

# A Self-Attention Mechanism for COVID-19 Detection using Chest X-Ray Images

Usman Muhammad, Md. Ziaul Hoque, Mourad Oussalah, Anja Keskinarkaus, Tapio Seppänen\*, Pinaki Sarder †

\* Center for Machine Vision and Signal Analysis (CMVS), University of Oulu, Finland

† Department of Pathology and Anatomical Sciences, University at Buffalo, USA

**Abstract**— COVID-19 is a rapidly spreading viral disease and has affected over 100 countries worldwide. The numbers of casualties and infected cases have been escalated particularly in vulnerable states with weakened healthcare systems. Recently, reverse transcription-polymerase chain reaction (RT-PCR) is the test of choice for diagnosing COVID-19. However, current evidence suggests that COVID-19 infected patients are mostly stimulated from a lung infection after coming in contact with this virus. Therefore, chest X-ray (i.e., radiography) and chest CT can be a surrogate in some countries where PCR is not readily available. This has forced the scientific community to detect COVID-19 infection from X-ray images and recently proposed machine learning methods offer great promise for fast and accurate detection. Deep learning with convolutional neural networks (CNNs) has been successfully applied to radiological imaging for improving the accuracy of diagnosis. However, the performance remains limited due to the lack of representative X-ray images available in public benchmark datasets. To alleviate this issue, we propose an attention mechanism for data augmentation in the feature space rather than in the data space using reconstruction independent component analysis (RICA). Specifically, a unified architecture is proposed which contains a deep convolutional neural network (CNN), an attention mechanism, and a bidirectional LSTM (BiLSTM). The CNN provides the high-level features extracted at the pooling layer where the attention mechanism chooses the most relevant features and generates low-dimensional augmented features. Finally, BiLSTM is used to classify the processed sequential information. We conducted experiments on two publicly available databases to show that the proposed approach achieves the state-of-the-art results with an accuracy of 97% and 84% while being able to generate explanations. Explainability analysis has been carried out using feature visualization through PCA projection and t-SNE plots.

**Index Terms**—COVID-19 detection, Attention mechanism, Feature augmentation, Transfer learning, BiLSTM

## I. Introduction

Coronavirus disease (COVID-19) is a viral respiratory disease that initially emerged in China when a cluster of patients with unknown pneumonia was reported in the capital of Hubei Province (Wuhan). The virus that caused the disease was identified to be severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses based on phylogeny, taxonomy, and established practice [26]. At the time of writing, the World Health Organization (WHO) reported that approximately 185 million people are affected and 4 million people have died worldwide. Moreover, animals such as cats and dogs have also been reported to be infected with SARS-CoV-2 in many

countries, including the United States. Thus, WHO declared this virus a “Public health emergency of international concerns” and classified it as a pandemic on March 2020 [3]. Most infected people develop mild to moderate illness and common symptoms are runny nose, body aches, cough, fever, sore throat, and shortness of breath [52].

Since the beginning of the pandemic, several diagnostic methods have been approved by several international and country-specific agencies. However, there is no clear consensus on the correct tests to be used related to any acute complaints to yield a correct diagnosis in a timely constraint. In EU member countries, there are 365 different commercialized devices that have been used for conducting a such research. Among them, 168 are Immunoassays, three are sequencing-based methods, 192 are PCR-based methods, and two commercialized tools are based on different medical devices [17]. WHO recommended RT-PCR test (developed by Corman), which is nowadays considered as the current standard for detecting a coronavirus infection. However, the false-negative rate was found to be approximately 20% to 40% in the infected cases in China due to inappropriate sample collection, personnel operation, storage, and low sensitivity test kits [15].

Along with laboratory testing, chest CT scans with the help of a radiologist can be considered as a screening tool with RT-PCR [57]. COVID-19 infected patients show ground-glass opacities in the periphery of both lungs, and appear more grey or hazy as opposed to the normally dark-appearing lungs. It is also stated that those patients who recovered from COVID-19 pneumonia, lung disease was observed ten days after the onset of symptoms [40]. In the early days of the pandemic, clinical centers in Wuhan were working with an insufficient number of often malfunctioning test kits, resulting in a concerning amount of false negatives. To counteract these challenges, doctors were persuaded to make diagnoses based only on laboratory and chest CT results [6]. In developing countries, such as India, where the number of test kits remains low, CT is also used for COVID-19 detection.

In addition to CT scans, chest X-ray scanning machines are easily accessible in almost all hospitals, and have a potential role in the diagnosis because X-ray images represent visual indexes linked with SARS-CoV-2 [57]. In Fig. 1, we visualize example images from normal, COVID-19, pneumonia, and bacterial pneumonia classes, taken from two X-ray image databases [11, 61, 40]. Thus, radiologic images obtained from COVID-19 cases with laboratory results may help in the

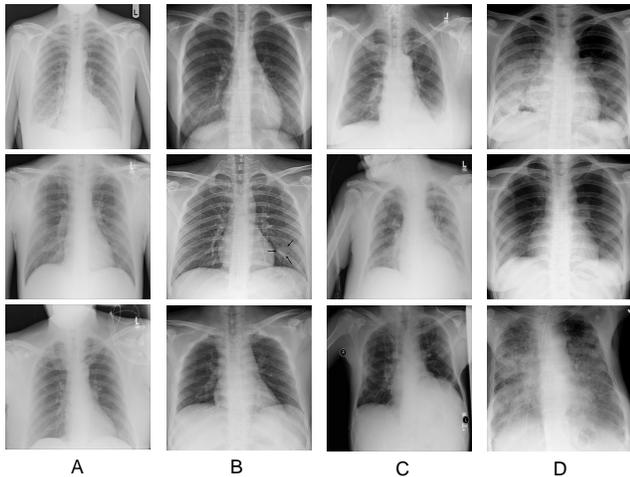


Figure 1: Samples images from chest X-ray database: (A) Normal, (B) COVID-19, (C) Pneumonia, (D) Bacterial pneumonia.

early detection of SARS-CoV-2. The study conducted on CT images by Kong *et al.* [32], demonstrates acute bilateral airspace opacities in infected patients. Zhao *et al.* [65] reported that most patients had a fever as the onset symptom. Based on the result of X-ray scans, ground-glass opacities (GGO) 87%, vascular enlargement in the lesion 72% or mixed GGO, and consolidation 65% appeared. Moreover, authors show that lesions present on CT images are more likely to have a peripheral distribution. Li *et al.* observed that chest CT had a small rate of missed diagnosis of COVID-19. GGOs and consolidation with or without vascular enlargement are common CT features of COVID-19 and may be useful as a standard method for the rapid diagnosis of COVID-19 [57]. Similarly, Zu *et al.* [69] concluded that 33 of chest CTs can have rounded lung opacities.

Machine learning (ML) techniques are attracting substantial interest in the medical field, where deep learning-based models have been successfully utilized in many healthcare applications such as depression detection [36], pain estimation [54], breast cancer detection [10], Alzheimer’s disease classification [18], and pneumonia detection from chest X-ray images [2]. Due to the increase in COVID-19 cases, healthcare systems have been overwhelmed and require alternative solutions for the automated diagnosis of COVID-19. In this regard, many attempts have been put forward to address such problems using radiology images [40, 6, 57, 9, 32]. However, it is not feasible to build a large labeled database for every disease, i.e., viral pneumonia, COVID-19, bacterial pneumonia, aspiration pneumonia, etc. Thus, the bias in small datasets and the lack of representative training and tuning data impair the performance of such deep learning models.

