ISSN 2007-9737

Radiology-Assisted Detection of Pleural Effusion Using a Balanced Baseline Model

Jorge Emmanuel Zamora Zamora[1], Rodolfo Salgado-Rivera[1], Gustavo Adolfo Alonso-Silverio[2], Elias Ventura-Molina[3], Cornelio Yáñez-Márquez[1,*], Antonio Alarcón-Paredes[1]

[1] Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

[2] Universidad Autónoma de Guerrero,
Facultad de Ingeniería,
Mexico

[3] Instituto Politécnico Nacional,
Centro de Innovación y Desarrollo Tecnológico en Cómputo,
Mexico

jzamoraz2024, rsalgador2024, cyanez, aalarcon}@cic.ipn.mx gsilverio@uagro.mx,
eventuram@ipn.mx

Abstract. Pleural effusion is a frequent and clinically relevant finding on chest radiographs in acute and critical care. An automated detection pipeline was developed using transfer learning with DenseNet121 on a strictly balanced NIH ChestX-ray14 cohort (3,955 Pleural Effusion / 3,955 No Finding; 7,910 images) partitioned patient-wise into 70% training (5,933), 10% validation (665), and 20% test (1,312). Training used restrained augmentation (RandomZoom 0.1, RandomContrast 0.1); validation and test images were not augmented. After feature-extraction pretraining, targeted fine-tuning was applied to the upper DenseNet121 blocks. Predictions were obtained from softmax class scores and metrics were computed from the resulting confusion matrices. On the test set, results were accuracy 0.824, precision 0.877, recall 0.755, specificity 0.893, and F1 0.811 (TP 497, FN 161, FP 70, TN 584). On the validation set, results were accuracy 0.824, precision 0.853, recall 0.758, specificity 0.883, and F1 0.803 (TP 238, FN 76, FP 41, TN 310). Grad-CAM analyses highlighted saliency over costophrenic recesses, basal opacities, and meniscus-like contours in correctly classified positives. These findings indicate stable discrimination with an operating profile that favors specificity, establishing a transparent and reproducible baseline for pleural-effusion detection on chest radiographs.

Keywords. Pleural Effusion detection, chest radiograph, transfer learning, denseNet121, CLAHE, Grad-CAM, medical image classification.

1 Introduction

Chest radiography is a first-line imaging modality in emergency, inpatient, and intensive care services because of its speed, broad availability, and favorable cost profile [9]. Within this landscape, pleural effusion is a frequent and clinically consequential finding that arises in heart failure, infection, malignancy, acute respiratory distress, and postoperative care. Timely recognition on chest X-ray informs triage, ventilatory decisions, and selection of downstream diagnostic tests [22, 26, 11]. Specific etiology remains a challenge, as bacterial profiles and culture outcomes vary significantly, necessitating robust diagnostic support [17]. Moreover, in trauma settings or complex pediatric cases, accurate detection is critical to manage complications like pneumothorax or empyema [25, 1].

Indeed, the broader landscape of healthcare is undergoing a transformation where artificial intelligence serves as a pivotal auxiliary tool, reducing analysis time and aiding decision-making across various domains, from surgery to diagnostics [16, 12]. The centrality of radiographic assessment persists even when advanced modalities are available, since radiographs often guide early decisions and monitoring in high-throughput settings [24].

The clinical burden of pleural effusion is reflected in practice patterns and outcomes research across critical and complex inpatient populations. Portable chest radiography is routinely used in intensive care to monitor respiratory status, including Acute Respiratory Distress Syndrome (ARDS), and support procedural decisions, which makes consistent interpretation pivotal for safe throughput and resource use [20, 23].

Observational cohorts document how pleural fluid assessment and drainage intersect with respiratory trajectories and complications in multimorbidity, highlighting the need for reproducible imaging and follow-up [7, 5]. Contemporary clinical reviews emphasize structured diagnostic approaches that integrate radiographic signs with context to mitigate delays and misclassification in busy pathways [22]. Historically, diagnosis on chest radiographs has relied on expert evaluation of costophrenic angle blunting, meniscus configuration, hemithorax opacification, and mediastinal displacement, interpreted alongside clinical information [22]. Performance is sensitive to projection and positioning, particularly with supine anteroposterior portable studies, and to overlapping parenchymal processes that can mask or mimic pleural fluid, which contributes to reader variability and site-to-site differences [24].

Advances in machine learning have established deep convolutional neural networks as strong image-level classifiers for chest radiography.

