Elucidation of Infection Asperity of CT Scan Images of COVID-19 Positive Cases: A Machine Learning Perspectives

Dasari Naga Vinod
SRM Institute of Science and Technology

S.R.S. Prabaharan (srsprabaharan1611@gmail.com)
SRM Institute of Science and Technology https://orcid.org/0000-0001-8155-0677

Research Article

Keywords: Coronavirus, Computed Tomography, Severity, Lung Disjunction, Machine Learning

Posted Date: April 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1418988/v1

Elucidation of Infection Asperity of CT Scan Images of COVID-19 Positive Cases: A Machine Learning Perspectives

Dasari Naga Vinod[1,2], S.R.S. Prabaharan[3,*]

[1]*Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, 603203, Kanchipuram, Chennai, TN, India.*

[2]*Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, 600062, Chennai, TN, India.*

[3]*Directorate of Research, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, 603203, Kanchipuram, Chennai, TN, India.*

[*] **srsprabaharan1611@gmail.com (Corresponding author)**

Abstract

Owing to the profoundly irresistible nature of the SARS-CoV-2 infection, an enormous number of individuals halt in the line for CT Scan assessment, which overburdens the medical practitioners, radiologists, and adversely influences the patient's remedy, diagnosis, as well as restraint of the epidemic. Medical facilities like intensive care systems and mechanical ventilators are restrained due to highly infectious diseases. It turns out to be very imperative to characterize the patients as per their asperity levels. This article exhibited a novel execution of image segmentation and a machine learning approach for Covid-19 contamination asperity identification. With the help of the image segmentation model and machine learning classifier, we can identify and classify Covid-19 individuals into three asperity classes such as early, progressive and advanced, with an accuracy of 96% using chest CT scan image database. Experimental outcomes on an adequately enormous number of CT scan images exhibit the adequacy of the machine learning mechanism developed and recommended to identify Coronavirus severity.

Keywords: *Coronavirus; Computed Tomography; Severity; Lung Disjunction; Machine Learning.*

1. Introduction

Covid-19 has evolved into a global epidemic, conferring a severe health hazard globally. Many nations have witnessed a 2-surge form of stated cases, through the 1st surge in spring, 2nd wave in late summer as well as autumn (Vahidy et al. 2020; Fan G et al. 2021). The second surge of Coronavirus had been foreseen months prior and had earlier existed in Asian regions (Saito et al. 2021). Due to the second wave of Covid-19, local governments have announced related confining protocols. However, experimental data in the second surge varies from the first wave parameters, such as age group and contamination vulnerability (Ballester et al. 2020; Long et al. 2020). Certainly, it has been recommended that the second surge might be associated with the presence of a new-fangled variant in SARS-CoV-2, described double mutant. The affinity and divergence among the features of the two waves remain highly anonymous. Patients with a sore throat and dry cough were more in 2nd surge. In the subsequent surge, the oxygen necessity, as well as ventilators requirement, are more. The contaminated older age group keeps on being vulnerable. The insignificant higher extent of patients is more young age group, over 70% of patients are below 40 in the second wave. Individuals between 30-45 age group have tested positive before a year with 21%. As of March 04, 2022, 44,01,80,084 cases of Covid-19 with 59,72,651 deceased have been recorded (Covid-19 Dashboard). The comparison between wave 1.0 and wave 2.0 of covid-19 and its impacts is exhibited in Table 1. In the first wave, the infection has transformed many times, and there are different unidentified changes. While it reveals to the second wave there is a double mutant and the UK, Brazilian, and South African variants that have shown higher contagiousness.

Table 1 Analogy of Surge 1.0 vs. Surge 2.0 of Covid-19 and its impacts

Parameters	Wave 1.0	Wave 2.0
Original form	SARS-CoV-2	Double mutant
Asymptomatic	Lower Proportion	Higher Proportion
Breathlessness	Lower	Higher
Symptoms	Sore throat, dry cough, fever, body pains	Fatigue, mental fatigue, allergy, Dyspnea
Oxygen requirement	Lower (41.5%)	Higher (54.5%)
Age group	Older age group	No difference

Recognizing the Covid-19 individuals who require intensive medical care utilizing automated asperity appraisal strategies by machine learning as well as deep learning turns out to be highly urgent in this epidemic and also for fast identification of Covid-19 disease among healthy (Vinod et al. 2020) and pneumonia individuals (Vinod et al. 2021) enabled by chest X-ray furthermore CT scan pictures. Ideal appraisal of Coronavirus individuals at the beginning phase is currently an emergency errand if infection movement, emergency time, as well as death ratio is anticipated to be limited. However, precisely enacting the infection asperity in X-ray and CT pictures is very challenging. A few authors tracked down that the majority of the Covid-19 individuals have non-severe (mild or progressive) signs (Verity et al. 2020). Current investigations showed that the fatality pace of non-asperity Covid-19 individuals is a lot larger (~20 times) than asperity ones (Chen et al. 2020). Other studies exhibited that initial recognition of Coronavirus individuals that can be stated to the progressive as well as advanced levels is important now that the average time in the 1st surge: shortness of breath is only 5 days as well as intense respiratory distress syndrome is around eight days (Shan et al. 2021). In addition, patient management, as well as remedy category, are mainly reliant upon the asperity of the

infection. CT and X-ray have been effective methods for categorizing lung reforms and anomalies because of Covid-19 contamination. The past analysis demonstrated the progressions of Covid-19 infection in the lungs because of "ground-glass opacity, crazy paving pattern, consolidation, vascular enlargement, lower lobe involvement, and bilateral infiltration" (Amyar et al. 2020). Researchers have exhibited that asperity evaluation from CT scan pictures of Covid-19 individuals is an excellent method that aids in battling this exceptionally infectious disease. The dominant research contributions of this article outlined:

- We have recommended a machine learning technique to recognize the asperity levels in the Covid-19 patients through CT Scan pictures.
- We have rigorously examined the Covid-19 CT scan pictures for patients intending to screen the early, progressive, and advanced stages. As an outcome, all the pictures exhibiting lesions were re-affirmed by experienced radiologists.
- The obtained results were analyzed with accuracy, receiver operating characteristics, as well as a confusion matrix.
- The recommended technique has obtained an accuracy of 96% for the qualitative study.

The following sections are described: Section 2 analyzes the current X-ray and CT scan pictures for severity examination utilizing machine learning and deep learning positioned CNN. Section 3 described the proposed model and the associated database. Section 4 exhibits the experimental outcomes and examination amidst the accomplishment of the categorization, and in the end, inferences are illustrated in Section 5.