A simple way to deal with these challenges consists in applying data augmentation techniques, which enable researchers to significantly increase the diversity of data, without collecting new data. However, augmented data that could be borrowed from unlabeled data [19], random erasing [67] or randomly masking regions [13] are heavily dependent on

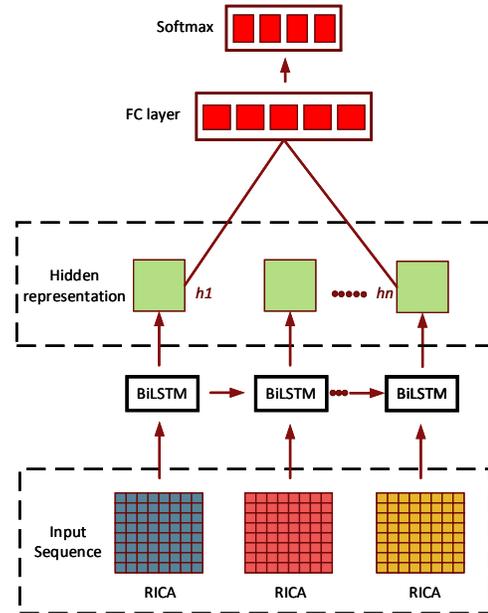


Figure 2: The main idea of the proposed work. RICA is designed to augment the feature space. The augmented features are used as a sequence input to train the BiLSTM network for final detection. Three different colors in RICA represent three augmented features.

training parameters. For instance, a slight rotation between 1 to 30 or a random cropping  $((288, 288) \rightarrow (224, 224))$ , could be useful on digit recognition tasks such as MNIST, but as the rotation degree increases, the label of the data is no longer preserved under post-transformation [51]. The dominant approaches such as Generative adversarial networks (GANs) [21], Bidirectional GANs [14], the DCGAN [42], Progressively Growing GANs [30], the CycleGAN [68], generate synthetic images but require careful domain adaptation to transfer the knowledge and features to the real image domain. On the other hand, existing CNN-LSTM based COVID-19 detection methods [5] treat the convolutional features as equally important and ignore the interference information (e.g., mutual exclusion and redundancy), which can prevent learning of long data sequences. Moreover, the high-dimensional vector generated by CNN can increase the network parameters of LSTM and make the network difficult to optimize.

Therefore, in this paper, an attention mechanism is proposed and integrated with CNN-BiLSTM architecture to approximate the real distribution in feature space rather than in data space, where the generated features not only maintain the label information but also promise diversity. Especially, reconstruction independent component analysis (RICA) has been utilized to augment feature space but without performing any training data augmentation strategy. An illustration of a such concept is provided in Fig 2. Besides, the interference information or redundancy is significantly eliminated by selecting the low-dimensional augmented features. On other other hand, it is worth mentioning that chest radiography analysis is known to have inherent limitation in early stages of COVID-19 detec-

tion, due to low sensitivity in GGO detection [69]. Moreover, recovered patients are likely to be protected against reinfection for several weeks but may still transmit the virus. However, well-trained deep learning methods can focus on anomalies that are not visible to the human eyes, and may encourage their applications in health care system. In overall, our main contributions in this paper are summarized as follows:

- 1) The proposed work explores the idea of feature augmentation in COVID-19 detection. To achieve this, reconstruction independent component analysis [33] is utilized, and the features are augmented from the pooling layer of a CNN model.
- 2) The need for annotated data in COVID-19 detection is mitigated through the proposed feature augmentation strategy.
- 3) The effectiveness and validation of our proposed method have been extensively explored on two publicly available datasets and achieve state-of-the-art results.
- 4) PCA and t-SNE feature visualization has been utilized to demonstrate the explainability of the proposed learning model.

The rest of this paper is organized as follows. Section 2 deals with literature review in the field. Section 3 details the proposed method highlighting the different phases of our proposal, including deep-feature extraction, attention-based learning module and the associated recurrent neural network. Section 4 emphasizes the experiment results, including data-set description, evaluation metrics, implementation details, results and explainability analysis. Finally conclusive statements and perspective works are provided in Section 5.

## II. Literature Review

In recent months, researchers have evaluated SARS-CoV-2 infected chest X-ray images using convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative adversarial networks (GANs) and encoder-decoder models. A brief overview about recent developments is provided in this section.

Narin et al. [38] utilized five pre-trained deep learning models (InceptionV3, Inception-ResNetV2, ResNet50, ResNet101 and ResNet152) to detect COVID-19, bacterial pneumonia and viral pneumonia. The final performance was 98% accuracy using pre-trained ResNet-50 model. Rajaraman et al. [44] proposed a method that increases training data using weakly labeled data augmentation. A stage-wise approach was used to train the CNN. The authors concluded weakly labeled data augmentation is superior in comparison to baseline non-augmented training.

Ozyurt et al. [41] used LBP, a traditional machine learning descriptor, and a feature selector method that selects most informative features together to achieve a better performance, achieving a 95.84% classification accuracy on CT images. Rahimzadeh and Attar [43] introduced a combined deep CNN to identify 11302 chest X-ray images. In their study, Xception and ResNet50V2 was used and claimed as a new strategy to address the unbalanced dataset problem. They reported an accuracy rate of 99.56%. Castiglioni et al. [8] utilized an

independent dataset of 110 patients suspected for COVID-19 infection, and developed a ten convolutional neural networks (CNNs) to evaluate the performance.

Multiple state-of-the-art deep learning models including DenseNet201, Resnet50V2 and Inceptionv3, were fine-tuned individually to make independent predictions in [12]. Then, a weighted average ensembling technique was used to combine them to achieve a classification accuracy of 91.62%. Similarly, Wang et al. [59] proposed a tailored deep convolutional neural network for classifying chest X-ray images. Hemdan et al. [24] examined seven different CNN architectures in their experiment, including DenseNet-121, VGG-19, ResNet-V2, Inception-V3, Xception, MobileNet-V2 and InceptionResNet-V2. Their work revealed that the DenseNet and VGG-19 models achieved the best performance with 91% accuracy for detecting COVID-19 and non-COVID-19 infections.

Berrimi et al. [7] fine-tuned two pre-trained models, DenseNet and InceptionV3 to classify both X-ray and CT chest scans. To increase the diversity of the training data, the images were rotated, zoomed, horizontally flipped, and shifted. Usman et al. [37] proposed to augment the feature space by using sparse filtering. Nour et al. [39] designed a five convolution layers (CNN) from the scratch. The extracted CNN features are then evaluated with traditional machine learning classifiers such as k-nearest neighbor, support vector machine (SVM), and decision tree. The authors concluded that the SVM classifier with an accuracy of 98.97% performs the best among all of them. Yoo et al. [64] used the pre-trained ResNet18 and different decision trees are utilized to detect CXR images as normal, tuberculosis, and COVID-19.

M. Turkoglu [56] employed the transfer learning approach by using the AlexNet architecture. To choose the most effective features, the Relief feature selection algorithm is used. Finally, the Support Vector Machine (SVM) is applied to detect COVID-19, and Pneumonia disease. An accuracy of 99.18% was reported. Sahlol et al. [45] proposed a combined approach where Inception model is utilized to extract the features and a swarm-based feature selection algorithm is applied to choose the most relevant features. Two public COVID-19 X-ray datasets are used and 99.18% accuracy was reported. Mesut et al. [55] utilized MobileNetV2 and SqueezeNet models to extract the deep features. Then, the Social Mimic optimization method is proposed and the features were combined and classified using SVM classifier. A slide image approach is used for image normalization in [25]. To overcome the limitation of chest X-ray samples, Karbhari et al. [29] proposed an Auxiliary Classifier Generative Adversarial Network (ACGAN) to generate synthetic images. Based on obtained images, Convolutional Neural Networks (CNNs) is utilized to detect COVID-19 in the CXRs.

Loey et al. [34] utilized a GAN architecture to synthesize auxiliary images as a motivation to overcome the issue of lack of datasets especially in chest X-rays images. Three deep transfer models are selected to detect four classes, i.e., the COVID-19, normal, pneumonia bacterial, and pneumonia virus. Googlenet performed the best in their work. The network consists of encoder and decoder is proposed in [16], to show that CORONA-Net performs the best for COVID-

19 detection. MASC-Net consists of a multi-input encoder-decoder, and introduced to automatically detect infected lung regions from COVID-19 chest CT scans [58]. 3D U-Net is proposed as encoder-decoder method in [62], where the multi-task learning is applied and compared with four transfer learning strategies. The authors concluded that using multiple lung lesion datasets can extract more general features.

We are motivated by the fact that chest X-ray images have been commonly used in most of the current works and have an important role in the diagnosis of COVID-19 detection. However, learning from imbalanced data or lack of necessary extracted features obtained from limited X-ray training samples cannot provide the expected performance in the COVID-19 detection. Thus, the proposed work focuses on a data augmentation strategy where the label preserved features are generated to improve the performance of deep learning model.