Foundational work demonstrated the feasibility of transfer learning from natural-image pretraining, followed by task-specific adaptation to thoracic findings under clinical protocols [9, 3]. This versatility of Convolutional Neural Networks (CNNs) has been corroborated in neuroimaging, where architectures like MobileNet have successfully detected complex conditions such as multiple sclerosis using MRI data [19], further validating the transfer learning paradigm in medical image analysis. In pleural effusion, recent studies report accurate detection with transfer learning and fine-tuning and introduce severity-aware objectives that better align outputs with management needs [8, 21]. Representation learning tailored to chest radiographs further suggests gains in data efficiency and robustness where labels are scarce or heterogeneous [2, 27].

Several challenges persist in this sense. Reported performance depends on dataset composition, label provenance, and evaluation choices, which complicates reproducibility and external validity across institutions and acquisition settings [13, 24]. Furthermore, a significant gap remains in evaluating the reliability of model predictions. Recent research emphasizes the importance of confidence calibration methods, such as Temperature Scaling and Mixup, to ensure that deep learning models in medical imaging are not just accurate but also trustworthy in their confidence estimates [14].

Many pipelines prioritize binary detection rather than clinically informative stratification of burden or laterality that would support decision-making and follow-up [8]. While our work focuses on data curation, other approaches emphasize advanced optimization strategies. For instance, recent studies have introduced novel metaheuristic algorithms for spectral band selection to reduce dimensionality and improve classification accuracy in hyperspectral imaging [15], highlighting the diverse methodological avenues to enhance model performance.

Differences between portable and standard acquisitions remain insufficiently characterized in model assessment, limiting guidance for deployment in emergency and critical care pathways [20, 27]. Comparative investigations focused on pleural effusion that systematically evaluate architectures and fine-tuning regimes under consistent, clinically aligned metrics are comparatively limited relative to broader thoracic classification efforts [18] [10].

The purpose of this study is to develop and rigorously evaluate a deep learning pipeline for pleural effusion detection on chest radiographs.

We establish a methodologically transparent baseline using a DenseNet121 architecture with transfer learning, focusing on balanced data curation, patient-wise partitioning, and a two-stage

fine-tuning protocol to ensure the results are both robust and reproducible [27, 8, 13].

1.1 Related Work

Computer-aided diagnosis in chest radiography has rapidly advanced due to deep learning.

Our research builds upon three key areas of existing literature: the application of transfer learning to medical imaging, specific advances in pleural effusion detection, and the persistent methodological challenges that motivate our work.

Early applications of deep learning to chest radiography established that convolutional neural networks pretrained on natural images could be adapted to detect thoracic findings directly from pixel data, introducing transfer learning as a practical strategy for limited medical labels [3, 6, 10]. Subsequent surveys and methodological reviews synthesized progress across datasets, architectures, and evaluation practices, emphasizing clear reporting of label derivation, standardized metrics, and validation on data unseen during model development to enable clinical translation [16, 27, 4]. Comparative studies further showed that careful transfer learning with progressive unfreezing tends to outperform training from scratch on chest X-ray classification, while revealing sensitivity to data curation choices and partitioning protocols that can affect reproducibility [2, 21].

Within pleural effusion specifically, two lines of research are prominent. The first targets binary detection in multi-label classifiers trained on public benchmarks such as ChestX-ray14, CheXpert, or MIMIC-CXR, where pleural effusion appears as a frequent label and benefits from ImageNet-initialized encoders with task-adapted augmentation and class-balancing [13, 14, 15]. Across this work, DenseNet and related modern backbones remain common baselines [2]. The second stream seeks more clinically aligned supervision beyond mere presence or absence. For instance, Huang et al. proposed a transfer-learning framework that estimates effusion severity, reporting that severity-aware targets can improve decision support compared with binary endpoints alone [8]. Similarly, multi-finding pipelines that include pneumothorax and pleural effusion demonstrate that shared representations across related findings can be effective, although evaluation becomes more complex as inter-label correlations grow [19].

Despite these advances, dataset provenance and labeling strategy are recurring determinants of external validity. Studies examining annotation granularity caution that weak or coarse supervision may limit generalization, particularly for findings like pleural effusion where laterality and burden influence management [21, 18]. This is compounded by the inherent limitations of radiography, which has interpretive pitfalls compared to other modalities [22, 26]. Furthermore, real-world acquisition variability, especially from portable imaging in intensive care, modifies finding appearance and view availability, motivating assessments that stratify by projection and device [24, 13].

To address the reliance on extensive manual labels, representation learning has emerged as a complementary direction. Self-supervised pretraining tailored to chest X-rays has shown gains in downstream classification, including pleural effusion, when labels are scarce [Huang et al., 2022]. Synthesizing these strands, the literature supports transfer-learning pipelines for pleural effusion detection but calls for evaluation protocols that account for label quality, acquisition variability, and external validation to ensure performance persists beyond the development cohort [27, 8, 19, 21, 2]. Our study addresses this gap by establishing a transparent and reproducible baseline using a strictly balanced cohort and patient-disjoint data splits.

