A Novel Web Platform for COVID-19 diagnosis using X-Ray exams and Deep Learning Techniques

Virgínia Xavier Nunes*†, Aldísio G. Medeiros*‡, Raylson Silva de Lima*‡, Luís Fabrício de F. Souza*‡
Victor Hugo C. de Albuquerque ‡, and Pedro P. Rebouças Filho*†‡

* Laboratório de Processamento de Imagens, Sinais e Computação Aplicada (LAPISCO),
† Federal Institute of Education, Science and Technology of Ceará (IFCE), Ceará, Brazil
‡ Federal University of Ceará (UFC), Fortaleza, Ceará, Brazil,
Email: {virginia.nunes, aldisio.medeiros, fabricio.freitas}@lapisco.ifce.br, raylson.lima@alu.ufc.br,
victor.albuquerque@ieee.org, pedrosarf@ifce.edu.br

2021 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-3900-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/IJCNN52387.2021.9533469

***Abstract*—Modern computer vision techniques applied to radiographic studies are presented as an alternative to assist the specialist in screening and diagnosing the respiratory syndrome (SARS-CoV-2), assisting in clinically severe cases, such as acute pneumonia, acute respiratory failure, organ failure, and death. This work proposes a screening method based on the Internet of Medical Things (IoMT) based on deep learning techniques for the classification of COVID-19 from chest X-ray (CXR) exams. The proposed system called Computer-Aided Remote medical diagnostics System (CARMEDSys) applied to the diagnosis of COVID-19 consists of three main stages: 1) segmentation of the lung region in X-ray images, 2) deep extraction of attributes from the filtered pulmonary area and 3) Prediction patient status with machine learning assistance. The performance of CARMEDSys was evaluated considering twelve different deep neural networks, via the transfer of learning. Besides, the performance of this approach is evaluated against recent studies for the classification of healthy patients, with pneumonia, or with COVID-19. The evaluation methodology considered two different sets of radiographic images, reaching Sensitivity (99.97%), F1-Score (99.43%), and Accuracy (98.89%) promising to distinguish patients with pneumonia and COVID-19 combining DenseNet201 as attribute extractor with Support Vector Machine with radial basis function, exceeding up to 12.31% sensitivity for prediction of COVID-19 recent related works.**

***Index Terms*—COVID-19, Coronavirus, Deep Learning, Chest X-ray images, IoMT**

I. Introduction

According to the World Health Organization (WHO), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is caused by a new coronavirus discovered in 2019. It is an infectious disease that, when it becomes more serious, is able to progress for critical pneumonia with severe respiratory failure, organ failure, and death [1].

Previous studies on influenza [2] showed that the risk of death for critically ill patients was higher in those who had some comorbidity, such as chronic obstructive pulmonary disease, cardiovascular disease, or hypertension. Likewise, people who had a history of previous illnesses increased their predisposition to develop respiratory failure by up to 4 times in cases of SARS-CoV-2 [3] coronavirus infection.

The high degree of contagion of the disease contributes to many infected people, and this significant increase in the demand for clinical care overburdens health facilities. Thus, health professionals are subjected to high workloads, directly compromising the cognitive performance, perception, and decision-making of specialists. Also, healthcare professionals are increasingly at risk of infection because they often deal with infected patients. In this way, computer-aided diagnostic systems (CAD) work by offering a tool to support this professional in the interpretation of X-ray exams, providing a second opinion regarding the findings in the exam, with efficiency and precision that do not depend on the workload, which the system will be exposed, in addition to contributing to the response time and diagnostic performance.

The application of the Internet of Things (IoT) concepts in medical diagnostics systems has demonstrated a significant increase in productivity, mobility, and monitoring of results. When using the Internet of Medical Things (IoMT) [4] in the development of systems, we will have greater assistance, considering that we will be able to connect several electronic devices and use the ecosystem of technologies from the Internet. IoMT adds, in addition to technology and greater autonomy on the part of professionals in the field, an innovative means of connecting medical assistance to patients who lack health services [5].

In this work, we propose a new solution based on CNNs in the classification of COVID-19 based on chest X-ray (CXR) exams. Fine-tuning and transfer learning techniques were used to detect, segment, and extract lung characteristics. Machine learning techniques were applied to solve the classification task. Two data sets were selected to solve this task. To solve the segmentation task, we used the Chest X-Ray Masks and Labels [6] image base containing 704 exams with their respective ground truth provided by specialists. For the classification task, we used the COVIDx [7] dataset composed of 16,756 chest radiographs in 13,645 patient cases.

Therefore, we combine the principles of IoMT with CAD systems to face the quarantine and pandemic of COVID-19, proposing a system to assist the medical diagnosis called Computer-Aided Remote Medical Diagnostics System (CARMEDSys), offering greater accessibility, flexibility, speed and accuracy of clinical diagnoses through digital

images. Specifically, this article offers the following study and contributions:

1) Fine-Tunning the Detectron2 convolutional neural network using the Chest X-Ray Masks and Labels dataset to detect and segment the lung region of CXR exams.
2) Perform two classification steps: in the first step, we have a binary classification to identify whether the exam is Normal (healthy) or Sick; if the exam classifies as a patient, the second classifier will diagnose in Pneumonia or COVID-19.
3) A quick and accessible approach to lung disease prediction useful for screening patients.
4) Proposal for a medical diagnosis aid system based on IoMT.

II. Related Works

Motivated by the need for a more agile diagnosis and less dependence on the radiologist to assess chest X-ray images of patients suspected of being infected by COVID-19, several studies based on deep learning have been proposed.

A. Convolutional Neural Networks

This subsection presents some recent work that proposes new techniques based on CNNs to detect COVID-19 using CXR.

