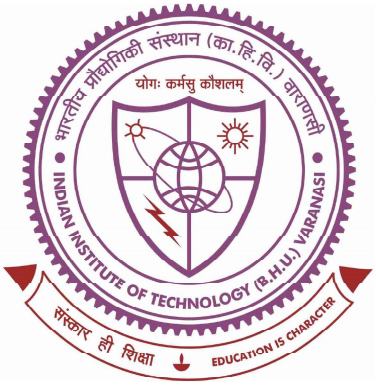
Transfer learning based multiclass classification of Covid-19 using chest X-Rays and weakly supervised model for cardiac MRI segmentation



The thesis submitted in partial fulfillment

for the award of degree

DOCTOR OF PHILOSOPHY

by

ALOK TIWARI

SCHOOL OF BIOMEDICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY

(BANARAS HINDU UNIVERSITY)

VARANASI-221005

Roll No.: 16021007

November 2023

List of Research Publications

---

[1] ***Tiwari, Alok, et al.***, "Deep learning-based automated multiclass classification of chest X-rays into Covid-19, normal, bacterial pneumonia and viral pneumonia." ***Cogent Engineering 9.1 (2022): 2105559***.

[2] ***Tiwari, Alok et al.*** , 'Detection of COVID-19 Infection in CT and X-ray Images Using Transfer Learning Approach'. ***Technology and Health Care***, vol. 30, no. 6, pp. 1273-1286, 2022

[3] ***Sharan, Tiwari Alok et al.***. "Cascaded Model (Conventional + Deep Learning) for Weakly Supervised Segmentation of Left Ventricle in Cardiac Magnetic Resonance Images" ***IETE Technical Review (2022)*** 0256-4602, 0974-5971

[4] ***Tiwari, Alok et al.***, "Application of Neuroimaging Tools in Identification of Pinpoint Location of Blockage", ***Springer Singapore (2019)*** , DOI: 10.1007/978-981-13-1453-7_7

Review Report 1 - Examiner - 1

Comments:

1. **It would be interesting to understand the quality of images in the database. Were the images all looking ideal, normalized wrt some parameters? Did the author carry out any quality check on the images and any pre-processing steps were used before using them for deep learning?**

Response:

The quality of images varies across the Chest X-Ray, ACDC, and SunnyBrook databases, each serving specific purposes in medical imaging research.

1. The ChestX-ray Dataset: Dataset has extensive pre-processing to avoid class imbalance due to insufficient sample images from one particular class. So, to achieve that, we have applied the data augmentation method to increase the number of sample images and make it a uniform distribution. In order to perform data augmentation, we have tried ImageDataGenerator module from Keras library by tuning various data augmentation parameters like rescale = 1/255, rotation_range = 20, width_shift_range = 0.2, height_shift_range = 0.2, and, also, we set horizontal_flip = True, as in chest X Ray, we are mainly focusing about the vertical area, so horizontal flip does not make any such difference and hence does not lose much information. In order to remove duplicate images, we searched for exact duplicates or near exact duplicates with the help of Pillow library in python. With the help of hashing, we successfully removed these kinds of duplicates. We have also selected the same number of images from each class to avoid overfitting or underfitting to one particular class. The network remains unbiased to any specific class when performing classification tasks.

2. ACDC Database: The ACDC (Automated Cardiac Diagnosis Challenge) database provides high-quality cardiac MRI images with fine pixel-level panoptic semantic annotations for adverse conditions. The emphasis is on detailed annotations and controlled imaging conditions .

3. SunnyBrook Database: The SunnyBrook Cardiac Data (SCD) contains cardiac cine-MRI images used in the 2009 Cardiac MR Left Ventricle Segmentation Challenge. It consists of 45 cine-MRI images from mixed sources. The quality of images may vary due to the different sources and the challenges associated with cardiac MRI acquisition .

In summary, the ACDC database is known for its high-quality cardiac MRI images with detailed annotations. The quality of images in ChexNet and SunnyBrook databases can vary depending on the specific dataset sources and imaging conditions .

2. **I request the author to comment in general how the contrast inhomogeneities, subject variations from patient to patient or changes in acquisition parameters can affect the outcomes of the training. In general, a comment on this for both chapters 3 and 4 would be useful for future researchers trying to curate a database or use the same database for training.**

Response:

Certainly, the factors you've mentioned—contrast inhomogeneities, subject variations, and changes in acquisition parameters—can significantly impact the outcomes of training deep learning models on medical image datasets like ChexNet. Here's an explanation of how each of these factors can affect training outcomes:

1. Contrast Inhomogeneities: Medical images, including X-rays, can exhibit variations in contrast due to factors such as the imaging equipment used, the patient's condition, and the imaging technique. These variations can impact the visual features that the model

learns during training. If not addressed, the model might struggle to generalize across images with differing contrast levels, potentially leading to reduced diagnostic accuracy. Pre-processing techniques that normalize contrast or apply histogram equalization can help mitigate these variations.

2. Subject Variations from Patient to Patient: Patients' physical characteristics, body positions, and underlying medical conditions can result in significant variability in the appearance of pathologies in medical images. Deep learning models trained on a dataset with diverse subject variations might inadvertently focus on irrelevant features or be sensitive to non-essential differences. Robust models need to learn the essential disease-related patterns while ignoring irrelevant variations. Data augmentation techniques, combined with careful curation of diverse patient cases, can help models become more resilient to subject variations.

