International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Covid-19 Progression Forecasting and Mortality Rate Analysis for Genome Clinical Characteristics and Chest CT Scan

B. Sandhiya[*1], S. Brindha[2]

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: The COVID-19 sickness has spread over the world as a result of SARS-CoV2 turning into a pandemic. To reduce the strain on healthcare systems, the decision-making model with prediction methodology for clinical data is required. The foremost objective of the proposed work is towards the analysis of COVID-19 disease mortality rate for patients with the help of genomic clinical data and predicting the progression levels of disease spread. Both the datasets are analyzed individually for the COVID and NON COVID patients. In Genomic clinical features, the significant mortality of disease is identified with the help of Chi-square test based on the P-value. The Chest CT scan utilizes 3D U-Net segmentation with a hybrid transfer learning algorithm such as ResNet50 and DenseNet121 along with metaheuristic optimization that reduces time complexity and classifies the disease progression levels of COVID-19. Jupyter Notebook simulations are used to implement the suggested technique. In addition, it is assessed using Accuracy, Sensitivity, Specificity, and Precision. Experimental results make it obvious that the purpose of a proposed prediction model has a 99.07% accuracy rate, a 97.16% Precision rate, a 97.66% Recall rate, and a 97.16% f1 – Score in identifying Progression of COVID-19 affected patients. The accuracy rate depends on the progression stages for No progression, Mild, and Highly Risk are 97.22%, 100%, and 100%. These results imply that medical specialists can benefit from the proposed methodology by using it to forecast COVID-19 disease for experimental prophecy investigation.

Keywords: *CLAHE, 3D U-Net Segmentation, Chi-Square Test, ResNet50 and DenseNet121.*

1. Introduction

On January 30, 2020, SARS-CoV-2 was classified as a worldwide health emergency of international concern by the World Health Organization (WHO). Every province in China has been affected by this pandemic, which has quickly expanded to the rest of the world. Over 158 countries and territories were impacted as of March 16, 2020, the date this article was written [1]. The recent standard intended for diagnosing COVID-19 is an optimistic nucleic acid test result through the technologies based on future generation or reverse transcription polymerase chain reaction (RT-PCR) in real time [2]. But in the context of clinical practice, there are comparatively many false-negative outcomes due to unstable specimen processing, which has made the outbreak worse [3]. Furthermore, not all hospitals have the stringent framework needed for SARS-CoV-2 laboratory testing assembled [4]. Identifying COVID-19 using a testing kit will provide false negative reports so the identification of disease needs a quick and efficient model to overcome these issues.

Certain patients have also been reported to exhibit other imaging characteristics, including consolidation, cavitation, and interlobular septal thickening. These COVID-19 imaging symptoms are nonspecific, nevertheless, and make it challenging to differentiate them from other types of pneumonia. Chest Radiography data are used to diagnose pulmonary-related diseases likes Asthma, Lung Cancer, COVID-19, and pneumonia [5]. The evolutionary model of the system will be based on the two datasets Genomic clinical features and Chest CT scan. As far as we are aware, not much research is directly focused on analysis of Genomic features of individuals with and without COVID-19.

In the earlier juncture, the Ground Glass Opacity is a main feature that is associated with the COVID-19 disease that is presented in the Chest CT data. Automated methods for locating such elusive abnormalities can improve early detection rates with high accuracy and help in diagnosis. The utilization of Deep Learning (DL) and Artificial Intelligence (AI) in problem-solving methodologies holds great promise in effectively tackling the current issues [6] CT scan imaging and genomic clinical data may both yield important biomarkers for COVID-19 illness. Radiography data and clinical features aid in the quick diagnosis of the illness.

Notwithstanding all of its advantages, DL presents difficulties at every stage of development and implementation. The underlying problem is that most health-related systems require real data before they can use synthetic data, and as Artificial Intelligence along with Deep Learning [7-11] designs acquire extra complexity,

[1] *SRM Easwari Engineering College, Chennai – 600069, INDIA*
ORCID ID : 0000-0002-2202-4287
[2] *SRM Easwari Engineering College, Chennai – 600069, INDIA*
ORCID ID : 0009-0007-0750-5182
** Corresponding Author Email: sandhiya5705@gmail.com*

algorithms get more difficult to identify with. Despite the complexity revealed by the existing research and terminations, both AI along with DL algorithms helps in solving disease classification limitations by utilizing image processing procedures [12]. These factors led to the construction of artificial images of the Chest CT scan data for patients with diseases both COVID-19 and non-COVID-19 utilizing a TL (Transfer learning) and Grasshopper optimization in this work. Below is a summary of this paper's main contributions.

- An enhanced technique for feature extraction, filtering, preprocessing and segmentation is developed to advance the competence of the model suggested and for effective disease progression analysis in the medical field.
- Hybrid Transfer Learning along with metaheuristic optimization algorithms are used in the direction of lessens the execution time with to advance the disease prediction efficiency.
- The unique statistical test model effectively signifies the mortality of the genomic clinical feature.
- An improved learning technique is modified to attain greater accuracy, and the methods are recommended when a labeled dataset requires qualified and suitable resources for learning or training.
- Comprehensive ablation tests to validate the different components of the COVID-19 progression forecasting for genomic clinical features and chest CT scan.

The relevant work is described in Section 2 of this paper, and Section 3 proposes a COVID-19 prediction model, whereas in Section 4, the suggested prediction model's outcome and discussion are assessed, respectively. The proposed work is concluded in Section 5, along with the scope for future work.

2. Related Works

This section gives a detailed study related to the existing work and its limitations. In [13] in order to forecast the early stages of the COVID-19 disease the author implemented an ensemble deep transfer learning methodology. By associating with the sensors that are used for medicinal purposes as well as an infected area of the chest X-ray modalities for COVID-19 diseases are taken using a deep ensemble model of a cloud server. In many review articles, COVID-19 disease predictions are diagnosed by physicians depending on the patient's symptoms. This process is considered to be low-performance and time-consuming. To overcome these issues the author [14] designed a model based on CNN and DarkNet which helps in processing the CT scan data and X-ray data. YOLO detection is also implemented to monitor the cases quickly.

