



# A Novel Weighted Consensus Machine Learning Model for COVID-19 Infection Classification Using CT Scan Images

Rohit Kumar Bondugula<sup>1</sup> · Siba K. Udgata<sup>1</sup> · Nitin Sai Bommi<sup>1</sup>

Received: 12 April 2021 / Accepted: 15 June 2021  
© King Fahd University of Petroleum & Minerals 2021

## Abstract

As COVID-19 has spread rapidly, detection of the COVID-19 infection from radiology and radiography images is probably one of the quickest ways to diagnose the patients. Many researchers found the necessity to utilize chest X-ray and chest computed tomography imaging to diagnose COVID-19 infection. In this paper, our objective is to minimize the false negatives and false positives in the detection process. Reduction in the number of false negatives minimizes community spread of the COVID-19 pandemic. Reducing false positives help people avoid mental trauma and wasteful expenses. This paper proposes a novel weighted consensus model to minimize the number of false negatives and false positives without compromising accuracy. In the proposed novel weighted consensus model, the accuracy of individual classification models is normalized. While predicting, different models predict different classes, and the sum of the normalized accuracy for a particular class is then considered based on a predefined threshold value. We used traditional Machine Learning classification algorithms like Linear Regression, Support Vector Machine, *k*-Nearest Neighbours, Decision Tree, and Random Forest for the weighted consensus experimental evaluation. We predicted the classes, which provided better insights into the condition. The proposed model can perform as well as the existing state-of-the-art technique in terms of accuracy (99.64%) and reduce false negatives and false positives.

**Keywords** Machine learning · Weighted consensus model · Chest CT scan · COVID-19

## 1 Introduction

The World Health Organization (WHO) has declared the novel coronavirus (COVID-19) disease to be a pandemic and has raised public health concerns around the world. COVID-19 has been linked to 123.87 million confirmed cases and 2.72 million deaths as of the 17th of March, 2021. [1]

COVID-19 is wide-spread and highly contagious which is transmitted directly from the infected people through direct contact and spreads indirectly through the air, surface, and the surroundings in which the infected persons come in con-

tact [2]. The disease causes viral pneumonia in the lungs, resulting in acute respiratory problems and creating a lesion on the lungs. It also causes a variety of symptoms like fever, dry cough, headache, tiredness, loss of taste and smell and dyspnea [2–4]. Moreover, the COVID-19 spread is more worsened by the certainty that most of the infected people are having asymptomatic symptoms [3]. Therefore, quickly diagnosing the infected person's symptoms and quarantining them is crucial to curb the spread of the disease.

The pandemic situation is affecting billions of people on a social, economic, and medical basis, creating dramatic changes in social relationships and educational environments and affecting many people's lives. We cannot blame the doctors since they are responsible for many people and have few resources. However, we can assist or ease the burden on them by developing a model that predicts whether a person is potentially positive or negative [5–8].

The healthcare industry is looking for advanced technologies that can monitor, detect, and diagnose infection and quickly control the COVID-19 pandemic spread. Internet of

✉ Siba K. Udgata  
udgata@uohyd.ac.in

Rohit Kumar Bondugula  
rohitkumarbondugula@gmail.com

Nitin Sai Bommi  
nitinsai.bommi@gmail.com

<sup>1</sup> School of Computer and Information Sciences, University of Hyderabad, Hyderabad, India



Medical Things (IoMT) is one such sophisticated technology that can monitor people by crowd screening, tracking, notifying, and detecting the virus and controlling the spread through contact tracing and alerting the healthcare authorities [5].

In today's medical practice, there are two primary types of diagnosis. The nasopharyngeal swab is used in real-time RT-PCR. The second category is imaging techniques, with CT scans outperforming chest X-rays. According to studies, chest CT is faster and more sensitive than the PCR process [9].

COVID-19 is often diagnosed with RT-PCR and serological testing [10]. However, these tests are difficult to conduct due to a lack of resources and qualified staff, particularly in late-stricken areas (e.g., Africa and Latin America). Furthermore, the sensitivity of PCR can be low [9, 11, 12]. Therefore, alternative methods to quickly diagnose the COVID-19 infection are crucially needed.

Detecting the disease at an early stage and instantaneously quarantining the person is vital to stop the disease's outspread because of the unavailability of the vaccine. The Chinese government announced that the diagnosis of the infection can be verified through RT-PCR [9]. However, RT-PCR takes more time for test and suffers from high false-negative [9, 13–16].

In this present pandemic situation, the low sensitivity of the RT-PCR cannot always be accepted. In a few cases, the infected people cannot get treatment on time, as it may not detect correctly. The infected people then may spread the infection to healthy people. It is noticed from the clinical reports of people who are infected that there are bilateral changes in Chest X-Ray and Chest CT scan images [13]. Hence, chest CT scan and X-Ray images are utilized as a substitute device to detect COVID-19 infection due to high sensitivity [3].

This paper's main objective is to perform the classification of the COVID-19 patients using the Chest CT scan images such that the false negatives are minimum or false positives are minimum depending on the requirement. We used the machine learning classification models to detect the COVID-19 infection using the CT scan in the proposed work. We propose the novel weighted consensus model where the image passes through the models governed by predefined rules during the current situation to reduce or minimize community spread and save people from false negatives.

The remainder of the paper is laid out as follows: Sect. 2 discusses the literature review in the area of COVID-19 classification. In Sect. 3, the proposed methodology of the classification model is discussed. In Sect. 4, a detailed explanation about the experimental setup is discussed, which is followed by the results and discussions in Sect. 5. Finally, in Sect. 6, we discuss the conclusions and future scope of the work.

## 2 Literature review

Recently, many researchers have appraised the medical imaging patterns and lesions on chest CT scan and chest X-ray for detecting the COVID-19 [17–26].

The Artificial Intelligence and radiology imaging of COVID-19 can be handy for accurate and timely diagnosis of disease [27]. Fang et al. [24] have studied the sensitivity of the chest CT scan and RT-PCR. Xie et al. [25] reported that the COVID-19 diagnosis was true negative for over 3% of the cases in the sample of 167 patients using RT-PCR. The sensitivity of the chest CT scan for COVID-19 infection detection is high compared to the RT-PCR based on the symptoms and travel history analysis of the patients [25]. From the clinical reports of people who are infected, it is observed that there are bilateral changes in CT scan images [13]. Therefore, a chest CT scan is used to diagnose the disease due to high sensitivity [3].