3. Changes in Acquisition Parameters: Different medical facilities might employ varied acquisition parameters such as X-ray machine settings, exposure levels, and positioning protocols. These parameter differences can introduce image variations that the model may not have encountered during training. Consequently, the model's performance might degrade when applied to images acquired with different parameters. To address this, a diverse dataset that incorporates a range of acquisition settings can aid the model in learning to handle variations effectively.

In addressing these challenges, researchers and authors often take the following steps:

- Data Augmentation: By applying various transformations during training, like rotations, flips, and deformations, models can learn to handle subject variations and changes in acquisition parameters.

- Normalization: Normalizing pixel values and contrast can help mitigate the effects of contrast inhomogeneities, making the data more consistent for the model.

- Data Diversity: Ensuring that the dataset encompasses a broad range of patient profiles, disease severities, and acquisition conditions can improve the model's generalization capabilities.

- Regularization Techniques: Techniques like dropout and batch normalization can help models become more robust to variations during training.

- Validation and Testing: Rigorous validation on diverse datasets, including data collected from different sources, can reveal how well the model generalizes across variations. Testing the model's performance on external datasets can provide insight into its real-world applicability.

In summary, understanding and addressing contrast inhomogeneities, subject variations, and changes in acquisition parameters are crucial for developing deep learning models on medical image datasets like ChexNet. A combination of pre-processing techniques, diverse data curation, and appropriate training strategies can help models navigate these challenges and produce more reliable and generalizable diagnostic outcomes.

3. **Various deep learning architectures have been explored for x Ray classification and segmentation. Chapter 3 describes a couple of models. The results reported in chapter 3 shows accuracies of the proposed model compared with reported literatures (table 3.1). The results show good comparison with the existing models with the same data. It would be good if the author can comment on what could improve the performance of a 4 class data.**

Response:

Certainly, improving the performance of a 4-class classification model on the ChexNet dataset, which includes Covid, normal, viral pneumonia, and bacterial pneumonia cases, involves a combination of data-related strategies and model optimization techniques. Here's a comprehensive response on how the model's performance can be enhanced:

1. Data Augmentation: Expanding the dataset through data augmentation techniques can improve model generalization. By applying transformations like rotations, flips, and scaling to the images, the model learns to recognize the same pathology from different perspectives, making it more robust to variations in presentation.

2. Balancing Class Distribution: Ensure that the dataset has a balanced representation of all four classes. If one class has significantly fewer samples, the model may struggle to learn patterns from that class. Techniques like oversampling, undersampling, or generating synthetic samples can help achieve a more equitable distribution.

3. Fine-Tuning and Transfer Learning: Start with a pre-trained model, such as one trained on a large dataset like ImageNet. Fine-tune the model's weights using the ChexNet dataset. This leverages the features learned from a larger dataset and can lead to improved performance on the target classes.

4. Architecture Selection: Experiment with different deep learning architectures suitable for image classification tasks. Popular architectures include Convolutional Neural Networks (CNNs) like ResNet, DenseNet, and Inception. The architecture's depth and complexity should match the size of the dataset and the complexity of the classification task.

5. Hyperparameter Tuning: Tune hyperparameters like learning rate, batch size, optimizer, and regularization strength. A grid search or random search approach can help identify the optimal combination of hyperparameters that yield the best performance.

6. Regularization Techniques: Implement regularization techniques like dropout and batch normalization to prevent overfitting and improve model generalization to unseen data.

7. Attention Mechanisms: Utilize attention mechanisms, such as self-attention or spatial attention, to enable the model to focus on more informative regions of the images. This can enhance the model's ability to capture subtle patterns.

8. Ensemble Learning: Combine predictions from multiple models to create an ensemble. Ensemble methods can improve performance by reducing bias and variance and capturing diverse patterns in the data.

9. Progressive Learning: Train the model progressively by starting with simpler tasks and gradually increasing the complexity. This approach allows the model to learn progressively more challenging patterns, leading to better convergence.

10. Transfer Learning with Domain Adaptation: If there are domain-specific variations in the ChexNet dataset, consider techniques like domain adaptation. This involves adapting the model to the target domain while leveraging knowledge from the source domain.

11. Interpretable AI Techniques: Employ techniques like Grad-CAM (Gradient-weighted Class Activation Mapping) to visualize and interpret the areas of the image that the model focuses on when making predictions. This can provide insights into the model's decision-making process.

12. Regular Monitoring and Model Updating: Continuously monitor the model's performance on validation and test sets. If the performance plateaus or degrades, consider updating the model with new data or refining the training process.

In conclusion, improving the performance of a 4-class classification model on the ChexNet dataset requires a multifaceted approach that includes data augmentation, balanced classes, architecture selection, hyperparameter tuning, regularization, and potentially advanced techniques like attention mechanisms and ensembling.

Experimentation, iterative refinement, and a deep understanding of the dataset's nuances are key to achieving optimal performance.

