International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *14 October 2023* ***Accepted:*** *24 October 2023*

Bidirectional Encoder Representations Transformers for Improving CNN-LSTM Covid-19 Disease Detection Classifier

Jinu Paulson Siluvai Rathinam[1], Angeline Prasanna Gopalan[2]
[1]Department of Computer Science,
AJK College of Arts and Science,
Coimbatore 641105, India.
e-mail: jinuplsn7phd@gmail.com
[2]Department of Computer Science,
AJK College of Arts and Science,
Coimbatore 641105, India.
e-mail: angelineprasanna@gmail.com

Abstract—Early identification of COVID-19 diseased persons is crucial to avoid and prevent the transmission of the SARS-CoV-2 virus. To achieve this, lung Computed Tomography (CT) scan segmentation and categorization models have been broadly developed for COVID-19 diagnosis. Amongst, Multi-Scale function learning with an Attention-based UNet and Marginal Space Deep Ambiguity-attentive Transfer Learning (MS-AUNet-MSDATL) framework is developed to concurrently segment the COVID-19 infected regions and classify their risk levels from the CT/Chest X-Ray (CXR) scans. This model utilizes Convolutional Neural Network–Long Short Term Memory (CNN-LSTM) as a classifier for proper recognition. Although CNN-LSTM efficiently learns the spatial and temporal data, but it highly ignores the pixels and their adjacent information which results in lower classification rate. So, in this paper, Bidirectional Encoder Representations from Transformers (BERT) is introduced along with CNN-LSTM in MS-AUNet-MSDATL model to resolve the above mentioned issues for efficient COVID-19 risk level classification. Initially, the segmented CT and CXR images from MS-AUNet is given as input to the BERT model. BERT structure consist of stack of transformer encoder layers which extracts fixed features from a pre-trained models to obtain the numerical representation of the given image. Then, the numerical expressions from the BERT are transformed to pre-learned CNN model for selecting the important features from the given representations. The LSTM model receives the CNN output and generates a new representation based on the data order. In addition, Fully Connected Layer (FCL) maps the CNN-LSTM results into categorization classes to learn pixels and their adjacent information for COVID-19 risk level detection and diagnosis. The complete work is termed as MS-AUNet-EMSDATL. Finally, the test findings shows that the MS-AUNet-MS-B-DATL achieves accuracy of 98.67% and 98.55% on CT and CXR images compared to the other existing frameworks.

Keywords- COVID-19, Bidirectional Encoder, Convolutional Neural Network, Long Short Term Memory, Fully Connected Layers.

I. INTRODUCTION

The outbreak of respiratory infections that started in 2020 was caused by a new coronavirus illness (COVID-19), which initially appeared and scattered fast around the world. As a result of this sickness, which seriously endangers a person's overall health risk, society is currently having a huge and devastating global medical crisis [1]. Pulmonary symptoms are the main cause of COVID-19-related rheumatic fever. It may also lead to enteric pathogens, which can bring on gastrointestinal issues including nausea, vomiting and diarrhoea [2].

According to projections from the World Health Organization (WHO), there were 6.51 million fatal cases of new coronary pneumonia worldwide as of September 26, 2022, with 612,236,677 events having been reported [3]. Because COVID-19 is so contagious, prompt identification of sufferers may aid in limiting the virus' development. Therefore, a timely and accurate COVID-19 diagnosis is crucial for the rapid detection of this infection [4]. The most effective method for COVID-19 detection is observation utilizing real-time reverse transcription polymerase chain reaction (RT-PCR). Using samples of oropharyngeal tissue, saliva and nasal swabs, it may infer the presence of COVID-19 [5]. However, it has a significant false negative rate when identifying early stage conditions.

Meanwhile, COVID19 symptoms as observed by medical imaging procedures like CT and CXR, display specific features that diverge from healthy instances or additional forms of pneumonia [6, 7]. The traditional diagnosis of irregularities by doctors necessitates a significant amount of time and is greatly impacted by their judgment, hence it is crucial to research and develop an effective and computerized system for clinical visual classification. It can support medical professionals in making accurate real-time diagnoses and treatment choices. Deep

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *14 October 2023* ***Accepted:*** *24 October 2023*

learning approaches were incredibly successful at separating the required Region-Of-Interest (ROI), like healthy and unhealthy pulmonary areas, from CT scans [8]. The correct delineation of the COVID-19 disease patch using deep learning approaches is vitally pertinent for quick detection and estimation by physicians [9].

For the efficient COVID-19 CT images segmentation, an Attention-based U-Net (AUNet) model [10] was constructed to reweight the characteristic representation and capture the rich contextual features. The spatial and channel attention units in U-Net apply an attention strategy. To capture properties at different dimensions, a residual unit with dilated convolutions was also used. Additionally, the tiny irregular patches in the CT scans were segmented using the focal Tversky error. In contrast, the lung CT images' segregation of unclear edges was unsatisfactory. Additionally, the tiny irregular patches in the CT scans were segmented using the focal Tversky error. In contrast, the lung CT images' segregation of unclear edges was unsatisfactory.

So, MS-AUNet-MSDATL framework [11] is developed to simultaneously segregate the COVID-19 infected ROIs and classify the disease risk levels from the CT/X-ray scans. In this framework, Multi-Scale function learning with AUNet is initially constructed to extract features at various locations. Then, an improved filter MSRF is added to the attention-based U-Net to improve segmentation effectiveness and acquire structural information from the CT images. Next, marginal learning of the bounding box variables is minimized into sub-spaces to detect the target tissues. Finally, the segmented ROIs from the MS-AUNet structure are given to the different pre-learned CNN models like VGG16, ResNet50, InceptionResNetV2 and DenseNet121 structures. These pre-learned models can hierarchically capture more informative and discriminative characteristics from the lung CT/X-ray scans. Those characteristics are provided to the CNN-LSTM classifier to classify the disease risk levels for proper diagnosis. Moreover, the epistemic ambiguity of categorization outcomes is measured to determine areas where the learning frameworks are not optimistic regarding their decisions. Thus, the measured ambiguities deliver useful data regarding where and how much the physician could believe the classifier forecasts for COVID-19 recognition and diagnosis.

