X-Rays and CT-Scans based COVID-19 Detection using Multi-Modal Multi-Task Learning Framework

Nikhilanand Arya[1,1,*], Kwanit Gupta[1,2], Sriparna Saha[1,3], Sandeep Bansal[1,4]

Abstract

The world is still fighting the worldwide epidemic (COVID-19) declared by World Health Organization (WHO) in march 2020. We face challenges in the identification and diagnosis of infected patients. Chest X-Rays and CT-Scans are the two major screening techniques, which render significant roles in COVID-19 detection and diagnosis. In recent researches, several deep learning-based uni-modal architectures are developed for the classification of COVID-19, which either utilize features of X-Rays or CT-Scans but not both together. As both these techniques capture imaging of chests, doctors recommend patients to go through any one of them during early stage of diagnosis. In this work, we propose the multi-modal multi-task learning framework for the classification of patients as COVID-19 and non-COVID-19 using either X-Rays or CT-Scans along with complementary information extracted from both of them. This is a three-stage architecture, we first used the various combinations of popular transfer learning methods for task-specific embedding generation using chest screenings, and then we further created a shared embedding having common information from X-Rays and CT-Scans. We have clubbed the shared embedding with the task-specific embedding to have the final classification. The architecture proves its efficacy with 98.23% and 98.83% accuracy in COVID-19 detection over X-Rays and CT-scans, respectively. This architecture is novel and claimed the highest classification accuracy when compared with other related researches.

Keywords: COVID-19, Chest X-Rays (CXRs), Computed Tomography (CT), Transfer Learning, Multi-Task Learning Framework, Adversarial Training, Wasserstein Distance.

1. Introduction

COVID-19, a lethal disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been affecting the entire world since December 2019. We have been living in a condition of caution and terror because of its terrible effects, which range from mild self-limiting respiratory tract infection to severe progressive pneumonia, multiorgan failure, and death [1, 2, 3]. World health infrastructure is under tremendous stress due to the large volume of COVID-19 infected patients and it is continuously increasing day by day.At the moment, the standard diagnostic approach for detecting viral nucleic acid is real-time reverse transcriptase polymerase chain reaction (RT-PCR). However, many hyperendemic localities or nations are unable to do enough RT-PCR testing for tens of thousands of suspected COVID-19 individuals. Furthermore, this detection procedure might take several hours or even days to complete. Simultaneously, in order to obtain valid test findings, the sample must be examined numerous times over a period of several days. SARS and COVID-19 are both corona-viruses from the same family, and numerous methods [4, 5, 6] have been developed for detecting SARS cases using chest imaging, as well as for detecting pneumonia in general [7]. The detection and diagnosis of COVID-19 patients using artificial intelligence based architectures with chest imaging can be very helpful in reducing the burden of health infrastructures and doctors.

1.1. *Background Literature*

CXRs and CT-Scans are the images, which contain information of any SARS related illness. To address the problem arising due to lack of reagents, several techniques [8, 9, 10] have been developed in the COVID-19 detection which are based on computed tomography (CT) images. For example, *Fang et. al.* [9] experimented with very small set of chest CT images having 51 patients in COVID-19 testing and attained a high sensitivity of 98%. Further,

*Corresponding author

Email addresses: nikhilaryan92@gmail.com (Nikhilanand Arya), gupta.45@iitj.ac.in (Kwanit Gupta), sriparna.saha@gmail.com (Sriparna Saha), skbansal@digitalindia.gov.in (Sandeep Bansal)

[1]Nikhilanand Arya is pursuing PhD in Computer Science Engineering, Indian Institute of Technology Patna, India.

[2]Kwanit Gupta is a B.Tech. student at Department of Electrical Engineering, Indian Institute of Technology Jodhpur, India.

[3]Dr. Sriparna Saha is an Associate Professor in the Department of Computer Science and Engineering, Indian Institute of Technology Patna, India.

[4]Sandeep Bansal is working as principal research scientist , PhD Cell In-Charge at Digital India Corporation (formerly Media Lab Asia) (A Research & Development company of Ministry of Electronics & Information Technology, Govt. of India).

Gozes et. al. [10] experimented with the role of deep learning techniques in COVID-19 detection task having CT images as source of information. At one hand CT-images were very helpful in COVID-19 detection task, it is a time consuming process when compared with X-ray imaging for the same task. We can not rely only on CT-images as COVID-19 detection system because under-developed regions of several countries lacks in the quality and quantity of CT scanners, thereby leading to an inappropriate detection of COVID-19. Hence, several efforts have been made utilizing deep learning-based approaches to classify the chest X-ray images into COVID-19 patient class or normal class. As, researchers are dealing with X-ray images, they proposed their architectures with CNNs as the baseline framework. For instance, *Apostolopoulos et. al.* designed the MobileNet [11] architecture for COVID-19 diagnosis using chest X-ray images. Their architecture claimed the accuracy of 96.78% in COVID-19 diagnosis. Similarly, *Narin et. al.* [12] also proposed the CNN based transfer learning approach. The reported accuracy values are 97% for InceptionV3 and 87% for Inception-ResNetV2, respectively. *El-Din Hemdan et al.* [13] conducted a comparative investigation of many classic deep learning classification frameworks, and pre-trained the model using the ImageNet dataset [14] to discriminate between normal and COVID-19 classes. They used a limited dataset with just 50 radiographs for their investigation, 25 of which were from healthy patients and 25 from COVID-19 positive patients. VGG19 and DenseNet performed similarly in the author's model, with F1-Scores of 0.89 and 0.91 for normal and COVID-19, respectively. The early studies for COVID-19 detection and diagnosis were limited due to lack of large and publicly available dataset. In recent researches, *Wang et al.* [15] developed a new model architecture, COVID-net. In their study, a larger dataset (COVIDx) consisting of 13,800 chest X-ray images of normal, pneumonia and COVID-19 patients is established and the COVID-net is trained and tested on COVIDx. The classification task of COVID-net is related to identifying X-ray images as normal, pneumonia and COVID-19. Motivated with the application of transfer learning architectures, *Farooq et al.* [16] fine-tuned the ResNet-50 architecture for the multi-class classification task of identifying chest X-rays into normal, COVID-19, bacterial pneumonia and viral pneumonia. The studies shows that the chest imaging are important for COVID-19 detection task while ignoring the clinical symptoms. *Yazeed et al.* [17] established the usefulness eight binary features: sex, age ≥ 60 years, known contact with an infected individual, and the appearance of five initial clinical symptoms in COVID-19 detection and diagnosis. They proposed a machine-learning approach using above mentioned features from 51,831 tested individuals (of whom 4769 were confirmed to have COVID-19).

