Covid-19 Detecting in Computed Tomography Lungs Images Using Machine and Transfer Learning Algorithms

Dalila Cherifi[1,*], Abderraouf Djaber[1], Mohammed-Elfateh Guedouar[1], Amine Feghoul[1], Zahia Zineb Chelbi[1] and Amazigh Ait Ouakli[2]
[1]Institute of Electrical and Electronic Engineering, University of Boumerdes, Algeria
[2]Chahids Mahmoudi Hospital, Radiology Service, Tizi Ouzou, Algeria
E-mail: da.cherifi@univ-boumerdes.dz, raouf23899@gmail.com, fateh05gdr@gmail.com, aminefeghoul55@gmail.com, zahiazina35@gmail.com, amazigh.aitouakli@hcm-dz.com
[*]Corresponding author

Keywords: COVID19, CT- Scans, KNN, SVM, RF, CNN, transfer learning, InceptionV3, ResNet50V2

Received: June 27, 2022.

Coronavirus disease 2019 (COVID-19), a rapidly spreading infectious disease, has led to millions of deaths globally and has had a significant impact on public healthcare due to its association with severe lung pneumonia. The diagnosis of the infection can be categorized into two main approaches, a laboratory-based approach and chest radiography approach where the CT imaging tests showed some advantages in the prediction over the other methods. Due to restricted medical capacity and the fast-growing number suspected cases, the need for finding an immediate, accurate and automated method to alleviate the overcapacity of radiology facilities has emerged. In order to accomplish this objective, our work is based on developing machine and deep learning algorithms to classify chest CT scans into Covid and non-Covid classes. To obtain a good performance, the accuracy of the classifier should be high so the patients may have a clear idea about their state. For this purpose, there are many hyper parameters that can be changed in order to improve the performance of the artificial models that are used for the identification of such illnesses. We have worked on two non-similar datasets from different sources, a small one consisting of 746 images and a large one with 14486 images. On the other hand, we have proposed various machine learning models starting by an SVM which contains different kernel types, KNN model with changing the distance measurements and an RF model with two different number of trees. Moreover, two CNN based approaches have been developed considering one convolution layer followed by a pooling layer then two consecutive convolution layers followed by a single pooling layer each time. The machine learning models showed better performance compared to CNN on the small dataset, while on the larger dataset, CNN outperforms these algorithms. In order to improve the performance of the models, transfer learning has also been used where we trained the pre-trained InceptionV3 and ResNet50V2 on the same datasets. Among all the examined classifiers, the ResNet50V2 achieved the best scores with 86.67% accuracy, 93.94% sensitivity, 81% specificity and 86% F1-score on the small dataset while the respective scores on the large dataset were 97.52%, 97.28%, 97.77% and 98%. Experimental interpretation advises the potential applicability of ResNet50V2 transfer learning approach in real diagnostic scenarios, which might be of very high usefulness in terms of achieving fast testing for COVID19.

Povzetek: Raziskava se osredotoča na razvoj algoritmov strojnega in globokega učenja za razvrščanje CT posnetkov prsnega koša v razrede Covid in ne-Covid. Rezultati kažejo, da je pristop prenosa učenja ResNet50V2 najbolj učinkovit za hitro testiranje COVID-19.

1 Introduction

SARS-CoV-2, the virus causing COVID-19, is a part of Coronaviruses family of related RNA viruses found in birds and mammalian species. It has been rapidly spreading in many countries around the world until it was declared a pandemic by the World Health Organization. This virus causes illness such as respiratory tract infections or gastrointestinal diseases, where the effect on the respiratory system can range from mild to lethal. The common symptoms vary from coughing, fever, fatigue, to the loss of taste or smell during the early phases. The virus is transmitted mainly through respiratory routes and has an incubation period of 2 to 14 days. The diagnosis of an infection are generally based on reverse transcription polymerase chain reaction (RT-PCR) methods. Moreover inefficiency, RT-PCR test kits are in vast deficiency. As a result, many infected cases cannot be timely recognized and continue to infect others unconsciously. Thus, alternative methods such as computed tomography (CT) scans have been used. CT imaging manifest clear radiological findings of COVID-19 abnormalities in lungs including Ground-glass opacities, consolidation, subpleural reticulation and they are hopeful in serving as a more efficient and accessible testing manner due to the wide availability of CT devices that can generate results at a fast speed. Due to the crisis caused by the current pandemic, computer-aided CT scan diagnosis /detection should be employed to help radiologists in the diagnosis process to avoid overwhelming the capacity of radiology facilities. Machine Learning (ML) is a major part of Artificial Intelligence (AI) which allows computers to learn on their own without being explicitly programmed. Deep Learning is a division of ML which develops a layered, hierarchical architecture of learning inspired by AI emulating the deep, layered learning process of the primary sensorial areas of the neocortex in the human brain called neurons. One of its types is the Convolutional Neural Networks (CNN) which is mostly used in image classification and image and video recognition. Furthermore, the feature extraction is integrated in the training process with CNNs, this independence from human effort and prior knowledge in feature design is a most important advantage.