### III. The Proposed Method

The general framework of the proposed approach is divided into three components: (1) extraction of deep features (2) an attention-based learning module and (3) a BiLSTM-based sub-network. We first describe the procedure of feature extraction for guiding the process of feature generation. Next, we explain the procedure of augmenting the training data in feature space. Finally, the structure of BiLSTM network is discussed. The overall procedure of the proposed approach is illustrated in Fig 3.

#### A. Deep features extraction

Inspired by the performance of deep learning models, we adapt ResNet-50, an architecture known for its stability and performance, to extract high-level features for our task [23]. The model is fine-tuned by replacing the last fully-connected layer with a new fully connected layer and setting the number of outputs equal to the number of classes in the dataset. We freeze the weights of the first ten layers so that the gradients of these do not need to be computed. This is motivated by the fact that earlier features of ResNet contain more generic features (e.g. color blob detectors or edge detectors) and make the remaining layers more specific to the details of the classes contained in the original dataset. The weights of the new fully-connected layer are increased by a learning factor 10 and a bias factor 20. By doing so, earlier layers do not change that much and swiftly learn the weights of the newer layer. The softmax loss function, as denoted in Eq.1, is used as the loss function.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (1)$$

where  $\vec{z}$  is input vector to the softmax function,  $z_i$  are the elements of the input vector,  $e^{z_i}$  is the standard exponential function, and  $K$  denotes the number of classes (in this paper,  $c = 3$  and  $c = 5$ ).

#### B. An attention-based learning module

Data augmentation in the image space is a well established technique that enhances the size and quality of training datasets such that deep learning models can robustly model the training data. However, feature augmentation has not yet acquired the same level of attention. This is crucial for applications like COVID-19 detection, where the number of training samples remain limited. To accomplish this, a feature augmentation is conducted based on reconstruction independent component analysis (RICA) [33]. The latter was designed to overcome the drawbacks of independent component analysis (ICA) by replacing ICA's orthonormality constraint with a soft reconstruction penalty, which turns out to be very useful in learning sparse features. Therefore, the idea behind our proposed attention mechanism is to extract more meaningful information from generated ones to correctly classify target samples.

In our case, RICA receives data as input from the last pooling layer of ResNet-50, then it converts it into a new lower-dimension representation. In order to apply a such transformation, RICA is utilized by using the following equation:

$$v = Zk \quad (2)$$

where  $v$  is the vector representing the CNN features,  $Z$  denotes the matrix, and  $k$  are the independent components for dimensionality reduction. The goal of RICA is to define the observed data  $v$  by mixing the components  $k$ . We need to determine both  $Z$  and  $k$  from the data  $v$  because we can not directly extract the sources  $k$ , nor know the mixing matrix  $Z$ . Let  $W$  be the inverse of  $Z$ , then the model can be expressed as:

$$k = Wv \quad (3)$$

Hence, using the original data  $v$ , the goal is to determine a set of vectors (corresponding to the column vectors of matrix  $W$ ) that will form the features  $k$  sparse; while being an orthonormal basis. In this regard, our matrix  $W$  will assign the data  $v$  to features  $k$ . Hence, the optimization problem defined by RICA becomes [33]:

$$\min_W \lambda \|(Wv)\|_1 + \frac{1}{2} \|W^T Wx - x\|_2^2 \quad (4)$$

where  $\lambda$  evaluates the objective and non-linear convex function and  $W \in \mathbb{R}^{q \times n}$  (where  $q$  denotes the features and  $n$  is the number of data vectors in  $v$ ).  $L1$  denotes the sparsity penalty and has a tied reconstruction matrix  $W$ . To decrease the computational cost of the optimization, L-BFGS method [46] is used as a constrained optimizer that results in fast convergence. Moreover, RICA can manage data with approximate whitening or even without whitening [33].

As illustrated in the formulation of RICA in equation (4), in the first part,  $\lambda$  represents the weight assigned to the sparsity constraint in relation to the recreation condition. The second part emphasizes accurate recreation of the original features by minimizing the recreation error  $\|W^T Wx - x\|$ . In this regard, we fix the feature dimension of pooling layer features equal to 400, and empirically set the  $\lambda$  weights to 80, 100, and 120 because the higher the weight we give to the sparsity

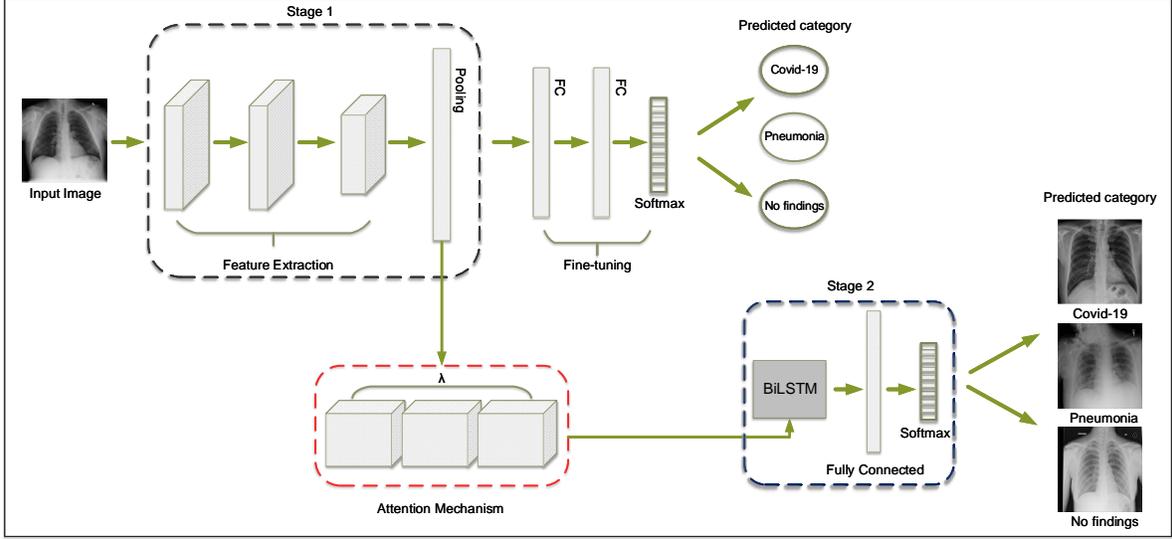


Figure 3: Flowchart of the proposed method. First, fine-tuning is performed to train a convolutional neural network (CNN) for COVID-19 detection. An attention mechanism based on reconstruction independent component analysis is used to leverage the most relevant CNN features and form a sequence of augmented features, which are then fed into a Bidirectional Long Short-Term Memory (LSTM) network for the final representation.

constraint the less precise will the recreation be, and vice-versa. Hence, we obtained three augmented feature vectors with different weights by keeping the same dimension 400 as mentioned above. Similarly, we repeat the same procedure to obtain three augmented features sets for the second dataset by setting the feature dimension to 500. The representations learned by the attention mechanism contain discriminative information related to the classes, which allows the network to accurately predict them.

### C. Recurrent Neural Network (RNN)

Mainstream CNN frameworks are related to conventional statistical models, thus lacking the capacity to map sequences to sequences. BiLSTM [48] is one kind of RNN, which has the ability to process sequences of arbitrary length, and has obtained surprising performance in text classification. However, the high dimensionality and sparsity of the data are one of the major challenges that limit its performance. Taking advantage of the low-dimensional RICA features, BiLSTM performs better than using the raw CNN features (we further discuss this argument in Section 4.4). The BiLSTM is implemented similarly to the standard bi-directional LSTM except that the input is based on three augmented features. We found that the proposed strategy calculated on each time step resulted in improved reconstructions, which we found to be vital to accomplish our feature augmentation process.

