

ORIGINAL RESEARCH



Machine learning techniques as an efficient alternative diagnostic tool for COVID-19 cases

Nicolás Bustos¹, Manuel Tello¹, Guillermo Droppelmann^{2,3}, Nicolás García^{2,4}, Felipe Feijoo¹, Víctor Leiva^{1,*}

¹School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, 2362807 Valparaíso, Chile

²Academic Unit, Clínica MEDS, 7691236 Santiago, Chile

³Health Sciences Ph.D. Program, Universidad Católica de Murcia, 30107 Murcia, Spain

⁴Departments of Radiology, Clínica MEDS, 7691236 Santiago, Chile

***Correspondence**

victor.leiva@pucv.cl;

victorleivasanchez@gmail.com

(Víctor Leiva)

Abstract

Background: The SARS-CoV-2 virus has demonstrated the weakness of many health systems worldwide, creating a saturation and lack of access to treatments. A bottleneck to fight this pandemic relates to the lack of diagnostic infrastructure for early detection of positive cases, particularly in rural and impoverished areas of developing countries. In this context, less costly and fast machine learning (ML) diagnosis-based systems are helpful. However, most of the research has focused on deep-learning techniques for diagnosis, which are computationally and technologically expensive. ML models have been mainly used as a benchmark and are not entirely explored in the existing literature on the topic of this paper.

Objective: To analyze the capabilities of ML techniques (compared to deep learning) to diagnose COVID-19 cases based on X-ray images, assessing the performance of these techniques and using their predictive power for such a diagnosis.

Methods: A factorial experiment was designed to establish this power with X-ray chest images of healthy, pneumonia, and COVID-19 infected patients. This design considers data-balancing methods, feature extraction approaches, different algorithms, and hyper-parameter optimization. The ML techniques were evaluated based on classification metrics, including accuracy, the area under the receiver operating characteristic curve (AUROC), F1-score, sensitivity, and specificity.

Results: The design of experiment provided the mean and its confidence intervals for the predictive capability of different ML techniques, which reached AUROC values as high as 90% with suitable sensitivity and specificity. Among the learning algorithms, support vector machines and random forest performed best. The down-sampling method for unbalanced data improved the predictive power significantly for the images used in this study.

Conclusions: Our investigation demonstrated that ML techniques are able to identify COVID-19 infected patients. The results provided suitable values of sensitivity and specificity, minimizing the false-positive or false-negative rates. The models were trained with significantly low computational resources, which helps to provide access and deployment in rural and impoverished areas.

Keywords

Artificial intelligence; Deep learning; PCR; ROC curve; R software; SARS-CoV-2; X-rays

1. Introduction

The Coronavirus (COVID-19) disease is caused by a novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and first began in the Wuhan City, Hubei Province, China, early in December 2019 [1, 2]. Although COVID-19 is a new virus, knowledge on this disease is daily generated. Several authors [1, 3] described SARS-CoV-2 as a type β -coronavirus cluster. COVID-19 has been defined as the third known zoonotic coronavirus, following the SARS and the Middle

East respiratory syndrome (MERS) [3–5]. The COVID-19 affects the human's lower respiratory tract, and it is often manifested as pneumonia [4]. Typical symptoms for COVID-19 positive patients include cough, fatigue, and fever, but other symptoms as diarrhea, digestive problems, dyspnoea, hemoptysis, headache, lymphopenia, and sand sputum production have been also identified [6–10]. Studies have shown that such symptoms likely appear after an incubation period of approximately 5.2 days [9, 11]. In contrast, the average time to death after symptoms appearance ranges from 6 to 41 days

with a mean of 14 days [9, 10].

The mentioned symptoms are common to other respiratory syndromes. Hence, detection and confirmation of a patient having a COVID-19 viral infection are performed using polymerase chain reaction (PCR) tests, which, to date, remains the standard gold examination [12]. However, there are several considerations and limitations to this type of diagnostic test. For instance, authors have reported that: (i) there are more than seven different SARS-CoV-2 nucleic acid PCR tests [13, 14]; (ii) when considering the viral load in samples, it has been shown that upper respiratory tract samples have their peak viral loads three days after symptom onset [15]; and (iii) as the SARS-CoV-2 epidemic evolves globally, we need better diagnostic tests that are rapid, reliable, validated and widely available. Other authors further elaborated on the accessibility to tests [16], including that the supply of the reagents does not satisfy the demand. Also, the results of PCR tests take one day or longer to be obtained after sampling (in contexts that are not overcrowded and saturated by infected people). Considering such limitations, some investigations have called for an initial screening with an auxiliary examination based on computed tomography (CT) or chest X-rays [12–14, 16–19]. Typical chest manifestations of COVID-19 diagnosed patients include consolidation, crazy paving pattern, ground-glass opacities, and reticular pattern, which are typical CT manifestations [19–21].

Artificial intelligence techniques have been proposed for screening, tracking, and predicting patient health outcomes and for developing drugs and vaccines through clinical trials [22, 23]. However, most of the research has been focused on the early detection and diagnosis of COVID-19 patients using deep learning (DL) techniques [18, 24–27].

Table 1 (Ref. [18, 24–26, 28–30]) presents a summary of some recent approaches used for COVID-19 diagnosis with medical images and techniques. Convolutional neural networks (CNN) have been used in recent literature on the topic. For instance, CNN models utilizing CT images of COVID-19, pneumonia, and healthy patients were considered in [26]. Similar work was conducted in [18, 24]. More recently, authors have derived hybrid techniques based on DL (CNN features) and support vector machines (SVM) algorithms, named convolutional support vector machines (CSVM) [28]. Also, COVID-19 patients using X-ray images were predicted employing CSVM [31]. These authors identified that CSVM with a polynomial kernel provided the best result.

Similarly, CSVM models with additional feature ranking techniques improved the classification performance for COVID-19 [25]. As a different approach, a high accuracy in predicting COVID19 could be reached by using transfer learning [27]. Combining architectures such as dense convolutional network (DenseNet), residual neural network (ResNet), and very deep CNN for large-scale image recognition (VGG), a high-performance model was proposed in [27], which reached an accuracy level of 97.1%. Deep features from 13 different CNN models with SVM as the classification technique were used in [29]. These features differ from previously mentioned works as they are implemented in two steps: first decomposing images using CNN and then classifying data with SVM. The hybrid

ResNet50 CNN and SVM model provided the best results based on accuracy, the area under the receiver operating characteristic curve (AUROC), F1-score, false-positive rate (FPR), false-negative rate (FNR), and sensitivity. It is also common to find other machine learning (ML) techniques without deep features [32]. These ML techniques are employed as a benchmark for more complex models, but their predictive power is not usually assessed due to their use as a benchmark and not as a classifier itself.

