Classification of COVID-19 using Deep Learning and Radiomic Texture Features extracted from CT scans of Patients Lungs

Jayalakshmi Mangalagiri[1*], Jones Sam Sugumar, Sumeet Menon[1], David Chapman[1], Yaacov Yesha[1], Aryya Gangopadhyay[1], Yelena Yesha[1,3], Phuong Nguyen[1,2*]

[1]*University of Maryland, Baltimore County*, Baltimore, MD, USA, [2]*OpenKneck Inc,* Halethorpe, MD, USA,
[3]*University of Miami,* Miami, FL, USA

**Corresponding author: jmangal1@umbc.edu, phuong3@umbc.edu*

2021 IEEE International Conference on Big Data (Big Data) | 978-1-6654-3902-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/BigData52589.2021.9671656

***Abstract*—COVID-19 is an air-borne viral infection, which infects the respiratory system in the human body, and it became a global pandemic in early March 2020. The damage caused by the COVID-19 disease in a human lung region can be identified using Computed Tomography (CT) scans. We present a novel approach in classifying COVID-19 infection and normal patients using a Random Forest (RF) model to train on a combination of Deep Learning (DL) features and Radiomic texture features extracted from CT scans of patient's lungs. We developed and trained DL models using CNN architectures for extracting DL features. The Radiomic texture features are calculated using CT scans and its associated infection masks. In this work, we claim that the RFs classification using the DL features in conjunction with Radiomic texture features enhances prediction performance. The experiment results show that our proposed models achieve a higher True Positive rate with the average Area Under the Receiver Curve (AUC) of 0.9768, 95% Confidence Interval (CI) [0.9757, 0.9780].**

Keywords— COVID-19, CT-Scans, Deep learning & Radiomic Feature extraction, classification, AUC-ROC.

I. INTRODUCTION

The novel Coronavirus alias COVID-19 has led to a global pandemic which has been changing the lives of the people drastically all over the world [1, 2]. Even though symptom tracking helps us to understand the disease, for diagnosing COVID-19 accurately, people must undergo an X-ray and/or CT scan which provides detailed images of the lungs [16, 17, 2, 18]. In addition, CT scans of the lung has been shown [1, 2, 3] to provide a significant adjunctive role in diagnosing, tracking progress and complication of COVID-19 as compared to other methods such as monitoring of temperature/respiratory symptoms and the current "gold standard", molecular testing, using sputum or nasopharyngeal swabs. More recently 2D/3D CT images have been used to detect COVID19 using Convolutional Neural Network (CNN) [1, 2, 10, 22, 23, 24]. The CNNs with 3D kernels have been shown to be more effective than 2D CNNs in extracting spatiotemporal features [18, 19]. Radiomic features extracted from CT scans have also proven to be effective in identifying the subtle disease characteristics which failed to be detected by human visualization [3, 5]. In the past, Radiomic texture features were identified based on the greater contribution value in differentiating indolent adenocarcinoma and Intralobular Adenocarcinoma [3]. The Radiomic features are used by the Radiologists to quantitatively analyze the presence and severity of abnormalities in the COVID-19 infected lungs such as ground glass opacities (GGO), Consolidation and Crazy-paving patterns. Recent study used feature selection methods for selecting Radiomic features and then used it for classification of COVID-19 infected and other pneumonia cases [21]. Authors used a segmentation algorithm to create COVID-19 infection masks for Radiomic feature extraction and achieved the AUC of 0.922 [21]. To detect COVID-19 infection, Radiomic features measured from the new segmentation methods were analyzed by using sophisticated statistical models with high interpretability [26]. The efficacy of Radiomics in diagnosing patients with COVID-19 and other types of viral pneumonia in addition to clinical symptoms and CT signs were studied in the past works [27] where the latest version of PyRadiomics was used for feature extraction.

In this work we propose to use Random Forest Classifier using the combination of extracted Radiomic texture features and Deep Learning features to classify COVID-19 virus infected patients from healthy cases. The texture features known as second-order statistical features are showing the interrelationship between voxels with similar or dissimilar grey level values which provide the measure of intra nodule heterogeneity. Analyzing Radiomic texture features from the segmented pulmonary parenchyma regions can assist radiologists in diagnosing COVID-19 [26]. Radiomic texture features extracted from the COVID-19 CT scans build the basic blocks towards our classification of COVID-19. The novelty of our work lies in the use of a combination of a set of DL features and Radiomic texture features to classify COVID-19 cases using RF to achieve high True Positive rate. Although many COVID-19 classifiers have been developed either using DL or Radiomics, to the best of our knowledge our study is the first to explore the combination of DL features with Radiomics texture features.

We will show the experimental analysis of the impact of both DL features and our selected Radiomic texture features

computed from the area of lung damage (COVID-19 segmentation masks) caused by this COVID-19 in identifying the COVID-19 cases. We will evaluate the performance of our 2D/3D integration of both DL and Radiomic texture feature classification models to classify COVID-19 infections using t-test, Area Under the Receiver (AUC) Operating Characteristics Curve (ROC) measures, True Positive Rate and False Positive Rate. We also conducted experiments using 2D images and compared them with that of the 3D images from the same number of patient's CT scans of lungs.

The major contribution in our research is as follows:

1. Extraction and selection of Radiomic texture features using 3D COVID-19 CT scans and associated COVID-19 infected segmentation mask data.
2. Developed a highly predictive 3D DL model for classification of COVID-19 infected patients using CT scans.
3. Classification of COVID-19 infected patients using both Extracted Radiomic texture features and DL features.

II. Background and Related Work

What are Radiomic Features?