Wang and Wong [7] proposed COVID-Net, a CNN to classify chest X-ray images as Normal, Pneumonia, or COVID-19. The work also presents a dataset of CXR images, named COVIDx, composed of 16,756 radiographs of 13,645 patients. More specifically, 8066 of the tests are healthy, 5526 have pneumonia without COVID-19, and 76 have COVID-19. CNN COVID-Net was trained and evaluated on the COVIDx dataset reaching 95%, 91%, and 80% sensitivity for healthy people, pneumonia, and Covid-19, respectively.

Farooq and Hafeez [8] proposed COVID-ResNet, a CNN originated from the pre-trained ResNet-50 network and followed by fine-tuning for the classification of COVID-19 in CXR images. The model can identify other types of illnesses such as bacterial pneumonia and viral pneumonia, in addition to detecting healthy patients from a normal X-ray. The authors used a version of the COVIDx data set as dataset containing 1203 radiographs of healthy patients (Normal), 931 radiographs of bacterial pneumonia, 660 radiographs of viral pneumonia, and 68 radiographs of patients positive for COVID-19.

B. Transfer Learning

Several works related to the detection of COVID-19 are presented in the current literature. The search for patterns in medical images using deep learning techniques brings new proposals in diagnosis by medical images for the detection of COVID-19 during the pandemic. This is the case of the study by Ioannis D. et al. [9] who proposed the automatic detection of COVID-19 in X-ray images using transfer learning through Convolutional Neural Networks, the study obtained 96% Accuracy, demonstrating the effectiveness in CNN Networks. However, the study used a limited set of medical images, compromising the generalization of the approach.

Narin et al. [10] proposed an approach based on the transfer of learning using the pre-trained models ResNet50, InceptionV3, and Inception-ResNetV2 to automatically predict patients with COVID-19 through chest X-ray images, achieving an accuracy of 98% in the best model. However, a notable disadvantage is the number of images used in the study, consisting of only 100 images in total: 50 images of normal patients' CXR and 50 images of patients with COVID-19.

C. Applications of IoMT Systems in Medical Images

The Internet of Medical Things (IoMT) can be defined as well-recognized Internet principles, tools, and techniques aimed especially at medical and health sectors and domains. The definition of IoMT proposes to integrate health resources and instruments to technology through Internet applications and electronic devices. This subsection features recent work that was developed by implementing the IoMT approach to tackle the COVID-19 pandemic and offer remote medical assistance to patients during this pandemic block.

Singh, Ravi Pratap, et al. [11] and Singh, Ravi Pratap, et al. [12] proposed an IoMT approach to provide medical treatment to orthopedic patients during the COVID-19 pandemic. The work offers better monitoring of patients in addition to making medical assistance in remote locations viable.

III. Materials and methods

In this section, we present the techniques that contributed to the formulation of the proposed approach. In the subsection III-A we present the exams dataset, followed by the main methods for extracting characteristics in subsection III-C, segmentation in subsection III-B, and classification models in subsection III-D. Finally, we present the evaluation metrics in subsection III-E proposed in this work.

A. Datasets

The data set used for the segmentation task was published and made available by Jaeger et al. [6]. The set consists of CXR images containing radiographs of healthy patients and patients with the manifestation of tuberculosis. In total, 406 radiographs are normal, and 394 radiographs have tuberculosis. The images are provided in Digital Imaging and Communications in Medicine (DICOM) format with different dimensions.

On the other hand, COVIDx, proposed by Wang and Wong [7], was used for the classification stage. This dataset consists of 16,756 chest radiography images from 13,645 patients. This set has three classes: Normal, Pneumonia, and COVID-19. Only AP radiographs were considered, where AP refers to the direction of the x-ray crossing the patient from the back towards the front of the chest. More specifically, the COVIDx data set consists of 8066 radiographic images of healthy

patients, 5526 patients with pneumonia, and 76 of patients with COVID-19.

In the current version of the data set used in this study, there were 8848 radiographs of healthy patients, 5,991 radiographs of patients with pneumonia, and 142 radiographs of patients infected with COVID-19. Clearly, the cases of COVID-19 are considerably less than the number of cases in the other classes. Therefore, we applied the augmentation technique to the set of instances of COVID-19.

The augmentation application was made in the training set after the separation of training and testing of the cases of COVID-19. In this sense, rotation techniques with angles within the range of [-20, 20], horizontal flip, and translation within the range of [-100, 100] pixels were applied. The set separation was 90% for training and 10% for testing each class at random.

B. Detectron2 for Lung Detection and Segmentation in X-ray scans

This subsection presents the convolutional neural network Detectron2, a recent Mask R-CNN proposed and maintained Facebook Artificial Intelligence Research (FAIR) [13].

CNN Detectron2, originated from the Mask R-CNN [14] framework, was proposed by the Facebook Artificial Research in Intelligence (FAIR) group. The framework is implemented in PyTorch. Detectron2 includes high-quality implementations of object detection algorithms. The choice of this framework for the detection and segmentation stage is due to its modularity, ease, and speed to perform the training, and high performance in working with medical images, as can be seen in Han, Tao et al. [15].

C. Deep Features Extractors

The application of deep learning models is still limited in the medical field, as there are not enough data sets to train these models from the beginning adequately [16]. To address this problem, an approach called Transfer Learning has been extensively investigated and applied to computer-assisted diagnosis [17]. This approach consists of transposing the knowledge acquired from one context to another. CNN models are trained in large datasets and with a large number of classes, such as ImageNet [18].

In this work, we consider the output of the last convolutional layer as deep attributes. The layer output is adjusted to a one-dimensional vector format. This vector represents the input image attributes and will be used as input for the classification algorithms. The CNN models for the deep extraction of attributes used in this work were: Xception [19], MobileNet [20], VGG [21], Inception-ResNet-V2 e InceptionV3 [22], DenseNet [23], NASNetMobile [24], e ResNet [25].

D. Classifiers

The technique of extracting attributes from CNNs combined with machine learning classifiers has been used successfully to solve robust classification tasks in medical images, as can be seen in the work of Nobrega et al. [17]. In order to obtain better results, we combined five classifiers with the attributes extracted from each CNN, the best combination being included in the approach solution.