The studies mentioned above demonstrate that DL (and CNN) is (are) highly effective for discriminating COVID-19 images. Nevertheless, SVM has been used as an alternative to a DL-based classifier, as CNN needs large data sets for training, validating, and reducing the overfitting [29]. There are still other drawbacks to be addressed when CNN-based models are considered. Firstly, CNN models heavily rely on high-performance computing, where complex models can take several hours to days to be trained, developed, analyzed, and implemented in a clinical context [33]. This is a critical issue in rural and impoverished areas that lack access to a proper technological infrastructure [34]. Secondly, commonly used algorithms and network architectures are complex to design and explain, which contrasts with desired biomedical and health contexts [35]. These drawbacks take significant importance in the current context of a pandemic. Notably, in the present world situation, a rapid response is needed in communities where the health systems are saturated, and access to PCR tests (or other tests, such as rapid tests) is limited [14–16]. Furthermore, it is commonly found in the literature that impoverished, rural, and small communities lack infrastructure access [34]. Significant socioeconomic differences and disparities have been found in the context of the COVID-19 pandemic [36–39]. Therefore, in such a context, simple, fast, and cheap tools, such as those provided by ML techniques, can be of great value.

The objective of this paper is to explore the capabilities of ML techniques, in comparison to DL, as valuable tools to diagnose coronavirus positive cases based on X-ray chest images, providing an accessible, affordable, and fast solution to the initial screening of COVID-19 patients. Note that our investigation takes a different approach to what has been studied in the literature. We develop a full factorial experiment to train a series of ML-based techniques, such as artificial neural networks (ANN), random forest (RF), and SVM. The design of experiment considers different data-balancing methods, cross-validation approaches, kernel functions, and hyper-parameter optimizations. For each technique studied here, the statistical experiment allows us to obtain predictive power metrics such as accuracy, sensitivity, and specificity. This experiment is crucial for re-sampling and data-balancing methods to accurately account for data unbalancing associated with low COVID-19 positive rates of detection.

Methods to handle unbalanced data can be classified into internal or external. On an internal level, one may tune the existing design to address the class imbalances, being its advantage that this level is specific to an algorithm. On the external level, different types of data sampling methods are used. The external approach is straightforward and can be applied to any algorithm, but it incurs overfitting or loss of

TABLE 1. Performance of the best models presented in the literature review.

ML technique	Image type	COVID-19 cases	Other cases	Performance metric			
				Sensitivity	Specificity	Accuracy	Others
CNN [18]	X-ray	231	1493 (pneumonia) 1583 (healthy)	92.11%	96.06%	92.18%	F1-score = 92.07%
CNN [24]	CT	1296	1735 (CAP*) 1325 (non-pneumonia)	90%	96%	96%	N/A
CSVM [25]	CT	53	247 (healthy)	98.93%	97.60%	98.27%	F1-score = 98.28%
CSVM [28]	CT	1262	1262 (non-COVID-19)	96.09%	94.10%	94.03%	F1-score = 88.15% Kappa = 88.07% AUROC = 0.9866
SVM using shrunken features [30]	CT and X-ray	101	159 (non-COVID-19)	91.88%	98.54%	94.23%	F1-score = 93.99%
Deep features and SVM classifier [29]	X-ray	127	127 (pneumonia) 127 (healthy)	95.33%	FPR = 2.33%	95.33%	F1-score = 95.34%
CNN [26]	CT	752	797 (pneumonia) 679 (regular control patients)	94.93%	91.13%	92.49%	AUROC = 0.9797

*where CAP: community-acquired pneumonia.

information [32, 40].

Unlike the existing literature, we use a design of experiment to estimate the classification power of ML techniques, which are often not fully optimized and explored by previous research. We demonstrate that the ML techniques can provide stable predictions with low and suitable FPR and FNR. The ML techniques that utilize few resources can be used as COVID-19 diagnostic tools in contexts where access to health infrastructure is lacking, such as in impoverished and rural regions or communities.

The rest of the paper is organized as follows. Section 2 presents the methods used in this research as well as details of the data employed for training and testing the proposed ML models. Section 3 reports the results from the design of experiment, while Section 4 concludes and discusses the main findings of our research.

2. Methods

This section describes the process followed to create a usable data set based on the analysis of CT images. Then, we provide the details (training, testing, and tuning) of the ML techniques used.

2.1 Data set: X-ray image source and database creation

The data set used for training models was created from a series of steps, including data (image) pre-processing, object detection, and feature extraction [41, 42]. Fig. 1 depicts the overall

process followed in this study. The image pre-processing step considers cutting, resizing, and formatting X-ray images. Object detection is focused on making a demarcation of the areas in the image where the objects of interest for the analysis are located. The extraction of characteristics considers two main features. First, we obtain textured X-ray images, which quantify the distribution of the intensities of the pixels within each object based on their average and standard deviation (SD). Among the most advanced texture feature extraction techniques, we consider pixel-pixel co-occurrence patterns [43] and granularity measures at different scales [44, 45]. The second class of features describes the X-ray image contour based on segmentation masks: the roughness or circularity of the contour [46]. More specifically, we create features that quantify the shape of the object, its image moments (which can be computed with or without reference intensities), and features that quantify the texture of the pixels. Then, the final data set considers the response variable (COVID-19 positive, healthy, or other pneumonia patients), 95 other covariates from X-ray image characteristics created from the feature extraction process, and patient's data (age and sex). The feature extraction process was developed with the EBImage package of the R software [47, 48]. The X-ray images were obtained from public repositories of COVID-19 (based on assembling of medical images in existing publications) and from the public chest X-ray database [49, 50].

As a result of the analysis of images, and a merge of these data with patient's characteristics, the data set finally contains: (i) three types of views related to posterior-anterior (PA),

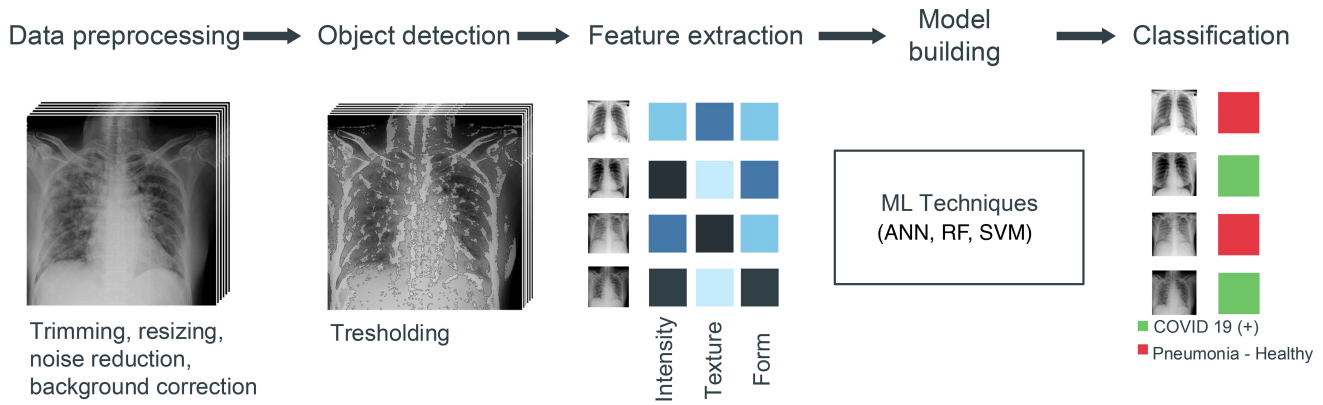


FIGURE 1. Overall data set and model building process.

anterior-posterior (AP), and left (L) parts; (ii) sex, age, and diagnosis (COVID-19, healthy, pneumonia, SARS, streptococcus, and others) of patients; and (iii) metrics of the images.