BiLSTM Networks capture each sequence vector based on the memory cell ( $C_e$ ), and compete for retaining dependencies between the elements in the input sequence. It is comprised of an input gate ( $i_e$ ), an output gate ( $o_e$ ) and a forget gate ( $g_e$ ). The input gate governs the information flow into the cell by multiplying the cell's non-linear transformation of inputs  $m_e$ . The output gate decides how much information from the cell

is used to compute the output activation of the LSTM unit. The forget gate regulates the extent to which a value remains in the cell. The LSTM unit updates for time step  $e$  are:

$$\begin{bmatrix} g_e \\ i_e \\ m_e \\ o_e \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \tanh \\ \sigma \end{bmatrix} H \cdot [p_{e-1}, x_e] \quad (5)$$

$$C_e = g_e \odot C_{e-1} + m_e \odot i_e \quad (6)$$

$$p_e = \tanh(C_e) \odot o_e \quad (7)$$

where  $x_e$  is the input at the current time-step,  $i_e$  is the current cell state,  $g$ ,  $i$  and  $m$  is the input gate activation, forget gate activation and output gate activation respectively,  $\sigma$  illustrates the logistic sigmoid function and  $\odot$  represents element-wise multiplication.

## IV. Experiments and Results

In this section, we first provide a brief description of two databases that are used to evaluate our method. Then, we present evaluation metrics, implementation details, and experimental results, which are later on discussed and compared to state-of-the-art methods.

### A. Datasets

*X-ray image dataset:* The first dataset used in the experiments is obtained from the open access sources provided by Cohen JP [11] and Wang et al.[61]. It contains three classes: pneumonia, "no-findings", and COVID-19, where pneumonia, and "no-findings" have 500 images each and COVID-19 has 125 X-ray scan images. A detailed explanation can be found

Table I: The performance comparison results for the three-class X-ray image database.

Method	Performance Metrics (%)				
	Sensitivity	Specificity	Precision	F1-Score	Accuracy
SqueezeNet	70.00	80.90	82.87	69.59	75.66
SqueezeNet+BiLSTM	89.00	93.60	93.30	90.40	91.89
SqueezeNet+SA+BiLSTM	<b>94.66</b>	<b>95.73</b>	<b>95.40</b>	<b>94.63</b>	<b>95.25</b>
GoogLeNet	84.66	88.93	83.41	83.06	87.25
GoogLeNet+BiLSTM	92.00	94.10	93.58	93.49	93.55
GoogLeNet+SA+BiLSTM	<b>93.66</b>	<b>96.73</b>	<b>96.67</b>	<b>94.66</b>	<b>96.10</b>
DenseNet201	89.66	92.03	87.04	88.18	90.81
DenseNet201+BiLSTM	93.33	94.86	91.27	92.42	94.96
DenseNet201+SA+BiLSTM	<b>94.00</b>	<b>95.20</b>	<b>96.02</b>	<b>94.99</b>	<b>96.40</b>
ResNet-50	86.33	90.86	86.45	86.30	88.62
ResNet-50+BiLSTM	88.41	94.86	93.48	92.11	94.62
ResNet-50+SA+BiLSTM	<b>93.33</b>	<b>96.46</b>	<b>96.34</b>	<b>94.33</b>	<b>97.26</b>

in [40].

*COVID-19 X-ray scan database:* The second dataset is collected from the open access source provided by Vantaggiato et al.[57], where two scenarios are utilized. In the first scenario, three classes (Normal, COVID-19, and Pneumonia) are provided in the dataset. For training, each class has 404 images. Validation and testing set contain 100 and 207 images, respectively. In the second scenario, two more classes were added by the authors yielding a five class model: Normal, COVID-19, Viral-Pneumonia, Bacterial-Pneumonia and Lung-Opacity. This dataset was acquired from different open access sources [11, 50, 27, 31].

### B. Evaluation Metrics

The performance of the proposed method is evaluated with respect to Sensitivity, Specificity, Precision, F-Score and Accuracy, defined using the equations below:

$$Sensitivity = \frac{TP}{TP + FN} * 100 \quad (8)$$

$$Specificity = \frac{TN}{TN + FP} * 100 \quad (9)$$

$$Precision = \frac{TP}{TP + FP} * 100 \quad (10)$$

$$F-Score = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity} * 100 \quad (11)$$

$$Accuracy = \frac{Number\ of\ correct\ prediction}{Total\ number\ of\ prediction} * 100 \quad (12)$$

Where TP, FP and FN represent the True Positive, False Positive and False Negative, respectively.

In addition to the above, the ROC (Receiver Operating Characteristics) curves have been utilized as well. This consists of graphs showing the variation of True Positive Rate (TPR) or, equivalently, *Sensitivity* as defined in Eq. (8), with respect to False Positive Rate (FPR), where FPR is defined as:

$$FPR = \frac{FP}{FP + TN} * 100 \quad (13)$$

Especially, ROC exhibits the performance of the underlined classification model at all classification thresholds.

Table II: The performance comparison with other deep learning methods using radiology images. The comparison results are obtained from [40].

Method	Type of Images	Number of Cases	Method Used	Accuracy (%)
Ioannis et al. [4]	Chest X-ray	224 COVID-19 (+) 700 Pneumonia 504 Healthy	VGG-19	93.48
Wang et al. [59]	Chest X-ray	53 COVID-19 (+) 5526 COVID-19 (-) 8066 Healthy	COVID-Net	92.40
Sethy et al. [49]	Chest X-ray	25 COVID-19 (+) 25 COVID-19 (-)	EsNet50+SVM	95.38
Hemdan et al. [24]	Chest X-ray	25 COVID-19 (+) 25 Normal	COVIDX-Net	90.00
Narin et al. [38]	Chest X-ray	50 COVID-19 (+) 50 COVID-19 (-)	Deep CNN	98.00
Ying et al. [53]	Chest CT	777 COVID-19 (+) 708 Healthy	DRE-Net	86.00
Wang et al. [60]	Chest CT	195 COVID-19 (+) 258 COVID-19 (-)	M-Inception	82.90
Zheng et al. [66]	Chest CT	313 COVID-19 (+) 229 COVID-19 (-)	UNet+3D Deep Network	90.80
Xu et al. [63]	Chest CT	219 COVID-19 (+) 224 Viral Pneumonia 175 Healthy	ResNet+Location Attention	86.70
Tulin et al. [40]	Chest X-ray	125 COVID-19 (+) 500 No Findings	DarkCovidNet	98.08
		125 COVID-19 (+) 500 Pneumonia 500 No Findings	DarkCovidNet	87.02
Proposed Method	Chest X-ray	125 COVID-19 (+) 500 No Findings	ResNET50+BiLSTM	<b>99.10</b>
		125 COVID-19 (+) 500 Pneumonia 500 No Findings	ResNET50+SA+BiLSTM	<b>97.26</b>

### C. Implementation details

All the X-ray images are resized to  $224 \times 224$  based on the size requirement of the model. No image data augmentation was applied, thus the features are augmented only in the feature space. The CNN is fine-tuned using stochastic gradient descent (SGD) with a learning rate of  $3e-4$ , mini-batch size of 32 and epochs of 5, and by shuffling every epoch. The feature vector is extracted from the output of the last pooling layer with a size of 2048. RICA is utilized to augment the pooling layer features into three augmented feature sets with dimension of 400 to 500 for X-ray image dataset and COVID-19 X-ray scan database, respectively. The performance also varies by varying the number of iterations done by RICA before stopping, and 80 to 120 number of iterations were used to extract the augmented features.

The BiLSTM sub-network is trained using the cross entropy loss and Adam optimizer is utilized by fixing the learning rate to 0.0001 and the hidden layer dimension to 60, and number of pochs to 150 on all databases. Initializing the BiLSTM sub-network with the right weights can be difficult because using the standard gradient descent from random initialization can hamper the training of BiLSTM. Therefore, we set the

Table III: The performance comparison results for the three-class COVID-19 database.