4. **Chapter 4 describes the transfer learning approach for detection of Covid 19 from CT/X rays and presents an efficient classification methodology for the precise identification of infection caused by Covid-19 using CT and X-ray images. Features were extracted using MobileNet V2 and classification was carried out through SVM. Figure 4.2 shows a flow chart with 'other lung diseases'. However, from the data used and results, it is not clear if it is a two class problem that classifies normal and covid cases or if other classes of diseases were also present in the training set. Please clarify this in the chapter. In Fact the title of the thesis itself says transfer learning approach to multiclass classification. Please clarify this in the thesis.**

Response:

In this paper, a different Chest X-Ray and CT dataset was used, as curated by the authors, and is available as open source on Kaggle . It has released 219 COVID-19, 1341 Normal, and 1345 Viral Pneumonia Chest X-ray (CXR) images. In this paper, our goal was to identify the presence of Covid-19, Normal vs Other lung diseases (Viral Pneumonia), and the term 'other lung diseases', have been used for Viral Pneumonia cases, and It is a multi class classification problem.

5. **Please shine more details on the database and type of data used. (what type of images, Classes of images in the database, etc) were used for training the Mobile Net SVM for transfer learning?**

Response:

The Curated database is of chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images, available on Kaggle. This COVID-19, normal, and

other lung infection dataset is released in stages. In the first release, it has released 219 COVID-19, 1341 normal, and 1345 viral pneumonia chest X-ray (CXR) images. In the first update, they have increased the COVID-19 class to 1200 CXR images. In the 2nd update, they have increased the database to 3616 COVID-19 positive cases along with 10,192 Normal, 6012 Lung Opacity (Non-COVID lung infection), and 1345 Viral Pneumonia images and corresponding lung masks.

6. **Chapter5 discusses a cascaded model for weakly supervised segmentation of the left ventricle in Cardiac MR Images. The approach here is very interesting. The author uses various traditional segmentation algorithms based on techniques like Seed Region growing, Random Walker, and K-Means clustering and the generated masks are used to train the deep learning model to get the final segmentation.**

Response:

Not Applicable (Thanks for the compliment)

7. **The proposed method is validated for the segmentation of the left ventricle in cardiac magnetic resonance images (ACDC Dataset). The method is again cross-validated with another dataset on which the model is not previously trained. The author might explain a little bit in detail whether a transfer of learning approach was actually implemented in this problem or the outcomes of the traditional segmentation results were used as ground truths to the DL algorithm.**

Response:

In this study, we present a novel approach for weakly supervised segmentation by combining established conventional segmentation techniques with contemporary deep neural network architectures. Our methodology entails a cascade of traditional methods including seed region growing, random walker, and K-means clustering, which operate on pixel classification based on intensity levels. These traditional methods are employed

to extract initial patch-based segmentations, yielding binary masks that are subsequently input into a newly proposed deep neural network. This network integrates an autoencoder with a U-Net style skip connection to accomplish the segmentation task. The autoencoder architecture aligns with U-Net's structure, comprising four layers and encoder-to-decoder skip connections to retain spatial-temporal feature maps. Distinctively, U-Net employs upsampling in the decoder to match feature map sizes, whereas the autoencoder employs pooling. Thus, our approach leverages an adapted U-Net architecture, customarily referred to as a variant of U-Net, for our segmentation task. It's noteworthy to mention that this adaptation, while utilizing the Unet framework, does not encompass pre-existing weights or parameters from previous tasks. Consequently, this process diverges from transfer learning, which specifically encompasses the utilization of prior knowledge to enhance performance on new tasks.

Review Report 1 - Examiner - 2

Comments:

1. Very thoroughly researched the scope of the problem.

2. Very clear summary of the findings in each area of the current imaging techniques.

3. Application of the new transfer learning was demonstrated.

4. Delineation of the limitations of the newer techniques of analysis.

5. Combining the other modalities of investigations with the new methodology would greatly enhance diagnostic value.

6. Liked the statement on future prospects.

List of points for clarification

1. **The x-ray images used in the manuscript, page 49, are from different age groups. Does it mean the pathology is similar among all ages? The criteria for diagnosis can be different. Please explain. You may be able to achieve higher accuracy and precision when similar ages are combined for the model building. It is not only the diagnostic criteria which can be generalized to other age groups because many times the pathological evolution can be different. So be careful, the model is not generalizable and applied to tests from the training group.**

Response:

The concerns raised align closely with considerations when working with the ChexNet dataset for multi-class classification involving four categories: Covid, normal, viral pneumonia, and bacterial pneumonia. The dataset incorporates X-ray images from diverse age groups, prompting a closer examination of the relationship between age, pathology presentation, and diagnostic accuracy.

1. Age-Related Pathology Variation: Notably, the ChexNet dataset comprises X-ray images from individuals spanning various age groups. This variation in age introduces the possibility of distinct pathology presentations across these groups. Pathologies might manifest differently due to developmental factors, changes in the immune system, or other age-related influences, potentially influencing the accuracy of diagnostic models.

2. Age-Dependent Diagnostic Criteria: Another pertinent consideration is the influence of age on diagnostic criteria. The criteria for identifying specific pathologies could indeed differ across age groups. Physiological and developmental disparities might lead to different diagnostic thresholds. Thus, a model trained predominantly on one age group might not seamlessly apply its learned criteria to others, which is a limitation for a generalized model.

3. Age-Constrained Model Performance: It's worth noting that training models on age-specific subgroups could enhance accuracy and precision for those groups. Similar age cohorts might exhibit consistent pathological patterns that the model can more effectively learn. However, this approach risks creating models that are optimized for narrow age ranges and struggle with unfamiliar age groups due to the complexities tied to age-related variations.

