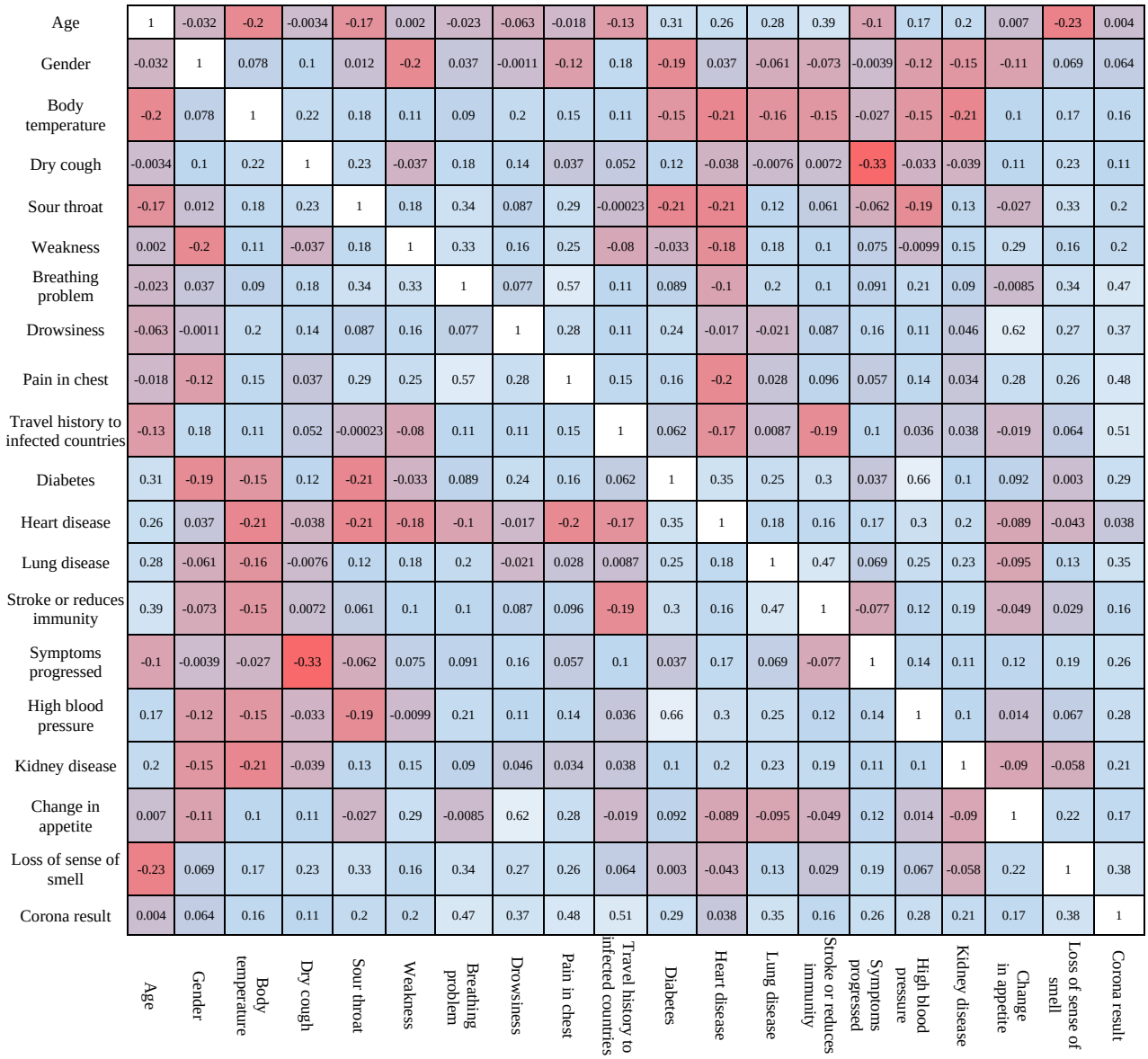

Table 3. Data statistics

Line Number	0	1	2	3	4	...	190	191	192	193	194
age	20	19	55	40	33	...	33.883516	40.155466	25.97856	37.539692	40.413249
gender	1	1	0	0	1	...	0.961172	0	1	0	0
body temperature	98.6	99	102	100	99.2	...	98.739781	99.828987	99.73384	100.07938	99.545426
dry Cough	0	1	1	0	0	...	1	0.844534	0	1	0
sour throat	0	0	1	0	1	...	0.038828	1	1	0.539692	0.413249
weakness	0	0	1	0	0	...	0.038828	1	0.74732	0.460308	0.413249
breathing problem	0	0	1	0	1	...	0.038828	1	1	1	0.413249
drowsiness	0	0	1	1	0	...	0.961172	0.844534	0.74732	0.539692	0.586751
pain in chest	0	0	1	1	0	...	0.038828	0.844534	1	0.539692	0.586751
travel history to infected countries	0	0	0	1	1	...	1	0	1	1	0.586751
diabetes	0	0	0	1	0	...	1	1	0	1	1
heart disease	0	0	0	0	0	...	0.961172	0.155466	0	0	0.413249
lung disease	0	0	0	0	1	...	0	1	0	0	0.413249
stroke or reduced immunity	0	0	0	0	1	...	0	0.844534	0	0	0
symptoms progressed	0	0	0	1	1	...	1	0.155466	0.74732	0.539692	1
high blood pressure	0	0	0	1	0	...	0.961172	1	0	1	1
kidney disease	0	0	0	0	0	...	0	0.155466	0.74732	0	0.413249