2.2 Machine learning techniques: training and testing

Different ML techniques were used for classification in this study, including ANN, RF, and SVM. A CNN was also used as a benchmark for comparison with ML techniques. All models were trained to predict patients with positive COVID-19 against healthy patients or with pneumonia. Performance evaluation metrics were used to assess the capability of ML techniques for classification. The mean and its confidence interval for the metrics were obtained using a k-fold-repeated-cross-validation (KFRCV) with five and ten as the number of folds, separately, and with ten repeats each. The data cleaning, training, and testing processes are depicted in Fig. 2. We use the R software and its libraries caret (for creating models), tidyverse (for data managing), e1071 (for calculating metrics), and pROC (for creating curves).

As shown in Fig. 2, we apply data-balancing after the data cleaning step (filtering and pre-processing). This balancing includes down-sampling, up-sampling, synthetic minority over-sampling technique (SMOTE) [51], and Rose. Note that, for the down-sampling method, we also use an external sampling approach to account for the reduced number of data points of the positive class (COVID-19). For each of the data-balancing methods, we perform a re-sampling algorithm based on KFRCV. Each model is trained with both five and ten folds. Additionally, the k-fold cross-validation approach is repeated ten times to estimate mean values and their confidence intervals for obtaining performance metrics accuracy, AUROC, sensitivity, and specificity. The training process is performed for different types of ANN, RF with varying numbers of trees, and SVM with different kernel functions. We perform a hyper-parameter optimization during re-sampling (tuning length = 15). We calculate the confusion matrix for each of the models to determine accuracy, AUROC, sensitivity, and specificity values. The details about each ML technique and its hyper-parameter optimization are discussed next.

ML hyper-parameters were tuned using grid and random search. RF models were created using different numbers of

trees (100, 300, 500, 1500, and 2000). For tuning parameters, the number of variables available for splitting varied randomly, at each tree node, ten times. For SVM, linear, polynomial, and radial kernel functions were compared. For SVM tuning parameters, different costs (for each kernel), sigma (for radial), degree, and scale (for polynomial) were implemented. These parameters change 15 times randomly. Table 2 presents the range and selection of the values of parameters in SVM and RF for better understanding. ANN models with both multi-layer and multi-decay search were developed. For tuning parameters, different numbers of layers, neurons, and decay were also considered. These numbers change throughout a grid search where, for multilayer, the number of layers (first layer) goes from 1 to 3, and the number of neurons from 1 to 10. For subsequent layers, we considered 0 to 10 for the number of neurons. For ANN with multi-decay search, we assumed one layer with neurons ranging from 1 to 10. We also used a grid search for the decay parameter of the ANN. In multi-layer ANN, decay takes five random numbers in the range 10^{-10} to 10. For the multi-decay ANN, 30 numbers were randomly chosen from the same range.

The CNN was built based on conventional structures, including input, three convolution layers (16, 32, and 64 filters with 3×3 pixels), max-pooling (2×2 pixels), and ending with a fully connected layer. A rectified linear unit was considered as the activation function, used after each convolution layer and for the fully connected layer. The hyper-parameters selection considered different configurations that were tested in models of varying complexity. The model configuration that presented the best performance in terms of the loss and accuracy functions was chosen. The model was implemented using the Google cloud platform (<https://cloud.google.com>, accessed: 5 June 2021).

3. Results

This section presents the results of our study, which include an exploratory data analysis, ML techniques for COVID-19 versus pneumonia patients and COVID-19 versus healthy patients, as well as the hyper-parameters selection for COVID-19 versus healthy cases.

TABLE 2. Range and selected values for the indicated ML technique, model, and parameter.

ML technique	Model*	Parameter	Range	Selected
SVM	Radial	C	[0.01, 788]	11.55
		Sigma	[5.87E-4, 0.1]	0.01
	Linear	C	[0.05, 880]	3.20
		C	[4.74E-2, 186]	0.99
	Polynomial	Scale	[1.56E-5, 1.73]	1.73
		Degree	{1, 2, 3}	1
RF	100			81
	300			74
	500	Mtry**	{1,...,97}	97
	1500			97
	2000			81

where Model*: refers to the kernel function or to the number of trees for the SVM or RF techniques, respectively; and Mtry**: is a parameter corresponding to the number of variables available for splitting at each tree node in RF.

TABLE 3. Summary statistics of the indicated data set.

Group	Female patients		Male patients		p-value		Total	
	N (%)	Age mean (SD)	N (%)	Age mean (SD)	% (b/z)	Age	N (%)	Age mean (SD)
Healthy	1345 (44)	44.9 (16.6)	1701 (56)	45.7 (16.4)	< 0.001(**)	0.170 ^(n.s.)	3046 (92.9)	45.3 (16.5)
Pneumonia	33 (53)	38.2 (19.1)	29 (47)	48.1 (18.0)	0.590 ^(n.s.)	0.041 ^(n.s.)	62 (1.9)	43.7 (18.6)
COVID-19	56 (38)	53.8 (20.9)	92 (62)	53.1 (20.5)	< 0.001(**)	0.842 ^(n.s.)	148 (4.5)	53.4 (20.7)
Total	1434 (43.6)	45.1 (16.8)	1822 (55.5)	46.2 (16.7)	< 0.001(**)	0.105 ^(n.s.)	3256 (100)	45.1 (16.7)

where ^(b/z): binomial or z-test, ^(W/t): Wilcoxon or t-test, ^(n.s.): non-significant, ^(**): significant at 1%.