The objective of this article is to develop and implement classification approaches using machine and deep learning models in order to discover presence of COVID-19 findings in CT lungs images. This article consists of four sections, after the introduction, the second section aims to give a brief description about the abnormal cases we are going to deal with for COVID-19 and CT findings on infected lungs. The third section describes the useful of covid-19 diagnosis with computer tomography (CT-Scan). The fourth section explains the used machine and deep learning classifiers, which are, K-Nearest Neighbors, Random Forest, Support Vector Machine, Convolutional Neural Network, InceptionV3 and ResNet50V2 models. Ultimately, the fifth part describes the experimental results achieved from using the algorithms presented in section two on both the small and large datasets we have chosen and then a discussion on the models, the results and the impact of the datasets.

2 Covid-19 diagnosis with CT-SCANs

In December 2019, a pneumonia outbreak was reported in Wuhan, China, then traced back to a novel strain of coronavirus diseases [1], subsequently given the interim name 2019-nCoV by the World Health Organization (WHO), later named SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) or COVID-19 (coronavirus disease 2019) [2]. On March 11, 2020, it was declared a pandemic that needs a global coordinated effort to stop the further spread of the virus. As of 25 June 2021, there have been 180,738,944 confirmed cases of COVID-19, including 33,915,110 deaths and 165,394,305 recovered cases, reported to WHO[3]. Coronaviruses are a family of related RNA viruses found in avian and mammalian species that resemble each other in morphology and chemical structure. These viruses cause illness such as respiratory tract infections or gastrointestinal diseases, where the effect on the respiratory system can range from mild to lethal. While more lethal varieties can cause SARS, MERS, and COVID-19, mild illnesses in humans include some cases of the common cold (which is also caused by other viruses). COVID-19 diagnosis involves analyzing samples to assess the current or past presence of SARS-CoV-2. There have been many ways to test the presence of the virus used by authorities around the world. The most common methods are: RT-PCR, X-ray, Computer Tomography (CT scan). A computed tomography (CT scan) is a medical imaging method used in radiology for diagnostic purposes to get exhaustive images of the body. This method is also used to detect the presence of Covid-19 related and anomalies in lungs where it shows a better sensitivity that the other methods [4], [5]. The chest imaging findings of COVID-19 were first published in January 2020 and included bilateral lung involvement and ground-glass opacities in the majority of hospitalized patients. Since then, a myriad of articles on chest CT findings in COVID-19 have been published at a rapid pace. The appropriate use of chest CT in patients with COVID-19 should be based on experience and, above all, the scientific evidence that has emerged since the outbreak of this disease, which keeps accumulating. CT scans are valuable for managing complex cases, assessing clinical deterioration, and excluding other diagnoses. They are particularly recommended for patients with critical illnesses who are suspected of having or have been tentatively diagnosed with COVID-19, as this enables swift decision-making for treatment and enhances the safety of healthcare professionals [4]. Several studies have been published reporting chest CT findings in COVID-19. However, many studies are limited by selection bias, potential blinding issues, and potential confounding of chest CT findings owing to the simultaneous presence of other lung diseases. Nearly all authors of studies who investigated the chest CT appearance of COVID-19 investigated CT performed in symptomatic patients. The pulmonary histology findings of COVID-19, which are characterized by acute and organizing diffuse alveolar damage, resemble those observed in other coronavirus infections, including SARS-CoV-1 and MERS-CoV. Accordingly, the reported chest CT abnormalities in COVID-19 are similar to those seen in infections with SARS-CoV-1 and MERS-CoV[4], [5]. The prevalence of chest CT abnormalities in COVID-19 is dependent on the stage and severity of the disease. Four evolutionary stages have been reported [4]: early phase, progression phase, peak phase, resolution phase. With the vast spreading of the virus in some regions, a new challenge is appearing when the prediction results of the imaging systems have to be automated by computers to save doctors time for supervising the critical cases. Here, the idea is to find optimum methods to get a better result in predicting the positive cases on CT images. Jocelyn Zhu et al. [6] employed deep learning convolutional neural networks to determine the severity of lung disease in COVID-19 infection using portable chest X-rays (CXRs), with radiologist scores of disease severity serving as ground truth. Singh D. et al. studied Classification of COVID-19 patients from chest CT images[7]. Perumal V. et al.[8] propose Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features. The proposed model produces precision of 91%, recall of 90% and accuracy of 93% by VGG-16 using transfer learning.

