MetaEfficientNet: A Few shot learning approach for lung disease classification

Shravani Nimbolkar[1], Anuradha Thakare[2], Subhradeep Mitra[3], Omkar Biranje[4], and Anant Sutar[5]
[1,2,3,4,5]Department of Computer Engineering, Pimpri Chinchwad College of Engineering,
Savitribai Phule Pune University
[1]shravani.nimbolkar17@pccoepune.org
[2]adthakare2014@gmail.com
[3] subhradeep.mitra17@pccoepune.org
[4]omkar.biranje17@pccoepune.org
[5]anant.sutar17@pccoepune.org

Abstract. Medical image diagnosing has been tremendously improved by making use of machine learning and deep learning algorithms in the last few decades. These algorithms have achieved human-like precision and accuracy in terms of identifying and distinguishing ailments particularly Lung related. This human-like accuracy comes at the cost of data. Conventional machine learning and deep learning methods achieve higher accuracy with vast amounts of data and their accuracy plummets when either there is data scarcity or data imbalance. To prevail over this issue we make use of Siamese net which is a metric spaced meta learning algorithm. Meta-learning mimics the idea of human learning from a minimal amount of data. Few-shot learning gains an upper hand over traditional algorithms by making use of a nominal amount of data. Using this approach we proposed the MetaEfficientNet and for the purpose of comparison, we used the CNN and VGG-16 based Siamese network as the baseline models. The proposed model achieved an AUC score of 0.9734 and accuracy of 97% which was best among the three models. Also, our proposed model achieved an accuracy score of 98% with a 3-way-8-shot approach and gave above 75% on newly given classes of lung diseases as well with a 6-way-8-shot learning approach.

Keywords: COVID-19 diagnosis, CXR images, Meta learning, Few-shot learning, Siamese network, EfficientNetB0.

1 Introduction

Lungs play a major role in the functioning of the body. it becomes very much necessary to ensure their proper functioning. Lung diseases are a big problem in today's society, where pollutants and dangerous viruses play a big part. The manual diagnostics of these kinds of diseases, although possible, has some drawbacks, and thus we think it’s the need of the hour to implement automated systems, using the latest available techniques. Lung Diseases also known as respiratory diseases registers as one of the leading causes of mortality today. Around 65 million people suffer from one of the other forms of lung disease and about 3 million die per year [1]. The severity of chronic respiratory diseases can be understood by considering the very latest example of COVID-19. There are different types of lung diseases and diagnosing them using CXR with the naked eye by a specialist or by a naive person may give us misleading results. Pneumonia and Tuberculosis may look similar to an untrained person but require completely different diagnosis and treatment. Early diagnosis, which is critical in the treatment of any disease, improves doctor-patient communication, enables for more focused diagnosis and therapy, and, most significantly, aids in the detection of early warning symptoms.

Medical imaging, such as X-ray scanning, bundled with Artificial Intelligence (AI) technologies, is an encouraging and logical tool for the detection and classification of Lung Diseases. Deep learning and Machine Learning methodologies have proved crucial in identifying such diseases. Methods such as CNN where we are able to diagnose ailments such as Tuberculosis [2], Pneumonia [3], COVID-19 [4], and many others are having human-like abilities to classify such diseases. Further to improve accuracy and save time and effort, the concept of transfer learning was introduced where it was able to classify disease using networks and weights. [5] contributed a large-scale survey of deep learning methods to identify lung diseases. Albeit such promising results, a problem that the above methods face is the quantity of data. Deep learning has conveniently set a new standard for object detection, however, to achieve such benchmarks it needs tons of data to build its intuitive behavior. Also in some domains, the availability of data is scarce. The solution for such problems is meta-learning. Meta-learning refers to adapting and learning quickly through the use of a few examples. One-shot and Few-shot learning use one or more samples to learn and take that information to achieve tasks such as detection and classification. In this paper, we propose a meta-learning approach to classify lung diseases. The following paper is divided into sections. Section 2 gives an account of related studies such as Transfer learning and Meta-learning techniques. Section 3 provides details on the proposed model. Section 4 gives the algorithm of the proposed mode. Section 5 gives evaluation parameters used. Section 6 discusses the result of our experimental studies and a comparative analysis.

2 Related Study

2.1 Transfer Learning

Transfer learning is an approach in deep learning wherein the pre-trained models are used as the first step in computer vision. It utilizes the information gathered from one task to solve tasks related to it. Pre-trained models are trained rigorously on a large image dataset like ImageNet dataset with l.2-million color images and 1000 classes, which is called a base dataset consisting of a wide variety of samples. The knowledge acquired throughout the process is then transferred while performing new tasks. Since developing neural network models on complex problems require high computations and time, transfer learning reduces the time it takes to construct and train a model by reusing modules of previously trained models. We are able to import transfer learning models and apply them to problems such as image classification, text classification, etc. Over the years, the pre-trained model approach has gained traction due to its ease and efficiency. First, we select a source model from all the available models. There are numerous models present such as VGG-16, ResNet, DenseNet, etc. Once we select our source model we then according to our convenience can modify or tune it. These features allow people to construct models that suit their needs and make such models flexible. Then we can use the tuned model to solve problems such as detection, classification or optimization.