There are two feature variations in CT scans of the patient's lungs. The features are either quantitative or qualitative in which radiologists review the pathophysiology of the lung nodules or damages. To extract the quantitative features, mathematical and data characterization techniques have been used. This process is called Radiomics, and the extracted quantitative features are the Radiomic features [5]. Radiomic features primarily embrace texture, shape, and gray level statistics of lung nodules. There are four groups of Radiomic features: 1) Shape/volumetric features, 2) First-order statistics features, 3)Second-order statistics features, 4) Transformed features: wavelet or Laplacian-of Gaussian filtration applied in 1) and 3) groups [6]. Radiomic texture features:(belong to the 3 group) contain gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), neighboring gray size zone matrix (NGSZDM), and gray-level dependence matrix (GLDM). Parekh et al. [6] defined "Radiomics is the high throughput extraction of quantitative features from radiological images creating a high dimensional data set followed by data mining for potentially improved decision support."

What are the Deep learning Features?

Deep Learning (DL) is seen as a hierarchical feature learning where DL models perform automatic feature extraction from raw data called feature learning [15]. DL representations are hierarchical and are of three feature types: low-level features, mid-level features, high-level features. These DL features extracted from several layers of the DL model are fed into the trainable classifier for classification purposes [15].

In recent years, the widely accessible large datasets and computationally powerful machines have paved the way for Deep convolutional neural networks to transcend the performance of many conventional computer vision algorithms [7]. Approaches depending on Deep convolutional neural networks (CNN) are very promising when the CT scans are available from large cohorts [8]. When compared to Radiomics approaches, Deep CNN methods usually need huge amounts of training data. After training, the CNN models can become very efficient in making predictions as they can be made directly from the image without having any need to extract the quantitative features before performing the classification.

Related work

In [10], the authors identify common and severe COVID-19 using Radiomic features. We also compute 149 features per patient from 3D CT scans of 38 COVID-19 patients then identify the top related Radiomic texture features for classification.

On the other hand, there is previous work that speckle-like textures perform classification of COVID-19 using Support Vector Machine (SVM) from a combination of X-ray and CT images of 126 patients using shrunken features [1]. Robust features were extracted using 4 different feature extraction techniques: Gray Level Co-occurrence Matrix, Local Binary Gray Level Co-occurrence Matrix, Gray Level Run Length Matrix, and Segmentation Based Fractal Texture Analysis feature extraction algorithms.

One of the past works demonstrated by Belkhatir et al. [11] uses a representative spatial texture feature and employs supervised image classification algorithms like SVM that helps to diagnose COVID-19 using 349 CT images that are reported COVID-19 positive and 397 CT images that are reported negative. Gray Level Co-occurrence Matrix textures were extracted using PyRadiomics for classifying images accordingly.

The authors in [21] extracted 108 features which include features in all four groups. We focus on using 7 Radiomic texture features. We have 3 texture features overlapping with the 108 radiomic features used in [21]. We combined these 7 Radiomic features with 64 extracted DL features as input for Random Forest classifiers. While authors in [21] use 108 radiomic features as input for ML classifiers.

Furthermore, there has also been relevant work shown by Mehta et al. where a 3D DL model was created for 3D image data and compared against the RF classification using a combination of volumetric Radiomic features, biomarkers, and DL features for lung cancer classification [5]. In Fact, the authors in [5] used only one radiomic feature i.e. the volumetric

radiomic features whereas, we have made use of seven radiomic texture features from the second order statistical features.

Both the selection of Radiomic features for analysis of COVID-19 and the combination Radiomic and DL features for classification have not been fully exploited yet to our best knowledge. Except, radiomic features extracted from CT scans have been combined with DL scores to classify COVID-19 critical cases from severe cases using Regression models [25]. Authors demonstrate that the Regression model achieves better performance with the highest testing AUC of 0.861 and better sensitivity when using both radiomic features and DL scores (merged model) [25]. This work in [25] by Li et al. is the existing research which is closely associated with our work.

However, these Radiomic texture features have been used to build a Machine Learning predictive model which performs better than radiologists in diagnosing non-small cell lung cancer [3]. In addition, using only the Radiomic texture features have shown to be statistically significant in differential diagnosis of COVID-19 and quantitatively informative in diagnosis of other lung diseases [26, 28, 29]. In addition, we use a set of high-level DL features while Li et al. just uses a single DF score [25].

In summary, from the previous works [5][25] we see that very limited work has been done on Radiomic features and its combination with deep learning. Thus, it is worthwhile to explore further down this area.

III. METHODOLOGY

In the methodology, as shown in Fig. 1, towards our first step, we have preprocessed the 3D data of COVID-19 CT scans which is collected from a public repository [19]. In Fig. 2, the dataset sample is shown.

After this, we extracted the Radiomic texture features from the 3D COVID-19 data using CT images and its segmented masks. In addition to this, we have also trained and extracted the DL features from our developed 3D DL CNN model (Fig. 3). Using the radiomic texture features and deep learning features as shown in Fig. 1, the RF model that is used to classify the patients into COVID-19 infection and non-COVID-19 classes.

A. Dataset Description

We have made use of COVID-19 public dataset consisting of a total of 38 patient's data which contains the CT-scans, its associated infection and lung masks [19]. The proportion of the COVID-19 infections in the CT scans of patients' lungs range from minor to moderately severe [0.01% to 59%] [30]. The COVID-19 masks were manually annotated by radiologists. Each patient's CT scans have resolution 512x512xZ dimension. We can call each patient's CT scans as 3D images or 3D volume. Where the Z is the number of CT slices ranges from 100-400 slices per patient with the slice thickness varying from 3.5mm to 7.5mm. The values in each CT scan represent radiodensity in the Hounsfield unit then were adjusted to the lung window [-1250,250] before usage.

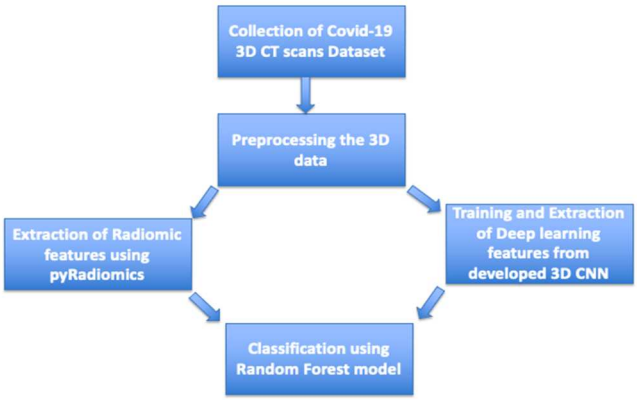

Fig. 1 Workflow diagram

In total, we have considered 38 patient's CT scans with 38 3D images, 3888 2D images to perform classifications. We conducted experiments using 3888 numbers of 2D images in which there are totally 300+ COVID-19 infections (COVID-19 infection masks) in 1800+ images.