Making an allowance for the untimely diagnosis of COVID-19 diseases are very significant, [15] the author proposed a new-fangled deep learning model described as the CovidDWNet representation which is assembled on the formation of depthwise dilated convolutions (DDC) in addition to feature reuse residual block (FRB). The CovidDWNet is evaluated with a gradient algorithm headed to get better precision of the structure. CovidDWNet+GB representation accomplishes the highest accuracy rate compared with other architectures and within a few seconds, it analyzes more than a thousand image datasets. In this paper [16] the author made an analysis using the known eleven Convolutional neural networks (CNN) model designed for categorizing the positive as well as negative cases of COVID-19 patients. Here, all the eleven model's performances are compared and discussed with each other models. To identify the efficient model the author implemented a smaller medical dataset.

2.1 Prediction based on Deep learning

In [17] the author implemented optimized deep learning DenseNet121 methodology that helps in diagnosing COVID-19 based on the gravitational search optimization (GSA) with an accuracy of 98%. This proposed implementation is used to choose the optimal values from the hyperparameter as well as compare them with the CNN architecture. [18] The reading looks at the prediction of COVID-19 disease supported on the chest X-ray dataset at a very early stage with minimal cost. On implementing deep learning algorithms, the COVID-19 datasets of Australia and Jordan achieved an accuracy of 94.8% and 88.43% respectively. For predicting the positive cases along with death cases based on the time series, three forecasting methods are implemented. The study [19] performs a predictive model estimated by a deep learning algorithm utilizing clinical data. The accuracy, precision, F1 score, recall, AUC, and precision of the six distinct deep learning algorithms CNN, LSTM, RNN, CNN LSTM, and CNNRNN are estimated for training purposes. In this instance, CNN LSTM outperforms other deep learning algorithms in terms of accuracy.

The COVID-19 thoracic dataset's features are extracted using a HOG extraction procedure in this paper [20]. With 85% tertiary classification accuracy, a convolutional neural network implements the training and testing dataset. To overcome the classification accuracy issues the author [21] proposed novel deep learning methods with an accuracy of 99.75% for evaluating the dataset for chest X-rays. The data feature is extracted by means of the deep feature extraction method. Afterward, an SVM classification procedure is used to choose the extracted feature using a

strong algorithm known as SDAR (Sequentially Discounting AutoRegressive). When compared to other methods, the classification accuracy is improved by using a Bayesian algorithm to optimize the classified hyperparameters.

To make the proposed work more effective and less complex the author [22] KNN classifiers are used to replace the final Softmax CNN layer. Three commanding evolutionary operators are integrated to expedite the process and maintain a balance between the phases of manipulation and exploration. This operator helps in attaining the optimal values significantly and augments the efficiency of an implemented classification representation. This paper [23] uses a novel CNN architecture for identifying COVID–19 based on two classes Normal and COVID. Based on the dataset for chest X-rays, the pre-trained models like MobileNetv2 and ResNet50 are estimated and compared with the existing algorithms. The proposed work achieves 96.71 % accuracy in predicting diseases.

Article [24] implements a deep convolutional neural network (DCNN) with the aim of mechanically spot the COVID-19 disease based on three categories using the X-ray image COVID-QU-Ex datasets. [25] The author uses an open-source dataset for predicting COVID-19 disease by using approaches based on deep learning methods. This method uses RNN and LSTM algorithms and reaches 98.58% and 93.45% accuracy. The study [26] implementation is based on segmentation workflow and scoring. Based on the latter integrated block the severity score of patients is determined. DeepLabV3+ and MA-Net is the best-performing algorithm which is proposed for lung segmentation and disease segmentation. This proposed work helps to reduce the time efficiency and enhance the performance by using hybrid transfer learning algorithms along with the 3D U-Net segmentation approach.

2.2 Machine Learning based classification

In [27] by using image techniques of pneumonia diagnosis the COVID-19 disease is predicted at a very early stage by the means of utilizing machine learning methods as well as deep learning algorithms. This algorithm helps within classifying the data between pneumonia and normal patients by using Random Forest and XGBoost with 97.3% and 97.7% accuracy. The imbalanced datasets are balanced with the SMOTE (synthetic minority oversampling technique) algorithm. Finally, the classified data is trained using the ResNet 152 network. The study [28] trains the data with both supervised and unsupervised learning in which comparatively the supervised algorithm provides better accuracy. In this work [29] the author introduced a divergent machine learning method that analyzes the open source data of Mexico and Brazil's COVID-19 disease. This data contains the patient's geographical condition, symptoms reports, risk factors, moreover social economic condition to forecast the death as well as recovery for patients. The data is applied to algorithms for machine learning (ML), such as random forests (RF), decision trees (DT), as well as logistic regression to comparatively categorize the data. At the final stage random forest produces good accuracy results in classifying the data.

The paper [30] helps in forecasting the COVID-19 using SVM model with the attribute's location, positive and negative cases, and recovery from diseases. And also monitors the latitude and longitude where they stay from 22 January 2020 to 25 April 2020. This review article [31] uses machine learning (ML) as well as deep learning (DL) representation comparatively for analyzing the COVID-19 disease. Some of the mathematical models have also been used for predicting the disease.

3. Proposed COVID-19 Prediction Methodology

The prediction model works with two different datasets one is a Genomic clinical feature and another is a Chest CT scan. The main intent of the proposed COVID-19 prediction methodology is to classify the progression levels of COVID-19 disease with multiclass outputs and to forecast the mortality of patients with binary classification. The below figure 1 outlines the work of the implementation. This section outlines the methodologies that are used to analyze the datasets are explained as follows.

