 Group N

GROUP N: Transforming Medical Imaging with Deep Neural Network-based Chest X-ray Analysis

1. Introduction

Medical Imaging is a critical component of modern healthcare that can aid medical professionals to make more informed diagnostic decisions. Chest X-rays are the most commonly used medical imaging modality and their interpretation can be a time-consuming, challenging, and error-prone process, even for expert radiologists to accurately diagnose diseases. The advancement of AI has triggered a significant transformation in the field of Healthcare, thus, our project aims to utilize state-of-the-art AI models, to assist in the diagnosis of respiratory and cardiovascular conditions in patients. We seek to train advanced Convolutional Neural Networks (CNNs) (AlexNet, ResNet18, and Inceptionv3), with Chest X-Ray images, to compare the performance metrics of these AI models to identify the underlying trends in the data and to diagnose abnormalities accurately.

Literature Review: Traditional Computer vision (CV) techniques have shown several limitations with respect to accuracy and performance. One of the major drawbacks is their reliance on custom or handcrafted features. In contrast, deep learning approaches such as CNNs can automatically learn relevant features directly from the image data and are superb when it comes to feature extraction.

In a study published in the journal 'Radiology', researchers compared the performance of traditional computer-vision methods and deep learning methods for the detection of lung nodules in chest CT scans. It was concluded that deep learning methods outperformed with fewer false positives and higher sensitivity [3]. Thus, CNNs for image classification tasks can bypass the need for complicated and expensive feature engineering [9].

Expected Results: The goal is to evaluate the models based on their metrics and draw insights into their performance on a particular dataset. Since the ResNet18 and Inceptionv3 have deep architectures we expect them to learn more features and perform better than the AlexNet. This can be verified by comparing the Accuracy, Precision, Recall, and F-Score of the models. However, the shallow nature of AlexNet and the lesser number of layers in ResNet18 help them train faster (training time per epoch) than Inceptionv3 and will thus be computationally less expensive.

2. Proposed Methodology

To perform disease classification from Chest X-rays, we chose the datasets publicly available from Kaggle namely, the Chest Xray Images, Chest X-Ray Dataset for Respiratory Disease Classification, NIH Chest Xray-14.

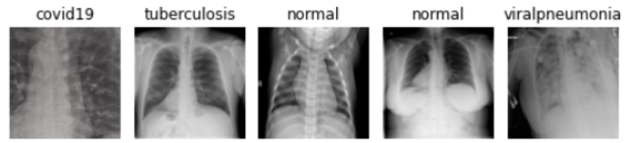

Figure 1. Random Image samples from the training set of D2

Dataset 1: This dataset consists of 5,856 chest X-ray images of patients with and without pneumonia. The data is present in a .jpeg format. It consists of 2 classes Normal (1,342) and Pneumonia (4,514), of size 1024*1024 [7].
Dataset 2: This dataset contains a total of 32,687 images with 5 classes of respiratory diseases, including COVID-19 (4,189), Lung-Opacity (6,012), Normal (10,192), Viral Pneumonia (7,397), and Tuberculosis (4,897). The data is in the form of a .npz file, which is a dictionary containing the image (224*224), image name, and image label. We have sampled 12,000 images from this dataset [1].
Dataset 3: The NIH Chest X-rays dataset is a collection of 112,120 frontal-view chest X-ray images of 30,805 unique patients, of which we sample 20,000 images. The metadata is in a .csv format and the dataset represents 15 different abnormalities diagnosed including Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, No Finding, Pleural thickening, Cardiomegaly, Nodule Mass, and Hernia [2] [8].

The models used have the following characteristics:
AlexNet: AlexNet was designed to work well on large datasets due to its ability to learn complex image features through its use of a combination of convolutional layers and max pooling layers. It has a total of 8 layers (5 convolutional and 3 fully connected) with 61,100,840 model parameters. It takes a 224*224 image as input [5].
ResNet18: ResNet18 is effective for small to medium-sized datasets and can achieve high accuracy with relatively low computational resources. It is known for its ability to overcome the vanishing gradient problem. It has 11,689,512 parameters, with 18 layers, and requires an input size of 224*224 [4].
Inceptionv3: It is a more recent architecture built to tackle the problem of computational efficiency in large datasets. It has a total of 159 layers but uses a combination of 1x1, 3x3, and 5x5 convolutions to extract the smallest of features from input images. It uses 23,851,784 model parameters and requires an input image size of 299*299 [6].