2. Literature Survey

Deep learning positioned infection asperity evaluation is more accurate and computable than radiologist evaluation, subjective reports. Regardless of profoundly encouraging verdicts of

deep learning techniques for Coronavirus infection asperity evaluation in X-ray pictures, a couple of analyses correlated with this subject have been stated.

He et al. (2021) recommended a synergetic learning technique to segment infection asperity into vulnerable / non-vulnerable, forming the evaluation into a 2-way categorization chore. Authors acquired 0.98 exactness utilizing 666 computed tomographic pictures. Zhu et al. (2020) employed the transfer learning mechanism amidst 131 X-ray pictures in 84 individuals to characterize Coronavirus individuals into four levels: gentle, moderate, severe, as well as critical. Albeit multiclass asperity evaluation was accomplished with the best technique producing an MAE of 0.85, the database is tiny enough to verify the Convolutional Neural Networks result broadly. In addition, they did not utilize the RoC assessment or precision accomplishment analysis measure.

Li et al. (2020) characterized infection asperity as vulnerable and non-vulnerable in 531 thick-area Computed Tomography samples utilizing a computerized deep learning technique. Authors employed binary imaging biomarkers: The anomaly part as well as normal contamination, for the asperity model's evaluation and achieved an AUC of 0.97. Tang et al. (2021) exhibited that machine learning techniques dependent on perceptible lineaments obtained in Computed Tomographic lung pictures can recognize asperity as well as non-asperity Coronavirus individuals. An exactness acquired to this binary categorization was 0.87, utilizing 176 Computed Tomographic pictures. Xiao et al. (2020) fostered a profound learning strategy positioned on residual CNN (ResNet34) to assess coronavirus infection asperity and further evaluate infection movement in Coronavirus individuals. They accomplished an exactness of 0.81 utilizing computed tomographic pictures of 408 Coronavirus individuals.

Previously, it is revealed that upon utilizing 729 Computed Tomography samples of COVID-19 individuals, a pre-trained deep neural network model was used to characterize infection

asperity and non-asperity (Yu et a. 2020) and they accomplished an exactness of 95.34 percent. An independent study carried out by Carvalho et al. (2020) disclosed the computer-aided prognosis involving the ANN model to analyze COVID-19 individuals into early, progressive, as well as advanced stages amidst an exactness of 82 percent utilizing hundreds of Computed Tomography scans (total of 229 images) of COVID-19 individuals. Also, it is recommended that a deep learning strategy was attempted for Coronavirus asperity characterization in terms of different stages such as early, progressive and advanced. Zhang et al. (2020) achieved the overall veracity of 91.6% utilizing 661 CT scans.

Emrah Irmak (2021) suggested that an automated convolution neural network (CNN) technique is developed and recommended to separate Covid-19 individuals into four asperity phases: early, progressive, advanced, as well as critical amidst an exactness of 95.5% utilizing chest X-ray pictures. Imaging lineaments had a powerful impact on the mechanism outcome, while an amalgamation of medical and imaging features produced a good execution overall. Albeit oversampling produced mixed outcomes, it accomplished the good technique execution in (Quiroz et al. 2021) investigation. Logistic regression technique classifying among mild and severe cases obtained better results for medical lineaments with the specificity of 90.6%, AUC of 0.848, and sensitivity of 45.5%, similarly in radiology lineaments specificity of 90.1%, AUC of 0.926, and sensitivity of 81.8%.

A typical pre-essential system for automatic prognosis of Coronavirus utilizes the segmentation of lung or lung lobe depending upon chest CT pictures. A few deep learning techniques suggest the division of lung in CT pictures with Coronavirus. Generally, U-Net (Ronneberger et al. 2015) disjoints lung areas and lung scores in Coronavirus applications (Zhang et al. 2020; Cao et al. 2020; Huang et al. 2020). Qi et al. (2020) used U-Net to depict the lesions in the lung. They elicited radiometric highlights of Coronavirus contracted individuals with the underlying seeds provided by a radiologist for foreseeing hospital stay. Likewise, a few variations of U-

Net analyses and evaluate the severity of Coronavirus. Jin et al. (2020) proposed a two-way channel to diagnose Coronavirus in CT pictures, and they employed U-Net++ (Zhou et al. 2018) to identify the entire lung region as well as segment injuries from lung areas. Shan et al. (2020) incorporated human-in-the-loop methodology into the training cycle of VB-Net (a variant of V-Net). This process is automatic to address the issues of lacking manual labels amid disjuncture in Computed Tomographic pictures.

Ahmad et al. (2022) present a high-level AI device to decide the seriousness of Coronavirus to gentle, moderate, and extreme from the lung CT pictures. The authors utilized a bunch of perceptible 1st and 2nd order factual surface highlights from each picture. The suggested model obtained veracity of 90.95%. Cai et al. (2020) describe the clinical information with the infection asperity, clinical treatment as well as clinical results were taken for each individual. Lesion volume, lung volume, the fraction of nonlesion lung volume, as well as non-lesion lung volume were calculated in Computed Tomographic films through a 2 U-Net mechanism validated for the disjunction of the lung as well as Coronavirus lesions of Computed Tomographic films. This mechanism achieved an Area Under Curve of 0.927 as well as 0.929 in the categorization of mild as well as asperity cases.

Table 2 Assessing COVID-19 Severity score using Chest CT traces

Score	**Degree of Opacity in the size of lung Entanglement**
0	No Opacity
1	Ground Glass Opacity
2	Consolidation in lung
3	White-out

The Coronavirus lung infection erects owing to the degree of imperviousness (opacity) and the size of lung entanglement (Wong et al. 2020). The degree of opacity estimated for each region

of the lung is represented as 0 - 3 (lung at the left side and the right side) as exhibited in Table 2.

3. Materials and Methods

3.1 Data Acquisition

The dataset utilized in this chore is an open-source GitHub archive (Processed dataset available) which immediately resides around 3000 CT scan Covid-19 pictures out of these 1000 for early cases, 1000 progressive cases, and 1000 for advanced cases separated with the help of lung segmentation method and classification with machine learning technique. The archive of the image dataset is open for lung disjuncture techniques. The whole image dataset is verified as well as elucidated, residing the verdict of the CT scan pictures.

3.2 Simulation Set-Up

The analysis of this examination is carried out on an NVIDIA GeForce MX350 by 8 GB RAM, Windows 10, as well as Intel Core i5 11th Gen GPU 4.2 GHz, whereas the simulation platform resides in Anaconda-Jupyter.