Method	Performance Metrics (%)				
	Accuracy	Precision	Sensitivity	Specificity	F1-Score(%)
SqueezeNet	66.60±0.44	61.25	49.75	74.87	48.57
SqueezeNet+BiLSTM	73.26±0.82	59.90	65.30	83.91	56.68
SqueezeNet+SA+BiLSTM	<b>76.57±1.47</b>	<b>64.57</b>	<b>67.28</b>	<b>84.01</b>	<b>63.29</b>
GoogLeNet	72.38±1.47	70.82	67.08	80.54	65.57
GoogLeNet+BiLSTM	76.49±1.47	67.24	73.97	85.62	62.19
GoogLeNet+SA+BiLSTM	<b>77.61±2.47</b>	<b>74.42</b>	<b>78.79</b>	<b>86.15</b>	<b>73.97</b>
DenseNet201	75.19±2.47	71.17	67.69	81.34	64.65
DenseNet201+BiLSTM	76.19±2.47	61.03	67.19	82.55	62.30
DenseNet201+SA+BiLSTM	<b>78.90±3.05</b>	<b>75.48</b>	<b>78.39</b>	<b>81.36</b>	<b>69.75</b>
ResNet-50	74.53±1.25	73.88	73.10	86.55	68.97
ResNet-50+BiLSTM	77.53±3.53	71.49	77.93	87.71	69.64
Ensemble-CNNs [57]	75.23±3.40	78.28	75.20	87.60	73.43
ResNet-50+SA+BiLSTM	<b>79.53±1.41</b>	<b>74.55</b>	<b>81.52</b>	<b>88.42</b>	<b>71.96</b>

Table IV: The performance comparison results for the five-class COVID-19 database.

Method	Performance Metrics (%)				
	Accuracy	Precision	Sensitivity	Specificity	F1-Score(%)
SqueezeNet	72.88±1.17	63.42	58.45	89.61	55.62
SqueezeNet+BiLSTM	79.55±2.17	61.73	67.98	91.22	55.41
SqueezeNet+SA+BiLSTM	<b>81.78±2.41</b>	<b>67.43</b>	<b>75.10</b>	<b>92.75</b>	<b>61.27</b>
GoogLeNet	73.98±1.88	68.01	63.28	90.82	62.55
GoogLeNet+BiLSTM	77.68±1.47	70.53	95.87	93.87	62.43
GoogLeNet+SA+BiLSTM	<b>80.77±1.13</b>	<b>73.62</b>	<b>77.40</b>	<b>94.03</b>	<b>71.36</b>
DenseNet201	82.51±2.40	73.02	72.36	93.09	70.14
DenseNet201+BiLSTM	81.10±1.54	84.54	84.99	96.29	83.95
DenseNet201+SA+BiLSTM	<b>84.10±1.90</b>	<b>83.76</b>	<b>87.18</b>	<b>96.22</b>	<b>83.25</b>
ResNet-50	79.64±1.84	75.94	71.49	92.87	68.29
ResNet-50+BiLSTM	83.64±1.84	67.14	77.93	92.41	60.71
Ensemble-CNNs [57]	81.00±2.39	82.99	82.96	85.24	81.49
ResNet-50+SA+BiLSTM	<b>84.64±2.55</b>	<b>76.32</b>	<b>83.65</b>	<b>90.82</b>	<b>73.88</b>

recurrent weights with *He* initializer [22] which performs the best in all scenarios of our experiments.

#### D. Experimental results

In order to detect COVID-19 on the first X-ray image database [11, 61], we split the original image dataset to eighty percent for training and twenty percent for testing in the same spirit as [40]. We report the results in terms of average classification accuracy, sensitivity, specificity, precision, and F1-score values in Table 1. First, we show the performance of the fine-tuned CNN (Resnet-50) and the three other popular state-of-the-art CNNs models; namely, Squeeze-Net, Google-Net, and DenseNet-201. Second, we compare the performance when using raw CNN features with BiLSTM. In the third scenario, we show the efficacy of the proposed attention mechanism with a combined CNN-BiLSTM. It can be noted

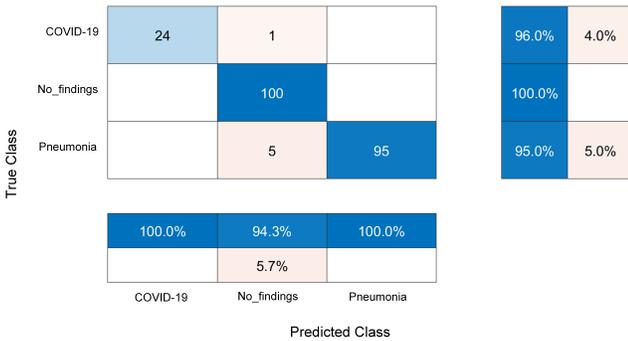


Figure 4: Confusion matrix of the three-class using ResNET50+SA+BiLSTM for X-ray image dataset [40]. The horizontal and vertical axis is for predicted and true classes, respectively.

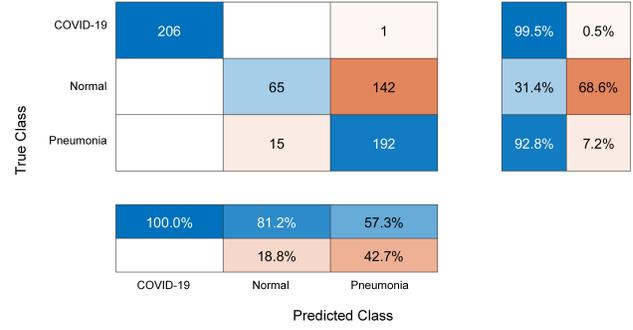


Figure 5: Confusion matrix of the three-class using ResNET50+SA+BiLSTM for COVID-19 X-ray scan database [57]. The horizontal and vertical axis is for predicted and true class, respectively.

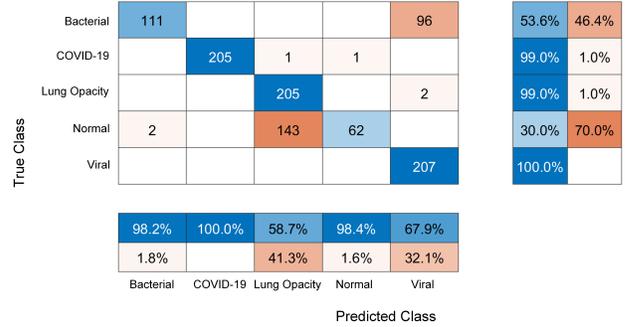


Figure 6: Confusion matrix of the five-class using ResNET50+SA+BiLSTM for COVID-19 X-ray scan database [57]. The horizontal and vertical axis is for predicted and true class, respectively.

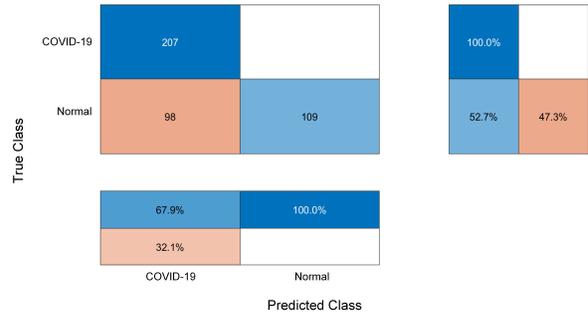
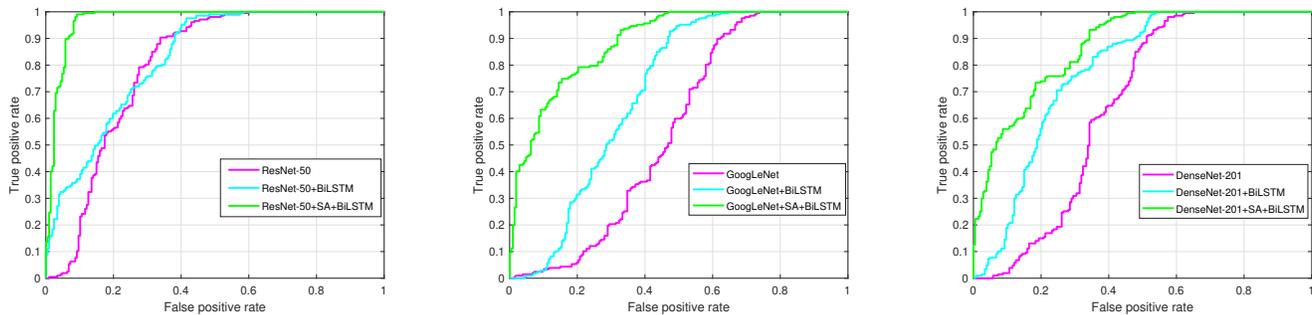
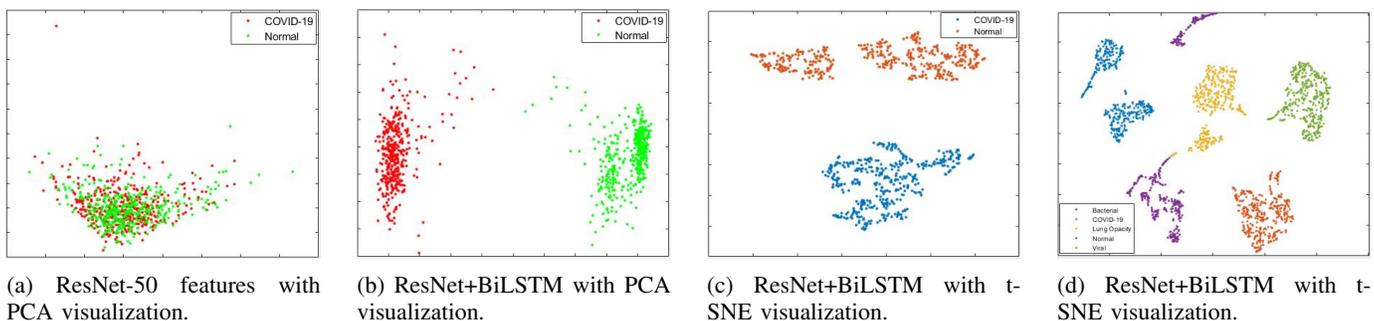