Methodology Our project uses Chest x-ray images on three datasets having a varying number of images, classes,

 Group N

and features, on the above CNN models. We first prepare the datasets and organize them to undergo pre-processing by resizing the images as per the model's input requirements, after which we convert the image to a Tensor, followed by image normalization, we achieve this by setting the mean to [0.485, 0.456, 0.406] and the standard deviation to [0.229, 0.224, 0.225] to standardize the pixel values and make the data more consistent. We then utilize the model architectures and train all three models on each of the datasets to evaluate model metrics with fixed hyperparameters. For **Dataset 1** (batch_size = 16, learning_rate = 0.001), **Dataset 2** (batch_size = 64, learning_rate = 0.0005), **Dataset 3** (batch_size = 256, learning_rate = 0.0005), and the loss-function = CrossEntropy.

To understand the impact of using pre-trained models we take the Inceptionv3 model trained on Dataset 3 and perform transfer learning onto Dataset 2, and we also use the ResNet18 model trained on Dataset 2 to transfer learn onto Dataset 1. We evaluate the metrics of all the models and compare performances across datasets.

The model's hyperparameters also contribute to its performance, to study their influence we take the AlexNet model trained on Dataset 2 and tune the model hyperparameter settings (learning_rate = 0.00005, 0.0005 0.005 and 0.05) and evaluate the performance.

3. Progress and Future Improvements

Attempts at solving the problem: Dataset 1 does not require any preparation as it is already in the required format and hierarchy. Dataset 2 contains files in a .npz format thus, we run a python script to convert each of these images into a .jpg format and also assign each of those to their respective class directories, however, we randomly take 40% of data to train the model. Dataset 3 contains 112,000 images so we run a python script to iterate over the .csv file to sample 1,300 images from each of the 15 classes to create a sample. The datasets after segregation will undergo model-specific preprocessing, using transforms by resizing the images based on the model's input requirements. We specified the sample ratio, batch size, and epochs after which, we split the dataset into a 70:20:10 ratio for train, testing, and validation sets and we finally load the dataset into batches. The development of the model was done by downloading the architecture for AlexNet and ResNet18 from PyTorch and using the specified hyperparameters, and the Adam optimizer, we train the model on Dataset 1 and Dataset 2. Post-training we run the model on the test data and plot a confusion matrix representing the model's prediction on new data.

Results: Initially, we decided to go with a low batch size of 8 for D2 which led to precision and recall of 0.5423 and 0.5521, this is due to underfitting because of using only 8 images for each epoch which is fewer than 2 images per class, thus the model does not learn enough features to classify. After this, we decided to increase the batch size to 64. After training both the model on D1 and The AlexNet on D2, the data in Table1 were obtained. It can be seen that ResNet18 performed slightly better than AlexNet on D1. Fig. 2 and Fig. 3 represent the training and validation metrics of AlexNet and ResNet18 on D1. Due to the computational complexity of the models and the lack of GPU, we were only able to train the ResNet18 on D2 for 7 epochs and the performance was tabulated. It is seen that the AlexNet outperforms at this point, however, no conclusions can be drawn as training is incomplete and the metrics steadily improving as evident from the accuracy and loss curve in 3 and thus not comparable.