3.3 Proposed Mechanism

The mechanism of the recommended technique is represented in Fig. 1, where the input is the Computed Tomographic Covid-19 positive images, and the outcome is the lung disjuncture as well as asperity evaluation of Covid-19 individuals (early or progressive or advanced). Each CT image is refined by various image pre-processing steps like resizing the images with 224 x 224 size, which will assist us in reducing common issues viz. bias, code complexity, training time, and applying min-max normalization to normalize the data.

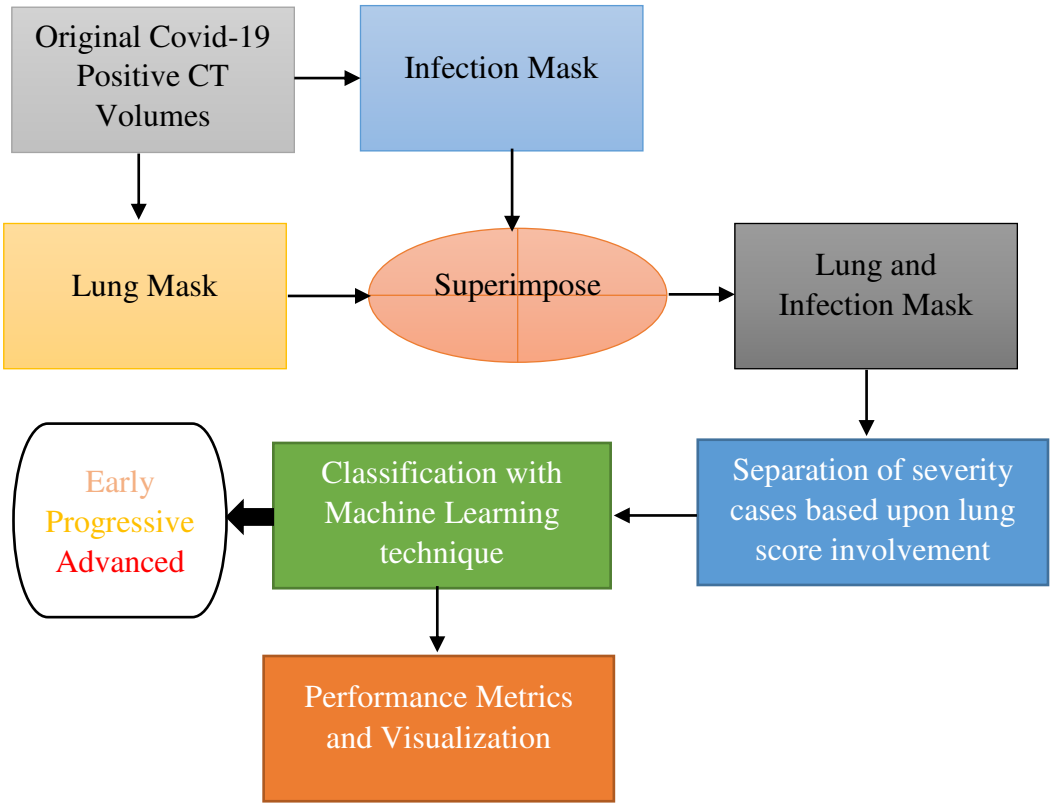


Fig. 1. A pipeline of the proposed Covid-19 infection severity identification model

The dataset's metadata has four major classes of pictures labeled as the original CT volumes, lung masks, infection masks, and superimpose of lung and infection masks for all images in the database, as shown in Fig. 1. These full-fledged pictures have corresponding copies under each label, highlighting various aspects of the original CT volume, evaluating the lung score involvement in each slice then separating the asperity stages in the CT scan image dataset. Employ multiple machine learning techniques to the image database to classify the three severity stages: early, progressive, and advanced cases. Random Forest classifier gives the best prediction accuracy compared to other machine learning and CNN techniques. Furthermore, corresponding performance metrics, confusion matrix, and receiver operating characteristics are drawn for analyzing the performance model.

Algorithm 1: Covid-19 Positive CT scan images for Pre-processing

Input : Input Image $I(e)_{(e)\,\varepsilon\,(1,2,\ldots n)}$

Output : Output Image $O(e)_{(e)\,\varepsilon\,(1,2,\ldots n)}$

```
Begin
    For each Input image, I do
        For {e} = 1 to n do
            Convert input image I into gray and resize the image with 224*224
            Apply min-max normalization and return the output image(O)
        End For
    End For
```

Where e is the Covid-19 +ve chest CT scan image dataset, as well as n is the total number of pictures.

Algorithm 2: Covid-19 Positive CT scan images for Severity Identification

Input : Input Image $O(e)_{(e)\,\varepsilon\,(1,2,\ldots n)}$

Output : O (E, P, A)

```
Begin
    For each Input image O, do
        For {e} = 1 to n do
            f ← Lung Mask
            g ← Infection Mask
            α ← f ⊕ g
            IF α ← 0 < x < 30, Then
                Label ← E,
            ELIF α ← 30 < x < 70,
```

Label ← P

ELSE

Label ← A

ENDIF

End For

End For

Where α is the superimposing of lung and infection masks, E represents early-stage, P represents the progressive stage, and A represents advanced stage.

Algorithm 3: Train the O (E, P, A) images enabled by Machine Learning Classifier

Input: Input Image O (E, P, A) $_{(E, P, A)\ \varepsilon\ (1, 2, \ldots n)^3}$, E = P = A, batch size = 100, testing ratio = 20%-30%-40%-50%, Training ratio = 80%-70%-60%-50%, Cross-validation = Folds (5, 10, 15, 20 ,25) number of trees =100, Classifier = Random Forest.

Output: Confusion Matrix, Visualization, Performance metrics.