3.1 Exploratory data analysis and global performance of ML techniques

Table 3 reports the descriptive statistics for the data under study. There are 3046 cases (instances) of healthy subjects (44% females), 62 cases of pneumonia (53% females), and 174 COVID-19 positive patients, with 148 cases with known sex (38% females). COVID-19 patients have a mean age of 53.4 years (SD = 20.7 years), pneumonia patients have a mean age of 43.7 years (SD = 18.6 years), and healthy patients have a mean age of 45.3 years (SD = 16.5 years). From this table, observe that ages and proportions of females and males are, in general, relatively similar in statistical terms with non-significant *p*-values at 1%, except for proportions in healthy and COVID-19 groups. However, a much larger number of healthy cases (compared to COVID-19 and pneumonia) is detected. Since the instances of COVID-19 and pneumonia are less, four different data-balancing methods were used: down-sampling, up-sampling, SMOTE, and rose. The data-balancing methods were applied before and during the KFRCV re-sampling process. For tuning, threshold, and model selection, we considered those that maximize the AUROC.

The classification performance of AUROC, sensitivity, and specificity metrics for each ML technique is shown in Fig. 3 through boxplots. The left panel depicts this performance to

discriminate among COVID-19 and healthy patients, whereas the right panel displays the results for COVID-19 and pneumonia cases. The figure shows the statistical distribution of the results obtained based on all repeated cross-validation processes and all instances of hyper-parameter search (the total number of trained and tested models). Note that very similar distributions are detected in most cases, except for the RF model when comparing COVID-19 and pneumonia groups, regarding the sensitivity and specificity metrics, where a higher variability and asymmetry are assessed.

3.2 Machine learning for COVID-19 versus healthy patients

Since the number of healthy patients is significantly greater than COVID-19 patients, we use a down-sampling method to create a balanced data set. Such a balancing method is repeated ten times. ANN, RF, and SVM models are applied using 5-fold and 10-fold cross-validation processes for each of the repeats. An alternative approach considers balancing in the cross-validation process, that is, each of the *k*-folds is balanced in terms of the number of COVID-19 and healthy cases so that this is an internal sampling. The classification performance for COVID-19 and healthy patients based on external and internal samplings is shown in Fig. 3-left panel.

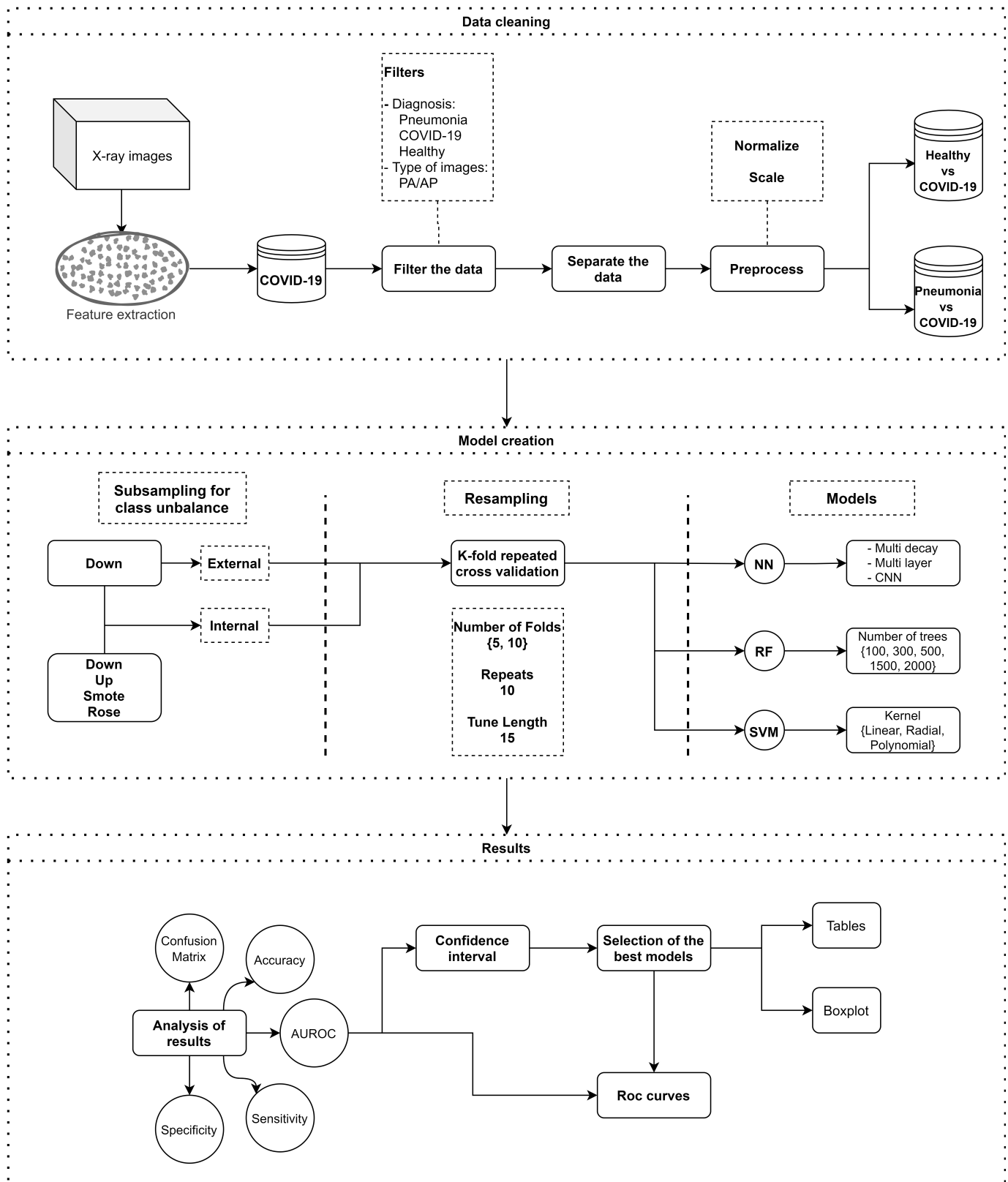


FIGURE 2. Algorithm followed for data analytics, model building, and sampling.

Table 4 reports the AUROC, sensitivity, and specificity averages obtained by each ML technique when comparing COVID-19 versus healthy patients. Results are averaged based on the cross-validation for each of the different hyper-parameters used to train the models. When an external balancing is considered, ANN, RF, and SVM models have a mean AUROC of 84.3%, 86.2%, and 86.1%, respectively.

