Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AI-POWERED DIAGNOSIS OF NEONATAL RESPIRATORY DISTRESS SYNDROME: A COMPARATIVE STUDY OF CNN MODELS ON CHEST X-RAYS USING CLINICAL MOROCCAN DATASET

SALMA SEKKAT[1,2,5], MOUNA ZOUINE[3], ZINEB CHEKER[4], OUSSAMA FANGACHI[1], KARIMA SAMMOUD[2], SAAD CHAKKOR[5], ABDALLAH OULMAATI[3], ADIL NAJDI[2]

[1]Department of Informatics and Technologies, Mohammed VI University Hospital of Tangier, Morocco
[2]Department of Epidemiology, Public Health, and Social Sciences, Faculty of Medicine and Pharmacy of Tangier, AbdelMalek Essaâdi University (UAE), Morocco
[3]Department of Neonatology, Mohammed VI University Hospital of Tangier, Morocco
[4]Laboratory of Innovative System Engineering, ENSA of Tetuan, Abdelmalek Essaadi University Tetuan, Morocco
[5]Laboratory of Information and Communication Technologies (LabTIC), National School of Applied Sciences of Tangier (ENSATg), University of Abdelmalek Essaadi. Route de Ziaten Km 10, Tanger Principale, Tangier B.P 1818, Morocco
E-mail: [1]sekkat.salma@etu.uae.ac.ma, [2]a.najdi@uae.ac.ma, [3]mounazouine@gmail.com, [4]z.cheker@gmail.com, [5]s.chakkor@uae.ac.ma, [3]abdallahoulmaati@hotmail.com

ABSTRACT

Early and accurate diagnosis of Neonatal Respiratory Distress Syndrome (RDS) is vital for improving outcomes in preterm infants. Traditional diagnostic methods often suffer from subjectivity and limited scalability, especially in resource-constrained settings. This study investigates the potential of convolutional neural networks (CNNs) for automated RDS detection from chest X-rays images, using a clinically annotated dataset sourced from multiple Moroccan hospitals. A comparative analysis of seven CNN architectures (InceptionV3, ResNet-50, DenseNet121, VGG16, EfficientNetV2, ConvNeXt, and a custom lightweight CNN) was conducted. Models were evaluated using precision, recall, F1-score, accuracy, and AUC. InceptionV3 achieved the highest performance with 93% accuracy and 96% recall for RDS, while the custom CNN demonstrated strong sensitivity (95% recall), making it well-suited for deployment in low-resource environments. The study underscores the effectiveness of deep learning for neonatal diagnostics and highlights the importance of context-specific AI solutions in enhancing early detection and decision support for RDS in clinical practice.

Keywords: *RDS, Chest X-rays Classification, Neonatology, Diagnostic Support System, CNNs, Deep Learning, InceptionV3, EfficientNetV2, ConvNeXt.*

1 INTRODUCTION

their dataset is from Moroccan hospitals of neonatal respiratory distress syndrome (RDS), a condition primarily affecting preterm infants due to surfactant deficiency [1], is critical for timely intervention and improved outcomes [2]. Conventional diagnostic methods rely on clinical signs, such as tachypnoea, nasal flaring, and grunting, supported by pulse oximetry, blood gas analysis, and radiographic scoring systems (Downes, Silverman-Anderson) [3,4]. While traditional models like SVMs, Random Forests, and logistic regression offer interpretability and acceptable performance on small datasets [3,4], they struggle with high-dimensional data, lack integration of multimodal inputs, and often rely on subjective assessments or invasive procedures [5,6]. Moreover, they are time-consuming and ill-suited for real-time monitoring, especially in resource-limited settings. In this context, artificial intelligence (AI), particularly deep learning using convolutional neural networks (CNNs), has emerged as a promising tool for enhancing neonatal diagnostics by enabling automated analysis of chest radiographs and physiological data [7,8]. CNN-based systems have