4. Model Generalization Challenge: The cautionary insight about model generalizability is pertinent. If a model trained predominantly on one age demographic is employed to analyze X-ray images from distinct age groups, its performance might diminish. Variances in pathology presentation and diagnostic criteria could compromise its efficacy, highlighting the need for models capable of accommodating diverse age-related factors.

5. Age-Related Pathological Evolution: Indeed, the evolution of pathologies can diverge across different age groups. The progression of diseases might differ in pediatric, adult, or elderly patients, warranting consideration of age-related nuances when constructing and assessing diagnostic models.

To recap, while pursuing heightened accuracy by focusing on specific age groups during ChexNet model training is appealing, **it's pivotal to strike a balance between precision and broad applicability.** Comprehensive evaluation across diverse age groups and external datasets is essential to ensure models perform adeptly across various cases. Acknowledging and addressing age-associated disparities in pathology presentation, diagnostic criteria, and pathological progression is key to developing resilient and dependable diagnostic models.

Suggestions for modification etc.

2. **Assumption of the diagnostic criteria is a static belief as our understanding about any disease increases so does the diagnostic criteria. With the changing diagnostic criteria over time how does the model capability improve with that change in time.**

Response:

Certainly, the query you've presented delves into an important aspect of diagnostic models, especially in the context of medical image analysis like the ChexNet dataset. Here's an explanation of how models' capabilities might evolve with changing diagnostic criteria over time:

1. Dynamic Nature of Diagnostic Criteria: Diagnostic criteria for medical conditions are not static; they evolve as our understanding of diseases improves. As medical research advances, new insights and discoveries lead to refined criteria for diagnosing various pathologies. This dynamic nature reflects the complexity of medical science and the ongoing pursuit of accuracy in diagnosis.

2. Model Adaptation to Changing Criteria: Deep learning models like those used in medical image analysis, including ChexNet, possess the potential to adapt to changing diagnostic criteria. However, this adaptation is contingent upon timely updates to the model's training data and retraining processes. When diagnostic criteria change, incorporating updated data that reflects these changes becomes essential to ensure the model remains accurate and aligned with current medical knowledge.

3. Continuous Learning and Retraining: One way to enable models to accommodate changing diagnostic criteria is through continuous learning and retraining. Models can be periodically retrained using new data that reflects the updated criteria. This process allows the model to learn the evolving patterns and characteristics associated with the revised criteria.

4. Incremental Improvement: Over time, as models are exposed to more varied and up-to-date data, their performance can incrementally improve. This improvement reflects the model's ability to learn and generalize from a broader range of cases. As new data with revised diagnostic criteria is introduced, the model may gradually refine its ability to accurately diagnose diseases based on the latest understanding.

5. Challenges and Caveats: Adapting models to changing criteria comes with challenges. Models need to be validated rigorously after updates to ensure they retain their generalization capabilities and do not overfit to specific changes. Striking a balance between retaining knowledge learned from previous data and incorporating new information is vital to maintaining robust model performance.

6. Collaboration with Medical Experts: Close collaboration between machine learning practitioners and medical experts is essential. Medical professionals can provide insights into changing diagnostic criteria, helping guide the model's retraining process effectively and ensuring it aligns with the current medical landscape.

In summary, while deep learning models have the potential to improve their capabilities with changing diagnostic criteria, this improvement requires a proactive approach involving continuous learning, retraining, validation, and collaboration between machine learning experts and medical professionals. By incorporating updated data and maintaining a flexible learning approach, models can adapt to evolving medical knowledge and contribute to more accurate and relevant diagnoses.

Review Report - 2

Dear Reviewer,

I wish to express my gratitude for your constructive feedback and insights. I have taken the time to thoroughly assess the research papers you shared, conducting an extensive literature review in the process. As a result, I have embarked on an in-depth exploration of the potential influence of age and gender-related biases on the classification of COVID-19 compared to other pneumonia types, utilizing X-ray images as our primary dataset.

I have carefully considered your suggestions regarding the comparison of X-rays of adult individuals with COVID-19, bacterial pneumonia, and viral pneumonia to adult "normal" X-rays, as well as comparing COVID-19, bacterial pneumonia, and viral pneumonia cases with normal X-rays of children of similar ages. These comparisons, undoubtedly, could provide deeper insights into the relationships between age, gender, and the diagnoses of these conditions.

I wholeheartedly acknowledge the significance of your comments, which underscore the critical importance of considering these demographic factors within the context of medical image analysis. Your dedication to advancing the quality of research in this field is deeply appreciated and serves as a motivating force for my work.

However, I would like to provide some context to explain why I missed these important key aspects in my thesis :

Dataset and Computational Challenges:

The dataset utilized for my research lacked age and gender-related metadata, except that it was mentioned as an adult chest X-Ray dataset. Unfortunately, specific age information was not available. This limitation was a consequence of the dataset's unavailability during the early stages of the COVID-19 pandemic when comprehensive data was not widely accessible, particularly with detailed age, gender, or demographic information.

It is worth noting that the dataset landscape has evolved since then, and more comprehensive datasets have become available, as evidenced by the studies you referenced in your comments. However, the challenges presented by these larger, more detailed datasets extend beyond just their availability. They pose significant computational demands and time constraints.