The number of patients for 3D COVID positive data is 19, and 19 patients' data are negative COVID data. Out of which, 30 patients were used for training, and 8 patients for testing COVID-19 cases (5 positive and 3 negative cases). In Fig. 2, the dataset sample is shown.

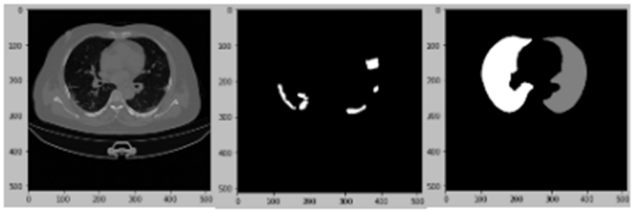

Fig .2 Lung image(left), covid-19 mask(middle), Lung mask for covid19 patient (right)

B. Deep Learning and Radiomic texture feature extraction

Our work made use of two types of features: (a). DL features, (b) Radiomic texture features.

(a) Deep Learning feature extraction

As step 1, we first developed our customized 10 layers 3D CNN model and trained for 32 epochs. The 3D CNN architecture consists of a Convolutional layer which is a multi-layer design that identifies and extracts the high-level features as we can see in Fig. 3. After the model was fully trained, given input CT scans, we extracted DL features from the second dense layer in the last with 64 units (high level DL features). These 64 DL features will be used as input features for the RF classifier. See section IV for details of our experiments.

(b) Radiomic texture feature extraction

As step 2, we have extracted the Radiomic texture features which are calculated by using the 3D public COVID-19 CT scans along with the associated infection masks [19].

We have calculated the subset of Radiomic texture features and selected 7 Radiomic texture features. These 7 Radiomic texture features have been shown to be reliable and quantitatively informative in differential diagnosis of COVID-19 and other lung diseases [3, 26, 28, 29].

(i). Gray Level Co-occurrence Matrix (GLCM): Gray Level Co-occurrence of matrix size N_g times N_g is the second-order joint probability of the image region. It is defined by the mask with the probability function, where (i, j) defines the number of times the combination of pixel i and j occurs in an image with a distance value and angle. Some of the features that we used in our study include [28]:

$$Maximum\ Probability\ = max(p(i,j))$$

Maximum probability is the appearance of the most pre-dominant pair of neighboring pixel intensities [28]

$$Joint\ Entropy\ = \ -\sum_{i=1}^{N_g} .\sum_{j=1}^{N_g} p(i,j)\ log_2(p(i,j) + \epsilon)$$

Joint Entropy is the measure of randomness in the neighborhood intensity values where, ϵ is the arbitrary small positive number ($\approx 2.2 \times 10^{-16}$) [28]

$$Sum\ Entropy\ =.\sum_{k=2}^{2N_g} p_{x+y}(k)\ log_2\ (p_{x+y}(k)\ +\ \epsilon)$$

Sum entropy is the sum of neighborhood intensity value differences [28]

where $p_{x+y}(k) = \sum_{i=1}^{N_g} .\sum_{j=1}^{N_g} p(i,j)\ \ ,i+j=k$ and $k = 2,3,\ldots,N_g$.

$$Joint\ Energy\ = \sum_{i=1}^{N_g} .\sum_{j=1}^{N_g} (p(i,j))^2$$

Energy is termed as a measure of homogeneous patterns. A greater energy shows that there are more examples of intensity pairs in the image that are close to each other with higher frequencies [28].

(ii). Gray Level Size Zone Matrix (GLSZM): Gray level Size zone matrices are useful to identify the textural homogeneity, and speckle like textures. It is used to differentiate gray scale granularity among different texture features. Some of the important GLSZM features include small area emphasis, large area emphasis, gray level non-uniformity, size-zone non-uniformity, zone percentage, gray level variances, and zone variance. We have used Gray Level Non-Uniformity and Size Zone Non-Uniformity in our experiments.

$$Gray\ Level\ Non-Uniformity = \frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_s} P(i,j))^2}{N_z}$$

GLN signifies the variability of gray-level intensity values in the image where a lower-value shows higher homogeneity of intensity values. [28]

$$Size\ Zone\ Non-Uniformity\ (SZN) = \frac{\sum_{i=1}^{N_s} (\sum_{j=1}^{N_g} P(i,j))^2}{N_z}$$

SZN measure occurrences of size zone quantities in the images [28] where,

N_g is the number of discrete intensity values in the image
N_s is the discrete zone sizes in an image
$P(i,j)$ is the size zone matrix

(iii). Gray Level Run Length Matrix (GLRLM): In a matrix of size N_g times N_g, Gray Level Run Length Matrix is the length of a number of pixels, which has the same gray level values. Some of the important features include short run emphasis, long run emphasis, gray-level non-uniformity, run-length non-uniformity, run percentage, gray level variance, and run variance. We used low-gray level run emphasis in our experiments

$$Low\ Gray\ Level\ Run\ Emphasis\ (LGLRE) = \frac{\sum_{i=1}^{N_g} .\sum_{j=1}^{N_r} P(\,i,j \mid \theta)/i^2}{N_r(\theta)}$$

LGLRE is the distribution of low gray level values where

N_g is the number of discreet intensity values in the image.
N_r is the number of run lengths in the image.
$N_r(\theta)$ is the number of run lengths along angle
$P(\,i,j \mid \theta)$ is the run length matrix for an arbitrary direction θ [28]

(iv). Gray Level Dependence Matrix (GLDM): Gray Level Dependence Matrix describes the gray level dependencies in an image. The number of same gray level data points connected by a distance which is dependent on the center of the matrix defining the value of Gray Level Dependence Matrix. Some of the important features include small dependence emphasis, larger dependence emphasis.