2 Materials and Methods

The end-to-end experimental workflow for this study is summarized in Figure 1. This block diagram provides a high-level overview of the four sequential stages of the methodology: (1) the curation of a balanced dataset from the NIH ChestX-ray14 collection and its subsequent patient-wise partitioning; (2) the standardized image preprocessing and augmentation pipeline; (3) the two-stage training and fine-tuning protocol for the DenseNet121 model; and (4) the final quantitative and qualitative evaluation on the held-out test set.

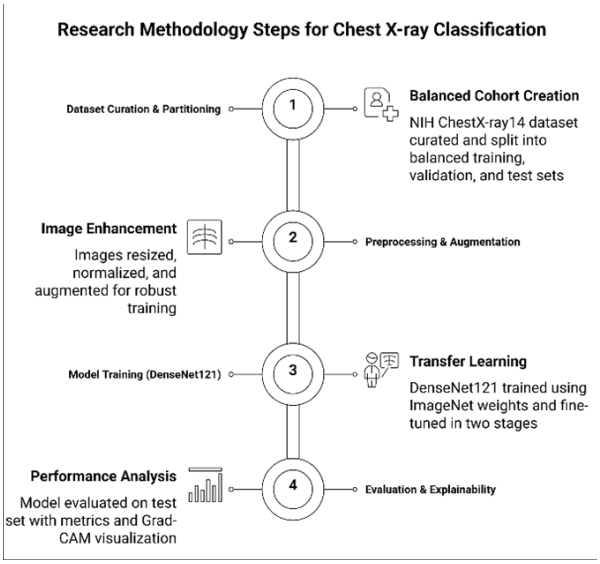

Fig. 1. Overview of the Experimental Workflow. High-level diagram of experimental workflow. The process comprises four main stages: (1) dataset curation and partitioning, (2) image preprocessing and augmentation, (3) model training, and (4) final performance evaluation

The following subsections detail each of these stages.

2.1 Dataset Curation and Partitioning

The study utilized the NIH ChestX-ray14 collection [18]. A balanced binary dataset was created by selecting all radiographs annotated exclusively with the label “Effusion” and an equal number of images labeled as “No Finding.” After removing any duplicate files, the complete cohort was partitioned patient-wise to prevent any subject overlap across subsets, ensuring an unbiased evaluation of the model’s generalization capability.

The dataset was deterministically split (seed = 45) into three disjoint sets: 70% for training (5,933 images), 10% for validation (665 images), and 20% for testing (1,312 images) Class balance was strictly maintained in the initial dataset definition, and labels were assigned as Pleural Effusion = 1 (positive) and No Finding = 0 (negative).

2.2 Image Preprocessing and Augmentation

A standardized preprocessing pipeline was applied to all images. First, contrast enhancement was performed offline using Contrast-Limited Adaptive Histogram Equalization (CLAHE) with a clip limit of 2.0 and a tile grid of 8×8. This operation enhances local contrast, which is particularly useful for accentuating the subtle radiographic signs of pleural effusion, such as meniscus-like fluid lines and opacities near the lung bases.

Following enhancement, all 1024×1024 PNG images were resized to the model’s expected input dimensions of 224×224 pixels. Since the pretrained model requires a three-channel input, the single-channel grayscale images were replicated across three channels. Finally, a standard preprocessing function specific to the DenseNet architecture was applied.

Data augmentation was applied exclusively to the training set to promote model robustness and reduce overfitting. The augmentations were limited to moderate geometric and photometric transformations, specifically RandomZoom (factor = 0.1) and RandomContrast (factor = 0.1). No augmentations were applied to the validation or test images.

2.3 Model Architecture and Training

2.3.1 Model Architecture

The core of the model is the DenseNet121 architecture, pretrained on the ImageNet dataset. This architecture was selected due to its parameter efficiency and demonstrated high performance in previous chest radiograph analysis benchmarks.

The top classification layer of the original model was removed, and its convolutional base was used as a feature extractor. A new classification head was added on top, consisting of a series of Dense, Batch Normalization, and Dropout layers, culminating in a final Dense layer with a single neuron and a sigmoid activation function for binary classification. L2 regularization was applied to the dense layers to mitigate overfitting.

2.3.2 Two-Stage Training Protocol

A two-stage fine-tuning procedure was implemented to effectively adapt the pretrained model to the task of pleural effusion detection. The initial stage consisted of feature extraction, where all layers of the DenseNet121 base were kept frozen. During this phase, only the newly added classification head was trained using the Adam optimizer with a learning rate of 1×10^{-4}. This approach allows the model to learn task-specific patterns from the new data without disrupting the robust, general-purpose features learned during pretraining.