	AlexNet		ResNet18	
	D1	D2	D1	D2
Accuracy	0.940	0.803	0.959	0.720
Precision	0.952	0.809	0.955	0.723
Recall	0.896	0.776	0.940	0.682
F-Score	0.919	0.786	0.947	0.688

Table 1. Evaluation Metrics of AlexNet and ResNet18 on Test data

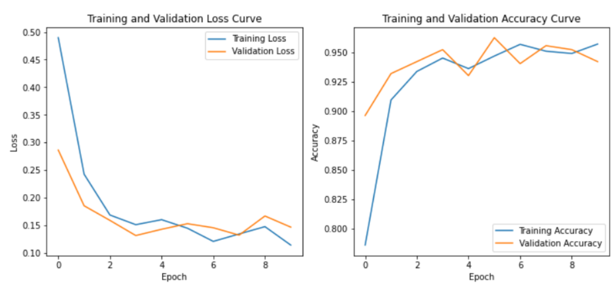

Figure 2. Training and Validation Metrics of AlexNet on D1

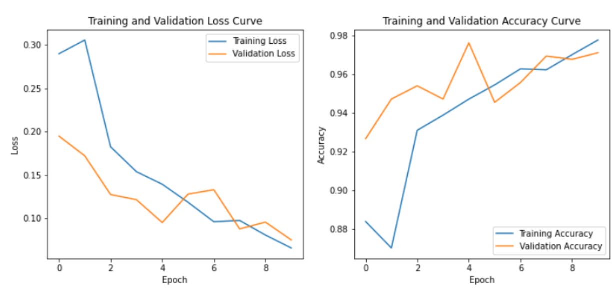

Figure 3. Training and Validation Metrics of ResNet18 on D1

Future Improvements We seek to train and evaluate the ResNet18 and AlexNet models on Dataset 3 along with developing the Inceptionv3 model followed by training it on all the datasets. We will also implement transfer learning, perform hyperparameter tuning as detailed in 2, and correspondingly evaluate the metrics for all the models. We finally compare the performances of all models across datasets to generate valuable insights.

References

[1] Arkaprabha Basu, Sourav Das, Susmita Ghosh, Sankha Mullick, Avisek Gupta, and Swagatam Das. Chest X-Ray Dataset for Respiratory Disease Classification, 2021. 1

[2] NATIONAL INSTITUTES OF HEALTH CHEST X-RAY DATASET. Nih chest x-rays — kaggle.com. https://www.kaggle.com/datasets/nih-chest-xrays/data. Accessed 10-Mar-2023. 1

[3] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014. 1

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[5] Alex Krizhevsky. One weird trick for parallelizing convolutional neural networks, 2014. 1

[6] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 1

[7] UJ-T. Chest xray images to diagnose pneumonia — kaggle.com. https://www.kaggle.com/datasets/paulti/chest-xray-images. Accessed 8-Mar-2023. 1

[8] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3462–3471, 2017. 1

[9] Samir S Yadav and Shivajirao M Jadhav. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big data*, 6(1):1–18, 2019. 1

