Multi-Disease Detection and Classification in Chest Radiography: A Deep Convolutional Neural Network Approach

Paul O. Okafor (MS)

Abstract

This research employs deep learning models, particularly YOLOv8, to enhance the detection and classification of thoracic pathologies using the NIH Chest X-ray Dataset, which contains 112,120 high-resolution images from 30,805 patients, annotated for 14 distinct pathologies using advanced Natural Language Processing techniques. The object detection model (YOLOv8) was trained on a subset of fewer than 1,000 images, limited to eight well-represented classes due to sparse bounding box annotations, while the classification model was trained on the entire dataset across all 14 pathologies, aiming to provide comprehensive disease detection capabilities. Training spanned 100 epochs with a focus on minimizing training and validation loss while optimizing top-1 and top-5 accuracy metrics. The performance evaluation involved comprehensive metrics including loss graphs, precision-recall, and F1 score analyses, which indicated successful model learning with some challenges in distinguishing between similar thoracic conditions. The study aimed to enhance the detection accuracy of medical imaging analyses, potentially aiding in faster and more accurate clinical diagnostics.

Keywords

Chest Radiography, Thoracic Pathologies, NIH Chest X-ray Dataset, Deep Learning, You Only Look Once (YOLO).

1. Introduction

Chest radiography stands as a cornerstone in the diagnostic assessment of thoracic diseases, leveraging its wide accessibility and non-invasive nature to provide critical insights into various pulmonary conditions. The utility of chest X-rays spans the detection of a broad spectrum of pathologies, from infectious diseases like pneumonia, which annually impacts hundreds of millions globally, to chronic conditions such as emphysema and cardiomegaly. Despite their ubiquity in clinical practice, the interpretation of chest radiographs is fraught with challenges. The subtlety of radiographic manifestations of different diseases, compounded by the variability in patient anatomy and the quality of radiographs, requires a high degree of expertise and poses a considerable risk for diagnostic errors.

The integration of deep learning into the realm of medical imaging has marked a transformative shift, offering promising solutions to augment the diagnostic process. Deep learning, particularly Convolutional Neural Networks (CNNs), has demonstrated exceptional prowess in extracting intricate patterns from images, surpassing traditional image processing and machine learning approaches in both accuracy and efficiency. These advancements hold significant promise for chest radiography, where the potential to automate and enhance the accuracy of disease detection could revolutionize patient care, particularly in resource-constrained settings where radiologist expertise is scarce.

However, the application of deep learning in this field is not without its hurdles. One of the most pressing challenges is the inherent imbalance in medical datasets, where certain conditions are significantly underrepresented. This skew in data distribution can lead to biased models that perform well on common conditions but falter on rarer, potentially more critical diseases. Moreover, the high degree of similarity between different thoracic pathologies in radiographic images complicates the task of multi-disease classification, necessitating models that can discern nuanced differences between disease signatures.

This research aims to delve into these challenges, proposing the development and optimization of deep CNN architectures tailored for the nuanced task of multi-disease detection and classification in chest X-rays, and aspiring to provide a level of diagnostic accuracy that can genuinely augment clinical decision-making and improve patient care in the field of radiology.

2. Problem Statement

The nuanced nature of thoracic diseases, coupled with the intrinsic challenges of chest X-ray interpretation, necessitates the development of sophisticated deep learning models capable of high precision and reliability. The task is further complicated by the inherent imbalance in disease prevalence within available datasets, which can skew model performance and impede the effective learning of rare disease features. Additionally, the high dimensionality and complexity of chest X-ray images demand models that can efficiently extract and utilize relevant features without succumbing to overfitting or computational inefficiencies.

This research aims to address these challenges by developing and optimizing deep convolutional neural network (CNN) architectures for the detection and classification of multiple diseases in chest X-rays. By leveraging advanced deep learning techniques, including attention mechanisms and novel loss functions, the proposed models seek to enhance diagnostic accuracy, offering valuable tools to support radiologists and improve patient outcomes.

3. Research Objectives and Significance

The objectives of this research include the following:

- To develop and optimize deep convolutional neural networks (CNN) models capable of accurately detecting and classifying multiple thoracic diseases in chest X-ray images.
- To systematically evaluate the performance of cutting-edge deep CNN architectures, such as variants of YOLOv8, tailored for the complexities of medical imaging data.
- To explore and implement novel training strategies and network adaptations to enhance model sensitivity and specificity in identifying a broad range of thoracic pathologies.

This research holds profound implications for the future of medical diagnostics, particularly in the realm of thoracic disease detection and classification through chest X-rays. By advancing deep convolutional neural network models, it aims to significantly improve diagnostic accuracy, reduce the burden on healthcare professionals, and enhance patient outcomes. The integration of state-of-the-art AI into medical imaging promises to democratize access to expert-level diagnostics, particularly in under-resourced settings, paving the way for a more efficient, accessible, and equitable global healthcare landscape.

4. Related Work

Rakshit et al. (2019) explored deep learning applications for thoracic disease detection using chest X-ray imagery, highlighting the challenges of working with deeply layered models that require extensive parameters. Their research focuses on automating disease classification with a simplified parameter model using transfer learning, specifically employing the ResNet18 architecture to create visual heatmaps for localized abnormality identification.