1.2. Motivation for the Proposed Architecture

All the existing works related to COVID-19 detection utilizing deep learning models have used single modality such as X-Rays, CT-scans or clinical symptoms. The unimodal architectures lack in combining important features related to COVID-19 detection task from different modalities. X-rays and CT-scans both capture images of chest. Fusion of information collected from different modalities can help in better prediction of the disease. This has motivated us to design the multi-modal framework for the COVID-19 classification. The multi-modal framework works only when we have all the modalities present for each and every sample. This is not the case in our dataset, as the datasets of X-Rays and CT-scans are taken from different sources. This limitation has further motivated us to utilize the multi-task learning framework. So, our final proposed architecture is the combination of multi-modal multi-task framework. The main contributions of this work are as follows:

- Firstly, transfer learning is adopted to find the best possible features of X-Rays and CT-scans. For this task we tested with twenty six different possible transfer learning architectures and selected top six best embedding generation architectures (presented in Table 3 and 4) based on AUC values.

- Secondly, we designed task specific feature extraction modules based on top six transfer learning architectures and shared feature extraction modules from the thirty six different combinations of task specific feature extraction modules and selected top six combinations (as presented in Table 8) based on the wasserstein distance.

- Finally, we used the combinations of task specific features and shared features to have the final classification between COVID-19 and Non-COVID-19. We selected **EfficientNetB0** as X-Ray Task Specific module, **EfficientN -etB1** as CT-Scan Task Specific module, **ResNet50** as X-Ray Shared Feature module and **ResNet50V2** as CT-Scan Shared Feature module followed by some hidden layers in the multi-modal multi-task learning framework for COVID-19 detection using X-Rays and CT-Scans with AUC values of **99.53%** and **99.85%**, respectively.

2. Methodology

In context of the proposed architecture, the suitable github repository (https://github.com/kwanit1142/Respiratory-Scans-based-COVID-19-Detection-using-Multi-Modal-Multi-Task-Learning-Framework) is made, which can be tried with user-customized dataset.

2.1. CXRs and CT-Scans Dataset

The Dataset[5] used in this study, originally consisted of 17099 images, divided into 9544 chest X-Rays [18][19] (https://github.com/ieee8023/covid-chestxray-dataset) and 7555 Computed Tomographic (CT) Scans[20] (https://github.com/UCSD-AI4H/COVID-CT). Furthermore, they were categorized into COVID and Non-COVID sub-classes[21], as seen from Table 1.

Table 1: Class-wise Dataset Distribution

Classes	**Chest X-Rays**	**CT-Scans**
COVID	4044	5427
Non-COVID	5500	2628

Despite the fact that images differed from each other in terms of orientations, shapes and sizes, some of them were found to be corrupted and poor in quality due to limited technical capabilities of scanning instruments, as seen in Fig. 1. So, we discarded those images and used the remaining 12840 images with train-test split of 80:20, having equal amount of chest X-Rays and CT-Scans, with uniform sub-classes variations, as seen from Table 2.

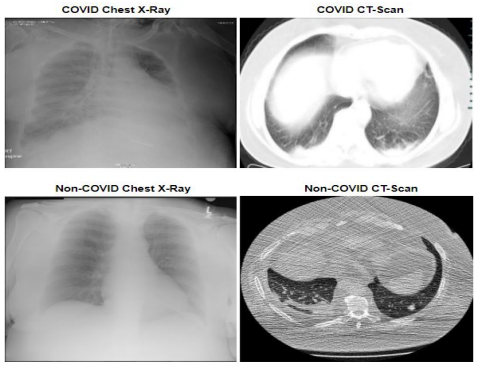

Figure 1: Distorted or downgraded Chest X-Ray and CT-Scan Images.

Table 2: Final Dataset Distribution with Train-test Split

Scan-Type	**Train Split**	**Test Split**
COVID X-Rays	3034	758
COVID CT-Scans	3034	758
Non-COVID X-Rays	2101	527
Non-COVID CT-Scans	2101	527

[5]https://www.kaggle.com/ssarkar445/covid-19-xray-and-ct-scan-image-dataset

2.2. Image Enhancement Techniques

For highlighting useful features and necessary contrast-based variations without loosing structural properties like Rib Cage's Shape, Tissue Patches, etc., we implemented some image-based pre-processing techniques, as following :-

2.2.1. Thresholding Methods

The motivation behind using thresholding methods comes from the fact that binary thresholding variants provide the necessary contrast depth to highlight certain segmented regions like lungs alignment, while adaptive thresholding variants bring out the internal network structures and tissue variations, especially in CT-Scans. In this technique, for each pixel of an image, if the intensity value is smaller than a certain threshold value, it is set to 0 (Black), otherwise to maximum value, i.e., 255 (White). Fig.2 and Fig.3 represent the outcomes of chest X-Ray and CT-Scan, respectively [22][23]. But, the final outcomes have certain irregularities like, some variants didn't show network arrangement, while some made the image loose its basic structural identity and also introduced grainy noises amidst the procedure.

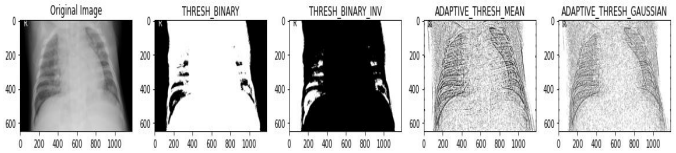

Figure 2: Thresholding Methods on a Chest X-Ray.