In attempt to resolve the above issues, this article adopts for transformer based model i.e BERT. BERT is developed based on DL models to train a bidirectional expression which effectively explores the information on the left and right sides of each layer concurrently. As a consequence, the pre-trained BERT model are finely calibrated with one additional layers for the optimal classification results. So considering the advantages of BERT model, it is integrated with MS-AUNet-MSDATL to learn the pixel and their adjacent information effectively for efficient COVID-19 risk level classification and diagnosis.

In MS-AUNet-EMSDATL, the segmented CT and CXR images from MS-AUNet-MSDL are fed into BERT model. BERT structure consist of stack of transformer encoder layers which extracts fixed features from a pre-trained models to obtain the numerical representation of the given image. Then, the numerical expressions from the BERT are translated to the CNN model to choose the most significant features from the considered representations. The output generated form the CNN is fed into the LSTM model to develop a new representation based on the data order. Additionally, FCL signifies the CNN-LSTM output into categorization classes for the purpose of COVID-19 risk level identification.

The following sections of this article are listed as follows: Section II examines past studies on COVID19 CT-CXR scan classification methods. The operation of the MS-AUNet-EMSDATL framework is devised in Section III, while its effectiveness is demonstrated in Section IV. Section V outlines the research's general findings and suggests future solutions.

II. LITERATURE SURVEY

Polat [12] proposed a multi-assigned semantic segmentation of chest CT images affected by the COVID-19 by utilizing the reorganized CNN model named DeepLabV3+. The feature interpretation capabilities of this model was improved using a pre-learned ResNet18 structure. However, the edge features of the ROI were ignored because subtle variation in lesion size towards the margins were undetected.

Shang et al. [13] created a Two-Stage Hybrid U-Net (TSH-UNet) to automatically separate COVID-19 pathogenic regions in CT images by recording the features of multiple layers by completely utilizing their context data. However, Segmentation error caused limited data diversity, making it difficult to separate minor lesions and use complicated convolution methods for micro characteristics.

Hossain et al. [14] trained TL on a deep CNN-based ResNet50 framework to classify COVID-19 patients using the CXR data. In this model, the ResNet50 layout was modified by adding two more FCL and employing different pre-learned weights. However, this method results in higher computational time and space.

Baghdadi et al. [15] constructed a technique for automatically categorizing COVID-19 on CT lung images using multiple pre-learned CNN and Sparrow Search Algorithm (SSA). The SSA was used to fine-tune several CNN and TL hyperparameters in order to get the best possible configuration for the pre-learned framework. However, an ensemble classifier was needed to increase the classification accuracy.

Meng et al. [16] designed a 2-phase TL detection system for Medical scans of COVID-19 (TL-Med). Initially, the Vision Transformer (ViT) pre-learning system was utilized to get generic characteristics from huge heterogeneous data and the medical characteristics were learned from large-scale

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *14 October 2023* ***Accepted:*** *24 October 2023*

homogeneous data. Moreover, 2-phase TL was applied to use the trained key characteristics and the actual data for COVID-19 identification. But, the accuracy

Gour and Jain [17] created a unique layered CNN structure for detecting COVID-19 disease using CXR and CT images. Various sub-models were acquired from the VGG19 and the Xception frameworks during learning. After that, those frameworks were combined by the softmax categorizer, which fuses the discriminating power of various CNN sub-models and identifies COVID-19 illness. But, it needs to categorize X-ray scans into the different classes of pneumonia.

Garg et al. [18] developed a new method depending on the Multi-Source Deep TL (MSDTL) to effectively monitor the prospective COVID-19 diseases. In this method, every province-related database was trained on a basic LSTM framework for forthcoming disease prediction in that domain. Also, the learned framework was fine-tuned by the MSDTL to achieve precise prediction. But, the efficiency was degraded while the gradient of disease range was extremely low.

Shaik and Cherukuri [19] designed a new ensemble model, which combines the power of various CNN structures before arriving at the final decision. Different pre-learned frameworks were applied and fine-tuned by the lung CT images. Then, those frameworks were utilized to build a robust ensemble categorizer which provides the final forecasting outcome. However, the recognition of various lung diseases from CT images was required to increase the accuracy.

Foysal et al. [20] created a Deformable Deep CNN (DDCNN) model to detect COVID-19 instances from CT scans. The model converted a 15-layer deep CNN model and adjusted the parameters and selected the 15-layered framework based on its maximum efficiency. The gradient class activation mapping (Grad-CAM) approach was utilized to monitor and validate the localization effort of targeted areas. But, this model results with higher categorization error.

Chetoui and Akhloufi developed PPFL for COVID-19 detection using ViT with Multi-Layer Perceptron (MLP) classification head. The model includes flatten, batch-normalization, dense and batch-normalization layers. The softmax function predicts normal or COVID-19 image classification, but lower results on smaller datasets.

Marefat et al. [22] used Compact Convolutional Transformers (CCT-COVID) to develop a transformer-based approach for autonomously identifying COVID-19 from CXR images. CCT was built on Compact Vision Transformers (CVT) and employs a convolutional tokenizer which results in the local data retention and the generation of finer tokens for classification tasks. However, this model results in high ambiguity issues.

Ren et al. developed a transformer-based method for classifying COVID-19 medical images. The model involves four stages like global self-attention, feature specifics extraction and fully linked and global average pooling layers. The first three phases capture vital feature information, while the fourth step extracts feature specifics. However, this approach has a slower convergence rate.

III. PROPOSED METHOD

In this part, the MS-AUNet-EMSDATL architecture is briefly described. Fig. 1 shows the MS-AUNet-EMSDATL block structure.