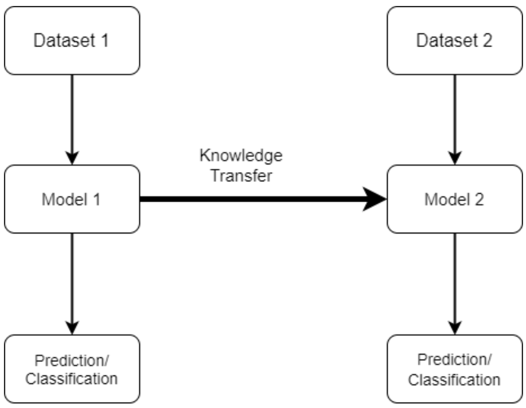

Fig. 1. Transfer Learning

The following study was conducted to bolster the importance of transfer learning. Hashmi MF et al.[6], proposed a high-performance model applicable to digital chest X-ray images. It introduced a weighted classifier approach, which combined weighted predictions from ResNet18, InceptionV3, DenseNet121, etc in an optimal way. To fine-tune the models, Transfer Learning was used. The Guangzhou Women and Children's Medical Center pneumonia dataset was used in this study. To increase the training dataset in a balanced way, partial data augmentation was employed. The model achieved a test accuracy of 98.43% and an AUC score of 99.76. A major drawback of this approach was the scarcity of available resources resulting in overfitting. And it adversely affected models' generalization ability. The diagnosis of the disease requires a deep understanding of the radiological features from Chest X-rays. S. V. Militante et al.[7], proposed a methodology that employs efficient approaches of 6 CNN models, to recognize and predict patients affected and unaffected with the disease, using a Chest X-ray image. The models used were: LeNet, VGG-16, StridedNet, GoogLeNet, AlexNet, and ResNet-50. The

Radiological Society of North America (RSNA) dataset, which consisted of a total of 28,000 chest X-Ray images, was used. The images were in JPEG format, having maximum dimensions of 1024 x 1024. All the images were categorized into two classes: infected and not infected. All the models attained a training accuracy of 96-98%, with GoogLeNet and LeNet obtaining the highest, i.e 98% accuracy for performance training. This study can also consider the optimization of hyper-parameters to improve the accuracy of the model. Ahsan et al.[8], proposed applying the VGG model on the chest X-ray dataset to spot if a patient is suffering from TB. The dataset used by the author consisted of 1324 CXRs and text data from Shenzhen and 276 from Montgomery datasets. The dataset was divided into training and testing datasets with a 75 to 25 ratio respectively. The VGG-16 model was used however instead of using a soft-max layer the author preferred to use a sigmoid layer as they only had two outputs. The model was able to attain 80% accuracy without augmentation and later achieved 81.2% with it. TFLearns Data Augmentation of TensorFlow was used for augmentation which helped in the making of new images and also alleviates data overfitting in the model. The author mentioned that the low accuracy was due to partial augmentation of the dataset and it could be improved by running the model on a high configuration system.

2.2 Meta-Learning

Meta-learning is a subfield of machine learning that focuses on classification and regression tasks and employs a learn-to-learn method. Because the algorithm learns from the output of other algorithms, it's also known as a meta learner. A successful meta-learning model should be able to adapt or generalize to new tasks that haven't been encountered yet during the training phase. There are three types in meta-learning: learning in metric space, learning the initializers, and learning the optimizers [9].

Learning in metric space. This method is similar to that used in closest neighbor algorithms. The model learns to find an optimal similarity distance function. It employs the kernel density estimation concept, in which the probabilities predicted for a particular input are determined as a weighted sum of its labels, with the weights generated by a kernel function. This approach attempts to determine the degree of similarity between a given image and the image in the support set, which are samples from the database. The few-shot learning is a very popular problem that is solved with metric-based learning.

Learning the initializers. The weights of a neural network are randomly initialized at the beginning of the training phase. These weights are changed throughout training by reducing losses via gradient descent. However, the convergence can occur early if the weights are initialized to their optimal values or close optimal values. We can find these initial optimal values for weights using algorithms like Reptile, Meta-SGD, and MAML.

Learning the optimizers. In this method, we train the optimizers. The neural networks are further optimized after the training phase by lowering losses and training on larger datasets. However, because few-shot or one-shot learning models are trained on a significantly smaller collection of data, gradient descent may fail. In optimization-based meta-learning, we can have the base model for learning and a meta-model for optimizing the base model.

2.3 One shot and few shot

One-shot and few shot models, like the human brain, can learn to differentiate between images belonging to multiple classes. With metric-based meta-learning, few-shot learning is widely employed. Few-shot learning comes in four flavors: zero-shot learning, one-shot learning, and N-shot learning. The N-way-K shot problem is classified as a typical few-shot learning problem where, the N refers to the number of class labels and K refers to the number of samples provided for every class. For the N-way-K shot problem, the task which includes N-way K samples is known as the support set. Few more examples of the same classes are included in the Query set which is used for performance evaluation of the task. One-shot can also be termed as N-way 1 shot problem where the model is fed with a single sample for each class. Siamese neural networks are the most commonly used metric-based one-shot and few shot learning algorithms as similarity networks. It consists of a pair of convolutional neural networks and outputs the similarity measures for given samples. (Shuti Jadon, 2020) has presented a detailed study of Deep learning in Few-shot learning [10].

2.4 Siamese Neural Network

Siamese networks are a class of neural networks and fall under the category of meta-learning or learning to learn. The network is a combination of the same embedding models which produce an embedding for each input image, i.e., each image is fed to one embedding model. The produced embeddings for each image is compared with embeddings of others as per the dissimilarity measure chosen. This dissimilarity measure is then passed to a contrastive loss function, which then updates the weights of the embedding model.

