Enhancing Chest X-Ray Diagnosis: Transfer Learning for Multi-Class Detection of Respiratory Conditions using CNNs

Tanzim Mostafa
Idaho State University
Pocatello, Idaho 83201
Email: tanzimmostafa@isu.edu

Shijon Das
Idaho State University
Pocatello, Idaho 83201
Email: shijondas@isu.edu

***Abstract*—This research aims to enhance chest X-ray diagnosis employing convolutional neural networks (CNNs). To this end, three datasets have been chosen to carry out our experiments. The final dataset is a combination of the first two datasets, comprising five distinct classes: Tuberculosis, COVID-19, Lung Opacity (Non-Covid Lung Infection), Viral Pneumonia, and Normal. Four CNN models, namely, Xception, InceptionResnetV2, EfficientnetV2B1, and Densenet169, have been evaluated on the three datasets. Leveraging the power of transfer learning and data augmentation, the models exhibited exceptional performance, achieving accuracy scores surpassing 90% across all classes. This work contributes valuable insights into the robustness of CNNs in multi-class classification tasks, emphasizing their efficacy in the accurate and reliable prediction of various respiratory diseases. The findings hold significant implications for advancing diagnostic capabilities in the field of medical imaging and contribute to the ongoing discourse on leveraging deep learning for enhanced healthcare outcomes.**

I. Introduction

Tuberculosis (TB) remains a global health crisis, claiming the lives of 1.8 million individuals annually worldwide. Recent data from the World Health Organization (WHO) indicates a staggering 9 million new cases globally, with over 80% concentrated in South-East Asia, Western Pacific, and Africa [1]. While TB is both curable and preventable, resource-poor and marginalized communities face challenges in diagnosis due to high prevalence and limited healthcare infrastructure.

The WHO's global TB report emphasizes the need for innovative approaches, such as mHealth, to enhance the diagnosis process and promote patient adherence to medical treatment. Despite advancements in computer techniques facilitating TB diagnosis in resource-poor areas, a significant gap persists between technological progress and clinical practices. Two major barriers contribute to this gap: firstly, the absence of large-scale, well-annotated, and publicly available X-ray image databases focused on diverse TB manifestations, and secondly, the lack of high-performance computing systems for accurate diagnosis through the analysis of chest X-ray images.

Several databases, including ImageCLEF, JSRT Digital Image Database, and ANODE Grand Challenge Database, contain TB images, yet they often concentrate on specific TB manifestations, such as pulmonary nodules [2]. The utilization of computer-aided chest radiography for TB screening and diagnosis has been restricted by modest sensitivity and specificity, as well as high inter- and intra-observer differences in reporting tuberculosis shapes on radiographic images [3; 4; 5].

Coronavirus disease (COVID-19), declared a pandemic by the World Health Organization (WHO)

on March 11, 2020, is an extremely contagious global health crisis caused by a novel coronavirus [6]. The pandemic declaration emphasized concerns regarding its rapid spread and severity, marking the first recorded pandemic caused by any coronavirus.

As of June 11, 2020, there have been over 7.5 million confirmed cases globally, with approximately 421 thousand deaths, 3.8 million recoveries, 3.2 million mild cases, and 54 thousand critical cases reported [7; 8].

Efficient screening and prompt medical response are crucial in combating COVID-19. Reverse Transcription Polymerase Chain Reaction (RT-PCR) is the primary clinical screening method, utilizing respiratory specimens for testing [9]. However, RT-PCR is manual, complex, labor-intensive, and time-consuming, with a positivity rate of only 63%. Additionally, shortages in supply contribute to delays in disease prevention efforts [10].

The high cost of RT-PCR kits (USD 120-130) and the need for specialized biosafety labs further hinder widespread testing [11; 12; 13] . This expensive and delayed screening tool exacerbates the spread of the disease, affecting not only low-income countries but also certain developed nations [11].

II. Motivation

In recent years, the application of deep learning techniques in medical image analysis has demonstrated unprecedented advancements, transforming the landscape of disease diagnosis and prognosis. This paradigm shift is particularly profound in the domain of pulmonary diseases, where the intersection of artificial intelligence and radiology holds immense promise. Our motivation for this research stems from the urgent need for robust and efficient diagnostic tools for two highly prevalent respiratory diseases: tuberculosis (TB) and coronavirus disease 2019 (COVID-19).

Tuberculosis remains a major global health concern, with an estimated 10 million new cases and 1.4 million deaths reported in 2019 alone according to the World Health Organization (WHO). Traditional diagnostic methods for TB often suffer from delays and inaccuracies, contributing to the persistence of the disease. Leveraging the power of deep learning to analyze X-ray images offers a compelling avenue for enhancing the speed and accuracy of TB diagnosis, thereby facilitating timely intervention and treatment.

The emergence of the COVID-19 pandemic has further underscored the critical importance of rapid and accurate diagnostic tools. With the overwhelming burden on healthcare systems worldwide, an automated and precise method for identifying COVID-19 cases from chest X-ray images is imperative. Deep learning algorithms, endowed with the ability to discern subtle patterns and anomalies, have exhibited remarkable success in various image-based tasks and present a strategic approach to address the challenges posed by the ongoing pandemic.