I trust this clarification provides a more accurate picture of the constraints I faced during the research phase, which limited my research to a more generalized version, rather than a specialized or micro study approach.

Action Points Taken (Thesis Modification):

I have implemented several revisions in my thesis, which I have appended to the end of this response letter, distinctly marking the key points I have incorporated for your convenience. Furthermore, I have integrated these modifications into the thesis itself, in accordance with your insightful recommendation to enhance its overall quality.

By keeping these things in mind, we can work towards creating more versatile and effective models for COVID-19 detection in both adult and pediatric populations, ultimately improving diagnostic accuracy and benefiting the field of medical imaging and healthcare as a whole. I am committed to further exploring these critical aspects in future research endeavors as well. I appreciate your understanding of the complexities involved in this situation.

Once again, I extend my heartfelt appreciation for your thoughtful review and for sharing valuable articles that support the consideration of age and gender-related biases in medical image analysis. Your feedback has been instrumental in shaping the trajectory of my research, and I look forward to your continued guidance.

Sincerely,

Alok Tiwari, Research Scholar
School of Biomedical Engineering,
IIT-BHU, Varanasi, UP

4. Incremental Improvement: Over time, as models are exposed to more varied and up-to-date data, their performance can incrementally improve. This improvement reflects the model's ability to learn and generalize from a broader range of cases. As new data with revised diagnostic criteria is introduced, the model may gradually refine its ability to accurately diagnose diseases based on the latest understanding.

5. Challenges and Caveats: Adapting models to changing criteria comes with challenges. Models need to be validated rigorously after updates to ensure they retain their generalization capabilities and do not overfit to specific changes. Striking a balance between retaining knowledge learned from previous data and incorporating new information is vital to maintaining robust model performance.

6. Collaboration with Medical Experts: Close collaboration between machine learning practitioners and medical experts is essential. Medical professionals can provide insights into changing diagnostic criteria, helping guide the model's retraining process effectively and ensuring it aligns with the current medical landscape.

In summary, while deep learning models have the potential to improve their capabilities with changing diagnostic criteria, this improvement requires a proactive approach involving continuous learning, retraining, validation, and collaboration between machine learning experts and medical professionals. By incorporating updated data and maintaining a flexible learning approach, models can adapt to evolving medical knowledge and contribute to more accurate and relevant diagnoses.

3.5.5 Evaluating model robustness: Age-Group Generalization

To acknowledge the importance of age and gender-related biases in medical imaging, and in order to validate our model on age-related variations [58], I have downloaded a smaller dataset sample from Kaggle, which is specifically mentioned as the children's chest X-ray dataset. The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal).

Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care.

For the analysis of chest X-ray images, all chest radiographs were initially screened for quality control by removing all low-quality or unreadable scans. The diagnoses for the images were then

graded by two expert physicians before being cleared for training in the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

I took a sample dataset from the above which includes 1400 images, which has been split into training (1120 images) and testing (280 images) sets with a split ratio of 0.75. It consists of three classes, namely - Normal, Viral Pneumonia, and Bacterial Pneumonia. I have trained the proposed model on this dataset, and the model's test performance results are as follows:

Classification Results (Normal Vs Pneumonia for children chest X-Ray dataset)

	Predicted Positive	Predicted Negative
Actual Positive	130	30
Actual Negative	35	85

Accuracy

$(TP + TN)/(TP + TN + FP + FN) = = (130 + 85) / (130 + 85 + 35 + 30) = 67.83$ %

Presicion

$TP / (TP + FP) = 130 / (130 + 35) = 78.79$ %

Recall

$TP / (TP + FN) = 130 / (130 + 30) = 81.25$ %

F1-Score

2 * (Precision * Recall) / (Precision + Recall)

$= 2 * (0.7879 * 0.8125) / (0.7879 + 0.8125) = 79.99$ %

3.5.6 Model Performance Comparison (Adult vs Children)

Upon scrutinizing the outcomes presented above and in direct contrast to the model's performance when applied to the adult dataset, it is evident that the corresponding metrics exhibit notable disparities [58] [59], [60], [61], [62], [63]. This disparity is visually represented in the accompanying table:

Dataset	Model	Train Accuracy	Val Accuracy	Test Accuracy
Adult Chest X-ray	ResNet	0.99	0.97	0.98
	VGG-16	0.97	0.96	0.96
	Inception Net	0.95	0.93	0.94
	Proposed Net	0.97	0.96	0.97
Children Xhest X-Ray	ResNet	0.62	0.58	0.61
	VGG-16	0.61	0.59	0.58
	Inception Net	0.59	0.57	0.56
	Proposed Net	0.67	0.61	0.59

In order to Justify poor model performance on a children's X-ray dataset compared to an adult chest X-ray dataset for COVID-19 detection can be attributed to several probable factors. Here is a proper justification for this difference in performance:

1. Age-Related Differences in Disease Presentation:

COVID-19 can manifest differently in children compared to adults. Children often exhibit milder symptoms, and their chest X-rays may not show the same level of pathology as those of adults [61, 62, 64]. This fundamental difference in disease presentation can make it more challenging to detect COVID-19 in children through imaging alone.

