Towards Automatic Diagnosis of the COVID-19 Based on Machine Learning

El Arbi Abdellaoui Alaoui\textsuperscript{1,2(✉)}, Stephane Cedric Koumetio Tekouabou\textsuperscript{3,5}, Ismail Ougamane\textsuperscript{1}, and Imane Chabbar\textsuperscript{4}

\textsuperscript{1} EIGSI-Casablanca, 282 Route of the Oasis, Casablanca, Morocco
\texttt{abdellaoui.e@gmail.com, ismail.ougamane@gmail.com}
\textsuperscript{2} Department Computer Science, Faculty of Sciences and Technologies, My Ismail University, Errachidia, Morocco
\textsuperscript{3} Laboratory LAROSERI, Department of Computer Science, Faculty of Sciences, Chouaib Doukkali University, B.P. 20, 24000 El Jadida, Morocco
\texttt{ctekouaboukoumetio@gmail.com}
\textsuperscript{4} Specialty hospital, CHU IBN SINA, Mohammed V University, Quartier-Souissi, B.P. 6220, Rabat, Morocco
\texttt{imana1chab@gmail.com}
\textsuperscript{5} Center of Urban Systems (CUS), Mohamed VI Politechnic University (UM6P), Hay Moulay Rachid, 43150 Ben Guerir, Morocco

\textbf{Abstract.} Since the beginning of the global health crisis attributed to the new coronavirus (COVID-19), several announcements of new diagnostic tests have been made. It has suddenly become complicated to provide a comprehensive overview detailing the specificity of each of them. In the absence of therapeutic drugs or vaccines specific for COVID-19, these tests are essential to detect patients at an early stage and to immediately isolate infected patients from the healthy population. Among the most commonly used tests are chest CT and laboratory tests (such as PCR) in the diagnosis of coronavirus 2019 (COVID-19). Analysis of imaging and laboratory test data from more than 1,000 patients shows that chest CT surpasses biological tests in the diagnosis of the epidemic associated with the new coronavirus, COVID-19. Thoracic computed tomography (CT) scans provide the best diagnosis for COVID-19 pneumonia, conclude these researchers from Huazhong University of Science (Wuhan, China) and Leiden University Medical Center (Netherlands). The researchers concluded that CT scans should be used as the primary screening tool for COVID-19. However, these techniques have the disadvantage of being slow and expensive, causing many patients to avoid screening if it is not free. In this paper, we propose an automatic decision model based on artificial intelligence to assist the physician during the screening for this pandemic in order to lighten the workload in hospitals. For this purpose, the main objective of our study is to build and test predictive diagnostic models based on machine learning that can analyze patient data in depth in order to separate coronavirus types. Thus, we can save time during the diagnostic process of this Covid-19 pandemic. The results obtained are very promising.
Keywords: Coronavirus · COVID-19 · Machine learning · CT chest sensors

1 Introduction

Corona-viruses are a large family of viruses that can cause a variety of illnesses in humans, ranging from the common cold to Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). A new corona-virus (COVID-19) that had not previously been identified in humans was identified in 2019 in Wuhan, China [12,17].

Corona-virus 2019 (also known as Covid-19 or SARS-CoV-2) is an infectious disease caused by the corona-virus SARS-CoV-2 belonging to the very large family of viruses [19]. These viruses are in constant mutation and evolution. It is during one of these mutations that it became capable of infecting humans. Unlike its predecessors, this virus appears to be particularly contagious. In fact, it has been found in many biological fluids and excretions (secretions from the mouth and nose, blood, stools, urine), which suggests that there is a risk of multiple transmission, especially since not all infected patients, especially the youngest, necessarily show symptoms. In 80% of cases, Covid-19 causes few problems and the patient recovers quickly, without the need for hospitalization. But in people who are already weakened - by chronic illness, immunosuppression, old age, etc., Covid-19 can become complicated and require hospitalization or even resuscitation. It is especially for them that everyone must take responsibility and respect the instructions given in case of suspicion of infection or containment measures if the infection is proven. Meanwhile, researchers around the world are working to find an effective treatment for the most fragile and, in the longer term, to find a vaccine. Indeed, it is because the Covid-19 is new and therefore we are not immune to it, that it can spread so rapidly around the world!

