AC-CovidNet: Attention Guided Contrastive CNN for Recognition of Covid-19 in Chest X-Ray Images

Anirudh Ambati
Computer Vision Group
IIIT Sri City
anirudh.a17@iiits.in

Shiv Ram Dubey
Computer Vision Group
IIIT Sri City
srdubey@iiits.in

Abstract

Covid-19 global pandemic continues to devastate health care systems across the world. In many countries, the 2nd wave is very severe. Economical and rapid testing, as well as diagnosis, is urgently needed to control the pandemic. At present, the Covid-19 testing is costly and time-consuming. Chest X-Ray (CXR) testing can be the fastest, scalable, and non-invasive method. The existing methods suffer due to the limited CXR samples available from Covid-19. Thus, inspired by the limitations of the open-source work in this field, we propose attention guided contrastive CNN architecture (AC-CovidNet) for Covid-19 detection in CXR images. The proposed method learns the robust and discriminative features with the help of contrastive loss. Moreover, the proposed method gives more importance to the infected regions as guided by the attention mechanism. We compute the sensitivity of the proposed method over the publicly available Covid-19 dataset. It is observed that the proposed AC-CovidNet exhibits very promising performance as compared to the existing methods even with limited training data. It can tackle the bottleneck of CXR Covid-19 datasets being faced by the researchers. The code used in this paper is released publicly at https://github.com/shivram1987/AC-CovidNet/.

1. Introduction

Coronavirus disease 2019 (Covid-19) has emerged very fast as an emergent health risk disease affecting the whole world. It is caused by Novel Coronavirus that leads to Pneumonia-like symptoms, including cough, fever, and/or difficulty in breathing. It has been observed that this infection spreads through the surfaces which might be infected from the infected person. Thus, once the spread rate increases, it becomes uncontrollable due to a chain reaction like infection spread. The spread of Covid-19 is categorized into different stages. The stage1 and stage2 refer to the small scale spread, whereas stage3 and beyond refer to the large scale spread due to chain reaction.

As per the World Health Organisation (WHO), Covid-19 was first observed on 31st December 2019 as the reported case from Wuhan city of China. The outbreak was declared a Public Health Emergency of International Concern on 30th January 2020. Countries like the USA, India, Brazil, Italy, and Spain are the worst affected ones with a huge number of infected cases. Covid-19 pandemic is being witnessed as the toughest time of the century during April-May 2021 due to its 2nd wave which has already entered into stage3/stage4 (i.e., community spread) of Covid-19 infection spread. As per the Covid-19 data released by WHO[1] on 23rd April 2021, the total number of confirmed cases has reached 143,445,675 with 3,051,736 deaths globally as of 22nd April 2021. Thus, the pandemic has led to a huge burden on the healthcare systems across the world.

As suggested by various professionals and healthcare agencies, the most efficient way to tackle this problem is to adapt mass testing, contact tracing, and isolation. Testing for Covid-19 is the most important part of the process and must be scaled as much as possible. The current RT-PCR testing [5] is reliable, but highly time-consuming, needs a lot of infrastructure, very expensive, and limitations of the testing kits. This is a major obstacle to effective containment of the virus leading to low testing rates in many countries. Thus, in the present situation, there is an urgent need to discover other ways of Covid-19 testing which might be more efficient, fast, and scalable.

It has been observed so far that this virus attacks the human respiratory system including the lungs which is responsible for taking in oxygen and expelling out carbon dioxide. Thus, leading to breathing difficulty in the patients[2]. The radiograph image of the lungs can be captured using different imaging tools such as CT-Scan and X-Ray. Getting the CT-Scan is again a costly and time-consuming process.

---

[1]https://covid19.who.int/
[2]https://www.who.int/health-topics/coronavirus

Moreover, only major hospitals have CT scanners. However, capturing an X-Ray is a very affordable as well as an efficient process. The covid-19 disease is caused by severe acute respiratory syndrome coronavirus-2 (SARS-COV-2) and the infected patients show distinct visual features in the Chest X-Ray (CXR) images. Hence, artificial intelligence based automated techniques can be utilized to detect the infection in CXR images. Such testing methods can be very fast, highly scalable, economical, and affordable within the existing infrastructure.