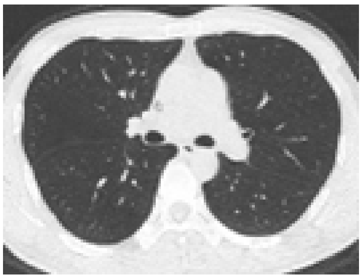

Figure 1: CT scan result of a normal lung [5]

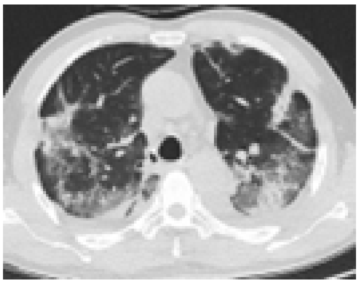

Figure 2: CT scan result of an abnormal lung [5]

3 Methodology

In this part, we are going to cover the algorithms used in the implementation of the experiments of section 4. The choice of the algorithm depends on the type of activity that needs to be automated [9], [10], [11] at hand, and the type of data. We can categorize any data science problem based on two key factors: how the algorithm is trained and the availability of output during training. Machine learning algorithms can be divided into different categories: supervised, unsupervised, semi-supervised learning and reinforcement learning. As we are interested in classification problems, we are going to introduce the appropriate machine learning algorithms that we have used.

3.1 Support Vector Machines algorithm

Support Vector Machines (SVM) are popular supervised learning models with associated learning algorithms that analyze data used for classification and regression problems analysis developed at AT&T Bell Laboratories by Vapnik with colleagues [12].

3.2 Random Forest algorithm

Random Forest is a popular tree-based machine learning algorithm that can be used for both Regression and Classification problems. It is a supervised learning technique and is based on the concept of ensemble learning, which is the process of combining multiple classifiers to solve a complex problem and to improve the performance of the model [13].

3.3 K-Nearest Neighbors algorithm

One more supervised learning technique that can be applied for both regression as well as classification problems is the k-nearest neighbors which is commonly referred as one of the simplest machine learning algorithms. K-Nearest Neighbors (K-NN) is a non-parametric learning algorithm, which means that it makes no assumptions about the underlying data and distribution[14]. It is also an instance-based algorithm that does not explicitly learn a model, but instead it memorizes the training instances which are consequently used as (knowledge) for the prediction phase [15]. Deciding the value of K is a very critical part of the K-NN algorithm because of prediction accuracy, and it should not be taken for granted [16], [17].

3.4 Convolutional Neural Network algorithm

Convolutional Neural Network (CNN, or ConvNet) is a type of deep learning model, inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. They are mostly used in image and video recognition, image classification, medical image analysis and natural language processing. From an architectural perspective, CNNs are regularized versions of multi-layer Perceptron's where the feature extractor is included in the training process[18]. This independence from prior knowledge and human effort in feature design is a major advantage. Convolution networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers. Training a CNN typically consists of two phases: A forward and a backward phase. During the forward phase and after initializing all the weights and kernels to some values, the input is then passed completely through the network. The received output of the network is compared to the desired output using a loss function, such as: Mean Squared Error (MSE), and Cross Entropy [19], [20].

3.5 Transfer learning

Transfer learning refers to the situation where a model developed for a task is reused as the starting point for a model on a second task in a similar domain. To train such a huge model, a lot of data is required. However, in many problems, the huge amount of data required to train the model is not available. Therefore, it is common to use transfer learning to generate generic features from a pre-trained deep-learning model and then use those features as an initialization for the task of interest.

3.5.1 ResNet-50V2

Residual Neural Networks (ResNets) introduced a novel architecture that fits stacked layers by utilizing skip-connections blocks, or shortcuts to jump over some layers to form the network. ResNet-50 is a type of this family of networks with a deep architecture using these blocks [21]. A pretrained version of the network trained on more than a million images from the ImageNet database can be loaded. The pretrained network can classify images into 1000 object categories. As a result, the network has learned rich feature representations for a wide range of images.

3.5.2 Inception-V3

Inception-v3 by Google is the third version in a series of Deep Learning Convolutional Architectures that provides a high-performance network with a relatively low computational cost. The key idea of Inception Nets is the use of inception module to design good local network topology (network within a network), these modules or blocks act as the multi-level feature extractor in which convolutions of different sizes are obtained to create a diversified feature map.

4 Experiments and results

In this part, we will present the implementation of Covid-19 identification from CT scan images using machine and deep learning models as an application of binary classification, where datasets were used to compare the results and the performance of these classification methods. We have implemented four classification methods which are: SVM, Random Forest, KNN and Convolutional Neural Networks. The implementation process passes through a training phase using the training dataset, after that comes the test phase to evaluate the obtained classifiers on a validation set that contains some test images. We have used the Python programming language in order to implement the algorithms of our classifiers. In addition, we have changed some parameters in order to see their effects on the performance of our models.