3.1 Data Collection of Genomic Clinical Features and Chest CT Images

In response to the COVID-19 crisis, the National Cancer Institute (NCI) has expanded the Cancer Imaging Archive (TCIA) to include COVID-19-related data. A compilation of radiographic and CT imaging studies along with the genomic clinical features of patients who tested positive for COVID-19 has been published in TCIA (https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70226443). Based on the Genomic clinical data the mortality rate and the significant manipulating variables are identified. The patient clinical data includes the following parameters Age, Sex, Race, and Latest weight and Height of patients. Along with the genomic features are also noted like Rheumatoid arthritis, systemic lupus erythematosus, extensive burns, asplenia and hyposplenia, tuberculosis, Medication-related illness Chicken pox, Herpes zoster, malnourishment, pregnancy, diabetes type I or diabetes type II, (CKD) chronic kidney disease, transplant, pre- and post-diagnosis dialysis, cancer, ICU Admit and Mortality: all these data were collected and used for analysis process.

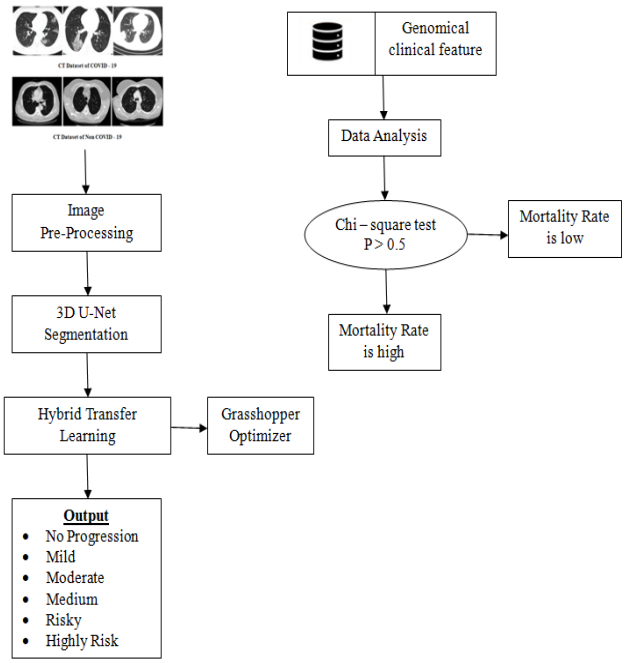
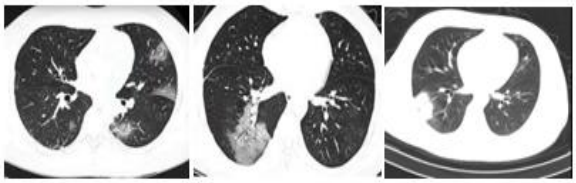
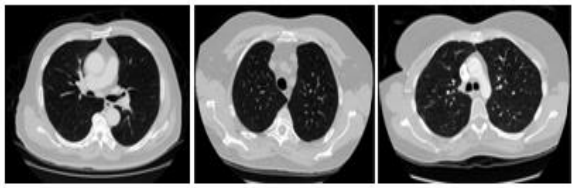


Fig. 1. COVID-19 prediction model (left) proposed model for Chest CT scan and (right) Proposed model for Genomic clinical feature

Secondly, the Chest CT image data of COVID-19 made known within Figure 2 be obtained from TCIA (https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70226443) meanwhile the Non COVID-19 Chest CT images shown in Figure 2 are downloaded from the open-source platform. Depending on both COVID and Non-COVID data the progression of the disease is forecasted with the assistance of hybrid transfer learning algorithms.



CT Dataset of COVID-19



CT Dataset of Non COVID-19

Fig. 2. Sample Image of Chest CT Scan data

3.2 Data Analysis

This section has two subsections that explain the analysis of data that supports genomic clinical data as well as Chest CT scan data. Both genomic clinical data and CT scan data are collected from the same patients with COVID-19 positive and negative cases. The statistical analysis chi-square test uses a conditional probability function that determines the significant condition based on the hypotheses h_0 and h_1.

3.2.1 Genomic Clinical Feature

In genomic clinical data, parameters like primary disease type that affects the respiratory and immune systems have been noted. In this file, a parameter called Mortality is included where this parameter helps in determining the mortal condition of the COVID-19 patients. If the mortal rate is '1' then the patient is considered to be COVID-19 and the patient is under critical conditions otherwise if the mortal rate is 0 then the patient is considered to be Non-mortal.

3.2.2 Chest CT data

The foremost endpoints of the implemented methodology are to discover out the progression levels of the infection spread percentage. [32] Lung CT image of the chest holds the following features Ground glass opacity is a hazy enlarged lung reduction that is not related to bronchial along with the vascular margins. Consolidation is obtained during the air replacement inside the lungs with blood, pus, and fluids. Crazy-paving is a thickening of the lobular septal with the filling of the alveolar. Finally, an air bronchogram Considering Ground glass opacity (GGO) as a primary feature for the proposed model COVID-19 positive and negative cases are determined. The increased GGO inside the lung lobe was estimated based on the degree of involvement with five progressions denoted in Table 1. No involvement is represented by no progression, if the value is ≤5% then it is mild progression, if a value is between 6-25% then moderate progression is determined, if a value is between 26-50% medium progression, if a value is between 51-75% risky progression is determined and lastly if the value is >75% then it is determined as highly risky.

COVID-19 Disease Progression	Infected Percentage
No Progression	0%
Mild Progression	1%-5%
Moderate Progression	6%-25%
Medium Progression	26%-50%
Risky	51%-75%
Highly Risk	>75%

Table 1. Disease Progression with an infected percentage

3.3 Data Preprocessing

Preprocessing is the method of enhancing the data quality that is suitable for the training phase. Every data should undergo a preprocessing phase before training the model. Based on the application, a characteristic of the preprocessing methodology differs [33].

3.3.1 Genomic Clinical Feature

Initially, the genomic clinical feature was cleaned and transformed. In data cleaning the imbalanced .csv file is made into balanced data by replacing the missing value either by 97 or 99. And also checks for the dependency between the rows and columns. Then removes the rows or columns that are independent and doesn't use them in the future. After removing noise and data inconsistency the data is transformed into yes or No representations that are replaced using 1 or 0 numbers respectively.