Supplementary Material

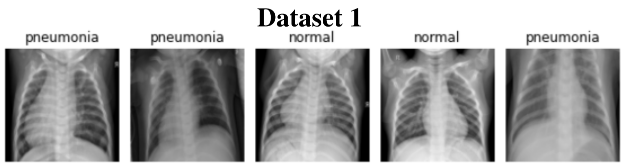

Figure 1. Random Image samples from training set of D1

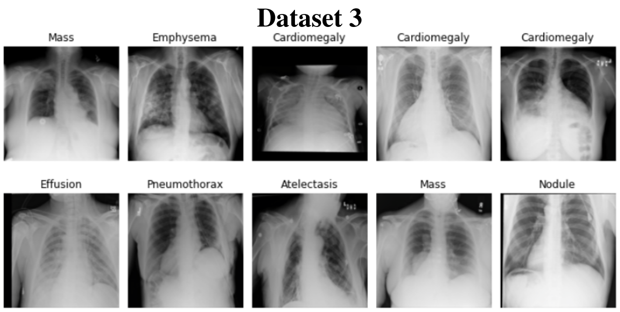

Figure 2. Random Image samples from training set of D3

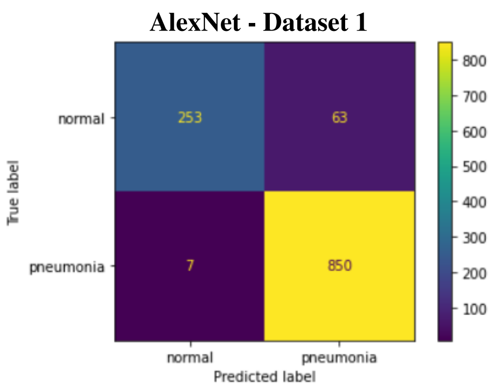

Figure 3. Confusion Matrix of AlexNet on D1-Test Data

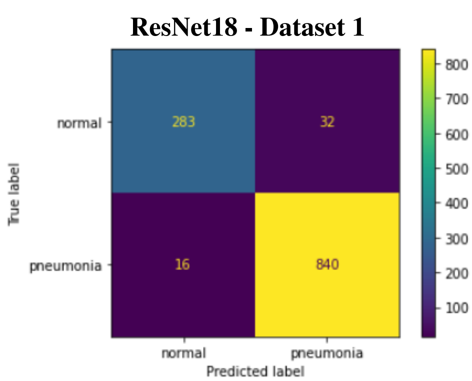

Figure 4. Confusion Matrix of ResNet18 on D1-Test Data

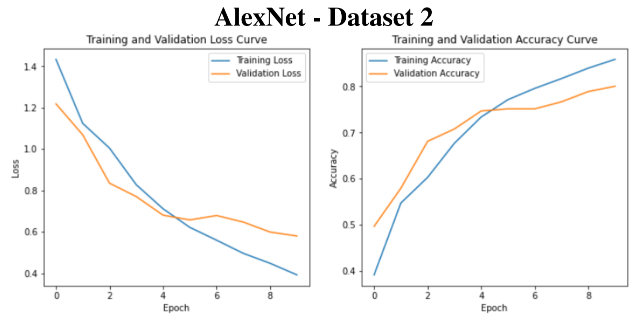

Figure 5. Training and Validation Metrics of AlexNet on D2

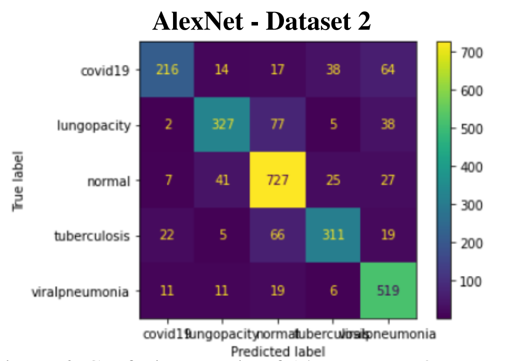

Figure 6. Confusion Matrix of AlexNet on D2-Test Data

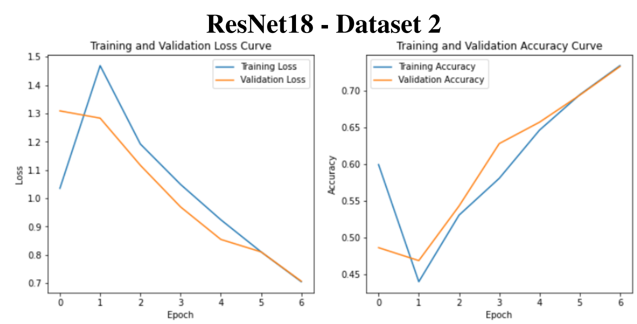

Figure 7. Training and Validation Metrics of ResNet18 on D2

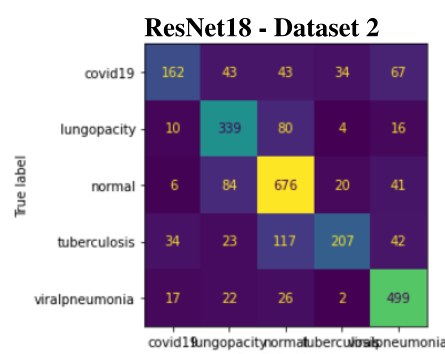

Figure 8. Confusion Matrix of ResNet18 on D2-Test Data