Following the initial phase, the process transitioned to fine-tuning. For this second stage, the top 60 layers of the DenseNet121 base were unfrozen, and training resumed with a significantly reduced learning rate of 1×10^{-6}. This step allows the model to subtly adjust its higher-level feature representations to be more specific to radiographic images, while the stable, low-level features in the earlier layers remain preserved. Throughout both stages, training was monitored using a suite of callbacks, including ModelCheckpoint to save the model with the highest validation accuracy, EarlyStopping to prevent overfitting, and ReduceLROnPlateau to adjust the learning rate if performance stagnated.

2.4 Evaluation and Explainability

The model checkpoint that achieved the highest validation accuracy was selected for final evaluation on the held-out test set. Performance was quantified using standard classification metrics: accuracy, precision, sensitivity (recall), specificity, F1-score, and ROC AUC, all derived from the confusion matrix.

To provide insight into the model's decision-making process, Gradient-Weighted Class Activation Mapping (Grad-CAM) was employed as a post-hoc explainability technique. Grad-CAM generates heatmaps that highlight the image regions most influential in each prediction. These maps were used for qualitative inspection to verify that the model's attention aligned with clinically relevant radiological signs of pleural effusion

Table 1. Validation and Test Performance

Metric	Validation	Test
Accuracy	0.8241	0.8240
Precision	0.8530	0.8770
Sensitivity (Recall)	0.7580	0.7550
Specificity	0.8830	0.8930
F1-Score	0.8030	0.8110
ROC AUC	0.8241	0.8205

2.5 Computational Resources

All experiments were conducted on a workstation equipped with an Intel Core i7 (13th Gen) CPU, 32 GB of RAM, and an NVIDIA GeForce RTX 4070 GPU. The software stack included Python with TensorFlow, Scikit-learn, and OpenCV libraries.

3 Results and Discussions

3.1 Metric Definitions and Formulas

Let TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives) denote confusion matrix counts, with Pleural Effusion as the positive class. Predicted labels were obtained via argmax over softmax scores. Metrics were computed as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{1}$$

$$\text{Sensitivity} = \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4}$$

$$\text{F1-Score} = \frac{2\,\text{Precision}\,\text{Recall}}{\text{Precision} + \text{Recall}}, \tag{5}$$

$$\text{F1-Score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \tag{6}$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{7}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}, \tag{8}$$

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d\text{FPR}. \tag{9}$$

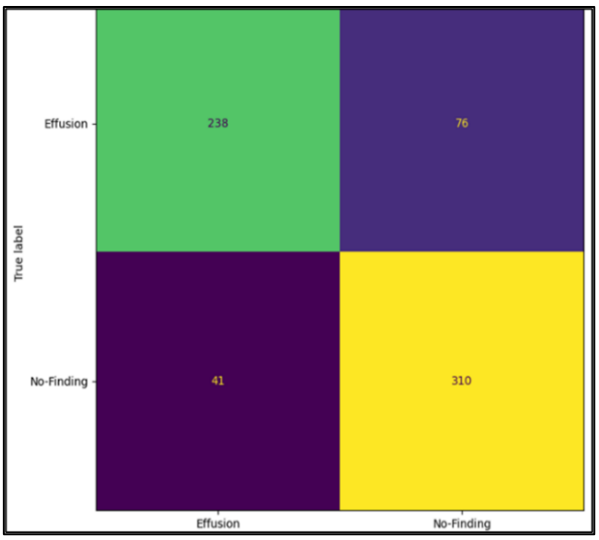

Fig. 2. Confusion matrix validation set

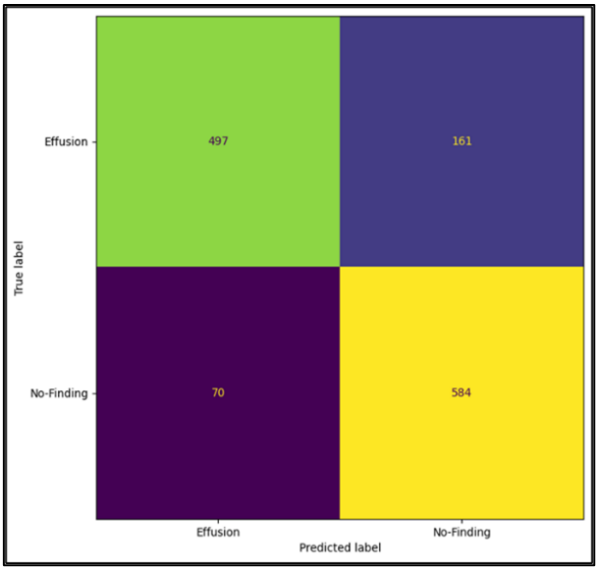

Fig. 3. Confusion matrix test set

3.2 Quantitative Performance

Performance between the validation and test sets remained highly stable, with negligible changes in accuracy (0.8241 vs 0.8240) and F1-score (0.8030 vs 0.8110), evidencing strong generalization.