In the absence of curative treatment or specific therapeutic vaccines for COVID-19, it is essential to detect the disease at an early stage and to be able to immediately isolate an infected patient. The latest guidelines recommend confirming the diagnosis of COVID-19 by RT-PCR (polymerase chain reaction from an RNA sample or gene sequencing) biological analysis of respiratory or blood samples prior to hospitalization. However, taking into account the possible hazards during sample collection and transport, as well as the performance of the kits, the total sensitivity of RT-PCR analysis for throat swab samples is estimated to be between 30% and 60%.

In a study of more than 1000 patients, published in the journal Radiology, chest CT surpassed laboratory tests in the diagnosis of coronavirus 2019 (COVID-19). The researchers concluded that CT should be used as the primary screening tool for COVID-19 [1]. Indeed, from January 6 to February 6, 2020, 1014 patients from Wuhan, China, who underwent both chest CT and RT-PCR testing were included. With RT-PCR as the reference standard, the performance of chest radiography in the diagnosis of COVID-19 was evaluated. In addition, for patients with multiple RT-PCR tests, the dynamic conversion of RT-PCR
results (negative to positive, positive to negative, respectively) was analyzed compared to serial chest scans for those with a time interval of 4 days or more [1].

In the face of the rapid spread of the coronavirus all means are good, including artificial intelligence (AI), which had already predicted the spread of the coronavirus at the end of December. Artificial intelligence (AI) can help us solve the urgent problems raised by the COVID-19 pandemic. It is not the technology itself that will make the difference, but rather the knowledge and creativity of the humans who use it.

Indeed, the COVID-19 crisis is likely to highlight some of the main shortcomings of AI. Machine learning, the current form of AI, works by identifying patterns in historical learning data. When used properly, AI has the potential to outpace humans not only in its speed, but also in detecting patterns in learning data neglected by humans [8,9,14,16].

Thus, the automation of diagnosis using new technologies and machine learning will make the diagnosis process faster and more reliable to prevent the loss of life of people infected by this virus. Indeed, artificial intelligence is one of the major topics of upheaval affecting our time. Rarely has a technological evolution generated so many opportunities for problem solving, so many changes in uses, so many fears. However, this is by no means a technological breakthrough. Artificial intelligence is part of the continuity of computer science, whose computing power is constantly growing, increased by the availability of large masses of data that the Internet world knows how to aggregate [2,13].

The objective of this paper is to design a simplified computer decision model to assist the physician during the pandemic screening process in order to ease the workload in hospitals. Indeed, the main objective of our study is to build and test models based on machine learning that are capable of in-depth analysis of patient data in order to predictively separate coronavirus types. Thus, we can save time during the diagnostic process of this Covid-19 pandemic and other types of this virus.

The rest of this paper is organized as following: We describe the proposed method to predict the type of coronavirus in Sect. 2 before analyzing the results obtained in the Sect. 3. Finally, we conclude our study in Sect. 4.

## 2 Methodology

In general, human learning is an adaptive process whereby the individual provides adequate responses to certain situations. In Psychology or Cognitive Sciences, the term “learning” refers to the process of increasing the efficiency of mental or behavioural activity as a result of experience.

Machine learning [11] is a sub-field of Artificial Intelligence (AI) whose objective is to study the means by which a machine can learn. Learning, in this context, means being able to adapt its behaviour in the presence of unknown situations (not foreseen by the designers of the machine) and being able to extract laws from databases of examples. Learning is therefore done through tools that
allow to acquire, extend and improve the knowledge available to the system. It consists of using computers to optimize an information processing model, according to certain performance criteria based on observations. Machine learning techniques are thus used for example for pattern recognition (writing, speech, vision), data mining (knowledge extraction), implementation of decision support tools, etc. The different methods of machine learning are classified into three groups: supervised learning [20], unsupervised learning and Semi-supervised learning (see Fig. 1): [18] (Fig. 2).

**Fig. 1.** Machine learning groups

**Fig. 2.** Automatic system for diagnosis of the COVID-19
2.1 Random Forest Classifier

The random forest algorithm [3] is a variant of bagging where a set of random trees close to the CART method is aggregated [4]. Usable in both regression and classification, this algorithm has shown very good performances in practice, especially for complex problems (nonlinear relations, interactions, high dimension, etc.).