Figure 7: Two class confusion matrix for COVID-19 X-ray scan database. The horizontal and vertical axis is for predicted and true class, respectively.



(a) The performance evaluation with the ResNet-50 model. (b) The performance evaluation with the GoogLeNet model. (c) The performance evaluation with the DenseNet-201 model.

Figure 8: The Receiver Operating Characteristics (ROC) curves of the COVID-19 X-ray scan database, where the true positive rate represents the y-axis and false positive rate represents the x-axis.



(a) ResNet-50 features with PCA visualization. (b) ResNet+BiLSTM with PCA visualization. (c) ResNet+BiLSTM with t-SNE visualization. (d) ResNet+BiLSTM with t-SNE visualization.

Figure 9: Two-dimensional scatter plots of ResNet features with PCA and t-SNE over COVID-19 X-ray database. The clusters correspond to two and five different classes available in the data.

that single CNN exhibits a limited performance to detect three classes: COVID-19, No-Findings, and Pneumonia. One of the reasons is that a number of training samples in the COVID-19 class resembles Pneumonia class and this is not enough to compare to the other two classes (No-Findings and Pneumonia), which might cause overfitting of the model.

Table 2 summarizes the comparison with the state-of-the-art works. The final performance is reported using a 5-fold cross-validation procedure and the comparison results are obtained from [40]. The proposed method provides a 97% accuracy which is 10% higher than the previously proposed method on the three classes, and 99% accuracy for the two-classes dataset. For further analysis, a confusion matrix is shown in Fig.4. It can be observed that the proposed approach classified COVID-19 better than the other two classes.

Table 3 illustrates the results of the three-class scenario on the second COVID-19 database [57]. The training set, validation set, and testing set is provided separately. Using the self-attention mechanism, the proposed method achieves 79% accuracy and improves 4% performance from the previous study [57]. In addition to the three-class scenario, the results of the five-class scenario is also reported in Table 4. It can be observed that the proposed method provides 84% accuracy which is 3% better than the previously proposed method in [57]. For further analysis, Figure 5 and 6 represent the confusion matrices of the three-class and the five-class, respectively. The main observation is that the proposed method attained 100%

accuracy for the detection of COVID-19 samples. A confusion matrix in Fig.7 is illustrated only for COVID-19 and Normal class with two rows and two columns showing the number of true positives, false negatives, false positives and true negatives. It shows that the model predicted all 207 COVID-19 X-ray images correctly, and no false negative is detected. In the case of a normal class, 98 images are misclassified while 109 images were correctly classified. All correct predictions are located in the diagonal of the table (highlighted in light blue and dark blue), so it is easy to visually inspect the table for prediction errors. Therefore, we can observe that the proposed framework is efficient to distinguish the COVID-19 samples in both datasets.

As our main focus is to detect COVID-19 samples, we use ROC curves for a two-class detection problem (COVID-19 vs Normal), in which only the true positive rate (TPR) and false positive rate (FPR) are needed. The best possible detection method would allow a learning curve in the upper left corner or coordinate (0,1) of the ROC space, depicting 100% sensitivity (no false negatives) and 100% specificity (no false positives). In Fig.8, the curves are visualized with raw CNN features, CNN with BiLSTM-based network, and finally with an attention mechanism. It could be observed that the proposed attention mechanism clearly improves the performance of a CNN-based BiLSTM architecture as exhibited by higher sensitivity rate.

Based on experimental analysis on both datasets, we con-

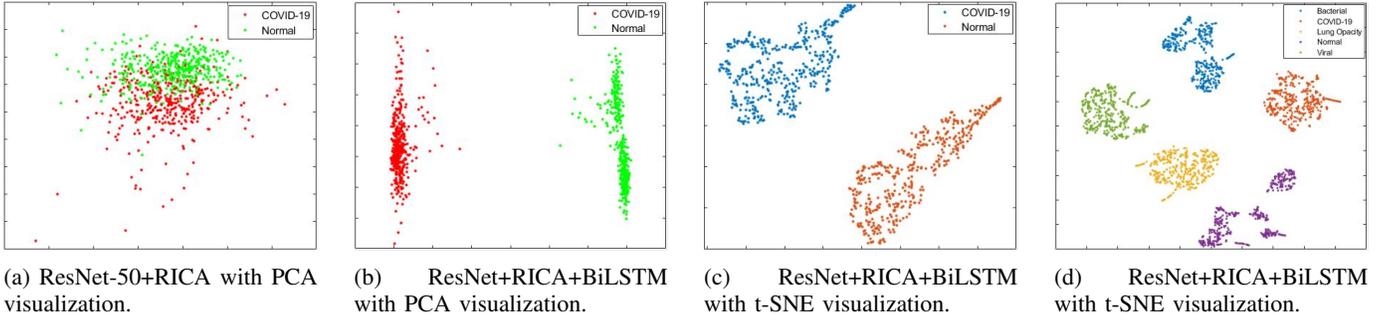


Figure 10: Two-dimensional scatter plots of augmented features with PCA and t-SNE over COVID-19 X-ray database. The clusters correspond to two and five different classes available in the data.

cluded that neither a single CNN model nor CNN-based BiLSTM achieves the best results for all the evaluation metrics. Therefore, the proposed attention mechanism is essential to produce a robust feature representation for COVID-19 detection.

### E. Explainability Analysis

One of the advantages of the proposed approach is that we can interpret the detection process of the model. Especially for each stage, we can see how the features are structured into the high-dimensional (ResNet) and the impact of the augmented feature space along the different classification stages. Taking this into consideration, we employed the PCA projection [28] and the "t-distributed stochastic neighboring embedding" (t-SNE) algorithm [35]. Strictly speaking, PCA offers a nice explanatory framework since its axes are made of a linear combination of the original dimensions, allowing comprehension of high dimension patterns. Similarly, t-distributed Stochastic Neighbor Embedding (t-SNE) creates low-dimensional representation of complex high dimensional data through a series of transformation and fine-tuned optimization procedures. In this respect, the projection results of either PCA or t-SNE provide a rough indication of the quality of the separation and supporting explainability through visual exploration.