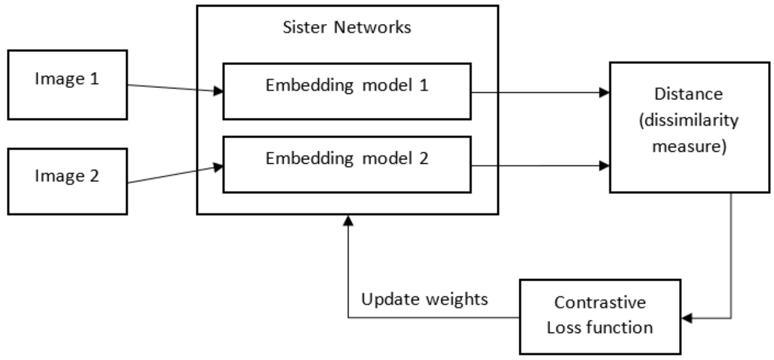

Fig. 2. General Structure of a simple Siamese Network for meta learning

Li et al. [11] discusses the schematics of Siamese networks on medical images taking the Euclidean distance between the embedding outputs produced by two images as a measure of their dissimilarity. This dissimilarity is used to train the embedding model to produce better embedding so that similar images have the least dissimilarity, and images from different classes will have higher dissimilarity. For this they used the ROP dataset which contained x-ray images of knee joints of patients with pain severity (0-5) of the area as the class or category of the x-ray. They achieved an AUC of 0.81 (95% CI 0.77 - 0.84) with Cohen's kappa 0.64 (95% CI 0.58 - 0.68) and an AUC of 0.90 (95% CI 0.88 - 0.92) with Cohen's kappa 0.66 (95% CI 0.61–0.71) using the Euclidean distance difference. The network trained for knee osteoarthritis achieved an AUC of 0.90 (95% CI 0.86 - 0.83) and 0.88 (95% CI 0.84 - 0.92) with Cohen's kappa 0.47 (95% CI 0.40 - 0.54) and 0.41 (95% CI 0.35 - 0.48) respectively, using the Euclidean distance difference. For ROP, the conventional neural network achieved a linear Kappa of 0.61 (95% CI 0.55 - 0.66). For knee osteoarthritis, the conventional neural network achieved a linear Kappa of 0.46 (95% CI 0.39 - 0.54). It found overlap in 95% in the confidence intervals of the Siamese network, thus stating the performance of Siamese networks on binary change detection.

Mohammad Shorfuzzaman et al. [12] pointed out the high data required for training the traditional deep learning algorithms, and if the dataset is insufficient, then it needs to be handled by using data augmentation techniques like GANs. The newly generated data if hand-tuned then has a high chance of over-fitting, or if generated through GANs face challenges in simulating real patient data which leads to unanticipated bias during model testing as the augmented data may be impossible to exist in nature. They used both loss functions, binary cross-entropy, and contrastive loss function. The chest x-ray images were rescaled to 100 x 100 and each pixel value was converted from [0, 255] to [0, 1]. The model was evaluated on different metrics and got the highest accuracy of 95.6% for the 3-way 10-shot learning approach using contrastive loss and 93.8% for the 3-way 10-shot learning approach using cross-entropy loss. The model was compared with other pre-existing models like InceptionV3, Xception, InceptionResNetV2 and VGG-16, the proposed model outperformed all of the mentioned models. The model was also evaluated on a 2-way 10-shot approach and got an accuracy of 96.5%.

K. Prayogo et al. [13] discusses the complex nature of the existing methods for the diagnosis of pneumonia disease. Most of the studies on image classification using deep learning approaches use CNN. CNN has been used in several problems like classifying stroke, type of muscle, and abdominal ultrasound images. They proposed a model for the classification of chest x-ray images into Normal, Bacterial Pneumonia, and Viral Pneumonia. The dataset used in this study was Labelled Optical Coherence Tomography (OCT) and Chest X-Ray Images, which contained 5863 images. They used a convolutional model as the embedding model and followed by a Fully connected network that accepts inputs from the embedding model and outputs the distance measure. For training the weights the cosine distance was used as the dissimilarity function. The model resulted in an accuracy of 80.03% for the 20-shots approach.

3 Proposed System

3.1 Data Acquisition

For this study, we used the COVID-19 Radiography dataset and Chest X-ray Images dataset [14, 15] which contains chest x-rays images categorized into three classes: Normal, Covid and Viral Pneumonia containing a total of more than 15,000 images.

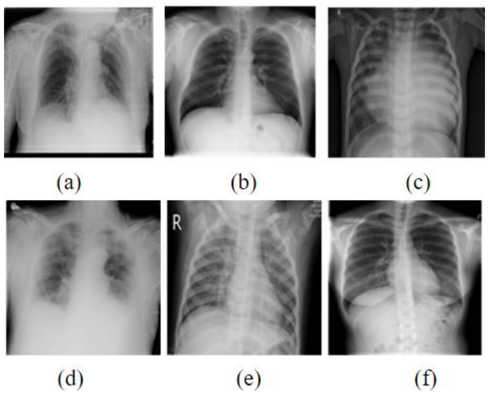

Fig. 3. Illustrates some sample images from the COVID19 Radiography and Pneumonia Dataset (a),(d) Covid, (b), (f) Normal, (c), (e) Viral Pneumonia

3.2 Siamese neural network using pre-trained models

In general, we use supervised metric-based learning to train the Siamese net to learn the image representation in feature space, and then use it for few shot learning without further training. The Siamese network is made up of two parallel pre-trained CNN embedding models, also known as sister networks, in our proposed architecture. The 'twin' here signifies that these two networks have the same parameters and weights. Each network creates an embedding vector of size 128 of its individual inputs after being given a pair of images. These sister networks are then trained to maximize the distance between input embeddings for different classes while minimizing the distance between embeddings for similar classes. An energy function sits atop the sister network, calculating the metric between the embeddings from these twin networks. In this paper, we have proposed a Siamese neural network which uses EfficientNetB0 architecture in the Siamese neural network as shown in Figure 4.