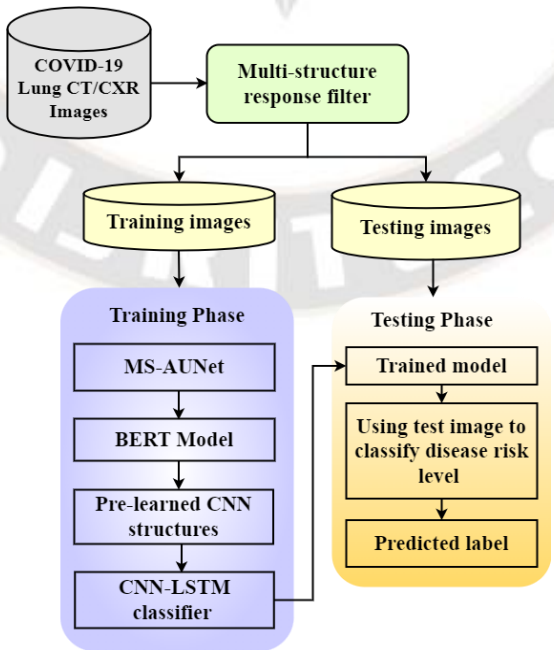

Figure. 1 Block structure of MS-AUNet-EMSDATL

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *14 October 2023* ***Accepted:*** *24 October 2023*

The key tasks in this framework include:

1. First, COVID-19 lung CT and X-ray scans are collected from different open sources. Such scans are enriched by the MSRF and segregated into the COVID-19-infected tissue ROIs and healthy ROIs at multiple scales by the MS-AUNet-MSDL.
2. Then, those infected tissue ROIs are fed into the BERT model for learning the pixel and its adjacent information to improve classification accuracy.
3. The output of BERT is fed into pre-learned CNN structures for extracting more relevant and discriminative features.
4. The CNN output is transformed to the LSTM model to from new representation of those features and get the trained classification models.

The trained classifiers are used to classify the test scans into low, medium and high-risk levels for proper COVID-19 diagnosis.

A. BERT Based Numerical Representation to learn pixel and its adjacent information

As illustrated in Fig. 2, the BERT is utilized in this framework to learn the pixel and its adjacent information effectively to increase the level of accuracy in COVID-19 detection. BERT is regulated according to the input representation which can be accepted by the perspective model. Tokenization, padding, numericalization and embeddings are the four components that comprises BERT. BERT tokenizes the input image pixel and then adds two different tokens, [CLS] at the beginning and [SEP] at the ending. Padding is essential to ensure that each input sentence has the same length after tokenization. A special token [PAD] is included whenever the phrase length exceeds tokens. In order to recognize the pixel's closest information, each token is converted into a non-negative integer during the numerical computation stage.

BERT employs a stack of transformer (Trm) encoder layers as its main structure. The integrated pixels varies in BERT which is denoted by E_1 , E_2 ,…., E_c. The encoded pixel sequences T_1, T_2 ,…., T_c have been processed through multiple Trm layers. Two sub-layers like position-wise feed-forward system and multi-head self-attention are composed of each encoder layer. Trm is employed with the residual technique and includes self-attention mechanism which results with rapid training time and robust expressive capacity. The Fig. 2 depicts the Trm structure in BERT.

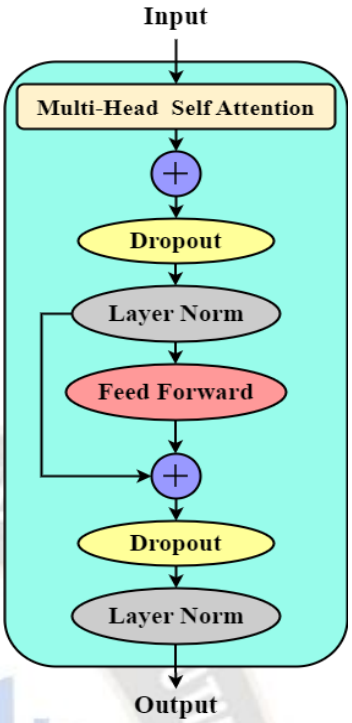

Figure. 2 Structure of Trm in BERT

This model composed of 12 encoder layers, 12 attention heads and a hidden size of 768. The pre-trained BERT model is processed by the encoder layer stack which adopts an input embedding to generate the numerical interpretations. The first layer encoder will evaluate each token's expressions and its results are fed into the second layer encoder as input. This process is continued until the 12^{th} encoder which is the final layer encoder. Each of them will be represented by contextualized embedding vectors from the 12^{th} encoder. The resulting matrix is $128 * 768$; 128 represents the number of tokens while 768 represents the hidden size.

1) Multihead Self-Attention (MHSA)

In the MHSA, an attention function maps an associated image, pixel values and neighbor information to an output which are all determined as vectors. The final product is produced as a weighted sum of the values with the weights assigned to each value being decided by the coherence function of the image with the pertinent pixel data. The attention mechanism is utilized for estimated dot-product attention as given in Eq. (1),

$$Attention\ (I, P, N) = softmax\left(\frac{IP^T}{\sqrt{D_P}}\right) Ni \quad (1)$$

In Eq. (1), I, P, Ni and D represents the corresponding image, pixel values, neighbour information and input data dimensions. The MHSA employs J head as $H_1, H_2, .., H_J$ and is denoted as

$$MHSA\ (A) = Concat\left(H_1, H_2, .., H_J\right) w^r \quad (2)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *14 October 2023* ***Accepted:*** *24 October 2023*

Where, w^r devises the learned parameter matrices and with similar dimensions using learned parameter matrices like $w^I, w^P, w^{Ni} \in \mathcal{R}^{D*D/J}$.

$$H_x = Attention\ (Aw_x^I, Aw_x^P, Aw_x^{Ni}) \quad (3)$$

Diverse attentions are learned by the various heads of the MHSA systems, where each head independently and concurrently completes its task. Assume that there are M attention heads in total and that each head has a global receptive field (GRF) in the nearest regions. The scaled product attention evaluates each attention pattern.

