IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2024.0429000*

# Segmented 3D Lung Cube Dataset and Dual-Model Framework for COVID-19 Severity Prediction

**MOHSIN ALI KHAN[1], ARSLAN SHAUKAT[1], ZARTASHA MUSTANSAR[2] and MUHAMMAD USMAN AKRAM[1]**

[1]Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
[2]School of Interdisciplinary Engineering and Sciences (SINES), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

Corresponding authors: Mohsin Ali Khan (e-mail: makhan.ce20ceme@student.nust.edu.pk) and Arslan Shaukat (arslanshaukat@ceme.nust.edu.pk)

**ABSTRACT** This research presents two key contributions aimed at improving COVID-19 severity prediction, specifically intubation or death within one month using 3D CT scan data. First, we introduce a novel dataset of 2,000 segmented 3D lung cubes meticulously curated from the STOIC dataset through a robust 10-step preprocessing and segmentation pipeline. It is evident that 3D CNNs outperform 2D CNNs in this domain, owing to their ability to capture inter-slice information in 3D images, while Vision Transformers excel in texture-based classification tasks. Therefore, as second contribution we propose two distinct methods for predicting COVID-19 severity, defined as intubation or death within one month. The first method employs a 3D-CNN pretrained on the MosMedData dataset, later fine-tuned on the STOIC dataset with two input layers: one for 3D lung images and another for age and gender metadata. The second method know as 3D-EffiBOT leverages a combination of 3D EfficientNetV2 and iBOT architectures to capture both 3D as well as 2D spatial features from volumetric CT scans. 3D EfficientNetV2 with weights obtained after inflating 2D ImageNet weights, was fine-tuned on the STOIC dataset using a dynamic layer unfreezing strategy, while iBOT was employed to extract 2D slice-level features from axial CT slices. Both models were trained using five augmentation techniques and evaluated using stratified 5-fold sampling to address class imbalance, achieving mean AUC score of 0.7862 and 0.7414 for 3D-EffiBOT and 3D-CNN respectively. This work demonstrates the effectiveness of hybrid architectures in medical imaging, offering a significant improvement over conventional method. The results suggest that combining advanced 3D and 2D feature extractors enhances diagnostic accuracy, providing a valuable tool for predicting severe COVID-19 outcomes. Future research directions include integrating patient pre-COVID medical history, expanding the model's application to other diseases, and exploring ensemble learning for improved performance across diverse populations.

**INDEX TERMS** 3D CT scans, COVID-19 severity prediction, Class imbalance, Deep learning, Dynamic augmentation, EfficientNetV2, Feature extraction, Fine-tuning, Hybrid model architecture, iBOT, Intubation prediction, Volumetric CT analysis

## I. INTRODUCTION

THE COVID-19 pandemic has underscored the need for innovative approaches in global health security, revealing critical gaps in early detection and response systems. Machine learning (ML) algorithms have the potential to analyze diverse datasets to predict and monitor disease outbreaks, enabling proactive interventions. Integrating ML into global health data systems can enhance real-time surveillance and early warning systems, crucial for timely containment measures. To date, extensive research has been conducted on COVID-19 using machine learning techniques for lung segmentation, detection, and severity analysis based on X-rays and CT scans [1]–[4]. In the current era of machine learning, innovative approaches are being developed in the field of medical imaging. Researchers worldwide are collaborating on a global scale, sharing open-source codes, models, and

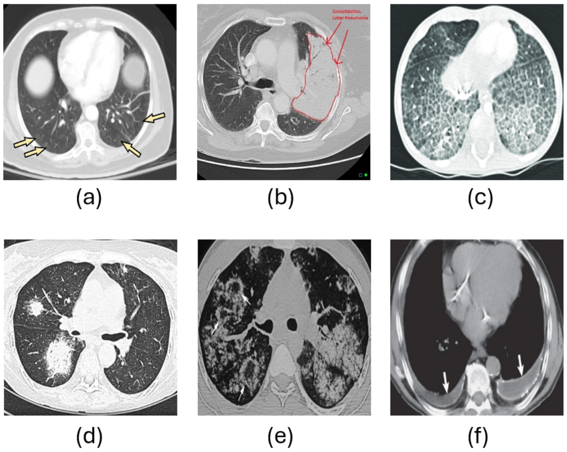


**FIGURE 1.** **Radiological features in 3D CT scans: (a) Ground-Glass Opacities (GGOs) [8], (b) Consolidation [9], (c) Crazy Paving Pattern [10], (d) Halo Sign [11], (e) Reverse Halo Sign [12] and (f) Pleural Effusion [13]**

datasets [5], [6]. CT scans are usually considered as the most accurate diagnostic technique for COVID-19 because of their highly sensitive nature in finding lung problems [7]. Research have proven that CT imaging can reveal early lung shifts even in patients without symptoms, and can spot consequences like acute respiratory distress syndrome (ARDS) and additional infections.

Radiological features like Ground-Glass Opacities (GGOs), consolidation, crazy paving pattern, halo sign, reverse halo sign, and pleural effusion are important in determining the severity of the disease and forecasting patient outcomes, including mortality, in the 3D volumetric CT scan classification of COVID-19 (See Fig. 1). When these all are put together, these imaging features make the COVID-19 classification more accurate. They also give us more information about how the disease develops with the passage of time and make it easier for us to group patients by intensity and estimated death risk. Understanding these patterns in a 3D dimensional context allows for a more complete assessment of the illness's scope and effects. Hence it is essential for making personalized treatment plans and increasing the survival rate of patients.

The primary objective of this research is to enhance clinical diagnostic systems for the benefit of society. By leveraging deep learning models on 3D volumetric images of CT scans, we aim to push the boundaries of what is currently possible in medical diagnostics. This research aims to enhance both the accuracy and early COVID-19 severity estimation while also offering a framework that can be adapted for future pandemics and other medical challenges. By using advanced machine learning techniques, we can create strong tools that provide real-time insights and help healthcare professionals make better decisions. This thesis makes following important contribution to medical imaging and machine learning, providing useful insights for future researchers and improving computer-assisted medical systems in health centers:

1) We proposed two models 3D-CNN and 3D-EffiBOT that went through pre-training and fine-tuning using transfer learning. This approach greatly improves the Area Under the Curve (AUC) severity score, boosting the models' predictive accuracy and reliability.
2) We compiled and meticulously pre-processed a dataset of 3D volumetric CT scans from 2,000 patients using STOIC dataset, focusing on lung segmentation. Each segmented lung cube has been verified through visual inspection, ensuring the accuracy and quality of the dataset. This dataset serves as a robust foundation for training and validating machine learning models.

## II. RELATED WORK

The development of COVID-19 indicates the importance of having reliable and accurate methods for detecting and forecasting disease and its severity. Machine learning models have shown a lot of promise, especially those that use 3D volumetric CT scan. A key obstacle in bringing machine learning to medical images is the limited number of medical datasets compared to normal computer vision datasets. 3D CT scans are particularly not plentiful, making it tough to train models from scratch. To handle this, researchers usually employ pretrained models that have been trained on big, varied datasets and then fine-tune these models on smaller medical datasets [1]–[4]. Two types of methods are mostly applied for the classifying of 3D chest CT images. The first method is known as slicing, in which we divide the 3D volume into 2D slices along any one axis. After that every slice is utilized later to train a 2D classifier. The second technique or method is 3D volumetric approach which involves utilizing the full and complete 3D volume representation of CT scan as input to a neural network based on 3D convolution processes.

In [14], six publically available datasets were used for analysis, including Mosmed [15], MedSeg, and MedSeg_1 [16] for infection area segmentation, while the SPGC [17] dataset is used to train the classification model. In order to assess the accuracy of the model, three additional datasets (LDCT, LDCT-PCR [18], and Mosmed) were used for assessment. The work uses an EfficientNetB51 to train the final classifier and a pretrained MobileNet for data augmentation during the classification phase. After preprocessing, the labeled slices are sent into the MobileNet to extract slice-wise feature maps, which are then run through a global-average-pooling layer of processing. In [19], Convolutional Neural Network (CNN) architecture with four convolutional layers, two dense layers, and flattening in between, is the recommended approach. This design is meant to group 2D slices of CT scans successfully using COV19-CT-DB database [20]. The preprocessing of these images includes anatomy-relevant features to focus on important areas and removal of non-relevant parts thus finally they increase the model overall accuracy. The CT slices are carefully labeled in order to give a good dataset for training the CNN. The training process employs a cross-validation methodology to ensure the model's stability and generalizability. The paper also emphasizes that a well annotated and well diverse dataset is very essential for effective model training. Furthermore, the model's simplicity and lightweight nature make it fit for usage in environments where computer power may be limited. In [21], a unique way is presented to increase the generalizing capabilities of 3D convolutional neural networks (3D CNNs) for COVID-19 detection. Using a greedy training method, ten distinct 3D CNN models are built throughout the process. This method generates a large number of different models, the best of which is selected based on performance metrics like accuracy and loss. Two datasets, SARS-CoV-2 CT [22] and Mosmed were used to train and validate the models. The results of the study proved the success of the suggested greedy training method. Specifically, the 3D CNN model labeled as Net5 had the best success. The researchers also conducted a comparison between their results and the most recent state-of-the-art algorithms, and the results showed that their model outperformed others, including EfficientNet-B0 and ResNet-50. These findings suggest that the greedy training technique could be a good substitute for well-established methods by offering a reliable way to identify viruses from CT scans.

The paper [23] presented an advanced neural network model based on the 3D conversion of the 2D ConvNeXt architecture to predict the severity of damage to the lung and identify COVID-19 infections using CT images. This research emphasizes the importance of modifying current 2D neural network designs to accommodate 3D medical imaging data. To address the limited number of medical datasets, it devised multiple pretraining strategies which were meant in order to increase model performance on 3D CT data. The model was trained and tested using the COV19-CT-DB database. This research revealed especially high effectiveness in recognizing extreme cases, which is essential to clinical applications. In [24] the implementation of deep learning models for classifying COVID-19 severity based on CT scans was examined which includes employing eight alternative Convolutional Neural Network (CNN) architectures: MobileNetv2, ResNet101, Xception, Inceptionv3, GoogleNet, EfficientNetb0, DenseNet201, and DarkNet53. The research developed a pipeline approach to aggregate the outputs of the top-performing CNN models to boost classification accuracy. The CT slices were preprocessed and scaled according to the specifications of each CNN architecture, and a 10-fold cross-validation technique was used to train and test the models. In [25] both 2D and 3D techniques for identifying COVID-19 in CT scans were evaluated to prove the usefulness of ensemble methods. According to the research, accuracy increased when 2D and 3D models were combined. First models were trained on IST-C and MosMedData datasets and then models were evaluated on the COVID-CT-MD. The research also indicates that attention processes and the use of LSTM for combining slice-level forecasts improved the accuracy of 2D models. Meanwhile 3D models advantageous in applying segmentation masks as input channels. This work highlights the need to improve the precision and robustness of COVID-19 detection systems by using both 2D and 3D data as well as ensemble techniques.