change in appetite	0	0	1	1	0	...	0.961172	0.844534	0.74732	0	0.586751
loss of sense of smell	0	0	0	0	1	...	1	1	1	0.460308	0.413249

2.1.4. Data Preparation

Actually, after understanding the content of our dataset and the various rows and columns parameters and values we are treating, we set it for the analysis and the modeling phases. Transferring data is the phase that takes the most time and also the least productive part but arguably the most significant section of the data mining process - As it is known: "garbage in, garbage out!".

This step of the CRISP-DM process may be divided into the three sub-steps:

• **Data Cleaning:** In this step we will choose how we can manage outliers, missing values, and other non-used data recognized in the precedent step.

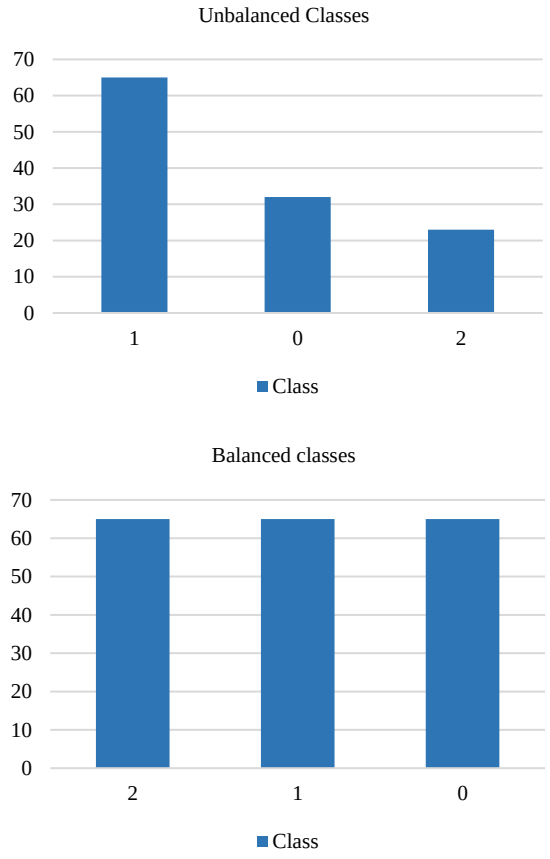

Figure 4. Unbalanced and balanced Dataset

Within our work, the dataset didn't contain any missed values and has only six duplicated lines. To enhance the quality of our data we delete duplicate, redundant, and all strongly correlated features.