The ResNet features in Fig.9 (a) indicate that both classes (COVID-19 vs Normal) are strongly correlated, which makes the BiLSTM hard to separate them as shown in Fig.9 (b) and Fig.9 (c). We also visualize the five class features in Fig.9 (d) and observed that the Normal class is still correlated with the Lung Opacity class that causes overfitting of CNN-BiLSTM architecture. Thus, the derived clusters indicate that the prior information obtained from raw CNN features causes to decrease the performance. On the other side, augmented features generated by RICA reduces the correlation between similar classes as shown in Fig.10 (a) and are able to capture more variability in the feature space. Moreover, it could be noticed from Fig.10 (b) and Fig.10 (c) that data points corresponding to Normal and COVID-19 are linearly separable which could potentially lead to better performance when training BiLSTM on low-dimensional data. Thus, the attention mechanism helps to overcome the overfitting issue and also separate the five classes efficiently in comparison to raw CNN features as shown in Fig.10 (d).

It should be noted that our exploration through visualization as a way to achieve explainability can be further expanded in different directions. First, projection quality metric can be used to assess the quality of each projection by PCA or t-SNE using. This includes global measures such that Normalized Stress, Distance Consistency, ClustMe [1], or local measures such as projection precision score [47] can contribute to shed light on the quality of such projections. Nevertheless, it should be noted that such assessment may also be misleading and cannot contribute towards comprehending *why such results occurred*. In this context, one shall mention the interesting work of Fujiwara et al. [20] who proposed a *contrasting clusters in PCA* (ccPCA) as a way to find out which dimensions contributed more to the formation of a selected cluster and why it differed from the rest of the dataset, based on information on separation and internal versus external variability.

## V. Conclusion

In this study, we address the problem of COVID-19 detection from X-ray images. For this purpose, a unified architecture constituted of a deep convolutional neural network (CNN), an attention mechanism, and a bidirectional LSTM (BiLSTM) is proposed. The CNN provides the high-level features extracted at the pooling layer where the attention mechanism selects the most relevant features and generates low-dimensional augmented features. Finally, BiLSTM is used to classify the processed sequential information. The proposed method provides an end-to-end structure without the need for manual feature extraction. We showed that the detection of COVID-19 was improved by using the low-dimensional augmented features through a reconstruction independent component analysis method. Extensive experiments on two publicly available COVID-19 X-ray image datasets where a comparison with state-of-art models; namely, Squee-ze-Net, Google-Net, and DenseNet-201 and recently published works showed that our newly designed CNN-based BiLSTM architecture outperformed several state-of-the-art models.

Especially, our model achieved a 97% accuracy which is 10% higher than the best performing model published so far in the literature on the three classes, and 99% accuracy for the two-classes dataset. While in the five class case, our model achieved 84% accuracy which is 3% better than the previously proposed method in [57]. At some other scenarios,

the developed model has demonstrated the ability to achieve 100% accuracy for the detection of COVID-19 samples. On the other hand, we also showed the possibility to utilize the componentwise property of the overall architecture where each stage (component) can be used to generate explanations that can be employed to comprehend the actions of the model. Especially, the attention mechanism provides an explainable AI model that can detect COVID-19 robustly. Explainability through PCA and t-SNE have also been explored and duly commented as well as highlighting the potential deficiencies that may restrict the ability of PCS or t-SNE projection to provide an answer to the "why" question in the explainability, while the prospect of a newly introduced ccPCA has been recognized. In the future, we plan to further robustify the feature selection method and RICA analysis in the attention mechanism layer in a way to enhance the explainability of the results and develop joint visualization approach that can comprehend both PCA, t-SNE projection outcomes with attention weights.

## VI. Declaration of Competing Interest

The authors have no conflict of interest that could have appeared to influence the work reported in this paper.

## VII. Acknowledgments

The research work of this paper was supported by the Center for Machine Vision and Signal Analysis (CMVS) in the Faculty of Information Technology and Electrical Engineering (ITEE), University of Oulu, Finland, and the Department of Pathology and Anatomical Sciences, University at Buffalo, USA. The authors are grateful to the Academy of Finland Profi5 DigiHealth project (#326291). The authors are also thankful to Samuel P Border (University at Buffalo, USA) and Jiehua Zhang (University of Oulu, Finland) for their valuable suggestions.

## References

- [1] Abbas, M.M., Aupetit, M., Sedlmair, M., Bensmail, H., 2019. Clustme: A visual quality measure for ranking monochrome scatterplots based on cluster patterns, in: *Computer Graphics Forum*, Wiley Online Library. pp. 225–236.
- [2] Al-Waisy, A., Mohammed, M.A., Al-Fahdawi, S., Maashi, M., Garcia-Zapirain, B., Abdulkareem, K.H., Mostafa, S., Kumar, N.M., Le, D.N., 2021. Covid-deepnet: hybrid multimodal deep learning system for improving covid-19 pneumonia detection in chest x-ray images. *Computers, Materials and Continua* 67.
- [3] Albahli, S., 2020. Efficient gan-based chest radiographs (cxr) augmentation to diagnose coronavirus disease pneumonia. *International journal of medical sciences* 17, 1439.
- [4] Apostolopoulos, I.D., Mpesiana, T.A., 2020. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine* 43, 635–640.
- [5] Aslan, M.F., Unlarsen, M.F., Sabanci, K., Durdu, A., 2021. Cnn-based transfer learning–ilstm network: A novel approach for covid-19 infection detection. *Applied Soft Computing* 98, 106912.
- [6] Bernheim, A., Mei, X., Huang, M., Yang, Y., Fayad, Z.A., Zhang, N., Diao, K., Lin, B., Zhu, X., Li, K., et al., 2020. Chest ct findings in coronavirus disease-19 (covid-19): relationship to duration of infection. *Radiology* , 200463.
- [7] Berrimi, M., Hamdi, S., Cherif, R.Y., Moussaoui, A., Oussalah, M., Chabane, M., 2021. Covid-19 detection from xray and ct scans using transfer learning, in: *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*, IEEE. pp. 1–6.
- [8] Castiglioni, I., Ippolito, D., Interlenghi, M., Monti, C.B., Salvatore, C., Schiaffino, S., Polidori, A., Gandola, D., Messa, C., Sardanelli, F., 2020. Artificial intelligence applied on chest x-ray can aid in the diagnosis of covid-19 infection: a first experience from lombardy, italy. *medRxiv* .
- [9] Chan, J.F.W., Yuan, S., Kok, K.H., To, K.K.W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C.Y., Poon, R.W.S., et al., 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The lancet* 395, 514–523.
- [10] Chouhan, N., Khan, A., Shah, J.Z., Hussnain, M., Khan, M.W., 2021. Deep convolutional neural network and emotional learning based breast cancer detection using digital mammography. *Computers in Biology and Medicine* 132, 104318.
- [11] Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., Ghassemi, M., 2020. Covid-19 image data collection: Prospective predictions are the future. *arXiv preprint arXiv:2006.11988* .
- [12] Das, A.K., Ghosh, S., Thunder, S., Dutta, R., Agarwal, S., Chakrabarti, A., 2021. Automatic covid-19 detection from x-ray images using ensemble learning with convolutional neural network. *Pattern Analysis and Applications* , 1–14.
- [13] DeVries, T., Taylor, G.W., 2017. Improved regularization of convolutional neural networks with cutout. *arXiv:1708.04552* .
- [14] Donahue, J., Krähenbühl, P., Darrell, T., 2016. Adversarial feature learning. *arXiv preprint arXiv:1605.09782* .
- [15] Dong, X., Cao, Y.y., Lu, X.x., Zhang, J.j., Du, H., Yan, Y.q., Akdis, C.A., Gao, Y.d., 2020. Eleven faces of coronavirus disease 2019. *Allergy* 75, 1699–1709.
- [16] Elbishlawi, S., Abdelpakey, M.H., Shehata, M.S., Mohamed, M.M., 2021. Corona-net: Diagnosing covid-19 from x-ray images using re-initialization and classification networks. *Journal of Imaging* 7, 81.
- [17] Falzone, L., Gattuso, G., Tsatsakis, A., Spandidos, D.A., Libra, M., 2021. Current and innovative methods for the diagnosis of covid-19 infection. *International journal of molecular medicine* 47, 1–23.
- [18] Feng, J., Zhang, S.W., Chen, L., Xia, J., Initiative,