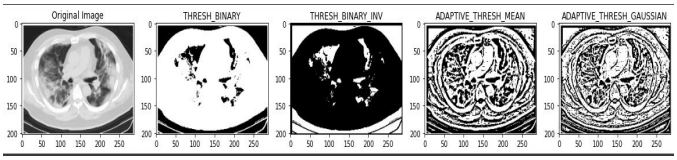

Figure 3: Thresholding Methods on a CT-Scan.

2.2.2. Smoothening Filters

In this method, a kernel (usually an all-ones matrix) is made to convolve through images. It helps in diffusing grainy noises which affect both the scan types in thresholding methods. It also preserves the basic structural identity of images. But, the final outcomes (Fig. 4 and Fig. 5) still had unclear internal network arrangements, without making any significant improvements in image quality.

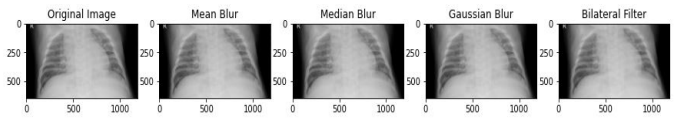

Figure 4: Smoothening filters on Chest X-Ray

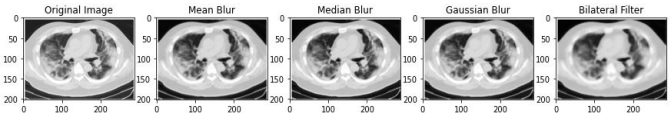

Figure 5: Smoothening filters on CT-Scan

2.2.3. Gradient Tools

This technique deals with 1st and 2nd order matrix-based differentiation, which enhances the boundary outlines of an image. It is done by extracting the gradient magnitudes and allocating them according to direction of neighbouring pixels and current pixel's position. Fig 6 and Fig.7 represent the outcomes[24][25][26] of chest X-Rays and CT-Scans, respectively. From the outcome images, it is evident that chest X-Rays became more deteriorated. Even "canny edge detection" is better than others, in showcasing the internal network details and outer boundary edges, it removed the opaqueness of patches and regions in both the scan types.

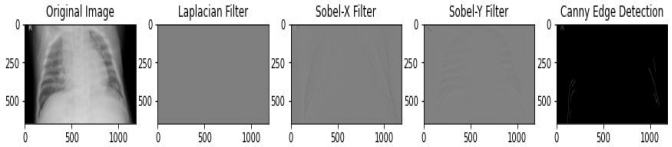

Figure 6: Gradient Tools on Chest X-Ray

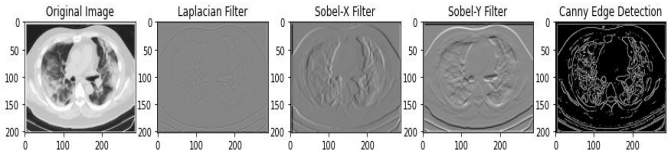

Figure 7: Gradient Tools on CT-Scan

2.2.4. Histogram Equalization

This technique has benefits of enhancing the necessary contrast variations which lead to unique characteristics like tissues distribution, network spread, segmented patches, etc. along with minimized background noises and sharpened boundary outlines at each scale. It works on a continuous distribution graph between intensity and number of pixels, in form of histograms. With the configuration options of **globally** [27] (Adjusting intensity histograms in context of each pixel for whole image) and **locally adaptive** [28] (Adjusting intensity histograms in context of each pixel for a certain image patch/window), it highlights useful contrast-based variations by cutting out the outlier intensities and re-distributing the remaining histograms into pre-defined range. Fig.8 and Fig.9 represent the outcomes of chest X-Rays and CT-Scans, respectively. We did not observe any drastic change in chest X-Rays, when compared with that of CT-Scans.

2.2.5. Morphological Operations[29]

This method provides the enhanced internal network regions and boundary outlines, especially in CT-Scans and minimises various noises with retained opaqueness in spatially-different patches.

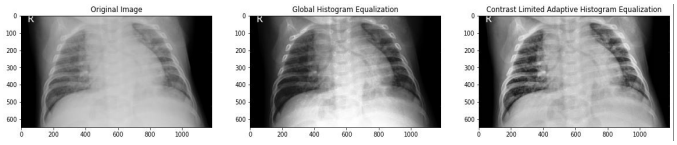

Figure 8: Histogram Equalizations on Chest X-Ray

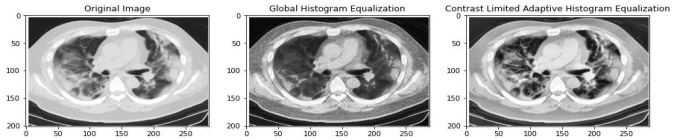

Figure 9: Histogram Equalizations on CT-Scan

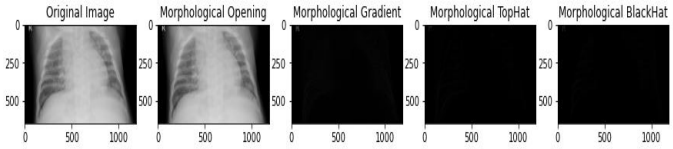

Figure 10: Morphological Operations on Chest X-Ray

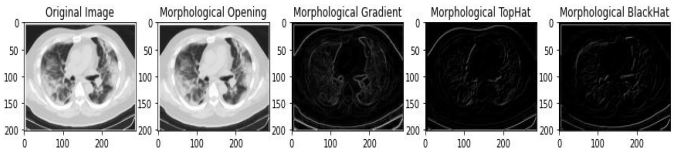

Figure 11: Morphological Operations on CT-Scans

It utilizes a Kernel (usually an all-ones matrix) to govern the nature of operation, performed on a binary image. Basic variants include erosion, which removes similar looking neighbour pixels and dilation, which fills similar neighbouring pixels. Other variants are different combinations of these operations. Fig.10 and Fig.11 represent the outcomes of chest X-Ray and CT-Scan, respectively. Final outcome images lost their basic structural identity, when compared to their original self. This phenomenon was observed more in chest X-Rays.

Considering the above-mentioned techniques and their proscons, we decided to go with the Contrast Limited Adaptive Histogram Equalization (CLAHE), as main pre-processing Technique. For CT-Scans, we also included morphological opening, in order to avoid false loops and refine the opaqueness of patches.