3.3 Confusion Matrix Analysis

Validation confusion matrix (Fig. 2): TP = 238, FN = 76, FP = 41, TN = 310. Test confusion matrix (Fig. 3): TP = 497, FN = 161, FP = 70, TN = 584.

Notably, the distribution of errors remains stable across both sets, confirming the model's robustness against overfitting. The high proportion of True Negatives (584 in the test set) reinforces the system's reliability in ruling out pathology, minimizing the risk of "alarm fatigue" in high-throughput clinical environments.

Specificity exceeds Sensitivity on both splits, indicating a decision profile oriented toward limiting false positives while preserving stable discrimination for Pleural Effusion. This emphasis is clinically compatible with reducing unnecessary follow-up in No Finding cases, with the option to recalibrate operating points in settings that prioritize recall.

3.4 Discussion

A deep learning model for pleural effusion detection was successfully developed, achieving stable and generalizable performance (accuracy ≈ 0.824) on a held-out test set. Beyond the performance metrics, the primary contribution of this work is the establishment of a transparent and reproducible baseline that directly addresses several key challenges identified in existing literature.

First, the critical issue of reproducibility was addressed. While many studies report high performance, those results are often difficult to replicate due to inconsistencies in data curation and partitioning. In this study, the use of a strictly balanced cohort and an explicit patient-wise 70/10/20 partition ensured that the model was evaluated on data from patients entirely unseen during training. This methodological rigor prevents data leakage and provides a more honest assessment of the model's ability to generalize, a crucial step for any potential clinical translation. The learning curves (Fig. 4) confirm that the two-stage training strategy achieved stable validation performance without overfitting, further bolstering the reliability of the findings.

Second, while the model was designed for binary classification, its operating profile provides important insights for clinical utility. The confusion

Fig. 4. Training and Validation Learning Curves. Accuracy and loss curves for the training and validation sets. The vertical line marks the transition to the fine-tuning stage (Stage 2). The stable validation performance indicates effective generalization without overfitting

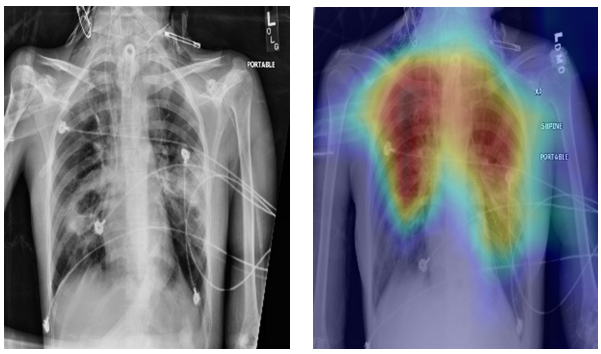

Fig. 5. Grad-CAM Saliency Map Visualization. A true positive case where the heatmap highlights model attention on the right costophrenic recess and lung base, confirming its focus on clinically relevant radiological signs of pleural effusion

matrices (Fig. 2 and Fig. 3) reveal a specificity-leaning profile, meaning the model is conservative in its predictions, thereby reducing false alarms on healthy patients. This trade-off is often desirable in screening workflows to avoid unnecessary follow-up tests. Although this study does not address clinically relevant factors like effusion volume or laterality—a known limitation in many detection pipelines—a robust baseline is now provided upon which such multi-class or severity-aware models can be reliably built and compared. By establishing this model with generalizable and explainable performance, we now have a reliable foundation to address more complex challenges, such as quantifying effusion volume or assessing robustness across different radiographic projections (AP vs. PA). These were identified as critical open problems in our introduction, and this work represents a necessary first step towards solving them.

Finally, qualitative evidence that the model learned clinically relevant features is provided by the Grad-CAM visualizations (Fig. 5). In true positive cases, model attention was correctly localized to the costophrenic recesses and lung bases, where pleural fluid typically accumulates. This alignment with radiological landmarks suggests that the model is not relying on spurious correlations or artifacts, a critical concern given the variability in image acquisition noted in the literature

3.5 Comparison with the State-of-the-Art

The primary objective of this study was to establish a methodologically robust and reproducible baseline. To properly contextualize our model's performance, it is essential to compare it with other works that have addressed the same task on the NIH ChestX-ray14 dataset. Table 2 presents a summary of our results alongside those from relevant state-of-the-art studies.