Tan and Liu [26] presented an improved design for COVID-19 diagnosis from CT-scan images utilizing a 3D CNN network coupled with BERT for classification. Resampling was used by the researchers to choose a predetermined number of slices from the CT volumes. These selected slices were subsequently categorized using the 3D CNN-BERT model which uses BERT for temporal pooling. Their approach comprises applying both normal morphology transforms and a UNet-based deep learning method for segmentation. In the study [27] by Hou et al., the authors offer a new approach for improving the accuracy of COVID-19 diagnosis using computed tomography (CT) scans. Mixup augmentation, a data augmentation technique that generates new training samples by interpolating between existing samples and contrastive learning together form the core of their methodology. The authors argue that this combination helps the model learn more robust features by enforcing similarity between augmented views of the same sample while distinguishing them from others, which is particularly beneficial given the limited availability of labelled COVID-19 CT images. To assess their approach, the authors used a large dataset that included both COVID-19 and non-COVID-19 CT images from two chest CT image datasets namely the COV19-CT-DB and MosMed database. Using a 3D ResNet50, the CMC-COV19D model was trained. To help the model acquire discriminative features, contrastive learning goals were added to the loss function. Additionally, the Mixup approach was utilized during training to produce blended samples, boosting the model's capacity to generalize across changes in the data.

The paper by Zunair et al. [28] examines the difficulties involved with processing volumetric CT scan data using deep learning algorithms, especially for the prediction of tuberculosis (TB). Traditional 2D convolutional neural networks (CNNs) generally fall short owing to their inability to use the depth information inherent in 3D data, resulting to unsatisfactory performance in medical picture processing. A 3D CNN architecture is deployed to process the entire volumetric data, capturing spatial context and depth information that is typically lost in 2D approaches. The study emphasizes various benefits of 3D CNNs, such as greater usage of spatial context and depth information, which are critical for precise illness prediction.

## III. MATERIALS AND METHODOLOGY

To predict severe COVID-19 infection, defined as incubation or death within one month, from computed tomography (CT) scans, two datasets of 3D CT scans were utilized: the "MOSMEDDATA: Chest CT Scans with COVID-19 Related Findings Dataset" [15] and the "Study of Thoracic CT in COVID-19: The STOIC Project" [29]. MosMed dataset comprises anonymised human lung computed tomography (CT) scans findings which are connected to COVID-19 (CT1-CT4) and which are normal (CT0). There were 1110 investigations including CT-0-254, CT-1-684, CT-2-125, CT-3-45, and CT-4-2. Secondly each file is stored in the NifTI format and preserved in the Gzip file which preserves complete volumetric

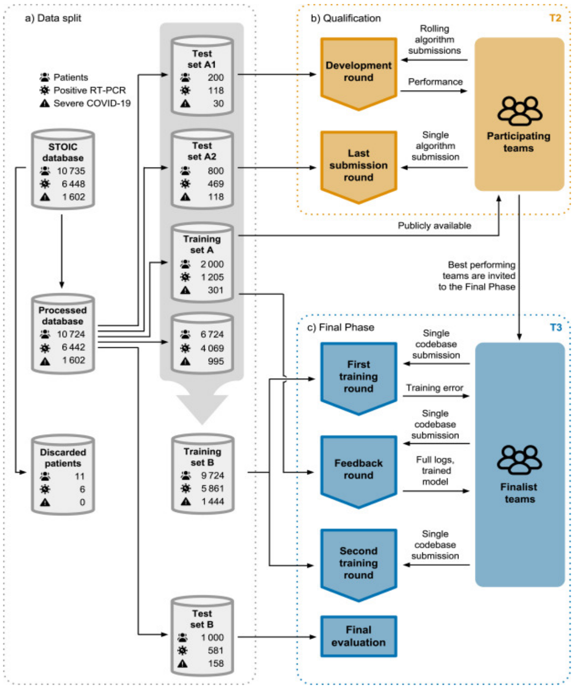


**FIGURE 2.** STOIC dataset schematic overview [30]

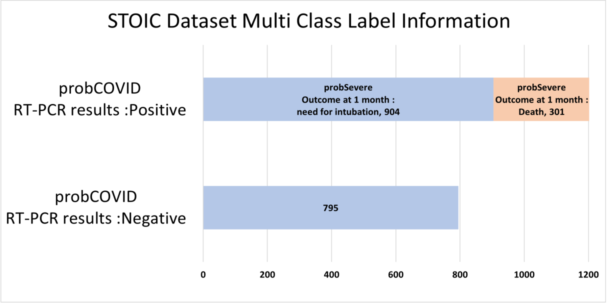
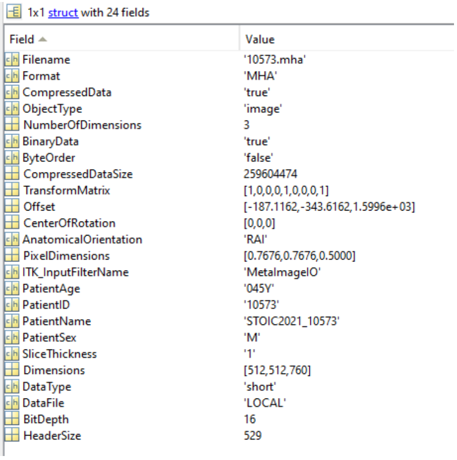


**FIGURE 3.** STOIC dataset Class Label Information

data required for correct analysis. The STOIC project dataset consists of thoracic 10735 volumetric 3D CT scans from COVID-19 patients out of which 2000 are available publicly as shown in Fig. 2, provided in high-resolution Meta-Image Medical Format. Till date this the largest volumetric 3D CT scans dataset collected. Each CT scan includes multiple slices with varying thickness, forming a 3D volumetric image of the thorax. The typical dimensions of these scans vary based on the patient's anatomy and scanning protocol but generally follow standardized medical imaging specifications. Each slice maintains consistent resolution and image quality, allowing for detailed analysis of lung structures and abnormalities.

The STOIC dataset includes annotations and classifications based on the probability as well as severity of lungs as involvement due to COVID-19 as shown in Fig. 3. The CT scans in the STOIC dataset were collected from multiple hospitals and medical centers across different regions of France. The collection period spans from March to May 2020, capturing data from the early stages of the pandemic. This period gives a comprehensive view of early stages of the virus on patients' lungs, including a variety of cases with different severity and disease progressions.


1x1 struct with 24 fields

| Field | Value |
|---|---|
| Filename | '10573.mha' |
| Format | 'MHA' |
| CompressedData | 'true' |
| ObjectType | 'image' |
| NumberOfDimensions | 3 |
| BinaryData | 'true' |
| ByteOrder | 'false' |
| CompressedDataSize | 259604474 |
| TransformMatrix | [1,0,0,0,1,0,0,0,1] |
| Offset | [-187.1162,-343.6162,1.5996e+03] |
| CenterOfRotation | [0,0,0] |
| AnatomicalOrientation | 'RAI' |
| PixelDimensions | [0.7676,0.7676,0.5000] |
| ITK_InputFilterName | 'MetaImageIO' |
| PatientAge | '045Y' |
| PatientID | '10573' |
| PatientName | 'STOIC2021_10573' |
| PatientSex | 'M' |
| SliceThickness | '1' |
| Dimensions | [512,512,760] |
| DataType | 'short' |
| DataFile | 'LOCAL' |
| BitDepth | 16 |
| HeaderSize | 529 |

**FIGURE 4.** Meta Data Associated to .mha File in STOIC dataset

### *A. PRE-PROCESSING AND SEGMENTATION*

During the initial inspection of the raw STOIC dataset, it was observed that all images in the .mha format contain embedded metadata, which includes both CT scan specifics and patient details as shown in Fig. 4. Among the metadata properties, several were deemed critical for preprocessing the .mha files. These include the transform matrix, offset, center of rotation, anatomical orientation, voxel dimensions, image dimensions, datatype, and bit depth. These properties are essential for accurately interpreting the CT scan data and ensuring consistent preprocessing.

Additionally, patient-specific information such as age and sex, embedded within the metadata, is vital for severity classification. These demographic factors influence COVID-19's progress and severity and provide a valuable framework for the model. Incorporating patient age and sex into the training process can enhance the model's predictive accuracy and generalizability. Therefore, these demographic factors were documented and used during the processing stage to refine the training dataset further and improve the model's capability to classify COVID-19 severity more effectively. This comprehensive approach ensures that both the technical and contextual aspects of the CT scans are leveraged, leading to more robust and reliable predictions. The steps outlined in Fig. 5 along with their corresponding output images at various stages of pre-processing as shown in Fig. 6 were followed to extract lung volumetric images from the raw CT scans:

#### 1) Image Data Type Handling

As shown in meta data, image types were 16 bit signed integer as voxel values were mentioned in Hounsfield Unit as a quantitative measure of radio density in CT images. So