Researchers have tried to explore the artificial intelligence based deep learning techniques for Covid-19 detection, such as COVID-Net [37], CovidAID [23], COVID-CAPS [1], CovXNet [22], DarkCovidNet [27] and Convolutional Neural Network (CNN) Ensemble [6]. The lack of sufficient data to train and test the models is the main problem in the development of the deep learning based models for Covid-19 detection in CXR images. Hence, it is an urgent requirement to develop a deep learning model that would learn distinctive features from the limited data. In order to learn the discriminative and localized features, we propose an attention guided contrastive CNN (AC-CovidNet) for Covid-19 recognition from CXR images. Following are the commitments of this work:

- We propose a novel AC-CovidNet deep learning framework for Covid-19 recognition in CXR images.

- The use of the attention module enforces the learning of localized visual features corresponding to Covid-19 symptoms.

- The contrastive loss increases the discriminative ability and robustness of the model by learning the similarity between Covid-19 infected samples and dissimilarity between Covid-19 positive and negative samples.

- The proposed AC-CovidNet model is able to learn the task specific important features from the limited training data.

- The impact of the proposed method is analyzed for different amount of training data w.r.t. the recent state-of-the-art models.

The remaining paper is organized as follows: Section 2 summarizes the related works; Section 3 illustrates the proposed AC-CovidNet model; Section 4 details the experimental settings; Section 5 presents the results and analysis; and finally Section 6 summarizes the findings with concluding remarks.

2. Related Works

It has been observed in the primary research conducted by Wang and Wong (2020) [37] that chest radiograph images can be used for the Covid-19 detection. It opened up the new urgent and demanding area of the possible usage of Artificial Intelligence for early, efficient and large-scale detection of viruses among people. Some companies around the world working in AI area have started working over AI based solutions for the Covid-19 detection using CT scans and X-Ray radiograph images[3]. They have started harnessing the power of deep learning to learn the infection pattern. An update over the essentials of Radiologists on Covid-19 is published by Kanne et al. [16]. Ng et al. released the imaging profile of the Covid-19 infection with radiologic findings [25]. Li et al. discovered the spectrum of CT findings and temporal progression of Covid-19 disease which reveals that this problem can be solved using imaging AI based tools [21]. Bai et al. performed a performance study of radiologists which can differentiate the Covid-19 from viral pneumonia on chest CT [3]. Deep CNN model is also employed by Narin et al. [24] to perform the automatic detection of Covid-19 from X-Ray images. Residual network based deep learning model [12], truncated inception network [7], and deep learning based anomaly detection [38] are also used for the screening of Covid-19 from radiographs. The uncertainty and interpretability are estimated by Ghosal and Tucker [14] in the deep learning models for Covid-19 detection.

A tremendous growth is observed in deep learning based Covid-19 detection from CXR images. In one of the first attempts, Chest X-Ray (CXR) radiograph images are used for Covid-19 detection using a Deep Learning based convolutional neural network (CNN) model COVID-Net [37]. A projection expansion projection extension module is used heavily in COVID-Net which is experimented on various configurations of the model. Authors used a human-machine collaborative design strategy to create COVID-Net architecture where human driven prototyping and machine based exploration is combined. A COVIDx dataset is also accumulated from various sources and being updated with new data. The dataset and models are publically released for further research[4]. Using COVID-Net model, 96% sensitivity is observed in [37] on a test set of 100 CXR images. A CovidAID model is proposed in [23] which is a pretrained CheXNet [30] - a 121 layer DenseNet [15] followed by a fully connected layer. Using the CovidAID model, 100% sensitivity is observed on a test set having 30 Covid-19 images. A capsule networks based deep learning model (COVID-CAPS) is investigated in [1]. Authors in [1] aim to prevent the loss of spatial information which is observed in CNN based methods. Using COVID-CAPS model, a sensitivity of 90% is reported on 100 test CXR images. CovXNet model [22] is proposed to use transferable multi-receptive feature optimisation. Basically, 4 different config-

---

urations of a network are utilized in CovXNet for training and prediction. Using CovXNet model, 91% sensitivity is reported on 100 test CXR images. The DarkNet architecture based DarkCovidNet model is introduced in [27] with You Only Look Once (YOLO) real time object detection system for Covid-19 detection. Using the DarkCovidNet model, a classification accuracy of 98.08% for binary classes and 87.02% for multi-class cases are observed. CNN ensemble of DenseNet201, Resnet50-v2 and Inception-v3 is utilized for Covid-19 recognition in [6]. Samples from only 2 classes (i.e., covid and non-covid) are used to train the CNN ensemble model. Using CNN ensemble model, a classification accuracy of 91.62% is reported.