**Algorithm 1. Random Forest (RF)**

**Input:**
- \( y \) the observation to predict;
- \( d_n \) the observation;
- \( M \) the number of Trees;
- \( b \in \mathbb{N} \) the number of candidate variables to cut a node.

**Output:** \( \hat{f}_{RF}(x, \mathcal{L}) = \arg\min_k (\#\{m : \hat{g}(x, \mathcal{L}_m^*) = k\}) \)

1: for \( k = 1 \) to \( B \) do
2: Draw a bootstrap \( \mathcal{L}_m^* \) sample in \( d_n \)
3: Construct a CART tree on this bootstrap sample \( \mathcal{L} \), each cutoff is selected by minimizing the cost function of CART over a set of \( b \) randomly selected variables among the \( p \). We note \( \hat{g}(., k) \) the built tree.
4: end for

2.2 ExtraTrees Classifier

ExtraTrees [7] is another comprehensive method specifically designed to use decision trees. Randomness is related to the way splits are calculated. In contrast to random forests, instead of looking for the most discriminating thresholds, thresholds are randomly selected for each attribute and the best of these randomly generated thresholds is chosen as the splitting rule.

**Bagging Classifier.** The bagging algorithm improves prediction efficiency if a disturbance in the learning base leads to significant changes in the constructed model. The general principle of this algorithm is to aggregate a collection of weak classifiers to obtain a better classifier. In general for classification, aggregation by majority vote:

\[
\hat{f}_{Bagg}(x, \mathcal{L}) = \arg \max_{y \in \mathcal{Y}} \sum_{t=1}^{T} I(\hat{g}_t(x) = y)
\]

2.3 AdaBoost Classifier

AdaBoost is a classification algorithm that aims to use basic classifiers to build a powerful set. In the end, the model generates an aggregated classification.
**Algorithm 2. Bagging Classifier (BC)**

**Input:**
- Data set $\mathcal{L} = \{(X_1, Y_1), ..., (X_n, Y_n)\}$;
- Base learning algorithm $\mathcal{Z}$;
- Number of base learners $t = 1, ..., T$;

**Output:** $\hat{f}_{Bagg}(x, \mathcal{L}) = \arg\max_{y \in \mathcal{Y}} \sum_{t=1}^{T} I(\hat{g}_t(x) = y)$

1: **for** $t = 1$ to $T$ **do**
2: $\hat{g}_t = (\mathcal{L}, \mathcal{L}^*)$
3: **end for**

For each basic model, the algorithm assigns a weight based on the individual performance of the model. The idea is that the classifiers will have to focus on observations that are difficult to classify correctly. Any machine learning algorithm can be used as a basic classifier if it accepts weights on the training set [10,15].

Initially all learning examples have the same weights $L_1(x) = \frac{1}{m}$. At each iteration (let us suppose that $m$ designates the number of the current iteration) we choose in $\hat{g}$ the classifier $\hat{g}_m$ which minimizes the classification error on the training data weighted by the $L_t$. Then we calculate $\alpha_t$, the weight of $\hat{g}_m$ in the final mix, update the weight $L_t$ to boost the elements that were misclassified and move on to the next iteration. The detailed algorithm is given below.

**Algorithm 3. AdaBoost Classifier (BC)**

**Input:**
- Data set $\mathcal{L} = \{(X_1, Y_1), ..., (X_n, Y_n)\}$;
- Base learning algorithm $\mathcal{Z}$;
- Number of base learners $t = 1, ..., T$;
- Initialize the weight distribution $L_1(x) = \frac{1}{m}$;

**Output:** $\hat{f}_{AdaB}(x, \mathcal{L}) = \text{sign}(\sum_{t=1}^{T} \alpha_t \hat{g}_t(x))$

1: **for** $t = 1$ to $T$ **do**
2: $\hat{g}_t = (\mathcal{L}, \mathcal{L}^*)$
3: $\epsilon_t = P_{x \sim L} (\hat{g}_t(x)) \neq f(x)$
4: **if** $\epsilon_t > 0.5$ **then**
5: $\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \epsilon_t}{\epsilon_t} \right)$
6: $L_{t+1}(x) = \frac{L_t(x)}{Z_t} \times \begin{cases} \exp(-\alpha_t) & \text{if } \hat{g}_t(x) = f(x) \\ \exp(\alpha_t) & \text{if } \hat{g}_t(x) \neq f(x) \end{cases} = \frac{L_t(x) \exp(\alpha_t \hat{g}_t(x) f(x))}{Z_t}$
7: **end if**
8: **end for**