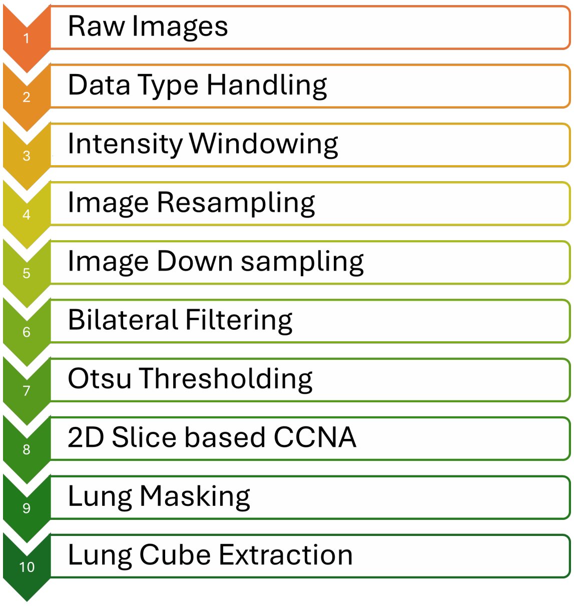


**FIGURE 5. Steps followed during pre-processing and segmentation**

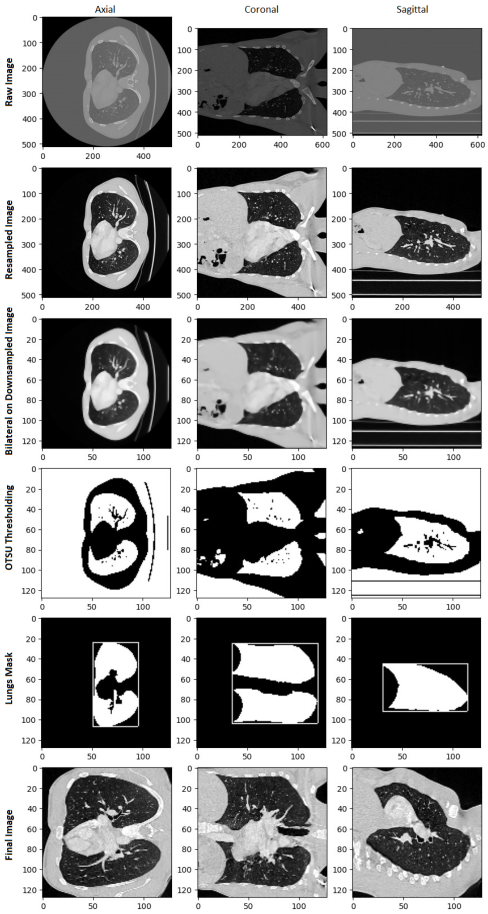


**FIGURE 6. Output Images on various stages of pre-processing and segmentation**

all images were assigned Float 32 data type before doing any manipulation.

### 2) Intensity Windowing

In literature review [31] it was noted that each body part has its own Hounsfield value representation in CT scan images and for COVID related research papers more or less -1000 to 400 range was used for lungs analysis therefore images voxels intensities were clipped and normalized to 0 to 1 from window specified above.

### 3) Image Resampling

All Images were resampled to Size [512, 512, 512] and Voxel Spacing [0.75, 0.75, 0.75]. During resampling if the image was stretched previously, boundary pixels were replicated after resizing.

### 4) Lungs Cube Extraction

All images were down sampled to [128, 128, 128], then bilateral filter is applied to preserve edges and smoothing. Then Otsu threshold is applied to do connected component analysis on Coronal, Axial and Sagittal plane along all slices to get the mask boundaries as shown in Fig. 7. An additional 8% padding is applied to each dimension of the lung cube as defined by the mask boundaries. This ensures that the affected areas near the lung surface are preserved while also providing some extra space to keep the lungs within the image during rotation transformations during training. The final corner coordinates of the cube were computed and were used to extract the lungs from resampled image of [512, 512, 512] and was reshaped to [128, 128, 128]. At the end final images were all same sizes.

As Fig. 7 displays hotspots indicating the regions with the maximum area at the slice level across each axis plane within the resampled images. In contrast the graphs in Fig. 8 illustrate the regions with the maximum area within the final lung cube volumetric image. These graphs demonstrate that the lung cubes for all patients are consistently centered, cropped, and reshaped, making them independent of the patient's actual body shape, gender, and age. The mean line with minimal variance confirms that all features are spatially well-aligned. This alignment facilitates faster training even with not complex machine learning models as computations grow exponentially with the inclusion of the third axis to cater the entire 3D volumetric information.

## B. IMAGE AUGMENTATION

To enhance network generalization, data augmentation plays a crucial role, with common techniques including flipping and rotating the images. Deep learning models benefit significantly from these augmentations, as they help improve the models' ability to generalize from the training data. In scenarios involving small datasets, pre-trained networks are often employed to avoid over-fitting. These networks are typically trained on large public datasets of 2-dimensional RGB images, which differ from medical CT scans that are

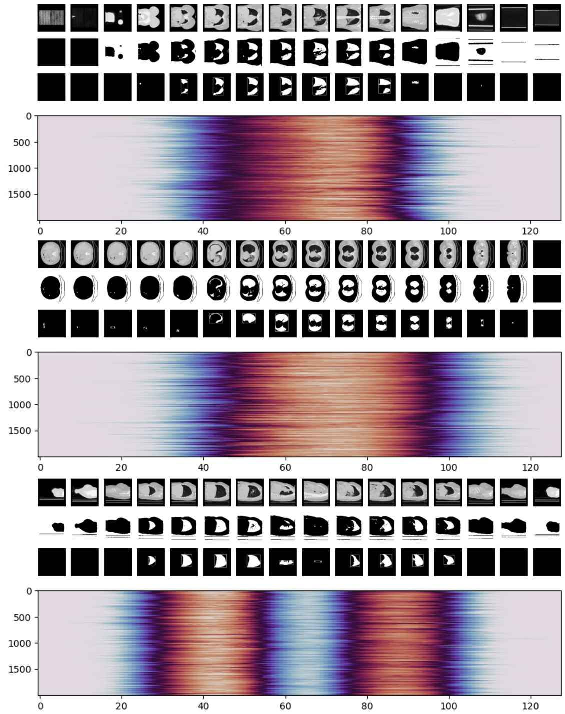


**FIGURE 7.** Connected Component Analysis on Coronal, Axial and Sagittal Slices after Otsu threshold and bilateral filter to the resampled 2000 volumetric images

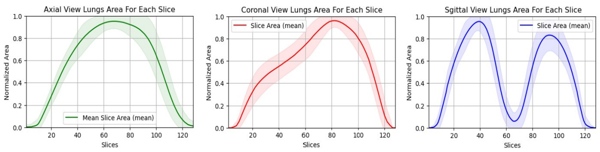


**FIGURE 8.** Graphs depicting the distribution of maximum area within the final lung cube highlighting consistent alignment and independence from patient specific factors

volumetric and gray-scale. Despite these differences, pre-trained networks and transfer learning substantially boost network generalization. The specific augmentation techniques applied during training for both the MosMedData and STOIC datasets were Gaussian Noise, Gaussian Smooth, Rotation Transform (-15 to 15 degree along any axis in 3D view), Elastic Transform and Flip Transform (mirror flipping only). Fig. 9 compares a normal volumetric image with the output of each augmentation method. For Rotation and Elastic Transform, spline interpolation of order 3 was employed to ensure smooth and realistic transformations. These augmentations play a crucial role in enhancing the model's robustness and generalization by simulating various real-world conditions and introducing variability into the training dataset. The visual comparison underscores the effectiveness of these techniques in preparing the dataset for more accurate and reliable predictions of COVID-19 severity from CT scans. The impact and benefits of these augmentation methods will help model to get better generalization.

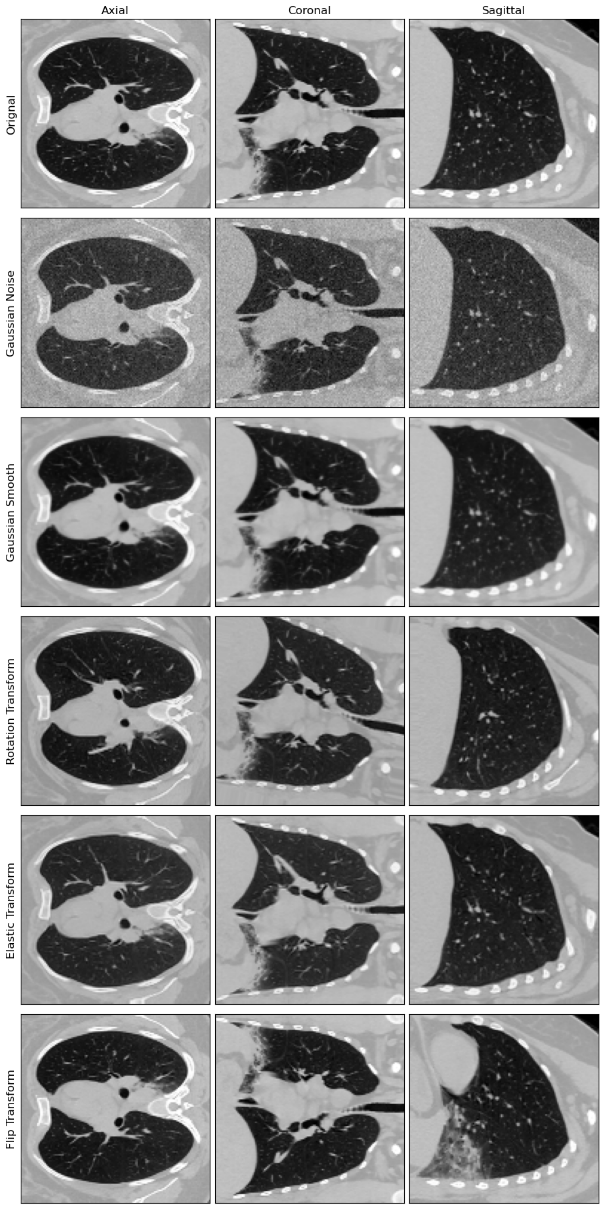


**FIGURE 9.** Comparison of augmentation techniques output employed with the original volumetric image

## IV. CLASSIFICATION NETWORKS

### *A. 3D-CNN*

To predict severe COVID-19 infection from CT scans, defined as intubation or death within one month, initially a 3D Convolutional Neural Network (CNN) was designed and trained on the MosMedData to create a pre-trained network. The network architecture consists of an input layer for images of shape [128, 64, 128], followed by four convolutional blocks. Each block contains a 3D convolution layer with filter sizes of [64, 128, 256, 512], activation relu ,a kernel size of [3, 3, 3], and a stride of [1, 1, 1]. These convolution layers are followed by 3D batch normalization layers and 3D pooling layers. MaxPooling is used for the pooling in each block. This architecture aims to capture complex spatial features from the CT scans, enhancing the model's ability to predict severe COVID-19 outcomes accurately. The detailed design of the model is shown in Fig. 10(a).