Begin

For Input O (E, P, A), do

x = concatenate {E, P, A} ε $(1,2,\ldots n)^3$ (inputs)

y = concatenate {E, P, A} ε $(1,2,\ldots n)^3$ (target labels)

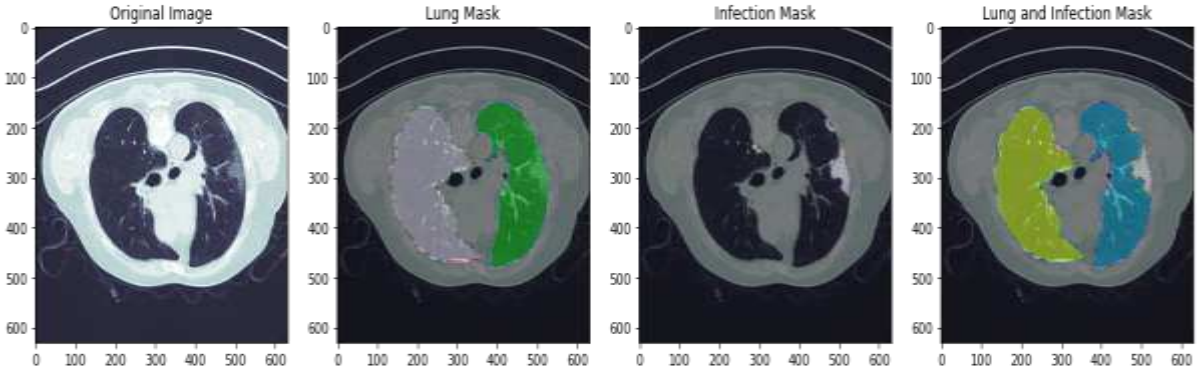
End For

Image pre-processing technique as described in algorithm-1, severity identification of the image database framework illustrated in algorithm-2, and finally, categorization as well as accomplishment metrics of the proposed technique as exhibited in algorithm-3.

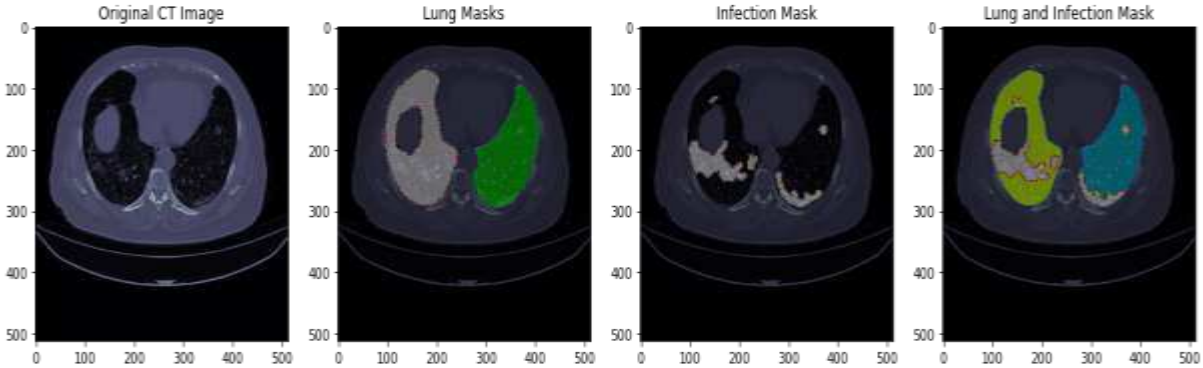
3.4 Covid-19 Lung Infection Asperity

Lung parenchyma, as well as Coronavirus contaminations disjuncture, were accomplished on Computed Tomographic scans utilizing novel deep learning techniques like U-Net. This

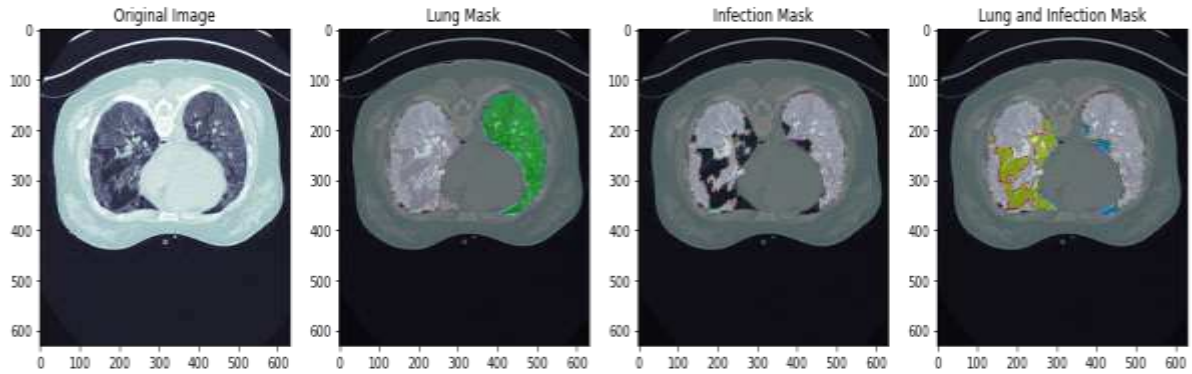
mechanism is refined to provoke a lung mask for the input Computed Tomographic slice. The disjunctor lung is then provided to other deep learning methods to recognize the contamination areas within the disjunct computed tomography picture. The generated contamination mask is utilized to identify coronavirus. Moreover, the coronavirus contamination lesion is assessed by calculating the % of the contamination lung pixels.



(a)



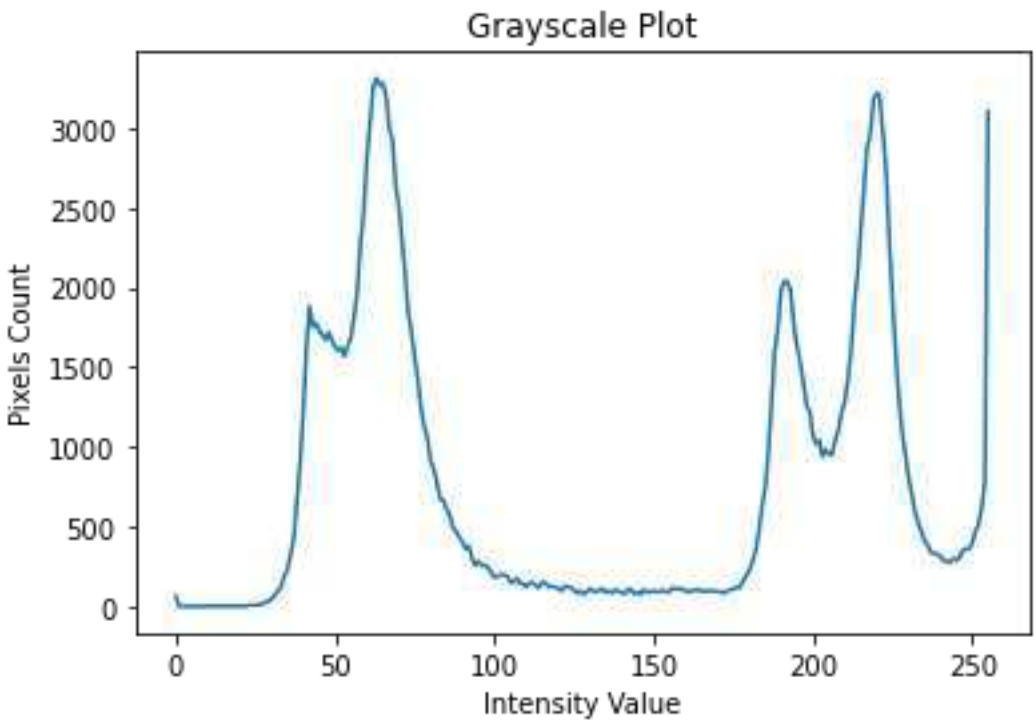
(b)