Yu-Dong Zhang et al. [28] proposed the DesneNet-OTLS method, which outperformed most of the state-of-the-art approaches in COVID-19 diagnosis. COVID-Net model [29] was developed to detect COVID-19 positive cases from chest radiography images which can achieve 80% sensitivity. Kermany et al. [30] used the ConvNet model for Chest X-ray and got training accuracy of 95.21% and validation accuracy of 95.31%. Xu et al. [13] employed a CNN model which differentiates COVID-19 pneumonia and viral pneumonia with maximum accuracy of 86.7%.

Wang et al. [14] used the CT images of infected patients and analyzed the radiographic changes. They developed a model that used the amended inception transfer learning technique with an accuracy of 89.5%. The extracted features from CT images are used for prior diagnosis. This method can diagnose faster and also performs better compared to Xu's model [13]. Qianqian Ni et al. [31] used a deep learning approach to identify COVID-19 pneumonia in chest CT images.

Rajaraman et al. [32] used a weakly-labeled data augmentation strategy on COVID-19 chest X-ray images. Novitarci DCR et al. [33] combined SVM and CNN to detect COVID-19 from chest X-ray images.

Khalifa et al. [34] used generative adversarial networks and transfer learning models like ResNet18, AlexNet and SqueezeNet to classify COVID-19.

Ozturk et al. [35] employed DarkNet on Chest X-ray images for the binary classification and multi-class classification with accuracy of 98.08% and 87% respectively.

Narin et al. [16] proposed DCNN-based transfer models for diagnosis of COVID-19 using the chest X-ray images. They have employed Inception-ResNetV2, InceptionV3, and ResNet50 models for good prediction. The latter model gave an accuracy of 98%, that is the so far better result for chest x-ray [13, 14].

Yu-Dong Zhang et al. [36] proposed a novel deep learning model that can diagnose COVID-19 on chest CT more accurately with a sensitivity of 93.28%, a specificity of 94.00%, and with an accuracy of 93.64%. Zhang [37] also proposed a novel seven-layered CNN-based innovative diagnosis model which is effective in detecting the COVID-19 in chest CT images and achieves a sensitivity of 94.44%, a specificity of 93.63%, and an accuracy of 94.03%.

Maier et al. [38] performed an analysis on chest X-ray images combining six different databases from open datasets to determine images of infected patients while distinguishing COVID-19 and pneumonia from 'no-findings' images. Saba et al. [39] proposed six models for the tissue characterization and classification of COVID-19 with pneumonia and achieved better results.

Qian Lie et al. [40] integrated an image preprocessing technique for anomaly detection with supervised deep learning models for chest CT scan based COVID-19 diagnosis. Menendez et al. [41] developed a web application COVID-19 TRAINING, for training and diagnosis of COVID-19 chest x-ray.

Guangyu Guo et al. [42] proposed IE-Net to eliminate the influence of the varied dimensions and diagnose the COVID-19 cases. IE-Net achieves 92.79% recall, 94.80% accuracy, 92.97% precision and 94.93% AUC for diagnosing COVID-19 cases from non-COVID-19. Umit Budak et al. [43] SegNet-based network model, which used the attention gate mechanism for the automatic segmentation of COVID-19 lesions in CT images which achieved sensitivity 92.73%, specificity 99.51%, and dice scores 89.61% respectively.

In the VSBN model, Wang et al. [44] proposed a novel VGG-style base network as the backbone network and a convolutional block attention module as the attention module. The model's sensitivity, accuracy, and F1 per class were all above 95%.

From the comprehensive review, it has been noticed that for early diagnosis of COVID-19 patients, chest X-ray and CT images can be used [45]. Therefore, in this paper, machine learning models are used to classify COVID-19 patients from CT images.

## 2.1 Research gaps in the existing literature

Although many researchers have contributed significantly to this research domain, we still found some gaps in the work. While discussing with the medical practitioners and healthcare front line workers, the following shortcomings in the literature are highlighted, and those are the following;

- Most of the work is focusing on maximizing the accuracy of their proposed method. Accuracy, though, is a crucial

performance evaluation parameter but can not be the only parameter.

- Few works also focused on the model's training time and testing time and tried to reduce the classification time without compromising accuracy.
- The existing works are not tuned to address the changing pattern of the COVID-19 spread.
- No existing work focuses on minimizing the false negatives or false positives without compromising on the accuracy of the model
- The medical practitioners do not appreciate the existing models as the practitioners are least bothered about the statistical accuracy but more concerned about false negatives or false positives depending on the situation.

## 2.2 Contributions of the present work

After a thorough review of the existing works and identifying the gaps in these works, we designed a model to address the gaps. We developed a more acceptable and realistic model. The main contributions are the following:

- 1 We introduced a novel weighted consensus model intending to lower the number of false negatives and false positives while maintaining accuracy.
- 2 The proposed model uses the best performing architecture together with a consensus algorithm to enhance the accuracy.
- 3 The proposed WCM model will also work for limited data samples as data augmentation technique can be used.
- 4 The proposed model is supposed to be accepted by the medical practitioners as it is designed according to their requirements.
- 5 The proposed model can also minimize false negatives or false positives without compromising each other much. This is possible by adaptive fine-tuning of the threshold values of the individual models used.

## 3 Proposed method

We used traditional Machine Learning classification algorithms to train the images. Five popular algorithms were used for classification which are described as follows.

### 3.1 Logistic regression

Logistic regression uses a logistic function that produces an output in the range [0, 1]. This algorithm is widely used to differentiate two classes linearly. It is an extension of linear regression with bounded output. The probability is the estimated output of the hypothesis.



### 3.2 Support vector machine

SVM's goal is to find a hyperplane in n-dimensional space that divides different categories. There are numerous ways to create a hyperplane that separates different groups. On the other hand, SVM attempted to optimize the distance between the hyperplane and the data points.

### 3.3 K-nearest neighbour

A non-parametric algorithm stores the input data and finds the difference between the input data and the data to be tested. The model then assigns a class based on the mode of the k nearest samples. When the input data is so huge, it becomes computationally expensive as it has to find the difference between every input and the test data.

### 3.4 Decision tree

This algorithm uses a tree-like structure to make decisions based on the input. It only contains conditional control statements. The model is prone to over-fitting as it tries to make conditions for every type of input.

### 3.5 Random forest

As the decision tree is prone to over-fitting, we try to generalise by constructing multiple decision trees and then considering the mode of them as the output class.

### 3.6 Proposed novel weighted consensus method

To ensure reliability and robustness of the prediction, we used five models as the base. Similar to the analogy where we consult another doctor for a second or a third opinion and then use the weightage of the suggestions given by different doctors, we also use five models performing at human-level accuracy (consulting five doctors) and then use the weightage of each prediction to finally declare the outputs.