III. Related Works

CXR images have been used by numerous research groups for classifying TB and non-TB cases using classical machine learning techniques [14; 15; 16; 17; 3; 18]. The parameters of deep-layered CNNs were varied for detecting tuberculosis in deep machine learning algorithms [19; 20; 21; 22; 23; 24]. Hooda et al. [19] obtained an accuracy of 82.09% while trying to classify CXR images into TB and non-TB categories using a deep learning approach. Evalgelista and Guedes [21] reported an accuracy of 88.76% using a computer-aided approach that is based on intelligent pattern recognition using CNNs for TB detection from chest X-ray images. Pasa et al. [22] reported an accuracy of 86.82% on the proposed deep network architecture which is optimized for the screening of tuberculosis. They also reported a tool for interactive visualization of TB cases. The performance of a pre-trained model, DenseNet, was evaluated by Nguyen et al. [23] for classifying normal and tuberculosis images from Shenzhen (CHN) and Montgomery County (MC) databases [3] using the fine-tuned model, and reported the Area Under the Curve (AUC) values of 0.94 and 0.82, respectively. Hernández et al. [24] proposed an ensemble of CNN models for the automatic classification of TB from X-ray images with an accuracy of 86

Recently, a large number of studies have focused on the detection of COVID-19 using X-ray images

through various AI-based techniques. Different approaches, such as transfer learning techniques, novel network architectures, and ensemble solutions, have been proposed to enhance network performances for the classification of COVID-19, normal, and other lung diseases.

Apostolopoulos et al. [25] achieved a notable accuracy of 96.78% in detecting COVID-19 from bacterial pneumonia and normal X-rays, utilizing a dataset comprising 1427 X-ray images. Similarly, Abbas et al. [26] reported an accuracy of 95.12% for classifying COVID-19, normal, and Severe Acute Respiratory Syndrome (SARS) chest X-ray images, using a pre-trained CNN model (DeTraC Decompose, Transfer, and Compose) on a modest database of 196 X-ray images. Minaee et al. [27] demonstrated specificity and sensitivity values of 90% and 97%, respectively, leveraging the ChexPert dataset [28]. Despite the modest size of the training dataset, these outcomes indicate the potential utility of deep machine-learning models in COVID-19 detection.

Khan et al. [29] explored a limited array of machine learning algorithms for a four-class classification scenario (COVID-19, bacterial pneumonia, viral pneumonia, and normal) using a notably small dataset. Goldstein et al. [30] developed a classifier for COVID-19 detection, employing a pre-trained deep learning model (ResNet50) and augmenting data through lung segmentation. Their model, trained on 2362 CXR images from four hospitals, achieved an accuracy of 89.7% and a sensitivity of 87.1%. Chowdhury et al. [31] proposed an ensemble of deep convolutional neural network (CNN) models named Efficient Convolutional Network (ECOVNet). By utilizing the transfer learning method on 16,493 CXR images, they achieved a remarkable accuracy of 97%. Ashfar et al. [32] reported an accuracy of 95.7% using a Capsule Network named COVID-CAPS, demonstrating the efficacy of unconventional CNN architectures on smaller datasets.

The use of transfer learning showed very promising results after it has been widely used in recent works for tackling such a problem.

IV. Methodology

Figure-2 shows the overall system diagram with the five class image classification problem for Dataset-C.

A. Dataset

The first Kaggle dataset, which is called Dataset-A, contains chest X-ray images for Tuberculosis(TB) positive cases along with Normal images. This dataset was created by a team of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh along with their collaborators from Malaysia in collaboration with medical doctors from Hamad Medical Corporation and Bangladesh. This dataset has 700 TB images along with 3500 normal images.

The second Kaggle dataset, which is called Dataset-B, also contains chest X-ray images for COVID-19 positive cases along with Normal, Lung Opacity which is the Non-COVID lung infection and Viral Pneumonia. There are 3616 COVID-19 positive cases along with 10,192 Normal, 6012 Lung Opacity (Non-COVID lung infection), and 1345 Viral Pneumonia chest X-ray images. This dataset too was compiled by the same group of collaborators as the first one.

The third dataset, which is called Dataset-C, is the combined dataset of the first two datasets along with the augmented images. After data augmentation, this new combined dataset contains a total of 32992 images. Data augmentation was applied to all the classes except Normal. As a result, the new combined dataset has 10692 COVID-19-positive cases along with 6012 Lung Opacity images,10192 Normal, 4014 Viral Pneumonia, and 2082 Tuberculosis images.

B. Data Preprocessing

Data preprocessing plays a key role in computer vision in enhancing the generalization capabilities of models and their performances. For dataaugmentation, roboflow has been used. Auto-Orient ensures a uniform orientation across all images. This uniform orientation eliminates potential biases introduced by variations in how the images were originally captured. This preprocessing creates a

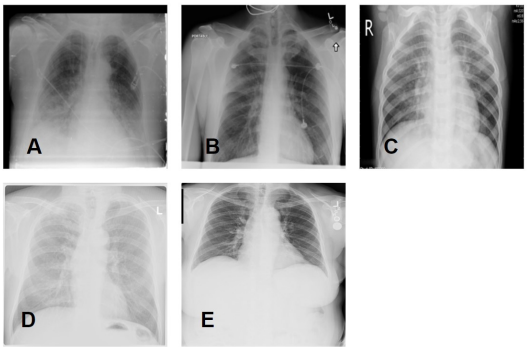

Fig. 1. Sample X-ray images from the dataset. COVID-19 positive X-Ray image(A), Lung Opacity X-Ray image(B), Viral Pneumonia X-Ray image(C), Tuberculosis X-Ray image(D), and Normal X-Ray image(E)

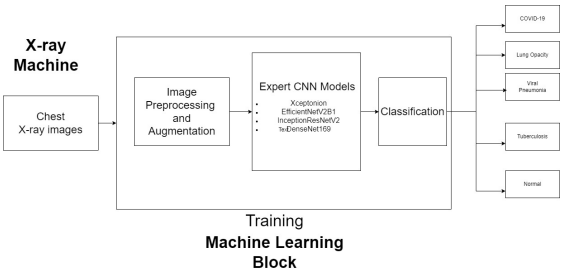

Fig. 2. Block diagram of the overall system.

more consistent and unbiased dataset which helps facilitates the robust training of the models.