Note that RF and SVM have very similar AUROC values (no statistical difference with a p -value = 0.33). However, both RF and SVM perform better than the ANN technique (p -value = 0.001 and 0.003, respectively). The mean sensitivity-specificity values for ANN, RF, and SVM were 74.9–77.9%, 74.2–81.1%, and 76.3–79.1%, respectively. The results obtained for the internal down-sampling do not report a

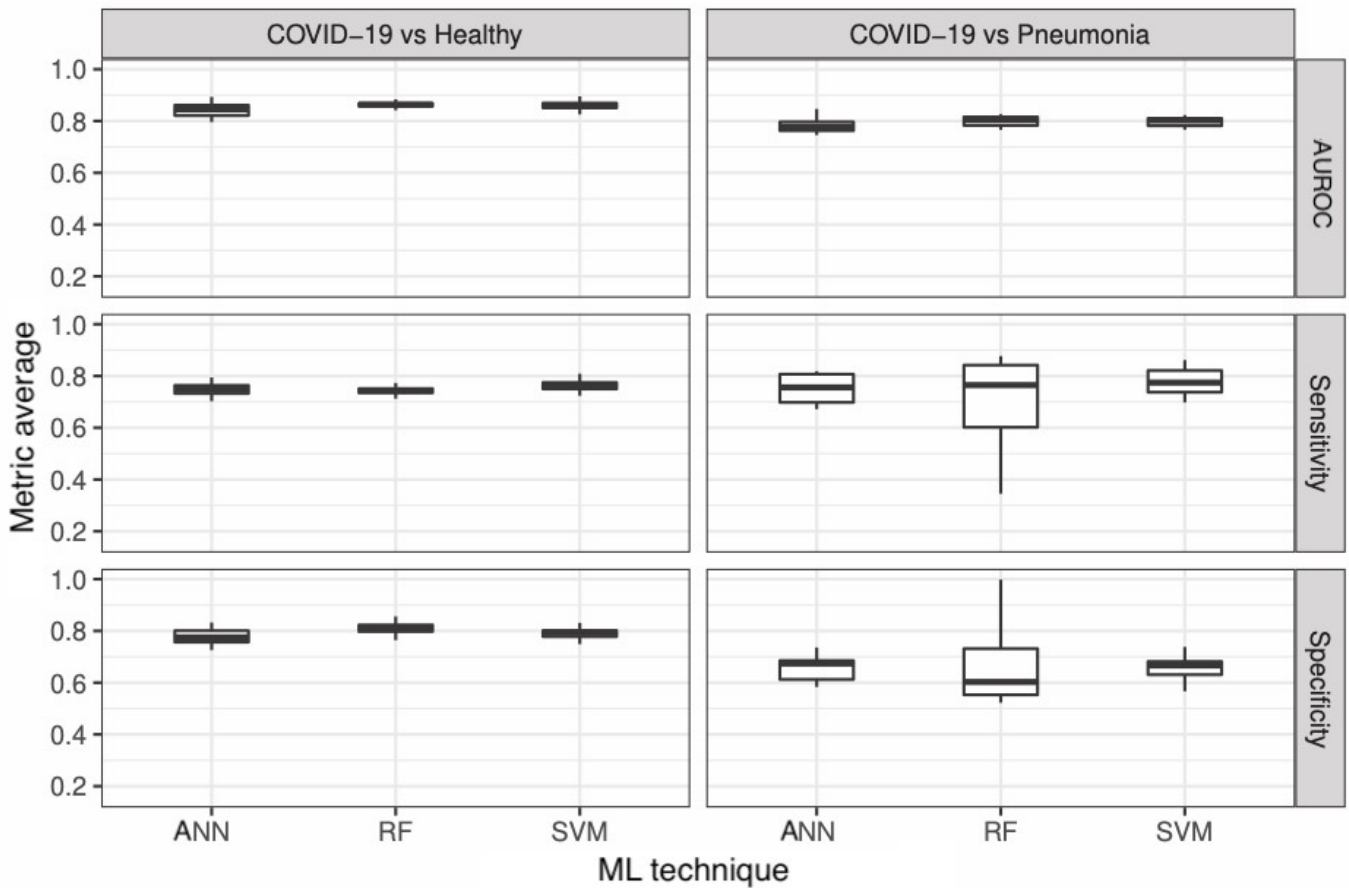


FIGURE 3. Boxplots of the global performance of the indicated ML technique, metric, and group comparison.

significant difference in relation to the external sampling.

TABLE 4. Results (in %) of the indicated sampling, ML technique, and metric for healthy patients.

Sampling	ML technique	AUROC	Sensitivity	Specificity
External	RF	86.23	74.18	81.05
	SVM	86.07	76.25	79.09
	ANN	84.33	74.94	77.87
Internal	RF	86.67	74.94	81.40
	SVM	85.99	76.32	78.80
	ANN	83.53	74.09	77.20

3.3 Machine learning for COVID-19 versus pneumonia patients

Classification models for pneumonia and COVID-19 patients were also developed. Note that the number of positive COVID-19 subjects is greater than pneumonia patients (146 and 62 cases). Nevertheless, the total number of cases (pneumonia and COVID-19) is not sufficiently large. Therefore, balancing methods (down-sampling, up-sampling, rose, and SMOTE) were applied only during the cross-validation re-sampling process (internal method).

Table 5 reports the AUROC, sensitivity, and specificity averages obtained by each ML technique when comparing

COVID-19 versus pneumonia patients. The best-balanced sensitivity-specificity is obtained, for all ML techniques, when the down-sampling method is utilized. High unbalanced sensitivity-specificity is observed particularly for the RF model with the rose method. In addition, the SMOTE provides good sensitivity scores, but it generates low specificity, meaning that it fails at distinguishing pneumonia cases from COVID-19 patients. Classification performance for COVID-19 and pneumonia cases shows higher variability (see Fig. 3) across all ML techniques when compared to the healthy subjects. The high variability observed for sensitivity and specificity results from the poor performance of some data-balancing methods, as observed in Table 5. ANN, RF, and SVM models have a mean AUROC of 78.3%, 80.0%, and 79.8%, respectively. RF and SVM have no statistical difference (p -value = 0.58), but both RF and SVM perform better than the ANN technique (p -value = 0.01 and 0.04, respectively). The mean sensitivity-specificity values for ANN, RF, and SVM were 74.9–65.3%, 64.6–69.7%, and 77.8–65.8%, respectively.

3.4 Hyper-parameters selection for COVID-19 versus healthy cases

Table 6 reports the metric performance for different models defined by hyper-parameters of the ML techniques. Results are presented for the down-sampling method as it reported the best results. Focusing on AUROC values, most models per-