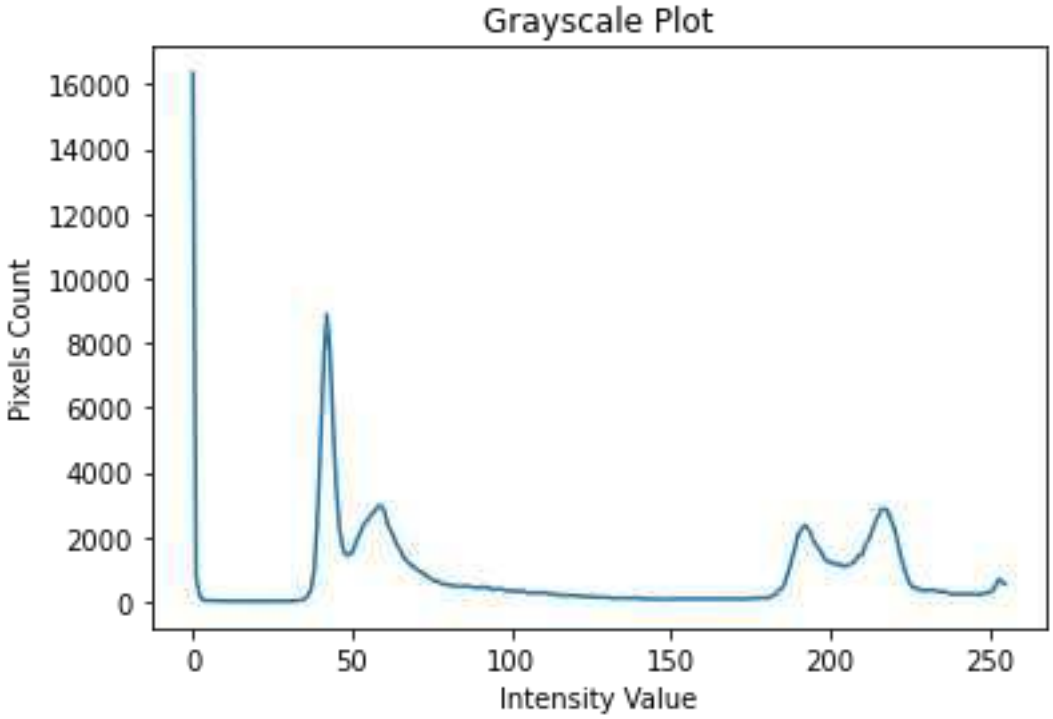
(c)

Fig. 1. A sample CT scan images with severity stages: (a) Lung involvement endures ground-glass opacity it exhibits individual in an early phase of Covid-19 contamination. (b) Lung involvement ensures consolidation; it demonstrates individuals in the progressive stage of Covid-19 contamination. (c) Lung involvement endures white out. It exists in individuals in the advanced phase of Covid-19 contamination.

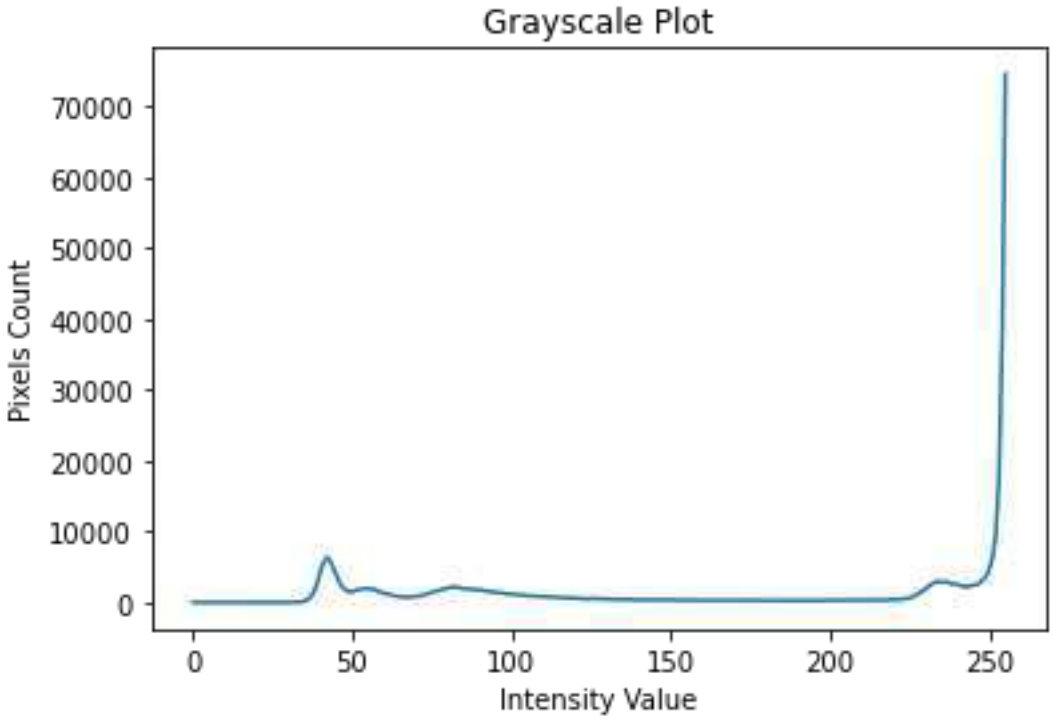
Lung entanglement score is estimated among 0 - 3 for each lung depending upon ground-glass opacity, consolidation, as well as white-out. Indicated as 0 for no involvement, 1 for 1% - 30% involvement ensures early stage, 2 for 31% - 70% involvement endures progressive stage and 3 for 71% - 100% involvement ensures an advanced stage. The color maps are applied to the lung and infection masks for better visualization of the lesion disjunction. The identification of Coronavirus severity depended on the prophecy maps produced with the lesion disjunction networks, described in Fig. 1. (a, b, c). The corresponding grayscale plots are exhibited in Fig. 2. (a, b, c), respectively.