Additionally, the augmentation techniques applied contributed significantly in enhancing the diversity of the dataset.. ”Horizontal Flip” introduces mirror images, aiding the model in learning invariant features regardless of object orientation. The ”Rotation” augmentation introduces variability in object angles, promoting the development of rotational invariance. These augmentations collectively enable the model to generalize better to unseen orientations. The adjustments in ”Hue” and ”Brightness” bring about variations in color and illumination, mitigating the impact of different lighting conditions. This is vital for the model to adapt to real-world scenarios where lighting may vary. Lastly, the ”Blur” augmentation, with a subtle 1-pixel blur, helps the model become less sensitive to fine details, fostering a degree of tolerance towards noise or slight blurriness in test images.

In essence, the carefully curated preprocessing and augmentations contribute to a more robust, diverse, and generalized dataset. These measures not only addresses the potential biases and inconsistencies but also equip the model to handle a spectrum of real-world conditions, enhancing its practical utility and performance in deployment scenarios. The detailed augmentation strategy ensures that the model learns features invariant to various transformations, ultimately contributing to its reliability and efficacy in scientific applications.

C. Model Specification

This section gives an overview of the models used in our experiment. Four CNN Models have been applied to Dataset-A, Dataset-B and Dataset-C, which are Xception, InceptionResNetV2, EfficientnetV2, and DenseNet.

1) Xception: Xception, an abbreviation for ”Extreme Inception,” is another deep learning architecture that builds on the inception module. It replaces the standard convolutional layers with depthwise separable convolutions. In Xception, each convolutional operation is split into depthwise and pointwise convolutions. This separation helps the model capture spatial and channel-wise dependencies more efficiently. Xception is known for its ability to achieve competitive performance with lower computational cost compared to traditional architectures.

2) InceptionResNetV2: InceptionResNetV2 is a deep neural network architecture that combines the concepts of both Inception and ResNet. It incorporates residual connections and inception modules to achieve a highly efficient and accurate model. The architecture includes multiple inception blocks, each comprising parallel convolutional layers with different kernel sizes. Residual connections are added to allow for smooth gradient flow during training. The model also utilizes batch normalization and dropout for regularization. The final layers typically involve a global average pooling layer followed by fully connected layers for classification.

3) EfficientnetV2: EfficientNetV2 is an advanced neural network architecture designed for efficient and scalable deep learning models. Building upon the success of its predecessor, EfficientNet, this version introduces several key improvements to enhance model performance and reduce computational demands. EfficientNetV2 employs a novel compound scaling method that optimizes the scaling coefficients for depth, width, and resolution across different layers of the network. This allows for more efficient use of resources, enabling the creation of powerful models with fewer parameters. The architecture also incorporates a new attention mechanism, helping the network focus on relevant information and improving its ability to capture complex patterns in data. EfficientNetV2 is particularly well-suited for tasks requiring high computational efficiency, making it a popular choice in various fields, including computer vision and natural language processing

4) DenseNet: DenseNet, or Densely Connected Convolutional Networks, introduces the concept of dense blocks, where each layer receives direct input from all preceding layers. This densely connected structure encourages feature reuse, reduces the number of parameters, and enhances the gradient flow through the network. DenseNet comprises densely connected blocks, each containing multiple convolutional layers. Transition layers, including batch normalization and pooling, are used to control the spatial dimensions. The final layers usually involve global average pooling and fully connected layers for classification.

D. Implementations

This section describes the implementation of the 4 models, namely Xception, InceptionResNet V2, EfficientNet V2, and DenseNet on each of the three datasets. To facilitate a performance comparison, we maintained consistency in the primary aspects of implementation across all four instances. Our implementation leveraged Tensorflow and Keras, executed within Google Colaboratory notebooks. The adoption of transfer learning, a widely recognized technique in deep learning, was employed. This involves utilizing pre-trained versions of the models as the starting point for another related task. Transfer learning has been shown to help improve the performance of the models and is especially useful given the huge compute resources required to train deep learning models.

The Xception model contains approximately 22.9 million parameters. Before being fed into our Xception network, the images were resized to 71x71 pixels. The InceptionResNet V2 model consists of around 55.9 million parameters. Before being fed into the InceptionResNet model, the input pixels had to be scaled between -1 and 1, and the images were resized to 96x96 pixels. The EfficientNet V2 B1 version has been used, which consists of 8.2 million parameters. For this case, the images input to the model had to be resized to 224x224 pixels. Finally, the DenseNet169 version has been implemented, containing 14.3 million parameters. For DenseNet169, the images were resized to 96x96 pixels before being fed into the model.