2) Feed-Forward Module (FFM):

The FFM is used to convert all heads to learn all pixel neighboring data through FCLs. A Rectified Linear Unit ($ReLU$) activation separates two linear modifications in the FFM. The FFM is a prevalent function throughout the $Trms$ so that the transformer configurations are comparable across the $Trms$. FFM is determined in Eq. (4),

$$FFM = \max\ (w_1 i\ +\ SE_1, 0)\, w_2 + b_2 \quad (4)$$

3) Layer Normalization:

A normalization strategy is developed to prevent the internal covariate shift during CNN-LSTM model training. The preceding Eq. (5) determines the overall structure of normalization.

$$\hat{\imath} = T.\frac{i - E(i)}{\sqrt{V(i) + \alpha}} + S \quad (5)$$

Where T and S are the learned scale and shift parameters respectively. The expression $E(i) = S$ and $V(\hat{\imath}) = T^2$ may be deduced with little effort. The input of the present layer with i_x input to the i^{th} neuron is represented as $i = (i_1, i_2, \ldots, i_D)$ for the objective of layer normalization. Layer normalization offers a plethora of benefit as it expedites better regulating tasks, rapid faster learning rate and serves as a regularizer throughout the procedure. Unlike batch normalization, it is not dependent on batch characteristics and may thus be employed in short batch settings. In BERT, the self-attention layers of the transformer serves as a GRF that allows each token to adjust with any other token. The self-attention layer performs an evaluation of an attention distribution across the input areas for each pixel in the region, allowing the distribution to accumulate knowledge about both the pixel and its neighbor's regions.

B. Joint Training of BERT with CNN-LSTM

MS-AUNet-EMSDATL feeds the BERT output values as input to the CNN-LSTM, making training easier. BERT may be trained in a single regions, i.e., single training, whereas CNN-LSTM is trained in many regions. The goal of joint training is to expose BERT to a wide range of target pixel regions such that it can be projected to generalize to formerly unidentified locations and concurrently acquire their neighboring information. Throughout a training repetition, each objective pixel will have a particular region assigned to it to achieve an optimal proportion among generalization capability and training efficacy. Assume three regions such as circle, square and line in combined training of BERT with CNN-LSTM which is not required to select all three areas for every target pixel rather than assigning one region randomly. For example, pixels U, K and Q are assigned to the circle, square and line region respectively.

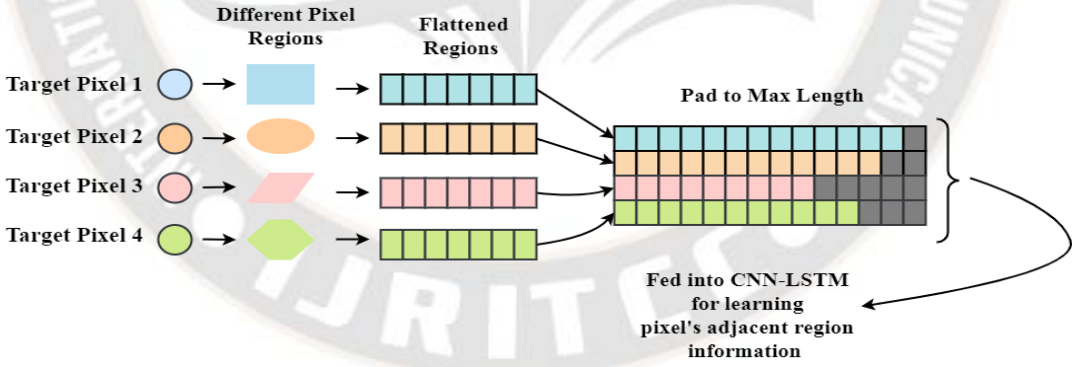

Figure. 3 Joint training of BERT with CNN-LSTM for learning pixel's neighbor information

In order to concurrently learn the pixel's adjacent information in the BERT and CNN-LSTM, these neighbor regions close to the pixels are first flattened into pixel sequences and then padded to a maximum length using fake pixels. The pad-to-max length method is based on language models that extend sentences by employing an expected token until their maximal values are reached. The tokens are considered in numerical may then be immediately loaded form CNN-LSTM for training the model to learn the information efficient which enhance the performances for COVID-19 detection. The Fig. 3 devises the joint learning task. In Fig. 3, the gray pixels depicts the fake pixels that are used to pad pixel data to a maximum length. Pixel sequences from various locations are concatenated and provided automatically into the CNN-LSTM for training.

The Fig. 4 defines the structure of BERT with CNN-LSTM for COVID-19 detection. The BERT is identified as the pixel-level data extractor in the suggested model, which is highly

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *14 October 2023* ***Accepted:*** *24 October 2023*

adaptable in learning the contexts from the pixels neighbor regions. The jointly trained BERT with CNN-LSTM has an effective generalization ability that allows it to learn the pixels nearby information. BERT with CNN-LSTM is pre-trained in some regions that may be fine-tuned or immediately utilized to improve prediction tasks on a region with a distinct form. Furthermore, the FCLs (softmax layer) transfers the CNN-LSTM output into categorization classes to learn pixels and its neighboring data for COVID-19 risk level identification.