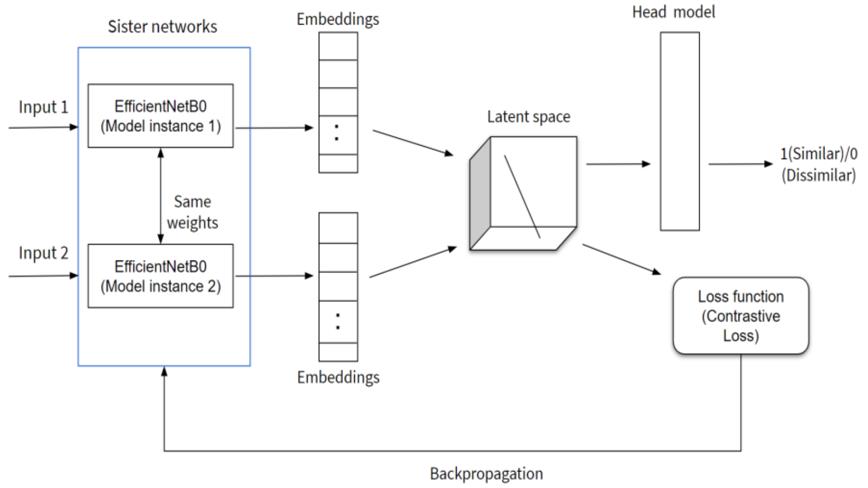

Fig. 4. The proposed architecture, Siamese network using EfficietNetB0 architecture. The blue box envelopes the sister network which is fed with a pair of images. The outputs of that network are the feature embedding of two images which are given to the loss function

Head model. The final result is produced by stacking the head model on top of the two sister networks, as shown in Figure 5. It has a Lambda layer that takes the embeddings vectors of size 128 from the sister models as input and computes the Euclidean distance between the embeddings. The lambda layer is followed by the Batch Normalization layer which normalizes the input coming from the Lambda layer. The Dense layer is the output layer with Sigmoid activation function which produces the final similarity score.

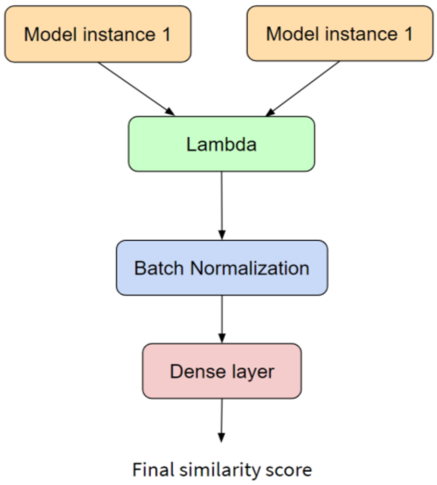

Fig. 5. The head model

Contrastive Loss function. For the proposed model, learning is done using the contrastive loss function. The choice of the loss function is a crucial step because in the training phase the loss function makes the base model learn to obtain the feature embedding of the input image. Various loss functions, such as Triplet loss, contrastive loss, and circle loss, are also used to train Siamese networks. The contrastive loss function, which we used in our study, requires a pair of images as input. The function penalizes the base model for generating false feature embeddings based on the label given to the pair of input images. In a nutshell, the loss comes out to be low for similar or closer embedding vectors for images of the same class and high for dissimilar embedding vectors for the images of different classes[16]. Mathematically it is given as:

$$Contrastive\ loss = Y(D)^2 + (1 - Y)max(margin - D, 0)^2 \quad (1)$$

In the above equation, Y is the true label which will be 1 when two input images will be the same and 0 if those are dissimilar. D is our Euclidean distance between two embeddings. $D = ||I_1 - I_2||_2$ where I_1 and I_2 are the two feature embeddings of two inputs. The margin is to hold a constraint over the loss; when two input values are dissimilar, and if their distance is greater than a margin, then they do not incur a loss.

Transfer learning and Siamese neural networks. Transfer learning discussed earlier in section 2.1. is an advanced and smart approach as compared to traditional Machine learning, wherein the pre-trained models are used for a new but similar task. Transfer learning works better than Machine learning if the model features gained from the previous task are generic enough. Transfer Learning provides the distinct benefit of speeding up the model training process and fine-tuning the results. In our study, we found that Transfer learning has a great potential to leverage the performance of Siamese neural networks and few-shot learning. We experimented with the two very famous pre-trained architectures and fine-tuned them to suit our CXR dataset to avoid training the network from scratch.

These two differ not only by their architecture and depth but also by their sizes. The VGG16 with size 549 MB due to its depth and a large number of fully connected layers is pretty heavier than EffiicentNetB0 of size 29 MB. The two architectures are discussed below in detail

VGG16. The VGG16 is famously used for image classification especially for medical imaging [17]. The VGG16 network is considered one of the most important CNNs for image classification because of its deep yet simple architecture, which gives it robustness against overfitting while providing good performance. Due to the smaller kernel size of VGG16, it can extract intricate features in the CXR which are crucial for image classification [18]. The VGG16 has 16 layers out of which 13 are convolutional layers with the Relu activation function and 3 are fully connected layers, with all kernel sizes of 3x3. Each convolution layer is followed by a max-pooling layer with all 2x2 kernel sizes. Convolution layers function as automatic feature extraction that stores training weights. The next layer is 3 fully connected layers (FC) which are the final layer as a classifier[19]. A detailed architecture of VGG16 is shown in figure 6.

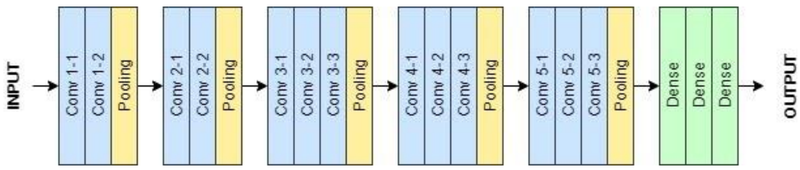

Fig. 6. Standard VGG16 pretrained model with 16 convolutional layers.[20]

EfficienNetB0. EfficientNet was first introduced by M. Tan and Le[21] in 2019. These models are based on simple and highly effective compound scaling methods. This method enables to scale up a baseline ConvNet to any target resource constraints while maintaining model efficiency, used for transfer learning datasets

[22]. EfficientNet has been proposed to improve the performance of CNNs by scaling in three dimensions, i.e., width, depth, and resolution using a set of fixed scaling coefficients that meet some specific constraints [23]. In general, EfficientNet models achieve both higher accuracy and better efficiency over existing CNNs such as AlexNet, ImageNet, GoogleNet, and MobileNetV2. The detailed architecture of EfficientB0 is shown in Figure 7.