Some of the implementations constant across all four models are described as follows. Each of the four models have been pre-trained on ImageNet. They have been configured to exclude the top classification layer, allowing for customization based on our specific task. The four architectures include global average pooling to reduce the spatial dimensions of the feature maps and mitigate overfitting. A dropout layer with a rate of 0.05 has been added to enhance the model's generalization ability by randomly deactivating a small fraction of neurons during training. For our first dataset, a dense layer with a single neuron and a sigmoid activation function is used as the last layer for all 4 models, to produce binary classification outputs. For the second dataset, the final layer is a dense layer consisting of 4 neurons, while for the last dataset, the final layer is changed to a dense layer of 5 neurons to make the 5 classifications. For the last two datasets, softmax activation has been employed across the 4 models. This architecture leverages transfer learning, capitalizing on the knowledge encoded in the pre-trained models, while tailoring the final layers for the specific classification task at hand. For the first dataset, binary cross entropy has been used as the loss function, since the goal

is to classify instances into one of two classes. For the second and third dataset, sparse categorical cross entropy is employed, since they are multi-class classification tasks. The models are then compiled using the Nadam optimizer with a learning rate of 1e-3. The choice of optimizer and loss function is pivotal for training neural networks, and Nadam is an optimizer that combines the benefits of Nesterov momentum and Adam optimization. Some of the other hyperparameters used include batch size of 32, and an early stopping callback with patience of 5, which halts training if the validation performance does not improve for a 5 consecutive epochs. The models are then restored to their best weights to ensure optimal performance. During the initial training phase, the pre-trained layers are frozen, to feature as extractors. These layers have learned hierarchical features from ImageNet, and by keeping them frozen, we preserve these learned representations. After that, each of the models are trained for a maximum of 100 epochs, by unfreezing the pre-trained layers.

V. Performance Evaluation Metrics

Four metrics have been used to evaluate the performance of the CNN models: Accuracy, Precision, Recall, and F1 Score.

1. **Accuracy:** Accuracy is a measure of the overall correctness of a classification model.

$$Accuracy = \frac{NumberofCorrectPredictions}{TotalNumberofPredictions} \tag{1}$$

2. **Precision:** Precision focuses on the accuracy of positive predictions, avoiding false positives.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{2}$$

3. **Recall (Sensitivity or True Positive Rate):** Recall measures the ability of a model to capture all relevant positive instances.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{3}$$

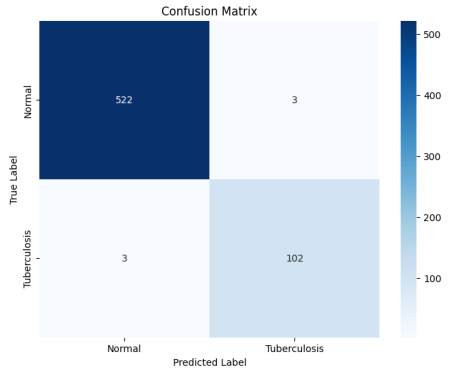

Fig. 3. Confusion matrix for Xception on test set of Dataset A

4. **F1 Score:** The F1 score is the harmonic mean of precision and recall, providing a balance between the two metrics.

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

VI. Results and Discussion

A. Dataset A

1) Xception: The Xception model achieved overall accuracy of 99.05%. Moreover, it's precision is 0.9714, recall is 0.9714, and F1 score is 0.9714. Overall, the Xception model performed very well regarding it's scores on all the metrics.

Figure 3 shows the confusion matrix of the Xception model on the test set of Dataset A. We can observe that our model has correctly classified 522 Normal images and 102 Tuberculosis images. On the other hand, it incorrectly predicted 3 Normal cases as Tuberculosis and 3 Tuberculosis images as Normal.

2) InceptionResNetV2: InceptionResNetV2 achieved overall accuracy of 98.89%. Also, it's precision is 0.9900, recall is 0.9429, and F1 score is 0.9659. Overall, the model also performed good scores on each of the metrics.

Figure 4 shows the confusion matrix of the InceptionResNetV2 model on the Dataset A's test set. We can observe that our model has correctly classified 524 Normal images and 99 Tuberculosis images. On the other hand, it incorrectly predicted

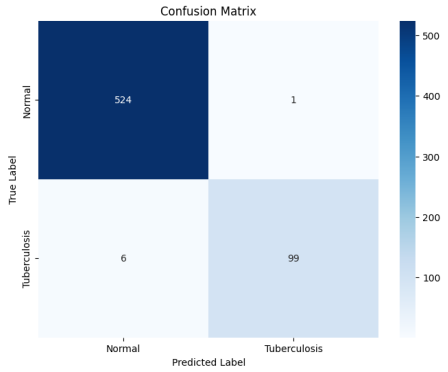

Fig. 4. Confusion matrix for InceptionResNetV2 on test set of Dataset A

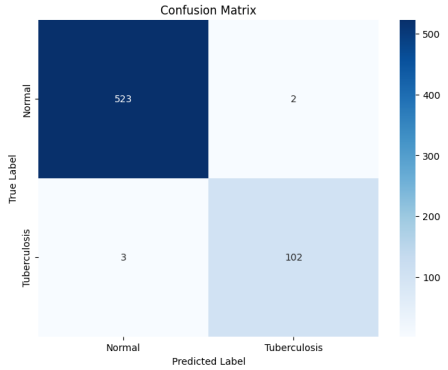

Fig. 5. Confusion matrix for EfficientNetV2B1 on test set of Dataset A

1 Normal case as Tuberculosis and 6 Tuberculosis cases as Normal.

3) EfficientNetV2B1: EfficientNetV2B1 obtained overall accuracy of 99.21%. It's precision is 0.9808, recall is 0.9714, and F1 score is 0.9761. This model also has good scores for the four metrics.