The image is first passed through all five models and the predictions of each model are saved. Now all the models' accuracies are summed and the weightage of each model is found by dividing the model's accuracy by the total accuracy (normalizing weights). This ensures that the weighted accuracies sum to 1. This gives the weightage of the model among the five models. If a model has high accuracy, the model also carries much weightage.

Once the weightage and the individual models' predictions are found, to predict if an image belongs to a class or not, we sum all the normalized weights predicting that class and consider the class where the weightage is maximum. Since we calculate the class with the maximum threshold,

we do not concentrate on a single class. The pseudo-code for the above explanation is given by Algorithm 1.

To have better control over the number of FPs and FNs, we set a threshold value and then decide if the image belongs to that class or not. As the threshold goes higher, for the image to be predicted as the main class, more individual models have to predict it as the main class. This ensures that even if a model mispredicts an image, there are other models whose weights are considered in classifying the image. Here we are mainly focusing on a single image by setting a threshold value. If it is below the value, we can declare that the image does not belong to the wanted class. Then we use the maximum threshold algorithm on the other two classes for the final output. The pseudo-code is given by Algorithm 2.

In algorithm 1, if two classes get the same model weightage, we can take a call of class precedence. Since FN of covid positive is dangerous than FP, priority is given to covid positive (class 2).

The flow of data and the model are presented in Fig. 1.

---

#### Algorithm 1 Maximum weightage algorithm

---

```

1:  $p_{pct}, p_{nct}, p_{nict} \leftarrow 0$ 
2: for model in models do
3:    $p = model.predict(image)$ 
4:   if  $p == 0$  then
5:      $p_{nct} += model\_weightage$ 
6:   else if  $p == 1$  then
7:      $p_{nict} += model\_weightage$ 
8:   else
9:      $p_{pct} += model\_weightage$ 
10:  end if
11: end for
12:  $m = \max(p_{nct}, p_{nict}, p_{pct})$ 
13: if  $m == p_{pct}$  then
14:    $final\_class = 2$ 
15: else if  $m == p_{nict}$  then
16:    $final\_class = 1$ 
17: else
18:    $final\_class = 0$ 
19: end if

```

---

## 4 Experiment design

### 4.1 Dataset description

The HUST-19 benchmark CT Scan dataset [46] was used in our experiment. They divided CT images into three categories:

- (i) non-informative CT (NiCT) images, in which the lung parenchyma was not captured for any decision,

**Algorithm 2** Threshold weightage algorithm

```

1: desired_class = 2
2: desired_threshold = 0.25
3: class_to_numerical = (nct → 0, nict → 1, pct → 2)
4: ppct, pnct, pnict ← 0
5: for model in models do
6:   p = model.predict(image)
7:   if p == 0 then
8:     pnct += model_weightage
9:   else if p == 1 then
10:    pnict += model_weightage
11:   else
12:    ppct += model_weightage
13:   end if
14: end for
15: if class_to_numerical > desired_threshold then
16:   final_class = desired_class
17: else
18:   final_class = class(maximumthreshold)
19: end if
    
```

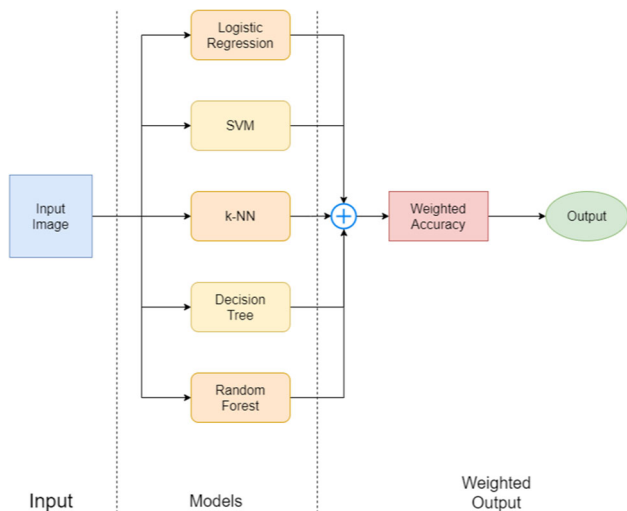


Fig. 1 Block diagram of the weighted consensus model

- (ii) positive CT (pCT) images, in which imaging features associated with COVID-19 pneumonia could be unambiguously discerned, and
- (iii) negative CT (nCT) images, in which imaging features in both lungs were irrelevant to COVID-19 pneumonia.

They manually labeled 19685 CT slices, which we trained using the three classes. We used 4001 pCT, 9979 nCT, and 5705 NiCT scan images to train the models. The number of image samples used in work is compared with the base paper in 1.

The distribution of data is visualized in Fig. 2 and some sample images are shown in Fig. 3.

Each image is loaded and resized into (150, 150) pixels to speed up training. If the images are loaded with higher resolution, the computational cost exceeds and if the images are loaded with lower resolution, the model may not capture

Table 1 Data statistics of the base paper and our data

	Base paper	Our data
Positive	4001	4001
Negative	9979	9979
Non-informative	5705	5705
Total	19685	19685

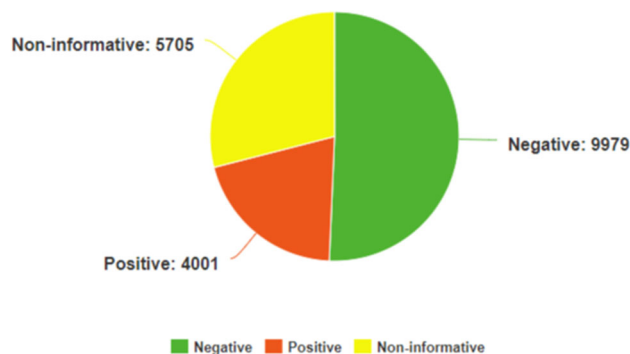


Fig. 2 Distribution of proportion of classes in the dataset

the essential features present in the dataset. The trade-off between computational cost and accuracy is balanced for better results.

The images are loaded with three channels and after resizing, the total shape of an image is 67500. Since the images do not present any RGB visuals to our naked eye, we converted the channels to 1 by loading the images as gray-scale images. Thereby saving space and speeding up computations.

While training, the traditional machine learning models expect a 2D array. So the images are flattened and sent as input.