From the above presented works, it is convincing that AI powered deep learning methods can play a vital role for the Covid-19 screening. The CT scans and CXR radiographs are used majorly for the imaging based techniques. Less attention has been given to CXR images so far due to the not so great generalization performance caused by the limited availability of data. Given the need to conduct the mass screening at affordable cost, the further fast research over CXR images of lungs is very much needed using the limited data. Thus, in this paper, we utilize the capability of attention mechanism and contrastive learning to tackle the learning with limited data for Covid-19 recognition from CXR images.

3. Proposed AC-CovidNet CNN Model

In this section, first we provide a brief of deep learning, attention mechanism and contrastive learning. Then we present the proposed AC-CovidNet architecture along with objective function.

3.1. Background

Deep learning has shown a great impact from last decade to solve many challenging problems [20]. Deep learning models consist of the deep neural networks which learn the important features from the data automatically. The training of the deep models is generally performed using stochastic gradient descent optimization [18], [10]. Convolutional neural network (CNN) based models have been used to deal with the image data, such as image classification [19], face recognition [33], image retrieval [9], hyperspectral image analysis [32], and biomedical image analysis [11].

Attention mechanism in deep learning facilitates to learn the localized features which is more important in the context of the problem of Covid-19 recognition from CXR images [2]. It is also discovered that the attention based model can outperform the plain neural network models [36]. The attention mechanism has been also utilized for different applications such as facial micro-expression recognition [13], breast tumor segmentation [35], and face recognition [31]. Thus, motivated from the success of attention mecha-
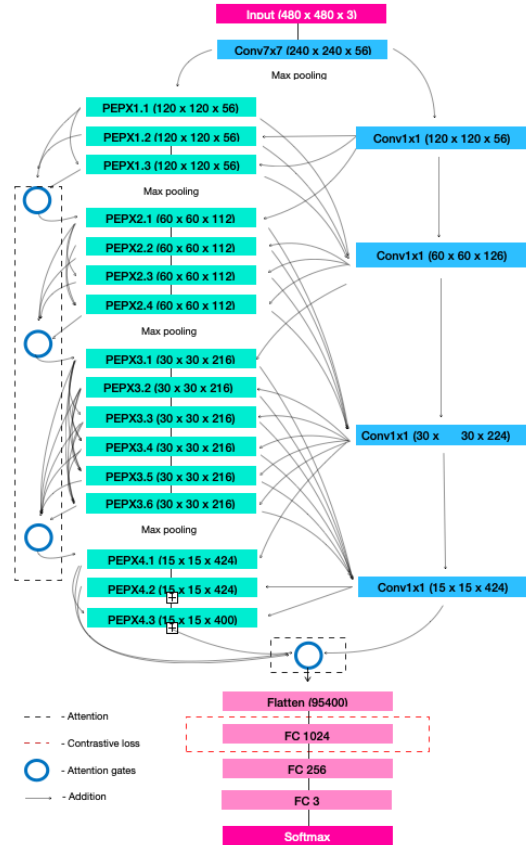

Figure 1. Architecture of the proposed AC-CovidNet

nisms for different applications, we utilize it in the proposed model for Covid-19 recognition.

Contrastive learning is the recent trend to learn the similarity and dissimilarity between the similar and dissimilar samples in the abstract feature space for visual representations [4], [17]. Generally, contrastive learning is dependent upon the feature similarity between positive pairs and negative pairs [34]. The contrastive learning has shown very promising performance for different problems, such as face generation [8], image-to-image translation [28], medical visual representations [39], and video representation [29]. Thus, motivated from the discriminative and robust feature representation by contrastive learning, we utilize it in the proposed method to learn the weights of the network.

3.2. Proposed Model

In this paper, we propose an attention guided contrastive CNN for Covid-19 recognition, named as AC-CovidNet. The architecture of the proposed AC-CovidNet model is illustrated in Fig. 1. The proposed model is based on the popular COVID-Net model [37]. It heavily uses light weight residual projection expansion projection extension (PEPX) mechanism. The PEPX component is shown in Fig. 2. This
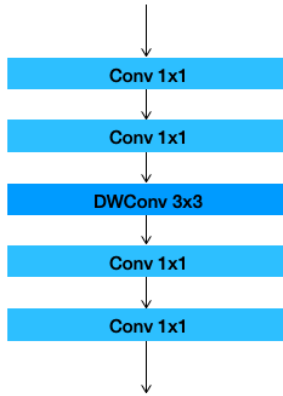
Figure 2. Projection Expansion Projection Extension (PEPX) Module. Here, DWConv represents the depth wise convolution operation.