Finally, we resized all the pre-processed images to 224 x 224 and stored in 3-channeled RGB format.

2.3. Embedding Generation via Transfer Learning

Since the availability of pre-processed image dataset is not sufficient on its own to have a better learning of the deep neural architectures, we require it to be expressed in simpler form, i.e., Image Embeddings (N-dimensional feature vector). This procedure was made easier using Transfer Learning [30], which suggests to use a pre-trained deep learning model for more specific domain field such as medical [31], language translation [32], etc., by taking help from the knowledge of a referential domain, like ImageNet[33], MS-COCO[34], etc. This can be easily visualized from Fig.12. More specifically, it is done by

freezing the base network (making it's weights non-trainable) and adding more neural layers for a defined course of action.

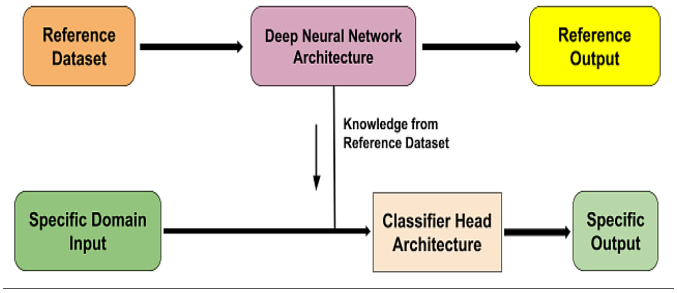

Figure 12: Transfer Learning

We experimented with twenty six different transfer learning models (VGGs[35], ResNets[36][37], DenseNets[38], EfficientNets[39], MobileNets[40][41], NASNets[42], InceptionNets[43][44] and Xception[45]) on the basis of their block architectures and information extraction methods. To understand more about the quality of embeddings, we also studied the extent of task performed from un-normalized output scores (Logits) as well. So, Table 3 and Table 4 show the test accuracy and test AUC (Area Under ROC) scores, found from logits of top-six model variants for chest X-Rays and CT-Scans, respectively.

Table 3: Top-six model variants according to Chest X-Rays un-normalized output scores

Transfer Learning Model	**Accuracy**	**AUC Score**
EfficientNetB0	94.94%	**0.995**
ResNet50	93.62%	0.9908
EfficientNetB1	92.61%	0.9884
DenseNet201	95.02%	0.9881
EfficientNetB2	95.33%	0.989
DenseNet169	**95.8%**	0.9901

Table 4: Top-six model variants according to CT-Scans un-normalized output scores

Transfer Learning Model	**Accuracy**	**AUC Score**
EfficientNetB1	90.66%	0.9825
ResNet101	91.28%	0.9699
EfficientNetB4	**92.92%**	**0.9891**
ResNet50	92.06%	0.9753
EfficientNetB7	91.28%	0.9696
ResNet50V2	92.53%	0.9447

We also fine-tuned the neural network architecture using some configurable parameters and set them according to Table 5. The classification results from these fine-tuned transfer learning models are presented in Table 9 and Table 10.

Table 5: Details of Configurable Parameters

Parameter Type	**Parameter Detail**
Pre-Trained Weights	ImageNet based Weights
Epochs	100
Batch Size	32
Internal Activation Function	Leaky ReLU
Classifier Activation Function	Sigmoid
Optimizer	Adam
Pooling Type	Global Average Pooling
Loss Function	Binary Cross Entropy
Output Vector Dimension	1280

2.4. Multi Task Learning Framework

Solely implementing the transfer learning models is not sufficient to have reliable results yet, because of the following reasons :-

1. Results were only based on nature of scans only. But, in real life cases, only a single scan type can't guarantee full-proof results.
2. Till Now, the image embeddings in raw form are generated, which are not capable of clearly differentiating between a COVID and non-COVID patient. Refining these is the necessity of the situation.
3. We didn't experiment on the common relations that a CT-Scan shares with chest X-Ray images like chest lesions, diaphragm alignment, etc.

To address them all, multi-task learning framework [46][47] is implemented after embedding generation procedure. A visual representation of the proposed architecture is shown in Fig.13, where it comprises of three main components, as follows:-

2.4.1. Task Specific Feature Extraction Module

After obtaining the embeddings from the fine-tuned Transfer Learning Models, those would be furnished according to the nature of scans, i.e., chest X-Ray specific embeddings and CT-Scan specific embeddings, via fully connected dense neural networks with the following characteristics :-

Pyramidal Architecture. The pyramidal design lowers the number of neurons in a layer as the neural network goes more deep, resulting in concise features within limited resources and computational necessities.

Leaky ReLU[48] as Internal Activation Function. It allows more generalized information pass, by preserving the negative quantities in a feature tensor, compared to ReLU [49], totally relying on positive quantities (Here, k is a parametric constant).

$$LeakyReLU(z) = \begin{cases} z & \text{z} \geq 0 \\ -kz & \text{otherwise} \end{cases} \quad (1)$$