4.1 Evaluation metrics

The aim of the different machine learning algorithms varies depending on the specific use case. Thus, the evaluation measures for classification tasks are different than the ones for other tasks. There are different scoring metrics which are useful to summarize the outcomes and evaluate the performance of the classifiers, based on four standard indicators: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Based on the indicators defined above, all the scoring metrics used for the model's evaluation are to be introduced [22]:

- **The confusion matrix:** is a square matrix that reports the counts of the four indicators predicted by the classifiers, as shown in the following table. The columns are indicating the predicted number of samples and the rows are showing what the actual or true class of the instances is.
- **Accuracy:** is the most usable evaluation metric. It provides general information about how many samples are misclassified and is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the sum of correct predictions by the total number of predictions.

$$Acc = (TP + TN)/(TP + FP + TN + FN) \quad (1)$$

- **Specificity:** is defined as ratio of the total number of true negative predictions to the total number of actual negatives.

$$Specificity = TN/(TN + FP) \quad (2)$$

- **Sensitivity or Recall (REC):** is defined as the ratio of true positive predictions to the total number actual positives.

$$REC(Sensitivity) = TP/(TP + FN) \quad (3)$$

- **Precision or (PRE):** is defined as the fraction of relevant examples (true positives) among all of the examples which were predicted to belong to a certain class.

$$PRE = TP/(TP + FP) \quad (4)$$

- **F1-score:** is similar to accuracy which seeks to create a balance between precision and recall scores. Therefore, this score takes both false positives and false negatives into account. Its relationship is represented by the following equation:

$$F1 = 2 * (PRE * REC)/(PRE + REC) \quad (5)$$

4.2 Dataset

The main bottleneck for the realization of this study is the lack of good quality comprehensive data sets, however, after an intense search, we decided to use to the following public datasets:

4.2.1 Small COVID-19 lung CT-Scans dataset

Possibly the first attempt to create Covid-19 CT scan data set was the COVID-19 Lung CT dataset [23], which consists of images mined from scientific articles, hospital donations and websites. Allowing to build a publicly available

COVID-CT dataset, containing 349 CT scans that are positive for COVID-19 from 216 patients and 397 CT scans regarding healthy and non-COVID patients. It is important to highlight that some images contain textual information which may interfere with model prediction.

4.2.2 Large COVID-19 lung CT-Scans dataset

This dataset is a public dataset [24] consisting of 14486 CT scans from 1070 patients, with 7593 CT scans of 466 patients that are positive for SARS-CoV-2 infection (COVID-19) and 6893 CT scan images of 604 non-infected patients. Data was collected from different public datasets. These datasets have been publicly used in COVID-19 diagnosis literature and have proven their effectiveness in deep learning applications. Therefore, the merged dataset is expected to improve the generalization ability of deep learning methods by learning from all these resources together.

4.3 Training

The data-set is split into multiple sets that are used in different stages of the training, validation and evaluation of the model. At first, we implemented machine learning models (SVM, KNN, RF) using Scikit-learn library and initially fit each on the training set. Since there are few parameters for tuning and we already decided on the models beforehand, the validation set is not needed. Nevertheless, it is possible to employ various methodologies to optimize the parameters; in our particular case, we manually adjusted parameters such as the kernel type in SVM, the number of trees in RF, and the distance metric in KNN. After that, the test set is used for performance evaluation. In the second step, we have implemented deep learning models (CNN, ResNet, Inception) with Keras library using the TensorFlow backend and compiled our model and have trained it on the training dataset. The validation set is then used for tuning the parameters of the model by examining many examples and attempting to find a model that minimizes the loss. Finally, the test dataset is used to provide a performance evaluation of a final model fit on the training dataset. Note that the test dataset has never been used in training neither validation, and it is also called a holdout dataset. The term ”validation set” is sometimes used instead of ”test set” in some literature if the original dataset was partitioned into only two subsets [25] which was in our machine learning case. However, in the correct usage, the validation set is a development set and the test set is the independent set used to evaluate the performance. For data augmentation we have applied different transformations to the training sets such as rotation, horizontal flip, and scaling. For the hyper-parameters, we have used learning rate, batch size, epoch and number of layers. These parameters had great influence on the results of the classifier.