3.3.2 Chest CT data

The COVID-19 prediction model evaluates image resizing, since the dataset images are different in size each image is to be resized 224*224*224 equally for trimming down the space and computational complexity. Then using Gaussian Filter the noise from the data is removed [34]. Followed by, increasing the image enhancement. CLAHE is an alteration of adaptive histogram equalization (AHE) that be useful for developing the local contrast of the image [35]. In the proposed work usage of CLAHE is categorized into the following steps

3.3.2.1 Tile generation- The default size of a grid is 8*8 based on this size the number of rows along with columns are generated into the form of a tile for Chest CT image data.

3.3.2.2 Histogram Equalization- Initially, for every tile the histogram is calculated as a set of bins. Then, the exceeded clip limit bin values are accumulated for excess calculation followed by the excess distribution where the values are transferred into other bins. Next, in excess redistribution depending on the image's exact value, specific bin values are pushed back to clip limits considering that clip limits are more efficient than specified limits. This process is done recursively until the excess values are insignificant. Finally scaling and mapping are calculated using the cumulative distribution function (CDF) expressed in the below Equ. (1).

$$\frac{Number\ of\ grey\ level}{Number\ of\ pixel\ in\ window}\sum_{k=0}^{n} Histogram\ of\ image\ with\ co-ord(x,y) \quad (1)$$

The calculated CDF values for Chest CT images are plotted and represented with blue color in Figure 3b

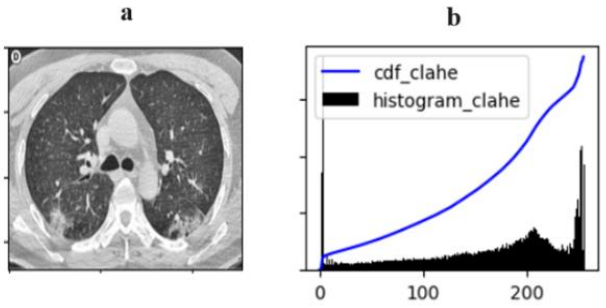


Fig. 3. (a) Image enhanced with CLAHE algorithm and **(b)** Histogram equalization of CLAHE algorithm with CDF

3.3.2.3 Bilinear Interpolation – It is a resampling process for generating new pixel values based on the four nearer pixel value average of distance weighted. In this way, the image contrast is enhanced, and an output of the CLAHE enhancement image is shown in Figure 3 with the CLAHE histogram graph symbolized in red color.

3.4 U-Net segmentation for Chest CT image

U-Net architecture is based on a deep learning (DL) model [36]. Earlier, the U-Net segmentation are been employed for 2D images later 3D U-Net segmentation was employed for 3D data also [37]. The preprocessed image of Chest CT data is taken as the input for segmentation. The prediction representation employs a 3D U-Net segmentation to separate the infected area of the lungs based on the ground glass opacity feature. The shared weight of convolution in the encoder in addition to the 3D U-Net decoder aids in obtaining the intra-slice features of the Chest Ct scan data. Initially, the 3D U-Net learns the GGO volume for training data. Then the infected area feature is separated from the lungs shown in Figure 4. U-Net segmentation has convolutional layers similar to the Convolutional Neural Network (CNN).

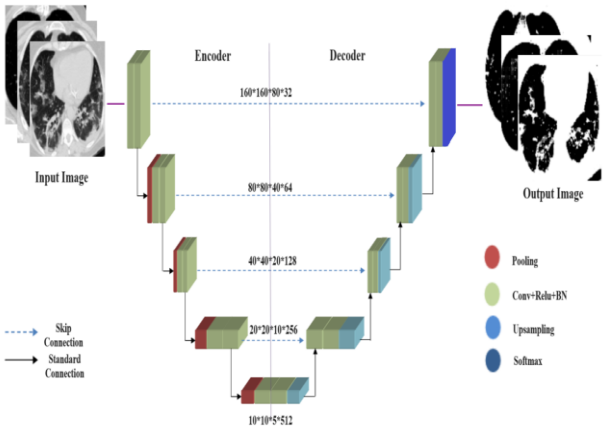


Fig. 4. 3D U-Net Segmentation for Chest CT image

Several kernels in this layer aid in the data's feature extraction. Each CNN neuron's output is determined by its nonlinear activation function, the Rectified Linear Unit action function (ReLU). The feature maps are down-

sampled using pooling and concatenated with the relevant features in the convolutional layer. Additionally, it aids in lowering the number of parameters CNN uses. A valued feature and an available average feature are typically provided by max-pooling and average pooling. Two convolutional layers of size 3x3x3 settle the analysis phase of the 3D U-Net of each layer, which is succeeded by the max-pooling of 2x2x2 with two strides and the ReLu activation.

The synthesis path includes a concatenation layer and transposed convolutions. The transposed convolutional layer, a type of deconvolutional layer, yields a dimension with the same size. To produce additional features for segmentation, the decoder concatenates the feature that convolved from the inferior layer with feature maps of even resolution. A concatenation layer reduces the number of output channels to the number of labels in the final layer. The voxels of the proposed input image are 128x128x32 and similarly, the voxels of a segmented output image are 128x128x32 in tri-directional respectively. In this proposed work the skip connections and standard connection are denoted. The skip connection has a sequential gradient flow starting from the layer to the finishing layer. Mean the standard connection used in denoting the down flow and up flow of the layers. This connection helps in enhancing the performance of the algorithm. To overcome distraction and trapped issues the randomized training set is shuffled to update and make the coefficients. The accuracy for the initial 25 epochs of 3D U-Net segmentation is symbolized in Graph 1a.