The classifiers used in this work were Naive-Bayes Classifier [26], Multi-Layer Perceptron (MLP) [27], k-Nearest Neighbors (kNN) [28], Random Forest [29], e Support Vector Machine [30] com kernel Linear, Polinomial e RBF.

E. Evaluation Metrics

Four evaluation metrics were used to verify the classifiers performance: Accuracy, F1-Score, Recall, and Precision. The calculation is done through the confusion matrix generated by the classifiers, according to True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) [31].

IV. A NOVEL WEB PLATFORM FOR A COVID-19 REMOTE DIAGNOSIS

IoMT-based systems are innovative because they integrate technologies with medical resources and equipment to assist and collaborate in healthcare professionals' workflow. Also, these systems provide ample connectivity and agility in accessing and exchanging important information for the patient's diagnosis.

This section describes the Computer-Aided Remote Medical Diagnostics System (CARMEDSys) platform inspired by IoMT to assist in the diagnosis of COVID-19. A workflow based on computer vision techniques built on a WEB ecosystem with communication through a mobile application is proposed. The proposed methodology presents, as a first stage, the detection, and segmentation of the pulmonary region, later, in the second stage, the extraction of characteristics from the exams and, finally, in the third stage, the diagnosis prediction. Different deep learning techniques are evaluated in each stage, such as CNN, deep extractors, and transfer of learning.

A. Remote Access Web Platform

CARMEDSys presents the following workflow: **Step 1** - Health professional accesses the WEB system and uploads the CXR image; **Step 2** - The image loaded and distributed to one of the processing nodes; **Step 3** - Detection and segmentation of the lung region; **Step 4** - Extraction of characteristics of the region of interest using deep extractors; **Step 5** - Prediction of the diagnosis through the attributes extracted in two steps, **Step 5.a** - Classification of the examination between healthy and ill, and **Step 5.b** - Classification of the examination between pneumonia and COVID-19. In Figure 1 presents a flowchart of the proposed methodology.

1) Step 1 - Healthy of Things System Access: The professional can access CARMEDSys through browsers or smartphones. If access via the browser, the user will be directed to the WEB version, then the authentication credentials will be requested. However, if the access is made through the smartphone, the user must install the mobile application that will be made available with authorization. In

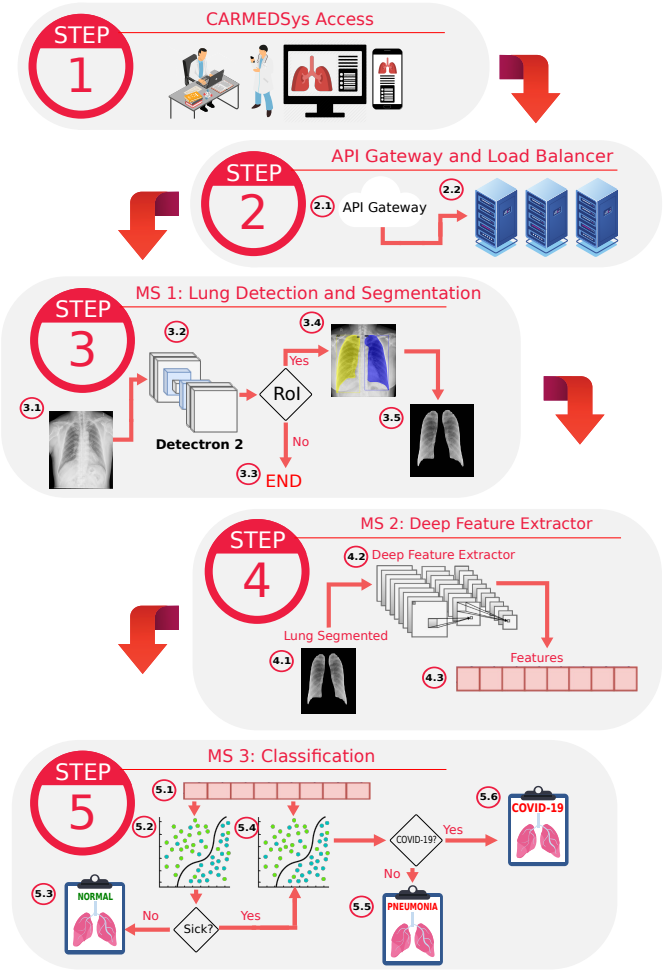

Fig. 1. The figure shows the architecture of the proposed application based on the IoMT. Step 1 illustrates the moment of access to the WEB system and mobile application by the health professional. Step 2 is responsible for distributing the flow of requests to ensure that services are processed quickly and safely. Step 3 consists of detecting and segmenting the CXR lung region using Detectron2. After the lung is segmented, in step 4, the attributes are extracted using deep extractors. Finally, in step 5, the classification process begins. The first classification identifies whether the image represents a healthy lung or any pathologies: Pneumonia or COVID-19. When the system prediction is Normal, the patient has no infection, and the process ends. If any manifestation of the disease is identified, a second classification is made to diagnose the examination in Pneumonia or COVID-19.

both cases, the user has the option to upload an CXR exam image. All accesses are recorded in an audit for control and security of access to patient data.

2) Step 2 - API Gateway and Load Balancer: The CARMEDSys architecture is structured in multiple processes that meet the various diagnostic prediction requests produced by the platform users. For processing to happen in a parallel and distributed way, thus offering fast and secure processing.