During pre-training, MosMedData annotations were assigned numeric values based on severity. As there were five classes present in MosMedData (Zero, Mild, Moderate, Severe and Critical), only Moderate, Severe and Critical are used for pretraining in order to classify severity. Moderate was assigned 0 and Severe and Critical CT images were labeled as 1. For severity classification this encoding ensured compatibility with the model's final dense layer, which has a single unit with a sigmoid activation function. The model was trained for 25 epochs with a batch size of 19, using the binary crossentropy loss function and the Adam optimizer with a learning rate of 0.0001. The same preprocessing and augmentation techniques detailed above were applied during training. This approach aimed to fine-tune the model's ability to predict COVID severity levels from CT scans by leveraging the diverse and annotated MosMedData, setting the stage for subsequent training on the final STOIC dataset to enhance the model's predictive performance.

After pre-training, the model was modified to include an additional input layer to incorporate age and gender characteristics from the metadata associated with the raw .mha image files in STOIC dataset. This enhancement allows the model to utilize these demographic factors alongside the CT scan data to predict whether subjects had a severe COVID-19 infection, defined as intubation or death within one month. The final architecture of the modified model is illustrated in the Fig. 10(b). The weights obtained during the pre-training on MosMedData were retained during this transition. The modified model was then further trained using the same preprocessing and augmentation techniques on new dataset derived from STOIC dataset after extensive pre-processing and segmentation. This integrated approach leverages both imaging and demographic data to improve the model's predictive accuracy and robustness, addressing the complexity and variability inherent in real-world clinical scenarios.

The dataset was divided into five stratified folds for cross-validation, with 80% allocated for training and 20% for validation and testing. For the severity classification task, CT volumetric images labeled as COVID-19 positive were used, with 904 images indicating the need for intubation and 301 images labeled with death as the outcome within one month. Given the significant class imbalance, the training data was balanced by over sampling of the minority class training samples such that their total numbers gets equal to the majority class. To further address this imbalance and improve generalization, data augmentation techniques were applied on-the-fly during model training. Each sample had a 50% chance of being augmented at each epoch, with one of five augmentation techniques selected randomly. This dynamic augmentation strategy ensured that each sample could be transformed differently in each epoch, aiding in the robust generalization of the 3D CNN model.

Finally training was conducted using a batch size of 25, with the binary cross-entropy loss function and the Adam optimizer employed for optimization. The learning rate was scheduled to decay exponentially starting at 0.00001 with a decay step of 200 and a decay rate of 0.96 in a staircase manner. In order to fine-tune and modify hyperparameters the model was trained for 25 epochs with validation accuracy and loss being recorded at each epoch. The goal of this training strategy was to maximize model performance and guarantee strong generalization to final data. It includes controlling class imbalance and augmenting data dynamically on the fly and closely observing validation measures.

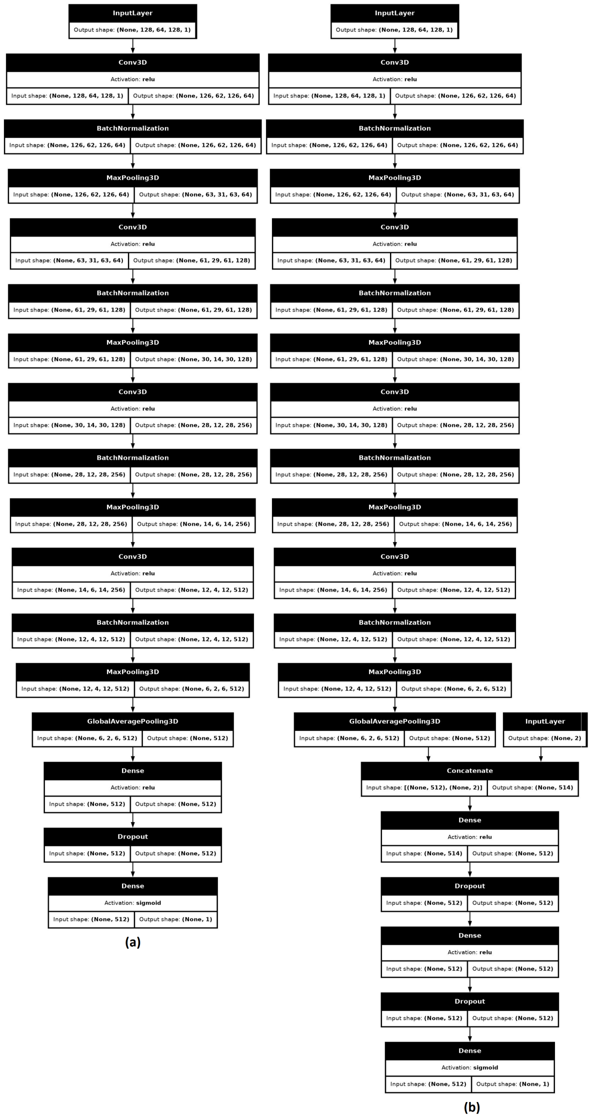

InputLayer
Output shape: (None, 128, 64, 128, 1)
Conv3D
Activation: relu
Input shape: (None, 128, 64, 128, 1)
Output shape: (None, 126, 62, 126, 64)
BatchNormalization
Input shape: (None, 126, 62, 126, 64)
Output shape: (None, 126, 62, 126, 64)
MaxPooling3D
Input shape: (None, 126, 62, 126, 64)
Output shape: (None, 63, 31, 63, 64)
Conv3D
Activation: relu
Input shape: (None, 63, 31, 63, 64)
Output shape: (None, 61, 29, 61, 128)
BatchNormalization
Input shape: (None, 61, 29, 61, 128)
Output shape: (None, 61, 29, 61, 128)
MaxPooling3D
Input shape: (None, 61, 29, 61, 128)
Output shape: (None, 30, 14, 30, 128)
Conv3D
Activation: relu
Input shape: (None, 30, 14, 30, 128)
Output shape: (None, 28, 12, 28, 256)
BatchNormalization
Input shape: (None, 28, 12, 28, 256)
Output shape: (None, 28, 12, 28, 256)
MaxPooling3D
Input shape: (None, 28, 12, 28, 256)
Output shape: (None, 14, 6, 14, 256)
Conv3D
Activation: relu
Input shape: (None, 14, 6, 14, 256)
Output shape: (None, 12, 4, 12, 512)
BatchNormalization
Input shape: (None, 12, 4, 12, 512)
Output shape: (None, 12, 4, 12, 512)
MaxPooling3D
Input shape: (None, 12, 4, 12, 512)
Output shape: (None, 6, 2, 6, 512)
GlobalAveragePooling3D
Input shape: (None, 6, 2, 6, 512)
Output shape: (None, 512)
Dense
Activation: relu
Input shape: (None, 512)
Output shape: (None, 512)
Dropout
Input shape: (None, 512)
Output shape: (None, 512)
Dense
Activation: sigmoid
Input shape: (None, 512)
Output shape: (None, 1)
(a)
InputLayer
Output shape: (None, 128, 64, 128, 1)
Conv3D
Activation: relu
Input shape: (None, 128, 64, 128, 1)
Output shape: (None, 126, 62, 126, 64)
BatchNormalization
Input shape: (None, 126, 62, 126, 64)
Output shape: (None, 126, 62, 126, 64)
MaxPooling3D
Input shape: (None, 126, 62, 126, 64)
Output shape: (None, 63, 31, 63, 64)
Conv3D
Activation: relu
Input shape: (None, 63, 31, 63, 64)
Output shape: (None, 61, 29, 61, 128)
BatchNormalization
Input shape: (None, 61, 29, 61, 128)
Output shape: (None, 61, 29, 61, 128)
MaxPooling3D
Input shape: (None, 61, 29, 61, 128)
Output shape: (None, 30, 14, 30, 128)
Conv3D
Activation: relu
Input shape: (None, 30, 14, 30, 128)
Output shape: (None, 28, 12, 28, 256)
BatchNormalization
Input shape: (None, 28, 12, 28, 256)
Output shape: (None, 28, 12, 28, 256)
MaxPooling3D
Input shape: (None, 28, 12, 28, 256)
Output shape: (None, 14, 6, 14, 256)
Conv3D
Activation: relu
Input shape: (None, 14, 6, 14, 256)
Output shape: (None, 12, 4, 12, 512)
BatchNormalization
Input shape: (None, 12, 4, 12, 512)
Output shape: (None, 12, 4, 12, 512)
MaxPooling3D
Input shape: (None, 12, 4, 12, 512)
Output shape: (None, 6, 2, 6, 512)
GlobalAveragePooling3D
Input shape: (None, 6, 2, 6, 512)
Output shape: (None, 512)
InputLayer
Output shape: (None, 2)
Concatenate
Input shape: [(None, 512), (None, 2)]
Output shape: (None, 514)
Dense
Activation: relu
Input shape: (None, 514)
Output shape: (None, 512)
Dropout
Input shape: (None, 512)
Output shape: (None, 512)
Dense
Activation: relu
Input shape: (None, 512)
Output shape: (None, 512)
Dropout
Input shape: (None, 512)
Output shape: (None, 512)
Dense
Activation: sigmoid
Input shape: (None, 512)
Output shape: (None, 1)
(b)

**FIGURE 10. Proposed 3D Convolutional Neural Networks (a) With Single Volumetric Input for pre-training (b) With additional input layer for Age and Gender to classify severity**

## B. 3D-EFFIBOT

In the search of open-source or publicly accessible models specifically fine-tuned for CT images an important one found was iBOT: Image BERT Pre-Training with Online Tokenizer [32] [33]. Initially pre-trained on the ImageNet-22K dataset iBOT ViT-L/16 was further fine-tuned for 35 epochs on a large set of 165,000 CT slices, representing about 4,000 patients from seven public datasets [34]. It should also be noted that iBOT ViT-L/16 was fine-tuned on axial view CT scan slices. By assuming that the iBOT ViT-L/16 with fine-tuned weights can be used as feature extractor it was further went into consideration that which group of slices should be selected to depict iBOT ViT-L/16 features for the complete 3D volumetric CT image. Finding a way to use iBOT ViT-L/16 features to represent the whole 3D volumetric CT images was a significant challenge since iBOT is confined to processing 2D slices. We chose to include 50% of the axial view slices that covered the maximum area in order to address this and provide a more representative selection. The number of slices were further reduced by half by eliminating neighboring slices that had almost identical information not only in terms of visual representation but also in terms of iBOT feature vector in order to prevent repeated behavior and excessive computations. Finally 32 slices were selected out of the original 128 slices and each slice was passed through iBOT to extract a feature vector of size [1, 1024]. After getting features of shape [32, 1024] for all slices respectively mean along first axis is calculated to combine the output of all the 32 slices and final shape obtained [1, 1024] for each patient respectively.