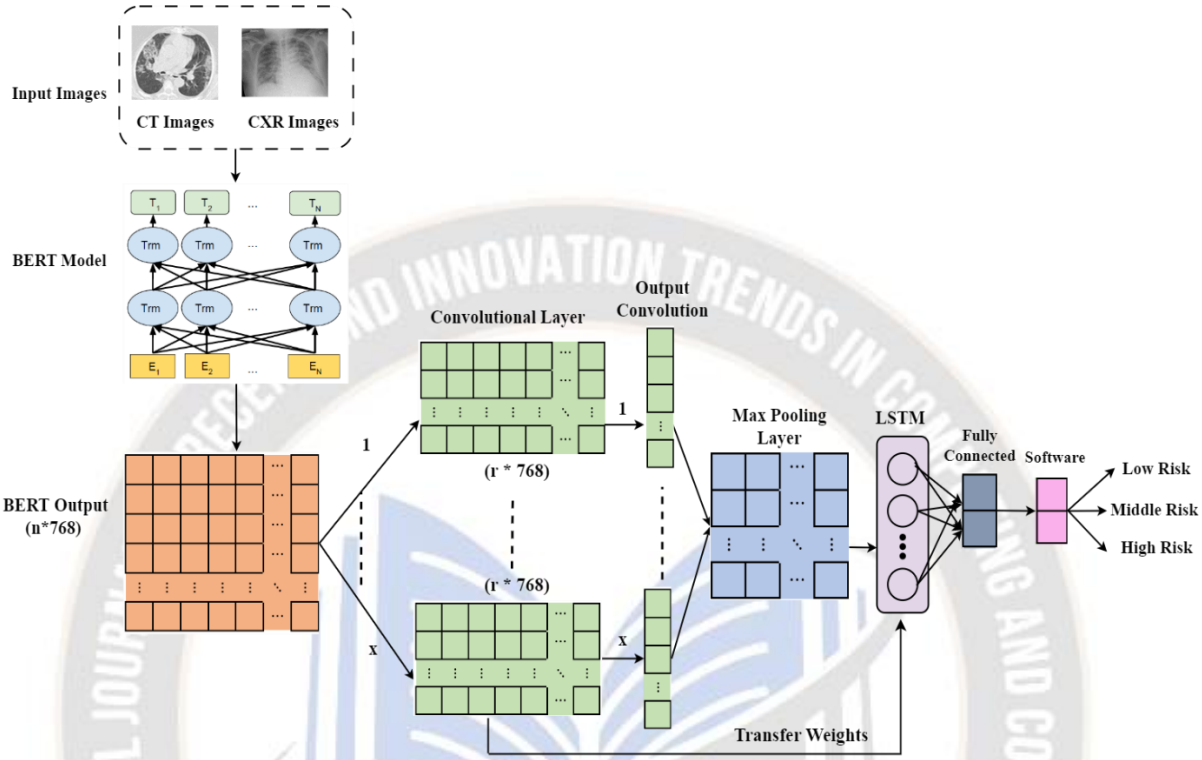

Figure.4 Structure of BERT with CNN-LSTM for COVID-19 detection

IV. RESULT AND DISCUSSION

In this part, TSH-UNet [13], CNN-SSA [15], MSDTL [18], ViT-PPFL [21], CCT-COVID [22] and MS-AUNet-EMSDATL [11] are used to evaluate the efficiency of the MS-AUNet-EMSDATL. The fore-mentioned models are implemented in MATLAB 2017b for conducting an evaluation in terms of accuracy, precision, recall and f-measure.

A. Dataset Description

The CXR image dataset [24] is utilized in this experiment. A total of 1200 CXR images from the COVID-19 and healthy classes i.e., 600 images from each class are randomly selected from this collection. The images of 500 and 100 number from each class are utilized for training and testing respectively. Additionally, the dataset Radiopaedia-COVID-19 CT Cases-2020 [25] is taken into account. 760 COVID-19 chest CT pictures and 760 regular chest CT images were obtained from this dataset. 610 photographs from each class are taken into account for training and 150 images from each class are taken into account for testing.

B. Accuracy

It is the ratio of correctly classified COVID-19 samples to the total number of examples examined. It is calculated by Eq. (6).

$$Accuracy = \frac{True\ Positive\ (TP)+True\ Negative\ (TN)}{TP+TN+False\ Positive\ (FP)+False\ Negative\ (FN)} \quad (6)$$

In Eq. (6), the number of samples successfully identified as COVID-19 is known as the TP. The number of COVID-19 samples that are mislabeled as healthy is known as FP. The number of healthy samples that are incorrectly labeled as COVID-19 is FN. The number of healthy samples that were accurately identified as healthy known as TN.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *14 October 2023* ***Accepted:*** *24 October 2023*

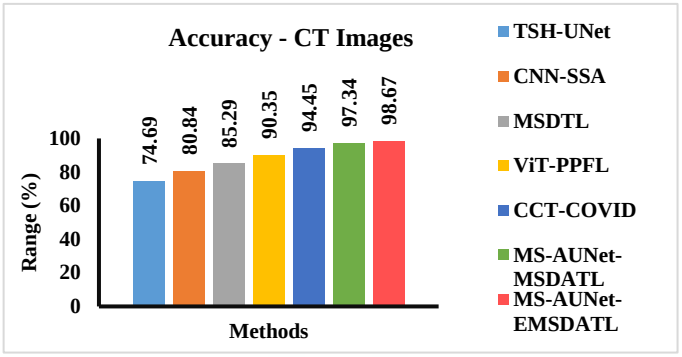

Figure. 5 Comparison of accuracy using CT images

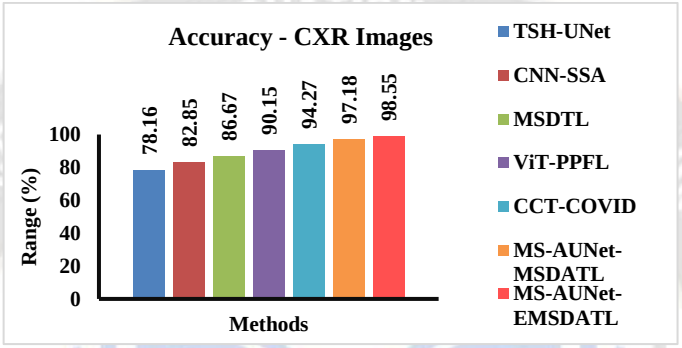

Figure. 6 Comparison of accuracy using CXR images

The accuracy attained by the various DL-based models preformed on the gathered CT and CXR images to determine COVID-19 infection risk levels are shown in Fig. 5 and 6. It finds that the MS-AUNet-EMSDATL accuracy utilizing CT image dataset is 32.11% higher than TSH-UNet, 22.06% higher than CNN-SSA, 15.69% higher than MSDTL, 9.21% higher than ViT-PPFL, 4.47% higher than CCT-COVID and 1.37% higher than MS-AUNet-MSDATL. Similarly, the accuracy of proposed MS-AUNet-EMSDATL is 28.09% greater than TSH-UNet, 18.95% greater than CNN-SSA, 13.71% greater than MSDTL, 9.32% greater than ViT-PPFL, 4.54% greater than CCT-COVID and 1.41% greater than MS-AUNet-MSDATL. This is because of localizing COVID-19 diseased tissues and learning more discriminative features from the collected lung scans for infection risk level categorization.