3) Step 3: Lung Detection and Segmentation Service: As shown in 1, step 3 is responsible for the first main stage of this solution and represents the phase of detection and segmentation of the pulmonary region. In III-B presented the neural network Detectron2, which is used here to detect and segment the lung region in X-ray images. From the dataset presented in the III-A, the network training phase used 704 CXR images. From the image submitted by the user in Stage 1 (in 1 -3.1), Detectron2 is applied to identify if there is a region of interest (RoI) in the image (in 1 -3.2). If not, there is no lung region in the image and the processing flow ends (in 1 -3.3). Otherwise, a pulmonary region was detected and segmented by Detectron (in 1 -3.4). Therefore, the result of this stage is the segmented lung region (in 1 -3.5) represented in an image free of findings that could act as noise, such as bone structure and organs adjacent to the lung.

4) Step 4: Deep Extraction: The pulmonary feature extraction step is constructed from pre-processed images from the previous stage (in 1 -4.1). Inspired by deep extractors, this phase of the methodology proposes using extracting features with CNN from CXR images. To find the best deep extractor, all CNN models described in the III-C were evaluated. These CNNs have pre-training weights in the ImageNet dataset.

5) Step 5 - Classification: Finally, the classification stage is performed using the deep attributes extracted in the previous step (in 1 - 5.1). According to 1 - 5.2 and 5.3, the methodology for predicting the diagnosis of patients consists of two steps. The objective of the first step is to classify a sample in Normal or Sick classes (in 1 - 5.3). If the patient is classified as Normal, the flow ends. However, if the predicted class is sick, the flow goes to the second sub-stage, which classifies the sample as pneumonia or COVID-19 (in 1 - 5.4).

V. Results and Discussions

The results is organized as follows: In the subsection V-A, we present the results of the proposed methodology as a classification binary problem for X-ray exams in Normal or Sick (Pneumonia + COVID-19), in this section the results of the best combination of deep extractor and classification algorithm with the proposed solution for the first stage. The second stage, evaluated in subsection V-B, corresponds to the stage where the CXR exam classified as sick, is classified as Pneumonia or COVID-19. Finally, the third stage, V-C, combine the models of stage one and two to produce a single automatic system for detecting COVID-19. In this last section, we validate our method of comparing our best results with the works proposed in the literature on the same data set.

A. First Stage: Normal vs. Sick Binary Classification

This stage is the first classification stage, it consists of classifying the segmented lung region. First, we organized the segmented dataset into two categories: healthy exams as Normal and exams with Pneumonia and COVID-19 as Sick.

Then, we apply deep extraction techniques using the Xception, ResNet50, MobileNet, VGG16, VGG19, InceptionV3, InceptionResNetV2, NASNetMobile, DenseNet121, DenseNet169 and DenseNet201 extractors described in III-C, generating 11 new datasets. We applied the classic machine learning methods Naive Bayes, MLP,

Nearest Neighbors, Random Forest, and SVM with the Linear, Polynomial, and RBF kernels, described in III-D.

The training set was normalized by an average of zero and unit variation, and the test set was also normalized using the same normalization rule as the training set. To find the best combination of hyperparameters, value ranges for the hyperparameters were generated for each classifier, and the grid search technique with cross-validation of k-folds was applied. The hyperparameters that achieved the highest accuracy in the validation set (average of 10-folds) were selected as the best hyperparameters.

The grid search technique with 10-folds validation was adopted to choose the best hyperparameters for each classifier. The number of neurons in the hidden layer of the MLP classifier varied within the range of [2, 1001] neurons. The *k* hyperparameter of the kNN classifier varied between [1, 9] selecting only odd values. For Random Forest, the maximum number of decision trees was 1500. For the SVM classifier with Linear Kernel, the range of *C* varied between $[2^{-5}, 2^{15}]$. The SVM with Polynomial Kernel, on the other hand, the degree of the polynomial varied between the values [3, 5, 7, 9] and the hyperparameter *C* between $[2^{-5}, 2^{15}]$. Finally, for SVM with Kernel RBF, the range of *C* and γ ranged from $[2^{-5}, 2^{15}]$ e $[2^{-15}, 2^{3}]$, respectively.

Table I presents the values for the metrics of Accuracy (ACC), Sensitivity (SEN), Positive Predictive Value (PPV), and F-Score for all the extractor-classifier combinations for the first classification step Normal vs. Sick. In the case of disease detection, the sensitivity metric is important because it reveals how sensitive the system is to the detection of sick patients. The greater the sensitivity, the greater the system's capacity to identify the patients who present the disease. The Positive Predictive Value (PPV) metric indicates the efficiency of the system in not detecting false positives.

Once the evaluation metrics were clarified in the paragraph above, to select the best extractor-classifier model, we chose the combination that reached the highest average between the F1-Score and Sensitivity metrics. We chose F1-Score because it represents a weighting between Sensitivity and Positive Predictive Value in a single value. We also chose Sensitivity because it represents the quality of the model in identifying infected patients. Thus, we weigh the F1-Score with Sensitivity. Selecting the best model only for the highest F1-Score was not very effective for our problem, as the F1-Score can be elevated due to a high PPV, but not sensitive to disease. However, it is better to treat a patient who eventually does not have COVID-19, but has been classified as if he did, than to stop treating a patient who has COVID-19 because he was wrongly classified without COVID-19.

Thus, considering the Table I, the combination that achieved the highest average between SEN and F1-Score and chosen as the model proposed for the first classification stage was the DenseNet121 extractor combined with the SVM Linear classifier, reaching 93.95% SEN, 91.03% PPV, 92.46% F1-Score and 91.40% ACC. The average value between SEN and F1-Score was 93,205. The value found for the hypertext parameter *C* of the SVM classifier with Linear Kernel was *C* = 4.

TABLE I
ACCURACY (ACC), SENSITIVITY (SEN), POSITIVE PREDICTIVE VALUE (PPV), AND F-SCORE ACHIEVED FOR EACH EXTRACTOR-CLASSIFIER COMBINATION FOR THE FIRST NORMAL VS. SICK BINARY CLASSIFICATION STEP IN THE VALIDATION SET.