Recent literature reports very high-performance metrics. For example, [2] achieved an outstanding AUC of 0.99 and an accuracy of 98% for pleural effusion using an EfficientNet-V1-B4 architecture. Similarly, [6] reported 95% accuracy using a modified VGG19 network. While these results are impressive, it is critical to note a key methodological distinction: many of these studies do not explicitly report employing a strict patient-wise partitioning scheme. As highlighted in the literature, this can

Table 2. Comparison with State-of-the-Art Studies

Study	Architecture	Patient-Wise Partitioning	AUC	Accuracy
[20]	EfficientNet-V1-B4	Not Reported	0.99	0.98
[21]	VGG19 (modified)	Not Reported	Not Reported	0.95
This Study	DenseNet121	Yes (Strict)	0.89	0.8240

lead to data leakage and potentially optimistic performance metrics.

Our model achieved an AUC of 0.89 and an accuracy of 0.824, which are competitive results obtained under a strict, reproducible, patient-wise validation protocol. This methodological rigor ensures that our baseline provides a more realistic and reliable estimate of generalization performance on unseen patient data. Therefore, our work serves as a trusted foundation, emphasizing the importance of the trade-off between achieving top-tier metrics and ensuring the methodological soundness required for clinical translation.

3.6 Limitations and Future Work

Several limitations of the present study should be acknowledged, which also provide clear directions for future research.

First, the model was developed and validated using data from a single public source, the NIH ChestX-ray14 dataset. Although this dataset is large and widely used, the model's performance on external data from different clinical sites, patient populations, or imaging hardware remains unevaluated. Future work should therefore focus on validating the model's external generalizability across diverse institutional settings.

Second, the implemented pipeline is limited to binary classification, predicting only the presence or absence of pleural effusion. This approach does not capture clinically significant parameters such as the volume of the effusion (e.g., small, moderate, or large) or its laterality (unilateral vs. bilateral), which are often critical for patient management. A natural extension of this research would be the development of multi-class models capable of quantifying effusion severity.

Finally, the analysis did not stratify performance based on the radiographic projection type (e.g., posteroanterior vs. anteroposterior portable views). The radiographic appearance of pleural effusion can vary significantly between these projections, potentially impacting model performance. Future studies could investigate the model's robustness across different acquisition conditions to better characterize its utility in varied clinical workflows, such as intensive care units where portable radiography is common

4 Conclusions

In this study, a deep learning model for pleural effusion detection was developed and rigorously evaluated. The primary contribution, however, is not merely the achievement of stable and generalizable performance (accuracy $\approx$ 0.824), but the establishment of a transparent and methodologically sound baseline that addresses a critical need for reproducibility in the field.

This was accomplished through a series of deliberate methodological choices. The use of a strictly balanced "Pleural Effusion vs. No Finding" cohort minimized the class imbalance bias common in clinical datasets. Furthermore, a patient-wise partitioning scheme was strictly enforced to prevent any data leakage between the training, validation, and test sets, ensuring an unbiased assessment of the model's performance on unseen patient data. The clearly defined two-stage training protocol, combining feature extraction with targeted fine-tuning, further contributes to the reproducibility of these results.

Qualitative validation through Grad-CAM analysis supported the clinical plausibility of the learned representations, showing that the model's attention correctly localized on relevant radiological signs. It is hoped that this methodologically transparent baseline will serve as a useful reference for future studies. Potential next steps could include extending the model to assess effusion severity and laterality, or undertaking the crucial external multicenter validation needed to evaluate its potential for clinical translation.

Acknowledgments

The authors gratefully acknowledge the support of the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) and the Laboratorio de Cómputo Inteligente del Centro de Investigación en Computación, Instituto Politécnico Nacional (CIC-IPN) also the Faculty of Engineering, of UAGro and the Centro de Innovación y Desarrollo Tecnológico en Cómputo, Instituto Politécnico Nacional (CIDETEC-IPN), whose resources and guidance were essential for completing this study.