Matsumoto et al. (2020) delved into the niche application of deep learning for diagnosing heart failure from chest X-rays, achieving a notable accuracy of 82% through data augmentation and transfer learning techniques. The use of heatmap visualizations to interpret model decisions is particularly relevant to our study, providing a methodological baseline that supports our findings on the importance of visual validation in improving model interpretability and trustworthiness among medical practitioners. Their work underscores the potential of specialized deep learning applications in enhancing diagnostic processes for specific cardiac conditions.

Zhou et al. (2019) addressed the challenge of cardiomegaly identification using the ChestX-ray8 dataset, leveraging transfer learning to attain an AUC of 0.87. Their results underscore the efficacy of deep learning methods in distinguishing specific pathologies, reinforcing the viability of our approach to employ similar techniques for broader thoracic condition identification. This study not only corroborates the effectiveness of transfer learning in medical imaging analysis but also enhances our understanding of its potential in refining diagnostic accuracies across varied thoracic diseases.

Wang et al. (2017) introduced the ChestX-ray8 database, focusing on the weakly-supervised classification and localization of common thorax diseases. Their pioneering work in utilizing hospital-scale databases for training deep learning models offers significant insights into managing large datasets and extracting valuable diagnostic information despite the challenges posed by loosely labeled data. The framework they developed for disease detection and localization using only image-level labels provides a foundational approach that informs our strategies for dealing with similar labeling limitations and underscores the feasibility of applying weak supervision in medical image analysis to achieve substantial advancements in automated diagnostics.

5. Model Development

5.1. Data

The foundation of this research is the NIH Chest X-ray Dataset, a comprehensive collection comprising 112,120 high-resolution frontal-view X-ray images sourced from 30,805 unique patients (Crawford 2018). Each image within this dataset has been meticulously labeled for the presence of 14 distinct thoracic pathologies through advanced Natural Language Processing (NLP) techniques applied to corresponding radiological reports. These pathologies include Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Mass, Nodule, Pleural Thickening, Pneumonia, and Pneumothorax, making this dataset an invaluable asset for training deep learning models aimed at multi-disease detection and classification in chest radiography. This rich dataset not only facilitates the training and validation of deep learning models for disease detection and classification but also presents challenges such as label imbalances and the high dimensionality of medical images, necessitating innovative computational approaches.

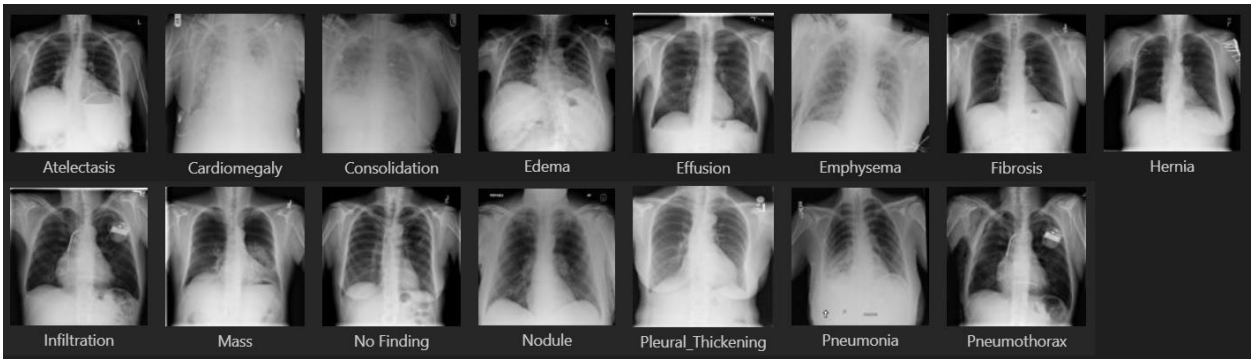

Figure 1: Chest Xray images showing the 14 distinct thoracic pathologies.

Table 1. Description of the Thoracic Pathologies

Thoracic Pathologies	Definition
Atelectasis	A lung condition characterized by the deflation of alveoli, leading to partially or completely collapsed lung tissue (Restrepo & Braverman 2015).
Cardiomegaly	An enlargement of the heart, often congenital, but can also result from conditions such as heart valve diseases, thyroid disorders, or systemic diseases like diabetes and HIV (Zhou et al. 2021).

Consolidation	Occurs when liquid replaces air in the lung tissue, leading to stiffened and swollen lung areas, often due to infections or injuries (Peng 2020).
Edema	The accumulation of excess fluid in the lungs, which can complicate breathing, typically associated with congestive heart failure or acute lung injury (Herrero et al. 2018).
Effusion	Known as pleural effusion, this condition involves excess fluid build-up in the pleural cavity, limiting lung expansion and affecting breathing (Zhou et al. 2021).
Emphysema	A lung condition marked by the destruction of alveoli, leading to larger air spaces and reduced surface area, causing breathlessness (Eriksson 2013).
Fibrosis	A progressive lung disease where lung tissue around the alveoli becomes stiff and scarred, making it difficult to breathe (Murtha et al. 2017).
Hernia	Occurs when lung tissue protrudes through the thoracic wall, usually without symptoms unless it becomes severe, potentially causing pain and swelling (Wong et al. 2018).
Infiltration	The accumulation of foreign substances in the lung, which can fill and block the lung spaces, posing severe health risks (Mégarbane & Chevillard 2013).
Mass	A significant opaque area on a chest X-ray, indicating a region in the lungs larger than 30 mm, possibly due to benign or malignant causes (de Paula et al. 2015).
Nodule	A small, often benign opaque area in the lung, typically less than 30 mm in diameter, which may be asymptomatic unless malignant (Rakshit et al. 2019).
Pleural Thickening	The scarring and hardening of the tissue surrounding the lungs, similar to fibrosis, and often detected late as it progresses (Rakshit et al. 2019).
Pneumonia	An inflammatory condition of the lung affecting primarily the alveoli, usually caused by infection with viruses or bacteria (Rakshit et al. 2019).
Pneumothorax	The presence of air in the pleural space causing lung collapse, which can disrupt normal breathing and lead to acute respiratory distress (Zhou et al. 2021).