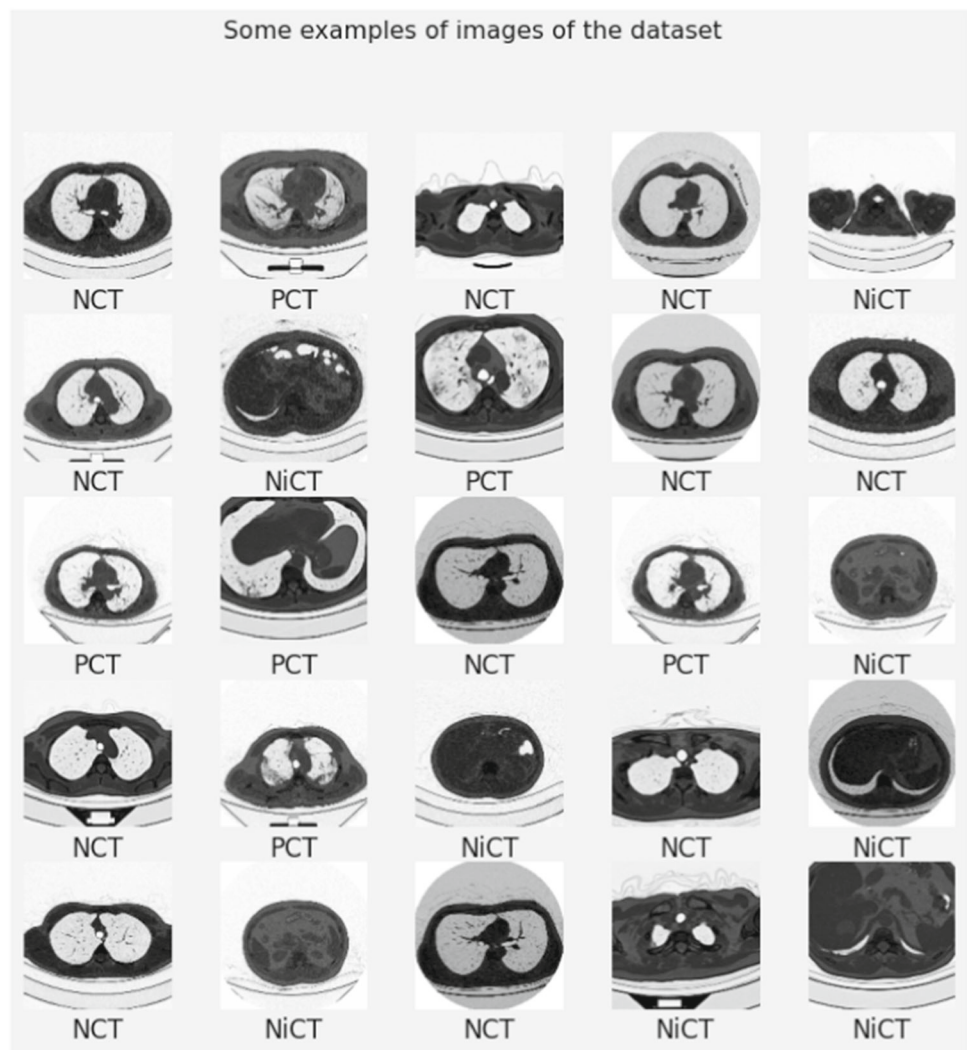
**4.2 Encoding the dataset labels**

The dataset contains 3 classes namely pCT (positive), nCT (negative) and NiCT (non-informative). Since the mathematical models cannot infer textual labels, they are encoded into numerical values. The order of labels is not mandatory as the label is a dependent variable. In the database, The total number of samples is 19685, out of which 4001 samples are labeled as positive, 9979 are labeled as negative and 5705 are non-informative.

**4.3 Splitting the dataset for training and testing**

The dataset is preprocessed and split into training and test sets in the ratio 9:1 as shown in the Table 2. Since there are many images, evaluating the performance on 1000-2000 images is optimal. After dividing the dataset, the test set is not modified and used to evaluate all the models.

**Fig. 3** Visualization of positive, negative and non-informative classes of the COVID-19 CT scan



**Table 2** Dataset samples after splitting into training and testing

Training set	Testing set
17716	1969

accelerates the training process and brings all the pixel data under a common scale.

#### 4.4 Data normalization

Images are made up of pixel value matrices. Color images include a separate array of pixel values for each color channel, such as red, green, and blue, whereas black and white images have a single matrix of pixels. Pixel values are frequently unsigned integers in the 0 to 255 range. Although these pixel values can be directly provided to neural network models in their raw format, this can cause problems like slow training. Instead, preprocessing the image pixel values before modeling, such as simply scaling pixel values to the range 0-1, might be quite beneficial. So we applied the min-max normalization technique to all the pixels in the image. This

## 5 Experimental results and analysis

To evaluate the performance of a classification model, several metrics such as classification accuracy, precision, F1-score, sensitivity, and specificity are used. We computed the metrics at different thresholds to observe the percentage of correctly classified classes and select the suitable one.

The training performance is evaluated using the following different performance metrics for each of the classes. The overall accuracy for all the classes is calculated as defined in Eqs. 1–6.

$$\text{Accuracy (each class)} = \frac{TN + TP}{TP + TN + FN + FP} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

**Table 3** The percentage of weightage of each Machine Learning algorithm and accuracy of all the models with training time and testing time

Models	Weightage (in%)	Accuracy (in%)	Training time (in s)	Testing time (in s)
Logistic regression	20.109	99.594	1569.287	0.135
SVM	20.109	99.594	983.467	146.367
kNN	20.068	99.391	37.807	689.590
Decision tree	19.606	97.105	283.367	0.169
Random forest	20.109	99.594	52.387	0.141
Weighted consensus	–	<b>99.645</b>	–	<b>0.413 (per sample)</b>

$$\text{sensitivity} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

where

TP—The original class is positive, as expected by the model.

FP—While the initial class was negative, the model expected a positive outcome.

TN—The original class is negative, as expected by the model.

FN—While the original class is positive, the model predicts a negative outcome.

The models are evaluated on 1969 images.

Since medical images are to be predicted, traditional performance metrics like accuracy alone are not enough. So we recorded the results over a wide range of metrics.