(a)


Grayscale Plot
Pixels Count
16000
14000
12000
10000
8000
6000
4000
2000
0
0
50
100
150
200
250
Intensity Value

(b)


Grayscale Plot
Pixels Count
70000
60000
50000
40000
30000
20000
10000
0
0
50
100
150
200
250
Intensity Value

(C)

Fig. 2. A sample CT scan image severity grayscale plots with 0-255 grayscale intensity values: (a) Early-stage contains 3000 pixels count (b) Progressive Stage contains 16000 pixels count (c) Advanced stage contains 60000 pixels count.

3.5 Random Forest Categorization

Random Forest (RF) is a superintended learning technique employed for categorization as well as regression disputes. Nevertheless, it is essentially used for categorization disputes. We envisage that a forest involves trees as well as higher trees, indicating a higher vital forest. Generally, the RF technique generates decision trees on data tests as well as consequently gets the forest from every one of them, ultimately electing the finest explication over voting. A group mechanism is finest than a solitary decision tree since it curtails the over-fitting by averaging the outcome. The mechanism of the RF classifier is illustrated in algorithm-4.

Algorithm 4: Random Forests for Regression or Classification

1. For r = 1 to S:
   i. Construct a bootstrap model K of size L from the training data.
   ii. Develop a random-forest tree F_r to reboot the data by iteratively recurrent the ensuing stages for every terminal node until the tiniest node size l_{min} is attained.
      (a) Choose n variables at random from the q variables.
      (b) Select the finest variable/split-point between the n.
      (c) Divide the node into two daughter nodes.
2. Result of the collection of trees $\{\boldsymbol{F_r}\}_{\boldsymbol{1}}^{\boldsymbol{S}}$.

To accomplish a prophecy at a new point y:

Regression: $\boldsymbol{\phi_{rf}^{S}(y) = \frac{1}{S}\sum_{r=1}^{S} F_r(y)}$.

Classification: Consider β_r (y) to be the class prophecy of the rth random forest tree.

$\boldsymbol{\beta}_{rf}^{S}(\boldsymbol{y})$ = majority vote $\{\boldsymbol{\beta}_{r}(\boldsymbol{y})\}_{1}^{S}$.

End For

Where S is the number of trees, $\boldsymbol{F_r}$ is the rth tree and β_r is the class value for rth tree.

RF classifier is an ensemble ML technique for better prediction than various machine learning classifiers adapted for our technique. The number of trees, as well as the endorsed number of precedents, are 100 is utilized in the RF categorization to course if the batch prophecy is being accomplished. Large / lesser precedents may be afforded, but this gives exertion a chance to define an afforded batch size. We enforced an RF classifier on subgroups of the CT scan image database to classify early, progressive and advanced cases of Covid-19 individuals.

4. Results and Analysis

Assessments like attainment assessment measurements should demonstrate the precision and legitimacy of picture categorization strategies. There are many notable and ordinarily utilized performance assessment measurements for image categorization disputes in the state of the art. These measurements are obtained from the confusion matrix, a table utilized to portray grouping methods. Exactness, explicitness, affectability, and accuracy are viewed as the most well-known execution assessment measurements. The exhibition assessment of the models in this article is made utilizing the previously mentioned measurements, the RoC curve. Equivalent principles to every one of these measurements can be found in Equations 1, 2, 3, 4, and 5.

$$\text{Accuracy} = \frac{\text{True Positive (TP)+True Negative (TN)}}{\text{True Positive (TP)+True Negative (TN)+Flase Positive(FP)+False Negative(FN)}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive +Flase Positive}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive+ False Negative}} \quad (3)$$

$$\text{F-Measure} = \frac{2 \text{ x Precision x Recall}}{\text{Precision+Recall}} \quad (4)$$

$$\text{Matthews Correlation Coefficient (MCC)} = \frac{\text{TP x TN} - \text{FP x FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (5)$$

We have enforced distinct machine learning classifiers such as Logistic Regression, RF, Multi-Layer Perceptron (MLP), Support Vector Machine, and deep learning classifiers like Convolution Neural Network (CNN) to foresee the severity of Covid-19 samples. Random Forest classifier obtained the best accuracy when correlated amidst various classifiers by train-test split mechanism as illustrated in Table 3.

Table 3 Performance metrics with different classifiers by train (80%) - test (20%) splitting technique

Classifiers	Accuracy (%)
Random Forest	**92.5**
Random Tree	90.5
Logistic Regression	91.7
SVM	82.8
MLP	86.3
CNN	79.6

The presentation of the recommended mechanism for Coronavirus severity order is assessed utilizing the twenty-five-fold cross-validation strategy for Coronavirus infection asperity appraisal. Dataset is parted into twenty-five parts. In every fold, 24 perceptions are utilized to prepare the classifier. The rest of the fold is utilized to verify the trained classifier. The analyses are rehashed multiple times. Grouping execution for the errand is assessed for every fold, as well as the normal order execution of the technique is determined. Execution measurements

are determined and exhibited in the confusion matrix as shown in Table 4. And the performance metrics of the model with each class as illustrated in Table 5.

Table 4 Confusion Matrix by 25-fold cross-validation mechanism

Class	Early	Advanced	Progressive	Support
Early	955	5	40	1000
Advanced	3	977	20	1000
Progressive	30	20	950	1000
			Total	3000

Table 5 Performance metrics in True Positive (TP), False Positive (FP), Precision, Recall, F-Measure, and MCC by Random Forest Classifier with 25-fold cross-validation mechanism.

Class	TP	FP	Precision	Recall	F-Measure	MCC
Early	0.95	0.017	0.96	0.95	0.96	0.93
Advanced	0.97	0.016	0.96	0.97	0.97	0.95
Progressive	0.95	0.032	0.93	0.94	0.94	0.91

Finally, we have applied the various splitting mechanisms as exhibited in Table 6 by random forest classifier. Compared to train and test split methods, cross-validation with fold-25 obtained the highest accuracy with 96% out of fold-5, fold-10, fold-15, and fold-20, respectively, as shown in Table 7.

Table 6 Classification performance for various Training and Testing Splits by Random Forest Classifier.