C. Precision

It is the number of COVID-19 cases that have been classified at the TP and FP rates. It It is calculated by Eq. (7).

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Fig. 7 and 8 portrays the precision for different models implemented on the considered CT and CXR images respectively to classify COVID-19 infection risk levels. It analyses that the precision of the MS-AUNet-EMSDATL by utilizing the CT image data is 27.77% greater than TSH-UNet, 17.64% greater than CNN-SSA, 12.21% greater than MSDTL, 6.79% greater than ViT-PPFL, 4.83% greater than CCT-COVID and 1.49% greater than MS-AUNet-MSDATL for COVID-19 infection risk level categorization. Also, the precision of the MS-AUNet-EMSDATL using X-ray image dataset is 31.19% greater than TSH-UNet, 24.88% greater than CNN-SSA, 18.59% greater than MSDTL, 12.88% greater than ViT-PPFL, 4.90% greater than CCT-COVID and 2.11% greater than MS-AUNet-MSDATL. This ensures that the precision of MS-AUNet-MSDATL is improved compared to the TSH-UNet, CNN-SSA, MSDTL, ViT-PPFL, CCT-COVID and MS-AUNet-MSDATL on both CT and X-ray images.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *14 October 2023* ***Accepted:*** *24 October 2023*

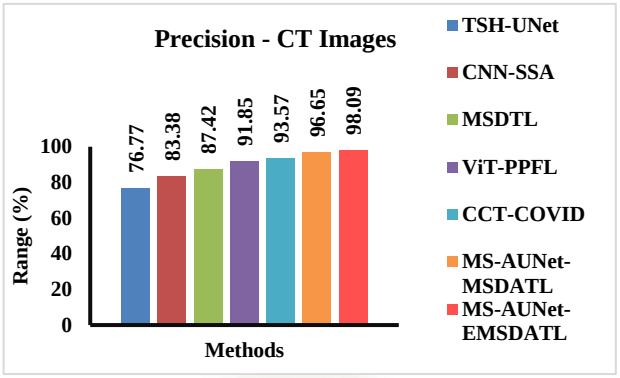

Figure. 7 Comparison of precision using CT images

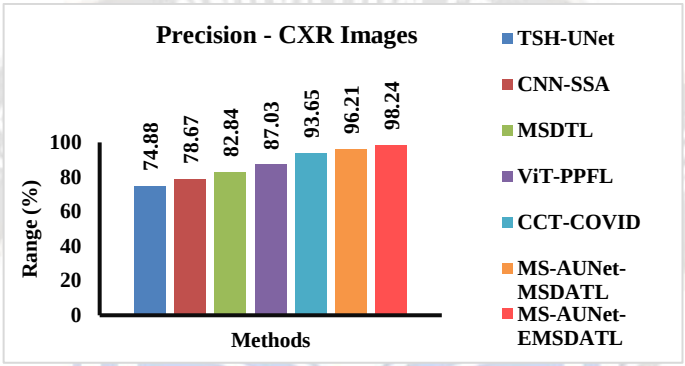

Figure. 8 Comparison of precision using CXR images

D. Recall

It is the percentage of COVID-19 cases that are properly classified at TP and FN rates. It is determined by Eq. (8).

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Fig. 9 and 10 show the recall for several DL-based models used to categorize COVID-19 infection risk levels on CT and CXR images. The recall of the MS-AUNet-EMSDATL using CT image dataset is 29.99% greater than TSH-UNet, 22.94% greater than CNN-SSA, 14.84% greater than MSDTL, 9.02% greater than ViT-PPFL, 4.29% greater than CCT-COVID and 1.48% greater than MS-AUNet-MSDATL for COVID-19 infection risk level categorization. Also, the recall of the MS-AUNet-EMSDATL using X-ray image dataset is 31.66% greater than TSH-UNet, 22.40% greater than CNN-SSA, 16.59% greater than MSDTL, 11.23% greater than ViT-PPFL, 4.78% greater than CCT-COVID and 1.27% greater than MS-AUNet-MSDATL. This guarantees that the recall of MS-AUNet-MSDATL is improved compared to the TSH-UNet, CNN-SSA, MSDTL, ViT-PPFL, CCT-COVID and MS-AUNet-MSDATL on both CT and X-ray images.

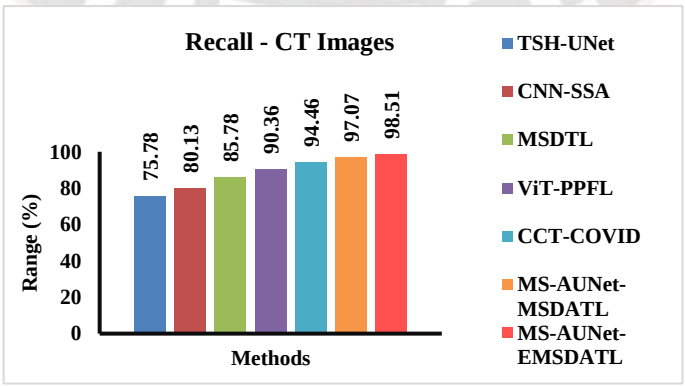

Figure. 9 Comparison of recall using CT images

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *14 October 2023* ***Accepted:*** *24 October 2023*