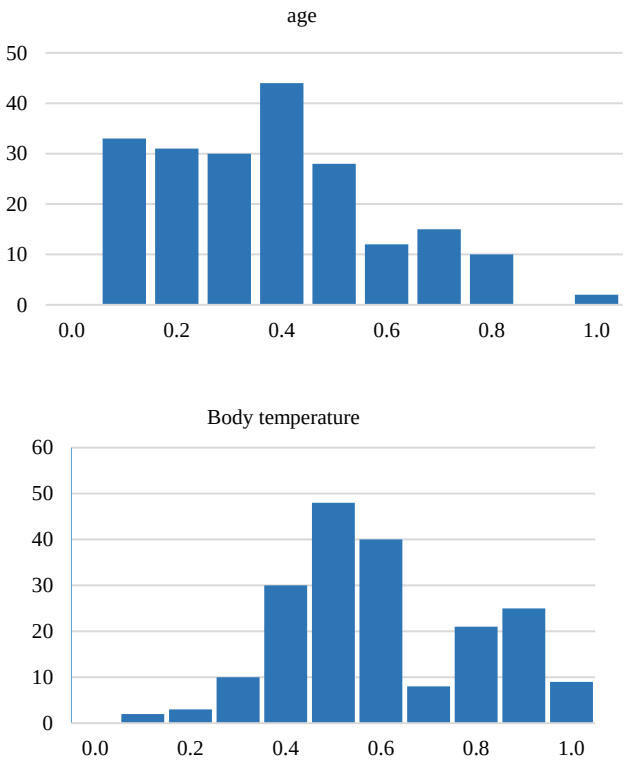

Figure 5. Normalizing age and body temperature data

• **Feature Scaling:** To normalize the range of independent variables or features of data we use the feature scaling method. In our dataset, the majority of parameters has 0 or 1 as value except age which is in years and body temperature which is in Fahrenheit.

To normalize these two parameters, we used Min-Max Scaling technique. The transformation is done using the following formula:

$$\text{Transformed.Value} = \frac{\text{Values} - \text{Mean}}{\text{Maximum} - \text{Minimum}} \quad (1)$$

After the normalization of our data all the feature values are included in a fixed interval [0, 1].

	age	gender	body temperature	Dry Cough	sour throat	weakness	breathing problem	drowsiness	pain in chest	travel history to infected countries	diabetes	heart disease	lung disease	stroke or reduced immunity	symptoms progressed	high blood pressue	kidney disease	change in appetide	Loss of sense of smell
0	0.032258	1.0	0.366197	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.016129	1.0	0.422535	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.596774	0.0	0.845070	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
3	0.354839	0.0	0.563380	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0
4	0.241935	1.0	0.450704	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0

Figure 6. Normalized dataset

• **Features Selection:** This process helps us to select a subset of suitable characteristics (variables, predictors) to construct our model.

To calculate features importance in our dataset, we worked with embedded method: we created a random forest classifier, trained the classifier and then printed results which are described in Figure 7.

Afterward, we create a picker object that use the random forest classifier to recognize features that have an importance of more than 0.03, trained the selector and printed results reported in Figure 8.

```
('age', 0.08329519452869709)
('gender', 0.04844310181860412)
('body temperature', 0.07915472932227977)
('Dry Cough', 0.030232099863868597)
('sour throat', 0.035672069145880926)
('weakness', 0.029121458240738366)
('breathing problem', 0.09952062239747847)
('drowsiness', 0.07292404500911387)
('pain in chest', 0.10844556032939802)
('travel history to infected countries', 0.11015544450709924)
('diabetes', 0.03388672934049531)
('heart disease', 0.015862307984062043)
('lung disease', 0.04817613657053333)
('stroke or reduced immunity', 0.031281597868061055)
('symptoms progressed', 0.040318335115133296)
('high blood pressure', 0.03240594663869564)
('kidney disease', 0.020261999410858787)
('change in appetite', 0.021169121510801064)
('Loss of sense of smell', 0.05967350039820053)
```

Figure 7. Dataset features values

- Age
- Gender
- body temperature
- dry cough
- breathing problem
- drowsiness
- pain in chest
- travel history to infected countries
- diabetes
- lung disease
- stroke or reduced immunity
- symptoms progressed
- high blood pressure
- loss of sense of smell

Figure 8. Most important dataset features

We notice that after testing our random forest classifier on the new limited dataset (which is constructed with only important feature) the accuracy of our model was 0.79 however the same classifier applied to the original dataset gave us an accuracy of 0.84.

Hence, we decided to work for the rest of our study with the whole features.

2.2. Modeling

This phase is in a big interaction with the data preparation phase. It consists of the selection and application of data mining techniques by adjusting their parameters to get optimal results. We choose to use multiples algorithms and we took our time to set the different settings and inputs data for each one to get the best response.

To build this phase of our project we followed these steps:

1. select the modeling techniques which meet our objectives.
2. Test method determination.
3. Model construction.

2.2.1. Select Modeling Techniques which Meet Objectives

We used supervised machine learning algorithms as well as a deep learning solution to achieve our goals: The Random Forest, Decision Tree, K-Nearest Neighbors (KNN), SVM and neural network, Naive Bayes.

2.2.2. Test Method Determination

It's often boils down to separating the data into a training set and a test set. As our dataset contains just 127 rows, we choose Stratified Random Sampling to split the data set according to the following percentages:

• 80% for the training set.
• 10% for the entire Test.
• 10% for the validation set.