Although iBOT is a useful 2D feature extractor but it is unable to capture the 3D spatial connections present between slices which are visible in the sagittal and coronal views and are necessary for accurate CT evaluation. Therefore more research was done to find models capable of including 3D contextual information while prediction. Several 2D models pre-trained on ImageNet, such as VGG, ResNet, DenseNet, MobileNet, EfficientNet, and ConvNeXt were explored. These models along with their versions have been modified for 3D use by inflating their 2D ImageNet weights [35] [36]. Among the most promising alternatives were ConvNeXt-Tiny and EfficientNetV2-B3. The choice of EfficientNetV2-B3 was made because, according to the official Keras website [37] it had slightly greater Top-1 accuracy on the ImageNet validation dataset and a more easy to handle parameter count (14 million as opposed to 31 million) than ConvNeXt-Tiny. EfficientNetV2-B3's less computational complexity makes it more appropriate for dealing with 3D volumetric images in a time-efficient way.

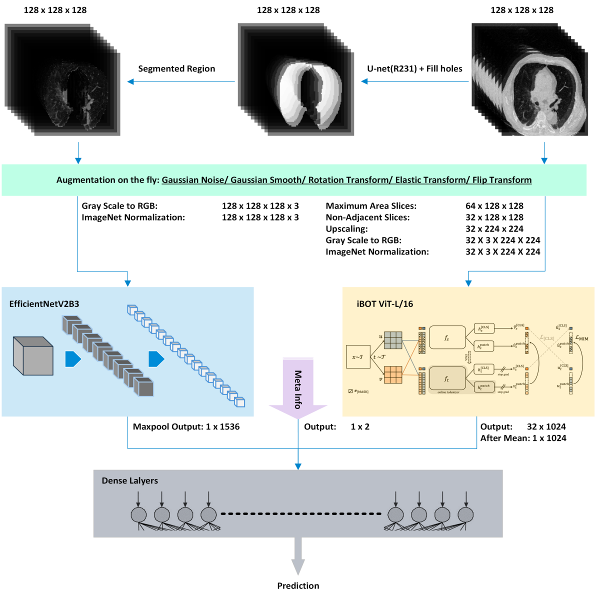

128 x 128 x 128
128 x 128 x 128
128 x 128 x 128
Segmented Region
U-net(R231) + Fill holes
Augmentation on the fly: Gaussian Noise/ Gaussian Smooth/ Rotation Transform/ Elastic Transform/ Flip Transform
Gray Scale to RGB: 128 x 128 x 128 x 3
ImageNet Normalization: 128 x 128 x 128 x 3
Maximum Area Slices: 64 x 128 x 128
Non-Adjacent Slices: 32 x 128 x 128
Upscaling: 32 x 224 x 224
Gray Scale to RGB: 32 X 3 X 224 X 224
ImageNet Normalization: 32 X 3 X 224 X 224
EfficientNetV2B3
iBOT ViT-L/16
Meta Info
Maxpool Output: 1 x 1536
Output: 1 x 2
Output: 32 x 1024
After Mean: 1 x 1024
Dense Lalyers
Prediction

FIGURE 11. Overview of the Training Cycle for 3D-EffiBOT Model

The proposed method 3D-EffiBOT combines the benefits of iBOT ViT-L/16 for 2D slice-level feature extraction with EfficientNetV2-B3's abilities to handle 3D contextual information. In order to improve further performance we applied a dynamic method to gradually unfreeze the top layers of the EfficientNetV2-B3 model during training which allows us for better fine-tuning on the testing dataset. On top of all we added a simple dense layer with 512 neurons followed by a dropout layer with a rate of 0.5 and a final dense layer with a single neuron with sigmoid activation function to predict severe COVID-19 outcomes which is defined as intubation or death within one month. Furthermore the model was trained using additional meta-information such as patient age and gender. The 3D-EffiBOT whole training cycle is represented in Fig. 11. This hybrid technique integrating iBOT ViT-L/16 and EfficientNetV2-B3 provides an optimal balance between performance and computational efficiency making it well-suited for clinical applications in severe COVID-19 prognosis from volumetric CT images. Initially lung masks for the lung cube dataset were obtained using a U-Net (R231) model [38] [39] followed by binary hole-filling to ensure the mask continuity. After that data augmentation techniques which are discussed earlier were applied to address the class imbalance in the dataset and to further improve model generalization. These augmentation techniques were applied on-the-fly during training such that each sample had a 50% probability of being augmented at each epoch. One of five augmentation techniques was randomly selected for each sample ensuring dynamic transformations across epochs. This strategy was designed to enhance the model's ability to generalize across diverse scenarios by presenting varied data representations to the model in every epoch.

Given that the iBOT ViT-L/16 model was fine-tuned on non-segmented 2D CT scan slices we opted to use non-segmented slices obtained from the augmented volumetric images. These slices were further shortlisted based on their maximum area followed by selection of non-adjacent slices to avoid redundancy. Later own those slices were up-scaled, represented in RGB format and normalized using ImageNet database statistics to align with iBOT's pretraining requirements. The feature vectors extracted from these augmented slices were plotted to visualize the variations introduced by different augmentation strategies on a single 3D volumetric image as shown in Fig. 12. On the other hand 3D EfficientNetV2-B3 whose weights were obtained by inflating 2D ImageNet weights was trained using segmented slices from the augmented volumetric images. Similar to iBOT the slices were represented in RGB format and normalized using ImageNet database statistics. A dynamic fine-tuning strategy was employed for EfficientNetV2-B3, starting with the top 20 layers being unfrozen and gradually increasing to 100 layers in increments of 20. The initial phase of training lasted for 10 epochs with all layers frozen, followed by 10 epochs of fine-tuning for each step as the top layers were progressively unfrozen.

The dataset for this study was divided into five stratified folds with 80% allocated for training and 20% for testing in each fold. For the task of severity classification, 1205 CT volumetric images labeled as COVID-19 positive were used which includes 904 images indicating the need for intubation and 301 images where death occurred within one month. Training was conducted over a total of 60 epochs with the first 10 epochs dedicated to transfer learning and the remaining 50 epochs allocated for dynamic fine-tuning. The model was trained using a batch size of 4 with the BinaryFocalCrossentropy loss function and optimized with the AdamW optimizer. The learning rate was dynamically adjusted starting at 0.00001 and decaying exponentially with a decay step of 500 and a rate of 0.96 following a staircase pattern. Additionally, a weighted loss strategy within the BinaryFocalCrossentropy function was adopted to mitigate the effects of class imbalance avoiding the need for oversampling which would increase computational costs per epoch.

This training strategy aimed to maximize model performance while ensuring robust generalization to unseen testing data. By dynamically augmenting data on the fly and controlling for class imbalance using weighted loss strategy the model was designed to handle the complexities of predicting severe COVID-19 outcomes from CT scans, such as intubation and death.

## V. EXPERIMENTAL RESULTS

This section focuses on the experimental results meant to predict from a computed tomography (CT) scan whether subjects had a serious COVID-19 infection which is described as intubation or death within one month using STOIC dataset. The STOIC dataset which was used to evaluate the pre-trained models is the main dataset for this study as MosMedData was only used for pretraining in 3D-CNN. Performance measures used to rate the model are Precision, Recall, F1 score and the Area Under the plot (AUC) of the Receiver Operating Characteristic (ROC) plot. The main goal of the STOIC2021 COVID-19 AI Challenge was to identify that which patients will have serious COVID-19 which is described as intubation or death within one month after the CT scan's collection. The goal was to measure model performance mainly using the Receiver Operating Characteristic (ROC) curve's Area Under the Curve (AUC). The AUC score is crucial for assessing how well the model can distinguish between highly and low severity patients. It also allows a simple comparison of the performance of the proposed model with other researchers' models. This comparison helps validate the effectiveness of the proposed model by measuring it against a benchmark that the challenge participants set.

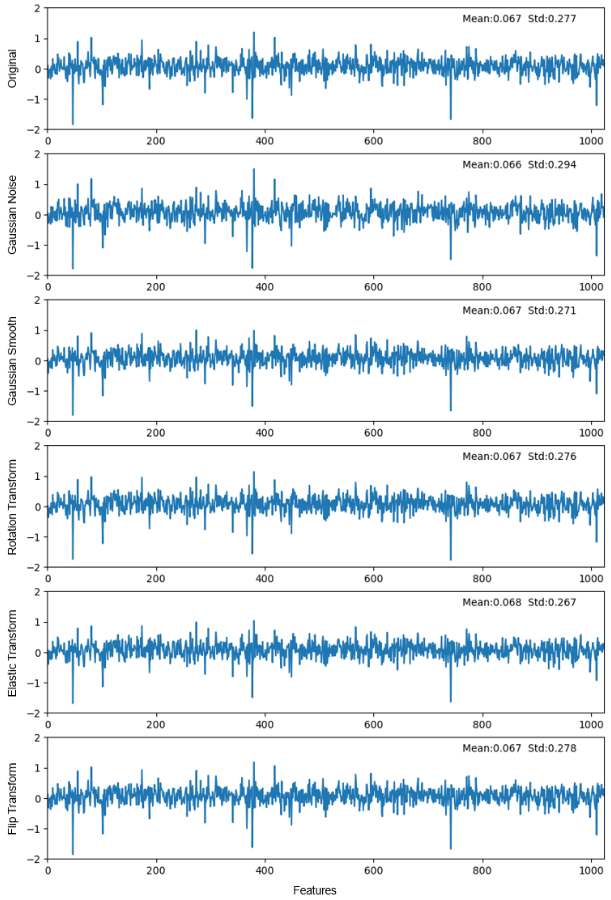


**FIGURE 12. iBOT Feature vectors visualizing the effects of different augmentation strategies on a single 3D volumetric image**