2.4 XGBoost Classifier

XGBoost [5] is an optimized way to realize the deTree Gradient Boosting algorithm. XGBoost has performed remarkably well in machine learning competitions because it effectively handles a wide variety of data types, relationships,
and distributions, as well as the large number of hyper-parameters that can be modified and set for improvement. This flexibility makes XGBoost a solid choice for regression, classification (multiclass and binary) and ranking problems. For the model evaluation the algorithm can independently determine the types of the loss functions. An additional adjustment term is added to the model, to reduce the risk of overloading. As a predictive value for regression of each tree, the algorithm use the mean score. For the \( m^{th} \) decision tree, its calculation formula can be expressed as:

\[
\hat{y} = \sum_{i=1}^{m} f_m(x_i), \quad f_m \in W, \tag{1}
\]

where \( f_m \) is a function in the functional space \( W \), \( m \) is the number of trees, and \( W \) is the space of all decision trees.

The objective function at the \( t \)-th iteration can be presented as:

\[
\Theta(t) = \Phi(t) + \Omega(t) = \sum_{i=1}^{n} \Phi(y_i, \hat{y}_i) + \sum_{k=1}^{t} \Omega(f_k), \tag{2}
\]

where \( n \) is the \( n^{th} \) prediction \( \hat{y}^{(t)}_i \) can be written as

\[
\hat{y}^{(t)}_i = \sum_{i=1}^{m} f_m(x_i) = \hat{y}^{(t-1)}_i + f_t(x_i) \tag{3}
\]

The regularization term \( \Omega(f_k) \) for a decision tree is defined by Chen and Guestrin [5] as follows:

\[
\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{i=1}^{m} \omega_j^2 \tag{4}
\]

where \( \lambda \) is a parameter to scale the penalty, \( \gamma \) is the complexity of each leaf, \( T \) is the number of leaves in a tree, and \( \omega \) is the vector of scores on the leaves. Then, the first-order along and the second-order Taylor expansions are taken to the loss function in XGBoost.

3 Experiments and Results

3.1 Data Set

The covid-chestxray-dataset data frames describe the health status of individual patients of novel coronavirus (COVID-19). The covid-chestxray-dataset data frame does not contain information from Clinical Staff, but it does contain actual ages and other informations of patients. The principal source for data about patients is the covid-chestxray-dataset in github [6] (Fig. 3).
Towards Automatic Diagnosis of the COVID-19 Based on Machine Learning

![Operating principle of boosting algorithms](image)

**Fig. 3.** Operating principle of boosting algorithms

| N  | Attribute                          | Description                                                                 | Type          |
|----|------------------------------------|-----------------------------------------------------------------------------|---------------|
| 1  | Patientid                          | ID of the patient                                                           | Numerical     |
| 2  | Offset                             | Number of days since the start of symptoms or hospitalization               | Numerical     |
|    |                                    | If a report indicates “after a few days”, then 5 days is assumed           |               |
| 3  | Sex                                | Man or woman                                                                | Categorical   |
| 4  | Age                                | Age of the patient in years                                                 | Numerical     |
| 5  | Finding                            | Type of pneumonia                                                           | Categorical   |
| 6  | Survival                           | If the patient is survived or not                                           | Categorical   |
| 7  | Intubated                          | If the patient was intubated or not                                         | Categorical   |
| 8  | intubation_present                 | If the patient is intubated in present or not                              | Categorical   |
| 9  | went_icu                           | If the patient went to intensive care unit in present or not                | Categorical   |
| 10 | needed_supplemental_O2             | If the patient needed supplemental oxygen or not                           | Categorical   |
| 11 | Extubated                          | If the patient was extubated or not                                         | Categorical   |
| 12 | Temperature                        | Body temperature of the patient                                             | Numerical     |
| 13 | pO2_saturation                     | Partial pressure of oxygen of the patient                                  | Numerical     |
| 14 | leukocyte_count                    | The percentage of leukocyte in blood                                        | Numerical     |
| 15 | lymphocyte_count                   | The percentage of lymphocyte in blood                                       | Numerical     |
| 16 | neutrophil_count                   | The percentage of neutrophil in blood                                       | Numerical     |
| 17 | View                               | Posteroanterior (PA), Anteroposterior(AP), AP Supine (APS), or Lateral (L) | Categorical   |
|    |                                    | For X-rays; Axial or Coronal for CTscans                                   |               |
| 18 | Modality                           | CT, X-ray, or something else                                                | Categorical   |
| 19 | Date                               | Date on which the image was acquired                                       | Date          |
| 20 | Location                           | The hospital where the patient is hospitalized                             | Categorical   |
| 21 | Folder                             | The folder of the placement of X-Ray of the patient                         | Categorical   |
| 22 | Filename                           | Filename of X-Ray of the patient                                            | Categorical   |
| 23 | Doi                                | Digital Object Identifier of data set                                      | Categorical   |
| 24 | Url                                | URI of origin data set                                                     | Categorical   |
| 25 | License                            | The type of licence of the origin data set                                  | Categorical   |
| 26 | clinical_notes                     | Clinical notes about the patient                                           | Text          |
| 27 | other_notes                        | Other clinical notes about the patient                                     | Text          |