2. Dataset Size and Diversity:

The size and diversity of the dataset play a crucial role in model performance. Adult chest X-ray datasets are typically more extensive and well-annotated than children's datasets [65] due to the higher incidence of chest X-ray imaging in adults. A smaller and less diverse dataset for children can limit the model's ability to generalize to a broader population.

3. Data Imbalance:

Children's X-ray datasets may suffer from class imbalance [65], where the number of COVID-19 cases is significantly lower than other conditions or normal cases. This imbalance can lead to bias in the model, making it more proficient at classifying the majority class (e.g., normal cases) and less accurate in identifying the minority class (COVID-19 cases).

4. Developmental Differences in Anatomy:

Children's anatomy differs [7] from that of adults, and their chest X-ray images can vary in terms of size, structure, and growth-related changes. Models trained primarily on adult datasets may

struggle to adapt to these age-specific anatomical variances, affecting their performance on pediatric X-rays.

5. Limited Availability of Pediatric Data:

Pediatric-specific COVID-19 datasets, especially those containing a large number of X-ray images, may be limited [66] in comparison to adult datasets. This scarcity of data can hinder model training and result in suboptimal performance.

6. Technical Challenges in Pediatric Imaging:

Capturing high-quality X-ray images of children can be technically challenging[67] due to their smaller size and higher likelihood of movement during the imaging process. Lower image quality can introduce noise and reduce the model's ability to detect subtle abnormalities.

7. Model Optimization for Age Groups:

Models optimized for adult chest X-ray analysis may not be suitable for pediatric cases without fine-tuning or transfer learning [68]. Age-specific model optimization is crucial to adapt to the unique characteristics of pediatric X-rays.

8. Disease Prevalence in the Population:

The prevalence of COVID-19 may differ between adults and children in a given population [69]. A lower prevalence in the pediatric population may result in fewer positive cases for model training, making it more challenging for the model to learn the patterns specific to pediatric COVID-19 cases.

In summary, the poorer model performance on a children's X-ray dataset compared to an adult chest X-ray dataset for COVID-19 detection can be attributed to a combination of differences in disease presentation, dataset characteristics, anatomical variances, and technical challenges. Addressing these challenges may require dedicated efforts, including the collection of more extensive and diverse pediatric datasets and the development of age-specific model optimizations.

3.5.7 Achieving the enhanced generalization capabilities of the model

To develop a model with enhanced generalization capabilities[70] across both adult and pediatric X-ray datasets, while also improving its performance in the context of COVID-19 classification, the following research avenues warrant exploration:

1. Comprehensive Age-Stratified Datasets:

Collect and curate large and comprehensive X-ray datasets that include both adults and children, stratified [71] by age groups. Ensure that these datasets capture a wide range of COVID-19 cases, other pneumonia types, and normal cases for each age group.

2. Anatomical Variation Modeling:

Investigate methods for modeling and accommodating the anatomical variations[72] between adults and children in chest X-ray images. This could involve developing age-specific preprocessing techniques and data augmentation strategies to standardize image size and features.

3. Transfer Learning and Fine-Tuning:

Explore transfer learning techniques by pretraining models on adult chest X-ray datasets and fine-tuning [73] them on pediatric datasets. Fine-tuning should consider the differences in age groups, enabling the model to adapt to both populations effectively.

4. Age and Gender Biases:

Investigate the impact of age and gender biases [61] in COVID-19 classification. Analyze how these biases affect model performance and develop techniques to mitigate bias-related challenges for both adults and children.

5. Ensemble Models:

Evaluate the effectiveness of ensemble models [74] that combine predictions from multiple models, each specialized for a specific age group. This can help improve overall model performance and robustness.

6. Data Augmentation and Synthesis:

Develop advanced data augmentation and synthesis techniques [75] specifically designed for pediatric X-ray images. These methods can help address the challenges posed by smaller datasets for children.

7. Explainable AI in Pediatric Medicine:

Incorporate explainable AI techniques to provide interpretable results for pediatric COVID-19 diagnosis [76]. This can enhance the trust and adoption of AI-based diagnostic tools in pediatric medicine.

8. Clinical Validation and Collaboration:

Collaborate with medical professionals to validate model performance in real clinical settings. Gather feedback from clinicians [77] to fine-tune the models for practical use and ensure alignment with clinical needs.

9. Multimodal Approaches:

Investigate the potential benefits of combining X-ray data with other clinical data modalities [78], such as patient history and laboratory results, to enhance diagnostic accuracy for both adults and children.

10. Robustness to Noise and Image Quality:

Develop models that are robust to variations in image quality and noise [79], which can be common in pediatric X-rays. This can involve preprocessing techniques to enhance image quality.

11. Ethical Considerations:

Address ethical considerations [80] related to pediatric data collection and model deployment, ensuring that privacy and informed consent protocols are in place.

12. Knowledge Sharing and Collaboration:

Foster collaboration among researchers, medical institutions, and data providers to share knowledge, datasets, and best practices for developing generalized models for COVID-19 detection across different age groups.

By considering the above-mentioned points, we can work towards creating more versatile and effective models for COVID-19 detection in both adult and pediatric populations, ultimately improving diagnostic accuracy and benefiting the field of medical imaging and healthcare as a whole.