For selecting the Radiomic texture features, image and associated mask dataset is accessed using SimpleITK library. The image and associated mask dataset are then converted to a series of arrays and loaded into the PyRadiomics library.

The parameters were set to extract the features and the available features in PyRadiomics are First order features, Gray Level Co-Occurrence Matrix, Gray Level Run Length Matrix, Gray Level Size Zone Matrix, and Gray Level Dependence

Matrix. A total of 149 features were found for each patient in 3D out of which the selected 7 texture features are listed as shown in Table I.

TABLE I SELECTED RADIOMIC TEXTURE FEATURES

RADIOMIC TEXTURE FEATURE	SPECIFIC FEATURE
Gray Level Co-Occurrence Matrix	Maximum Probability
Gray Level Co-Occurrence Matrix	Joint entropy
Gray Level Co-Occurrence Matrix	Sum Entropy
Gray Level Co-Occurrence Matrix	Joint Energy
Gray Level Size Zone Matrix	Gray Level Non-Uniformity
Gray Level Size Zone Matrix	Size Zone Non-Uniformity
Gray Level Run Length Matrix	Low Gray Level Run Emphasis

C. 3D Dataset and Preprocessing

Our 3D dataset consists of 38 patients where each patient has multiple slices ranging between 100-400. Out of the 38 patients, there are 19 COVID-19 positive patients and 19 non-COVID-19 patients. To achieve a faster DL model's computation time using GPU, for each patient, we have resized the entire 3D CT images in 512x512xZ dimension into 100x100x20 volume where Z is the number of slices that varies from patient to patient. The third dimension is reduced to 20 by chunking multiple continuous slices into 20 chunks then average. Because the patient CT scans can have resolution from 512x512x600 to 1024x1024x600 or larger in Z dimension leading to hundreds of millions of voxels radiodensities in Hounsfield units (HU) per patient. Also, the CNN architectures for training 2D 512x512 slices typically require ~10GB of GPU memory already, and thus training a deep learning model using entire 3D CT scan volume is impractical [18]. Thus, the input 3D volume for each patient has 100x100x20 dimensions, each associated with a label: either 1 (COVID-19 positive) or 0 (COVID-19 negative or non-COVID-19) by Radiologists. These data will be used as input to train our 3D CNN model.

The Radiomic texture feature extraction has been done using full 3D CT scans without using resize and chunking. Therefore, there is no impact on the performance of the model using Radiomic texture features. The segmentation annotations are manually done by the experienced senior radiologists [30]. An open-source python package called PyRadiomics is used to extract 3D features for each image against the associated masks where the extracted features are stored in a data frame for feasibility.

Table II shows the range of values for each of 7 Radiomic texture features for COVID-19 infection and non-COVID-19 cases extracted from 3D CT scans.

TABLE II CALCULATED RADIOMIC TEXTURE FEATURES (VALUES IN RANGE) FROM 3D CT SCANS

Radiomic Feature	COVID data	Non-COVID data
GLCM-Joint Energy	0.001 - 0.644	0.0018 - 0.375
GLCM-Joint Entropy	1.058 - 10.297	1.5 - 9.457
GLCM-Max probability	0.0025 - 0.789	0.0057 - 0.375
GLCM-Sum Entropy	0.988 - 6.405	0.5 - 5.96
GLRLM-Low Gray Level Run Emphasis	0.0018 - 0.753	0.0044 - 0.368
GLSZM-Gray Level Non-Uniformity	1 - 932.158	1 - 1674.54
GLSZM-Size Zone Non-Uniformity	1 - 10837.43	3 - 11403.46

D. Architecture & Training of our 3D CNN model

Our 3D CNN architecture comprises 10 layers with input size 100x100x20 after preprocessing our 3D data for computational feasibility (Fig. 3). It uses RELU as the activation function where the last layer uses sigmoid function with cross entropy as our loss function. The expected output classification from this architecture is in binary format (COVID-19 /non-COVID-19). The first layer of our 10-layer 3D-CNN architecture consists of a 3D-convolution layer (100x100x20, 32 feature maps) followed by activation, batch normalization, and max pooling. The second layer of our architecture consists of a 3D-convolution layer followed by activation, batch normalization and max-pooling. Similarly, we have the odd numbered layers exactly similar to the first layer and the even numbered layers similar to the second layer except the last 10th layer does not have the max-pooling layer at the end but has a flattening layer instead. This combination of convolution layers is then followed by 2 dense layers along with batch normalization and dropout followed by a final dense layer.

E. RF Classifier

RF works on the principle of collective decisions. Multiple decision trees collectively form a RF. The RF classifier is an ensemble of N trees $\{T_i(X), ..., T_N(X)\}$, where $X = \{x_1, ..., x_p\}$ is a p-dimensional feature vector describing the input to be classified. This type of an ensemble produces N outputs $\{\hat{Y}_1 = T_1(X), ..., \hat{Y}_N = T_N(X)\}$ and these N outputs are later aggregated to predict the class of $\hat{Y}$ of the input. A single feature forms a branch of each decision tree. An input is given to multiple decision trees, and the collective majority will be the output of the RF algorithm. For example, 7 selected Radiomic texture features (see table I for the list of features) extracted from CT scans are given as input to the RF model, output predicts either patient having COVID-19 infection or non-COVID-19 case. We have used the RF classifier to form an ensemble with the 3D-CNN model. The number of trees in

the forest was set to 5000. The random state was set to 42 when using the bootstrapping method to fit the model.