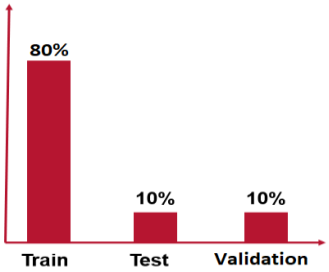

Figure 9. Data splitting

2.2.3. Model Construction

As said before we choose to apply six different algorithms to our dataset. In the following subsections we will explain each one of them:

Bayesian networks

It is founded on Bayes' theorem that determines the conditional probability according to Equation (2).

$$P(A_i / B) = P(A_i) \times P(B / A_i) / \sum_j P(A_j) \times P(B / A_j) \quad (2)$$

KNN

The K-Nearest Neighbours is a classic classification algorithm that relies exclusively on the selection of classification metric. It is "non-parametric" (just k must be set) and is based on training data

Decision tree

A decision tree is a graphical representation for a decision tool as a set of choice tree:

• The 1st classifier: DecisionTreeClassifier (*criterion* = "gini", *random_state* = 100, *max_depth* = 4, *min_samples_leaf* = 5).
• The 2nd classifier: DecisionTreeClassifier (*criterion* = "entropy", *random_state* = 100, *max_depth* = 4, *min_samples_leaf* = 5).

SVM (Support Vector Machine)

SVMs are a group of machine learning algorithms that resolve regression and classification problems. It helps to resolve linear and non-linear and non-linear problems. Idea of those algorithms is to separate data by a line of hyperplane.

Random Forest

The "Random Forest" (Forest of Decision Trees) algorithm is a classification algorithm that help to improve single decision tree's performance by reducing its variance. I it combines many decision trees in a bagging-type approach.

• *n_estimators* (Number of trees in the forest) = 10,000.

Deep Learning: Neural Networks

A neural network is made up of elementary components called neurons. Modeling a biological neuron, the basic component of the network is a cell with n inputs E_1, E_2 ... E_n, and one or more outputs. Each entry E_i has a weight W_i. The neuron combines the n inputs as a linear function. The $sigma W_i * E_i$, then apply a transfer function f to the result to get the output. The function f is generally, a threshold function, which completely changes the output, if a small modification is applied to the inputs. In this work the different parameters of our neural network are as below:

- Inputs: 19 neurons (19 characteristics): age, body temperature, gender, dry cough, weakness, sour throat, breathing problem, drowsiness, travel history to infected countries, pain in chest, diabetes, lung disease, heart disease, stroke or reduced immunity, symptoms progressed, high blood pressure, kidney disease, loss of sense of smell, change in appetite.
- Output: 3 neurons (3 outputs) Corona result: 0 represent low risk, 1 represent medium risk, 2 is high risk.
- Layers: The network consists of 2 hidden layers containing 32 neurons and two dropout layers.
- The Hyper-parameters: Epochs = 120, Batch size = 32.
- The cost function and the optimizer: Categorical cross entropy, Adam.
- Activation functions: RELu, Softmax.

Layer(type)	Output Shape	Param#
dense_185(Dense)	(None, 19)	380
dense_186(Dense)	(None, 32)	640
dropout_93(Dropout)	(None, 32)	0
dense_187(Dense)	(None, 32)	1056
dropout_94(Dropout)	(None, 32)	0
dense_188(Dense)	(None, 3)	99

Total params: 2,175 Trainable params: 2,175 Non-trainable params: 0

Figure 10. Summary of deep learning model

2.3. Evaluation

2.3.1. Machine Learning Algorithms

To evaluate our model, we proceed by comparing different model performances. Figure 11 summarize the results of machine learning algorithms. As we can deduce the best score is given by random forest algorithm. In our study and to be sure from the result we developed a decision algorithm, which combine the most important three algorithms with the highest results, which is based on vote principle.

2.3.2. Deep Learning Model

As we know the confusion matrix is the best tools a performance measurement for our model classification. In our work we use this matric to measure precision, recall, F1-score and accuracy (Figure 12).

	Accuracy	precision	recall	f1-score
Decision Tree	85%	86%	87%	86%
Naïve Bayes	80%	78%	78%	78%
KNN	85%	89%	80%	80%
Random forest	95%	96%	93%	94%
SVM	85%	87%	88%	85%

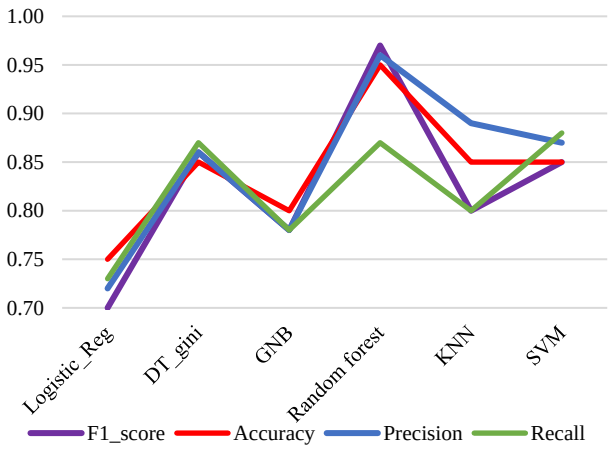

Figure 11. Comparison of model performance

Confusion Matrix: [8 0 0]
[1 4 0]
[0 0 7]