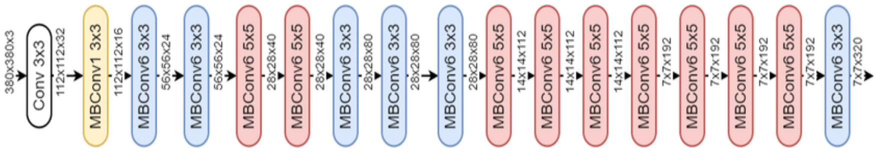

Fig. 7. The architecture of EfficientNet-B0 [24]

CNN baseline model. Convolutional Neural Networks (CNNs) were specifically created for working on image data. CNN's tries to extract the features in the image using various filters and kernels. The design is inspired by the visual cortex of the human brain. When an image is fed to the Convolutional neural net it processes the image and captures the spatial and temporal dependencies as the filters

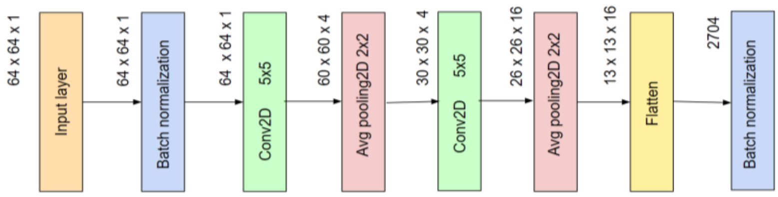

Fig. 8. Architecture of CNN baseline model

convolute across the entire image. As a baseline model, we have used a simple CNN, illustrated in Figure 7, as the embedding models for our Siamese net. The CNN consists of two 2D convolutional layers, two Average pooling layers, two batch normalization layers, and one Dense layer at the end with tanh activation function.

4 Methods

We have to classify the CXR images into one of the three classes namely Normal, COVID and Pneumonia. After initializing model parameter p and input data (pairs of images), the model is getting trained for *nepochs* with batch size N. In our proposed model, we have used EfficientNetB0 based Siamese net which takes a pair of images as input and gives the embeddings of each image which are denoted as *em1, em2*. The Euclidean distance between the two embeddings is calculated which is denoted as *dist*. The head model which is denoted as *H* receives the *dist* as input and outputs a similarity score given as *predicted_similarity* in the algorithm below. This score together with the true label for the pair of images is given to the Contrastive loss function for calculating the loss *L* and the model parameter is updated with new weights *w*.

4.1 Algorithm for training

Input: Dataset D, Batch size N, Number of epochs nepochs, fine-tuned EfficientB0 model M with parameter p, Head model H, Loss L, margin m

Initialize: img1, img2, image_pairs, true_similarity_label, predicted_similarity, dist for training

image_pairs, true_similarity_label = get_pairs(D)
$p_0 = w_0$

For i do nepochs
    for b do getBatches()
        img1, img2 = image_pairs
        em1 = M(img1)
        em2 = M(img2)
        $dist_b$ *= euclidean((em1,em2))*
        predicted_similarity = head_model($dist_b$*)*
        L_b *= ContrastiveLoss(predicted_similarity, true_similarity_label,m)*
        Update parameter p_b *with new weight, w*
    end for
end for

4.2 Evaluation metrics

In the confusion matrix of a multi class classification problem, the columns are for actual class labels and rows are for predicted class labels. From this accuracy matrix, we can calculate different performance measures such as precision, recall, f1 score and AUC.

Precision. Precision is nothing but the exactness of the classifier, The number of true positive labels divided by total positive labels. In our case for COVID-19 classification the Precision would be calculated as true positive for COVID-19 divided by the total positives for COVID-19. Total The formula for precision is given as:

$$Precision = \frac{n(True\ Positive\ labels)}{n(True\ Positive\ labels) + n(False\ Positive\ labels)} \quad (2)$$

Recall. The recall is another measure of performance which is also called sensitivity or true positive rate is the measure of classifiers completeness. It can be calculated by dividing the number of True positive labels by the sum of true positive and false negative labels. In our case for COVID-19 classification, the recall would be the true positive cases of COVID-19 divided by the same and false negative cases of COVID 19. The formula for Recall is given as :

$$Recall = \frac{n(True\ Positive\ labels)}{n(True\ Positive\ labels) + n(False\ Negative\ labels)} \quad (3)$$

F1-Score. F1 score or F score also called as the harmonic mean of Precision and Recall is a way to combine Precision and Recall. F1 score is often reliable when an uneven class distribution seeks a balance between Precision and Recall especially when there are more labels of True Negatives. F1 - score for multi class classification is we don’t compute overall F1-score, instead we calculate F1-score per class in one vs rest manner.

$$F1\ score = 2 * \frac{Precision(class{=}a) * Recall(class{=}a)}{Precision(class{=}a) + Recall(class{=}a)} \quad (4)$$

Area Under Curve (AUC). AUC is the area under the ROC curve, is that measure of the flexibility of a classifier to tell apart between classes and is employed as a summary of the ROC curve. If AUC = 1, then the classifier is in a position to discern between all the Positive and therefore the Negative class points accurately. However if AUC had been 0, then the classifier would not have been able to classify accurately and would have classified all Positives as Negatives and all Negatives as Positives.