3.5.8 Discussion and Conclusion

In this paper, we have proposed a transfer-learning-based model, which discusses three case studies

1. Four-Class Classification
2. Three-Class Classification
3. Two-Class Classification

[55] M. T. Law, A. L. Traboulsee, D. K. Li, R. L. Carruthers, M. S. Freedman, S. H. Kolind, and R. Tam, "Machine learning in secondary progressive multiple sclerosis: an improved predictive model for short-term disability progression," *Multiple Sclerosis Journal–Experimental, Translational and Clinical*, vol. 5, no. 4, p. 2055217319885983, 2019.

[56] N. Patricia and B. Caputo, "Learning to learn, from transfer learning to domain adaptation: A unifying perspective," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1442–1449, 2014.

[57] J. D. Fuhrman, N. Gorre, Q. Hu, H. Li, I. El Naqa, and M. L. Giger, "A review of explainable and interpretable ai with applications in covid-19 imaging," *Medical Physics*, vol. 49, no. 1, pp. 1–14, 2022.

[58] H. Ieki, K. Ito, M. Saji, R. Kawakami, Y. Nagatomo, K. Takada, T. Kariyasu, H. Machida, S. Koyama, H. Yoshida, R. Kurosawa, H. Matsunaga, K. Miyazawa, K. Ozaki, Y. Onouchi, S. Katsushika, R. Matsuoka, H. Shinohara, T. Yamaguchi, S. Kodera, Y. Higashikuni, K. Fujiu, H. Akazawa, N. Iguchi, M. Isobe, T. Yoshikawa, and I. Komuro, "Deep learning-based age estimation from chest X-rays indicates cardiovascular prognosis," *Communications Medicine*, vol. 2, p. 159, Dec. 2022.

[59] K. Hammoudi, H. Benhabiles, M. Melkemi, F. Dornaika, I. Arganda-Carreras, D. Collard, and A. Scherpereel, "Deep Learning on Chest X-ray Images to Detect and Evaluate Pneumonia Cases at the Era of COVID-19," *Journal of Medical Systems*, vol. 45, p. 75, July 2021.

[60] C. Solomou and D. Kazakov, "Utilizing chest x-rays for age prediction and gender classification," in *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 356–361, IEEE, 2021.

[61] L. Álvarez Rodríguez, J. D. Moura, J. Novo, and M. Ortega, "Does imbalance in chest X-ray datasets produce biased deep learning approaches for COVID-19 screening?," *BMC Medical Research Methodology*, vol. 22, p. 125, Dec. 2022.

[62] A. J. Larrazabal, N. Nieto, V. Peterson, D. H. Milone, and E. Ferrante, "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis," *Proceedings of the National Academy of Sciences*, vol. 117, pp. 12592–12594, June 2020.

[63] L. Seyyed-Kalantari, H. Zhang, M. B. McDermott, I. Y. Chen, and M. Ghassemi, "Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations," *Nature medicine*, vol. 27, no. 12, pp. 2176–2182, 2021.

[64] L. Seyyed-Kalantari, H. Zhang, M. B. A. McDermott, I. Y. Chen, and M. Ghassemi, "Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations," *Nature Medicine*, vol. 27, pp. 2176–2182, Dec. 2021.

[65] B. Garcia Santa Cruz, M. N. Bossa, J. Sölter, and A. D. Husch, "Public Covid-19 X-ray datasets and their impact on model bias – A systematic review of a significant problem," *Medical Image Analysis*, vol. 74, p. 102225, Dec. 2021.

[66] B. T. Jankowitz and P. D. Adelson, "Pediatric traumatic brain injury: past, present and future," *Developmental neuroscience*, vol. 28, no. 4-5, pp. 264–275, 2006.

[67] N. Raschle, J. Zuk, S. Ortiz-Mantilla, D. D. Sliva, A. Franceschi, P. E. Grant, A. A. Benasich, and N. Gaab, "Pediatric neuroimaging in early childhood and infancy: challenges and practical guidelines," *Annals of the New York Academy of sciences*, vol. 1252, no. 1, pp. 43–50, 2012.

[68] L. Matrajt, J. Eaton, T. Leung, and E. R. Brown, "Vaccine optimization for covid-19: Who to vaccinate first?," *Science Advances*, vol. 7, no. 6, p. eabf1374, 2021.

[69] K. Yuki, M. Fujiogi, and S. Koutsogiannaki, "Covid-19 pathophysiology: A review," *Clinical immunology*, vol. 215, p. 108427, 2020.

[70] K. B. Ahmed, G. M. Goldgof, R. Paul, D. B. Goldgof, and L. O. Hall, "Discovery of a generalization gap of convolutional neural networks on covid-19 x-rays classification," *Ieee Access*, vol. 9, pp. 72970–72979, 2021.

[71] A. Vicco, C. McCormack, B. Pedrique, I. Ribeiro, G. N. Malavige, and I. Dorigatti, "A scoping literature review of global dengue age-stratified seroprevalence data: estimating dengue force of infection in endemic countries," *medRxiv*, pp. 2023–04, 2023.

[72] A. Alraddadi, "Literature review of anatomical variations: clinical significance, identification approach, and teaching strategies," *Cureus*, vol. 13, no. 4, 2021.

[73] G. Vrbančič and V. Podgorelec, "Transfer learning with adaptive fine-tuning," *IEEE Access*, vol. 8, pp. 196197–196211, 2020.