Accuracy: 95.0

Report	Precision	Recall	F1-score	Support
0	0.89	1.00	0.94	8
1	1.00	0.80	0.89	5
2	1.00	1.00	1.00	7
micro avg	0.95	0.95	0.95	20
macro avg	0.96	0.93	0.94	20
weighted avg	0.96	0.95	0.95	20

Figure 12. Classification report

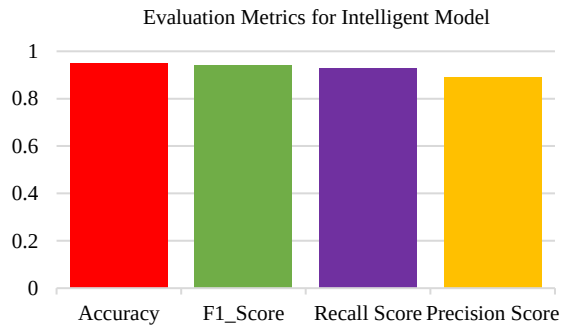

Figure 13. Evaluation metrics for the deep learning model

The accuracy given by our neural network is 95% with only 10% error. Those results are so encouraging, but to be sure that we didn't have an overfitting during our model training, we visualize the training loss vs validation loss (Figure 14). We can see in the graph that validation loss is slightly higher than training loss. We decide to stop the training at epoch 120, since afterwards no improvement is obtained and the two graphs have almost been set on the same values.

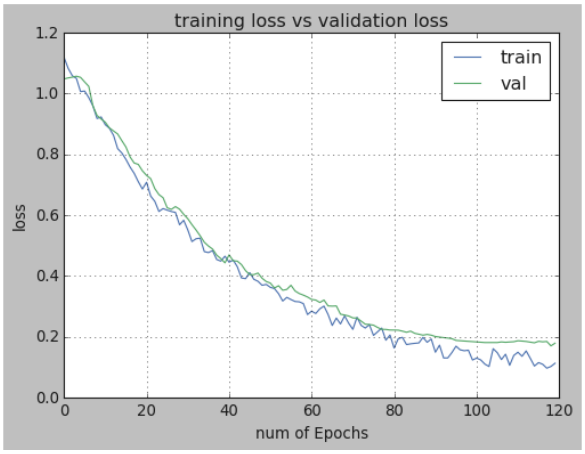

Figure 14. Training loss vs validation loss curve

These results allow us to deduce that the performance of our deep learning model exceeds those of machine learning algorithms. Besides, the training loss Vs validation loss curve shows that we don't have underfitting neither overfitting.

2.4. Deployment

Once we complete our analysis, we pass on to the last step of the CRISP-DM process ' Deployment '.

Due to the difficulty of integrating machine learning algorithms into the mobile application. We thought of converging towards deep learning, thanks to the Tensorflow Lite solution, which facilitates integration since it includes an execution environment on which we can run pre-existing models and a suite of tools that can be used to prepare our models for use on mobile and in-vehicle devices.

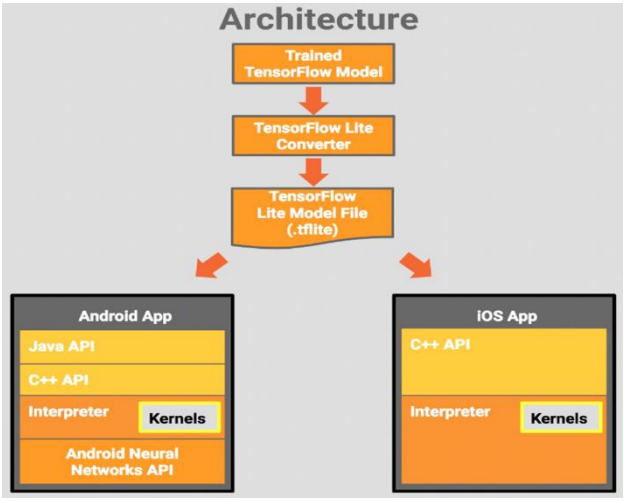

Figure 15. TensorFlow architecture

After integrating the IA model into our mobile application. Patient and doctors are enabling to pursue the COVID-19 status from their own devices. Figure 16 shows some of the mobile application interfaces.