Building on the detailed characterization of the NIH Chest X-ray Dataset, further exploratory data analysis underscores the dataset's demographic diversity and its implications for model development. The gender distribution within the dataset (figure 2) exhibits a moderate skew, with males representing 54.0% and females accounting for 46.0% of the 30,805 unique patients. This near parity is critical, as it facilitates the development of an unbiased model, capable of generalizing across different genders. Considering the anatomical and physiological differences between genders that may affect the radiographic manifestations of thoracic pathologies, the balanced gender representation ensures that the model's training data encapsulates a broad spectrum of clinical presentations.

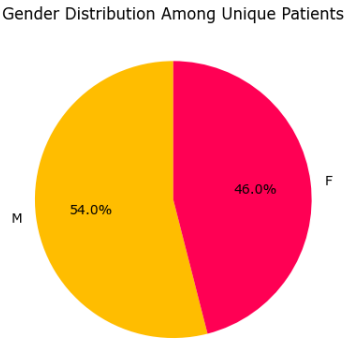

Figure 2. Gender Distribution among unique Patients.

Furthermore, the age distribution across various pathologies, provides pivotal insights into the epidemiological characteristics of the patient population. The boxplot of patient age distribution by disease (figure 3) indicates a significant variation in median ages and age ranges across different pathologies. For example, conditions such as Cardiomegaly and Emphysema predominantly affect older age groups, suggesting a potential need for stratified sampling or targeted data augmentation to enhance the model's performance in these subgroups. In contrast, diseases like Hernia do not show a pronounced age preference, underscoring the necessity for the model to maintain high performance across diverse age groups (figure 3). This demographic analysis not only enriches our understanding of the dataset but also informs several critical aspects of model training. By acknowledging and adjusting for the inherent demographic variabilities in age and gender among the patients, the model can be better tailored to detect and classify thoracic pathologies accurately across the general population. Such considerations are indispensable for ensuring that the deep learning model is not only effective but also equitable in its diagnostic capabilities.

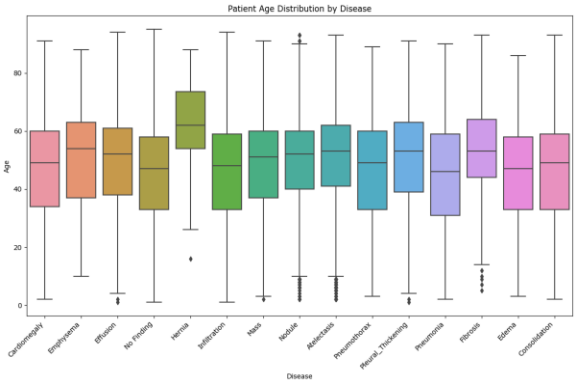

Figure 3. Patient Age Distribution by Disease

The age distribution (figure 4) histogram illustrates a prominently bimodal distribution with peaks around the mid-50s and another smaller peak around the early 20s. This distribution highlights the diversity in the patient age range, which spans from newborns to the elderly, with a significant concentration in the middle-aged demographic. This bimodality suggests that different age-specific features may be prevalent in the X-ray images, which could necessitate age-adjusted approaches in the model training process. For instance, age-related changes in lung structure and the increased likelihood of certain pathologies in older populations may require tailored image preprocessing or model architecture adjustments to optimize diagnostic performance across such a varied age spectrum.

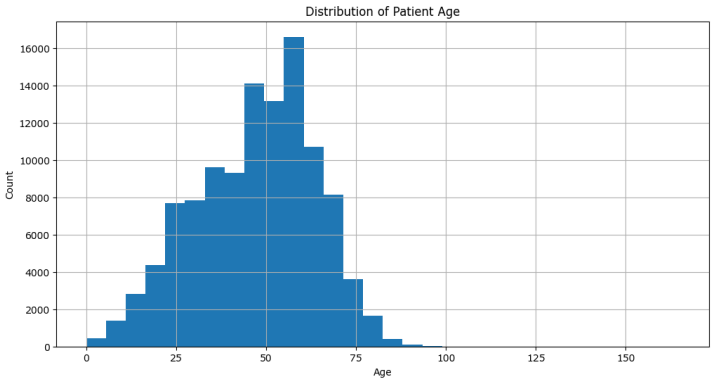

Figure 4. Distribution of Patient Age

The analysis of disease label distribution (figure 5a) within the dataset reveals a heavy skew towards certain conditions such as Infiltration, Effusion, and Atelectasis, while others like Hernia and Pneumonia are relatively underrepresented. This imbalance poses significant challenges for model training, as it could lead to a bias towards more frequently labeled conditions. To address this, techniques such as weighted loss functions or oversampling of underrepresented classes might be employed to ensure that the model does not overlook rarer but clinically significant abnormalities.