Pseudocode 1

FUNCTION: Calculating the percentage of COVID-19 Virus spread in the lungs

INPUT: j=0, k=0

for i=0 to enumerate(infection spread)

percentage = infect * 100

if maximum value (prediction class =1.0 and j<1) then

return COVID-19 positive

return percentage

else

if maximum value (prediction class = 0 and k<1) then

return COVID-19 negative

return percentage

end if

end for

Pseudocode 2

FUNCTION: Estimating COVID-19 Infection spread

def infection spread():

repeat n times:

target_infection = select_target_infection()

if no target_infection then

return

infect_code(target_infection)

return

The 3D U-Net segmentation helps in segmenting the COVID-19 spread of infected areas as of the Chest CT scan. Post segmentation the infected area percentage is calculated based on the above Pseudocode 1 and Pseudocode 2. The progression levels of the disease are determined depends upon the percentage of the infection spread. The stages of the progression according to the infection spread percentage are represented in Table 1 which is described in the Data analysis section 3.2.

3.5 Hybrid Transfer Learning

In the proposed methodology, a Hybrid Transfer Learning algorithm is implemented along with the metaheuristic optimization algorithm. This proposed methodology along with optimization techniques helps in choosing exact classification results. The architecture of our model is shown in Figure 5. DenseNet121 and ResNet50 are chosen in the parallel network to share their respective advantages: both are important for learning good representations. The DenseNet121 and ResNet50 modules were utilized to combine their taken features. To achieve this, we took each module's middle layers and eliminated its final fully-connected (FC) layer. Next, the feature maps created by these two modules with the number of channels changed using a convolutional layer to make sure they were the same size. The self-attention module to the model to boost our network's capacity for representing the COVID-19 virus's spread in the lungs and bypass the sample similarity constraint, as the SARS virus contains a significant quantity of discriminative information about the spread of diseases.

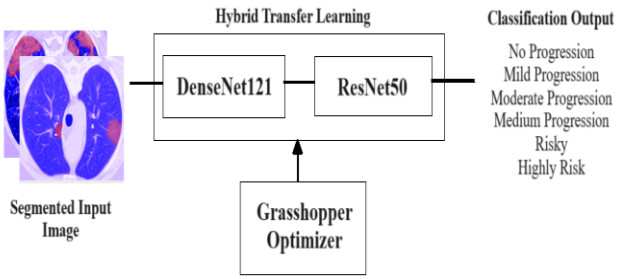


Fig. 5. Outline of Hybrid Transfer Learning with Metaheuristic Optimizer

Grasshopper block in Figure 5 aids in selecting the best optimal solution for our multi-classification problem. The social structure and hunting strategy of grasshoppers in the wild are modeled by the Grasshopper Optimization Algorithm (GOA) algorithm [38]. Each grasshopper in the population represents a solution in this population-based algorithm. It's easy; all we have to do is figure out three forces to determine the grasshopper's position denoted in Equ. 2. The solution is subject to three forces: gravity, wind advection, and social interaction with other grasshoppers.

$$Position\ of\ Grasshopper_i = soc\_inter_i + grav_i + wind\_adv_i \quad (2)$$

Lastly, to complete our Covid-19 disease progression classification task, we successively stacked two FC layers. The dropout technique was applied prior to the final FC layer in order to reduce the over fitting of the model. Although CNNs are very useful in many applications, metaheuristic optimization techniques are used to increase the effectiveness and accuracy of CNNs. In order to concentrate on the concern of limited data in the target area, Transfer Learning (TL) settle down the requirement to facilitate the training as well as test data be independent also identically distributed. This enables it to apply to a specific domain task the knowledge acquired from a related domain. The successful application of TL within biomedicine by some recent studies [39] encouraged us to apply TL to Covid-19 data prediction as well.

Where, the high level feature extracted are strongly dependent on the task and dataset chosen, similarly CNNs extract standard low-level features that are not dependent on the dataset they are used on. Nevertheless, ResNet50 [40] and DenseNet121 [41], which were pre-trained on a dataset, have acquired sufficient knowledge of low-level features like color, geometry, and texture. Moreover, characteristics similar to these can be found in Chest CT data. Additionally, taking this into account, we implanted the middle layer parameters both of these previously trained models into our model, enabling our network could focus more intently on acquiring high-level features from CT images in order to improve performance on our classification task. The grasshopper is applied on the CNN's algorithms during the feature extraction and classification process so that the extracts result of the COVID-19 would be chosen with minimal time and with better accuracy.

3.6 Evaluation platform and parameters

Performance analysis estimation is provided using the constraints like accuracy, sensitivity, precision, and specificity defined in the Equ. (3), (4), (5), and (6) (https://www.analyticsvidhya.com/blog/2021/06/classification-problem-relation-between-sensitivity-specificity-and-accuracy/). True Positive (TP): The infection's progression has been appropriately determined. (TN) True Negative: The output of non-progression is correctly predicted (FP) False Positive: An incorrect prediction of non-progression output. (FN) False Negative: An incorrect diagnosis of the infection's progression

$$Accuracy = \frac{TN+TP}{(TN+TP+FN+FP)} \quad (3)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (4)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (5)$$

$$F1-Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (6)$$

4. Result and Discussion

4.1 Descriptive statistics

The mortality rate for every feature of clinical data is estimated based on the conditional probability represented in Table 1 [42]. Here, to determine the significance rate of the mortality cases, the Chi-squared test was utilized. The chi-squared test helps in deciding the existence and non-existence of the primary diseases. To satisfy the significant condition the 'P' values of primary disease are estimated. If, the P value is greater than 0.05, it indicates a high patient mortality rate. If, P values < 0.05 then the mortality rate of the patient is low or no mortality. For each parameter, the mean, standard deviation, and P values are estimated and denoted in Table 2.