**Table 1.** Description
3.2 Data Processing

In the original data set, there are 21 columns before the processing, but after the cleaning of the data set (drop the columns that doesn’t contribute to the target variable ‘finding’ Table 1) the remain columns are five: offset, sex, age, modality and finding, and here the distribution of each column (see Fig. 4), we can conclude that random variable ‘age’ follows a normal distribution, the majority of patient are male.

![Distribution of variables](image)

To summarize the data processing the “finding” is the target column/variable, the columns like ‘survival’, ‘temperature’ doesn’t contribute to the target variable ‘finding’. So we can remove it from the data, for columns ‘Age’, ‘sex’ and ‘offset’ has less number of missing value, we have to impute them using different techniques.

3.3 Experimental Protocol

We define the workflow of the project that includes all steps required to build the machine learning project (see Fig. 5), we can divide the workflow of our project into four stages:

- Gathering data and Data pre-processing
- Researching the model that will be best for the type of data
- Training, cross-validate and testing the model
- Evaluation
All the experimentations were done on python using models from the Scikit-learn library which are very efficient for predictive classification problems. The computer used is a HP computer with 16 GB of RAM and a dual NVDIA graphics card. In order to choose an optimal approach, we compared the performance of different models using three main measures: (TP: True Positives, TN: True Negatives, FN: False Negatives, FP: False positives):

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{6}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{7}
\]

### 3.4 Results Analysis

To show the reliability of our model, we compare the different algorithms, the validation performances and the test performances. We calculated the Accuracy, Precision and Recall of each algorithm on these two scores. The values of the performance metrics are grouped below in the Table 2. The performance metrics associated with each metric are illustrated graphically.

From Table 2, we can see that for this metric, the bagging methods BC and ET give the best scores for both validation (0.9722, and 0.9709 respectively) and testing (0.9682 for both BC and ET). They are followed by the individual algorithms DT then KNC give respectively 0.9708 then 0.9396 for the validation and 0.9682 then 0.9318 for the test. Finally, the algorithms based on boosting, in particular XGBR, which gives 0.9620 and 0.9682 for validation and 0.9682 for the test. Whose values are summarized in the Table 2, we can see that overall...
Table 2. Performance Results based in accuracy, precision and recall

| Algorithm                 | Accuracy Validation score | Precision Validation score | Accuracy Test score | Precision Test score | Recall Validation score | Recall Test score |
|---------------------------|---------------------------|---------------------------|---------------------|----------------------|-------------------------|------------------|
| Random Forest (RF)        | 0.9680                    | 0.8501                    | 0.8492              | 0.7968               |                         |                  |
| Bagging Classifier (BC)   | 0.9722                    | 0.8535                    | 0.8979              | 0.7968               |                         |                  |
| Extra Trees (ET)          | 0.9709                    | 0.8493                    | 0.8979              | 0.7968               |                         |                  |
| XGBoost (XGB)             | 0.9620                    | 0.8376                    | 0.8979              | 0.7968               |                         |                  |
| Decision Tree (DT)        | 0.9708                    | 0.8483                    | 0.8979              | 0.7968               |                         |                  |
| K-Neighbors Classifier (KNC) | 0.9396                  | 0.7791                    | 0.7020              | 0.7430               | 0.7215                  |                  |

The bagging methods always give the best scores and therefore the lowest errors both in validation and in test. The values obtained for BC, ET and RF are respectively 0.8535, 0.8493 and 0.8501 in validation and 0.8979 for each in test for precision.