4.4 Experiments

4.4.1 COVID-19 classification using SVM

After applying some pre-processing, we build our classifier then we fit it in the training data. For our case, the data were randomly shuffled and divided into two subsets, the first one consisting of 80% of the initial data which is reserved for training the model and the second one of 20% to test how well is the model. After that, we test our model on the testing data then we compute and print the accuracy. Finally, we print the confusion matrix of the correct and incorrect predicted cases depending on the categories of the data. Also, we print the sensitivity and the specificity: In case of changing the kernel type, the SVM scores will be affected. We have used the three main kernels introduced in the previous part and the obtained results are presented in Tables 1,2 . From the values shown on Table 3 for the small dataset, we can notice that the Poly kernel gives a slightly better accuracy than the RBF one while the latter better predicts the non-Covid cases since its specificity is higher on this small dataset. On the other hand, the linear kernel yielded lower accuracy and F1-score, a known characteristic due to its simplicity, which results in lower accuracy. We can notice from the table 2 that for the large dataset, the different kernels of the SVM algorithm affect the sensitivity, specificity and accuracy, where the Poly SVM outperformed both the Linear and the RBF SVMs in both datasets, making it the most suitable for this kind of classification.

Table 1: SVM evaluation metrics using different kernels for small dataset.

Kernels	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
Poly	80.0	77.14	82.86	79
RBF	78.0	71.23	84.42	76
Linear	68.0	62.34	73.97	67

Table 2: SVM evaluation metrics using different kernels for large dataset.

Kernels	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
Poly	90.29	88.78	91.91	90
RBF	88.59	87.61	89.78	89
Linear	87.62	88.05	87.09	88

4.4.2 COVID-19 classification using KNN

Our second model to be used is the KNN, similarly we apply some preprocessing on the data set then we fit our KNN classifier on the preprocessed data. Notice that we used

Manhattan since it is already known as the best distance parameter to use in KNN classification. We apply the same process of random shuffling and splitting the data into two subsets, mirroring the methodology used in the SVM experiment. To show the effect of the metrics, we change the metric to Euclidean distance and repeat the same process of defining the model and fitting it to the training set, followed by a final evaluation on the test set. Both metrics results were compared. From Table 3, for the small dataset, we can confirm that Manhattan distance gives better scores than the other distance types. Overall, even though the accuracy of our classifier is 78%, KNN is not good at classifying positive cases. As for the large dataset, the data presented in Table 4 highlights the superior performance of the Manhattan distance over the Euclidean distance. However, accuracy alone does not provide a complete picture of the model's predictive capabilities. When examining sensitivity, it becomes apparent that it is relatively lower when compared to specificity.

Table 3: KNN evaluation metrics using different distance types for small dataset.

Distance type	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
Manhattan	78.0	62.5	89.53	71
Euclidean	74.67	48.43	94.19	62

Table 4: KNN evaluation metrics using different distance types for large dataset.

Distance type	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
Manhattan	84.29	81.43	87.43	84
Euclidean	83.79	79.52	88.48	83

4.4.3 COVID-19 classification using Random Forest

The third model is the random forest (RF). We apply the same processing as we did for the previous experiments. The results from evaluating the model on the small dataset using two different numbers of trees are summarized in Table 5. We can notice that 120 trees would give a better performance of the model than 60 trees. In addition, we have tried to increase the number of estimators to 1000 and more but there was no big difference in the accuracy. The model was implemented on the large dataset and the obtained scores are presented in Table 6. We initially experimented with 120 estimators and subsequently raised the count to 240, resulting in a modest improvement in accuracy. Nevertheless, further increasing the number of estimators did not yield substantial differences in accuracy.

Table 5: RF evaluation metrics using different number of trees for small dataset.

Number of trees	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
120	78.57	79.41	77.78	79
60	78.0	76.56	79.07	75

Table 6: RF evaluation metrics using different number of trees for large dataset.

Number of trees	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
120	89.52	89.05	90	90
240	89.77	89.06	90.5	90

4.4.4 COVID-19 classification using CNN

Moving to the deep learning models, we have started by developing a Convolutional Neural Network with four convolution layers. After doing some preprocessing and augmentation to the data, we split it into three subsets: a training set with 80% of the original set, a validation and testing sets each of 10%. We have chosen these values since this dataset is small and the model needs enough training samples. After testing the model on the testing set, we have obtained the confusion matrix below which allows us to compute the evaluation metrics presented on table 7. As we found low evaluation scores, we have tried to improve the model by changing the architecture and the parameters a few times. However, the lack of information in this dataset prevented any improvement in the accuracy. In order to improve the model scores, we have developed a better architecture and applied it on the large dataset using two consecutive convolution layers followed by batch normalization, a single max pooling layer and a dropout to prevent over-fitting. This process has been repeated twice to finally have six convolution layers and only three pooling layers. We can notice an improvement compared to the previous CNN model. The evaluation metrics for both of the CNN models are presented in Table 8. It is worth noting that the new model exhibits superior performance, achieving accuracy rates above 90% and a well-balanced sensitivity and specificity.

Table 7: CNN evaluation metrics using 4 convolution layers for small dataset.