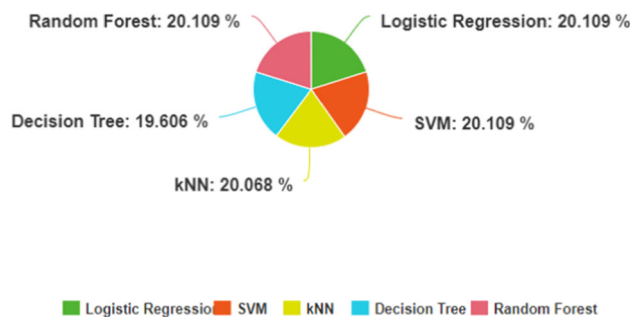
### 5.1 Execution time for training and testing

The models are trained and tested on the processor Intel Core i7-8700 CPU @ 3.20GHz\*12, 16GB of RAM. The time taken to train 17716 images and test 1969 images are recorded and shown in Table 3. The proposed model relies on the 5 base models. So the training time is the sum of training times of all the models.

### 5.2 Model weightage and accuracy

As proposed, the models’ accuracies are normalized and weights are calculated. The weights, along with accuracies, are shown in Table 3. The distribution of weights can be visually seen in Fig. 4.

The overall accuracy (99.645) is found to be more than the individual models’ accuracies. This highlights that even if one model mispredicts a test sample, other models collec-



**Fig. 4** Block diagram of the models’ weights

**Table 4** Sensitivity ( $C_1$ ) and Specificity ( $C_2$ ) values of Weighted Consensus Model of the three classes with different threshold values for the CT Scan Image dataset

Threshold	nCT		NiCT		pCT	
	$C_1$	$C_2$	$C_1$	$C_2$	$C_1$	$C_2$
0.1	1	0.97	1	0.98	0.99	0.99
0.2	1	0.99	1	0.99	0.99	0.99
0.3	0.99	0.99	1	0.99	0.99	0.99
0.4	0.99	0.99	1	0.99	0.99	0.99
0.5	0.99	0.99	0.99	0.99	0.99	1
0.6	0.99	1	0.99	0.99	0.99	1
0.7	0.99	1	0.99	0.99	0.98	1
0.8	0.99	1	0.98	0.99	0.98	1
0.9	0.97	1	0.96	0.99	0.96	1

tively add to the correct weightage, thereby classifying the sample correctly.

### 5.3 Sensitivity and specificity analysis of results

Sensitivity is a metric that calculates the number of correctly defined positive groups (i.e., the proportion of people that have a disease (affected) who are correctly identified as having the condition). Specificity, on the other hand, is a measure of how many negative groups were correctly defined. The sensitivity and specificity values are recorded at different thresholds ranging from 0.1 to 0.9. See Table 4. The optimal

**Table 5** Performance report of the Machine learning models for classification of CT scan images

Models	Class	Precision Accuracy (in%)	Recall Accuracy (in%)	F1 score Accuracy (in%)	Support
Logistic regression	0	1.00	1.00	1.00	982
	1	0.99	1.00	0.99	562
	2	1.00	0.99	1.00	425
	Accuracy	–	–	1.00	1969
	Macro avg	1.00	1.00	1.00	1969
	weighted avg	1.00	1.00	1.00	1969
SVM	0	1.00	0.99	1.00	982
	1	0.99	1.00	0.99	562
	2	1.00	1.00	1.00	425
	Accuracy	–	–	1.00	1969
	Macro avg	1.00	1.00	1.00	1969
	weighted avg	1.00	1.00	1.00	1969
<i>k</i> -NN	0	1.00	0.99	1.00	982
	1	0.98	0.99	0.99	562
	2	1.00	1.00	1.00	425
	Accuracy	–	–	0.99	1969
	Macro avg	0.99	0.99	0.99	1969
	weighted avg	0.99	0.99	0.99	1969
Decision tree	0	0.98	0.98	0.98	982
	1	0.97	0.98	0.97	562
	2	0.97	0.97	0.97	425
	Accuracy	–	–	0.97	1969
	Macro avg	0.97	0.97	0.97	1969
	weighted avg	0.97	0.97	0.97	1969
Random forest	0	0.99	1.00	1.00	982
	1	0.99	0.99	0.99	562
	2	1.00	0.99	0.99	425
	Accuracy	–	–	0.99	1969
	Macro avg	1.00	0.99	0.99	1969
	weighted avg	0.99	0.99	0.99	1969
Weighted consensus	0	1.00	1.00	1.00	982
	1	0.99	1.00	0.99	562
	2	1.00	1.00	1.00	425
	Accuracy	–	–	1.00	1969
	Macro avg	1.00	1.00	1.00	1969
	weighted avg	1.00	1.00	1.00	1969

threshold values vary for different classes. If the base models were to be considered alone, a default threshold of 0.5 might have been chosen. We have great control over the hyperparameter (threshold) with different threshold values and can be used under different circumstances. The optimal threshold was found to be 0.6 for nCT, 0.4 for NiCT and 0.5 for pCT. (The optimal values may be from one's perspective).

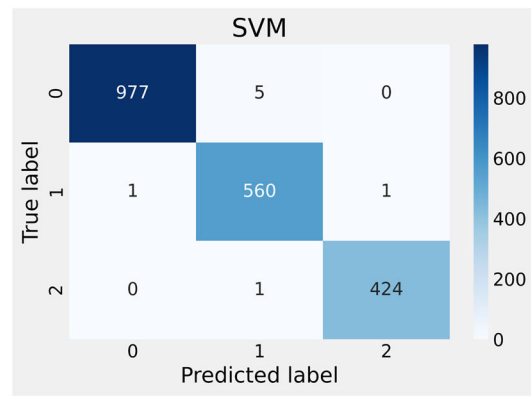
#### 5.4 Classification analysis reports

The overall performance report containing different basic metrics like precision, recall, F1-score, support is shown in Table 5. The table gives a clear picture of the performance of the individual models, which forms the basis for choosing them in the proposed weighted consensus model. The three classes 0, 1 and 2 correspond to negative, non-informative and positive classes, respectively.