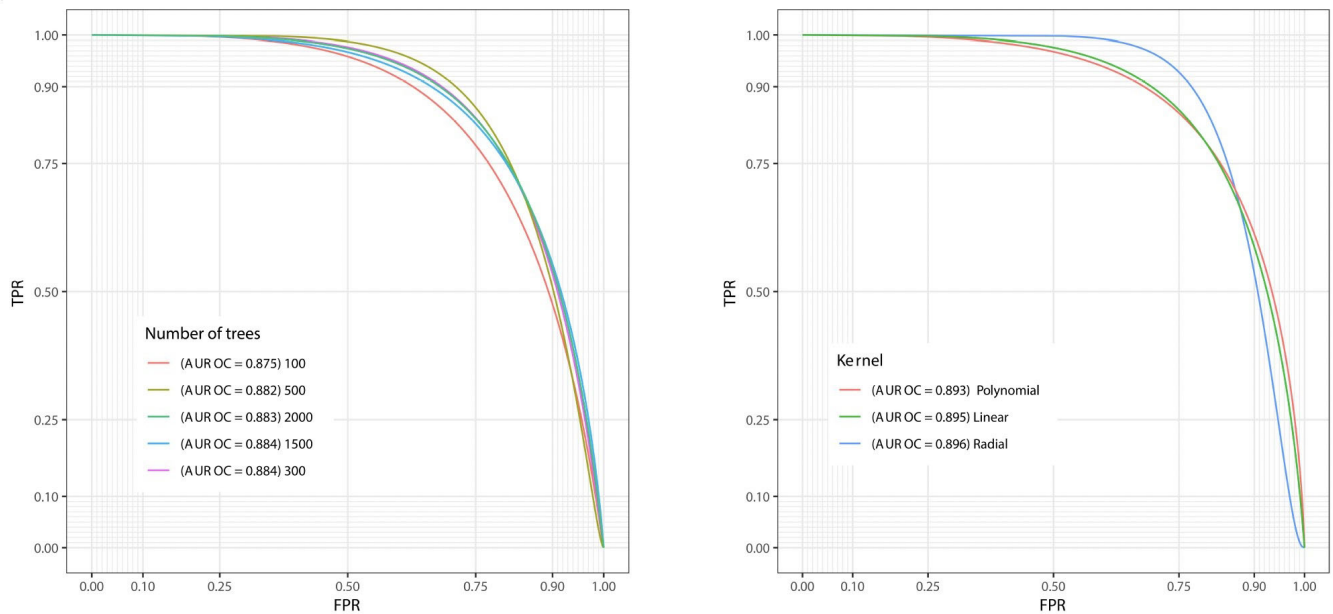


FIGURE 4. ROC for RF (left) and SVM (right) with the indicated number of trees and kernel.

TABLE 5. Results (in %) of the indicated balancing, ML technique, and metric for pneumonia vs COVID-19.

Balancing	ML technique	AUROC	Sensitivity	Specificity
Down-sampling	RF	79.41	69.73	72.79
	SVM	80.10	74.09	70.86
	ANN	77.81	70.21	69.59
Up-sampling	RF	82.08	86.78	53.46
	SVM	81.59	80.97	65.76
	ANN	79.63	80.35	60.85
Rose	RF	77.09	18.97	99.29
	SVM	77.70	72.09	67.15
	ANN	76.27	68.50	68.76
SMOTE	RF	81.54	82.87	59.26
	SVM	79.67	83.90	59.45
	ANN	78.55	80.57	60.42

form similarly, with the best results being obtained by SVM, independently of the chosen kernel function used. Results are described as the mean value with its 95% confidence interval. Also, note that such an interval is narrow, in most cases, indicating a stable performance for most models. A second important observation is that all SVM models presented the best sensitivity-specificity balancing.

Interestingly, both SVM and RF perform better than CNN (for example, AUROC value of 0.89 for SVM against 0.84 for CNN). ANN (multi-decay, multi-layer) and CNN are trained using 5-fold cross-validation due to computational burden. In contrast, SVM and RF are trained and evaluated using 10-fold cross-validation. The high AUCROC scores obtained by both RF and SVM are depicted in Fig. 4. Note that SVM models

show a better balance between the true positive rate (TPR) and FPR, resulting in a higher AUROC score.

For a better model performance analysis, Table 7 presents accuracy and F1-score. Despite being well-known performance metrics, models are selected by maximizing the AUROC. Therefore, these values do not represent the best performance dimensions of the process.

Computational times were extracted for the best models and portrayed in Table 6. Besides the computational burden mentioned, Table 8 also shows that ANN models use almost 500% higher computational time than the next most lasting model (RF-2000). Note that computational times for CNN are not reported. CNN models were trained using the Google cloud platform with significantly different computational technical characteristics.

4. Discussion and final remarks

Conventional chest radiography can be helpful to identify different patterns of lung alterations in COVID-19 patients, allowing us to grade disease severity based on total lung involvement. In addition, X-ray has a shallow radiation dose, and it is more accessible and cheaper than other diagnostic imaging methods, such as CT scans. Furthermore, trained professionals are needed to interpret the results. Thus, the support of tools from artificial intelligence for fast and accurate diagnosis is crucial, especially when resources are scarce, and response times must be immediate to avoid the spread of the disease. This research demonstrated the capabilities of ML techniques to identify COVID-19 cases from healthy or pneumonia patients using X-ray images. The flexibility of simple models and accessibility to X-ray images are needed, particularly for health systems of hospitals located in rural zones that lack infrastructure. This limitation is common in developing countries, as in Latin America, where the number

TABLE 6. Results (in %) for the mean and its 95% CI of the indicated ML technique, model, and metric.

ML technique	Model	k-folds	AUROC	Sensitivity	Specificity
SVM	Radial	10	0.90 (0.90, 0.90)	0.79 (0.77, 0.81)	0.83 (0.81, 0.85)
	Polynomial	10	0.89 (0.88, 0.90)	0.81 (0.79, 0.83)	0.81 (0.79, 0.82)
	Linear	10	0.90 (0.88, 0.90)	0.80 (0.78, 0.82)	0.81 (0.79, 0.82)
RF	1500	10	0.89 (0.87, 0.90)	0.76 (0.74, 0.78)	0.82 (0.80, 0.83)
	2000	10	0.89 (0.87, 0.90)	0.76 (0.74, 0.77)	0.82 (0.80, 0.84)
	300	10	0.88 (0.87, 0.90)	0.76 (0.74, 0.78)	0.82 (0.80, 0.84)
	500	10	0.88 (0.87, 0.89)	0.76 (0.74, 0.79)	0.82 (0.80, 0.84)
	100	10	0.88 (0.87, 0.89)	0.75 (0.74, 0.77)	0.81 (0.79, 0.83)
ANN	Multi-decay	5	0.89 (0.88, 0.90)	0.80 (0.77, 0.81)	0.83 (0.81, 0.85)
	Multi-layer	5	0.86 (0.85, 0.88)	0.76 (0.73, 0.78)	0.81 (0.79, 0.83)
	Convolutional	5	0.85 (0.82, 0.87)	0.76 (0.71, 0.80)	0.74 (0.68, 0.79)

TABLE 7. Results (in %) for the mean and its 95% CI of the indicated ML technique, model, and metric.