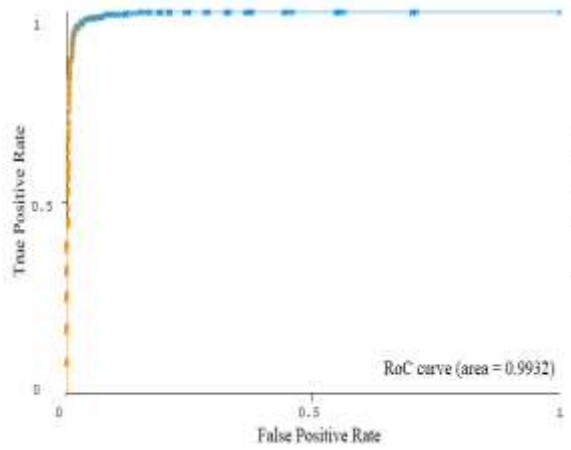
Train-Test Splitting (%)	Image Count (Train -Test)	Accuracy
90 - 10	2700-300	94.3
80 - 20	2400-600	92.5
70 - 30	2100-900	91
60 - 40	1800-1200	92.3
50 - 50	1500-1500	90.5

Table 7 Accuracy metrics for various folds by Random Forest Classifier

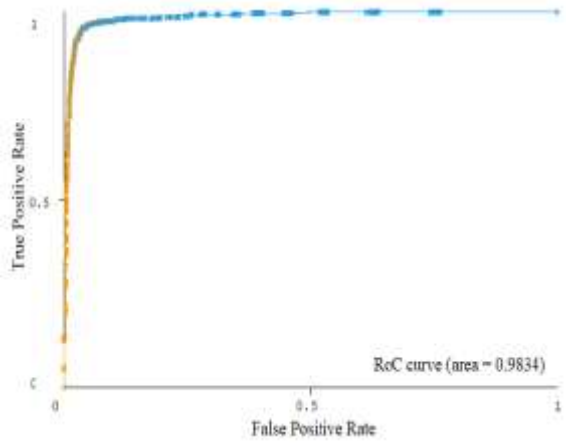
Cross-Validation	Accuracy (%)
Fold-5	93.8
Fold-10	94.8
Fold-15	94.8
Fold-20	95.4
Fold-25	**96**

During training the model, the verification samples obtained around 0.4 loss rate as well as 96% veracity. The loss rate exhibited a strong match among training as well as testing, affirming that our model is not experiencing overfitting or underfitting in an image database. Then, the RoC curve was obtained to analyze the accuracy of the mechanism, as exhibited in Fig. 4 (a, b, and c). An investigation of CT Scan images by early, progressive, and advanced

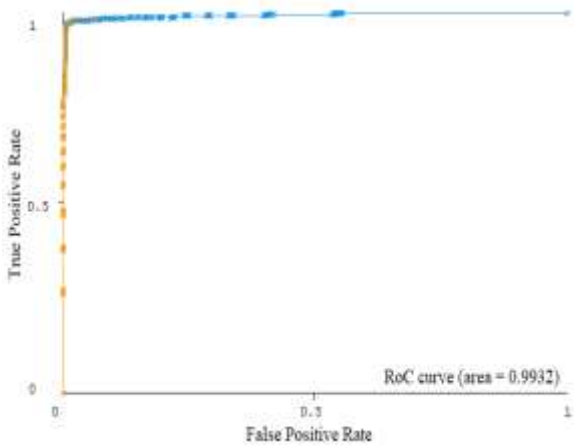
levels correctly identified with 96% precision and 97% recall scores when assessed on 3000 CT Scan images via cross-validation method (Fold-25).



(a)



(b)



(c)

Fig. 4 (a) Receiver Operating Characteristics graph exhibits early is the target class. (b) The Receiver Operating Characteristics graph provokes progressive is the target class. (c) The Receiver Operating Characteristics graph, amidst advanced cases, is the target class.

Table 8 Analogy of the recommended method with state-of-the-art techniques

Reference	Mechanism	Image Category	Accuracy (%)
Emrah Irmak (2021)	CNN	X-Ray	95.52
Juan Carlos Quiroz et al. (2021)	Logistic Regression	CT Scan	81.8
Xiao et al. (2020)	ResNet-34	CT Scan	81.9
Carvalho et al. (2020_	Artificial Neural Network	CT Scan	82
Zekuan Yu et al. (2020)	Dense Net-201	CT Scan	95.3
Proposed Method	**Random Forest**	**CT Scan**	**96**

Artificial Intelligence (AI) aided machine learning positioned analysis. A novel severity identification mechanism is proposed and legitimized to classify the Covid-19 containment individuals rendering their asperity situations early vs. progressive vs. advanced. Empirical outcomes advise that the recommended tool can precisely predict disease severity in Covid-19 individuals utilizing Chest CT scan imaging, offering promise for medical prognosis as well as early hospitalization. Various authors have also recently identified the Covid-19 infection enabled by CNN and machine learning models, as shown in Table 8.

5. Conclusion

In some instances, lack of abilities such as inexperience and false staging of radiologists to determine the infection vulnerability can prompt more deceased Covid-19 individuals. This article proposes novel employment of image lung segmentation and machine learning classifier via random forest mechanism to classify Covid-19 individuals into three asperity levels: early

vs. progressive vs. advanced with an accuracy of 96%. A seamless execution was achieved by applying a recognized, significant, and balanced image database compared to various existent techniques, as described in Table 6. The fully robotized system with an end-to-end mechanism without the necessity of physical element eradication. It is accepted that this analysis has an extraordinary potential to ease up the responsibility of the over-burden forefront radiologists and speed up the prognosis, treatment of patients, and consequently facilitate the restraint of the epidemic.

Acknowledgment

The authors are thankful to SRMIST management for the support. One of the authors (DNV) is grateful to SRMIST for the Research assistantship. DNV is grateful to Dr. S. Vinayagam & Dr. V. Dinesh Ram at SRM Medical College Hospital and Research Centre for their expert guidance as well as aid in analyzing the CT scan image database.

Author contributions

DNV – Formulated, co-designed, and performed the simulation work, and co-written the manuscript.

SRSP – Senior author who conceived the idea and co-written the manuscript.

Funding None

Data availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The authors declared that they have no conflicts of interest to this work.

Ethical approval This article does not deal with any ethical problems.

Informed consent We declare that all authors have informed Consent.

References

Amini, N., & Shalbaf, A. (2022). Automatic classification of severity of COVID-19 patients using texture feature and random forest based on computed tomography images. *International Journal of Imaging Systems and Technology*, *32*(1), 102-110.

Amyar, A., Modzelewski, R., Li, H., & Ruan, S. (2020). Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. *Computers in Biology and Medicine*, *126*, 104037.

Ballester-Arnal, R., & Gil-Llario, M. D. (2020). The virus that changed Spain: impact of COVID-19 on people with HIV. *AIDS and Behavior*, *24*(8), 2253-2257.