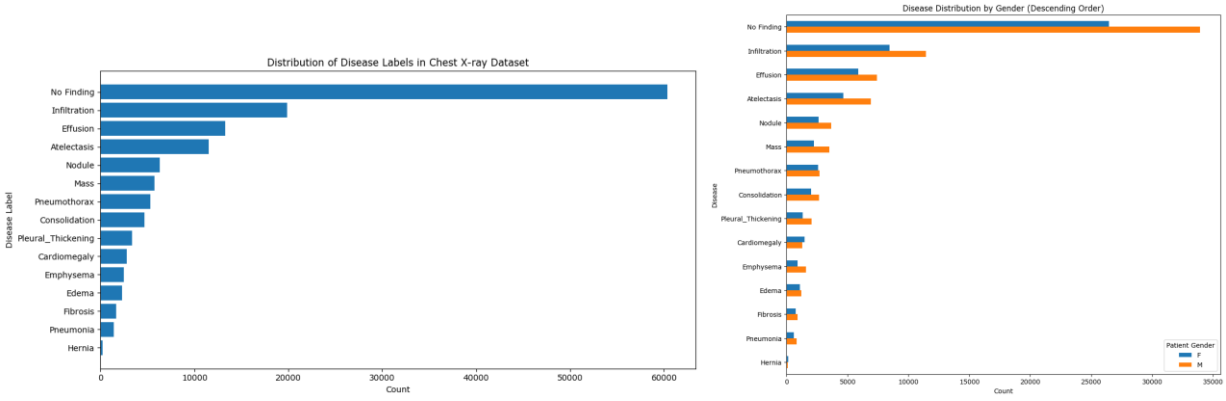

Figure 5. Distribution of Disease labels in Chest X-Ray Dataset

The disease distribution by gender (figure 5b) further enriches the demographic analysis by showing distinct patterns in disease prevalence between males and females. Notably, conditions such as Emphysema and Cardiomegaly exhibit a higher incidence in males, while others like Pneumonia and Fibrosis are more evenly distributed between genders. This disparity necessitates the incorporation of gender-specific considerations in the model to prevent gender bias in diagnostic predictions.

5.2. Pre-processing

The raw data, including two critical CSV files—one containing bounding boxes and the other classification information alongside demographics—are loaded. the data is immediately organized for further processing. Instead of random assignment, images are methodically sorted into training, validation, or testing sets according to pre-existing lists specified in the “train_val_list” and “test_list”, ensuring that the distribution of images adheres to designed experimental parameters, which may aim to balance pathology representation or demographic diversity. Paths for these categorized images are then meticulously extracted and stored in

a dedicated column within the data frame, optimizing retrieval and access during model training and validation phases. This organization facilitates efficient access to the image files during the model training and validation processes (table 2).

Table 2. Table showing the class-wise number of images.

Diseases	Train	Test	Validation
Atelectasis	6,640	3279	1640
Cardiomegaly	1,366	1069	341
Effusion	6,993	4658	1666
Infiltration	11,037	6112	2745
Mass	3,258	1748	776
Nodule	3,743	1623	965
Pneumonia	692	555	184
Pneumothorax	2085	2665	552
Consolidation	2,266	1815	586
Edema	1,115	925	263
Emphysema	1,133	1093	290
Fibrosis	992	435	259
Pleural Thickening	1800	1143	442
Hernia	112	86	29
No disease	40,411	9861	10089
No. of Images	**83643**	**37067**	**20827**

For the bounding boxes (which were annotated for only 984 Chest Xray images), essential transformations are carried out to meet the specific requirements of different models. In the case of YOLOv8, the bounding box specifications are converted from the standard (x, y, width, height) format to a format focused on the center coordinates (x_center, y_center) along with width and height. Additionally, scaling plays a critical role in this phase. Both the images and their corresponding bounding box coordinates are scaled relative to the model's expected input size. This normalization is crucial for maintaining consistency across different image sizes and dimensions, which is vital for the accurate detection and classification of thoracic pathologies.

The preprocessed data is formatted into structures that are compatible with YOLOv8 classification and detection. This conversion involves organizing the image data and annotations in a way that can be efficiently interpreted and utilized by the models during training for both image classification and object detection.

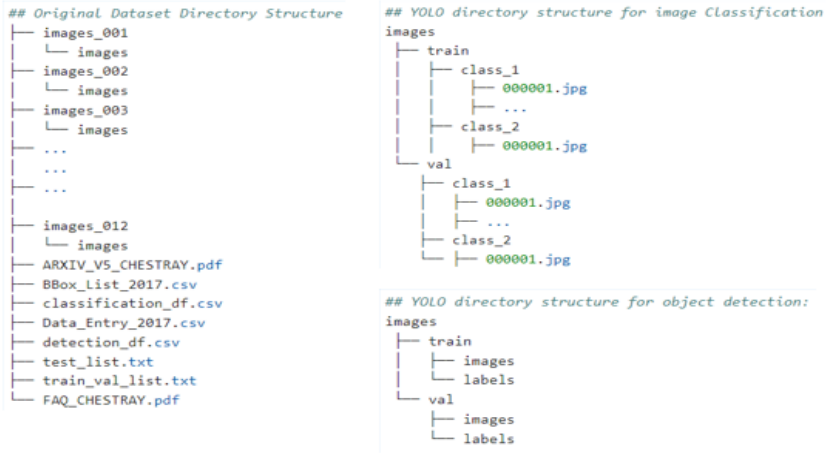

Figure 6. Data directory structure for YOLOv8 Classification and Detection.