architecture also uses selective long range connectivity in model which improves the representational capacity. It also facilitates the training of the model easier. However, the extensive use of these long range connections may bring a lot of redundant low level features. In order to resolve this issue, the proposed model uses an attention mechanism. Attention helps the model to prioritize the regions of important w.r.t. the problem being solved. Attention is also useful to suppress the activations from the redundant features from the initial layers and helps to focus on the important features that are required to solve the given problem. We use the attention gates in the proposed architecture as suggested in [26], at various layers in the COVID-Net architecture where many long range connections are used. This improves the sensitivity as the model attends better to the important visual features of infected regions in CXR images due to Covid-19. As the difference between Covid-19 and Pneumonia features is very subtle, we propose to use the supervised contrastive loss. The contrastive loss facilitates the network to increase the distance between the learnt representation of the classes as much as possible.

3.2.1 Architecture

The proposed AC-CovidNet architecture is an extension of COVID-Net by utilizing the attention gates where multiple long range connections converge. The details of the architecture of the proposed model is illustrated Fig. 1. We use PEPX layers and attention gates heavily as a part of our architecture. We also use the contrastive loss function while training. The model is first trained using supervised contrastive learning before final fine tuning using supervised learning.

3.2.2 PEPX Layer

The architecture of a Projection-Expansion-Projection-Extension (PEPX) is shown in Fig. 2. The idea of this module is to project features into a lower dimension using the first two conv1x1 layers, then expand those features using a depthwise convolution layer (DWConv3x3) and project into lower dimension again using two conv1x1 layers. Thus, the PEPX layer leads to the efficient model by reducing the number of parameters and operations.

3.2.3 Attention Gate

We use attention gates in the proposed AC-CovidNet model at various layers as depicted in Fig. 1. The block diagram of the attention module is shown in Fig. 3. Features from multiple layers are passed through conv1x1x1 and are added together. Then, the aggregated features are passed through Relu activation function followed by conv1x1x1 and then sigmoid activation function. The feature map output of the sigmoid layer is then passed through a resampler. The output of the resampler is added to the features from the nearest input layer to the attention module in order to produce the output of the attention gate.

3.2.4 Network Details

The following are the main components of the proposed AC-CovidNet model. a) Encoder network, $E(.)$: It is the AC-CovidNet model as discussed above after removing the last three layers which form the classifier network. The encoder network maps the input image into a vector of size 1024. b) Projection network, $P(.)$: This network projects the output of the encoder network into a vector of size 128. This is a multilayer perceptron with input size of 1024, one hidden layer of length 512, and an output vector of size 128. This network is used while training the encoder and later discarded. c) Classifier network, $C(.)$: This is the last three layers of AC-CovidNet that take a 1024 length output of the encoder as the input and produces an output of size $c = 3$ (corresponding to the classes, Normal, Pneumonia, and Covid-19).

First we train the encoder network $E(.)$ with supervised contrastive loss (SupConLoss) using the projection network $P(.)$. Then we discard the projection network and add the classifier network $C(.)$ to the encoder network. Then we freeze the weights of the encoder model and train the classifier with categorical cross entropy loss. The training of the encoder of the proposed model is summarized in Algorithm 1.

3.3. Objective Function

We use the supervised contrastive learning method [17] to train the encoder network for feature extraction. We train
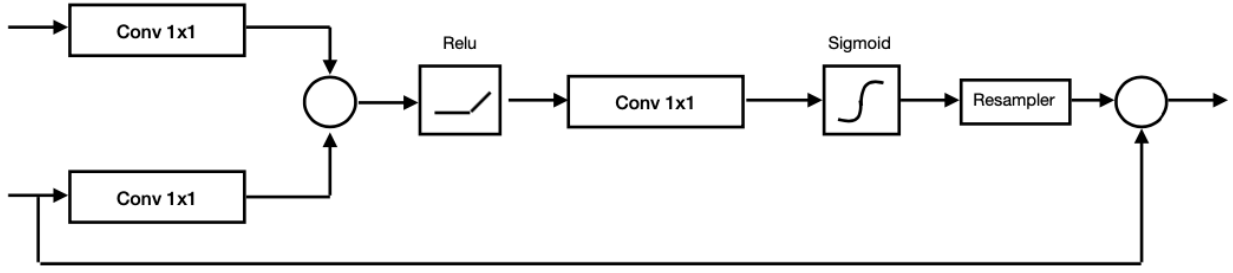
Figure 3. The block diagram of an attention gate.