Model	Classifier	ACC(%)	SEN(%)	PPV(%)	F-Score(%)
Xception	Bayes	80.07	84.43	80.45	82.40
	MLP	89.94	89.46	93.67	91.52
	Nearest Neighbors	82.14	79.16	93.90	85.90
	Random Forest	87.07	88.23	89.65	88.93
	SVM Linear	89.40	90.53	91.26	90.89
	SVM Polynomial	90.47	89.81	94.25	91.97
	SVM RBF	84.94	84.40	90.80	87.48
ResNet50	Bayes	79.74	81.30	84.48	82.86
	MLP	87.14	88.07	90.00	89.02
	Nearest Neighbors	83.94	80.86	94.71	87.24
	Random Forest	85.80	88.51	86.78	87.63
	SVM Linear	87.47	88.83	89.65	89.24
	SVM Polynomial	87.54	91.09	87.01	89.00
	SVM RBF	88.67	88.88	91.95	90.39
MobileNet	Bayes	87.40	89.54	88.62	89.08
	MLP	89.34	91.18	90.34	90.76
	Nearest Neighbors	87.60	84.33	96.55	90.03
	Random Forest	90.07	91.48	91.37	91.43
	SVM Linear	91.13	91.63	93.21	92.42
	SVM Polynomial	89.94	90.80	91.95	91.37
	SVM RBF	91.13	92.21	92.52	92.36
VGG16	Bayes	85.87	87.55	88.16	87.85
	MLP	90.13	89.84	93.56	91.66
	Nearest Neighbors	86.67	83.50	95.97	89.30
	Random Forest	89.80	92.32	89.88	91.08
	SVM Linear	90.80	94.41	89.42	91.85
	SVM Polynomial	91.53	91.69	93.90	92.78
	SVM RBF	89.47	94.16	87.24	90.57
VGG19	Bayes	84.94	86.42	87.81	87.11
	MLP	90.40	91.15	92.41	91.78
	Nearest Neighbors	85.80	82.68	95.51	88.64
	Random Forest	89.60	91.41	90.57	90.99
	SVM Linear	90.93	91.99	92.41	92.20
	SVM Polynomial	91.07	92.20	92.41	92.30
	SVM RBF	91.67	92.96	92.64	92.80
InceptionV3	Bayes	80.34	85.98	78.96	82.32
	MLP	87.20	93.23	84.02	88.39
	Nearest Neighbors	81.61	83.44	85.17	84.30
	Random Forest	74.75	95.37	59.31	73.13
	SVM Linear	84.67	95.45	77.24	85.38
	SVM Polynomial	88.67	92.27	87.81	89.98
	SVM RBF	85.80	88.51	86.78	87.63
InceptionResNetV2	Bayes	81.27	83.20	84.82	84.00
	MLP	88.87	87.59	94.13	90.74
	Nearest Neighbors	83.87	80.42	95.40	87.27
	Random Forest	86.60	88.40	88.50	88.45
	SVM Linear	89.94	91.27	91.37	91.32
	SVM Polynomial	81.61	88.07	78.96	83.27
	SVM RBF	90.60	91.65	92.18	91.91
NASNetMobile	Bayes	74.08	78.73	75.74	77.21
	MLP	86.54	89.57	86.89	88.21
	Nearest Neighbors	80.87	81.04	87.47	84.13
	Random Forest	82.54	87.81	81.14	84.34
	SVM Linear	78.81	90.35	71.03	79.53
	SVM Polynomial	80.94	80.86	87.93	84.25
	SVM RBF	82.34	92.30	75.86	83.28
DenseNet121	Bayes	86.14	87.91	89.19	88.18
	MLP	91.27	92.96	92.64	92.80
	Nearest Neighbors	87.40	84.70	95.51	89.78
	Random Forest	88.47	92.44	87.24	89.76
	SVM Linear	**91.40**	**93.95**	**91.03**	**92.46**
	SVM Polynomial	87.20	90.16	87.47	88.79
	SVM RBF	90.47	93.32	90.00	91.63
DenseNet169	Bayes	86.00	87.16	88.96	88.05
	MLP	90.87	90.94	93.56	92.23
	Nearest Neighbors	87.54	85.16	95.05	89.84
	Random Forest	86.60	92.82	83.33	87.82
	SVM Linear	89.94	90.16	92.75	91.44
	SVM Polynomial	90.20	92.28	90.68	91.47
	SVM RBF	91.40	91.95	93.33	92.64
DenseNet201	Bayes	88.40	89.01	91.26	90.12
	MLP	91.60	91.24	94.59	92.88
	Nearest Neighbors	86.74	83.65	95.86	89.34
	Random Forest	85.54	94.78	79.42	86.42
	SVM Linear	91.47	92.93	92.29	92.61
	SVM Polynomial	90.67	91.10	92.98	92.03
	SVM RBF	89.07	93.79	86.89	90.21

B. Second Stage: Pneumonia vs. COVID-19 Binary Classification

The Second Stage corresponds to the second classification stage, which consists of classifying the segmented lung region from the segmentation stage and classified as Sick in the first classification stage. In this step, another classifier trained to identify Pneumonia without COVID-19 and Pneumonia with COVID-19 is applied to classify the exams that showed Pneumonia manifestation in the first classification step.

Initially, we selected from the segmented COVIDx set only radiography exams categorized as Pneumonia and COVID-19.

Similar to the first classification step, we applied the deep extraction techniques to the images with only the segmented pulmonary region using the Xception, ResNet50, MobileNet, VGG16, VGG19, InceptionV3, InceptionResNetV2, NASNetMobile, DenseNet121, DenseNet169 and DenseNet201 extractors present in subsection III-C, generating 11 new datasets. After extracting the dataset, we apply the classic Naive Bayes, MLP, Nearest Neighbors, Random Forest, and SVM machine learning methods with the Linear, Polynomial, and RBF kernels, described in subsection III-D.