5.3. Model

The YOLOv8 architecture was utilized for training two distinct models for thoracic pathology detection and classification. The detection model was trained using a dataset of fewer than 1000 images, annotated with bounding boxes to identify 8 different thoracic pathologies. This model underwent rigorous training and evaluation over more than 70 epochs, with continuous monitoring of metrics such as precision and recall. Conversely, the classification model was trained on a significantly larger dataset of 112,120 high-resolution frontal-view X-ray images, representing 14 distinct pathologies. This model's performance was evaluated over 100 epochs, with insights drawn from detailed performance graphs. Both models leveraged YOLOv8's advanced features to optimize accuracy and efficiency in real-time pathology identification.

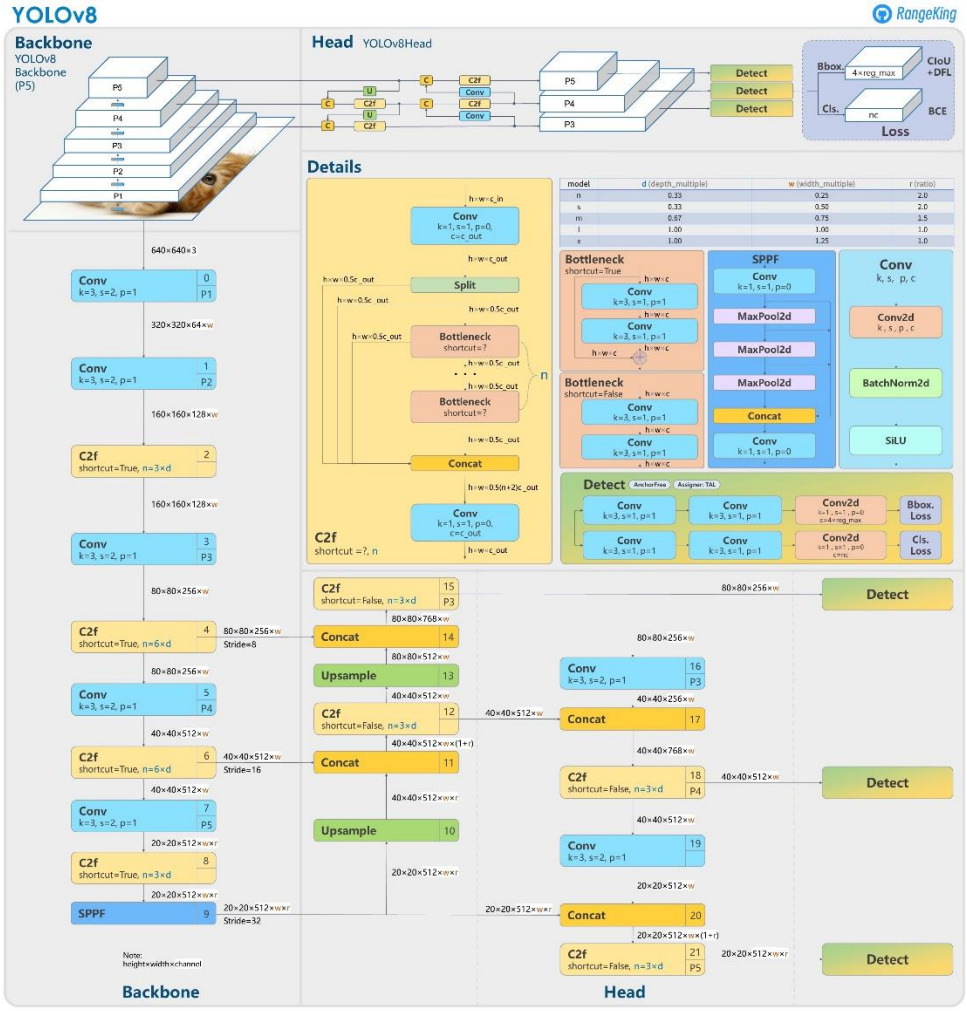

Figure 7. YOLOv8 Architecture

YOLOv8 represents a significant advancement in the YOLO series, specifically designed for tasks such as object detection, classification, and instance segmentation. This latest iteration by Ultralytics introduces several architectural refinements that not only enhance performance but also increase the model's efficiency, making it highly suitable for real-time applications.

One of the key features of YOLOv8 is its anchor-free detection mechanism, which simplifies the model by eliminating the need for predefined anchor boxes. This modification leads to a more streamlined detection process and reduces the computational complexity. The anchor-free approach directly predicts the center

c_x, c_y of an object with a bounding box defined as (b_x, b_y, b_w, b_h) where (b_x, b_y) is the top-left corner of the box, and (b_w, b_h) are its width and height. This is expressed mathematically as:

$$c_x = b_x + \frac{b_w}{2}, \qquad c_y = b_y + \frac{h}{2}$$

This change enhances the model's ability to generalize better across different scenarios by focusing on the geometric center of objects.