Covid-19 Dashboard is available for public use at https://news.google.com/covid19/map?hl=enIN&gl=IN&ceid=IN%3Aen&mid=%2Fm%2F03rk0. Accessed November 30, 2021.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., ... & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The lancet*, *395*(10223), 507-513.

Carvalho, A. R. S., Guimarães, A., Werberich, G. M., de Castro, S. N., Pinto, J. S. F., Schmitt, W. R., & Rodrigues, R. S. (2020). COVID-19 chest computed tomography to stratify severity and disease extension by artificial neural network computer-aided diagnosis. *Frontiers in Medicine*, 862.

Cai, W., Liu, T., Xue, X., Luo, G., Wang, X., Shen, Y., ... & Liang, T. (2020). CT quantification and machine-learning models for assessment of disease severity and prognosis of COVID-19 patients. *Academic radiology*, *27*(12), 1665-1678.

Cao, Y., Xu, Z., Feng, J., Jin, C., Han, X., Wu, H., & Shi, H. (2020). Longitudinal assessment of COVID-19 using a deep learning–based quantitative CT pipeline: illustration of two cases. *Radiology: Cardiothoracic Imaging*, *2*(2), e200082.

Fan, G., Yang, Z., Lin, Q., Zhao, S., Yang, L., & He, D. (2021). Decreased case fatality rate of COVID-19 in the second wave: a study in 53 countries or regions. *Transboundary and emerging diseases*, *68*(2), 213-215.

He, K., Zhao, W., Xie, X., Ji, W., Liu, M., Tang, Z., & Shen, D. (2021). Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. *Pattern recognition*, *113*, 107828.

Huang, L., Han, R., Ai, T., Yu, P., Kang, H., Tao, Q., & Xia, L. (2020). Serial quantitative chest CT assessment of COVID-19: a deep learning approach. *Radiology: Cardiothoracic Imaging*, *2*(2), e200075.

Irmak, E. (2021). COVID-19 disease severity assessment using CNN model. *IET image processing*, *15*(8), 1814.

Jin, S., Wang, B., Xu, H., Luo, C., Wei, L., Zhao, W., & Xu, W. (2020). AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks. *MedRxiv*.

Long, S. W., Olsen, R. J., Christensen, P. A., Bernard, D. W., Davis, J. J., Shukla, M., ... & Musser, J. M. (2020). Molecular architecture of early dissemination and massive second wave of the SARS-CoV-2 virus in a major metropolitan area. *MBio*, *11*(6), e02707-20.

Li, Z., Zhong, Z., Li, Y., Zhang, T., Gao, L., Jin, D., & Tang, Y. (2020). From community-acquired pneumonia to COVID-19: a deep learning–based method for quantitative analysis of COVID-19 on thick-section CT scans. *European radiology*, *30*(12), 6828-6837.

Quiroz, J. C., Feng, Y. Z., Cheng, Z. Y., Rezazadegan, D., Chen, P. K., Lin, Q. T.,& Cai, X. R. (2021). Development and validation of a machine learning approach for automated severity assessment of COVID-19 based on clinical and imaging data: retrospective study. *JMIR Medical Informatics*, *9*(2), e24572.

Qi, X., Jiang, Z., Yu, Q., Shao, C., Zhang, H., Yue, H., & Ju, S. (2020). Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: A multicenter study. *MedRxiv*.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., & Shi, Y. (2020). Lung infection quantification of COVID-19 in CT images with deep learning. *arXiv preprint arXiv:2003.04655*.

Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., & Shi, Y. (2021). Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. *Medical physics*, *48*(4), 1633-1645.

Saito, S., Asai, Y., Matsunaga, N., Hayakawa, K., Terada, M., Ohtsu, H., ... & Ohmagari, N. (2021). First and second COVID-19 waves in Japan: a comparison of disease severity and characteristics. *Journal of Infection*, *82*(4), 84-123.

Tang, Z., Zhao, W., Xie, X., Zhong, Z., Shi, F., Ma, T., ... & Shen, D. (2021). Severity assessment of COVID-19 using CT image features and laboratory indices. *Physics in Medicine & Biology*, *66*(3), 035015.

The processed dataset is available for public use at https://www.kaggle.com/abhishek0389/eda-covid-segmentation/data?select=ct_scans, https://github.com/UCSD-AI4H/COVID-CT. Accessed June 2021.

Vinod, D. N., & Prabaharan, S. R. S. (2020). Data science and the role of Artificial Intelligence in achieving the fast diagnosis of Covid-19. Chaos, Solitons & Fractals, 140, 110182.

Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., ... & Ferguson, N. M. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases*, *20*(6), 669-677.

Vinod, D. N., Jeyavadhanam, B. R., Zungeru, A. M., & Prabaharan, S. R. S. (2021). Fully automated unified prognosis of Covid-19 chest X-ray/CT scan images using Deep Covix-Net model. *Computers in biology and medicine*, 104729.

Vahidy, F. S., Drews, A. L., Masud, F. N., Schwartz, R. L., Boom, M. L., & Phillips, R. A. (2020). Characteristics and outcomes of COVID-19 patients during initial peak and resurgence in the Houston metropolitan area. *Jama*, *324*(10), 998-1000.

Wong, H. Y. F., Lam, H. Y. S., Fong, A. H. T., Leung, S. T., Chin, T. W. Y., Lo, C. S. Y., ... & Ng, M. Y. (2020). Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology*, *296*(2), E72-E78.

Xiao, L. S., Li, P., Sun, F., Zhang, Y., Xu, C., Zhu, H., ... & Zhu, H. (2020). Development and validation of a deep learning-based model using computed tomography imaging for predicting disease severity of coronavirus disease 2019. *Frontiers in bioengineering and biotechnology*, 898.

Yu, Z., Li, X., Sun, H., Wang, J., Zhao, T., Chen, H., ... & Xie, Z. (2020). Rapid identification of COVID-19 severity in CT scans through classification of deep features. *BioMedical Engineering OnLine*, *19*(1), 1-13.

Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., & Wang, X. (2020). Deep learning-based detection for COVID-19 from chest CT using weak label. *MedRxiv*.

Zhang, Y., Wu, H., Song, H., Li, X., Suo, S., Yin, Y., & Xu, J. (2020). COVID-19 pneumonia severity grading: test of a trained deep learning model.

Zhu, J., Shen, B., Abbasi, A., Hoshmand-Kochi, M., Li, H., & Duong, T. Q. (2020). Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. *PloS one*, *15*(7), e0236621.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3-11). Springer, Cham.