ML technique	Model	Accuracy	F1-score
SVM	Radial	81.20 (80.03, 82.38)	80.65 (79.35, 81.95)
	Polynomial	80.71 (79.39, 82.02)	80.59 (79.20, 81.99)
	Linear	80.31 (78.89, 81.73)	80.08 (78.57, 81.58)
RF	100	76.81 (75.26, 78.35)	76.06 (74.45, 77.67)
	300	76.46 (75.02, 77.90)	75.69 (74.16, 77.23)
	500	76.75 (75.31, 78.19)	76.00 (74.46, 77.55)
	1500	76.67 (75.22, 78.11)	75.80 (74.24, 77.36)
	2000	76.49 (75.07, 77.92)	75.73 (74.19, 77.27)

TABLE 8. Runtimes (in seconds) of overall cross-validation for the indicated ML technique, model, and metric.

ML technique	Model	Runtime (in seconds)
SVM	Radial	103.11
	Linear	876.14
	Polynomial	87.83
RF	100	111.20
	300	323.56
	500	535.30
	1500	1598.52
	2000	2138.20
NN	Multi-decay	9537.21
	Multi-layer	9496.19
	Convolutional	-

of COVID-19 cases has reached dramatically high levels in countries like Argentina, Brazil, Chile, Colombia, and Peru. Besides, the lack of appropriate infrastructure generates an increasing need for diagnosis techniques that can be efficiently utilized and updated.

In this context, ML techniques offer the appropriate flexi-

bility without a compromise in predictive accuracy. However, ML models have not been thoroughly studied in the existing literature, as they are generally used as benchmarks for more complex algorithms, such as CNN or other DL models. Nevertheless, often CNN models are an impractical solution due to their high computational requirements and lack of access to data and infrastructure in rural areas of developing countries. Thus, the results obtained in our study offer higher discriminatory capability in comparison to other standard clinical-based diagnosis techniques. For instance, scientists from Johns Hopkins Medicine [52] have shown that, over the four days of infection before the typical time of symptom onset, the mean FNR is 67% with a 95% confidence interval of 27% to 94%. Another recent systematic review reinforces the need for repeated testing in patients with suspicion of COVID-19 infection, given that up to 29% of patients could have an initial PCR false-negative result [53]. Furthermore, PCR's limitations include the need for higher-level laboratory facilities, proper swab sample techniques, clinician experience, and error-free patient samples [54, 55]. Hence, if clinical suspicion is high, an infection should not be ruled out based on a PCR test only [52]. Therefore, as shown in this study, ML models are suitable tools that provide comparable or better results than clinical tests. Thus, ML models can be used as a support tool for fast initial screening of patients when diagnosing for COVID-19.

The full factorial design of experiment used in this paper

provided a more stable discriminatory accuracy or classification power of ML techniques. Indeed, we accounted for design parameters, such as different data balancing methods, that must be studied in a context that arises in a COVID19 pandemic (a large number of PCR tests and a small number of positive cases). Among the evaluated models, SVM obtained the best results when correctly identifying infected patients. SVM with different kernel functions provided AUROC values as high as 90% with well-balanced sensitivity-specificity scores of approximately 80% with narrow confidence intervals.

It is important to note that the results presented in our study can be further improved. Our findings were obtained from a relatively small set of positive infected patients (174) with publicly available X-ray exams. Our findings must be validated and improved with access to both larger sets of images and exams with better image quality. Therefore, there seems to have great opportunities to further enhance the accuracy of ML techniques' results. This is significantly important since ML is transforming the current medical practice and plays a critical role in the digital healthcare era [56]. These ML techniques help doctors to improve their diagnostic ability, impacting patients' prognostic, quality of life, and effective treatment prescriptions. Moreover, the use of such techniques has also implications for the health systems, reducing exam time, saving costs for the medical practice, and improving daily productivity [57].

AUTHOR CONTRIBUTIONS

NB, MT, and FF designed the study, collected the data, and created the models and codes. NB, FF, and VL performed the data analysis. NB, MT, FF, and VL drafted the paper. GD and NG performed a clinical development and validated the analysis and results. NB, FF, and VL revised the paper critically for important intellectual contents. All authors approved the final version to be published.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

ACKNOWLEDGMENT

The authors thank four Reviewers for their constructive comments on an earlier version of this paper.

FUNDING

This research was partially supported by grants FONDECYT 11181176 (F. Feijoo) and 1200525 (V. Leiva) from the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge, and Innovation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Sun P, Lu X, Xu C, Sun W, Pan B. Understanding of COVID-19 based on current evidence. *Journal of Medical Virology*. 2020; 92: 548–551.
- [2] Velavan TP, Meyer CG. The COVID-19 epidemic. *Tropical Medicine & International Health*. 2020; 25: 278–280.
- [3] Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*. 2020; 92: 418–423.
- [4] Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, *et al.* World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*. 2020; 76: 71–76.
- [5] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, *et al.* A novel coronavirus from patients with pneumonia in China. *The New England Journal of Medicine*. 2020; 382: 727–733.
- [6] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020; 395: 497–506.
- [7] Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, *et al.* Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: A descriptive, cross-sectional, multicenter study. *American Journal of Gastroenterology*. 2020; 115: 766–773.
- [8] Ren LL, Wang YM, Wu ZQ, Xiang ZC, Guo L, Xu T, *et al.* Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chinese Medical Journal*. 2020; 133: 1015–1024.
- [9] Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity*. 2020; 109: e102433.
- [10] Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus in Wuhan, China. *Journal of Medical Virology*. 2020; 92: 441–447.
- [11] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*. 2020; 382: 1199–1207.
- [12] Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, *et al.* Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*. 2020; 296: E115–E117.
- [13] Al-Tawfiq JA, Memish ZA. Diagnosis of SARS-CoV-2 infection based on CT scan vs RT-PCR: reflecting on experience from MERS-CoV. *Journal of Hospital Infection*. 2020; 105: 154–155.
- [14] Wang Y, Kang H, Liu X, Tong Z. Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak. *Journal of Medical Virology*. 2020; 92: 538–539.
- [15] Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, *et al.* SARS-CoV-2 Viral load in upper respiratory specimens of infected patients. *New England Journal of Medicine*. 2020; 382: 1177–1179.
- [16] Dai W, Zhang H, Yu J, Xu H, Chen H, Luo S, *et al.* CT Imaging and differential diagnosis of COVID-19. *Canadian Association of Radiologists Journal*. 2020; 71: 195–200.
- [17] Garg M, Prabhakar N, Gulati A, Agarwal R, Dhooira S. Spectrum of imaging findings in pulmonary infections. Part 1: bacterial and viral. *Polish Journal of Radiology*. 2019; 84: e205–e213.
- [18] El Asnaoui K, Chawki Y. Using X-ray images and deep learning for automated detection of coronavirus disease. *Journal of Biomolecular Structure and Dynamics*. 2020; 1–12.
- [19] Ye Z, Zhang Y, Wang Y, Huang Z, Song B. Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. *European Radiology*. 2020; 30: 4381–4389.
- [20] Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, *et al.* Chest CT Findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. *Radiology*. 2020; 295: e200463.
- [21] Ng M, Lee EYP, Yang J, Yang F, Li X, Wang H, *et al.* Imaging Profile of the COVID-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*. 2020; 2: e200034.
- [22] Feijoo F, Palopoli M, Bernstein J, Siddiqui S, Albright TE. Key indicators of phase transition for clinical trials through machine learning. *Drug Discovery Today*. 2020; 25: 414–421.