Algorithm 1: Training AC-CovidNet

E: Encoder Network, P: Projection Network
C: Classifier Network
$SupCon$: Supervised Contrastive Loss
$CrossEntropy$: Cross Entropy Loss
while $epochs --$ **do**
    // Stage 1
    for $batch\ X\ in\ Data$ **do**
        Initialize Z as null;
        for $x_i\ in\ X$ **do**
            $h_i = E(x_i)$;
            $z_i = P(h_i)$;
            $Z.append(z_i)$;
        end
        $\mathcal{L}_{SC} = SupCon(Z_i)$;
        Update E and P to minimise $\mathcal{L}_{SC}$;
    end
end
Discard $P(.)$ and freeze weights of $E(.)$;
Final model, $M(\boldsymbol{x}) = C(E(\boldsymbol{x}))$;
while $epochs --$ **do**
    // Stage 2
    for $batch\ X\ in\ Data$ **do**
        $\hat{Y} = M(X)$;
        $\mathcal{L}_{CE} = CrossEntropy(Y, \hat{Y})$;
        Update M to minimise $\mathcal{L}_{CE}$;
    end
end

---

the classifier network using the cross entropy loss function after freezing the encoder network.

3.3.1 Supervised Contrastive loss

Contrastive loss is most commonly used in unsupervised and self-supervised learning. In order to adapt this method to supervised learning and take advantage of the labels available, the supervised contrastive learning has been investigated in [17]. This loss is used to train the encoder network of the proposed AC-CovidNet.

Consider $\boldsymbol{z}_\ell = \mathrm{P}\left(\mathrm{E}\left(\boldsymbol{x}_\ell\right)\right)$ where $\boldsymbol{x}$ is the input and $A(i)$ is a set of all indices (I) except i. Then, the supervised contrastive loss is given as,

$$\mathcal{L}_{SC} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_p / \tau\right)}{\sum_{a \in A(i)} \exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_a / \tau\right)}$$

where, $\tau \in \mathcal{R}^+$ is scalar temperature parameter, $P(i) \equiv \left\{p \in A(i) : \tilde{\boldsymbol{y}}_p = \tilde{\boldsymbol{y}}_i\right\}$ is the set of indices of all the positive samples other than i, and $|P(i)|$ is its cardinality. Note that $\tilde{\boldsymbol{y}}$ represents the class label.

3.3.2 Cross Entropy Loss

Cross entropy loss is used to train the classifier which takes input from the feature extractor. It produces the output probability corresponding to three class from the softmax activation function. The cross entropy loss is given as,

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{3} y_i^j \cdot \log \hat{y}_i^j$$

where $\hat{y}_i^j$ and y_i^j are the output probability and ground truth value, respectively, for j^{th} class corresponding to i^{th} sample and N represents the number of samples in a batch.

4. Experimental Setup

4.1. Dataset Used

We use COVIDx dataset [37] in this work to experiment with the proposed AC-CovidNet model. It is the largest

Table 1. Datasets configurations used for training and testing w.r.t. the number of samples in Covid-19, Pneumonia and Normal categories.

S. No	Configuration	Covid-19		Pneumonia		Normal	
		#Training Images	#Test Images	#Training Images	#Test Images	#Training Images	#Test Images
I	COVIDx-v1	517	100	7966	100	5475	100
II	COVIDx-v2	467	150	7916	150	5425	150
III	COVIDx-v3	417	200	7866	200	5375	200

Table 2. The results comparison in terms of the sensitivity.

Model	COVIDx-v1			COVIDx-v2			COVIDx-v3		
	Covid	Pneumonia	Normal	Covid	Pneumonia	Normal	Covid	Pneumonia	Normal
COVID-Net [37]	96	89	95	94.66	87.33	95.33	94	88.5	94
CovidAID [23]	93	95	73	91	92.66	70.66	88	92.5	74
Ensemble Learning CNNs [6]	89	91	87	87.33	92	86	83	91	89
DarkCovidNet [27]	86	88	92	84.66	84	89.33	84.5	89.5	90.5
Covid-Caps [1]	90	87	77	88	86	74	86.5	84.5	76
CovXnet [22]	91	71	86	89.33	69.33	86.66	88	70	84
CovidNet + attention	95	86	95	95.33	82.66	95.33	94	87	94.5
CovidNet + contrastive loss	96	89	95	96	87.33	94	95	90	95
AC-CovidNet	**96**	**88**	**95**	**96.66**	**87.33**	**95.33**	**96.5**	**87.5**	**95**