The training set was normalized by an average of zero and unit variation, and the test set was also normalized using the same normalization rule as the training set. To find the best combination of hyperparameters, ranges of values for the hyperparameters were generated for each classifier equivalent to that used in the first classification step. The hyperparameters that achieved the highest accuracy in the validation set (average of 10-folds) were selected as the best hyperparameters.

Table II presents the values for the metrics of Accuracy (ACC), Sensitivity (SEN), Positive Predictive Value (PPV), and F-Score for all extractor-classifier combinations for the second classification step Pneumonia vs. COVID-19 binary. Exactly as in the first classification stage, the best extractor-classifier model selected as the most adequate to solve the problem was the one that reached the highest average between the metrics F1-Score and SEN. Choosing the best model considering only the highest F1-Score value is not recommended since its high value may be due to the high value of PPV, however, the SEN is low, leading the system not to identify the people really infected.

Thus, considering the Table II, the combination that achieved the highest average between SEN and F1-Score and chosen as the model proposed for the second classification stage was the DenseNet201 extractor combined with the SVM RBF classifier, reaching 99.67% SEN, 99.19% PPV, 99.43% F1-Score and 98.89% ACC. The average value between SEN and F1-Score was 99.55. The values found for the hyperparameters C and γ of the SVM classifier with Kernel RBF was $C = 32$ and $\gamma = 2$.

TABLE II
ACCURACY (ACC), SENSITIVITY (SEN), POSITIVE PREDICTIVE VALUE (PPV), AND F-SCORE ACHIEVED FOR EACH EXTRACTOR-CLASSIFIER COMBINATION FOR THE SECOND STAGE OF BINARY CLASSIFICATION PNEUMONIA VS. COVID-19 IN THE VALIDATION SET.

Model	Classifier	ACC(%)	SEN(%)	PPV(%)	F-Score(%)
Xception	Bayes	78.60	98.98	78.96	87.84
	MLP	96.03	99.66	96.27	97.94
	Nearest Neighbors	92.23	99.30	92.71	95.89
	Random Forest	95.72	99.00	96.60	97.78
	SVM Linear	96.83	99.17	97.57	98.36
	SVM Polynomial	96.51	99.66	96.76	98.19
	SVM RBF	98.09	99.34	98.70	99.02
ResNet50	Bayes	87.32	98.90	88.02	93.15
	MLP	95.72	99.00	96.60	97.78
	Nearest Neighbors	91.28	99.29	91.74	95.37
	Random Forest	96.67	98.69	97.89	98.29
	SVM Linear	95.87	99.16	96.60	97.86
	SVM Polynomial	97.93	99.18	98.70	98.94
	SVM RBF	98.73	99.03	99.67	99.35
MobileNet	Bayes	80.66	99.60	80.58	89.08
	MLP	96.98	99.83	97.08	98.44
	Nearest Neighbors	95.24	99.66	95.46	97.52
	Random Forest	96.98	99.01	97.89	98.45
	SVM Linear	91.91	99.82	91.90	95.70
	SVM Polynomial	98.41	99.83	98.54	99.18
	SVM RBF	98.41	99.67	98.70	99.18
VGG16	Bayes	96.19	98.05	98.05	98.05
	MLP	96.35	99.50	96.76	98.11
	Nearest Neighbors	96.35	99.66	96.60	98.11
	Random Forest	94.13	99.48	94.49	96.92
	SVM Linear	86.21	99.81	86.08	92.44
	SVM Polynomial	93.66	99.65	93.85	96.66
	SVM RBF	97.78	99.67	98.05	98.85
VGG19	Bayes	96.98	98.07	98.86	98.46
	MLP	94.77	99.32	95.30	97.27
	Nearest Neighbors	94.92	99.49	95.30	97.35
	Random Forest	95.40	99.16	96.11	97.61
	SVM Linear	87.00	99.62	87.05	92.91
	SVM Polynomial	96.35	99.50	96.76	98.11
	SVM RBF	98.57	99.35	99.19	99.27
InceptionV3	Bayes	88.58	98.57	89.64	93.89
	MLP	96.19	99.50	96.60	98.02
	Nearest Neighbors	91.12	99.29	91.58	95.28
	Random Forest	97.14	98.23	98.86	98.54
	SVM Linear	96.51	99.50	96.92	98.19
	SVM Polynomial	94.61	99.49	94.98	97.18
	SVM RBF	98.09	99.02	99.02	99.02
InceptionResNetV2	Bayes	87.63	99.09	88.18	93.32
	MLP	92.39	99.30	92.88	95.98
	Nearest Neighbors	90.49	99.29	90.93	94.93
	Random Forest	95.24	98.83	96.27	97.54
	SVM Linear	89.69	99.64	89.80	94.46
	SVM Polynomial	92.39	99.13	93.04	95.99
	SVM RBF	96.35	99.33	96.92	98.11
NASNetMobile	Bayes	74.16	98.92	74.43	84.94
	MLP	95.87	99.33	96.44	97.86
	Nearest Neighbors	88.74	99.63	88.83	93.92
	Random Forest	95.87	99.16	96.60	97.86
	SVM Linear	88.74	98.92	89.48	93.96
	SVM Polynomial	88.58	99.63	88.67	93.83
	SVM RBF	95.56	99.49	95.95	97.69
DenseNet121	Bayes	97.30	97.92	99.35	98.63
	MLP	94.61	99.65	94.82	97.18
	Nearest Neighbors	94.61	99.32	95.14	97.19
	Random Forest	95.40	98.67	96.60	97.62
	SVM Linear	90.01	99.64	90.12	94.64
	SVM Polynomial	96.19	99.50	96.60	98.02
	SVM RBF	98.25	99.51	98.70	99.10
DenseNet169	Bayes	97.93	98.09	99.83	98.95
	MLP	92.55	99.48	92.88	96.06
	Nearest Neighbors	93.50	99.31	94.01	96.59
	Random Forest	84.94	99.24	85.27	91.73
	SVM Linear	89.69	99.46	89.96	94.47
	SVM Polynomial	97.14	99.34	97.73	98.53
	SVM RBF	97.62	99.34	98.22	98.77
DenseNet201	Bayes	96.51	98.06	98.38	98.22
	MLP	93.81	99.31	94.33	96.76
	Nearest Neighbors	94.13	99.48	94.49	96.92
	Random Forest	97.46	98.54	98.86	98.70
	SVM Linear	89.69	99.11	90.29	94.49
	SVM Polynomial	96.51	99.66	96.76	98.19
	SVM RBF	**98.89**	**99.67**	**99.19**	**99.43**