6. Results and Discussion

6.1. YOLOv8 Detection Model

The YOLO detection model was rigorously trained on a dataset comprising fewer than 1000 images with 8 distinct thoracic pathologies, each annotated with bounding boxes to facilitate precise object localization. The training process was monitored across various metrics, including losses and performance indicators such as precision and recall, with the results graphically represented over the span of >70 epochs. Notably, the training and validation losses, subdivided into box loss, class loss, and object loss (train/box_loss, train/cls_loss, train/df_loss, and their validation counterparts), displayed a consistent downward trend, indicative of the model's improving accuracy in detecting and classifying objects within the images as training progressed.

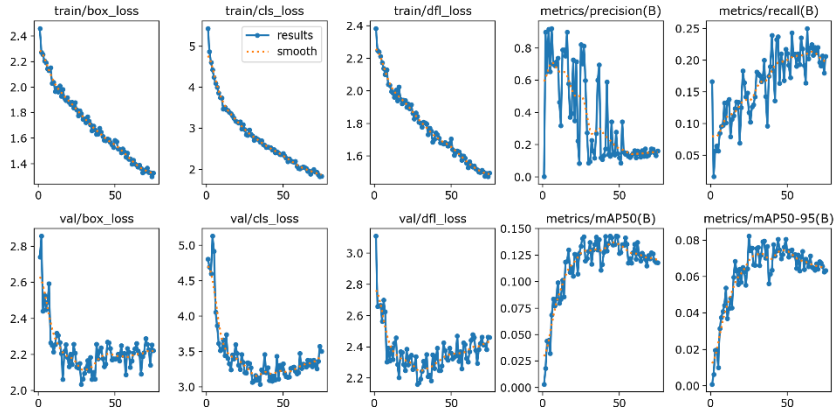

Figure 8. YOLO detection results.

The train/box_loss and val/box_loss, which measure the error in predicting the bounding box coordinates relative to the ground truth, declined sharply, suggesting enhanced spatial recognition capabilities of the model. Similarly, train/cls_loss and val/cls_loss exhibited a decrease, reflecting the model's growing proficiency in accurately classifying the objects within those bounding boxes. However, fluctuations in val/df_loss hint at challenges the model faced in differentiating objects from the background, underscoring the complexity of the task given the limited dataset size.

The metrics of precision and recall for both training and validation phases, along with the mean average precision scores (mAP@0.5 and mAP@0.5:0.95), provide a more nuanced insight into model performance. While precision saw considerable variance, indicating variability in the model's ability to correctly identify positive samples, recall demonstrated a gradual increase, affirming the model's enhanced sensitivity in detecting all relevant objects. Noteworthy is the improvement in mean average precision, particularly mAP@0.5, which showed a significant enhancement, supporting the model's utility in practical scenarios where accurate detection is critical. These trends collectively suggest a promising direction for further refining the model's capabilities, potentially through enriched training data or advanced architectural adjustments.