Number of conv layers	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
4	73.33	68.57	80.0	72

Table 8: CNN evaluation metrics using different architectures for large dataset.

Model	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
First CNN	92.27	95.52	88.7	93
New CNN	93.65	91.30	96.23	94

4.4.5 COVID-19 classification using InceptionV3

Moving to the first transfer learning models, we split the dataset into similar subsets as the CNN experiment. Moreover, fully connected layers of flatten, dense and activation are subsequent to the InceptionV3 model. After finishing the tasks noted above, we train the model in order to get the accuracy and the loss graphs. As we have the confusion matrix, we can summarize the evaluation scores in the following Table 9,10. From the values shown on Table 9, we can notice that the scores for the small dataset differ dramatically due the fact that the testing set is very small. On the other hand, we obtained the same accuracy as the Poly SVM model. As for the large dataset, we notice a better performance on the training and validation sets for this model compared with previous models. From the values shown on Table 10, we can notice that all scores are high enough but there is a gap between the sensitivity and the specificity values.

Table 9: InceptionV3 model evaluation metrics for small dataset.

Model	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
Inception V3	80.0	90.0	73.33	78

Table 10: InceptionV3 model evaluation metrics for large dataset.

Model	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
Inception V3	97.17	95.15	99.35	97

4.4.6 COVID-19 classification using Resnet50V2

Our second transfer learning model is ResNet50V2. Similarly, we apply preprocessing and augmentation on the dataset and we follow the succeeding procedures. The performance of this model on the small dataset was better than the previous ones and increasing the number of epochs yielded good results. From the confusion matrix, we can summarize the evaluation scores in Table 11. We noticed that ResNet50V2 model achieved slightly superior performance than all the machine and deep learning models. The performance of this model on the large dataset is as good as the InceptionV3 one. Examining Table 12, we observe that the ResNet50V3 model exhibits the highest performance, as it achieves superior accuracy and effectively balances sensitivity and specificity.

Table 11: ResNet50V2 model evaluation metrics for small dataset.

Model	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
ResNet50V2	86.67	93.94	81.0	86

Table 12: ResNet50V2 model evaluation metrics for large dataset.

Model	Accu. (%)	Sens. (%)	Spec. (%)	F1-score (%)
ResNet50 V2	97.52	97.28	97.77	98

4.5 Discussions

Machine and deep learning algorithms were implemented with the purpose of classifying the images of two datasets into Covid and non-Covid classes. For the small dataset, we have started with machine learning algorithms since they are easy to implement and simple to tune. We found that Poly SVM has the highest accuracy of 80%. However RF shows a better balance between accuracy (78.57%), sensitivity (79.41%) and specificity (77.78%) among all the classifiers. In contrast, KNN model was the worst one with an accuracy of 78% and a substantial disparity between sensitivity (62.5%) and specificity (89.53%). This imbalance in sensitivity can be interpreted as the model's cautious approach in identifying positive results, which ultimately impacts its effectiveness in virus detection. CNN also exhibited subpar performance, achieving an accuracy of 73.33%. It's important to note that these results do not suggest that machine learning algorithms are superior in binary classification tasks; rather, the performance was limited by the dataset's insufficient size for training the CNN model effectively. This problem led us to use transfer learning models such as InceptionV3 which achieved an accuracy identical to that of Poly SVM at 80%. However, it exhibited a substantial disparity between sensitivity (90%) and specificity (73.33%). On the other hand, ResNet50V2 achieved better results, with an accuracy of 86.67%, a sensitivity of 93.94%, and a specificity of 81%. Based on all the evaluation metrics mentioned here, we have shown that the ResNet50V2 is performing better than the other models on this dataset. Similarly in the case of the larger dataset, our machine learning experiments produced an improved accuracy of 90.29% with the SVM model. In addition, the RF was the most balanced model giving an 89.77% accuracy, 89.06% sensitivity and 90.5% specificity when using 240 trees. As for the deep learning models, the CNN model yielded an accuracy of 92.27%, while the second approach, involving the use of two consecutive convolutional layers before each pooling layer, achieved a slightly higher accuracy of 93.65% compared to the first approach. Moreover, the transfer learning models InceptionV3 and ResNet50V2 resulted in excellent performance with an accuracy of 97.17% and 97.52% respectively. Based on this marginal difference, checking the balance of these two models will exhibit which one is better. As we can notice in the tables 10 and 12, the ResNet50V2 has better scores with a sensitivity of 97.28% and a specificity of 97.77% while the respective ones of the InceptionV3 are 95.12% and 99.35%. These results lead to conclude that the ResNet50V2 is the best model to be used on this large dataset too. This is a consequence of its deep architecture that contains over 23 million trainable parameters. Coming to the comparison of the datasets, we have found that machine learning models are performing better than deep learning ones on the COVID-19 Lung CT Scans dataset as the lack of enough information caused by the small number of samples prevent the deep learning models from having better scores. In contrast, the deep learning models showed better performance with the Large COVID-19 CT scan slice dataset. Furthermore, transfer learning algorithms also worked better on the large dataset and the results were convincing compared to the ones obtained from the small dataset. The obtained results were compared to the literature review using the same dataset [8], [26]. We can observe that the ResNet50, InceptionV3 and CNN based models used in our work were trained on the large dataset which was used from the challenge of WHO made in 2020 lunching world competition. Our dataset has been processed and optimized by the world health organization engineers and this make it more reliable to be used as a baseline in the comparison process. The obtained results using the CT scans imaging provide higher results of accuracy of 99.91%, and Sensitivity, Specificity and F1 scores respectively equals to 99.82% 99.97% and 99.90% comparatively by the results obtained by Baghdadi et al.[26]. The authors proposed WOANet and achieved Accuracy, Sensitivity, Specificity, Precision, and F1 score of 98.78%, 98.37%, 99.19%, 99.18% as presented in the table 13. In the both works, the results are closed, however, Baghdadi et al.[26] add another preprocessing step which may exclude some significant features from the data and extend the processing pipeline duration.