For the recall, the values obtained for BC, ET and RF are respectively 0.8482, 0.8463 and 0.8492 in validation and 0.7968 for each in test. These better performances are followed by those of the boosting methods, in particular XGB with respectively 0.8336 in validation and 0.7968 in test. Finally, the individual algorithms (DT and KNC) which give the lowest score corresponding to the highest errors, for the precision and recall metrics, the values obtained for these two metrics are respectively 0.8483, 0.7791, 0.8183, 0.7430 in validation then 0.8979, 0.7020, 0.7968, 0.7215 in test.

4 Conclusion

The current Covid-19 pandemic requires a rapid and effective diagnostic strategy for patient management. Recall that the probability of being detected positive is related to viral load and depends on the duration of symptoms and the severity of the disease. Thoracic CT scans are quick and relatively easy to perform. Recent research has revealed that the sensitivity of Chest CT for COVID-19 infection was 98% compared to the sensitivity to RT-PCR of 71%. The researchers concluded that Chest CT should be used as the primary screening tool for COVID-19. Artificial intelligence can help fight the coronavirus, if applied creatively.

Indeed, the main objective of our study is to build and test models based on machine learning that are able to analyze patient data in order to predictively separate coronavirus types. Thus, we can save time during the diagnostic process of this Covid-19 pandemic and other types of this virus. The results of this paper showed that our learning machine based model can accurately detect COVID-19 and other classes of coronaviruses with an accuracy of 97%.
References

1. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., Xia, L.: Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 296(2), 200642 (2020)
2. Anastassopoulou, C., Russo, L., Tsakris, A., Siettos, C.: Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS One 15(3), e0230405 (2020)
3. Breiman, L.: Random forests. Mach. Learn. 45(1), 5–32 (2001)
4. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. CRC Press, Boca Raton (1984)
5. Chen, T., Guestrin, C.: XGboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
6. Cohen, J.P., Morrison, P., Dao, L.: COVID-19 image data collection. arXiv preprint arXiv:2003.11597 (2020)
7. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Mach. Learn. 63(1), 3–42 (2006)
8. Karada˘g, K., Tenekeci, M.E., Ta¸saltın, R., Bilgili, A.: Detection of pepper fusarium disease using machine learning algorithms based on spectral reflectance. Sustain. Comput.: Inf. Syst. 28, 100299 (2019)
9. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., et al.: Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology 296(2), 200905 (2020)
10. Li, X., Wang, L., Sung, E.: Adaboost with SVM-based component classifiers. Eng. Appl. Artif. Intell. 21(5), 785–795 (2008)
11. Mitchell, T.M.: Does machine learning really work? AI Mag. 18(3), 11–11 (1997)
12. Muhammad, L., Islam, M.M., Sharif, U.S., Ayon, S.I.: Predictive data mining models for novel coronavirus (COVID-19) infected patients recovery. SN Comput. Sci. 1(4), 206 (2020)
13. Narin, A., Kaya, C., Pamuk, Z.: Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849 (2020)
14. Nguyen, T.T.: Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research directions (2020)
15. Schapire, R.E., Freund, Y., Bartlett, P., Lee, W.S., et al.: Boosting the margin: a new explanation for the effectiveness of voting methods. Ann. Stat. 26(5), 1651–1686 (1998)
16. Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., Shen, D.: Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. arXiv preprint arXiv:2004.02731 (2020)
17. Shinde, G.R., Kalamkar, A.B., Mahalle, P.N., Dey, N., Chaki, J., Hassamien, A.E.: Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN Comput. Sci. 1(4), 1–15 (2020)
18. Springenberg, J.T.: Unsupervised and semi-supervised learning with categorical generative adversarial networks. arXiv preprint arXiv:1511.06390 (2015)
19. Yang, Y., Peng, F., Wang, R., Guan, K., Jiang, T., Xu, G., Sun, J., Chang, C.: The deadly coronaviruses: the 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. J. Autoimmun. 109, 102434 (2020)
20. Zhou, Z.H.: A brief introduction to weakly supervised learning. Natl. Sci. Rev. 5(1), 44–53 (2018)