References

1. **Abubakar Fi, Rufai Ai, Ahmed Hk, Adamu A, Ukwuni Si (2024).** Pattern and outcome of paediatric pleural effusion seen at Usmanu Dafodiyo University Teaching Hospital, Sokoto: A 5-year retrospective study. Ibom Medical Journal, Vol. 17, No. 1, pp. 42–48. DOI: 10.61386/imj.v17i1.375.
2. **AbuKaraki, A., Alrawashdeh, T., Abusaleh, S., Alksasbeh, M. Z., Alqudah, B., Alemerien, K., Alshamaseen, H. (2024).** Pulmonary Edema and Pleural Effusion Detection Using EfficientNet-V1-B4 Architecture and AdamW Optimizer from Chest X-Rays Images. Computers, Materials & Continua, Vol. 80, No. 1, pp. 1055–1073. DOI: 10.32604/cmc.2024.051420.
3. **Chow, L. S., Tang, G. S., Solihin, M. I., Gowdh, N. M., Ramli, N., Rahmat, K. (2023).** Quantitative and Qualitative Analysis of 18 Deep Convolutional Neural Network (CNN) Models with Transfer Learning to Diagnose COVID-19 on Chest X-Ray (CXR) Images. SN Computer Science, Vol. 4, No. 2, pp. 141. DOI: 10.1007/s42979-022-01545-8.
4. **Daoud, M. I., Alrahahleh, Y., Abdel-Rahman, S., Alsaify, B. A., Alazrai, R. (2021).** COVID-19 Diagnosis in Chest X-ray Images by Combining Pre-trained CNN Models with Flat and Hierarchical Classification Approaches. 2021 12th International Conference on Information and Communication Systems (ICICS), IEEE, Valencia, Spain, pp. 330–335. DOI: 10.1109/ICICS52457.2021.9464532.
5. **Duanmu, H., Ren, T., Li, H., Mehta, N., Singer, A. J., Levsky, J. M., Lipton, M. L., Duong, T. Q. (2022).** Deep learning of longitudinal chest X-ray and clinical variables predicts duration on ventilator and mortality in COVID-19 patients. BioMedical Engineering OnLine, Vol. 21, No. 1, pp. 77. DOI: 10.1186/s12938-022-01045-z.
6. **Fauzan, A. R., Wahyuddin, M. I., Ningsih, S. (2021).** Pleural Effusion Classification Based on Chest X-Ray Images using Convolutional Neural Network. Jurnal Ilmu Komputer dan Informasi, Vol. 14, No. 1, pp. 9–16. DOI: 10.21609/jiki.v14i1.898.
7. **Fjæreide, K. W., Petersen, P. L., Mahdi, A., Crescioli, E., Nielsen, F. M., Rasmussen, B. S., Schjørring, O. L. (2023).** Pleural effusion and thoracentesis in ICU patients: A longitudinal observational cross-sectional study. Acta Anaesthesiologica Scandinavica, Vol. 67, No. 7, pp. 943–952. DOI: 10.1111/aas.14258.
8. **Huang, T., Yang, R., Shen, L., Feng, A., Li, L., He, N., Li, S., Huang, L., Lyu, J. (2022).** Deep transfer learning to quantify pleural effusion severity in chest X-rays. BMC Medical Imaging, Vol. 22, No. 1, pp. 100. DOI: 10.1186/s12880-022-00827-0.
9. **Irmici, G., Cè, M., Caloro, E., Khenkina, N., Della Pepa, G., Ascenti, V., Martinenghi, C., Papa, S., Oliva, G., Cellina, M. (2023).** Chest X-ray in Emergency Radiology: What Artificial Intelligence Applications Are Available?. Diagnostics, Vol. 13, No. 2, pp. 216. DOI: 10.3390/diagnostics13020216.
10. **Khanal, B., Singh, A., Paul, S., Das, R. (2021).** Multi-Class Classification on Chest X-Ray Images Using Convolution Neural Network. 2021 IEEE 18th India Council International Conference (INDICON), IEEE, Guwahati, India, pp. 1–5. DOI: 10.1109/INDICON52576.2021.9691747.
11. **Khodeary, A., Mohamed, T., Alkhayat, K., Sayed, S. (2022).** Bacterial Profile and

Antibiotics Susceptibility Pattern of Pleural Effusion Isolates from Sohag University Hospital. Egyptian Journal of Medical Microbiology, Vol. 31, No. 3, pp. 29–35. DOI: 10.21608/ejmm.2022.247177.

12. **Li, D., Li, S. (2022).** An artificial intelligence deep learning platform achieves high diagnostic accuracy for Covid-19 pneumonia by reading chest X-ray images. iScience, Vol. 25, No. 4, pp. 104031. DOI: 10.1016/j.isci.2022.104031.

13. **Luo, L., Chen, H., Xiao, Y., Zhou, Y., Wang, X., Vardhanabhuti, V., Wu, M., Han, C., Liu, Z., Fang, X. H. B., Tsougenis, E., Lin, H., Heng, P.-A. (2022).** Rethinking Annotation Granularity for Overcoming Shortcuts in Deep Learning–based Radiograph Diagnosis: A Multicenter Study. Radiology: Artificial Intelligence, Vol. 4, No. 5, pp. e210299. DOI: 10.1148/ryai.210299.

14. **López-Miguel, N., Diaz-Hernández, R., Altamirano-Robles, L. (2025).** Confidence Calibration of CNNs in Medical Image Databases. Computación y Sistemas, Vol. 29, No. 1, pp. 7–14. DOI: 10.13053/cys-29-1-5510.