Table 13: Results Comparison with related work on large COVID-19 CT-scan slice dataset.

Author	Model	Accu. (%)	Sens. (%)	Spec. (%)	F1-Score
Perumal V. et al.[8].	VGG-16	93	90	91	-
Baghdadi et al.[26].	WOANet	98.78	98.37	99.74	99.74
Our Work	ResNet50 V2	99.91	99.82	99.97	99.90

5 Conclusion

An experimental evaluation of various Machine and Deep Learning based image classification approaches are presented in this work, in order to identify COVID-19 positive cases from chest CT scan images. Additionally, the proposed models present comparable results on two different datasets in order to show the effect of the dataset quality and size on the predictive performance of the models. Nevertheless, different values for hyper-parameters were used to obtain the optimum results of each algorithm. We established that the Machine Learning algorithms (SVM, RF and KNN) may have a high performance on small data in the presence of enough training data. Nevertheless, when it comes to accuracy, CNN outperforms these algorithms. However, it's important to note that accuracy alone may not provide a comprehensive measure of the model's overall robustness. While a model may achieve higher accuracy, it may struggle to effectively capture the nuances in the data, resulting in weaker performance when faced with data variations, as seen in the case of KNN's predictive ability for positive cases. This underscores the significance of considering accuracy, sensitivity, and specificity values collectively when assessing the model's suitability for disease screening. The proposed Transfer Learning approaches (ResNet50V2 and InceptionV3) improved on all the performance measures as they achieved very impressive results. With all the records that have been seen, the ResNet50V2 model was found to be the best among all the models on both small and large datasets. Furthermore, it was able to outperform the highest reported accuracy of 86% and F1-score of 85% on the small dataset as it achieved 86.67% accuracy, 93.94% sensitivity and 81% specificity and 86% F1-score. For the large dataset it achieved 97.52% accuracy, sensitivity of 97.28% and a specificity of 97.77%. Therefore, we improved the scores of the COVID-19 Lung CT Scans dataset compared to the highest reported accuracy of 86% and F1-score of 85% [41]. Many experiments on different machine and deep learning models were implemented in order to classify the positive and negative COVID-19 cases using two datasets. Thus, the results were varying according to different parameters and characteristics of each model. We found that SVM and RF models show good accuracy and balance of the scores respectively for the small dataset, while the transfer learning models obtained better performance when working on large dataset. Furthermore, we have demonstrated the good effect of using two consecutive convolutional layers before every pooling layer on the evaluation scores. Our analysis concluded that the developed methods improve significantly the COVID-19 detection in CT images and suggests to be considered as a clinical option.

References

[1] World Health Organization, ”Origin of SARS-CoV-2,” 26 March 2020. [Online]. Available: https://apps.who.int/iris/handle/10665/332197.

[2] World Health Organization, ”WHO Director-General’s remarks at the media briefing on 2019-nCoV on 11 February 2020”. [Online]. Available: https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020. https://doi.org/10.1007/s11263-007-0056-x

[3] Worldometer, COVID-19 CORONAVIRUS PANDEMIC,[Online]. Available: https://www.worldometers.info/coronavirus/.

[4] Martínez Chamorro, E., Díez Tascón, A., Ibáñez Sanz, L., Ossaba Vélez, S., and Borruel Nacenta, S. Radiologic diagnosis of patients with COVID-19. Radiología (English Edition), 63(1): 56–73, 2021. https://doi.org/10.1016/j.rxeng.2020.11.001.