open-source dataset for Covid-19 Chest X-Ray (CXR) images currently available. This dataset is a combination of several other open datasets, including Covid-19 Image Data Collection, Covid-19 Chest X-ray Dataset Initiative, ActualMed Covid-19 Chest X-ray Dataset Initiative, RSNA Pneumonia Detection Challenge dataset, and Covid-19 radiography database. This dataset contains three classes of CXR images, i.e., Covid-19, Pneumonia and Normal. It has a total of 14,258 images from 14,042 patients. It has 7,966 Normal images, 5,475 Pneumonia images and 517 Covid-19 images in the training set and 100 images from each category in the test set. In our experiments we create three different sets of data by varying the distributions of samples in training and test sets. We name the original COVIDx dataset as COVIDx-v1 by following the above said distribution. COVIDx-v2 is the COVIDx dataset with 150 test images from each category and 7,916 Normal images, 5,425 Pneumonia images and 467 Covid-19 images in the training set. Similarly, COVIDx-v3 has 200 test images from each category and 7,866 Normal images, 5,375 Pneumonia images and 417 Covid-19 images in the training set. These configurations are summarized in Table 1 w.r.t. the number of samples in Covid-19, Pneumonia and Normal categories.

4.2. Training Settings

The proposed AC-COVIDNet model is pretrained on imagenet as suggested by [37]. Then the entire model is trained in two stages. In the first stage, the encoder network (i.e., feature extractor) is trained using contrastive loss function as suggested in [17] for feature extraction, so that the distance between the learnt features is optimized. In the second stage, the feature extractor is frozen and trained by adding a classifier on the top with cross entropy loss function. We train the proposed model as well as the state-of-

the-art models on all three variations of COVIDx dataset, i.e., with 100, 150 and 200 test images of different categories. The models are trained using Adam optimiser [18]. The learning rate is set as 1.7e-4. The batch size of 64 is used. We use Relu activation function in every layer of the network and softmax in the last layer. Max-pooling is used after every batch of PEPX layers. The three versions of the COVIDx dataset are used for training and the sensitivity of the classes are compared. Covid-19 sensitivity is the percentage of instances with Covid-19 that are correctly identified. The model is trained using the computational resources provided by Google Colab. Keras deep learning library is used with tensorflow as a backend.

5. Experimental Results and Analysis

We test the proposed AC-CovidNet model on all three configurations of the COVIDx dataset and calculate the sensitivity for Covid-19 and other classes. In order to demonstrate the superiority of the proposed method, we also compute the results using state-of-the-art deep learning based Covid-19 recognition models, such as CovXNet [22], COVID-CAPS [1], CNN Ensemble [6], DarkCovidNet [27], COVID-Net [37], and CovidAID [23]. The results in terms of the sensitivity for the Covid-19 class are reported in Table 2. On configuration I (i.e., COVIDx-v1 dataset with 100 test images), configuration II (i.e., COVIDx-v2 dataset with 150 test images), and configuration III (i.e., COVIDx-v3 dataset with 200 test images), the observed Covid-19 sensitivity using the proposed AC-CovidNet model is 96%, 96.66%, and 96.5%, respectively. It can be observed in Table 2 that the proposed model outperforms the remaining models over all three settings of the COVIDx dataset.

Note that the proposed model is able to achieve better results than the other compared models because the proposed model learns the Covid-19 specific features using the attention module and increases the separation between different classes in feature space using the contrastive loss. In the configuration I of the dataset (with 100 test images), the performance of the proposed model is better than Cov-XNet, COVID-CAPS, CNN Ensemble, DarkCovidNet and CovidAID models and same as COVID-Net. Thus, in order to demonstrate the advantage of the proposed model, we compare the results with less number of training samples and more test samples. Basically, it depicts the generalization capability of the proposed model. On an expectation, a bigger test set can reflect the better generalization of the deep learning models. Thus, we experiment with configuration II having 467 training samples and 150 test samples of Covid-19 category. It can be seen in Table 2 that the proposed model is able to retain the similar performance by correctly classifying 145 Covid-19 images out of 150 in the dataset. However, the performance of other models dropped significantly. We also test the performance by further reducing the number of training samples and increasing the number of test samples in configuration III with 417 training images and 200 test images from Covid-19 category. It can be noticed that the performance of the proposed AC-CovidNet model is still similar. However, the other models drastically fail to generalize in case of limited training set. Thus, it clearly indicates that the proposed AC-CovidNet model is able to capture the robust and discriminative features pertaining to the Covid-19 infection and generalize well even with the limited training data. It also shows the positive impact of attention modules and contrastive learning for the Covid-19 recognition from CXR images.