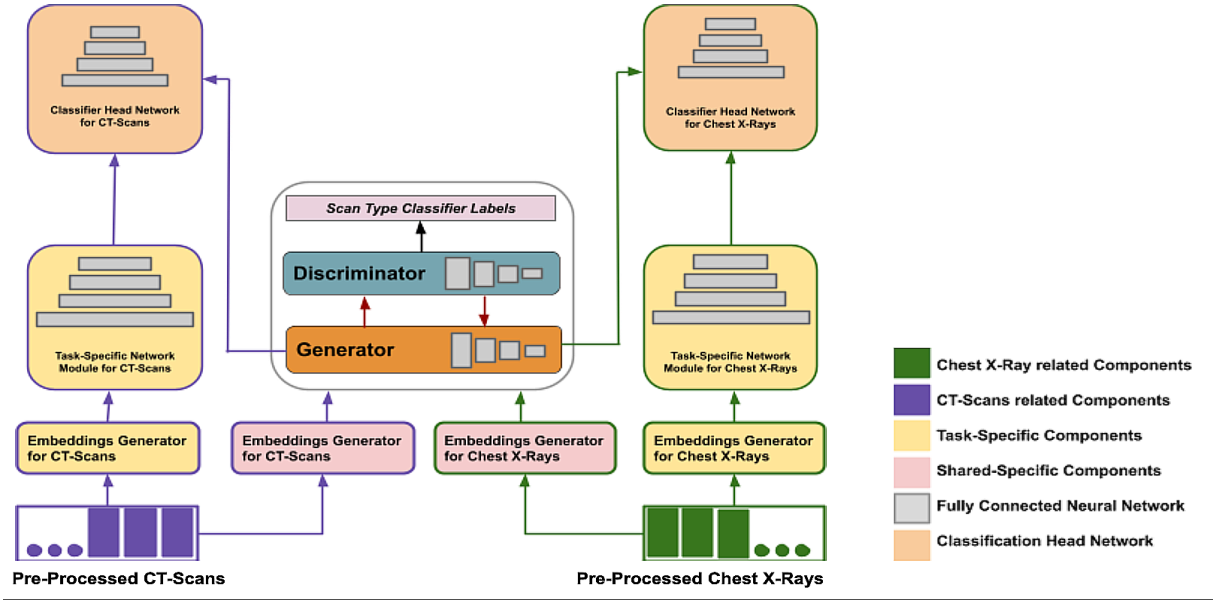

Figure 13: Multi-Task Learning Framework

Batch Normalization[50]. The batch normalization helps to convert each feature of a N-featured vector space, into its respective normal distribution, that results in a standardized built and avoid gradient-related issues like overshooting, vanishing, etc. (Here, γ, β are learnable constants and μ_z, σ_z are mean and variance, respectively.)

$$BatchNorm(z) = \gamma(\frac{z - \mu_z}{\sigma_z}) + \beta \tag{2}$$

Dropout Regularization [51]. The dropouts in between hidden layers eliminate useless/null neurons in a layer via probabilistic manner, such that the neural network avoids overfitting and becomes more generalized in terms of prediction task.

Table 6 and Table 7 represent the task-specific results for chest X-Ray and CT-Scan embeddings, respectively.

Table 6: Task-Specific Metrics for Chest X-Rays Un-Normalized Outputs

Embeddings Model	**Accuracy**	**AUC Score**
EfficientNetB0	**98.04%**	**0.9974**
ResNet50	96.81%	0.9965
EfficientNetB1	96.42%	0.9952
DenseNet201	95.33%	0.9935
EfficientNetB2	95.02%	0.9892
DenseNet169	92.30%	0.9700

2.4.2. Shared Feature Extraction Module

In a real life case, a patient doesn't necessarily posses both chest X-Ray and CT-Scan. In order to extract COVID-indicative features and compensate for other scan types, we have also used common relations that a chest X-Ray and CT-Scan share to each other, via special neural network, with the following properties :-

Table 7: Task-Specific Metrics for CT-Scans Un-Normalized Outputs

Embeddings Model	**Accuracy**	**AUC Score**
EfficientNetB1	**96.96%**	**0.9965**
ResNet101	96.11%	0.9938
EfficientNetB4	96.03%	0.9904
ResNet50	94.71%	0.9911
EfficientNetB7	92.92%	0.9707
ResNet50V2	92.76%	0.9690

Adversarial Training[52]. This part of the shared feature extraction module comprised of two deep neural networks, i.e., generator and discriminator, they work together with the former generating a similar looking output as that of a reference and the later distinguishing between them. Here, both the models try to minimize their losses that lead into min-max competition between them. Instead of real and fake, the discriminator would try to differentiate between chest X-Rays and CT-Scans, which leads to generator intermixing the embeddings of both scan types.

Neural Network Architecture. Followings are the unique features of generator and discriminator architectures :-

- Pyramid-shaped fully connected dense neural networks
- Leaky ReLU as internal activation function
- Adam as optimizer
- 1000 epochs for training
- Binary cross-entropy losses came in range of 10^{-7}.

Wasserstein Distance[53] as Evaluation Metric. Analogous to Euclidean distance, it determines the similarity between images, using their embeddings as inputs. Ideally, 0 indicates

that they would be exactly similar to each other and in our use case, higher the distance would lead to more intermixing between chest X-Rays and CT-Scans features. Following equation represents the wasserstein distance, with μ as mean and C as covariance matrices for embeddings.

$$WD = |\mu_1 - \mu_2|^2 + Tr(C_1 + C_2 - 2(C_1C_2)^{1/2}) \quad (3)$$

Table 8 illustrates the values of wasserstein distance for top-six combinations of chest X-Ray and CT-Scan embeddings.

Table 8: Wasserstein Distances for Combinations of Chest X-Ray and CT-Scans Embeddings

Chest X-Ray Embeddings	**CT-Scan Embeddings**	**Wasserstein Distance**
EfficientNetB0	ResNet50V2	316075.9
EfficientNetB1	ResNet50V2	312193.2
DenseNet169	ResNet50V2	321268.2
EfficientNetB2	ResNet50V2	318603.1
DenseNet201	ResNet50V2	317371.7
ResNet50	ResNet50V2	**325268.7**

Classifier Neural Network Heads. After travelling through the above-mentioned modules, the respective feature matrices merge together and pass through their corresponding classifier heads, with following features :-

- Pyramidal-shaped fully connected dense neural networks
- Leaky ReLU as internal activation function
- Binary cross-entropy as loss function
- Adam as optimizer
- Sigmoid as classifier activation function.