### *A. 3D-CNN RESULTS*

Several significant insights into the model's performance are revealed by the Fig. 13 and 14 in this research. First, the tiny learning rate of 0.00001, which declines further in a staircase pattern, is responsible for the steady rise in both training and validation accuracy from the first epoch onward. Given that the model was pre-trained, which supplied starting weights and biases that aided in early categorization, this gradual rise in accuracy is to be anticipated. The accuracy stabilized as the training went on, especially between epochs 20 and 25, showing that the model was successfully learning from the data. The loss graphs similarly show this stability, with the mean validation loss decreasing across all folds throughout the same time frame. This suggests the model had reached a saturation limit, beyond which more training would cause over-fitting. In order to avoid over-fitting and guarantee that the model retained its generalization ability, training was

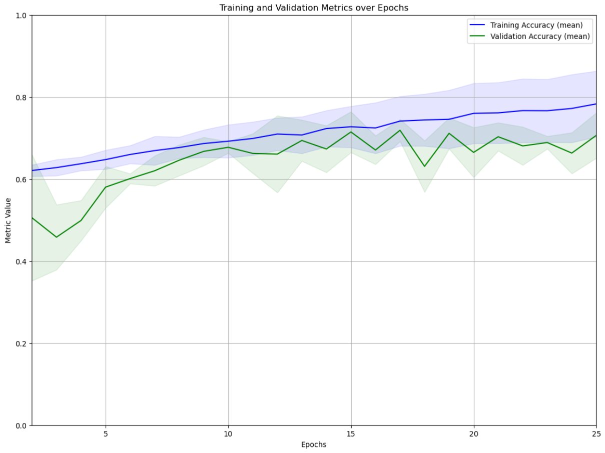


**FIGURE 13.** **3D-CNN Training and Validation Accuracy**

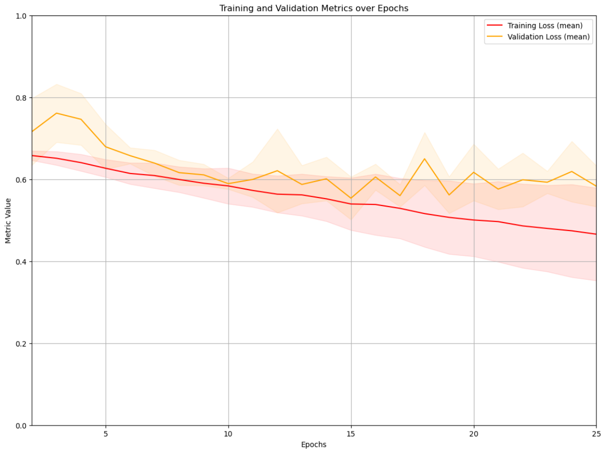


**FIGURE 14.** **3D-CNN Training and Validation Loss**

terminated after 25 epochs.

Accuracy by itself could be deceptive given the notable imbalance in the original dataset. In order to address this, extra metrics were computed for each fold throughout the five-fold cross-validation, as shown in Table 1, including AUC, Precision, Recall, and F1 Score. The research also emphasizes how data imbalance affects the performance of the model. Although the training data was balanced to decrease bias toward the negative class, the validation dataset remained uneven, which is visible in the confusion matrix (Fig. 15).

**TABLE 1.** **3D-CNN Scores For Each Fold**

| Fold | AUC Severity | Precision | Recall | F1 Score |
|---|---|---|---|---|
| 1 | 0.8025 | 0.4787 | 0.7500 | 0.5844 |
| 2 | 0.7722 | 0.5211 | 0.6167 | 0.5649 |
| 3 | 0.6779 | 0.3878 | 0.6333 | 0.4810 |
| 4 | 0.7540 | 0.4388 | 0.7167 | 0.5443 |
| 5 | 0.7005 | 0.4045 | 0.5902 | 0.4800 |
| Mean | 0.7414 | 0.4462 | 0.6614 | 0.5309 |

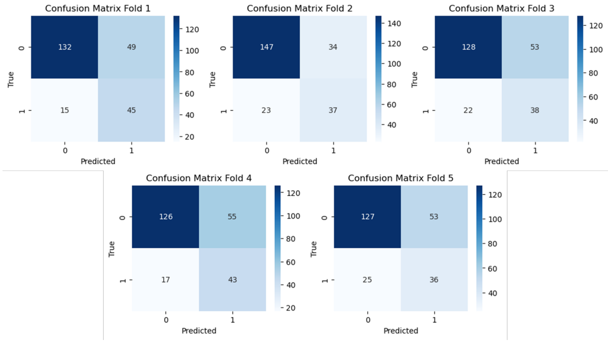


**FIGURE 15.** **3D-CNN Confusion Matrix**

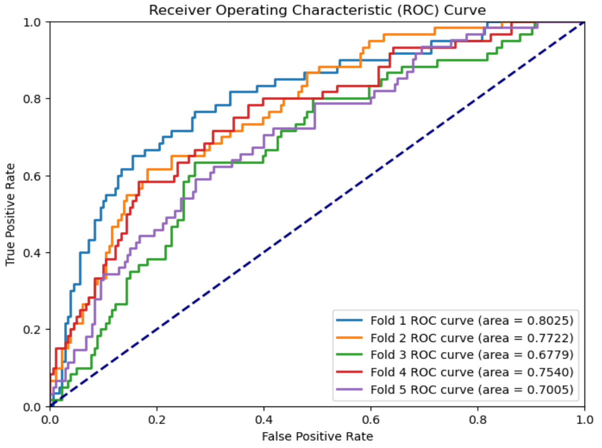


**FIGURE 16.** **3D-CNN ROC Curve**

Due to the smaller number of samples in the severe class that resulted in death as opposed to those who required intubation within a month, the real positive rate was lower than the true negative rate. The AUC for each fold is further shown by the ROC curve (Fig 16), where Fold 1 once again displays the largest area.

## B. 3D-EFFIBOT RESULTS

The dataset was divided into five stratified folds with 80% allocated for training and 20% for testing. Fig. 17 illustrates the training accuracies and loss curves for all folds as no separate validation dataset was used during full model training. Initial experiments using a validation set for a few epochs provided insights into the optimal number of epochs required to train both the transfer learning and fine-tuning components of the models. Based on those experiments it was later decided to limit the training to a maximum of 60 epochs such that for the first 10 epochs only the top dense layers were trained with all other layers frozen hence facilitating effective transfer learning. Afterward layers were unfrozen incrementally starting with 20 layers for each subsequent 10-epoch block reaching a total of 100 unfrozen layers (approximately 25% of the total layers in the pre-trained EfficientNetV2 model).

Given that EfficientNetV2 and iBOT were both pre-trained

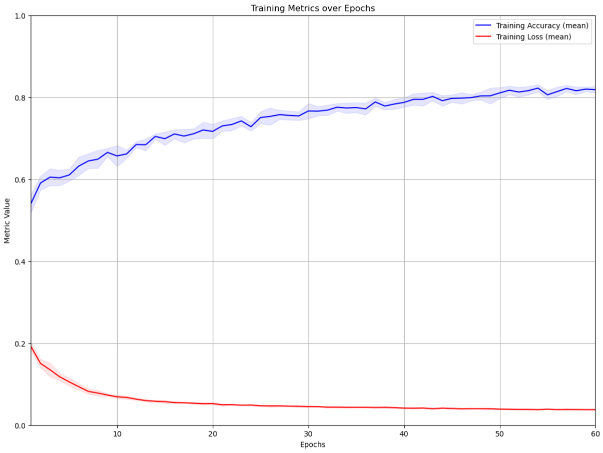


**FIGURE 17.** 3D-EffiBOT Training and Validation Accuracy

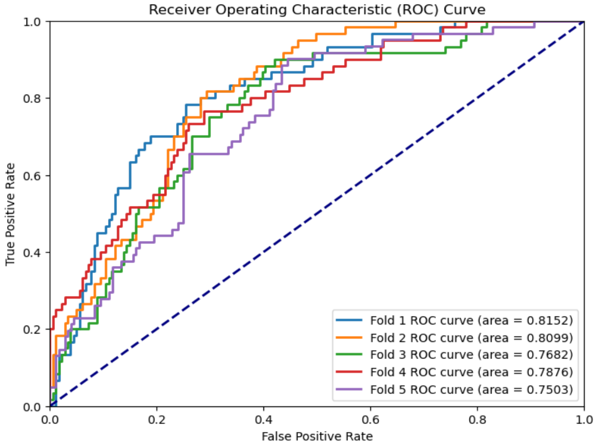


**FIGURE 19.** 3D-EffiBOT ROC Curve

**TABLE 2.** 3D-EffiBOT Scores For Each Fold

| Fold | AUC Severity | Precision | Recall | F1 Score |
|---|---|---|---|---|
| 1 | 0.8152 | 0.5054 | 0.7833 | 0.6143 |
| 2 | 0.8099 | 0.4848 | 0.8000 | 0.6037 |
| 3 | 0.7682 | 0.4607 | 0.6833 | 0.7202 |
| 4 | 0.7876 | 0.4835 | 0.7333 | 0.5827 |
| 5 | 0.7503 | 0.4512 | 0.6066 | 0.5037 |
| Mean | 0.7862 | 0.5663 | 0.7213 | 0.6049 |

models (iBOT was fine-tuned too on 2D CT scans slices) no additional pre-training was necessary using others datasets like MosMedData. The gradual learning rate decay starting from 0.00001 and decreasing in a staircase manner contributed to the steady rise in training accuracy as shown in Fig. 17. The pre-trained weights and biases also facilitated early and accurate categorization boosting initial training performance.

The results presented in Table 2 demonstrate the superiority of 3D-EffiBOT hybrid architecture by combining EfficientNetV2 and iBOT over 3D-CNN, with an approximate 4% increase in AUC score. The average AUC across all five folds was 0.7862 as compared to 0.7414 achieved by 3D-CNN. This improvement highlights the efficacy of using the hybrid model though the data still lacked critical patient pre-COVID medical history which could potentially enhance the model's understanding of patient immunity and comorbidity. Incorporating such information in future research could lead to more accurate prognosis. Fig. 18 underscores the dataset imbalance which motivated the use of a weighted loss strategy within the BinaryFocalCrossentropy function. The AUC scores followed a similar pattern across folds suggesting consistent stratified fold distribution in both 3D-CNN and 3D-EffiBOT with Fold 1 achieved the highest AUC. This pattern is further confirmed by the ROC curves shown in Fig. 19 where Fold 1 has the largest area under the curve. The overall results suggest that while the 3D-EffiBOT yields significant improvements over 3D-CNN further refinement particularly through the inclusion of more comprehensive patient metadata could lead to even better performance.