[5] Kwee, T. C. and Kwee, R. M. (2020). Chest CT in COVID-19: What the Radiologist Needs to Know. RadioGraphics, 40(7):1848–1865. https://doi.org/10.1148/rg.2020200159

[6] Zhu, J., Shen, B., Abbasi, A., Hoshmand-Kochi, M., Li, H., and Duong, T. Q. Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. PloS one, 15(7): e0236621, 2020. https://doi.org/10.1371/journal.pone.0236621

[7] Singh, D., Kumar, V., Vaishali, and Kaur, M. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks. European Journal of Clinical Microbiology and Infectious Diseases, 39(7): 1379–1389, 2020. https://doi.org/10.1007/s10096-020-03901-z

[8] Perumal, V., Narayanan, V., and Rajasekar, S. J. S. Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features. Applied Intelligence, 51(1), 341–358, 2020. https://doi.org/10.1007/s10489-020-01831-z

[9] Hayashi, C. What is Data Science? Fundamental Concepts and a Heuristic Example. Data Science, Classification, and Related Methods, 40–51, 1998. https://doi.org/10.1007/978-4-431-65950-1_3

[10] Cherifi, D., Falkoun, N., Ouakouak, F., Boubchir, L., and Nait-Ali, A. EEG Signal Feature Extraction and Classification for Epilepsy Detection. Informatica, 46(4), 2022. https://doi.org/10.31449/inf.v46i4.3768

[11] Cherifi D., Cherifi Z., and Belkadi W. Brain Tumor Classification Using Convolutional Neural Networks and Transfer Learning. In International Conference on Artificial Intelligence, Renewable Energetic Systems, pp. 37-48. Cham: Springer International Publishing, 2022.

[12] Chaudhari, R. S., Ghorpade, S. J., and Patil, S. S. SVM Classifier: Identify Linear Separability Of NAND And NOR Logic Gates. ECS Transactions, 107(1):18281–18292,2022. https://doi.org/10.1149/10701.18281ecst

[13] Ho, T. K. (2002). A Data Complexity Analysis of Comparative Advantages of Decision Forest Constructors. Pattern Analysis and Applications, 5(2):102–112. https://doi.org/10.1007/s100440200009

[14] Atawodi, Ilemona S., A Machine Learning Approach to Network Intrusion Detection System Using K Nearest Neighbor and Random Forest, 2019. Master’s Theses. 651. https://aquila.usm.edu/masters_theses/651

[15] Kramer, O. K-Nearest Neighbors. Intelligent Systems Reference Library, 13–23, 2013. https://doi.org/10.1007/978-3-642-38652-7_2.

[16] K. Zakka, A Complete Guide to K-Nearest-Neighbors with Applications in Python and R, 2016. [Online]. Available: https://kevinzakka.github.io/2016/07/13/k-nearest-neighbor/.

[17] Keogh, E. Nearest Neighbor. Encyclopedia of Machine Learning and Data Mining, 897–897, 2017. https://doi.org/10.1007/978-1-4899-7687-1_579

[18] Guernine, A., and Kimour, M. T. Optimized Training for Convolutional Neural Network Using Enhanced Grey Wolf Optimization Algorithm. Informatica, 45(5), 2021. https://doi.org/10.31449/inf.v45i5.3497

[19] Rojas, R. Modular Neural Networks. Neural Networks, 411–425, 1996. https://doi.org/10.1007/978-3-642-61068-4_16

[20] Arabshahi, P. (1996). Fundamentals of Artificial Neural Networks [Book Reviews]. IEEE Transactions on Neural Networks, 7(3), 793. https://doi.org/10.1109/tnn.1996.501738

[21] He, K., Zhang, X., Ren, S., and Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp.770-778, 2016. https://doi.org/10.1109/cvpr.2016.90

[22] Jayaswal, Vaibhav. Performance Metrics: Confusion Matrix, Precision, Recall, and F1 Score. Medium, Towards Data Science 15, 2020.

[23] Blanche, Luis, and A. Lorenzo. COVID-19 lung CT scans. Kaggle, 2020. [Online]. Available: https://www.kaggle.com/luisblanche/covidct.

[24] Maftouni, M. Large COVID-19 CT scan slice dataset. Kaggle,2021. [Online]. Available: https://www.kaggle.com/maedemaftouni/large-covid19-ct-slice-dataset.

[25] J.Brownlee, Jason. What is the difference between test and validation datasets. Machine Learning Mastery, 2017.

[26] Baghdadi, N. A., Malki, A., Abdelaliem, S. F., Magdy Balaha, H., Badawy, M., and Elhosseini, M. An automated diagnosis and classification of COVID-19 from chest CT images using a transfer learning-based convolutional neural network. Computers in Biology and Medicine, 144, 2022. https://doi.org/10.1016/j.compbiomed.2022.105383