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

shown high accuracy in detecting respiratory conditions like pneumonia and COVID-19 [9–10], and are increasingly applied to neonatal care, including gestational age estimation and lung ultrasound interpretation [11,12,13]. However, despite RDS being a leading cause of neonatal morbidity and mortality in regions like Morocco [14], its early detection through AI remains underexplored. Studies such as those by Jang et al. [15] and Gravina et al. [12] have proposed AI models for predicting respiratory conditions, yet limitations in dataset size, generalizability, and clinical integration persist. Similarly, while newer architectures like EfficientNetV2 and ConvNeXt show promise in pediatric imaging [16,17], their application to infant RDS has not been fully evaluated. This study addresses this critical gap by developing and evaluating a CNN-based framework for automated RDS diagnosis from neonatal chest X-rays. Using a curated dataset collected from Moroccan hospitals and labeled by neonatologists, we conduct a comparative analysis of multiple CNN architectures, including InceptionV3, ResNet-50, DenseNet121, VGG16, EfficientNetV2, ConvNeXt, and a custom CNN model. Models are assessed using precision, recall, F1-score, and accuracy to determine their clinical viability for early detection and decision support. By focusing on a real-world, resource-constrained setting, this research aims to provide an innovative, scalable AI solution to improve RDS diagnosis and reduce diagnostic variability. Despite recent advances in deep learning for respiratory disease detection, the specific application to neonatal respiratory distress syndrome (RDS) remains underexplored. The burden of RDS in Morocco and other low-resource settings highlights the urgent need for automated, context-sensitive diagnostic tools. Previous studies have mainly focused on pneumonia or COVID-19, using datasets that do not capture neonatal-specific conditions, which limits their generalizability. To address this gap, this study formulates the following problem statement: early diagnosis of neonatal RDS is hindered by subjective interpretation of chest radiographs and their dataset is from Moroccan hospitals tailored to regional clinical contexts.

Previous studies applying CNNs to neonatal and pediatric imaging have mostly focused on pneumonia or COVID-19 detection, often using datasets from high-resource countries. While these works demonstrate the diagnostic potential of CNNs, none has addressed neonatal RDS in low-resource contexts. Our work differs in both motivation and findings: it provides the first comparative evaluation of CNNs for neonatal RDS using a clinician-validated Moroccan dataset, and introduces a lightweight CNN trained from scratch to meet the needs of low-resource healthcare systems.

Previous studies applying CNNs to neonatal and pediatric imaging have mostly focused on pneumonia or COVID-19 detection, often using datasets from high-resource countries. While these works demonstrate the diagnostic potential of CNNs, none has addressed neonatal RDS in low-resource contexts. Our work differs in both motivation and findings: it provides the first comparative evaluation of CNNs for neonatal RDS using a clinician-validated Moroccan dataset, and introduces a lightweight CNN trained from scratch to meet the needs of low-resource healthcare systems.

Accordingly, this research poses two main questions:

1. Which convolutional neural network (CNN) architecture provides the best trade-off between accuracy, sensitivity, and clinical applicability for neonatal RDS diagnosis?

2. Can a lightweight CNN trained on localized Moroccan data achieve comparable performance to pretrained models in low-resource settings?

The contributions of this paper are threefold: (i) the first comparative benchmarking of multiple state-of-the-art CNNs for neonatal RDS diagnosis in a North African clinical setting, (ii) the design and evaluation of a lightweight CNN trained from scratch on a clinician-validated dataset, and (iii) a critical discussion of clinical implications, limitations, and open research issues. The remainder of this paper is organized as follows: Section II reviews related literature; Section III details the dataset, models, and evaluation process; Section IV presents and discusses the results; and Section V concludes with key findings, limitations, and directions for future work. The remainder of this paper is structured as follows: Section II reviews relevant literature on AI applications in neonatal and respiratory care, identifying key gaps and justifying the study's focus. Section III describes the methods and materials, including dataset construction, preprocessing steps, CNN architectures, and performance metrics. Section IV presents the experimental results and comparative analysis of the models. Section V discusses the implications of the findings, highlights model performance in a clinical context, and considers limitations. Finally, Section VI concludes the study and proposes future directions, including potential clinical integration and further validation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2 RELATED WORK