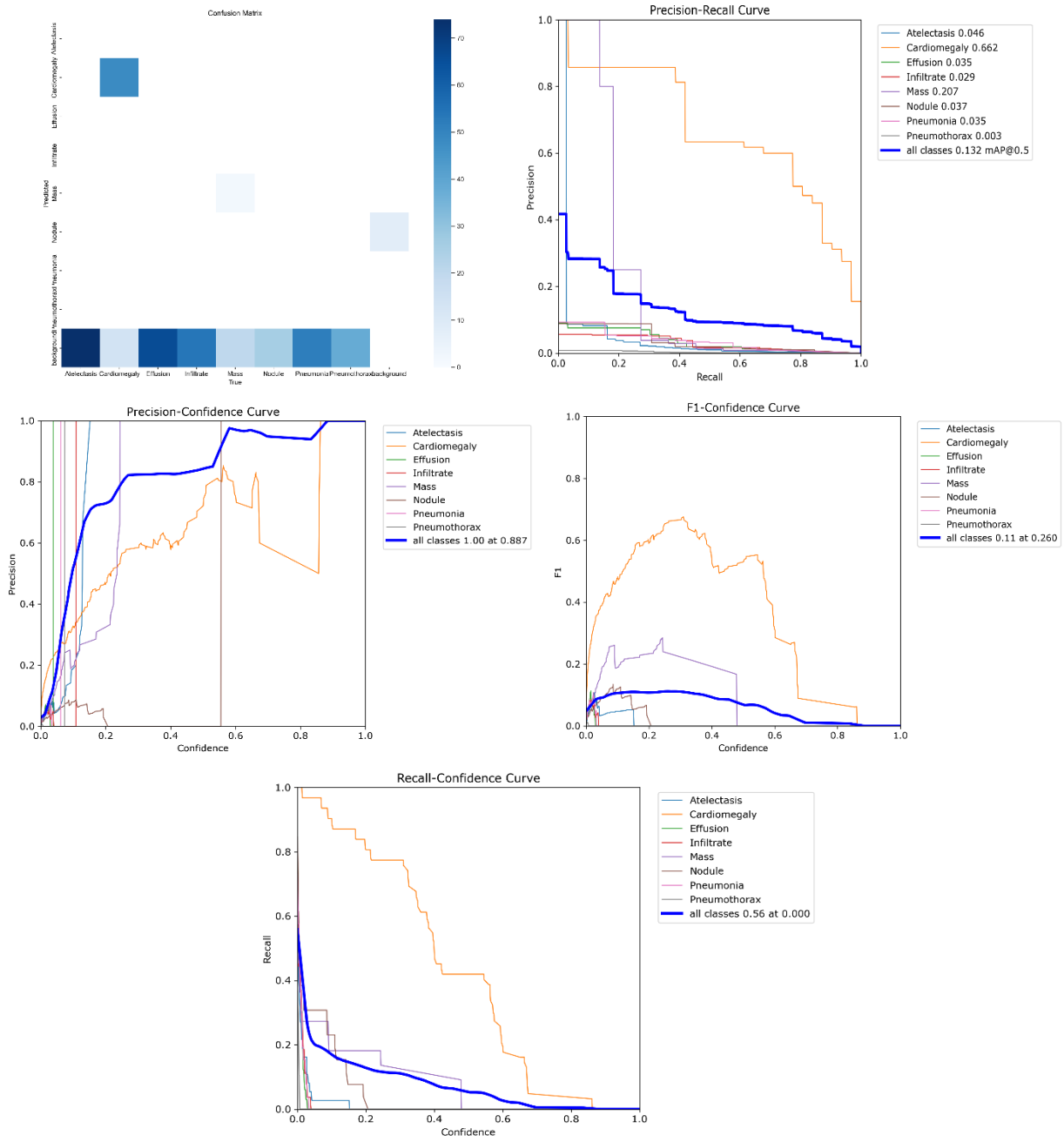

Figure 9. Model Performance Metrics for YOLO Detection on Thoracic Pathologies

The performance metrics above illustrate the nuanced performance of the YOLO model across multiple thoracic conditions. The confusion matrix indicates a strong tendency for the model to correctly identify 'No Finding' but also reveals significant misclassifications, particularly among less prevalent classes such as 'Nodule' and 'Pneumothorax', suggesting that the model struggles with distinguishing between less common pathologies. The Precision-Recall curves further delineate this challenge, showing generally low precision and recall for most conditions except 'Cardiomegaly', which enjoys higher precision yet moderate recall, indicating a reliable detection but limited sensitivity across the dataset.

Precision-Confidence and F1-Confidence curves offer deeper insights into the model's confidence thresholds. Notably, while 'Cardiomegaly' maintains high precision across confidence levels, other classes

see a stark drop in both precision and F1 score as confidence decreases. This characteristic underscores a critical trade-off in the model's confidence calibration; higher thresholds may yield reliable predictions for certain conditions but at the cost of overall sensitivity and F1 scores, particularly for rarer conditions. Such results highlight the need for refined training or post-processing techniques to balance detection sensitivity with specificity, especially in a medical imaging context where accurate and confident predictions are paramount.

6.2. YOLOv8 Classification Model

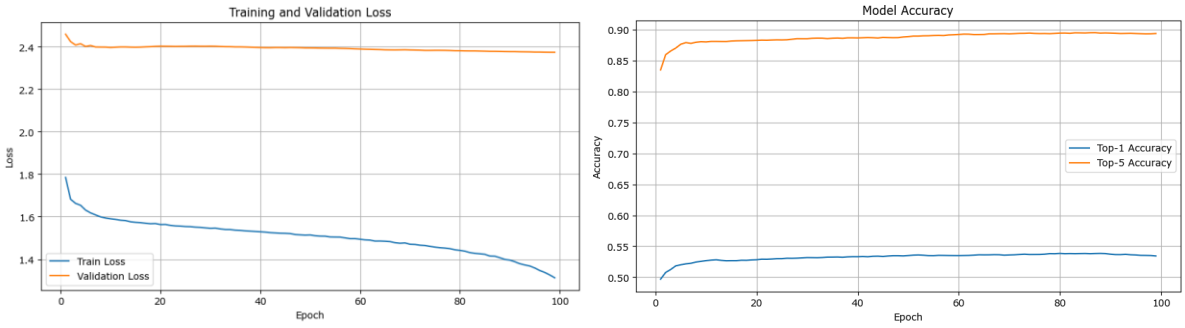

Figure 10. Plots of Training and Validation Loss

The performance graphs from the classification model trained over 100 epochs on a comprehensive dataset of 112,120 high-resolution frontal-view X-ray images and 14 distinct thoracic pathologies reveal several critical insights. The Training and Validation Loss graph shows a significant decline in training loss over the epochs, suggesting effective learning and model fitting to the dataset. However, the validation loss remains relatively flat and high after an initial drop, indicating potential issues such as overfitting or a mismatch in the distribution between the training and validation sets.

The Model Accuracy graph illustrates that while the Top-5 accuracy is impressively high and stable across epochs, indicating that the true class is almost always within the model's top five predictions, the Top-1 accuracy remains relatively low. This disparity suggests that while the model is generally good at narrowing down the possible classes to the correct answers, pinpointing the exact class directly is challenging, possibly due to the similarities among the different pathologies.

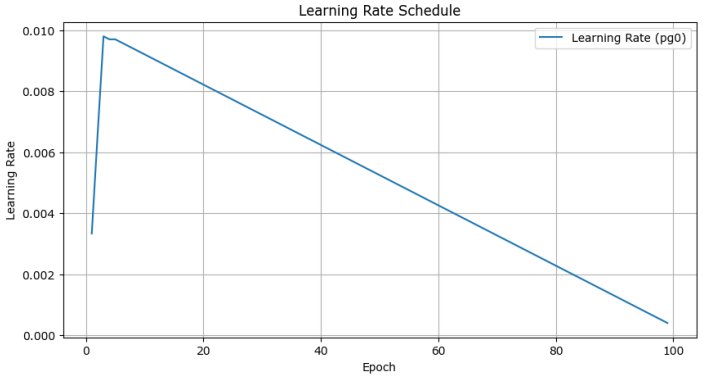

Figure 11. Learning rate Schedule.