Figure 5 shows the confusion matrix of the EfficientNetV2B1 model on the Dataset A's test set. We can observe that it correctly classified 523 Normal images and 102 Tuberculosis images. However, it incorrectly classified 2 Normal case as Tuberculosis and 3 Tuberculosis images as Normal.

4) DenseNet169: DenseNet169 obtained an overall accuracy of 99.68%, precision of 0.9905, recall of 0.9905 and f-1 score of 0.9905. These values show that this model is also performing well on the dataset.

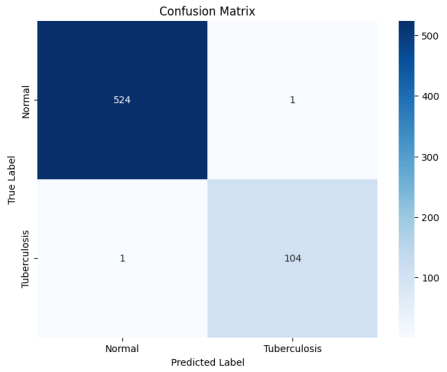

Fig. 6. Confusion matrix for DenseNet169 on test set of Dataset A

Figure 6 shows the confusion matrix of the DenseNet169 model on the test set of Dataset A. We can observe that it correctly classified 524 Normal images and 104 Tuberculosis images. However, it incorrectly classified 1 Normal and 1 Tuberculosis class.

5) Comparison: Table I shows the results for all the models on each of the four metrics. We can observe that while all the models perform well on the dataset, DenseNet169 has the highest overall accuracy, precision, recall, and F-1 score.

TABLE I
RESULTS FOR DATASET A

Model	Accuracy (%)	Precision	Recall	F1
Xception	99.05	0.9714	0.9714	0.9714
InceptionResNetV2	98.89	0.9900	0.9429	0.9659
EfficientNetV2	99.21	0.9808	0.9714	0.9761
DenseNet169	99.68	0.9905	0.9905	0.9905

B. Dataset B

1) Xception: Xception model achieved overall Accuracy of 91.21%. The table II presents the performance metrics of the Xception model for a multi-class classification task across different classes. The columns display key evaluation metrics, such as precision, recall, and F1-score, which provide insights into the model's classification accuracy, ability to

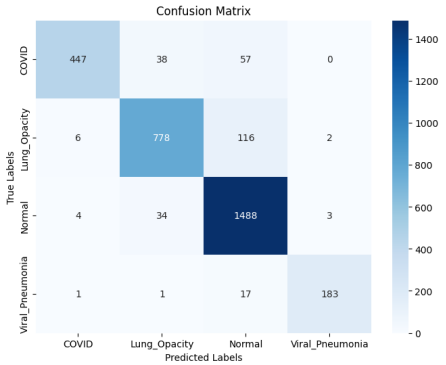

Fig. 7. Confusion matrix for Xception on test set of Dataset B

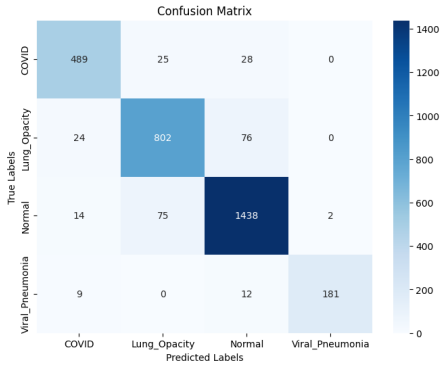

Fig. 8. Confusion matrix for InceptionResNetV2 on test set of Dataset B

identify relevant instances, and overall balanced performance. For instance, the Xception model demonstrates high precision for the ”COVID” and ”Viral Pneumonia” classes, indicating a low rate of false positives. The recall values, representing the model's ability to capture true positives, vary across classes, with the ”Normal” class achieving particularly high recall. The F1-score, a harmonic mean of precision and recall, offers a balanced measure of the model's overall effectiveness. The table serves as a concise summary of the Xception model's classification performance, offering a detailed assessment of its strengths and areas for improvement across various classes in the multiclass classification problem.

Figure 7 shows the confusion matrix for Xception on Dataset B, which provides a detailed breakdown of the model's predictions, highlighting areas of correct classification and instances of misclassification across the four classes.

TABLE II
CLASSIFICATION REPORT OF XCEPTION ON DATASET B

Model	Precision	Recall	F1-Score
COVID	0.98	0.82	0.89
Lung Opacity	0.91	0.86	0.89
Normal	0.89	0.97	0.93
Viral Pneumonia	0.97	0.91	0.94

2) InceptionResNetV2: InceptionResNetV2 obtained 91.65% overall accuracy on Dataset B. The table II, shows the classification report of InceptionResNetV2 on Dataset B. The InceptionResNet model demonstrated strong performance across multiple key metrics, as evidenced by the provided classification results. It achieved high precision, recall, and F1-score values for the diverse classes within the dataset. Specifically, the model showcased an impressive precision of 0.99 for the ”Viral Pneumonia” class, indicating its ability to accurately identify instances of this category.

Figure 8 shows the confusion matrix for InceptionResNetV2 on Dataset B.