[74] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018.

[75] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, p. 100258, 2022.

[76] M. Karim, T. Döhmen, D. Rebholz-Schuhmann, S. Decker, M. Cochez, O. Beyan, *et al.*, "Deepcovidexplainer: Explainable covid-19 predictions based on chest x-ray images," *arXiv preprint arXiv:2004.04582*, vol. 28, 2020.

[77] D. Brasfield, G. Hicks, S. Soong, J. Peters, and R. Tiller, "Evaluation of scoring system of the chest radiograph in cystic fibrosis: a collaborative study," *American Journal of Roentgenology*, vol. 134, no. 6, pp. 1195–1198, 1980.

[78] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, "Improving the performance of cnn to predict the likelihood of covid-19 using chest x-ray images with preprocessing algorithms," *International journal of medical informatics*, vol. 144, p. 104284, 2020.

[79] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Making image quality assessment robust," in *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 1718–1722, IEEE, 2012.

[80] L. M. Connelly, "Ethical considerations in research studies," *Medsurg nursing*, vol. 23, no. 1, pp. 54–56, 2014.

[81] M. J. Rawle, D. L. Bertfield, and S. E. Brill, "Atypical presentations of COVID-19 in care home residents presenting to secondary care: A UK single centre study," *AGING MEDICINE*, vol. 3, pp. 237–244, Dec. 2020.

[82] S. Tripathi, A. Verma, and N. Sharma, "Augmented Deep Learning Architecture to Effectively Segment the Cancerous Regions in Biomedical Images," in *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, (Gunupur Odisha, India), pp. 1–6, IEEE, Dec. 2020.

[83] T. S. Sharan, S. Tripathi, S. Sharma, and N. Sharma, "Encoder Modified U-Net and Feature Pyramid Network for Multi-class Segmentation of Cardiac Magnetic Resonance Images," *IETE Technical Review*, pp. 1–13, Aug. 2021.

Response to Reviewer

1. **Thanks for your responses.**

   Response: Thanks for your acknowledgement.

2. **Particularly impressed by your extended literature research beyond the articles I had attached.**

   Response: I appreciate your positive assessment of the extended literature research in my thesis. It was a deliberate effort to provide a comprehensive foundation for my work, and I'm pleased that you found it impressive.

3. **Your responses as well as incorporation into the text of the thesis gave a better response.**

   Response: Thank you for recognizing the value of my responses and their integration into the thesis. Your feedback encourages me to continue refining my work to provide better and more meaningful responses.

4. **Hopefully your future research will continue to exemplify the fact that age and gender are important facets of disease presentation from the medical perspective.**

   Response : I share your view on the importance of considering age and gender in disease presentation, and I'm committed to ensuring that my future research reflects this significance. Your expectation is duly noted and will guide my future work.

5. **It is important to recognize the difference in the results of comparing normal adult chest x-rays with diseased adult chest x-rays and comparing normal Pediatric age chest x-rays with diseased Pediatric age group chest x-rays. The accuracy table on page 70 is reflective of the comparison. Although the Pediatric age accuracy is considerably less than that compared for the adults.**

   Response: Your observation about the distinction in results between adult and pediatric chest X-rays is crucial. I concur that this distinction is reflected in the accuracy table on page 70, where pediatric accuracy is notably lower. This underscores the need for more comprehensive datasets, especially in a populous country like India, to enhance predictability and accuracy.

6. **By sheer numbers of the population of India there should be more data sets of x-rays available for both adults and children for better predictability.**

   Response: As discussed in point 5, there is definitely a need for more comprehensive datasets, especially in a populous country like India, to enhance predictability and accuracy for both adults and children datasets.

7. **Even in the children there could be variation in normal x-ray from age 1 day to age 20 years, the span of Pediatric age group. Whereas the age from day 1 through 28 days is considered neonatal age; infancy is up to 1 year of age, toddler, children and adolescent. And each age group will have**

different characteristics. For example; x-ray at 1 hr of age will look different from x-ray at 4 hrs of age on day 1 of life.

8. **Involvement of medical specialists in the area should be consulted for better application of the AI models.**

Response (7 and 8): I appreciate your insight into the variation in normal X-rays across different pediatric age groups. Your suggestion to involve medical specialists is well-founded, and I believe that their expertise is essential for the effective application of AI models in this context. This interdisciplinary approach is indeed valuable.

9. **Hope to see more research from you in future in newborn age group patients.**

Response : I'm grateful for your hope to see more of my research in the newborn age group. This is a challenging and important area, and your encouragement motivates me to delve deeper into this field in the future.

10. **On the page 70 there is a typo in the Pediatric analysis result, "Xhest" should be "Chest".**

Response: Thank you for pointing out the typo on page 70. Your attention to detail is much appreciated, and I made the necessary corrections as shown below-

Dataset	Model	Train Accuracy	Val Accuracy	Test Accuracy
Adult Chest X-ray	ResNet	0.99	0.97	0.98
	VGG-16 Net	0.97	0.96	0.96
	Inception Net	0.95	0.93	0.94
	Proposed Net	0.97	0.96	0.97
Children Chest X-ray	ResNet	0.62	0.58	0.61
	VGG-16 Net	0.61	0.59	0.58
	Inception Net	0.59	0.57	0.56
	Proposed Net	0.67	0.61	0.59