5 Experimentation & Results

5.1 Preprocessing

The data was taken from two datsets viz. Chest Xray Pneumonia dataset and

COVID19 Radiography database avaiable on Kaggle. The resultant dataset was a balanced one. In the preprocessing step, the CXR images were first converted into grayscale format and then resized to 64 x 64 for faster processing and faster training of the model. The resized image was then normalized using rescaling by Min-Max scaler (minimum value is 0 and maximum value is 255 for every pixel) so that higher and lower valued pixels will not generate bias in the training of the model. The Dataset was found to be imbalanced which was made balanced using sampling methods.

5.2 Training

We evaluated CNN(baseline), MetaEfficientNet, VGG16 and compared their results. We trained all the three models on the benchmark datasets and observed their losses and accuracies throughout the training and testing phase. EfficienetB0 consiste of total 237 layers, based on the experimentation and literture study we chose to keep initial 120 layers as non-trainable and rest as trainable. The standard number of epochs were 8 with a batch size of 16. The train and test ratio was 80:20, 10 % of the training data was taken out for validation. We transformed the train, validation, and test sets into pairs of images before feeding the data to the models. The generated pairs are labeled either dissimilar(i.e. 0) or similar(i.e. 1). If both of the image samples belong to the same class then the label for the pair would be 1 and 0 otherwise. Taking the pairs of images as input the model was trained followed by some hyperparameter tuning for better results. CNN based Siamese neural network is our baseline model which yielded test accuracy of 79.67% with an AUC score of 0.87. Although VGG16 gave test accuracy upto 89.34 %, but its fine tuning took more, thus, EfficientNetB0 was proven to be more time efficient and also lightweight than VGG16. The EfficientNetB0 substantially outperformed the other two models, with training and testing accuracy of 98 % and 97 % respectively and AUC score of 0.97. comparison on the basis of accuracy and AUC score is presented in Table 1. Furthermore, detailed comparison of the models using various performance measures such as recall, precision and f1-score for similar and dissimilar pairs of images is given in Table 2. From Figure 9, it can be observed that the proposed model can predict the similarity scores with greater precision when compared to the baseline model.

Table 1. Performance comparison of EfficientNetB0 with VGG16 and baseline model with respect to the accuracy and AUC score

Model	Training accuracy	Testing accuracy	AUC score
CNN based Siamese model (baseline)	79.38%	79.37 %	0.8750
VGG16 based Siamese model	92.29%	89.34 %	0.9560
EfficientNet based Siamese	**98.23 %**	**97.23 %**	**0.9734**

Table 2. Performance comparison of EfficientNetB0 with VGG16 and baseline model based on Recall, Precision, F1-score for similar and dissimilar pairs of images.

Model	Recall		Precision		F1-score	
Pair label	Dissimilar	Similar	Dissimilar	Similar	Dissimilar	Similar
CNN based Siamese model (baseline)	0.68	0.91	0.88	0.74	0.77	0.82
VGG16 based Siamese model	0.92	0.92	0.92	0.92	0.92	0.92
MetaEfficient Net (Proposed)	**0.96**	**0.98**	**0.96**	**0.98**	**0.97**	**0.97**

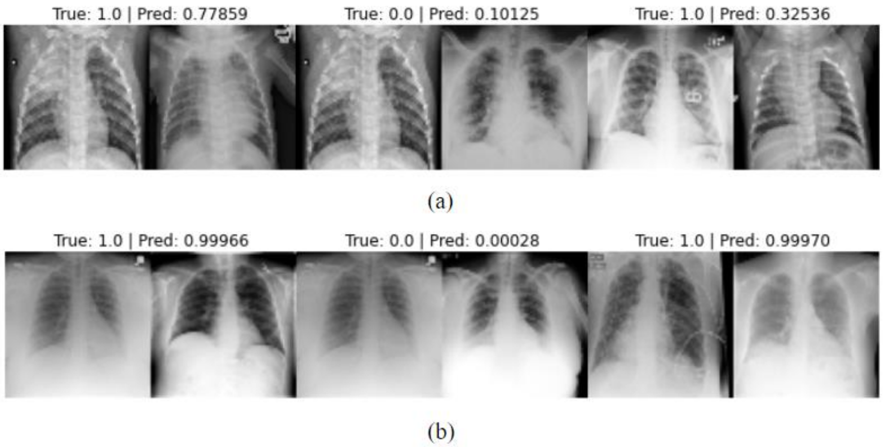

Fig. 9. Similarity score prediction, a) Results of CNN based Siamese net (baseline), b) Results of EfficienNetB0 based Siamese net (Proposed)

One of the criteria used to evaluate a model's performance is how effectively it can minimize the loss. We have used the Contrastive loss function for all the three Siamese neural nets. The CNN Siamese model could reduce the loss upto 0.1573 in training and 0.1545 in validation phase. The VGG16 performed slightly better as it was able to converge the loss upto 0.0646 in training and 0.0841 in the validation phase. However, the EfficientNetB0 model successfully minimized the training loss upto 0.0162 and validation loss upto 0.0380. According to the line graph shown in figure 10 the training loss of EfficientNetB0 appears to drop steadily throughout the training. Even though the validation loss exhibits a zig-zag line indicating little fluctuations from one epoch to another, the proposed model's validation loss appears to be minimum among the three. In contrast to EfficientNetB0, the line graph of

CNN baseline having negligible gradient did not converge unlike other models.