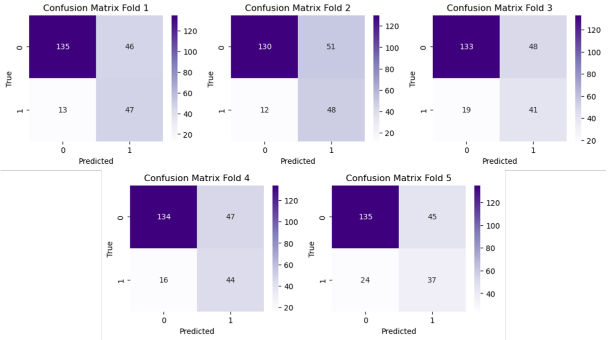
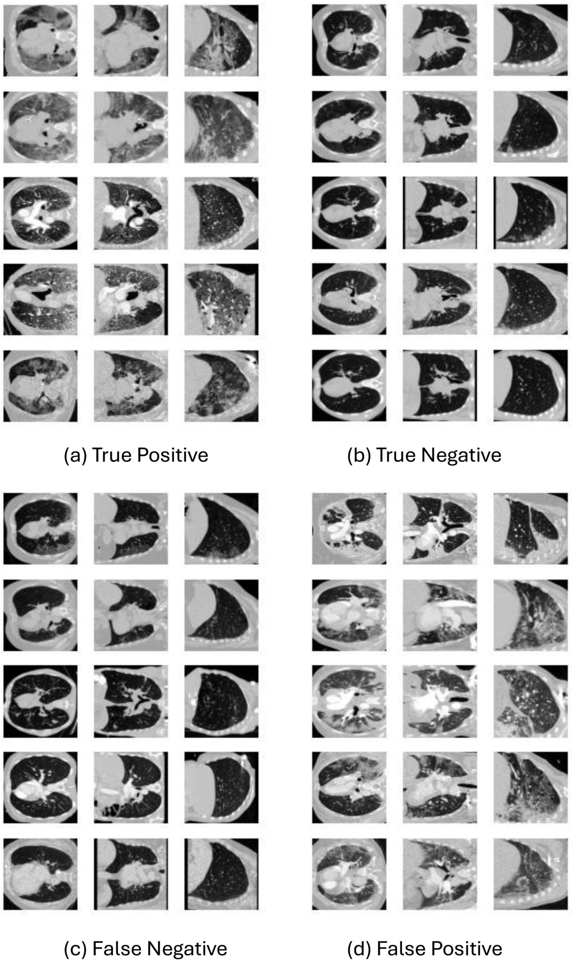


**FIGURE 18.** 3D-EffiBOT Confusion Matrix

### *C. COMPREHENSIVE ANALYSIS OF RESULTS AND THEIR IMPLICATIONS*

In order to increase model accuracy and generality, this study highlights how crucial it is that other factors such as patient medical history should be taken into account in further studies. In addition to the results in metrics, Fig. 20 clearly states that there is dire need of patient pre-covid health history for consideration during prediction of death in 1 month as true positive images clearly states the presence of fluid in lungs due to extreme severity which lead to death, but also false positive images predicted by the model clearly display the excessive abnormality present in the lungs, these abnormalities could be lung cancer or other lung disease, if not COVID extreme severity as labeled by the researchers. This same phenomena can also be seen in True Negative and False Negative COVID severe patients. As both show similar lung views. In conclusion of STOIC challenge it was observed that some patients died due to other health issues in addition to COVID as pre-covid weak immune system in the body was responsible for these deaths. This factor if included by the researchers while preparation of vast dataset,


(a) True Positive

(b) True Negative

(c) False Negative

(d) False Positive

**FIGURE 20. Examples of 3D CT scans which are correctly or incorrectly classified**

**TABLE 3. Qualification (Last Submission) Leader-board [40]**

| Positon | User(Team) | AUC Severity |
|---|---|---|
| 1st | simon.j | 0.8044 |
| 2nd | lorenjul (Code 1055) | 0.7868 |
| 3rd | titericz | 0.7837 |
| 4th | etro | 0.7752 |
| 5th | miriamelia (uaux2) | 0.7662 |
| Proposed | 3D-CNN | 0.7414 |
| Proposed | 3D-EffiBOT | 0.7862 |

will not only help future analyst for accurate severity and death prediction but also help other researchers to develop AI that could analyse 3D CT scans by going more deep into the features for broader disease prediction, This would be great as many diseases shares same kind of symptoms in human body.

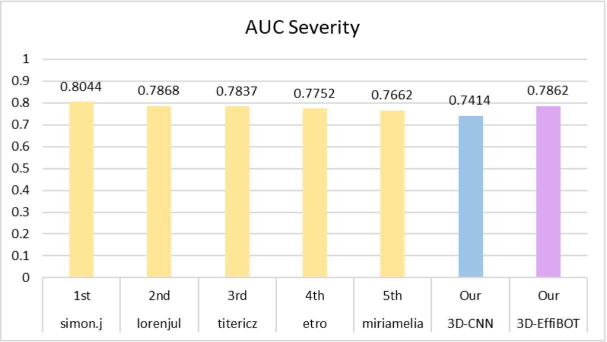


**FIGURE 21. AUC Severity Comparison Chart with participants**

Table 3 and Fig. 21 compare Qualification Leader-board AUC Severity displayed on website [40]. These scores are for reference to compare proposed model results. Although the results shown in Table 3 of participants were computed on different testing dataset which is not public yet (see Fig. 2) unlike our proposed methods 3D-CNN and 3D-EffiBOT which are evaluated on public data set after stratified 5-fold sampling. As T3 approach [30] is implemented during this STOIC challenge, therefore the Qualification Leader-board scores were collected and computed after Last submission round of the challenge in qualification phase. This means that these results of all participants were reported after training on public dataset and testing on private dataset Test set A2 (see Fig. 2). Meanwhile our proposed method were solely relies on public dataset Training set A, therefore in our case availability of training data was less as compared to the availability of training dataset to the participants at the time of STOIC competition. Please also note that there is also Final Leader-bard available on the website [40], scores of which are slightly higher than that of Qualification Leader Board, but this leader board is obtained after completing T3 approach which means all the qualified participants selected from the Qualification round were further supplied with another huge size of training dataset privately (Training set B of 9724 3D CT images) therefore final scores were increased by 1.1% to that qualification round scores.

In conclusion, as this research solely relies on small amount of training dataset (Training set A of 2000 3D CT images) out of large STOIC Database of 10735 3D CT images, therefore there is always a chance to further increase the model performance by including patients health history as immune system of the human body varies from patient to patient due many factors like age, gender, genetic issues, ill due to other disease and region specific conditions associated with patient.

## VI. CONCLUSIONS

In conclusion, this research presents a comprehensive approach to predicting severe COVID-19 infection from CT scans using 3D-CNN and 3D-EffiBOT. By leveraging transfer learning and fine-tuning, we were able to enhance model generalization and achieve early accuracy improvements. The segmentation and preprocessing techniques employed, combined with the use of five augmentation strategies, addressed the challenges posed by imbalanced severity classes and variations in raw CT data. The final models after training with a carefully scheduled learning rate demonstrated a robust performance with mean AUC of 0.7414 and 0.7862 for 3D-CNN and 3D-EffiBOT respectively. This underscores the effectiveness of CNNs and VITs hybrid architecture in capturing volumetric information providing a significant advantage. Our findings highlight the potential of using hybrid models in clinical applications for early prediction of severe COVID-19 outcomes, thereby offering valuable insights for future research and development in medical image analysis.

Looking ahead, future work will focus on enhancing the robustness and scalability of our 3D models for predicting severe COVID-19 infections from CT scans. One area of exploration is the integration of additional clinical data, such as laboratory test results and patient medical histories, to further improve prediction accuracy and provide a more holistic view of patient health. More complex data augmentation methods and sophisticated transfer learning approaches may also be able to improve model generalization and alleviate class imbalance problems. Exploring the potential of ensemble learning approaches, where many models are integrated to increase prediction performance, is another intriguing area. Furthermore, broadening the dataset to include a more heterogeneous population from various geographic regions may enhance the model's suitability and dependability for varying patient demographics. Ultimately, the development of user-friendly interfaces and the implementation of real-time prediction capabilities will be crucial stages in converting our study into useful, actionable tools that healthcare providers may use to manage COVID-19 and perhaps other respiratory disorders.

## REFERENCES

[1] H. Hassan, Z. Ren, H. Zhao, S. Huang, D. Li, S. Xiang, Y. Kang, S. Chen, and B. Huang, "Review and classification of ai-enabled covid-19 ct imaging models based on computer vision tasks," *Computers in Biology and Medicine*, vol. 141, pp. 105 123–105 123, 2 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521009173

[2] M. Jalali Moghaddam and M. Ghavipour, "Towards smart diagnostic methods for covid-19: Review of deep learning for medical imaging," *IPEM-Translation*, vol. 3-4, p. 100008, 11 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667258822000061

[3] T. Zhou, F. Liu, H. Lu, C. Peng, and X. Ye, "A review of deep learning imaging diagnostic methods for covid-19," *Electronics*, vol. 12, pp. 1167–1167, 02 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/5/1167

[4] A. S. Althenayan, S. A. AlSalamah, S. Aly, T. Nouh, and A. A. Mirza, "Detection and classification of covid-19 by radiological imaging modalities using deep learning techniques: A literature review," *Applied Sciences*, vol. 12, pp. 10 535–10 535, 10 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/20/10535

[5] "Hugging face – the ai community building the future." Huggingface.co, 01 2022. [Online]. Available: https://huggingface.co/datasets?sort=trending&search=covid19

[6] "Find open datasets and machine learning projects | kaggle," Kaggle.com, 2024. [Online]. Available: https://www.kaggle.com/datasets?search=covid19&tags=12107-Computer+Science

[7] M. Alam, M. U. Akram, and W. Fareed, "Deep learning-based analysis and classification of covid patients through ct images," 02 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10136653