2.5. Evaluation Metrics

In order to measure the robustness of results, following evaluation metrics were used where TP stands for True Positives, FP for False Positives, TN for True Negatives and FN for False Negatives. $AUC - Score$ is a quantitative metric which tells about the classifier model's ability to discriminate between positive class and negative class. Specifically, it is the Area under ROC Curve, which is calculated by choosing different probabilistic thresholds and finding out the values of TPR (True Positive Rate) and FPR (False Positive Rate), as follows :-

$$TPR = \frac{TP}{TP + FN} \quad (4)$$

$$FPR = \frac{FP}{TN + FP} \quad (5)$$

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

$$Sensitivity = \frac{TP}{TP + FN} \quad (7)$$

$$Specificity = \frac{TN}{TN + FP} \quad (8)$$

$$F1 - Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} = 2\frac{Precision * Recall}{Precision + Recall} \quad (9)$$

3. Results

Table 9: Evaluation Results for the Chest X-Ray Classification Task on COVID Detection

Transfer Learning Model	**Acc**	**AUC**	**Sn**	**Sp**	**F1-Score**
EfficientNetB0	94.47%	0.9925	97.90%	92.10%	93.53%
DenseNet169	94.01%	0.9763	96.76%	91.84%	92.78%
EfficientNetB2	94.79%	0.9892	98.85%	90.65%	93.09%
EfficientNetB1	93.85%	0.9883	97.52%	**92.89%**	93.85%
DenseNet201	92.37%	0.9926	**99.42%**	87.50%	91.41%
ResNet50	**95.18%**	**0.9934**	99.04%	92.50%	**94.37%**

Table 10: Evaluation Results for the CT-Scan Classification Task on COVID Detection

Transfer Learning Model	**Acc**	**AUC**	**Sn**	**Sp**	**F1-Score**
EfficientNetB4	**94.24%**	0.9842	95.04%	92.76%	92.49%
ResNet50	93.46%	**0.9890**	93.33%	**94.86%**	92.97%
ResNet101	92.84%	0.9783	94.28%	92.89%	92.17%
ResNet50V2	94.01%	0.9578	96.19%	90.52%	91.65%
EfficientNetB1	93.7%	0.9862	**97.52%**	91.57%	**93%**
EfficientNetB7	91.91%	0.9660	94.28%	90.26%	90.49%

Table 9, Table 10 and Table 11 represent the classification task results for COVID-19 detection using Chest X-Rays and CT-Scans, by Task Specific Transfer Learning models and Multi-Modal Multi-Task Learning Framework, respectively. If we consider Table 9, then the best performing Chest X-Ray Task Specific Transfer Learning architecture is **ResNet50** with AUC-score of 99.34% and accuracy of 95.18%, while Table 10 shows **EfficientNetB1** as the best CT-Scan Task Specific Transfer Learning architecture with AUC value of 98.62% and sensitivity of 97.52%. If we move further from Uni-Modal Classification to Multi-Modal Classification Task where one modality is the Task Specific Features and other is the Shared common features between all the Task Specific modules, then the prediction accuracy has improved from 94.47% to 98.28% and 93.70% to 98.83% for the chest X-Ray and CT-Scan based classifications, respectively. These improvements are the results of comparative analyses between **EfficientNetB0** from Table 9 and combinations of **EfficientNetB0** with **ResNet50** from Table 11 followed by **EfficientNetB1** from Table 10 and combinations of **EfficientNetB0** with **ResNet50V2** from Table 11.

Fig 14 and Fig 16 represent the Confusion Matrices for best variants of Task Specific Transfer Learning models and Multi-Task Learning Framework, applied on Chest X-Rays, respectively. Fig 15 and Fig 17 represent the Confusion Matrices for

Table 11: Evaluation Results for Adversarial and Non-Adversarial Multi-Task Classification on COVID-19 Detection

Task Specific and Shared Features	Acc	AUC	Sn	Sp	F1 Score	Acc	AUC	Sn	Sp	F1 Score
	(Adversarial)					(Non-Adversarial)				
X-Ray: EfficientNetB0 and ResNet50 **CT-Scan: EfficientNetB1 and ResNet50V2**	**98.28%** **98.83%**	0.9953 0.9985	98.09% 97.71%	**98.42%** 99.6%	97.9% **98.55%**	96.11% **98.21%**	0.9915 **0.994**	99.04% 97.52%	94.07% **98.68%**	95.41% **97.8%**
X-Ray: EfficientNetB0 and EfficientNetB2 **CT-Scan: EfficientNetB1 and ResNet50V2**	97.58% 98.52%	**0.9968** 0.9969	98.66% 96.76%	96.84% **99.73%**	97.09% 98.16%	**97.98%** 95.95%	**0.9974** 0.9599	98.28% 95.23%	**97.76%** 96.44%	**97.54%** 95.05%
X-Ray: EfficientNetB0 and DenseNet169 **CT-Scan: EfficientNetB1 and ResNet50V2**	97.19% **98.83%**	0.9911 **0.9993**	99.61% 97.71%	95.52% 99.6%	**99.6%** **98.55%**	95.41% 97.82%	0.9643 0.9805	99.43% **98.47%**	92.63% 97.36%	94.65% 97.36%
X-Ray: EfficientNetB1 and DenseNet169 **CT-Scan: EfficientNetB4 and ResNet50V2**	97.04% 97.51%	0.9818 0.9864	**99.81%** 98.09%	95.13% 97.1%	96.5% 96.99%	96.26% 95.56%	0.9729 0.9548	99.43% 93.33%	94.07% 97.10%	95.6% 94.50%
X-Ray: EfficientNetB1 and ResNet50 **CT-Scan: EfficientNetB4 and ResNet50V2**	96.96% 96.96%	0.991 0.9769	**99.81%** **98.85%**	95% 95.65%	96.4% 96.37%	96.58% 96.89%	0.9874 0.9784	**100%** 95.61%	94.21% 97.76%	95.97% 96.16%
X-Ray: ResNet50and EfficientNetB2 **CT-Scan: ResNet101 and ResNet50V2**	96.11% 97.27%	0.9919 0.9939	99.61% 96.19%	93.68% 98.02%	95.43% 96.65%	95.80% 94.86%	0.9929 0.9566	**100%** 96.00%	92.89% 95.13%	95.10% 94.55%