C. Comparison against recent methods

To validate our experiments, in this subsection we compare our approach which consists of the integration of the segmentation stages, followed by the classification stage of the pulmonary region between Normal vs. Sick, concluding with the classification stage between Pneumonia vs. COVID-19, with the proposed method by Wang and Wong [7]. The combinations between the DenseNet121 extractor with the SVM Linear classifier for the first classification step and the DenseNet201 extractor with the SVM RBF classifier for the second classification step were chosen to obtain the best performance considering the average between SEN and F1-Score metrics in the validation set. Thus, we seek to ensure that the system operates in its best configuration, identifying the largest number of infected and reducing the number of false positives in new tests as much as possible, giving greater importance to recognizing the disease.

Although the current version of the COVIDx dataset acquired to carry out the experiments of this work is larger than the version used in the work of Wang and Wong [7], the distribution of images by class to perform the training of the models was similar. More specifically, the number of images in the Normal class used by Wang and Wong was 7966, while the number of images for the Normal class used in this work

was 7978 images, a difference of only 12 images. Also, the number of images in the Pneumonia class used by Wang and Wong for training was 5426, while the number of images in the Pneumonia class used in this work was 5373, a total of 53 fewer images.

In addition, the number of images of the class COVID-19 used by Wang and Wong were 76 images, while the number of images with the manifestation of COVID-19 used in this work was 142 and, after augmentation, the number of images with COVID-19 it became 2064. However, the author's Wang and Wong in their work, applied the augmentation technique. They did not provide the new distribution of exams by class, thus hindering a more appropriate and fair comparative study.

Also, note that our test set size is larger, making our assessment results more general. The number of images defined for testing by Wang and Wong was not the same as those chosen for this work. Unlike Wang and Wong [7], who selectively selected only 100 images from the Normal class, 100 images from the Pneumonia class and ten images from the COVID-19 class, this work randomly separated 10% of the total number of images from each class, resulting in a much larger number of images per class for testing. More specifically, 870 images of Normal radiography were used for testing in this work, while Wang and Wong used only 100 images, a difference of 770 fewer images.

Considering the Pneumonia class, 618 images were used for testing in this work, while Wang and Wong used only 100 images, that is, 518 fewer images to be tested. As for the COVID-19 class, although the number of images for training is 2064, augmentation images did not leak into the test set, thus avoiding possible anomalies in the test set. Therefore, only 13 images were randomly separated from the test set. Similarly, Wang and Wong [7] selected only ten images for the COVID-19 test.

Table III shows the results generated by our best model based on this study compared to the work of Wang and Wong [7]. Our experiments used a much larger number of samples for the test set than the number of samples from Wang and Wong [7] considering the same COVIDx data set. Despite this, our work surpassed the Sensitivity metric of the Pneumonia and COVID-19 classes by 2.2% and 12.31%, respectively, compared to the work of Wang and Wong [7]. This shows our model's high performance and efficiency in identifying patients positive for COVID-19, and even patients with Pneumonia.

Considering the Positive Predictive Value metric, our model obtained 89% for Pneumonia and 71% for COVID-19. These metrics' values suffered considerably due to the imbalance of the classes in the test set of our distribution. Even so, we overcame the work of Wang and Wong [7] with 4.03% in the PPV metric for the Normal class.

Figure 2 shows the confusion matrix obtained by our best approach in the test set. Through the confusion matrix, we can carefully observe and discuss the results achieved.

We can observe, through the confusion matrix in Figure 2, that our approach failed to correctly classify only a small sample of COVID-19, classifying it as Pneumonia. However, our model did not erroneously classify any sample of COVID-19 as Normal, an important result considering that this would result in discharging a sick patient without adequate follow-up or treatment.

TABLE III
COMPARATIVE TABLE OF THE METRICS OBTAINED IN THE TEST SET BY THE PROPOSED METHOD WITH THE COVIDX DATASET COMPARED WITH THE WORK OF WANG AND WONG [7].

Method	**Our Approach**		Wang and Wong [7]	
	SEN(%)	PPV(%)	SEN(%)	PPV(%)
COVID-19	92.31	71	80	88.9
Pneumonia	93.2	89	91	93.8
Normal	92	95.33	95	91.3

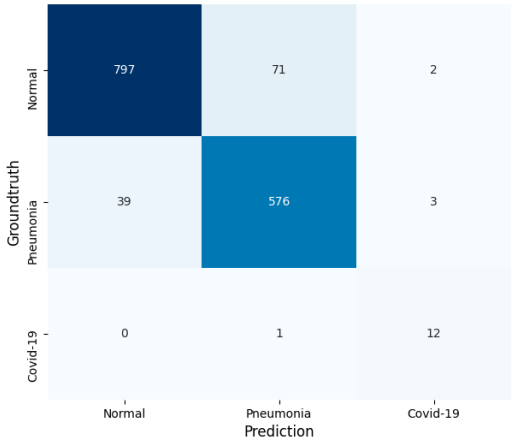

Fig. 2. Confusion matrix obtained by the approach proposed in the test set.

Still, in Figure 2, we can see 576 Pneumonia samples were classified correctly. Only three samples of Pneumonia were identified as COVID-19. Still, a very low number of false positives is considered for COVID-19.