TABLE III
CLASSIFICATION REPORT OF INCEPTIONRESNETV2 ON DATASET B

Model	Precision	Recall	F1-Score
COVID	0.91	0.90	0.91
Lung_Opacity	0.89	0.89	0.89
Normal	0.93	0.94	0.93
Viral_Pneumonia	0.99	0.90	0.94

3) EfficientNetV2B1: The model EfficientNetV2B1 achieved 92.35% accuracy on Dataset B. EfficientNetV2B1, as applied to the multiclass classification task, demonstrated strong performance across various metrics for each class, as can be observed from table IV

From Figure 9 we can observe the confusion matrix for EfficientNetV2B1 on Dataset B.

4) DenseNet169: DenseNet169 achieved overall accuracy of 92.44% on Dataset B. DenseNet169,

TABLE IV
CLASSIFICATION REPORT OF EFFICIENTNETV2B1 ON DATASET B

Model	Precision	Recall	F1-Score
COVID	0.94	0.96	0.95
Lung_Opacity	0.90	0.88	0.89
Normal	0.92	0.94	0.93
Viral_Pneumonia	0.98	0.89	0.93

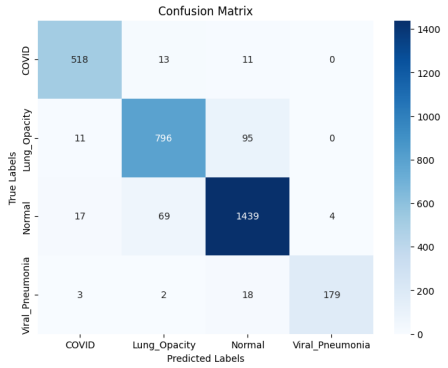

Fig. 9. Confusion matrix for EfficientNetV2B1 on test set of Dataset B

also demonstrated strong performance across various metrics for each class, as can be observed from table V

TABLE V
CLASSIFICATION REPORT OF DENSENET169 ON DATASET B

Model	Precision	Recall	F1-Score
COVID	0.95	0.96	0.95
Lung Opacity	0.87	0.91	0.89
Normal	0.94	0.92	0.93
Viral Pneumonia	0.94	0.97	0.95

Figure 10 provides the confusion matrix for DenseNet169 on Dataset B.

5) Comparison: Overall, all the models performed similarly well. But, based on accuracy values, DenseNet169 achieved the highest accuracy of 92.44%.

C. Dataset C

1) Xception: Xception model achieved overall Accuracy of 95.15%. The table VI presents the performance metrics of the Xception model for a five class classification task across different classes.

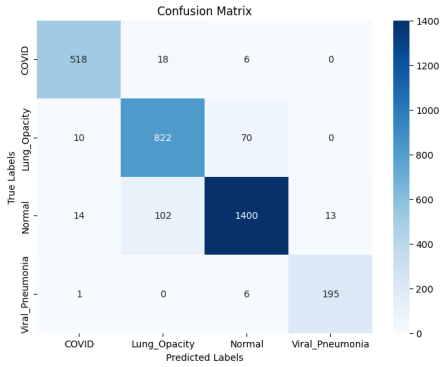

Fig. 10. Confusion matrix for DenseNet169 on test set of Dataset B

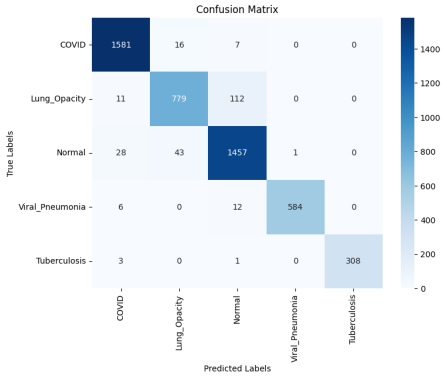

Fig. 11. Confusion matrix for Xception on test set of Dataset C

Figure 11 shows the confusion matrix for Xception on Dataset C, which provides a detailed breakdown of the model's predictions, concerning areas of correct classification as well as misclassification across the five classes.

TABLE VI
CLASSIFICATION REPORT OF XCEPTION ON DATASET C

Model	Precision	Recall	F1-Score
COVID	0.97	0.99	0.98
Lung Opacity	0.93	0.86	0.90
Normal	0.92	0.95	0.93
Viral Pneumonia	1.00	0.97	0.98
Tuberculosis	1.00	0.99	0.99

2) InceptionResNetV2: InceptionResNetV2 obtained 94.93% overall accuracy on Dataset C. The

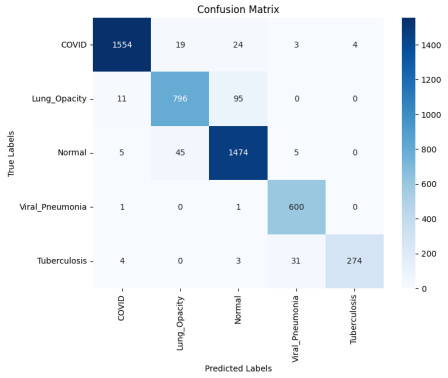

Fig. 12. Confusion matrix for InceptionResNetV2 on test set of Dataset C

table VI, shows the classification report of InceptionResNetV2 on Dataset C. Figure 12 shows the confusion matrix for InceptionResNetV2 on Dataset C.