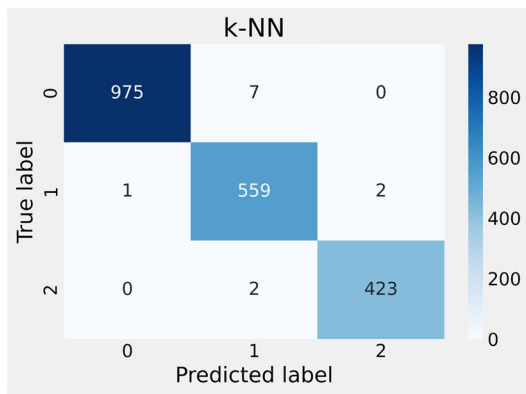




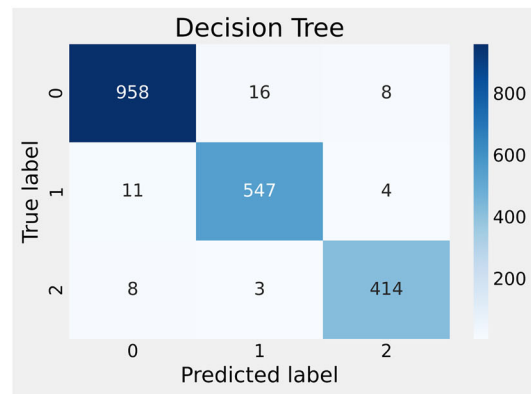
(a) Logistic Regression



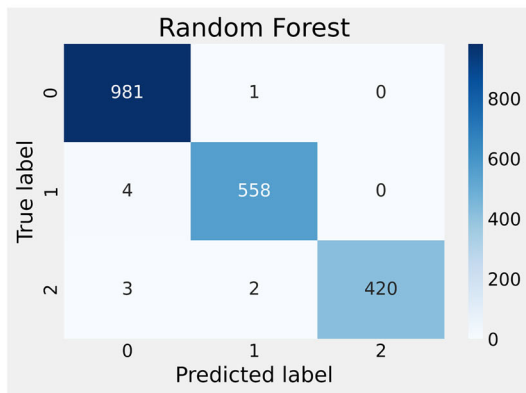
(b) SVM



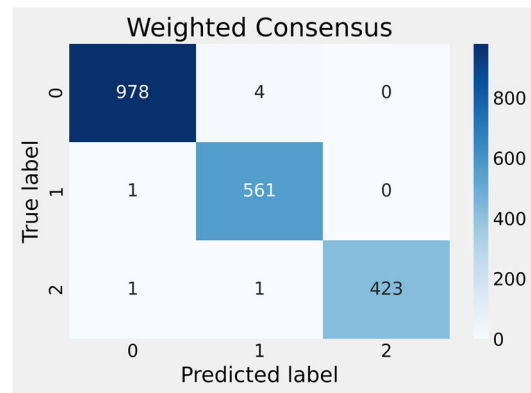
(c) k-NN



(d) Decision Tree



(g) Random Forest



(f) Weighted Consensus

Fig. 5 Confusion matrix analysis of COVID-19 chest CT scan

We used scikit-learn API to generate the classification report. By default, scikit-learn rounds off to 2 decimal places. We achieved accuracies close to 99.5% in Logistic regression, SVM and Random Forest models. So the accuracy metrics in the classification reports show 1.00. Support is

the number of actual occurrences of the class in the specified dataset. The support for different classes is close to the number of testing samples for that class. This is obvious by the accuracy, precision and recall.

**Table 6** Comparison of three classes with base paper [46] and the proposed model

Class	HUST-19(AUC)	WCM (AUC)
Positive versus (negative and non-informative)	0.991	0.9976
Negative versus (positive and non-informative)	—	0.9970
Non-informative versus (positive and negative)	0.994	0.9973
Average	—	0.9973

## 5.5 Confusion matrices of the classification results

The confusion matrices of individual models on test images are shown in Fig. 5. We adaptively chose threshold values to capture the best confusion matrices for each model using optimum sensitivity and specificity values from Table 5.

## 5.6 False positives and false negatives observed during classifications

The main goal of this work is to reduce the number of FPs or FNs while taking into account the trade-off. Therefore we computed the number of FNs and FPs at the end of each stage to show the efficacy of our proposed model. Tables 7, 8 and 9 show the FNs and FPs of different classes predicted on the weighted consensus model.

As shown in Table 7 for nCT scan, the FPs decrease as the threshold increases. This indicates that an image will be classified as positive only if it crosses that threshold value. This will reduce the chances of falsely predicting positive values. Similarly, FNs increase as the threshold increases.

The optimal threshold can be taken with different thresholds and their corresponding FPs and FNs, depending on the situation.

## 5.7 Results and discussions

In this paper, we propose a new weighted consensus model based on five machine learning classifiers, including Logistic Regression, SVM, KNN, Decision Tree, and Random Forest, to accurately predict classes while reducing false positives and false negatives. In the CT Scan medical data collection, we can tune the model to predict at different thresholds in three different groups, such as nCT, NiCT, and pCT.

In the nCT CT Scan class from Table 7, it can be observed that as we increase the threshold value, FP decreases, and in contrast to it, FN increases. Finally, at 0.5, we got significant values of FP and FN. Similarly, the effect of threshold values are shown for NiCT and pCT CT Scan images in Table 8, Table 9 respectively. Similar behavior is observed in both cases. In all the three classes (nCT CT Scan, NiCT CT scan and pCT CT Scan), the FP numbers decrease and FN numbers increase with an increase in threshold values. At a certain threshold value of 0.5, FP and FN numbers are observed to

**Table 7** False negatives and False positives of nCT CT Scan

Threshold	False positives	False negatives
0.1	23	0
0.2	8	0
0.3	6	2
0.4	4	3
0.5	2	4
0.6	0	6
0.7	0	7
0.8	0	8
0.9	0	28

**Table 8** False negatives and False positives of NiCT CT Scan

Threshold	False positives	False negatives
0.1	24	0
0.2	10	0
0.3	9	0
0.4	8	0
0.5	5	1
0.6	4	3
0.7	3	4
0.8	1	7
0.9	1	20

**Table 9** False negatives and False positives of pCT CT Scan

Threshold	False positives	False negatives
0.1	14	1
0.2	2	1
0.3	1	1
0.4	1	1
0.5	0	2
0.6	0	4
0.7	0	5
0.8	0	5
0.9	0	13

be minimum. We considered this threshold value for NiCT and pCT classes. Eventually, we got significant FP numbers and FN numbers on all three classes of CT scan data set at 0.5. Therefore, we conclude that the threshold value can be chosen to be 0.5 for this study. The overall performance of the proposed algorithm in terms of sensitivity and specificity values corresponding to all three classes of CT scan dataset for different threshold values are reported in Table 4.

For evaluating classification algorithms and models, apart from accuracy, log-loss is one of the most widely used metrics as it imposes a significant loss on wrong predictions. We found the log-loss for the Weighted Consensus model on the test set to be **0.1227** which is considered as good in all standard literature. This emphasizes the robustness and reliability of the proposed model.

With an accuracy of 99.645% and prediction time of 0.413 seconds per sample, the model is highly robust, promising and can be deployed for instant predictions.

HUST-19 [46] achieved an AUC value of 0.994 in distinguishing NiCT images from pCT and nCT images; and an AUC value of 0.991 in predicting pCT images for image-based prediction. The proposed weighted consensus model performed better with higher AUC scores for all “one-vs-rest” classes compared to Table 6.