Number of Sample N=100								
GENOMIC CLINICAL FEATURES	Mortality Rate High				No Mortality			
	N	Mean	Standard deviation	P value	N	Mean	Standard deviation	P value
RACE	32	1.4	0.502625	0.730892	68	3.2	0.712561	<0.05
OBESITY	46	1.9	0.307794	0.888491	54	3.6	0.507894	<0.05
TUBRCULOSIS	22	1.85	0.366348	0.440078	78	3.4	0.521689	<0.05
SYSTEMIC_LUPUS_ERYTHMATOSUS	19	1.85	0.366348	0.440078	81	3.4	0.521689	<0.05
RHEUMATOID_ARTHRITIS	21	1.75	0.444262	0.214659	79	3.3	0.689781	<0.05
EXTENSIVE_BURNS	1	1.9	0.307794	0.888491	99	3.1	0.456978	<0.05
ASPLENIA_&_HYPOSPLENIA	3	1.95	0.223607	0.280590	97	3.12	0.689455	<0.05
MEASLES	18	1.9	0.307794	0.109868	82	3.18	0.698451	<0.05
CYTOMEGALOVIRUS	16	2.29	0.451326	0.440078	84	2.74	0.678412	<0.05
CHICKEN_POX	32	2.26	0.478951	0.462946	68	2.71	0.568974	<0.05
HERPES_ZOSTER	15	1.76	0.307982	0.833359	85	3.2	0.578941	<0.05
MALNUTRITION	2	1.048	0.250264	0.440078	98	4.19	0.456872	<0.05
CURRENT_PREGNANT	5	2.97	0.498712	0.157674	95	2.1	0.689781	<0.05
CHRONIC_KIDNEY_DISEASE	11	1.95	0.223607	0.380004	89	3.19	0.689710	<0.05
DIABETIES	41	2.186	0.524241	0.462946	59	2.91	0.789221	<0.05
HEMODIALYSIS	4	2.298	0.517544	0.280590	96	3.48	0.789532	<0.05
CANCER	5	1.9	0.307794	0.888491	95	3.98	0.578933	<0.05
ICU_ADMIT	6	32.7	17.12598	0.088750	94	63.7	46.55939	<0.05

Table 2. Descriptive statistics of genomic clinical feature with N=100

$$Mean = \frac{A\ sum\ of\ the\ Sample\ Data}{Total\ number\ of\ Sample\ Data} \quad (7)$$

$$Standard\ Deviation = \sqrt{\frac{\Sigma(Sample\ Data\ Value - Mean)^2}{Number\ of\ Sample\ Data}} \quad (8)$$

The mean and standard deviation formulas, found in Equ. (7) And (8) are used to evaluate the descriptive statistics values for genomic clinical features. The predictive model built for the genomic clinical feature is implemented by using the Jupyter Notebook platform.

4.2 Performance analysis of Chest CT data

The effectiveness of statistical analysis is evaluated for finding the disease spread for patients with COVID-19 infection by the means of performance metrics. The Chest CT scan images are used to train as well as test the proposed Hybrid transfer learning with grasshopper Optimizer in conjunction with a standard pre-trained model ResNet 50 along with DenseNet121 for the advancement of COVID-19 prediction. By estimating the confusion matrix values the f1-score, accuracy, precision, and recall are determined for the classification results. The analysis of performance is evaluated for the proposed model and also for the DenseNet121 and ResNet50 without optimization represented in the Table 3 and 4.

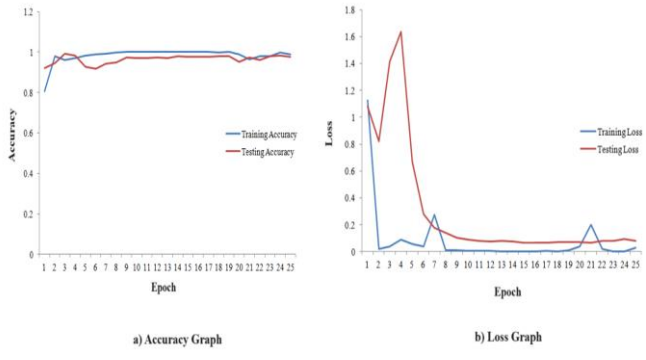
	Accuracy	Precision	Recall	F1-Score
No Progression	100%	1	1	1
Mild	100%	1	1	1
Moderate	97.22%	1	0.86	0.92
Medium	100%	1	1	1
Risky	97.22%	0.83	1	0.91
Highly Risk	100%	1	1	1

Table 3. Performance metric tables of proposed methodology

	Accuracy	Precision	Recall	F1- Score
No Progression	94.21%	0.83	0.83	0.83
Mild	100%	1.0	1.0	1.0
Moderate	94.12%	0.83	0.83	0.83
Medium	94.12%	0.83	0.83	0.83
Risky	94.12%	0.75	0.75	0.75
Highly Risk	100%	1.0	1.0	1.0

Table 4. Performance analysis table of ResNet50 and DenseNet121 without Optimizer

In Graph 1b it is represented that there is a loss in the curve during the initial stages of training data and vice versa when the training ends. The positive COVID-19 cases are low when compared with NON COVID-19 cases. The accuracy result during the initial epochs may be low, as shown in Graph 1a, but it may be possible to reduce the loss function from high to low point by looking at all the chest CT images at each epoch. The sample result of the disease progression with infected percentage is shown in below figure 6. Here, the first column shows the categories of COVID-19 progression in the lung followed by the second column that displays the original Chest CT scan image. In third column displays the Image preprocessed result followed by the 3D U-net segmented. The lung segmentation in this column is indicated by the blue color, and the COVID-19 disease's life-threatening infection spread and progression percentage are indicated by the red color.