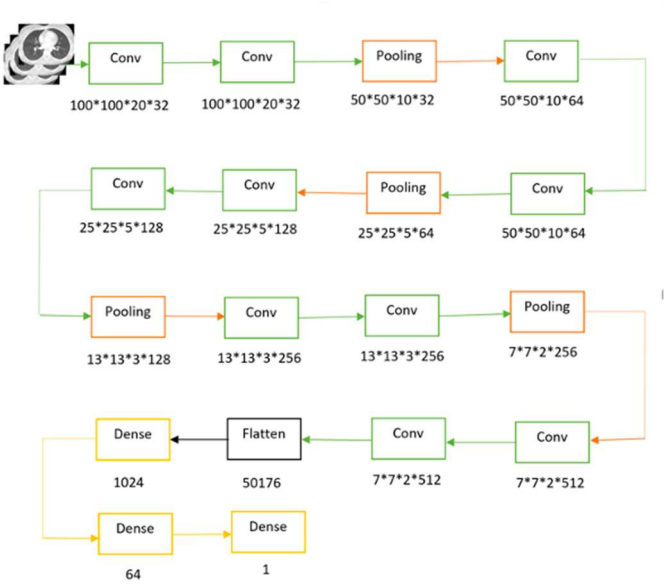

Fig. 3 3D CNN Architecture

F. Performance Metrics

We employed one of the important evaluation metrics called AUC-ROC. The AUC-ROC curve measures the performance of a binary classification problem at different threshold settings. The True Positive Rate and The False Positive Rate needed to be calculated for the ROC plot [13][14]. The probability curve is the ROC, and the degree or measure of separability is represented by AUC [13][14]. This evaluation metric tells us the capability of our model in distinguishing between different classes (COVID-19 or non-COVID-19). The better performance curve should hug tightly to the top left corner of the ROC plot. The higher the AUC, the better is our model in predicting 1s as 1s and has a lower chance of predicting the wrong class. In addition, we use a *t-test* to determine if there is a significant difference between mean value of the AUCs obtained from the experiments which use Radiomic texture features alone and the combination of DL and Radiomic texture features (our proposed solution).

IV. Experimental Design and Results

We performed 4 experiments to evaluate the performance of COVID-19 classification using our DL using 3D CNN only, RF, Images, and the combination of DL and Radiomic texture features as follows:

- Experiment 1 (Exp#1): RF classification using 3D Radiomic texture features only.
- Experiment 2 (Exp#2): 3D CNN DL classification using 3D CT scans only.
- Experiment 3 (Exp#3): RF classification using extracted DL features only.
- Experiment 4 (Exp#4): Our proposed RF classification using a combination of extracted 3D DL and Radiomic texture features.

We also presented the results of 4 additional experiments using the same design as experiment #1 to #4 above using 2D CT scans and 2D Radiomic texture features extracted from using 3888 2D CT images.

For each experiment, the RFs were trained and tested 1000 times. Then the True Positive Rate, False Positive Rate and AUC performance on the testing dataset were used for ROC/AUC plots.

Experiment 1 - RF Classification using 3D Radiomic texture features only

In this experiment we have performed classification using the extracted 3D Radiomic texture features using the RF classifier. The input is 7 Radiomic texture features shown in table I, to classify if the patient has COVID-19 or is a healthy patient. After extracting the 3D Radiomic texture features using PyRadiomics, the 7 texture features were stored and the data was split into 80% patient's data for training and 20% test data. The RF model with 5000 trees was trained and tested on the separated 20% test data. The overall workflow for creating RF classification using the Radiomic texture features extracted from 3D CT scans is shown in Fig. 4.

The test ROC is plotted in Fig. 5. It reaches a True Positive Rate of approximately 1 when the False Positive Rate is 0.18 with AUC of 0.9918.

Experiment 2 - 3D CNN DL classification using 3D CT scans only.

In experiment 2, we performed classification using a 3D CNN model with 3D CT scans only. The overall architecture of the 3D CNN model is shown in Fig. 3. The 3D CT scans were pre-processed, and the final dimension was (100x100x20). These data cubes are the input for the 3D CNN model to classify if the patient has COVID-19 or is a healthy patient. 80% percent of the patient's data was used for training and 20% for testing. The test ROC/AUC results are plotted in Fig. 8.

Experiment 3 - RF classification using extracted DL features only.

In experiment 3, we performed RF classification using DL features extracted from the trained 3D CNN model. From the 3D CNN model fully trained using 3D CNN data (in experiment 1), the DL features were extracted from the second dense layer which has 64 units given the 3D input dataset. Thus, the input for the RF model has a vector of 64 values (64 dense layer units). The input data is split into 80 % patient's data for training and 20% test data. The overall workflow for creating

this model is shown in Fig. 6 . The test ROC/AUC results and its comparison is described in Fig. 8 .

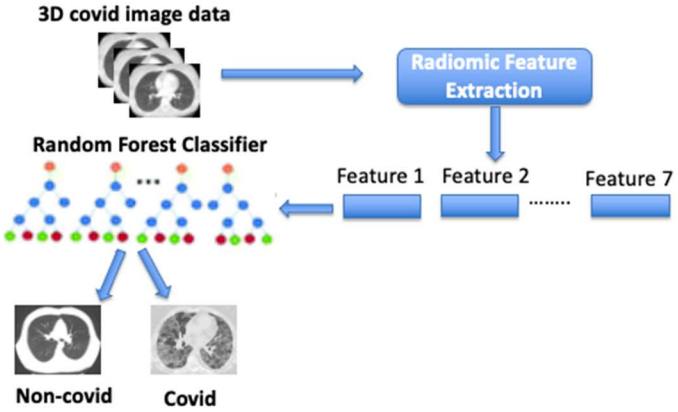

Fig. 4 RF Classification using 7 Radiomic texture features extracted from 3D CT scans.