Considering the Normal class, Figure 2 shows that our model correctly identified 797, or 91.60% of all Normal samples, failing to classify 73 samples as Sick. Considering these 73 Normal samples wrongly classified, the probability of the model classifying as Pneumonia was 97.26% against 2.74% for COVID-19, indicating very few false positives for COVID-19.

The proposed method proved to be superior in identifying patients with the manifestation of COVID-19 and Pneumonia through X-ray examinations compared to the work of Wang and Wong [7]. In addition, our approach has ensured greater generalization by experimenting with our method on a much larger set of tests than the compared work, providing greater confidence.

VI. CONCLUSION AND FUTURE WORKS

This work aimed to develop an IoMT-based medical aid system to detect COVID-19 via X-ray exams. Our approach is based on techniques of deep learning and transfer learning. The approach had two main stages: after the identification and automatic segmentation of the pulmonary region, we

have the classification of patients as healthy or sick based on radiographic exams; then, patients classified as sick are classified between paneumonia or COVID-19.

The results were very satisfactory, considering the set of CXR COVIDx images. Our best model obtained 92.31% sensitivity for the COVID-19 class, 93.2 % for the Pneumonia class, and 92% for the Normal class, exceeding up to 12.31% sensitivity for prediction of covid19 recentrelated works

The proposed method works with a total classification time of fewer than three seconds. This is especially important for scenarios in which response time is essential to start the treatment flow, increasing healthcare professionals' productivity and efficiency.

For future studies, we propose to try our approach on a larger set of CXR images, mainly containing more real samples of COVID-19, to balance the classes. Also, we propose to investigate new CNNs for the segmentation of the lung region performed in the first stage. We intend to apply other machine learning techniques for the classification task, such as Mixing Gaussians and Mixing Specialists.

ACKNOWLEDGMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nıvel Superior - Brasil (CAPES) - Finance Code 001". Also Pedro Pedrosa Rebouças Filho acknowledges the sponsorship from the Brazilian National Council for Research and Development (CNPq) via Grants Nos. 431709/2018-1 and 311973/2018-3.

REFERENCES

[1] Marco Cascella, Michael Rajnik, Arturo Cuomo, Scott C Dulebohn, and Raffaela Di Napoli. Features, evaluation and treatment coronavirus (covid-19). In *Statpearls [internet]*. StatPearls Publishing, 2020.

[2] Dominik Mertz, Tae Hyong Kim, Jennie Johnstone, Po-Po Lam, Stefan P Kuster, Shaza A Fadel, Dat Tran, Eduardo Fernandez, Neera Bhatnagar, Mark Loeb, et al. Populations at risk for severe or complicated influenza illness: systematic review and meta-analysis. *Bmj*, 347:f5061, 2013.

[3] Wei-jie Guan, Wen-hua Liang, Yi Zhao, Heng-rui Liang, Zi-sheng Chen, Yi-min Li, Xiao-qing Liu, Ru-chong Chen, Chun-li Tang, Tao Wang, et al. Comorbidity and its impact on 1590 patients with covid-19 in china: A nationwide analysis. *European Respiratory Journal*, 2020.

[4] Gulraiz J Joyia, Rao M Liaqat, Aftab Farooq, and Saad Rehman. Internet of medical things (iomt): applications, benefits and future challenges in healthcare domain. *J Commun*, 12(4):240–7, 2017.

[5] Noemi Scarpato, Alessandra Pieroni, Luca Di Nunzio, and Francesca Fallucchi. E-health-iot universe: A review. *management*, 21(44):46, 2017.

[6] Stefan Jaeger, Sema Candemir, Sameer Antani, Yì-Xiáng J Wáng, Pu-Xuan Lu, and George Thoma. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery*, 4(6):475, 2014.

[7] Linda Wang and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. *arXiv preprint arXiv:2003.09871*, 2020.

[8] Muhammad Farooq and Abdul Hafeez. Covid-resnet: A deep learning framework for screening of covid19 from radiographs. *arXiv preprint arXiv:2003.14395*, 2020.

[9] Ioannis D Apostolopoulos and Tzani A Mpesiana. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, page 1, 2020.

[10] Ali Narin, Ceren Kaya, and Ziynet Pamuk. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *arXiv preprint arXiv:2003.10849*, 2020.

[11] Ravi Pratap Singh, Mohd Javaid, Abid Haleem, Raju Vaishya, and Shokat Al. Internet of medical things (iomt) for orthopaedic in covid-19 pandemic: Roles, challenges, and applications. *Journal of Clinical Orthopaedics and Trauma*, 2020.

[12] Ravi Pratap Singh, Mohd Javaid, Abid Haleem, and Rajiv Suman. Internet of things (iot) applications to fight against covid-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 2020.

[13] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019.

[14] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[15] Tao Han, Virgínia Xavier Nunes, Luís Fabrício De Freitas Souza, Adriell Gomes Marques, Iágson Carlos Lima Silva, Marcos Aurélio Araujo Ferreira Junior, Jinghua Sun, and Pedro P Rebouças Filho. Internet of medical things—based on deep learning techniques for segmentation of lung and stroke regions in ct scans. *IEEE Access*, 8:71117–71135, 2020.

[16] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016.

[17] Raul Victor Medeiros da Nóbrega, Solon Alves Peixoto, Suane Pires P da Silva, and Pedro Pedrosa Rebouças Filho. Lung nodule classification via deep transfer learning in ct lung images. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, pages 244–249. IEEE, 2018.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[19] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[20] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[21] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[22] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*, 2017.

[23] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[24] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8697–8710, 2018.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[26] S Theodoridis and K Koutroumbas. Pattern recognition–fourth edition, 2009.

[27] Simon Haykin. *Neural Networks and Learning Machines, 3/E*. Pearson Education India, 2010.

[28] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[29] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[30] Vladimir Vapnik and Vlamimir Vapnik. Statistical learning theory wiley. *New York*, 1:624, 1998.

[31] Mohammad Hossin and MN Sulaiman. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2):1, 2015.