15. **Medjahed, S. A., Boukhatem, F. (2025).** Binary Coronavirus Disease Optimization Algorithm for Spectral Band Selection. Computación y Sistemas, Vol. 29, No. 1, pp. 501–510. DOI: 10.13053/cys-29-1-5054.

16. **Qassim Al-Asiri, M. B., Al-Asmari, A. A. (2024).** Role of Artificial Intelligence in Supporting the Performance of Surgical Robots: A Survey. Computación y Sistemas, Vol. 28, No. 3, pp. 1499–1508. DOI: 10.13053/cys-28-3-5181.

17. **Ranjit, S., Singh, A. K., Shrestha, I., Twayana, A. R., Bhandari, P., Siwakoti, S., Singh, S. (2022).** Negative Pleural Fluid Cultures among Patients with Pleural Effusion in a Tertiary Care Hospital: A Descriptive Cross-sectional Study. Journal of Nepal Medical Association, Vol. 60, No. 249, pp. 461–464. DOI: 10.31729/jnma.6758.

18. **Saleem, H. N., Sheikh, U. U., Khalid, S. A. (2021).** Classification of Chest Diseases from X-ray Images on the CheXpert Dataset. In **Mekhilef, S., Favorskaya, M., Pandey, R. K., Shaw, R. N.**, editors, Innovations in Electrical and Electronic Engineering, Vol. 756. Springer Singapore, Singapore, pp. 837–850. DOI: 10.1007/978-981-16-0749-3_64. Series Title: Lecture Notes in Electrical Engineering.

19. **Saurabh, S., Gupta, P. K. (2024).** Detection and Classification of Multiple Sclerosis from Brain MRIs by Using MobileNet 2D-CNN Architecture. Computación y Sistemas, Vol. 28, No. 3, pp. 1229–1242. DOI: 10.13053/cys-28-3-4197.

20. **Scott, J., Waite, S., Napolitano, A. (2021).** Restricting Daily Chest Radiography in the Intensive Care Unit: Implementing Evidence-Based Medicine to Decrease Utilizationt. Journal of the American College of Radiology, Vol. 18, No. 3, pp. 354–360. DOI: 10.1016/j.jacr.2020.05.035.

21. **Shamrat, F. J. M., Azam, S., Karim, A., Ahmed, K., Bui, F. M., De Boer, F. (2023).** High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. Computers in Biology and Medicine, Vol. 155, pp. 106646. DOI: 10.1016/j.compbiomed.2023.106646.

22. **Shen-Wagner, J., Gamble, C., MacGilvray, P. (2023).** Pleural Effusion: Diagnostic Approach in Adults. American Family Physician, Vol. 108, No. 5, pp. 464–475.

23. **Sjoding, M. W., Taylor, D., Motyka, J., Lee, E., Co, I., Claar, D., McSparron, J. I., Ansari, S., Kerlin, M. P., Reilly, J. P., Shashaty, M. G. S., Anderson, B. J., Jones, T. K., Drebin, H. M., Ittner, C. A. G., Meyer, N. J., Iwashyna, T. J., Ward, K. R., Gillies, C. E. (2021).** Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. The Lancet Digital Health, Vol. 3, No. 6, pp. e340–e348. DOI: 10.1016/S2589-7500(21)00056-X.

24. **Tejani, A., Dowling, T., Sanampudi, S., Yazdani, R., Canan, A., Malja, E., Xi, Y., Abbara, S., Peshock, R. M., Kay, F. U. (2024).** Deep Learning for Detection of Pneumothorax and Pleural Effusion on Chest Radiographs: Validation Against Computed Tomography, Impact on Resident Reading Time, and Interreader Concordance. Journal of Thoracic Imaging, Vol. 39, No. 3, pp. 185–193. DOI: 10.1097/RTI.0000000000000746.

25. **Tran, J., Haussner, W., Shah, K. (2021).** Traumatic Pneumothorax: A Review of Current Diagnostic Practices And Evolving Management. The Journal of Emergency Medicine, Vol. 61, No. 5, pp. 517–528. DOI: 10.1016/j.jemermed.2021.07.006.

26. **Yiu, W., Hannan, L. M., Tacey, M., Suleiman, M., Muruganandan, S. (2025).** Impact of a specialised pleural medicine unit in the management of pleural infection. Internal Medicine Journal, Vol. 55, No. 5, pp. 734–740. DOI: 10.1111/imj.70028.

27. **Çallı, E., Sogancioglu, E., Van Ginneken, B., Van Leeuwen, K. G., Murphy, K. (2021).** Deep learning for chest X-ray analysis: A survey. Medical Image Analysis, Vol. 72, pp. 102125. DOI: 10.1016/j.media.2021.102125.

Article received on 06/09/2025; accepted on 05/01/2026.
**Corresponding author is Cornelio Yáñez-Márquez.*