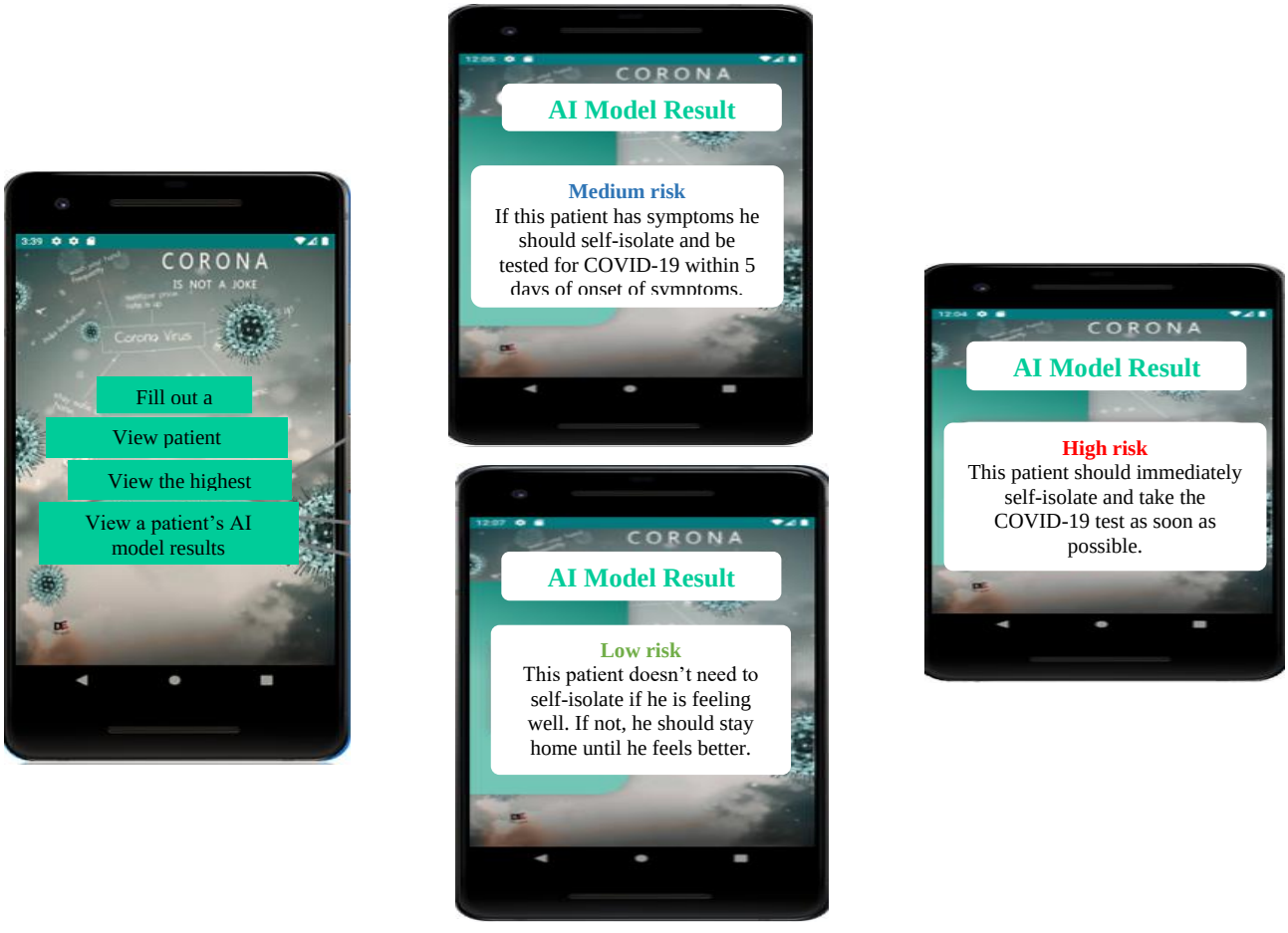

Figure 16. Deep learning mobile application interfaces

3. CONCLUSION

In this paper, we have proposed a mobile application with deep learning model for diagnosis and prognosis of the covid-19 infection. Based on the major symptoms such as body temperature, dry cough, breath problem, drowsiness, pain in chest, diabetes, lung disease, loss of sense of smell, high blood pressure- we made our prediction model, which can give as output the level of risk to be a coronavirus positive case. We used multiple machine learning algorithms (Bayesian network, KNN, SVM, Random forest) and we made a vote algorithm to choose the best predictive result.