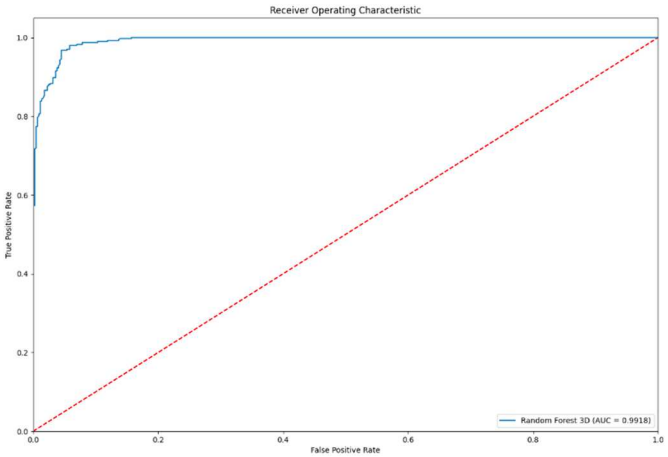

Fig. 5 The best test AUC-ROC curve for 3D DL features plus 3D Radiomic texture features experiment using 3D dataset

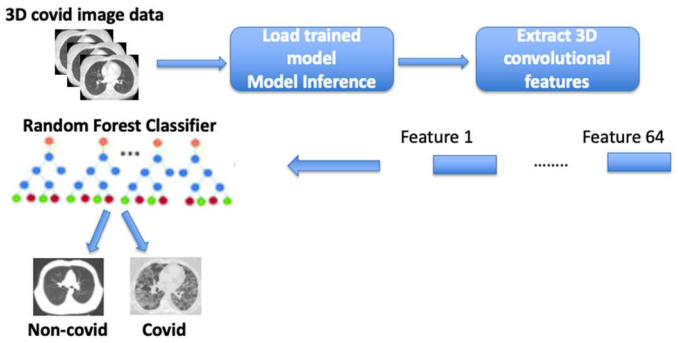

Fig. 6 Classification using 3D CNN DL features using RF

Experiment 4 Our proposed RF classification using a combination of extracted 3D DL and Radiomic texture features.

In this experiment, we performed RF classification using extracted DL combined with Radiomic texture features. There are 7 Radiomic texture features extracted, and the list of features are shown in Table I. The DL features are the same as the DL feature, we used in experiment 3. Since the DL features are a vector of 64 values and Radiomic texture features have a length of 7, we duplicate the Radiomic texture features 7 times before combining them. Thus, we have combined a vector of 407 features as input for the RF classifier. The input data is split into 80% of the patient's data for training and 20% for testing. The overall workflow is described in Fig. 7. The ROC/AUC results are plotted in Fig. 8.

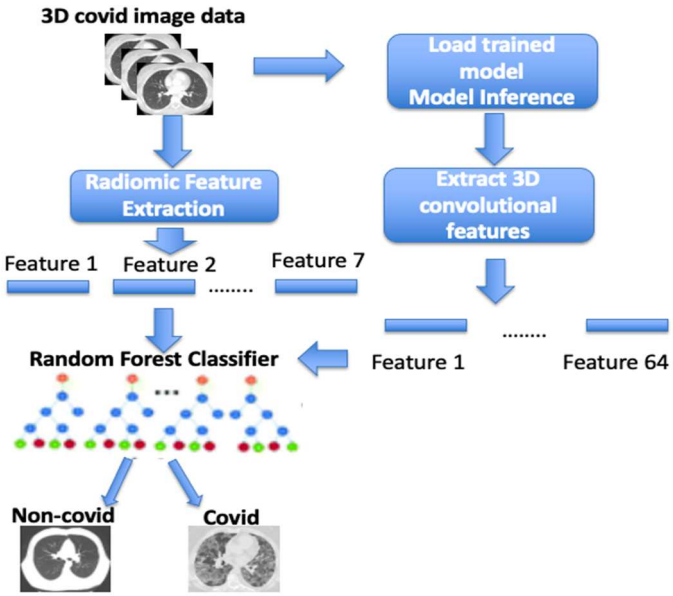

Fig. 7 Classification using 3D CNN Deep learning features plus 3D Radiomic features using RF.

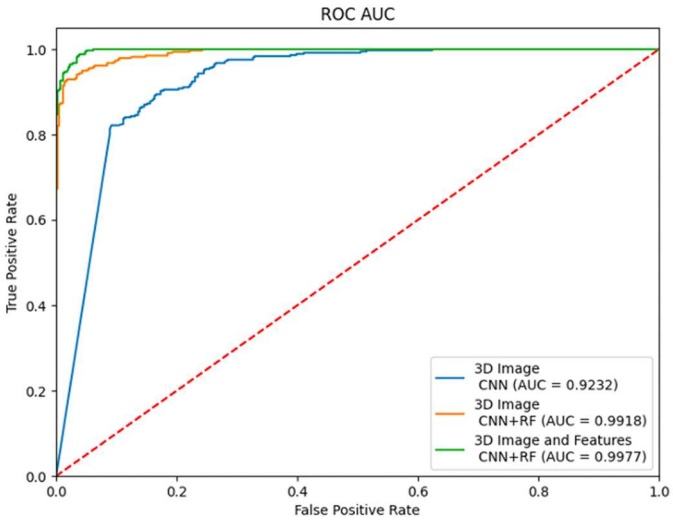

Fig. 8 The best Test ROC/AUC curves for Experiment 2, 3, 4. Our proposed model ROC is green.

In Fig. 8, the classification of 3D images using just the 3D CNN model (blue) only gives us the best AUC of 0.9232 whereas, the classification using 3D images using Exp#3 3D CNN and RF (orange) produces a better AUC of 0.9918. The

ROC curves show that the RF using extracted DL features (orange) gives a much better threshold of True Positive Rate with lower False Positive Rate than the 3D CNN model only (blue). Evidently, our **proposed model's ROC curve (green)**, hugging the top left corner, reaches quickly with high True Positive Rate and lower False Positive Rate for the experiment number 4. It gives approximately 90% True Positive rate with almost 0% False Positive Rate. Our combined model produces the best AUC value (0.9977). This justifies that our novel RF classifier using the combined DL and Radiomic texture features (Exp#4) yields the best performance compared with other models.