[8] C. Soler-Luna, D. Reynoso-Saldana, M. I. Burgos, and C. H. Gutierrez, "Unexpected ground-glass opacities on abdominopelvic ct of a patient with a negative sars-cov-2 antigen test result and no respiratory symptoms upon admission," *Cureus*, 10 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7575311/

[9] "Pneumonia ct - wikidoc," Wikidoc.org, 2022. [Online]. Available: https://www.wikidoc.org/index.php/Pneumonia_CT

[10] "Figure 8: A newborn with the typical "crazy-paving" appearance of..." ResearchGate, 2021. [Online]. Available: https://www.researchgate.net/figure/A-newborn-with-the-typical-crazy-paving-appearance-of-pulmonary-alveolar-proteinosis_fig4_348715362

[11] D. J. Bell, "Halo sign (chest)," Radiopaedia, 2022. [Online]. Available: https://radiopaedia.org/articles/halo-sign-chest-3

[12] V. N. Maturu and R. Agarwal, "Reversed halo sign: A systematic review," *Respiratory Care*, vol. 59, pp. 1440–1449, 04 2014. [Online]. Available: https://rc.rcjournal.com/content/59/9/1440

[13] U. Themes, "Pleural effusion," Radiology Key, 07 2019. [Online]. Available: https://radiologykey.com/pleural-effusion-2/

[14] A. Thyagachandran, A. Balachandran, and H. A. Murthy, "Identification and severity assessment of covid-19 using lung ct scans," *IEEE Access*, vol. 11, pp. 124 542–124 555, 01 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10309131

[15] M. S. P, A. A. E, P. N. A, V. A. V, L. N. V, G. V. A, B. I. A, G. P. B, G. A. V, and C. V. Yu, "Mosmeddata: Chest ct scans with covid-19 related findings dataset," arXiv.org, 2020. [Online]. Available: https://arxiv.org/abs/2005.06465

[16] MedSeg, H. B. Jenssen, and T. Sakinis, "Medseg covid dataset 1," *figshare*, 01 2021. [Online]. Available: https://figshare.com/articles/dataset/MedSeg_Covid_Dataset_1/13521488/2

[17] S. Heidarian, P. Afshar, N. Enshaei, F. Naderkhani, M. J. Rafiee, A. Oikonomou, A. Shafiee, F. B. Fard, K. N. Plataniotis, and A. Mohammadi, "Spgc-covid dataset," *figshare*, 09 2021. [Online]. Available: https://figshare.com/articles/dataset/SPGC-COVID_Dataset/16632397/1

[18] S. Heidarian, "Covid-19 low-dose and ultra-low-dose ct scans," IEEE DataPort, 05 2021. [Online]. Available: https://ieee-dataport.org/open-access/covid-19-low-dose-and-ultra-low-dose-ct-scans

[19] K. Morani and D. Unay, "Deep learning-based automated covid-19 classification from computed tomography images," *Computer Methods in Biomechanics and Biomedical Engineering Imaging & Visualization*, vol. 11, pp. 2145–2160, 06 2023. [Online]. Available: https://arxiv.org/abs/2111.11191

[20] D. Kollias, A. Arsenos, L. Soukissian, and S. Kollias, "Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis," arXiv.org, 2021. [Online]. Available: https://arxiv.org/abs/2106.07524

[21] M. Alwan, T. Hasan, and Z. Kamil, "Gta 3d-dld: Greedy training approach for 3d deep learning diagnosis based covid-19 ct scan," ResearchGate, 12 2022. [Online]. Available: https://www.researchgate.net/publication/366696862_GTA_3D-DLD_Greedy_Training_Approach_for_3D_Deep_Learning_Diagnosis_Based_COVID-19_CT_Scan

[22] PlamenEduardo, "Sars-cov-2 ct-scan dataset," Kaggle.com, 2020. [Online]. Available: https://www.kaggle.com/datasets/plameneduardo/sarscov2-ctscan-dataset

[23] D. Kienzle, J. Lorenz, R. Schön, K. Ludwig, and R. Lienhart, "Covid detection and severity prediction with 3d-convnext and custom pretrainings," arXiv.org, 2022. [Online]. Available: https://arxiv.org/abs/2206.15073

[24] "A novel study for automatic two-class and three-class covid-19 severity classification of ct images using eight different cnns and pipeline algorithm | adcaij: Advances in distributed computing and artificial intelligence journal," Usal.es, 2022. [Online]. Available: https://revistas.usal.es/cinco/index.php/2255-2863/article/view/28715/30170

[25] S. Atito, M. C. Yavuz, M. U. Şen, F. Gülşen, O. Tutar, B. Korkmazer, C. Samancı, S. Şirolu, R. Hamid, A. E. Eryürekli, T. Mammadov, and B. Yanikoglu, "Comparison and ensemble of 2d and 3d approaches for covid-19 detection in ct images," *Neurocomputing*, vol. 488, pp. 457–469, 02 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523122200162X

[26] W. Tan and J. Liu, "A 3d cnn network with bert for automatic covid-19 diagnosis from ct-scan images," *arXiv (Cornell University)*, 10 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9607390

[27] J. Hou, J. Xu, R. Feng, Y. Zhang, F. Shan, and W. Shi, "Cmc-cov19d: Contrastive mixup classification for covid-19 diagnosis," 10 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9607776

[28] H. Zunair, A. Rahman, N. Mohammed, and J. P. Cohen, "Uniformizing techniques to process ct scans with 3d cnns for tuberculosis prediction," arXiv.org, 2020. [Online]. Available: https://arxiv.org/abs/2007.13224

[29] "Study of thoracic ct in covid-19: The stoic project | radiology," Radiology, 2023. [Online]. Available: https://pubs.rsna.org/doi/full/10.1148/radiol.2021210384

[30] L. H. Boulogne, J. Lorenz, D. Kienzle, R. Schön, K. Ludwig, R. Lienhart, S. Jégou, G. Li, C. Chen, Q. Wang, D. Shi, M. Maniparambil, D. Müller, S. Mertes, N. Schröter, F. Hellmann, M. Elia, I. Dirks, M. N. Bossa, A. D. Berenguer, T. Mukherjee, J. Vandemeulebroucke, H. Sahli, N. Deligiannis, P. Gonidakis, N. D. Huynh, I. Razzak, R. Bouadjenek, M. Verdicchio, P. Borrelli, M. Aiello, J. A. Meakin, A. Lemm, C. Russ, R. Ionasec, N. Paragios, B. van Ginneken, and M.-P. Revel, "The stoic2021 covid-19 ai challenge: Applying reusable training methodologies to private data," *Medical Image Analysis*, vol. 97, p. 103230, 10 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841524001555

[31] "Fig. 3. hounsfield unit scale, mapping values to the represented..." ResearchGate, 2023. [Online]. Available: https://www.researchgate.net/figure/Hounsfield-Unit-scale-mapping-values-to-the-represented-tissue-and-an-axial-slice-of-a_fig2_369975157

[32] J. Zhou, C. Wei, H. Wang, W. Shen, C. Xie, A. Yuille, and T. Kong, "ibot: Image bert pre-training with online tokenizer," arXiv.org, 2021. [Online]. Available: https://arxiv.org/abs/2111.07832

[33] "bytedance/ibot: ibot :robot:: Image bert pre-training with online tokenizer (iclr 2022)," GitHub, 2022. [Online]. Available: https://github.com/bytedance/ibot

[34] S. Jégou, "Weights of two vit-l models," *Zenodo*, 05 2022. [Online]. Available: https://zenodo.org/records/6547999

[35] R. Solovyev, A. A. Kalinin, and T. Gabruseva, "3d convolutional neural networks for stalled brain capillary detection," *Computers in Biology and Medicine*, vol. 141, pp. 105 089–105 089, 11 2021. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0010482521008830?via%3Dihub

[36] "Zfturbo/classification_models_3d: Set of models for classifcation of 3d volumes," GitHub, 04 2022. [Online]. Available: https://github.com/ZFTurbo/classification_models_3D

[37] K. Team, "Keras documentation: Keras applications," Keras.io, 2024. [Online]. Available: https://keras.io/api/applications/#usage-examples-for-image-classification-models

[38] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental*, vol. 4, 08 2020. [Online]. Available: https://eurradiolexp.springeropen.com/articles/10.1186/s41747-020-00173-2

[39] "Johof/lungmask: Automated lung segmentation in ct," GitHub, 04 2024. [Online]. Available: https://github.com/JoHof/lungmask

[40] "Stoic2021 - covid-19 ai challenge - grand challenge," grand-challenge.org, 2021. [Online]. Available: https://stoic2021.grand-challenge.org/evaluation/quallification-last-submission/leaderboard/

**MOHSIN ALI KHAN** received the B.S. in electrical engineering from University Of Central Punjab (UCP), Lahore, Pakistan and M.S. degree in computer engineering from Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan. His research interests includes deep learning, machine learning, digital image processing and medical image analysis.

**ARSLAN SHAUKAT** received the B.S. and M.S. degrees in computer engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2003 and 2005, respectively, and the Ph.D. degree in computer science from The University of Manchester, U.K., in 2010. He is currently an Associate Professor with the Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, NUST (CEME NUST). He has published various research papers in refereed journals and conference proceedings. His research interests include machine learning, pattern recognition, digital image, and speech processing. He has been a member of technical program committees of numerous international conferences and a reviewer of international journals. He was a recipient of academic awards, including the Best Teacher Award in 2018 and the Best Research Paper Award in 2019.

**ZARTASHA MUSTANSAR** received the Ph.D. degree from The University of Manchester, U.K. She was selected by Microsoft Research Cambridge (MSR) to pursue research in physical sciences and engineering in Manchester. She is currently employed as an Assistant Professor with the Research Center for Modeling and Simulation (RCMS), NUST. She has published 32 research papers in various peer-reviewed journals and conferences. Her research interest includes biomechanical engineering, especially associated with health care.

**MUHAMMAD USMAN AKRAM** (Senior Member, IEEE) received the B.S. degree (Hons.) in computer system engineering and the master's and Ph.D. degrees in computer engineering from the College of Electrical and Mechanical Engineering, National University of Sciences and Technology (NUST), Rawalpindi, Pakistan, in 2008, 2010, and 2012, respectively. He is currently as Professor with the College of Electrical and Mechanical Engineering, NUST. He has over 200 international publications in well reputed journals and conferences. His main areas of research interests include biomedical imaging and image processing.