The impact of the attention and contrastive mechanism is also investigated by considering the only attention and only contrastive mechanism with base network CovidNet (i.e., CovidNet + attention and CovidNet + contrastive loss, respectively). As shown in the results using these methods in Table 5, the performance of AC-CovidNet, which uses both attention and contrastive mechanisms, is improved than the models which use only attention and only contrastive mechanisms for Covid-19 recognition. We also report the results for other two classes in Table 5, i.e., Pneumonia and Normal. It is observed that the performance of the proposed model is comparable to the state-of-the-art for Pneumonia and Normal classes.

6. Conclusion

In this paper, an AC-CovidNet model is proposed for Covid-19 recognition from chest X-Ray images. The proposed model utilizes the attention module in order to learn the task specific features by better attending the infected regions in the images. The proposed model also utilizes contrastive learning in order to achieve the better separation in the feature space by increasing the discriminative ability and increasing robustness. The results are computed over three different configurations of Covid-19 dataset with varying number of training and test samples. The results are also compared with six recent state-of-the-art deep learning models. It is noticed that the proposed AC-CovidNet model outperforms the existing models in terms of the sensitivity for the Covid-19 category. Moreover, it is also observed that the performance of the proposed model is consistent with a limited training set. Whereas, the existing methods fail to do so. It shows the better generalization capability of the proposed method. The future work includes the utilization of recent development in deep learning to solve the Covid-19 recognition problem from chest X-Ray images with better performance.

References

[1] Parnian Afshar, Shahin Heidarian, Farnoosh Naderkhani, Anastasia Oikonomou, Konstantinos N Plataniotis, and Arash Mohammadi. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. *Pattern Recognition Letters*, 138:638–643, 2020. 2, 6

[2] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015*, 2015. 3

[3] Harrison X Bai, Ben Hsieh, Zeng Xiong, Kasey Halsey, Ji Whae Choi, Thi My Linh Tran, Ian Pan, Lin-Bo Shi, Dong-Cui Wang, Ji Mei, et al. Performance of radiologists in differentiating covid-19 from non-covid-19 viral pneumonia at chest ct. *Radiology*, 296(2):E46–E54, 2020. 2

[4] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 3

[5] Victor M Corman, Olfert Landt, Marco Kaiser, Richard Molenkamp, Adam Meijer, Daniel KW Chu, Tobias Bleicker, Sebastian Brünink, Julia Schneider, Marie Luisa Schmidt, et al. Detection of 2019 novel coronavirus (2019-ncov) by real-time rt-pcr. *Eurosurveillance*, 25(3):2000045, 2020. 1

[6] Amit Kumar Das, Sayantani Ghosh, Samiruddin Thunder, Rohit Dutta, Sachin Agarwal, and Amlan Chakrabarti. Automatic covid-19 detection from x-ray images using ensemble learning with convolutional neural network. *Pattern Analysis and Applications*, pages 1–14, 2021. 2, 3, 6

[7] Dipayan Das, KC Santosh, and Umapada Pal. Truncated inception net: Covid-19 outbreak screening using chest x-rays. *Physical and engineering sciences in medicine*, 43(3):915–925, 2020. 2

[8] Yu Deng, Jiaolong Yang, Dong Chen, Fang Wen, and Xin Tong. Disentangled and controllable face image generation via 3d imitative-contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5154–5163, 2020. 3

[9] Shiv Ram Dubey. A decade survey of content based image retrieval using deep learning. *arXiv preprint arXiv:2012.00641*, 2020. 3

[10] Shiv Ram Dubey, Soumendu Chakraborty, Swalpa Kumar Roy, Snehasis Mukherjee, Satish Kumar Singh, and Bidyut Baran Chaudhuri. diffgrad: an optimization method for convolutional neural networks. *IEEE transactions on neural networks and learning systems*, 31(11):4500–4511, 2019. 3

[11] Shiv Ram Dubey, Swalpa Kumar Roy, Soumendu Chakraborty, Snehasis Mukherjee, and Bidyut Baran Chaudhuri. Local bit-plane decoded convolutional neural network features for biomedical image retrieval. *Neural Computing and Applications*, pages 1–13, 2019. 3