TABLE III. Summary of our experimental results

Experiments (Best Testing performance)	2D model (AUC)	3D model (AUC)
Exp#1 RF Classifier using Radiomic texture Features only	0.9856	*0.9918*
Exp#2 DL CNN	0.9467	0.9232
Exp#3 RF classifier using DL features	0.9704	0.9918
Exp#4 RF classifier using the combination of both DL and Radiomic texture Features	0.9912	**0.9977 (our model)**

Table III shows the AUC of the 4 experiments which we described above. In addition, we conducted the same experimental design for 2D CNN model, training using 2D input data of 3888 CT scans, and Radiomic texture features (extracted from the 2D images). The results for the 2D experiments are shown in table III. The AUC results in table III shows that 3D models performed better than that of 2D models.

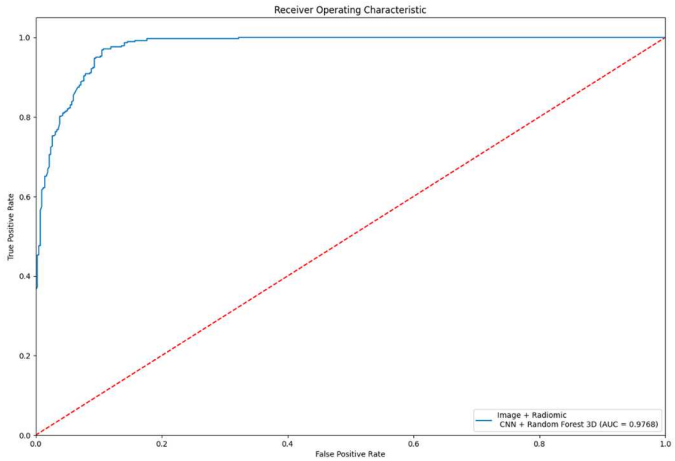

Fig. 9 Average Test AUC-ROC for RF Classifier using 3D Deep Learning plus Radiomic texture features (Exp#4).

Fig. 9 shows the Average AUC-ROC curve of our proposed model (Exp#4) with a RF classifier using a combination of DL and Radiomic texture features. The average value with confidence interval is Average AUC = 0.9768, 95% Confidence Interval [0.9756, 0.9779].

The effectiveness of using DL features combined with the Radiomic texture features in comparison of using Deep learning features alone is studied by using *t-test* on the obtained mean value of the AUCs obtained from the two experiments (Exp#2 and Exp#4). The equation for calculating the t-test value is as follows:

$$t = \frac{m - \mu}{s / \sqrt{n}}$$

where, m is the mean and μ is the theoretical value, s is the standard deviation and n is the variable set size. The obtained t-test value is 2.356, and the *p-value* <0.02. The p value of less than 2% is statistically significant, which means that our RF model using the combination of DL features and the Radiomic texture features is better than using DL CNN only (DL model trained using images). Experiments have been conducted using AMD 1885 MHz 32 cores, 658 GB, 3 NVIDIA GeForce RTX, each GPU has 11 GB memory. It took ~4 hours to train a DL model.

In summary, both *t-test* values and ROC/AUC results in Fig. 8 and Fig. 9 shows that our RF model using the combination of Radiomic texture features and DL features gives the best performance with high True Positive Rate.

III. V. Discussion

We have a limited total of 38 patients for 3D CNN models. However, the 2D CNN models are trained using 3888 2D CT images. In which there are 300+ COVID-19 infected areas (COVID-19's segmented annotations) in 1800+ images. The infected COVID-19 positive cases range from minor to moderately severe [0.01% to 59%]. We showed that the performance of 3D CNN models and 2D models show consistent improvements when compared between experiments #1, #2, #3, and #4 (see results in table III). With the increase in the size with diversity of patients' dataset, the performance of the models might drop, and further DL techniques can be explored for these classification tasks [2, 9, 18, 20]. As future work, morphological and pathological features can be extracted and added to enhance the classification models. Further experiments can be done by considering classifying COVID-19 infection against other lung diseases such as Pneumonia, Lung Cancer etc. In addition, the 3D CT scans dimensions were resized and chucked to 100x100x20 per patient for training DL models because of the burden of computing time and limitation of GPU's memory. Dimension reduction comes with the loss of information issues. We recommend another method to overcome this issue where the 3D CT scans are divided into multiple non overlapping sub 3D volumes for computation [18].

From our results, we performed four different combinations of experiments that led to evaluating different models. We recommend performing experiment # 1 (RF Classifier using Radiomic texture features only) when there is no segmentation or computation power to handle Deep Learning 3D CNN. We recommend performing experiment # 4 (RF classifier using the combination of both extracted DL and Radiomic texture Features) to get the best performance. We expect to see the adoption of our proposed model with further validation using much larger cohorts with diverse patient's CT scans.

IV. VI. Conclusion

In this work, we present a novel approach for classification of COVID-19 infected cases by incorporating a RF model to train on a combination of DL and Radiomic texture features that were extracted from CT scans of patients' lungs. In addition, we have developed and trained customized DL models using 10 layers of 3D CNN architecture for extracting DL features. Radiomic texture features combined with extracted DL features boost the performance of the prediction compared with just using DL or Radiomic texture classifiers. Our proposed 3D model has achieved a highly True Positive rate with the average AUC of 0.9768, 95% CI [0.9757, 0.9780].

Acknowledgment

This research was supported by NSF RAPID award titled: Deep learning Models for Early Screening of COVID-19 using CT Images, award # 2027628.

References

[1] S. Ozturk, U. Oz Kaya, and M. Barstugan, "Classification of Coronavirus (COVID-19) from X-ray and CT Images using Shrunken Features," International Journal of Imaging Systems and Technology, 31(1), 5-15.

[2].S. Menon, J. Galita, D. Chapman, A. Gangopadhyay, J. Mangalagiri, P. Nguyen, and M. Morris, "Generating Realistic COVID19 X-rays with a Mean Teacher+Transfer Learning GAN", arXiv preprint arXiv:2009.12478, 2020.