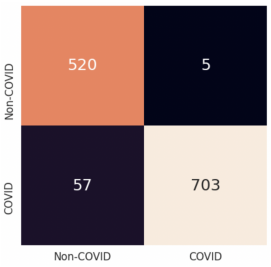

Figure 14: ResNet50 on chest X-Rays

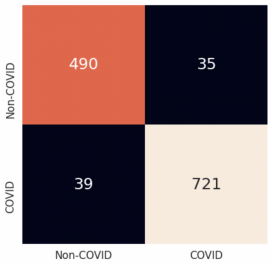

Figure 15: EfficientNetB4 on CT-Scans

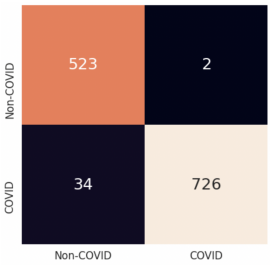

Figure 16: EfficientNetB0 and ResNet50 on chest X-Rays

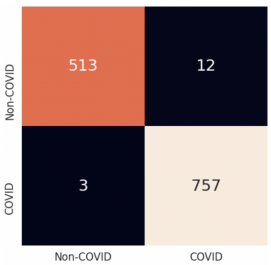

Figure 17: EfficientNetB1 and ResNet50V2 on CT-Scans

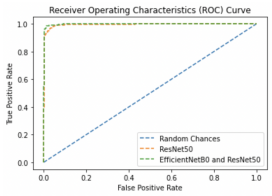

Figure 18: Receiver Operating Characteristics for chest X-Rays

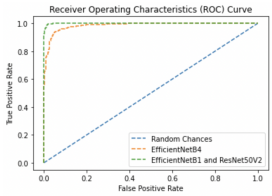

Figure 19: Receiver Operating Characteristics for CT-Scans

best variants of Transfer Learning model and Multi-Task Learning Framework, applied on CT-Scans, respectively. Fig 18 and Fig 19 represent the Receiver Operating Characteristics (ROC) Curves, which compares the best variants of Transfer Learning models and Multi-Task Learning Frameworks for Chest X-Rays and CT-Scans, respectively. The statistical significance of the proposed architecture is validated using the **ANOVA Test**. For this test, we have used the **prediction probabilities** of our proposed model along with other transfer learning based state-of-the art models. **f-value** and **p-value** are **0.831** and **0.587** for X-Ray based COVID-19 classification and **0.107** and **0.999** for CT-Scan based COVID-19 classification, respectively.

To show the efficacy of adversarial framework in shared feature extraction task, we have performed the comparative study of COVID-19 detection in both frameworks, with and without adversarial setups. From Table 11, it is clear that the adversarial architecture is providing better shared features which helps in improving the final COVID-19 detection measures as compared to non-adversarial architecture.

4. Conclusion

In this paper, we have proposed a Multi-Modal Multi-Task deep learning method using concatenation of extracted features from two different Task Specific Transfer Learning models and Shared Features from the combination of these Task Specific Transfer Learning models for classifying COVID-19 patients with two modalities, CT-Scan and X-Rays. These modalities are frequently used to diagnose diseases that attack the human respiratory system and each of these imaging provides plenty of information related to the infected chest. In this research, we have used two open-source datasets and further balanced the number of samples allocated for each class. The final proposed architecture for the chest X-Rays is **EfficientNetB0** as Task Specific module integrated with **ResNet50** as Shared Feature module and for CT-Scans is **EfficientNetB1** as Task Specific module in integration with **ResNet50V2** as Shared Feature Extraction module. The proposed architecture improves the prediction capability by integrating the shared features from both chest imaging as complementary information related to COVID-19 infection. It can work even though patients are having only chest X-Rays or CT-Scans. In future scope of this work, we can add clinical bio-markers as additional source of information. The more availability of complete multi-modal COVID-19 dataset (i.e., all patients are having X-Ray, CT-scan and clinical details) will motivate us to develop a complete multi-modal framework and perform the comparative study with the proposed architecture.

References

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in wuhan, china, The Lancet 395 (10223) (2020) 497–506. doi:https://doi.org/10.1016/S0140-6736(20)30183-5.
URL https://www.sciencedirect.com/science/article/pii/S0140673620301835

[2] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study, The Lancet 395 (10223) (2020) 507–513. doi:https://doi.org/10.1016/S0140-6736(20)30211-7.
URL https://www.sciencedirect.com/science/article/pii/S0140673620302117

[3] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng, Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China, JAMA 323 (11) (2020) 1061–1069. arXiv:https://jamanetwork.com/journals/jama/articlepdf/2761044/jama_wang_2020_oi_200019.pdf, doi:10.1001/jama.2020.1585.
URL https://doi.org/10.1001/jama.2020.1585

[4] M. Hosseini, M. Zekri, Review of medical image classification using the adaptive neuro-fuzzy inference system, Journal of medical signals and sensors 2 (2012) 49–60. doi:10.4103/2228-7477.108171.

[5] C. Quek, W. Irawan, E. Ng, A novel brain-inspired neural cognitive approach to sars thermal image analysis, Expert Systems with Applications 37 (4) (2010) 3040–3054. doi:https://doi.org/10.1016/j.eswa.2009.09.028.
URL https://www.sciencedirect.com/science/article/pii/S0957417409008094

[6] X. Xie, X. Li, S. Wan, Y. Gong, Mining X-Ray Images of SARS Patients, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 282–294. doi:10.1007/11677437_22.
URL https://doi.org/10.1007/11677437_22

[7] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, A. Y. Ng, Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning (2017). arXiv:1711.05225.

[8] T. Cherian, E. K. Mulholland, J. B. Carlin, H. Ostensen, R. Amin, M. de Campo, D. Greenberg, R. Lagos, M. Lucero, S. A. Madhi, K. L. O'Brien, S. Obaro, M. C. Steinhoff, Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies, Bull World Health Organ 83 (5) (2005) 353–359.

[9] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR, Radiology 296 (2) (2020) E115–E117.

[10] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, E. Siegel, Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis (2020). arXiv:2003.05037.

[11] I. Apostolopoulos, M. Tzani, Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, Australasian physical & engineering sciences in medicine / supported by the Australasian College of Physical Scientists in Medicine and the Australasian Association of Physical Sciences in Medicine 43 (03 2020). doi:10.1007/s13246-020-00865-4.

[12] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks (08 2021). doi:10.1007/s10044-021-00984-y.

[13] E. E.-D. Hemdan, M. A. Shouman, M. E. Karar, Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images (2020). arXiv:2003.11055.

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255. doi:10.1109/CVPR.2009.5206848.