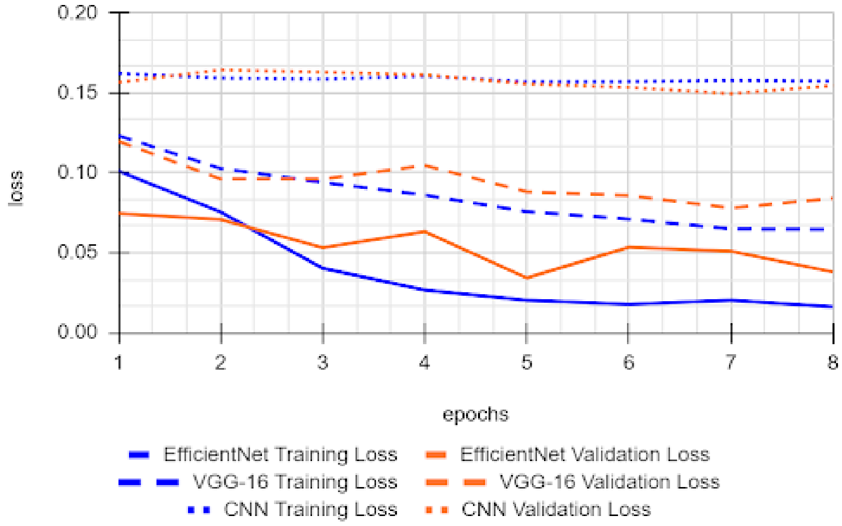

Fig. 10. Training and validation losses of CNN Siamese(baseline), VGG16 Siamese and MetaEfficientNet (Proposed)

5.3 3-way-8-Shot Learning testing with MetaEfficientNet

After fine tuning the proposed model on training data for classifying similar and dissimilar images, we took the samples from our dataset to build our support set. Based on our experimentation study, we chose a 3-way-8-shot approach to classify the query images. Support set is the set of images which is used to compare against the incoming test image or query image (Figure 11). For this study, we have used 8 images for each class of "Normal", "COVID" and "Pneumonia" in the support set. Query Set is the test set which needs to be classified. We have used the support set to gain the average similarity of the test images with each of the support images from every single class. The average similarity value of the test image with the image in the support set of each class is calculated, after which the test image is classified into the class which shows the highest similarity with the Query image.

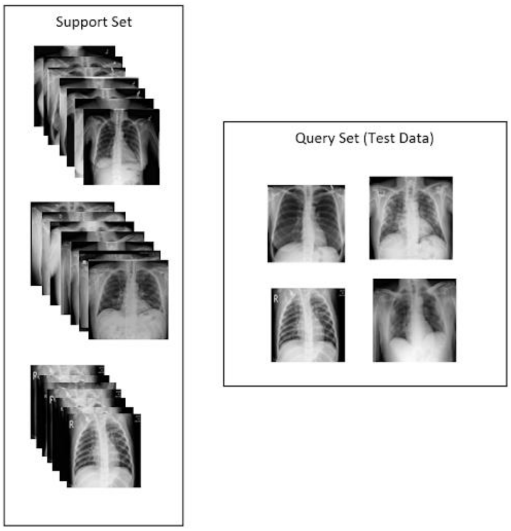

Fig. 11. Support Set and Query set. Above figure illustrates the support set which consists of 8 samples of three classes for the 3-way-8-shot learning and a query set or the test set having randomly chosen images which are to be classified.

Table 3. Comparison of Proposed model with state of the art methods

Reference no	Approach	Pre-trained model size	Accuracy	Precision	Recall	F1-score
[25]	Siamese + DenseNet121	33 MB	96.4 %	0.965	0.962	0.959
[12]	Siamese + VGG16 (MetaCovid 3-way-10-Shot)	528 MB	96.5%	0.980	0.970	0.974
MetaEfficientNet	**Siamese + EfficientNetB0 (3-way-8-Shot)**	**29 MB**	**97.23 %**	**0.980**	**0.976**	**0.976**

While dealing with medical problems, the recall which is also known as sensitivity is one the most crucial performance measure for the model. It is used for those Machine learning and deep problems where the false negative rate has to be minimum. Recall is defined as the number of positive labels predicted out of all the actual number of positive labels, therefore it is also referred to as the true positive rate. Our 3-way-8-shot model yielded an average recall of 0.97. It gave the highest recall of 0.99 for the class Viral Pneumonia, followed by COVID19 0.98, followed

Table 4. Performance comparison of MetaEfficientNet(6-way-8-shot learning) for different lung diseases

Disease	Accuracy
Cardiomegaly	0.77
Effusion	0.775
Fibrosis	0.78
İnfiltraton	0.66
Pneumothorax	0.76
Pnemonia	0.75

by Normal 0.96 and a testing accuracy of 98%. Table 3 shows the performance comparison of our robust MetaEfficientNet model for 3-way-8-shot with other state of the art methods using metrics like precision, recall and accuracy and also with respect to the sizes of the pre-trained models used. After getting a satisfactory performance over the three classes we introduced six new classes i.e. Cardiomegaly, Effusion, Fibrosis, Infilteration, Pneumothorax and Pneumonia to our MetaEfficentNet, out of which five classes were completey unknown and Pneumonia was one known class. The samples for all the new classes was taken from a different pool of CXR images called NIH Chest X-Ray dataset. The accuracies obtained for each class is listed in Table 4. We further experimented with different values of k to see if the performance of the model can be improved. Illustration of the same is shown below in Table 5.