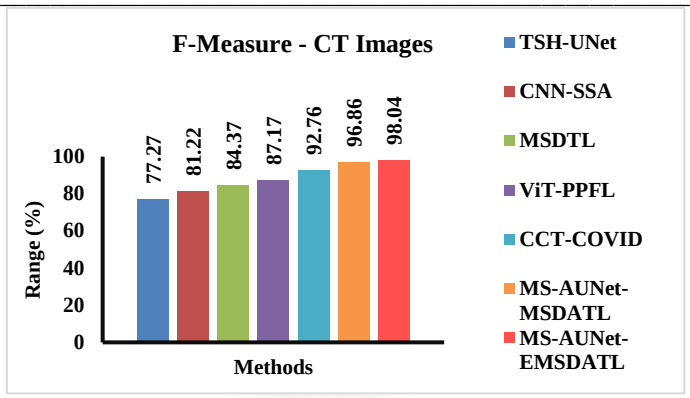

Figure. 10 Comparison of recall using CXR images

E. F-measure

It is determined by Eq. (9) as,

$$F - measure = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} \quad (9)$$

Fig. 11 & 12 depicts the F-measure values of various TL-based models implemented on the considered databases to classify COVID-19 infection risk levels. It addresses that the F-measure of the MS-AUNet-EMSDATL using CT image is 26.88% greater than TSH-UNet, 20.71% greater than CNN-SSA, 16.20% greater than MSDTL, 12.47% greater than ViT-PPFL, 5.69% greater than CCT-COVID and 1.22% greater than MS-AUNet-MSDATL for COVID-19 infection risk level categorization. Also, the F-measure of the MS-AUNet-EMSDATL using X-ray image dataset is 25.12% greater than TSH-UNet, 20.89% greater than CNN-SSA, 15.69% greater than MSDTL, 12.62% greater than ViT-PPFL, 6.89% greater than CCT-COVID and 2.06% greater than MS-AUNet-MSDATL. This is because of using an ensemble classifier, i.e. BERT and CNN-LSTM with TL, whereas the other models employ classical DL classifiers for the classification tasks.

Thus, these findings proved that the proposed MS-AUNet-EMSDATL using both CT and X-ray image datasets increases accuracy, precision, recall and f-measure compared to the TSH-UNet [13], CNN-SSA [15], MSDTL [18], ViT-PPFL [21], CCT-COVID [22] and MS-AUNet-EMSDATL [11] models efficiently. The proposed MS-AUNet-EMSDATL can be useful for physicians to provide an accurate diagnosis by recognizing COVID-19 risk levels.

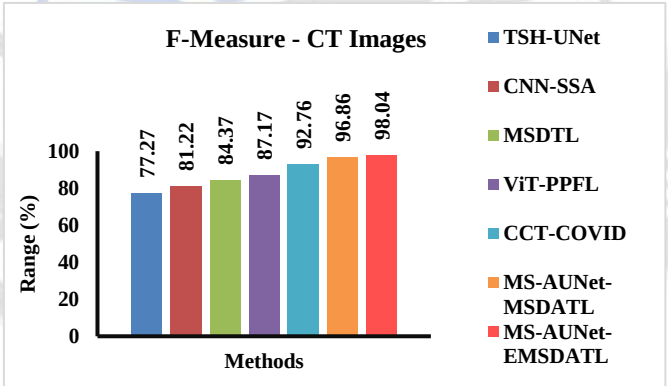

Figure. 11 Comparison of F-measure using CT images

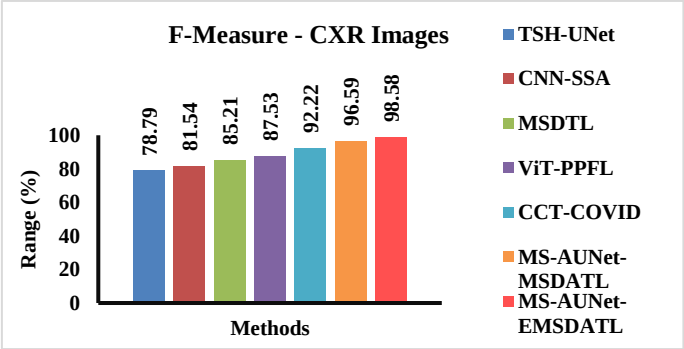

Figure. 12 Comparison of F-Measure using CXR images

V. CONCLUSION

In this paper, MS-AUNet-EMSDATL is proposed to learn pixels and their adjacent information for efficient COVID-19 risk level detection. The method introduces BERT and CNN-LSTM in MS-AUNet-MSDATL model to improve COVID-19 risk level classification. The BERT structure extracts fixed features from pre-trained models, transforms them into CNN models and sends the CNN output to LSTM models. The LSTM output is then mapped into classification classes, learning pixels and their adjacent information for COVID-19 risk level detection and diagnosis. The FCL converges the LSTM output into categorization classes for effective COVID-19 risk level identification and diagnosis. At last, the experimental test reveals that the proposed MS-AUNet-EMSDATL achieves accuracy of 98.67% and 98.55% on both CT and CXR scan while compared to the different classification frameworks.

REFERENCES

[1] S. H. Ahmed, A. Sahi, R. A. Al-Roomi and I. Al-Karkhi, "Introduction to COVID-19, history, impact, symptoms and prevention," Pakistan. J, Med. Health Sci, vol. 14, no. 2, pp. 1528-1534, 2020.

[2] N. S. R. Castro, I. C. Hernández, M. E. G. Reyes, M. Hernández, C. Romero, H. Colín and F. Reyes, "Clinical Signs and Symptoms Associated with COVID-19: A Cross Sectional Study," Int. J. Odontostomat, vol. 16, no. 1, pp. 112-119, 2022,doi:10.4067/S0718-381X2022000100112.

[3] R. B. Patel and B. B. Patel, "An Analysis of the COVID-19 Situation in India in Terms of Testing, Treatment, Vaccine Acceptance and National Economic Performance," International Journal of Public Health, vol. 67, no. 1604975, 2022, doi:10.3389/ijph.2022.1604975.

[4] Z. Luo, M. J. Y. Ang, S. Y. Chan, Z. Yi, Y. Y. Goh, S. Yan and X. Liu, "Combating the coronavirus pandemic: early detection, medical treatment, and a concerted effort by the global community," Research, 2020, doi:10.34133/2020/6925296.