Graph 1. Accuracy and Loss curves 3D U-Net segmentation

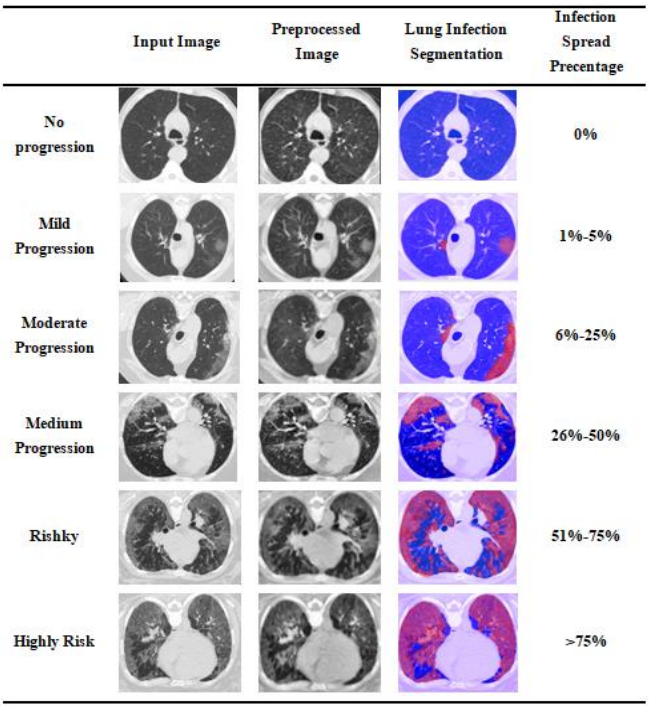
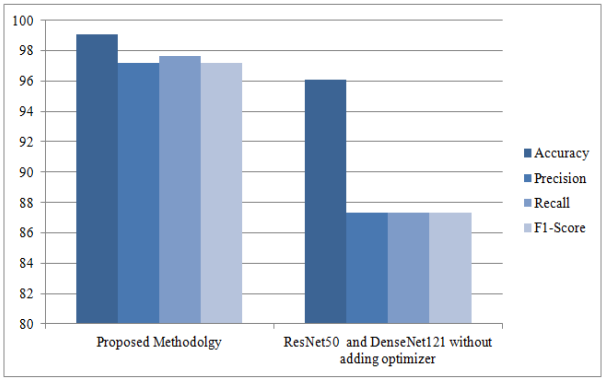


Fig. 6. Sample output of COVID-19 Disease progression and Infection spread percentage.

The optimization technique combined with hybrid transfer learning is demonstrated to classify the disease progression in a Chest CT scan. The suggested work achieves the mean accuracy of 99.07%, 97.16% precision, 97.66% recall, and 97.16% F1-Score. The performance analysis performed with the dataset reveals the robustness of the suggested algorithm. Our suggested works provide better accuracy and less time complexity when compared to the ResNet and DenseNet model without using optimizer shown in graph 2.



Graph 2. Performance analysis of proposed model and existing model without adding optimizer

5. Conclusion

Due to the emergency need that arose to battle against the deadly virus named COVID-19. The proposed work invented a robust way of segmenting the disease and predicting its progression percentage for Chest CT scan data. Meanwhile, for Genomic clinical features, the statistical analysis method is used for categorizing illnesses that are not COVID-19 and those that are COVID-19. The Chi-square test is applied on the road to classify the genomic clinical data based on the p-values obtained. Next for the Chest CT scan, the data is segmented using a 3D U-Net algorithm then the optimizer namely grasshopper with pre-trained hybrid transfer learning is used as a part of the implementation work. The CT data is tested and classified into six progression stages No progression, mild, moderate, medium, Risky, and highly risk progression based on the percentage of the infection spread. By using 3D U-Net segmentation along with hybrid TL combined with a metaheuristic optimization algorithm the performance of the parameter is achieved in a higher range. The ResNet and DenseNet without optimizer model is used for the comparison. The optimization algorithm with the hybrid Pre-Trained model surpasses other existing algorithms. The proposed model achieves an accuracy of 99.07%. Future work will be based on associating the features of the Genomic clinical data along with Chest CT scan features for determining the patient's survival rate.

6. References and Footnotes

Acknowledgements

The Indian Council of Medical Research (ICMR) provided funding for this project with RFC Number BMI/Adhoc/23/2022-23

Author contributions

B Sandhiya1: Conceptualization, Investigation, Data curation and Validation, **S Brindha2:** Data curation, Methodology, Writing-Original draft preparation, Software, Validation, Visualization, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Organization WH (2020) Novel coronavirus (2019-nCoV) situation reports. Available via https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/. Accessed 16 Mar 2020

[2] Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497-506

[3] Chung M, Bernheim A, Mei X et al (2020) CT imaging features of 2019 novel coronavirus (2019-nCoV).
Radiology. https://doi.org/10.1148/radiol.2020200230

[4] Lu H, Rui H, Pengxin Y, Shaokang W, Liming X (2020) A correlation study of CT and clinical features of different clinical types of 2019 novel coronavirus pneumonia. Chin J Radiol 54:E003–E003

[5] Churruca, M, Martínez-Besteiro, E, Couñago, F, and Landete, P (2021). COVID-19 pneumonia: a review of typical radiological characteristics. *World J Radiol* 13:327–43. doi: 10.4329/wjr.v13.i10.327

[6] Wang, S, Kang, B, Ma, J, Zeng, X, Xiao, M, Guo, J, et al. (2021). A deep learning algorithm using CT images to screen for Coronavirus disease (COVID-19). *Eur Radiol* 31:6096–104. doi: 10.1007/s00330-021-07715-1

[7] Sandhiya, B., and S. Kanaga Suba Raja. "Deep Learning and Optimized Learning Machine for Brain Tumor Classification." Biomedical Signal Processing and Control 89 (2024): 105778.

[8] Sandhiya, B., S. Kanaga Suba Raja, "Brain tumour segmentation and classification with reconstructed MRI using DCGAN." Biomedical Signal Processing and Control 92 (2024).

[9] Sandhiya, B., and S. Kanaga Suba Raja. "Multi-classification of Brain Tumor MRI Images Using Deep CNN Features and ML Classifiers." In International Virtual Conference on Industry, pp. 139-151. Singapore: Springer Nature Singapore, 2021.