TABLE VII
CLASSIFICATION REPORT OF INCEPTIONRESNETV2 ON DATASET C

Model	Precision	Recall	F1-Score
COVID	0.99	0.97	0.98
Lung Opacity	0.93	0.88	0.90
Normal	0.92	0.96	0.94
Viral Pneumonia	0.94	1.00	0.97
Tuberculosis	0.99	0.88	0.93

3) EfficientNetV2B1: The model EfficientNetV2B1 achieved 95.82% accuracy on Dataset C. EfficientNetV2B1, as applied to the multiclass classification task, demonstrated strong performance across various metrics for each class, as can be observed from table VIII

From Figure 13 we can observe the confusion matrix for EfficientNetV2B1 on Dataset B.

4) DenseNet169: DenseNet169 achieved overall accuracy of 95.31% on Dataset C. DenseNet169, also demonstrated strong performance across various metrics for each class, as can be observed from table IX

Figure 14 provides the confusion matrix for DenseNet169 on Dataset C.

5) Comparison: For Dataset C, overall all the models performed similarly well. However, based on accuracy EfficientNetV2B1 achieved the best performance of 95.82%.

TABLE VIII
CLASSIFICATION REPORT OF EFFICIENTNETV2B1 ON DATASET C

Model	Precision	Recall	F1-Score
COVID	0.99	0.99	0.99
Lung Opacity	0.93	0.89	0.91
Normal	0.93	0.95	0.94
Viral Pneumonia	0.98	0.99	0.98
Tuberculosis	0.98	0.99	0.98

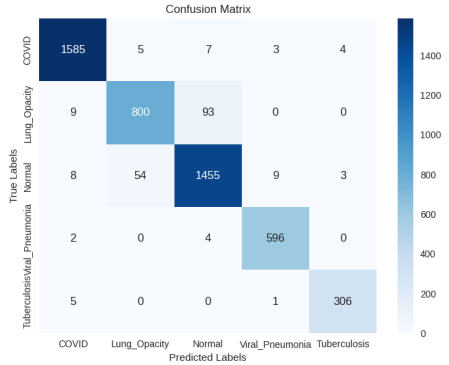

Fig. 13. Confusion matrix for EfficientNetV2B1 on test set of Dataset C

D. Comparison Accross Three Datasets

Table X shows the overall accuracy of each of the models on the three datasets. We can observe DenseNet169 is performing the best for the binary classification task on Dataset A, and four class classification in Dataset B. However, EfficientNetV2B1 has the highest overall accuracy for the combined dataset C containing five classes.

Another thing we can observe is that, after increasing the size of the dataset through various

TABLE IX
CLASSIFICATION REPORT OF DENSENET169 ON DATASET C

Model	Precision	Recall	F1-Score
COVID	0.99	0.98	0.98
Lung Opacity	0.87	0.92	0.90
Normal	0.94	0.92	0.93
Viral Pneumonia	0.99	0.99	0.99
Tuberculosis	1.00	0.98	0.99

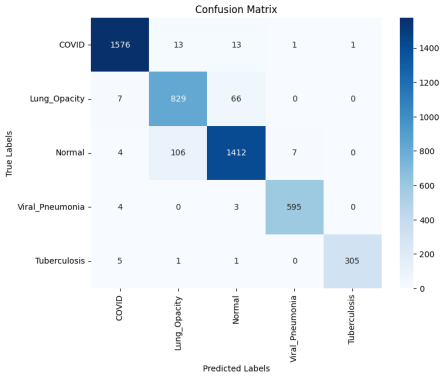

Fig. 14. Confusion matrix for DenseNet169 on test set of Dataset C

data augmentation techniques, overall all the models achieved better accuracy for Dataset C than Dataset B.

TABLE X
MODEL ACCURACY ON DATASET A, DATASET B, AND DATASET C

Model	**A (%)**	**B (%)**	**C (%)**
Xception	99.05	91.21	95.15
InceptionResNetV2	98.89	91.65	94.93
EfficientNetV2B1	99.21	92.35	95.82
DenseNet169	99.68	92.44	95.31

VII. CONCLUSION

In conlusion, for this research we explored the detection of COVID-19, Viral Pneumonia, Lung Opacity and Tuberculosis from Chest X-Ray images using CNNs. We used three datasets for carrying out our experiments: dataset A, containing Tubercolis and Normal images, dataset B containing Covid, Viral Pneumonia, Lung Opacity, and Normal images, and finally dataset C, which is a combination of the first two datasets. Additionally, several data augmentaion techniques have been applied to dataset C. Four CNN models, namely, Xception, InceptionResNetV2, EfficientNetV2B1, and DenseNet169 have been trained and evaluated on the three datasets, taking advantage of transfer learning. It was found out that DenseNet169 performed best on dataset A and dataset B, with accuracy scores of 99.68% and 92.44%, respectively. On the other hand, EfficientNetV2B1 performed the best on dataset C, obtaining accuracy of 95.82%. It has also been obserrved that there was a significant increase in accuracy for all models on dataset C, compared to dataset B, which can be attributed to the data augmentation techniques that have been employed on dataset C.

REFERENCES

[1] World Health Organization et al., "Global tuberculosis report 2016," 2016.

[2] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, and et al., "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.

[3] S. Jaeger, A. Karargyris, S. Antani, and G. Thoma, "Detecting tuberculosis in radiographs using combined lung masks," in *Engineering in Medicine and Biology Society*. IEEE, 2012.

[4] T. Xu, I. Cheng, R. Long, and M. Mandal, "Novel coarse-to-fine dual scale technique for tuberculosis cavity detection in chest radiographs," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, p. 3, 2013.

[5] Y.-L. Song and Y. Yang, "Localization algorithm and implementation for focal of pulmonary tuberculosis chest image," in *Machine Vision and Human-Machine Interface*. IEEE, 2010.