[5] R. J. Lu, L. Zhao, B. Y. Huang, F. Ye, W. L. Wang and W. J. Tan, "Real-time reverse transcription-polymerase chain reaction assay panel for the detection of severe acute respiratory syndrome coronavirus 2 and its variants," Chinese Medical Journal, vol. 134, no. 17, pp. 2048-2053, 2021, doi:10.1097/CM9.0000000000001687.

[6] D. Kollias, A. Arsenos and S. Kollias, "Ai-mia: Covid-19 detection and severity analysis through medical imaging," In European Conference on Computer Vision (pp. 677-690). Cham: Springer Nature Switzerland, 2022, October, doi:10.1007/978-3-031-25082-8_46.

[7] R. Rehouma, M. Buchert and Y. P. P. Chen, "Machine learning for medical imaging-based COVID-19 detection and diagnosis," International Journal of Intelligent Systems, vol. 36, no. 9, pp. 5085-5115, 2021, doi:10.1002/int.22504.

[8] D. Yang, C. Martinez, L. Visuña, H. Khandhar, C. Bhatt and J. Carretero, "Detection and analysis of COVID-19 in medical images using deep learning techniques. Scientific Reports, vol. 11, no. 1, pp. 19638, 2021, doi:10.1038/s41598-021-99015-3.

[9] M. S. Ahmed and A. M. Fakhrudeen, "Deep learning-based COVID-19 detection: State-of-the-art in research. International Journal of Nonlinear Analysis and Applications, vol. 14, no. 1, pp. 1939-1962, 2023, doi:10.22075/ijnaa.2022.7119.

[10] T. Zhou, S. Canu and S. Ruan, "Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism. International Journal of Imaging Systems and Technology, vol. 31, no. 1, pp. 16-27, 2021, doi:10.1002/ima.22527.

[11] J. P. S. Rathinam and A. P. Gopalan, "Multi-Scale Learning with Attention-based UNet and Marginal Space Deep Ambiguity Transfer Learning for Lung Disease Prediction," International Journal of Intelligent Engineering & Systems, vol. 16, no. 4, 2023,doi: 10.22266/ijies2023.0831.45.
H. Polat, "Multi-task semantic segmentation of CT images for COVID-19 infections using DeepLabV3+ based on dilated residual network. Physical and Engineering Sciences in Medicine, pp. 1-13, 2022, doi:10.1007/s13246-022-01110-w.

[12] Y. Shang, Z. Wei, H. Hui, X. Li, L. Li, Y. Yu, and Y. Zha, "Two-stage hybrid network for segmentation of COVID-19 pneumonia lesions in CT images: a multicenter study," Medical & Biological Engineering & Computing, vol. 60, no. 9, pp. 2721-2736, 2022, doi:10.1007/s11517-022-02619-8.

[13] M. B. Hossain, S. H. S. Iqbal, M. M. Islam, M. N. Akhtar and I. H. Sarker, "Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images," Informatics in Medicine Unlocked, vol. 30, pp. 1-10, 2022, doi:10.1016/j.imu.2022.100916.

[14] N. A. Baghdadi, A. Malki, S. F. Abdelaliem, H. M. Balaha, M. Badawy and M. Elhosseini, "An automated diagnosis and classification of COVID-19 from chest CT images using a transfer learning-based convolutional neural network," Computers in Biology and Medicine, vol. 144, pp. 1-17, 2022, doi:10.1016/j.compbiomed.2022.105383.

[15] J. Meng, Z. Tan, Y. Yu, P. Wang and S. Liu, "TL-Med: A two-stage transfer learning recognition model for medical images of COVID-19," Biocybernetics and Biomedical Engineering, vol. 42, pp. 842-855, 2022, doi:10.1016/j.bbe.2022.04.005.

[16] M. Gour and S. Jain, "Automated COVID-19 detection from X-ray and CT images with stacked ensemble convolutional neural network," Biocybernetics and Biomedical Engineering, vol. 42, no. 1, pp. 27-41, 2022 doi:10.1016/j.bbe.2021.12.001.

[17] S. Garg, S. Kumar and P. K. Muhuri, "A novel approach for COVID-19 Infection forecasting based on multi-source deep transfer learning," Computers in Biology and Medicine, vol. 149, pp. 1-16, 2022, doi:10.1016/j.compbiomed.2022.105915.

[18] N. S. Shaik and T. K. Cherukuri, "Transfer learning based novel ensemble classifier for COVID-19 detection from chest CT-scans," Computers in Biology and Medicine, vol. 141, pp. 1-8, 2022 doi:10.1016/j.compbiomed.2021.105127.

[19] M. Foysal, A. B. M. Hossain, A. Yassine and M. S. Hossain, "Detection of COVID-19 Case from Chest CT Images Using

Deformable Deep Convolutional Neural Network," Journal of Healthcare Engineering, 2023 doi:10.1155/2023/4301745.

[20] M. Chetoui and M. A. Akhloufi, "Peer-to-Peer Federated Learning for COVID-19 Detection Using Transformers," Computers, vol. 12, no. 5, pp. 106, 2023, doi:10.3390/computers12050106.

[21] A. Marefat, M. Marefat, J. Hassannataj Joloudari, M. A. Nematollahi and R. Lashgari, "CCTCOVID: COVID-19 detection from chest X-ray images using Compact Convolutional Transformers," Frontiers in Public Health, vol. 11, pp. 1025746, 2023, doi:10.3389/fpubh.2023.1025746.

[22] K. Ren, G. Hong, X. Chen and Z. Wang, "A COVID-19 medical image classification algorithm based on Transformer," Scientific Reports, vol. 13, no. 1, pp. 5359, 2023, doi:10.1038/s41598-023-32462-2.

[23] https://github.com/dara1400/Covid19-Xray-Dataset

[24] Radiopaedia-COVID-19 CT Cases-2020 dataset. www.radiopaedia.org