The base paper [46] used HUST-19 to predict whether an image is COVID-19 positive, negative, or non-informative, with an AUC of 0.919. However, the weighted consensus model was able to perform with an accuracy rate of 0.996 and an average AUC score of 0.997. Therefore, under experimental conditions, the proposed weighted consensus algorithm provides more reliable results by outperforming the existing results.

## 6 Conclusions and future scope

This paper presented a weighted consensus model for classifying and identifying possible COVID-19 infection from CT scan images with outstanding accuracy. The proposed model performs as good as the existing best methods in terms of accuracy. Still, it is also quite fast as we normalized the images. The novel proposed method can minimize the false negatives and false positives depending on the requirements. This model will control the spread of infection by minimizing false negatives and reducing patients' mental trauma by minimizing false positives when the situation improves.

In the future, we want to extend this model to include continuous and periodic feedbacks to improve efficiency and make the model more robust. We are also collecting data locally from the hospitals and will train the model for better accuracy and robustness, which will be more acceptable in local conditions. Besides, we plan to use this proposed model

to identify other diseases and develop this as a more general model.

## Declarations

**Conflict of interest** The authors declare that there is no conflict of interest in this work.

**Ethical approval** We have used the secondary data available in the public domain and have not conducted any experiments involving human beings in this study.

## References

1. W. H. Organization. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/> (2019)
2. Novel, C.P.E.R.E.: The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi* **41**(2), 145 (2020)
3. Wang, D.; Hu, B.; Hu, C.; Zhu, F.; Liu, X.; Zhang, J.; Wang, B.; Xiang, H.; Cheng, Z.; Xiong, Y.; et al.: Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *Jama* **323**(11), 1061–1069 (2020)
4. Xie, Z.: Pay attention to SARS-CoV-2 infection in children. *Pediatr. Investig.* **4**(1), 1–4 (2020)
5. Udgata, S.K.; Suryadevara, N.K.: *Internet of Things and Sensor Network for COVID-19*. Springer, Singapore (2021)
6. Udgata, S.K.; Suryadevara, N.K.: COVID-19, sensors, and Internet of Medical Things (IOMT). In: *Internet of Things and Sensor Network for COVID-19*. Springer Briefs in Applied Sciences and Technology, pp. 39–53. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-7654-6\\_3](https://doi.org/10.1007/978-981-15-7654-6_3)
7. Udgata, S.K.; Suryadevara, N.K.: Advances in sensor technology and IOT framework to mitigate COVID-19 challenges. In: *Internet of Things and Sensor Network for COVID-19*. Springer Briefs in Applied Sciences and Technology, pp. 55–82. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-7654-6\\_4](https://doi.org/10.1007/978-981-15-7654-6_4)
8. Udgata, S.K.; Suryadevara, N.K.: COVID-19: challenges and advisory. In: *Internet of Things and Sensor Network for COVID-19*. Springer Briefs in Applied Sciences and Technology, pp. 1–17. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-7654-6\\_1](https://doi.org/10.1007/978-981-15-7654-6_1)
9. Ai, T.; Yang, Z.; Hou, H.; Zhan, C.; Chen, C.; Lv, W.; Tao, Q.; Sun, Z.; Xia, L.: Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* **296**(2), E32–E40 (2020)
10. Zhao, J.; Yuan, Q.; Wang, H.; Liu, W.; Liao, X.; Su, Y.; Wang, X.; Yuan, J.; Li, T.; Li, J.; et al.: Antibody responses to SARS-COV-2 in patients with novel coronavirus disease 2019. *Clin. infect. Dis.* **71**(16), 2027–2034 (2020)
11. Fang, Y.; Zhang, H.; Xie, J.; Lin, M.; Ying, L.; Pang, P.; Ji, W.: Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* **296**(2), E115–E117 (2020)
12. West, C.P.; Montori, V.M.; Sampathkumar, P.: COVID-19 testing: the threat of false-negative results. *Mayo Clin. Proc.* **95**(6), 1127–1129 (2020)
13. Xu, X.; Jiang, X.; Ma, C.; Du, P.; Li, X.; Lv, S.; Yu, L.; Chen, Y.; Su, J.; Lang, G.; Li, Y.; Zhao, H.; Xu, K.; Ruan, L.; Wu, W.: Deep learning system to screen coronavirus disease 2019 pneumonia. *arXiv* (2020)