[10] Sandhiya, B., and S. Kanaga Suba Raja. "Deep Learning Framework for Brain Tumor Classification through Histopathological Image." In 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), pp. 1-6. IEEE, 2022. (Scopus)

[11] Sandhiya, B., R. Priyatharshini, B. Ramya, S. Monish, and Gopal R. Sai Raja. "Reconstruction, identification and classification of brain tumor using gan and faster regional-CNN." In 2021 3rd international conference on signal processing and communication (ICPSC), pp. 238-242. IEEE, 2021. (Scopus)

[12] Manapure, P, Likhar, K, and Kosare, H (2020). Detecting COVID-19 in X-ray images with keras, tensor flow, and deep learning. *Assessment*, Vol. 2.

[13] Yeruva, Ajay Reddy, et al. "Covid-19 Disease Detection using Chest X-Ray Images by Means of CNN." 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS). IEEE, 2022.

[14] Al Shehri, Waleed, et al. "A Novel COVID-19 Detection Technique Using Deep Learning Based Approaches." Sustainability 14.19 (2022): 12222.

[15] Celik, Gaffari. "Detection of Covid-19 and other pneumonia cases from CT and X-ray chest images using deep learning based on feature reuse residual block and depthwise dilated convolutions neural network." Applied Soft Computing 133 (2023): 109906.

[16] Yang, Yuan, et al. "A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions." Computers in Biology and Medicine 139 (2021): 104887.

[17] Ezzat, Dalia, Aboul Ella Hassanien, and Hassan Aboul Ella. "An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization." Applied Soft Computing 98 (2021): 106742.

[18] Alazab, Moutaz, et al. "COVID-19 prediction and detection using deep learning." International Journal of Computer Information Systems and Industrial Management Applications 12.June (2020): 168-181.

[19] Alakus, Talha Burak, and Ibrahim Turkoglu. "Comparison of deep learning approaches to predict COVID-19 infection." Chaos, Solitons & Fractals 140 (2020): 110120.

[20] Chen, Joy Iong-Zong. "Design of accurate classification of COVID-19 disease in X-ray images using deep learning approach." Journal of ISMAC 3.02 (2021): 132-148.

[21] Demir, Fatih, Kürşat Demir, and Abdulkadir Şengür. "DeepCov19Net: automated COVID-19 disease detection with a robust and effective technique deep learning approach." New Generation Computing 40.4 (2022): 1053-1075.

[22] Jalali, Seyed Mohammad Jafar, et al. "X-ray image based COVID-19 detection using evolutionary deep learning approach." Expert Systems with Applications 201 (2022): 116942.

[23] Sahin, M. Emin. "Deep learning-based approach for detecting COVID-19 in chest X-rays." Biomedical Signal Processing and Control 78 (2022): 103977.

[24] Sanida, Theodora, et al. "A Robust Hybrid Deep Convolutional Neural Network for COVID-19 Disease Identification from Chest X-ray Images." Information 14.6 (2023): 310.

[25] Alassafi, Madini O., Mutasem Jarrah, and Reem Alotaibi. "Time series predicting of COVID-19 based on deep learning." Neurocomputing 468 (2022): 335-344.

[26] Danilov, Viacheslav V., et al. "Automatic scoring of COVID-19 severity in X-ray imaging based on a novel

deep learning workflow." Scientific reports 12.1 (2022): 12791.

[27] Kumar, Rahul, et al. "Accurate prediction of COVID-19 using chest X-ray images through deep feature learning model with SMOTE and machine learning classifiers." MedRxiv (2020): 2020-04.

[28] Kwekha-Rashid, Ameer Sardar, Heamn N. Abduljabbar, and Bilal Alhayani. "Coronavirus disease (COVID-19) cases analysis using machine-learning applications." Applied Nanoscience 13.3 (2023): 2013-2025.

[29] Iwendi, Celestine, et al. "COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients." Journal of Experimental & Theoretical Artificial Intelligence (2022): 1-21.

[30] Singh, Vijander, et al. "Prediction of COVID-19 corona virus pandemic based on time series data using Support Vector Machine." Journal of Discrete Mathematical Sciences and Cryptography 23.8 (2020): 1583-1597.

[31] Saleem, Farrukh, et al. "Machine learning, deep learning, and mathematical models to analyze forecasting and epidemiology of COVID-19: A systematic literature review." International journal of environmental research and public health 19.9 (2022): 5099.

[32] Koo, Hyun Jung, et al. "Radiographic and CT features of viral pneumonia." Radiographics 38.3 (2018): 719-739.

[33] Rahman, Tawsifur, et al. "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images." Computers in biology and medicine 132 (2021): 104319.

[34] D'Haeyer, Johannes PF. "Gaussian filtering of images: A regularization approach." Signal processing 18.2 (1989): 169-181.

[35] Zuiderveld, Karel. "Contrast limited adaptive histogram equalization." Graphics gems (1994): 474-485.

[36] Du, Getao, et al. "Medical image segmentation based on u-net: A review." Journal of Imaging Science and Technology (2020).

[37] Çiçek, Özgün, et al. "3D U-Net: learning dense volumetric segmentation from sparse annotation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Springer International Publishing, 2016.

[38] Saremi, Shahrzad, Seyedali Mirjalili, and Andrew Lewis. "Grasshopper optimisation algorithm: theory and application." Advances in engineering software 105 (2017): 30-47.

[39] Pio G, Mignone P, Magazzù G, Zampieri G, Ceci M, Angione C. Integrating genome-scale metabolic modelling and transfer learning for human gene regulatory network reconstruction. Bioinformatics. 2022;38(2):487–93.

[40] Mukti, Ishrat Zahan, and Dipayan Biswas. "Transfer learning based plant diseases detection using ResNet50." 2019 4th International conference on electrical information and communication technology (EICT). IEEE, 2019.

[41] Nandhini, S., and K. Ashokkumar. "An automatic plant leaf disease identification using DenseNet-121 architecture with a mutation-based henry gas solubility optimization algorithm." Neural Computing and Applications (2022): 1-22.

[42] Kwak, Sanggyu. "Are Only p-Values Less Than 0.05 Significant? A p-value Greater Than 0.05 Is Also Significant!." Journal of Lipid and Atherosclerosis 12.2 (2023): 89.