[3] F. Xu, W. Zhu, Y. Shen, J. Wang, R. Xu, C. Outesh, L. Song, Y. Gan , C. Pu, and H. Hu, "Radiomic-Based Quantitative CT Analysis of Pure Ground-Glass Nodules to Predict the Invasiveness of Lung Adenocarcinoma," Frontiers in oncology, 10, 872.

[4] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. Van Stiphout, P. Granton, and H.J. Aerts, "Radiomics: extracting more information from medical images using advanced feature analysis," European Journal of Cancer, vol. 48, pp. 441–446, 2012.

[5] K.S. Mehta, A. Jain, J. Mangalagiri, S. Menon, P. Nguyen, and D. R. Chapman, " Lung Nodule Malignancy Suspicion Classifier Using Biomarkers, Radiomics And Image Features,"Journal of Digital Imaging 2021.

[6]V. Parekh, MA. Jacobs, "Radiomics: a new application from established techniques", vol. 1, Taylor and Francis Ltd., 207–226, 2016.

[7].A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A Deep neural network architecture for real-time semantic segmentation," Available at: https://arxiv.org/pdf/1606.02147.pdf, 2016.

[8] JL. Causey, J. Zhang, S. Ma, B. Jiang, JA. Qualls, DG. Politte, F. Prior, S. Zhang, X. Huang, "Highly accurate model for prediction of lung nodule malignancy with CT scans", Sci Rep 8, 1, 2018.

[9] P .Nguyen, D. Chapman, S. Menon, M. Morris, Y. Yesha, "Active semi-supervised expectation maximization learning for lung cancer detection from Computerized Tomography (CT) images with minimally label training data," Proc. SPIE 11314, Medical Imaging 2020: Computer-Aided Diagnosis, 113142E (16 March 2020); https://doi.org/10.1117/12.2549655

[10] W. Wei, X. Hu, Q. Cheng, Y.M. Zhao, and Y. Q. Ge," Identification of Common and Severe COVID-19: The Value of CT Texture Analysis and Correlation with Clinical Characteristics," European radiology, *30*(12), 6788-6796.

[11] Z. Belkhatir, R. S. Jose Estepar, and A. R. Tannenbaum,"Supervised Image Classification Algorithm using Representative Spatial Texture Features: Application to COVID-19 Diagnosis using CT Images," medRxiv.

[12] J.Ali, R. Khan, N. Ahmad, and I. Maqsood," Random forests and decision trees," International Journal of Computer Science Issues (IJCSI), 9(5), 272.

[13] Narkhede, S. (2018). Understanding auc-roc curve. *Towards Data Science*, *26*, 220-227.

[14] Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*.

[15] "What is Deep learning?". https://machinelearningmastery.com/what-is-deep-learning/. Access date : June,2020.

[16] J. Zhao, Y. Zhang, X. He, and P. Xie, "COVID-ct-dataset: a ct scan dataset about COVID-19," arXiv preprint arXiv:2003.13865, 490.

[17] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie, "Sample-efficient Deep learning for COVID-19 diagnosis based on CT scans," medrxiv.

[18] J. Mangalagiri, D. Chapman, A. Gangopadhyay, Y. Yesha, J. Galita, S. Menon, Y. Yesha, B. Saboury, M. Morris, P. Nguyen, "Toward Generating Synthetic CT Volumes using a 3D-Conditional Generative Adversarial Network", The 2020 International Conference on Computational Science and Computational Intelligence Symposium on Health Informatics and Medical Systems (CSCI-ISHI), Las Vegas 2020.

[19] "COVID-19 CT segmentation dataset". http://medicalsegmentation.com/covid19/ Access date: Jan, 2021

[20] S. Menon, D. Chapman, P. Nguyen, Y. Yesha, M. Morris, and B. Saboury, "Deep Expectation-Maximization for Semi-Supervised Lung Cancer Screening," Proceedings of ACM SIGKDD 2019, Anchorage, Alaska.

[21] X. Zhang, D. Wang, J. Shao, S. Tian, W. Tan, Y. Ma and Z. Zhang, "A Deep learning integrated Radiomics model for identification of coronavirus disease 2019 using computed tomography," Scientific reports, 11(1), 1-12. https://doi.org/10.1038/s41598-021-83237-6

[22] X. Wu, H. Hui, M. Niu, L. Li, L.Wang, B. He, and Y. Zha, "Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study," European Journal of Radiology, 128, 109041.https://doi.org/10.1016/j.ejrad.2020.109041

[23] A. Shalbaf and M. Vafaeezadeh, " Automated detection of COVID-19 using ensemble of transfer learning with Deep convolutional neural network based on CT scans," International journal of computer assisted radiology and surgery, 16(1), 115-123.https://doi.org/10.1007/s11548-020-02286

[24] S. Wang, Y. Zha, W. Li, Q. Wu, X. Li, M. Niu, and J. Tian. "A fully automatic Deep learning system for COVID-19 diagnostic and prognostic analysis,"European Respiratory Journal, 56(2).

[25] C. Li, D. Dong, L. Li, W. Gong, X. Li, Y. Bai and J. Tian, "Classification of severe and critical COVID-19 using Deep learning and Radiomics," IEEE journal of biomedical and health informatics, 24(12), 3585-3594.

[26] C. Zhao, Y. Xu, Z. He, J. Tang, Y. Zhang, J. Han, and W. Zhou," Lung Segmentation and Automatic Detection of COVID-19 Using Radiomic Features from Chest CT Images," Pattern Recognition, 108071.

[27] L. Wang, B. Kelly, E.H. Lee, H. Wang, J. Zheng, W. Zhang and J. Song, "Multi-classifier-based identification of COVID-19 from chest computed tomography using generalizable and interpretable Radiomics features," European journal of radiology, 136, 109552.

[28] van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research, 77(21), e104–e107.

[29] Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology. 278:563–577

[30] J. Ma, Y. Wang, X. An, C. Ge, Z. Yu, J. Chen, and X. Yang (2021). Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation. *Medical physics*, *48*(3), 1197-1210.