[6] W. H. Organization. (2020) Who director-general's opening remarks at the media briefing on covid-19–11 march 2020. [Online]. Available: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020

[7] ——, *Global COVID-19 Report*, Geneva, Switzerland, 2020.

[8] Johns Hopkins University (JHU) Medicine. (2020) Coronavirus covid-19 global cases by the center for systems science

and engineering (csse) at johns hopkins university (jhu). [Online]. Available: https://coronavirus.jhu.edu/map.html

[9] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, and W. Tan, "Detection of SARS-CoV-2 in different types of clinical specimens," *JAMA*, vol. 323, no. 18, pp. 1843–1844, 2020.

[10] T. Yang, Y.-C. Wang, C.-F. Shen, and C.-M. Cheng, "Point-of-care RNA-Based Diagnostic Device for COVID-19," *MDPI Sensors*, vol. 20, no. 10, 2020.

[11] Al Jazeera News. (2020) India's poor testing rate may have masked coronavirus cases. [Online]. Available: https://www.aljazeera.com/news/2020/03/india-poor-testing-rate-masked-coronavirus-cases-200318040314568.html

[12] ——. (2020) Bangladesh scientists create $3 kit. can it help detect covid-19. [Online]. Available: https://www.aljazeera.com/news/bangladesh-scientists-create-3-kit-detect-covid-19-200323035631025.html

[13] N. Wetsman. (2020) Coronavirus testing shouldn't be this complicated. [Online]. Available: https://www.theverge.com/coronavirus-testing-pcr-diagnostic-point-of-care-techonology

[14] H. Das and A. Nath, "An efficient detection of tuberculosis from the chest x-rays," *Int. J. Adv. Res. Comput. Sci. Manage. Studies*, vol. 3, no. 5, pp. 149–154, May 2015.

[15] J. Melendez, C. I. Sánchez, R. H. H. M. Philipsen, P. Maduskar, R. Dawson, G. Theron, K. Dheda, and B. van Ginneken, "An automated tuberculosis screening strategy combining x-ray-based computer-aided detection and clinical information," *Sci. Rep.*, vol. 6, no. 1, p. 25265, Jul 2016.

[16] N. Singh and S. Hamde, "Tuberculosis detection using shape and texture features of chest x-rays," in *Innovations in Electronics and Communication Engineering (Lecture Notes in Networks and Systems)*, H. Saini, R. Singh, G. Kumar, G. Rather, and K. Santhi, Eds., 2019, vol. 65.

[17] B. van Ginneken, S. Katsuragawa, B. M. ter Haar Romeny, K. Doi, and M. A. Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 139–149, 2nd Quart. 2002.

[18] J. Melendez, C. I. Sánchez, R. H. Philipsen, P. Maduskar, and B. van Ginneken, "Multiple-instance learning for computer-aided detection of tuberculosis," *Proc. SPIE*, vol. 9035, Mar 2014.

[19] R. Hooda, S. Sofat, S. Kaur, A. Mittal, and F. Meriaudeau, "Deep-learning: A potential method for tuberculosis detection using chest radiography," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep 2017, pp. 497–502.

[20] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, Aug 2017.

[21] L. G. C. Evalgelista and E. B. Guedes, "Computer-aided tuberculosis detection from chest x-ray images with convolutional neural networks," in *Proc. Anais do XV Encontro Nacional de Inteligência Artif. e Computacional (ENIAC)*, Oct 2018, pp. 518–527.

[22] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient deep network architectures for fast chest x-ray tuberculosis screening and visualization," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec 2019.

[23] Q. H. Nguyen, B. P. Nguyen, S. D. Dao, B. Unnikrishnan, R. Dhingra, S. R. Ravichandran, S. Satpathy, P. N. Raja, and M. C. H. Chua, "Deep learning models for tuberculosis detection from chest x-ray images," in *Proc. 26th Int. Conf. Telecommun. (ICT)*, Apr 2019, pp. 381–386.

[24] A. Hernández, Á. Panizo, and D. Camacho, "An ensemble algorithm based on deep learning for tuberculosis classification," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.*, Manchester, U.K. 2019, pp. 145–154.

[25] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physics and Engineering Sciences in Medicine*, vol. 43, no. 3, pp. 635–640, 2020.

[26] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network," *Applied Intelligence*, vol. 51, pp. 854–864, 2021.

[27] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour Soufi, "Deep-covid: predicting covid-19 from chest x-ray images using deep transfer learning," *Medical Image Analysis*, vol. 65, p. 101794, 2020.

[28] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute *et al.*, "Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 590–597.

[29] A. I. Khan, J. I. Shah, and M. M. Bhat, "Coronet: a deep neural network for detection and diagnosis of covid-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine*, p. 105581, 2020.

[30] E. Goldstein, D. Keidar, D. Yaron, G. Shachar, A. Blass, L. Charbinsky *et al.*, "Covid-19 classification of x-ray images using deep neural networks," 2020.

[31] N. K. Chowdhury, M. M. Rahman, N. Rezoana, M. A. Kabir *et al.*, "Ecovnet: an ensemble of deep convolutional neural networks based on efficientnet to detect covid-19 from chest x-rays," 2020.

[32] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "Covid-caps: a capsule network-based framework for identification of covid-19 cases from x-ray images," *Pattern Recognition Letters*, vol. 138, pp. 638–643, 2020.