14. Wang, S.; Kang, B.; Ma, J.; Zeng, X.; Xiao, M.; Guo, J.; Cai, M.; Yang, J.; Li, Y.; Meng, X.; et al.: A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *MedRxiv* (2020)
15. Sethy, P.K.; Behera, S.K.: Detection of coronavirus disease (COVID-19) based on deep features. Preprints **2020030300**, 2020 (2020)
16. Narin, A.; Kaya, C.; Pamuk, Z.: Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks. arXiv preprint [arXiv:2003.10849](https://arxiv.org/abs/2003.10849) (2020)
17. Chung, M.; Bernheim, A.; Mei, X.; Zhang, N.; Huang, M.; Zeng, X.; Cui, J.; Xu, W.; Yang, Y.; Fayad, Z.A.; et al.: CT imaging features of 2019 novel coronavirus (2019-NCOV). *Radiology* **295**(1), 202–207 (2020)
18. Li, X.; Zeng, X.; Liu, B.; Yu, Y.: COVID-19 infection presenting with CT halo sign. *Radiol. Cardiothorac. Imaging* **2**(1), e200026 (2020)
19. Venugopal, V.K.; Mahajan, V.; Rajan, S.; Agarwal, V.K.; Rajan, R.; Syed, S.; Mahajan, H.: A systematic meta-analysis of CT features of COVID-19: lessons from radiology. *medRxiv* (2020)
20. Singh, D.; Kumar, V.; Kaur, M.: Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks. *Eur. J. Clin. Microbiol. Infectious Dis.* 1–11 (2020)
21. Simpson, S.; Kay, F.U.; Abbara, S.; Bhalla, S.; Chung, J.H.; Chung, M.; Henry, T.S.; Kanne, J.P.; Kligerman, S.; Ko, J.P.; et al.: Radiological society of north America expert consensus statement on reporting chest CT findings related to COVID-19. Endorsed by the society of thoracic radiology, the American college of radiology, and RSNA. *Radiology: Cardiothoracic Imaging* **2**(2), e200152 (2020)
22. Ng, M.-Y.; Lee, E.Y.; Yang, J.; Yang, F.; Li, X.; Wang, H.; Lui, M.M.-S.; Lo, C.S.-Y.; Leung, B.; Khong, P.-L.; et al.: Imaging profile of the COVID-19 infection: radiologic findings and literature review. *Radiol. Cardiothorac. Imaging* **2**(1), e200034 (2020)
23. Song, F.; Shi, N.; Shan, F.; Zhang, Z.; Shen, J.; Lu, H.; Ling, Y.; Jiang, Y.; Shi, Y.: Emerging 2019 novel coronavirus (2019-NCOV) pneumonia. *Radiology* **295**(1), 210–217 (2020)
24. Fang, Y.; Zhang, H.; Xu, Y.; Xie, J.; Pang, P.; Ji, W.: CT manifestations of two cases of 2019 novel coronavirus (2019-NCOV) pneumonia. *Radiology* **295**(1), 208–209 (2020)
25. Xie, X.; Zhong, Z.; Zhao, W.; Zheng, C.; Wang, F.; Liu, J.: Chest CT for typical 2019-NCOV pneumonia: relationship to negative RT-PCR testing. *Radiology* 200343–200343 (2020)
26. Minaee, S.; Kafieh, R.; Sonka, M.; Yazdani, S.; Soufi, G.J.: Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. arXiv preprint [arXiv:2004.09363](https://arxiv.org/abs/2004.09363) (2020)
27. Santosh, K.: Ai-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multi-tudinal/multimodal data. *J Med syst* **44**(5), 1–5 (2020)
28. Zhang, Y.-D.; Satapathy, S.C.; Zhang, X.; Wang, S.-H.: COVID-19 diagnosis via densenet and optimization of transfer learning setting. *Cogn. Comput.* 1–17 (2021)
29. Wang, L.: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. arXiv preprint [arXiv:2003.09871](https://arxiv.org/abs/2003.09871) (2020)
30. Kermany, D.; Zhang, K.; Goldbaum, M.: Labeled optical coherence tomography (OCT) and chest X-ray images for classification. *Mendeley Data* **2** (2018)
31. Ni, Q.; Sun, Z.Y.; Qi, L.; Chen, W.; Yang, Y.; Wang, L.; Zhang, X.; Yang, L.; Fang, Y.; Xing, Z.; et al.: A deep learning approach to characterize 2019 coronavirus disease (COVID-19) pneumonia in chest CT images. *Eur. Radiol.* **30**(12), 6517–6527 (2020)
32. Rajaraman, S.; Antani, S.: Weakly labeled data augmentation for deep learning: a study on COVID-19 detection in chest X-rays. *Diagnostics* **10**(6), 358 (2020)
33. Novitasari, D.C.R.; Hendradi, R.; Caraka, R.E.; Rachmawati, Y.; Fanani, N.Z.; Syarifudin, A.; Toharudin, T.; Chen, R.C.: Detection of COVID-19 chest X-ray using support vector machine and convolutional neural network. *Commun. Math. Biol. Neurosci.* **2020**, Article-ID (2020)
34. Khalifa, N.E.M.; Taha, M.H.N.; Hassanien, A.E.; Elghamrawy, S.: Detection of coronavirus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest X-ray dataset. arXiv preprint [arXiv:2004.01184](https://arxiv.org/abs/2004.01184) (2020)
35. Ozturk, T.; Talo, M.; Yildirim, E.A.; Baloglu, U.B.; Yildirim, O.; Acharya, U.R.: Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput. Biol. Med.* **103792** (2020)
36. Zhang, Y.-D.; Satapathy, S.C.; Liu, S.; Li, G.-R.: A five-layer deep convolutional neural network with stochastic pooling for chest CT-based COVID-19 diagnosis. *Mach. Vis. Appl.* **32**(1), 1–13 (2021)
37. Zhang, Y.-D.; Satapathy, S.C.; Zhu, L.-Y.; Górriz, J.M.; Wang, S.-H.: A seven-layer convolutional neural network for chest CT based COVID-19 diagnosis using stochastic pooling. *IEEE Sens. J.* (2020)
38. Maior, C.B.; Santana, J.M.; Lins, I.D.; Moura, M.J.: Convolutional neural network model based on radiological images to support COVID-19 diagnosis: evaluating database biases. *Plos One* **16**(3), e0247839 (2021)
39. Saba, L.; Agarwal, M.; Patrick, A.; Puvvula, A.; Gupta, S.K.; Carriero, A.; Laird, J.R.; Kitas, G.D.; Johri, A.M.; Balestrieri, A.; et al.: Six artificial intelligence paradigms for tissue characterisation and classification of non-COVID-19 pneumonia against COVID-19 pneumonia in computed tomography lungs. *Int. J. Comput. Assist. Radiol. Surg.* 1–12 (2021)
40. Liu, Q.; Leung, C.K.; Hu, P.: A two-dimensional sparse matrix profile densenet for COVID-19 diagnosis using chest CT images. *IEEE Access* **8**, 213718–213728 (2020)
41. Fernández-Miranda, P.M.; Bellón, P.S.; Del Barrio, A.P.; Iglesias, L.L.; García, P.S.; Aguilar-Gómez, F.; González, D.R.; Vega, J.: Developing a training web application for improving the COVID-19 diagnostic accuracy on chest X-ray. *J. Dig. Imaging* 1–15 (2021)
42. Guo, G.; Liu, Z.; Zhao, S.; Guo, L.; Liu, T.: Eliminating indefiniteness of clinical spectrum for better screening COVID-19. *IEEE J. Biomed. Health Inf.* (2021)
43. Budak, Ü.; Çibuk, M.; Cömert, Z.; Şengür, A.: Efficient COVID-19 segmentation from CT slices exploiting semantic segmentation with integrated attention mechanism. *J. Dig. Imaging* (2021)
44. Wang, S.-H.; Fernandes, S.; Zhu, Z.; Zhang, Y.-D.: Avnc: attention-based VGG-style network for COVID-19 diagnosis by CBAM. *IEEE Sensors J.* (2021)
45. Liu, K.-C.; Xu, P.; Lv, W.-F.; Qiu, X.-H.; Yao, J.-L.; Jin-Feng, G., et al.: CT manifestations of coronavirus disease-2019: a retrospective analysis of 73 cases by disease severity. *Eur. J. Radiol.* 108941 (2020)
46. Ning, W.; Lei, S.; Yang, J.; Cao, Y.; Jiang, P.; Yang, Q.; Zhang, J.; Wang, X.; Chen, F.; Geng, Z.; et al.: ICTCF: an integrative resource of chest computed tomography images and clinical features of patients with COVID-19 pneumonia. (2020)