- A.D.N., et al., 2021. Alzheimer’s disease classification using features extracted from nonsubsampling contourlet subband-based individual networks. *Neurocomputing* 421, 260–272.
- [19] Fu, Y., Hospedales, T.M., Xiang, T., Gong, S., 2015. Transductive multi-view zero-shot learning. *IEEE transactions on pattern analysis and machine intelligence* 37, 2332–2345.
- [20] Fujiwara, T., Kwon, O.H., Ma, K.L., 2019. Supporting analysis of dimensionality reduction results with contrastive learning. *IEEE transactions on visualization and computer graphics* 26, 45–55.
- [21] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. *arXiv preprint arXiv:1406.2661* .
- [22] He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034.
- [23] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- [24] Hemdan, E.E.D., Shouman, M.A., Karar, M.E., 2020. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:2003.11055* .
- [25] Hoque, M.Z., Keskinarkaus, A., Nyberg, P., Seppänen, T., 2021. Retinex model based stain normalization technique for whole slide image analysis. *Computerized Medical Imaging and Graphics* 90, 101901.
- [26] of the International, C.S.G., et al., 2020. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2. *Nature microbiology* 5, 536.
- [27] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al., 2019. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 590–597.
- [28] Jolliffe, I.T., 2002. Principal components in regression analysis. *Principal component analysis* , 167–198.
- [29] Karbhari, Y., Basu, A., Geem, Z.W., Han, G.T., Sarkar, R., 2021. Generation of synthetic chest x-ray images and detection of covid-19: A deep learning based approach. *Diagnostics* 11, 895.
- [30] Karras, T., Aila, T., Laine, S., Lehtinen, J., 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* .
- [31] Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., et al., 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172, 1122–1131.
- [32] Kong, W., Agarwal, P.P., 2020. Chest imaging appearance of covid-19 infection. *Radiology: Cardiothoracic Imaging* 2, e200028.
- [33] Le, Q., Karpenko, A., Ngiam, J., Ng, A., 2011. Ica with reconstruction cost for efficient overcomplete feature learning. *Advances in neural information processing systems* 24, 1017–1025.
- [34] Loey, M., Smarandache, F., M Khalifa, N.E., 2020. Within the lack of chest covid-19 x-ray dataset: a novel detection model based on gan and deep transfer learning. *Symmetry* 12, 651.
- [35] Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-sne. *Journal of machine learning research* 9.
- [36] de Melo, W.C., Granger, E., Hadid, A., 2020. A deep multiscale spatiotemporal network for assessing depression from facial dynamics. *IEEE Transactions on Affective Computing* .
- [37] Muhammad, U., Holmberg, T., de Melo, W.C., Hadid, A., 2019. Face anti-spoofing via sample learning based recurrent neural network (rnn), in: *BMVC*, p. 113.
- [38] Narin, A., Kaya, C., Pamuk, Z., 2021. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Analysis and Applications* , 1–14.
- [39] Nour, M., Cömert, Z., Polat, K., 2020. A novel medical diagnosis model for covid-19 infection detection based on deep features and bayesian optimization. *Applied Soft Computing* 97, 106580.
- [40] Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Acharya, U.R., 2020. Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in biology and medicine* 121, 103792.
- [41] Ozyurt, F., Tuncer, T., Subasi, A., 2021. An automated covid-19 detection based on fused dynamic exemplar pyramid feature extraction and hybrid feature selection using deep learning. *Computers in Biology and Medicine* 132, 104356.
- [42] Radford, A., Metz, L., Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* .
- [43] Rahimzadeh, M., Attar, A., 2020. A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2. *Informatics in Medicine Unlocked* 19, 100360.
- [44] Rajaraman, S., Antani, S., 2020. Weakly labeled data augmentation for deep learning: A study on covid-19 detection in chest x-rays. *Diagnostics* 10, 358.
- [45] Sahlol, A.T., Yousri, D., Ewees, A.A., Al-Qaness, M.A., Damasevicius, R., Abd Elaziz, M., 2020. Covid-19 image classification using deep features and fractional-order marine predators algorithm. *Scientific Reports* 10, 1–15.
- [46] Schmidt, M., . Minfunc (a matlab function for unconstrained optimization of differentiable real-valued multivariate functions using line-search methods)(2005). URL <http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html> .

- [47] Schreck, T., Von Landesberger, T., Bremm, S., 2010. Techniques for precision-based visual analysis of projected data. *Information Visualization* 9, 181–193.
- [48] Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* 45, 2673–2681.
- [49] Sethy, P.K., Behera, S.K., Ratha, P.K., Biswas, P., 2020. Detection of coronavirus disease (covid-19) based on deep features and support vector machine .
- [50] Shih, G., Wu, C.C., Halabi, S.S., Kohli, M.D., Prevedello, L.M., Cook, T.S., Sharma, A., Amorosa, J.K., Arteaga, V., Galperin-Aizenberg, M., et al., 2019. Augmenting the national institutes of health chest radiograph dataset with expert annotations of possible pneumonia. *Radiology: Artificial Intelligence* 1, e180041.
- [51] Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *Journal of Big Data* 6, 1–48.
- [52] Singhal, T., 2020. A review of coronavirus disease-2019 (covid-19). *The indian journal of pediatrics* 87, 281–286.
- [53] Song, Y., Zheng, S., Li, L., Zhang, X., Zhang, X., Huang, Z., Chen, J., Wang, R., Zhao, H., Zha, Y., et al., 2021. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* .
- [54] Tavakolian, M., Hadid, A., 2019. A spatiotemporal convolutional neural network for automatic pain intensity estimation from facial dynamics. *International Journal of Computer Vision* 127, 1413–1425.
- [55] Toğaçar, M., Ergen, B., Cömert, Z., 2020. Covid-19 detection using deep learning models to exploit social mimic optimization and structured chest x-ray images using fuzzy color and stacking approaches. *Computers in biology and medicine* 121, 103805.
- [56] Turkoglu, M., 2021. Covidetectionet: Covid-19 diagnosis system based on x-ray images using features selected from pre-learned deep features ensemble. *Applied Intelligence* 51, 1213–1226.
- [57] Vantaggiato, E., Paladini, E., Bougourzi, F., Distanti, C., Hadid, A., Taleb-Ahmed, A., 2021. Covid-19 recognition using ensemble-cnns in two new chest x-ray databases. *Sensors* 21, 1742.
- [58] Vasilescu, V., Neacșu, A., Chouzenoux, E., Pesquet, J.C., Burileanu, C., 2021. A deep learning approach for improved segmentation of lesions related to covid-19 chest ct scans, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 635–639.
- [59] Wang, L., Lin, Z.Q., Wong, A., 2020. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports* 10, 1–12.
- [60] Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Cai, M., Yang, J., Li, Y., Meng, X., et al., 2021a. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). *European radiology* , 1–9.
- [61] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M., 2017. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106.
- [62] Wang, Y., Zhang, Y., Liu, Y., Tian, J., Zhong, C., Shi, Z., Zhang, Y., He, Z., 2021b. Does non-covid-19 lung lesion help? investigating transferability in covid-19 ct image segmentation. *Computer Methods and Programs in Biomedicine* 202, 106004.
- [63] Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Ni, Q., Chen, Y., Su, J., et al., 2020. A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* 6, 1122–1129.
- [64] Yoo, S.H., Geng, H., Chiu, T.L., Yu, S.K., Cho, D.C., Heo, J., Choi, M.S., Choi, I.H., Cung Van, C., Nhung, N.V., et al., 2020. Deep learning-based decision-tree classifier for covid-19 diagnosis from chest x-ray imaging. *Frontiers in medicine* 7, 427.
- [65] Zhao, W., Zhong, Z., Xie, X., Yu, Q., Liu, J., 2020. Relation between chest ct findings and clinical conditions of coronavirus disease (covid-19) pneumonia: a multicenter study. *American Journal of Roentgenology* 214, 1072–1077.
- [66] Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W., Wang, X., 2020. Deep learning-based detection for covid-19 from chest ct using weak label. *MedRxiv* .
- [67] Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y., 2020. Random erasing data augmentation, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 13001–13008.
- [68] Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232.
- [69] Zu, Z.Y., Jiang, M.D., Xu, P.P., Chen, W., Ni, Q.Q., Lu, G.M., Zhang, L.J., 2020. Coronavirus disease 2019 (covid-19): a perspective from china. *Radiology* 296, E15–E25.