We also used a deep learning neural network; it gives us very interesting results. We use this last model in our mobile application. In our next works, we will connect our application with sensors to get parameters automatically and we will develop the central system, which will give deep analysis and real time decisions.

Our global aim is to create a telehealth (or telemedicine) platform using artificial intelligence technologies to boost and encourage the clinical health care from a distance. The goal is for the AI platform to recommend treatment, sent recommendations, determine if the patient need to meet his doctor or no. We begin with cardiac patient and then we can generalize our models for other diseases and specialties.

REFERENCES

[1] Worldometers. Covid-19 Coronavirus Pandemic, https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?.

[2] Welcome Trust, "Sharing Research Data and Findings Relevant to the Novel Coronavirus (COVID-19) Outbreak", 2020, https://wellcome.

[3] L. Wynants, B. Van Calster, G.S. Collins, R.D. Riley, G. Heinze, E. Schuit, M. Smeden. "Prediction Models for Diagnosis and Prognosis of Covid-19", Systematic Review and Critical Appraisal. BMJ, Vol. 369, No. 369, p. 1328, 2020.

[4] J. Peto, A. Nisreen, M. Keith, N.A. Godfrey, R.A. Burgess, J.H. David, E. Riboli, P. Romer, et al., "Universal Weekly Testing as the UK COVID-19 Lockdown Exit Strategy", The Lancet 395, No. 10234, pp. 1420-1421, 2020.

[5] M.C. Patel, et al., "Asymptomatic SARS-CoV-2 Infection and COVID-19 Mortality during an Outbreak Investigation in a Skilled Nursing Facility", Clinical Infectious Diseases, Vol. 71, No. 11, pp. 2920-2926, 2020.

[6] G. Xu, Y. Yang, Y. Du, et al., "Clinical Pathway for Early Diagnosis of COVID-19", Updates from Experience to Evidence-Based Practice, Clinical Reviews in Allergy & Immunology, Vol. 59, No. 1, pp. 89-100, 2020.

[7] J.T. Guzik, S. A. Mohiddin, A. Dimarco, et al., "COVID-19 and the Cardiovascular System: Implications for Risk Assessment, Diagnosis, and Treatment Options", Cardiovascular Research, Vol. 116, No. 10, pp. 1666-1687, 2020.

[8] Y. Zheng, Y. Ma, J. Zhang, et al., "COVID-19 and the Cardiovascular System", Nature Reviews Cardiology, Vol. 17, No. 5, pp. 259-260, 2020.

[9] C. Song, J. Xu, J. He, et al., "COVID-19 Early Warning Score: A Multi-Parameter Screening Tool to Identify Highly Suspected Patients", MedRxiv, 2020.

[10] M. Mbida, A. Ezzati, "Artificial Intelligence Auscultation System for Physiological Diseases", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 49, Vol. 13, No. 4, pp. 97-103, December 2021.

[11] C. Feng, L. Wang, X. Chen, et al,. "A Novel Triage Tool of Artificial Intelligence-Assisted Diagnosis Aid System for Suspected COVID-19 Pneumonia in Fever Clinics". MedRxiv, p. 2020.03.19.20039099, 2021.

[12] A. Lopez Rincon, A. Tonda, D. Lucero Mendoza Maldonado, et al., "Specific Primer Design for Accurate Detection of SARS-CoV-2 Using Deep Learning - Preprint", Bull World Health Organ. E-pub., Vol. 27, pp. 1-16, 2020.

[13] A. Lopez Rincon, A. Tonda, L. Mendoza Maldonado, et al., "Classification and Specific Primer Design for Accurate Detection of SARS-CoV-2 Using Deep Learning", Scientific Reports, Vol. 11, No 1, pp. 1-11, 2021.

[14] X. Qi, Z. Jiang, Q. Yu, et al., "Machine Learning-Based CT Radiomics Model for Predicting Hospital Stay in Patients with Pneumonia Associated with SARS-CoV-2 Infection", A Multicenter Study, MedRxiv, Vol. 8, No. 1, pp. 859-859, 2020.

[15] V. Jain, A. Jain, V. Garg, et al., "Siamese Neural Networks for Pandemic Detection Using Chest Radiographs", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 51, Vol. 14, No. 2, pp. 104-110, June 2022.

[16] J. Gong, J.Ou, X. Qiu, et al., "A Tool to Early Predict Severe 2019-Novel Coronavirus Pneumonia (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong", China, JmedRxiv, 2020.

[17] F. Caramelo, N. Ferreira, B. Oliveiros, "Estimation of Risk Factors for COVID-19 Mortality-Preliminary Results", MedRxiv, 2020.