Table 4. MetaEfficientNet accuracies for new classes

	6-way-6-shot	6-way-11-shot	6-way-13-shot
Cardiomegaly	83	77.5	96.5
Effusion	80	75.5	83.5
Fibrosis	83.5	77.5	81.5
Infiltration	76.5	76	76.5
Pneumothorax	83.5	70	85
Pneumonia	83	76.5	88

6 Conclusion

Lung diseases are still among the most fatal diseases in the world. In 2020 we saw how COVID-19 was able to bring an entire planet to a halt. Current diagnostic tools, albeit accurate, are not very efficient and cheap. Conventional methods of machine learning and deep learning have set impressive benchmarks but at the cost of the requirement of immense data. In this paper, we have successfully proposed a MetaEfficientNet model for the Chest X-ray image classification. We performed a comparative analysis between our proposed model and baseline model CNN and VGG-16, our proposed model successfully achieved an accuracy of 97%. The experimentation results revealed that the MetaEfficientNet for 3-way-8-shot learning worked better than the state-of-the-art models. Moreover, when introduced with few samples of new classes of lung diseases, the 6 way-8-shot model gave comparable accuracies and performed significantly well on the unseen data of new classes. Thus our proposed model, with just a small fraction of data is capable of producing good results. Additionally this few shot learning approach can also be extended to other categories of diseases like cancer or tumors using CT scan data.

References

1. Forum of International Respiratory Societies. The Global Impact of Respiratory Disease – Second Edition. Sheffield, European Respiratory Society, 2017.
2. Hwang, E.J.; Park, S.; Jin, K.N.; Kim, J.I.; Choi, S.Y.; Lee, J.H.; Goo, J.M.; Aum, J.; Yim, J.J.; Park, C.M. Development and Validation of a Deep Learning—based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. Clin. Infect. Dis. 2019, 69, 739–747.
3. Tobias, R.R.; De Jesus, L.C.M.; Mital, M.E.G.; Lauguico, S.C.; Guillermo, M.A.; Sybingco, E.; Bandala, A.A.;Dadios, E.P. CNN-based Deep Learning Model for Chest X-ray Health Classification Using TensorFlow. InProceedings of the 2020 RIVF International Conference on Computing and Communication Technologies, RIVF 2020, Ho Chi Minh, Vietnam, 14–15 October 2020 .
4. Ahsan, M.M.; Alam, T.E.; Trafalis, T.; Huebner, P. Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and Non-COVID-19 patients. Symmetry 2020, 12.
5. Kieu STH, Bade A, Hijazi MHA, Kolivand H. A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy, Issues and Future Directions. Journal of Imaging. 2020; 6(12):131. https://doi.org/10.3390/jimaging6120131
6. Hashmi MF, Katiyar S, Keskar AG, Bokde ND, Geem ZW. Efficient Pneumonia Detection in Chest X-RAY Images Using Deep Transfer Learning. Diagnostics (Basel). 2020;10(6):417. Published 2020 Jun 19. doi:10.3390/diagnostics10060417.
7. S. V. Militante, N. V. Dionisio and B. G. Sibbaluca, "Pneumonia Detection through Adaptive Deep Learning Models of Convolutional Neural Networks," 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, 2020, pp. 88-93, doi: 10.1109/ICSGRC49013.2020.9232613.
8. Ahsan, Mostofa & Gomes, Rahul & Denton, Anne. (2019). Application of a Convolutional Neural Network using transfer learning for tuberculosis detection. 427-433. 10.1109/EIT.2019.8833768.
9. Sudharsan Ravichandiran, Hands-On Meta Learning with Python, edited by Pavan Ramchandani, et al., Packt Publishing Ltd, December 2018, www.packtpub.com
10. Jadon, S. (2020). An Overview of Deep Learning Architectures in Few-Shot Learning Domain. ArXiv, abs/2008.06365
11. Li, M.D., Chang, K., Bearce, B. et al. Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging. npj Digit. Med. 3, 48 (2020). https://doi.org/10.1038/s41746-020-0255-1

12. Mohammad Shorfuzzaman, M. Shamim Hossain, MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients, Pattern Recognition, Volume 113, 2021, 107700, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2020.107700.
13. K. Prayogo, A. Suryadibraya, and J. Young, “Classification of pneumonia from x-ray images using siamese convolutional network,” Telkomnika (Telecommunication Computing Electronics and Control), vol. 18, no. 3, pp. 1302–1309, 2020.
14. M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M. T. Islam, “Can AI help in screening Viral and COVID-19 pneumonia?” IEEE Access, Vol. 8, 2020, pp. 132665 - 132676.
15. Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), “Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification”, Mendeley Data, V2, doi: 10.17632/rscbjbr9sj.2
16. Hands-On Meta Learning with Python Meta learning using one-shot learning, MAML, Reptile, and Meta-SGD with TensorFlow Sudharsan Ravichandiran
17. Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.
18. Sitaula, C., Hossain, M.B. Attention-based VGG-16 model for COVID-19 chest X-ray image classification. Appl Intell 51, 2850–2863 (2021). https://doi.org/10.1007/s10489-020-02055-x
19. Ibrahem Kandel, Mauro Castelli, The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset, ICT Express, Volume 6, Issue 4, 2020, Pages 312-315, ISSN 2405-9595, https://doi.org/10.1016/j.icte.2020.04.010.
20. https://neurohive.io/en/popular-networks/vgg16/
21. Tan, Mingxing, and Quoc Le. "EfficientNet: Rethinking model scaling for convolutional neural networks." In International Conference on Machine Learning, pp. 6105-6114. PMLR, 2019.
22. Gonçalo Marques, Deevyankar Agarwal, Isabel de la Torre Díez, Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network, Applied Soft Computing, Volume 96, 2020, 106691, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2020.106691.
23. Linh T. Duong, Phuong T. Nguyen, Claudio Di Sipio, Davide Di Ruscio, Automated fruit recognition using EfficientNet and MixNet, Computers and Electronics in Agriculture, Volume 171, 2020, 105326, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2020.105326.
24. Putra, Tryan & Rufaida, Syahidah & Leu, Jenq-Shiou. (2020). Enhanced Skin Condition Prediction Through Machine Learning Using Dynamic Training and Testing Augmentation. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2976045.
25. Jadon, Shruti. (2021). COVID-19 detection from scarce chest x-ray image data using few-shot deep learning approach. 1. 10.1117/12.2581496.