[12] Muhammad Farooq and Abdul Hafeez. Covid-resnet: A deep learning framework for screening of covid19 from radiographs. *arXiv preprint arXiv:2003.14395*, 2020. 2

[13] Viswanatha Reddy Gajjala, Sai Prasanna Teja Reddy, Snehasis Mukherjee, and Shiv Ram Dubey. Meranet: Facial micro-expression recognition using 3d residual attention network. *arXiv preprint arXiv:2012.04581*, 2020. 3

[14] Biraja Ghoshal and Allan Tucker. Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection. *arXiv preprint arXiv:2003.10769*, 2020. 2

[15] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018. 2

[16] Jeffrey P Kanne, Brent P Little, Jonathan H Chung, Brett M Elicker, and Loren H Ketai. Essentials for radiologists on covid-19: an update—radiology scientific expert panel, 2020. 2

[17] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33, 2020. 3, 4, 5, 6

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 3, 6

[19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012. 3

[20] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 3

[21] Mingzhi Li, Pinggui Lei, Bingliang Zeng, Zongliang Li, Peng Yu, Bing Fan, Chuanhong Wang, Zicong Li, Jian Zhou, Shaobo Hu, et al. Coronavirus disease (covid-19): spectrum of ct findings and temporal progression of the disease. *Academic radiology*, 27(5):603–608, 2020. 2

[22] Tanvir Mahmud, Md Awsafur Rahman, and Shaikh Anowarul Fattah. Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization. *Computers in biology and medicine*, 122:103869, 2020. 2, 6

[23] Arpan Mangal, Surya Kalia, Harish Rajgopal, Krithika Rangarajan, Vinay Namboodiri, Subhashis Banerjee, and Chetan Arora. Covidaid: Covid-19 detection using chest x-ray. *arXiv 2004.09803*, 2020. 2, 6

[24] Ali Narin, Ceren Kaya, and Ziynet Pamuk. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *arXiv preprint arXiv:2003.10849*, 2020. 2

[25] Ming-Yen Ng, Elaine YP Lee, Jin Yang, Fangfang Yang, Xia Li, Hongxia Wang, Macy Mei-sze Lui, Christine Shing-Yen Lo, Barry Leung, Pek-Lan Khong, et al. Imaging profile of the covid-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*, 2(1):e200034, 2020. 2

[26] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. 4

[27] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U Rajendra Acharya. Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in biology and medicine*, 121:103792, 2020. 2, 3, 6

[28] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European Conference on Computer Vision*, pages 319–345. Springer, 2020. 3

[29] Rui Qian, Tianjian Meng, Boqing Gong, Ming-Hsuan Yang, Huisheng Wang, Serge Belongie, and Yin Cui. Spatiotemporal contrastive video representation learning. *arXiv preprint arXiv:2008.03800*, 2020. 3

[30] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, 2017. 2

[31] Yongming Rao, Jiwen Lu, and Jie Zhou. Attention-aware deep reinforcement learning for video face recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 3931–3940, 2017. 3

[32] Swalpa Kumar Roy, Gopal Krishna, Shiv Ram Dubey, and Bidyut B Chaudhuri. Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(2):277–281, 2019. 3

[33] Yash Srivastava, Vaishnav Murali, and Shiv Ram Dubey. Hard-mining loss based convolutional neural network for face recognition. In *Fifth IAPR International Conference on Computer Vision and Image Processing (CVIP)*, 2020. 3

[34] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. What makes for good views for contrastive learning. *arXiv preprint arXiv:2005.10243*, 2020. 3

[35] Aleksandar Vakanski, Min Xian, and Phoebe E Freer. Attention-enriched deep learning model for breast tumor segmentation in ultrasound images. *Ultrasound in Medicine & Biology*, 46(10):2819–2833, 2020. 3

[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017. 3

[37] Linda Wang, Zhong Qiu Lin, and Alexander Wong. Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10(1):19549, Nov 2020. 2, 3, 5, 6

[38] Jianpeng Zhang, Yutong Xie, Yi Li, Chunhua Shen, and Yong Xia. Covid-19 screening on chest x-ray images using deep learning based anomaly detection. *arXiv preprint arXiv:2003.12338*, 27, 2020. 2

[39] Yuhao Zhang, Hang Jiang, Yasuhide Miura, Christopher D Manning, and Curtis P Langlotz. Contrastive learning of medical visual representations from paired images and text. *arXiv preprint arXiv:2010.00747*, 2020. 3