- [23] Vaishya R, Javaid M, Khan IH, Haleem A. Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2020; 14: 337–339.
- [24] Li L, Qin L, Xu Z, Wang X, Yin Y, Kong B, *et al.* Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*. 2020; 296: E65–E71.
- [25] Özkaya U, Öztürk Ş, Barstugan M. Coronavirus (COVID-19) classification using deep features fusion and ranking technique. In *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*. Springer, Cham. 2020; pp. 281–295.
- [26] Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, *et al.* Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell*. 2020; 181: 1423–1433.
- [27] Polat C, Karaman O, Karaman C, Korkmaz G, Balci MC, Kelek SE. COVID-19 diagnosis from chest X-ray images using transfer learning: Enhanced performance by debiasing dataloader. *Journal of X-Ray Science and Technology*. 2021; 29: 19–36.
- [28] Özkaya U, Öztürk S, Budak S, Melgani F, Polat K. Classification of COVID-19 in chest CT images using convolutional support vector machines. *arXiv*. 2021. (in press)
- [29] Sethy PK, Behera SK, Ratha PK, Biswas P. Detection of coronavirus disease (COVID-19) based on deep features and support vector machine. Preprint. 2021. (in press)
- [30] Öztürk S, Özkaya U, Barstugan M. Classification of coronavirus (COVID-19) from X-ray and CT images using shrunken features. *International Journal of Imaging Systems and Technology*. 2021; 31: 5–15.
- [31] Maia M, Pimentel JS, Pereira IS, Gondim J, Barreto ME, Ara A. Convolutional support vector models: prediction of coronavirus disease using chest X-rays. *Information*. 2020; 11: 548.
- [32] Palacios CA, Reyes-Suarez JA, Bearzotti LA, Leiva V, Marchant C. Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. *Entropy*. 2021; 23: 485.
- [33] Park J, Lee J, Sim D. Low-complexity CNN with 1D and 2D filters for super-resolution. *Journal of Real-Time Image Processing*. 2020; 17: 2065–2076.
- [34] Valdivia M. Public health infrastructure and equity in the utilization of outpatient health care services in Peru. *Health Policy and Planning*. 2002; 17: 12–19.
- [35] Lan K, Wang D, Fong S, Liu L, Wong KKL, Dey N. A Survey of Data Mining and Deep Learning in Bioinformatics. *Journal of Medical Systems*. 2018; 42: 139.
- [36] Gray DM, Anyane-Yeboah A, Balzora S, Issaka RB, May FP. COVID-19 and the other pandemic: populations made vulnerable by systemic inequity. *Nature Reviews Gastroenterology and Hepatology*. 2020; 17: 520–522.
- [37] McMahon DE, Peters GA, Ivers LC, Freeman EE. Global resource shortages during COVID-19: bad news for low-income countries. *PLoS Neglected Tropical Diseases*. 2020; 14: e0008412.
- [38] Martinez DA, Hinson JS, Klein EY, Irvin NA, Saheed M, Page KR, *et al.* SARS-CoV-2 Positivity rate for latinos in the Baltimore–Washington, DC Region. *The Journal of the American Medical Association*. 2020; 324: 392–395.
- [39] Thakur N, Lovinsky-Desir S, Bime C, Wisnivesky JP, Celedón JC. The structural and social determinants of the racial/ethnic disparities in the US COVID-19 pandemic. what’s our role? *American Journal of Respiratory and Critical Care Medicine*. 2020; 202: 943–949.
- [40] Banerjee P, Dehnbostel FO, Preissner R. Prediction is a balancing act: importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical data sets. *Frontiers in Chemistry*. 2018; 6: 362.
- [41] Sommer C, Gerlich DW. Machine learning in cell biology-teaching computers to recognize phenotypes. *Journal of Cell Science*. 2013; 126: 5529–5539.
- [42] Sun Y, Li L, Zheng L, Hu J, Li W, Jiang Y, *et al.* Image classification base on PCA of multi-view deep representation. *Journal of Visual Communication and Image Representation*. 2019; 62: 253–258.
- [43] Haralick RM. Statistical and structural approaches to texture. *Proceedings of the IEEE*. 1979; 67: 786–804.
- [44] Chen YQ, Nixon MS, Thomas DW. Statistical geometrical features for texture classification. *Pattern Recognition*. 1995; 28: 537–552.
- [45] Chebira A, Barbotin Y, Jackson C, Merryman T, Srinivasa G, Murphy RF, *et al.* A multiresolution approach to automated classification of protein subcellular location images. *BMC Bioinformatics*. 2007; 8: 210.
- [46] Liu S, Mundra PA, Rajapakse JC. Features for cells and nuclei classification. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, Boston, MA, USA. 2011.
- [47] Pau G, Fuchs F, Sklyar O, Boutros M, Huber W. EBIImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics*. 2010; 26: 979–981.
- [48] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria. 2020. Available at: <https://www.R-project.org/> (Accessed: May 20 2020).
- [49] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017.
- [50] Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. COVID-19 image data collection: prospective predictions are the future. *ArXiv*. 2020. (in press)
- [51] Chawla N, Bowyer K, Hall L, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. 2002; 16: 321–357.
- [52] Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based SARS-CoV-2 tests by time since exposure. *Annals of Internal Medicine*. 2020; 173: 262–267.
- [53] Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, Zambrano-Achig P, Del Campo R, Ciapponi A, *et al.* False-negative results of initial TR-PCR assays for COVID-19: a systematic review. *PLoS ONE*. 2020; 15: e0242958.
- [54] Bachelet VC. Do we know the diagnostic properties of the tests used in COVID-19? A rapid review of recently published literature. *Medwave*. 2020; 20: e7890.
- [55] Watson J, Whiting PF, Brush JE. Interpreting a COVID-19 test result. *British Medical Association*. 2020; 369.
- [56] Aykroyd RG, Leiva V, Ruggeri F. Recent developments of control charts, identification of big data sources and future trends of current research. *Technological Forecasting and Social Change*. 2019; 144: 221–232.
- [57] Mesko B, Gyorffy Z. The rise of the empowered physician in the digital health era. *Journal of Medical Internet Research*. 2019; 21: e12490.

How to cite this article: Nicolás Bustos, Manuel Tello, Guillermo Droppelmann, Nicolás García, Felipe Feijoo, Víctor Leiva. Machine learning techniques as an efficient alternative diagnostic tool for COVID-19 cases. *Signa Vitae*. 2021. doi:10.22514/sv.2021.110.