[18] X. Ma, et al., "Development and Validation of Prognosis Model of Mortality Risk in Patients with COVID-19", Epidemiology and Infection, Vol. 148, No. 168, pp. 1-7, 2020.

[19] J. Lu, et al., "ACP Risk Grade: A Simple Mortality Index for Patients with Confirmed or Suspected Severe Acute Respiratory Syndrome Coronavirus 2 Disease (COVID-19) during the Early Stage of Outbreak in Wuhan, China", MedRxiv, 2020.

[20] L. Yan, H. Zhang, Y. Xiao, et al., "Prediction of Criticality in Patients with Severe Covid-19 Infection Using Three Clinical Features: A Machine Learning-Based Prognostic Model with Clinical Data in Wuhan", MedRxiv, Vol. 27, 2020.

[21] L. Karthik, G. Kumar, T. Keswani, et al., "Protease Inhibitors from Marine Actinobacteria as a Potential Source for Antimalarial Compound", PloS one, .Vol. 9, No 3, p. e90972, 2014.

[22] Y. Shi, X. Yu, H. Zhao, et al., "Host Susceptibility to Severe COVID-19 and Establishment of a Host Risk Score / Findings of 487 Cases Outside Wuhan", Critical Care, Vol. 24, No 1, pp. 1-4, 2020.

[23] "Cross-Industry Standard Process for Data Mining" https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining.
[24] A.H. Odeh, M.A. Odeh, "Increasing the Efficiency of Online Healthcare Services Software and Mobile Applications Using Artificial Intelligence Technology", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 44, Vol. 12, No. 2, pp. 16-22, September 2020.
[25] https://es.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining.
[26] https://www.kaggle.com

BIOGRAPHIES

Khaoula Slime was born in Meknes, Morocco in 1991. She received the D.Eng. degree in Computer Engineering from Faculty of Science and Technology SETTAT, Hassan I University, Morocco in 2015. Currently, she is an employee at Ibn Rochd University Hospital Center, Casablanca, Morocco. In addition, she is preparing a Ph.D. in big data and machine learning in University of Hassan II, Casablanca, Morocco. Her research interests are in artificial intelligence in health sector and how we can increase healthcare services using new technologies.

Abderahim Maizate was born in Casablanca, Morocco in 1979. He received his Engineering Diploma in Computer Science from Hassania School of Public Works, Morocco in 2004 and DESA degree from ENSIAS, Morocco in 2007. Then, he received his Ph.D. degree from Chouaib Doukkali University, El Jadida, Morocco. He is currently, a Professor Researcher at Hassan II University, Casablanca, Morocco. His research interests are wireless communication, mobile communication, wireless sensor networks, quality of service (QoS) guarantees and big-data.

Oumaima Bouaissi was born in Casablanca, Morocco in 1997. She received her master degree in data science and big data engineering from Faculty of Science Ben M'sik, Hassan II University, Csablanca, Morocco. Currently, she is preparing a Ph.D. diploma at National School of Computer Science and Systems Analysis, Mohammed V University, Morocco. Her research interests are in artificial intelligence in health sector.

Adil Hilmani was born in Morocco. He received his diploma in network and telecommunication engineering from Engineering School, University of Seville, Spain. He received the Ph.D. degree in computer engineering and computer science from Hassan II University, Casablanca, Morocco in 2021. Currently, he is a lecturer at Office of Professional Training and Job Promotion (OFPPT), Kenitra, Morocco. He is a temporary lecturer at IT Department, Ibno Tofail University, Kenitra, Morocco. His research interests include mobile networks and computing, wireless sensor networks and embedded systems software for IoT.

Charif Mahmoudi was born in Morocco. He received the M.Sc. and Ph.D. degrees in computer engineering from University of Paris-EST, France in 2009 and 2014, respectively. After his PostDoc at National Institute of Standards and Technology, he joined Siemens Corporate Technology as an IoT Software Architect. He participated as consultant software architect to several successful telecommunication projects within France Telecom and Bouygues Telecom and contributed to several research projects with deployments in several countries. His areas of research are on distributed systems, cloud computing, mobile computing and Internet of things.

Najat Mouine was born in Morocco. She received the M.D., Ph.D., FESC Cardiologist and is specialist in Cardiac Imaging and cardiac rehabilitation Young National Ambassador of EAPC Member of SFC, BSC and SMC. Currently, she is Assistant Professor in cardiology, specialist in cardiac imaging and cardiac rehabilitation, head of Clinical Cardiology Department, HMIMV, Rabat, Morocco.