The Learning Rate Schedule graph presents a steady decrease in the learning rate, following a planned decay schedule. This controlled reduction helps in fine-tuning the model's weights as training progresses, likely aiding in stabilizing the training convergence as seen in the other metrics. These observations collectively emphasize the complexity of the task at hand and the model's ability to learn general patterns effectively, though they also highlight the need for further tuning or possibly enhancing the model architecture to improve direct class identification accuracy.

7. Conclusion

The project demonstrated the feasibility of using the YOLOv8 deep learning architectures to detect and classify thoracic pathologies in chest X-ray images effectively. The YOLO-based object detection model, despite the limited number of annotated images, showed promising preliminary results in accurately identifying and localizing thoracic diseases within the eight trained classes, illustrating the model's effectiveness in handling specific pathologies with available bounding box annotations. On the other hand, the classification model, trained on the full dataset, achieved high top-5 accuracy, indicating its proficiency in identifying the correct pathologies among the top probable categories. However, the top-1 accuracy remains an area for improvement, reflecting the ongoing challenge of distinguishing between closely related thoracic conditions. The exploratory data analysis revealed critical insights into demographic impacts on disease presentation, which were incorporated into the training strategy to mitigate bias and enhance model generalizability. Future initiatives will focus on expanding the annotated dataset for object detection and refining the classification model to boost its diagnostic precision, ultimately aiming to support radiologists in making more accurate and timely assessments.

References

Crawford, C. 2018. *NIH Chest X-rays Dataset*. National Institutes of Health Chest X-ray Dataset and 1 Collaborator. Available at: https://www.kaggle.com/datasets/nih-chest-xrays/data [3rd May 2024].

de Paula, M.C.F., Escuissato, D.L., Belém, L.C., Zanetti, G., Souza Jr, A.S., Hochhegger, B., Nobre, L.F. and Marchiori, E., 2015. Focal pleural tumorlike conditions: nodules and masses beyond mesotheliomas and metastasis. *Respiratory medicine*, 109(10), pp.1235-1243.

Eriksson, S., 2013. Emphysema before and after 1963 historic perspectives. COPD: *Journal of Chronic Obstructive Pulmonary Disease*, 10(sup1), pp.9-12.

Herrero, R., Sanchez, G. and Lorente, J.A., 2018. New insights into the mechanisms of pulmonary edema in acute lung injury. *Annals of Translational Medicine*, 6(2).

Matsumoto, T., Kodera, S., Shinohara, H., Ieki, H., Yamaguchi, T., Higashikuni, Y., Kiyosue, A., Ito, K., Ando, J., Takimoto, E. and Akazawa, H., 2020. Diagnosing heart failure from chest X-ray images using deep learning. *International Heart Journal*, *61*(4), pp.781-786.

Mégarbane, B. and Chevillard, L., 2013. The large spectrum of pulmonary complications following illicit drug use: features and mechanisms. *Chemico-biological interactions*, 206(3), pp.444-451.

Murtha, L.A., Schuliga, M.J., Mabotuwana, N.S., Hardy, S.A., Waters, D.W., Burgess, J.K., Knight, D.A. and Boyle, A.J., 2017. The processes and mechanisms of cardiac and pulmonary fibrosis. *Frontiers in physiology*, 8, p.777.

Peng, R., Zhang, L., Zhang, Z.M., Wang, Z.Q., Liu, G.Y. and Zhang, X.M., 2020. Chest computed tomography semi-quantitative pleural effusion and pulmonary consolidation are early predictors of acute pancreatitis severity. *Quantitative imaging in medicine and surgery*, 10(2), p.451.

Rakshit, S., Saha, I., Wlasnowolski, M., Maulik, U. and Plewczynski, D., 2019. Deep learning for detection and localization of thoracic diseases using chest x-ray imagery. In Artificial Intelligence and Soft Computing: 18th International Conference, ICAISC 2019, Zakopane, Poland, June 16–20, 2019, Proceedings, Part II 18 (pp. 271-282). *Springer International Publishing*.

Restrepo, R.D. and Braverman, J., 2015. Current challenges in the recognition, prevention and treatment of perioperative pulmonary atelectasis. *Expert review of respiratory medicine*, 9(1), pp.97-107.

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. and Summers, R.M., 2017. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097-2106).

Wong, M., Reyes, J., Lapidus-Krol, E., Chiang, M., Humpl, T., Al-Faraj, M., Ryan, G. and Chiu, P.P., 2018. Pulmonary hypertension in congenital diaphragmatic hernia patients: prognostic markers and long-term outcomes. *Journal of pediatric surgery*, 53(5), pp.918-924.

Zhou, L., Yin, X., Zhang, T., Feng, Y., Zhao, Y., Jin, M., Peng, M., Xing, C., Li, F., Wang, Z. and Wei, G., 2021. Detection and semiquantitative analysis of cardiomegaly, pneumothorax, and pleural effusion on chest radiographs. *Radiology: Artificial Intelligence*, 3(4), p.e200172.

Zhou, S., Zhang, X., & Zhang, R. 2019. Identifying Cardiomegaly in ChestX-ray8 Using Transfer Learning. *Studies in health technology and informatics*, *264*, 482–486. https://doi.org/10.3233/SHTI190268.