[15] L. Wang, Z. Lin, A. Wong, Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, Scientific Reports 10 (11 2020). doi:10.1038/s41598-020-76550-z.

[16] M. Farooq, A. Hafeez, Covid-resnet: A deep learning framework for screening of covid19 from radiographs (2020). arXiv:2003.14395.

[17] Y. Zoabi, S. Deri-Rozov, N. Shomron, Machine learning-based prediction of COVID-19 diagnosis based on symptoms, NPJ Digit Med 4 (1) (2021) 3.

[18] J. P. Cohen, P. Morrison, L. Dao, Covid-19 image data collection, arXiv (2020).
URL https://github.com/ieee8023/covid-chestxray-dataset

[19] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, M. Ghassemi, Covid-19 image data collection: Prospective predictions are the future, arXiv 2006.11988 (2020).
URL https://github.com/ieee8023/covid-chestxray-dataset

[20] J. Zhao, Y. Zhang, X. He, P. Xie, Covid-ct-dataset: a ct scan dataset about covid-19, arXiv preprint arXiv:2003.13865 (2020).

[21] S. Sarkar, Covid 19 xray and ct scan image dataset (Jan 2021).
URL https://www.kaggle.com/ssarkar445/covid-19-xray-and-ct-scan-image-dataset

[22] D. Bradley, G. Roth, Adaptive thresholding using the integral image, Journal of graphics tools 12 (2) (2007) 13–21.

[23] N. Otsu, A Threshold Selection Method from Gray-level Histograms, IEEE Transactions on Systems, Man and Cybernetics 9 (1) (1979) 62–66. doi:10.1109/TSMC.1979.4310076.
URL http://dx.doi.org/10.1109/TSMC.1979.4310076

[24] P. J. Burt, E. H. Adelson, The laplacian pyramid as a compact

image code, in: Readings in computer vision, Elsevier, 1987, pp. 671–679.
[25] N. Kanopoulos, N. Vasanthavada, R. L. Baker, Design of an image edge detection filter using the sobel operator, IEEE Journal of solid-state circuits 23 (2) (1988) 358–367.
[26] J. Canny, A computational approach to edge detection, IEEE Transactions on pattern analysis and machine intelligence (1986) 679–698.
[27] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations, Computer vision, graphics, and image processing 39 (3) (1987) 355–368.
[28] A. M. Reza, Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement., VLSI Signal Processing 38 (1) (2004) 35–44.
URL http://dblp.uni-trier.de/db/journals/vlsisp/vlsisp38.html#Reza04
[29] M. L. Comer, E. J. Delp III, Morphological operations for color image processing, Journal of electronic imaging 8 (3) (1999) 279–289.
[30] S. Bozinovski, A. Fulgosi, The influence of pattern similarity and transfer learning upon training of a base perceptron b2, in: Proceedings of Symposium Informatica, 1976, pp. 3–121.
[31] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding transfer learning for medical imaging, arXiv preprint arXiv:1902.07208 (2019).
[32] B. Zoph, D. Yuret, J. May, K. Knight, Transfer learning for low-resource neural machine translation, arXiv preprint arXiv:1604.02201 (2016).
[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, Ieee, 2009, pp. 248–255.
[34] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: common objects in context, CoRR abs/1405.0312 (2014). arXiv:1405.0312.
URL http://arxiv.org/abs/1405.0312
[35] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.
URL http://arxiv.org/abs/1409.1556
[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, CoRR abs/1512.03385 (2015). arXiv:1512.03385.
URL http://arxiv.org/abs/1512.03385
[37] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, CoRR abs/1603.05027 (2016). arXiv:1603.05027.
URL http://arxiv.org/abs/1603.05027
[38] G. Huang, Z. Liu, K. Q. Weinberger, Densely connected convolutional networks, CoRR abs/1608.06993 (2016). arXiv:1608.06993.
URL http://arxiv.org/abs/1608.06993
[39] M. Tan, Q. V. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, CoRR abs/1905.11946 (2019). arXiv:1905.11946.
URL http://arxiv.org/abs/1905.11946
[40] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, CoRR abs/1704.04861 (2017). arXiv:1704.04861.
URL http://arxiv.org/abs/1704.04861
[41] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, L. Chen, Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation, CoRR abs/1801.04381 (2018). arXiv:1801.04381.
URL http://arxiv.org/abs/1801.04381
[42] B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le, Learning transferable architectures for scalable image recognition, CoRR abs/1707.07012 (2017). arXiv:1707.07012.
URL http://arxiv.org/abs/1707.07012
[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, CoRR abs/1512.00567 (2015). arXiv:1512.00567.
URL http://arxiv.org/abs/1512.00567
[44] C. Szegedy, S. Ioffe, V. Vanhoucke, Inception-v4, inception-resnet and the impact of residual connections on learning, CoRR abs/1602.07261 (2016). arXiv:1602.07261.
URL http://arxiv.org/abs/1602.07261
[45] F. Chollet, Xception: Deep learning with depthwise separable convolutions, CoRR abs/1610.02357 (2016). arXiv:1610.02357.
URL http://arxiv.org/abs/1610.02357
[46] S. Yadav, S. Ramesh, S. Saha, A. Ekbal, Relation extraction from biomedical and clinical text: Unified multitask learning framework, IEEE/ACM Transactions on Computational Biology and Bioinformatics (2020).
[47] S. A. Qureshi, G. Dias, M. Hasanuzzaman, S. Saha, Improving depression level estimation by concurrently learning emotion intensity, IEEE Computational Intelligence Magazine 15 (3) (2020) 47–59.
[48] A. L. Maas, A. Y. Hannun, A. Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. icml, Vol. 30, Citeseer, 2013, p. 3.
[49] A. F. Agarap, Deep learning using rectified linear units (relu), arXiv preprint arXiv:1803.08375 (2018).
[50] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, CoRR abs/1502.03167 (2015). arXiv:1502.03167.
URL http://arxiv.org/abs/1502.03167
[51] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research 15 (1) (2014) 1929–1958.
[52] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680.
[53] J. Brownlee, How to evaluate generative adversarial networks (Jul 2019).
URL https://machinelearningmastery.com/how-to-evaluate-generative-adversarial-networks/