Although artificial intelligence (AI) has been increasingly applied to respiratory disease diagnosis, its specific use for neonatal respiratory distress syndrome (RDS) remains limited, particularly in comparative studies involving convolutional neural networks (CNNs) such as DenseNet121, VGG16, ResNet-50, and InceptionV3 trained on large, annotated chest X-ray datasets. Most existing research has concentrated on other respiratory conditions, notably COVID-19, yielding useful insights into CNN performance but limiting their applicability to neonatal RDS. For example, Jang et al. [15] proposed an ensemble deep learning model to predict RDS in very low birth weight infants using a large, multicenter Korean dataset; however, concerns regarding generalizability and the retrospective design remain. Similarly, Gravina et al. [12] and Bassiouny et al. [13] applied deep learning techniques to lung ultrasound (LUS) images for neonatal respiratory assessment, showcasing potential but constrained by limited sample sizes and implementation complexity. Meanwhile, Ervural and Ceylan [b] investigated thermal imaging combined with CNNs to distinguish between respiratory and non-respiratory neonatal conditions, achieving promising results in a small cohort, though their method lacks scalability and robustness across diverse populations. In their systematic review, Rashid et al. [18] and Tran et al. [19] emphasized the necessity for standardized protocols, multicenter validation, and real-world testing to strengthen AI's clinical relevance in diagnosing respiratory distress. Complementary findings by Hiles et al. (2017) [20] highlighted the diagnostic value of LUS over chest X-ray (CXR), with sensitivity and specificity rates of 97% and 91%, respectively. Nevertheless, their review identified key limitations including operator dependency, study heterogeneity, and the lack of high-quality randomized trials. In the context of COVID-19, comparative CNN-based studies by Shazia et al. [9], Khan and Aslam [21], and Visuña et al. [22] demonstrated the performance of architectures such as DenseNet121, ResNet-50, and InceptionV3. Despite these successes, unresolved challenges, such as data imbalance, limited multimodal integration, and inadequate external validation, still hinder clinical translation. Additionally, research by Manokaran et al. [23] and Ghoshal et al. [10] reinforced the capability of EfficientNet and DenseNet variants for pneumonia detection, while underlining the importance of ensuring generalization across different diagnostic contexts. Furthermore, recent studies have explored cutting-edge CNN architectures for neonatal diagnostics. Notably, Ceylan et al. [24] proposed a deep CNN (C+EffxNet) model for detecting multiple neonatal respiratory disorders, reporting accuracy above 97%. Likewise, Yildirim et al. [25] applied a similar hybrid CNN approach to classify RDS, ARDS, TTN, pneumothorax, and BPD, achieving a high kappa agreement. However, these works often rely on public datasets or remain focused on broader classifications without targeted analysis of RDS. In parallel, Liu et al. [26], Ahmad et al. [27], and Vergine et al. [28] evaluated manual and semi-automated LUS and CXR approaches for RDS detection. While these studies confirm high diagnostic accuracy (often exceeding 90%), they also expose drawbacks such as the lack of automation, operator dependency, and variability in interpretation. Haider et al. [29], in turn, examined the use of AI-enhanced imaging for early detection of neonatal respiratory complications, integrating AI with clinical imaging pathways. Although this approach supports early intervention, it requires further validation for RDS-specific applications. Together, these contributions underscore the diverse methodological landscape and the ongoing tension between automation, clinical accuracy, and deployment feasibility. To address these gaps, the present study offers a systematic evaluation of both pretrained and custom CNN models applied to a Moroccan clinical dataset. By focusing on automated detection of RDS from chest radiographs in neonates, this work highlights the importance of localized, real-world data and underscores the clinical potential of context-sensitive AI solutions in neonatology. In summary, while prior studies have demonstrated the potential of CNNs for pediatric pneumonia or COVID-19, very few have specifically focused on neonatal RDS, and none to our knowledge have comparatively evaluated multiple architectures on real-world Moroccan clinical data. This work fills that gap by benchmarking six pretrained CNNs and one custom lightweight model on a dataset curated and validated by neonatologists, thereby providing context-specific insights that go beyond incremental improvements and contribute novel evidence to the field.

3 METHODS AND MATERIALS

3.1 Study Design and Process

This study employs a quantitative retrospective approach to evaluate the performance of AI models in diagnosing Infant Respiratory Distress Syndrome (RDS) using chest X-ray images. A comparative analysis of multiple convolutional neural network (CNN)-based architectures is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

conducted to determine the most effective model for classifying RDS and non-RDS cases. Effectively, this work was conducted from November 2022 to April 2024 at the Tangier University Center. The study population consisted of 1300 chest X-rays of premature newborns extracted from their patient records via the center's hospital information system. The newborns were admitted to the neonatal intensive care unit (NICU) for any reason. The ground truth for the diagnosis of RDS in our study was not based solely on raw medical record data. Each chest X-ray image included in the dataset was clinically reviewed and classified by a specialized neonatology team. The diagnosis of RDS was established according to standardized clinical criteria, including the presence of respiratory signs such as tachypnoea (> 60 breaths/min), nasal flaring, chest retractions, and grunting, consistent with international guidelines. Only cases that met these criteria and were confirmed by clinical judgment and radiographic assessment were labeled as RDS. Images from patients with other respiratory conditions or congenital anomalies were excluded to ensure diagnostic clarity.

In Phase 1, the radiographic images were classified by a neonatology team as either RDS or N-RDS. Respiratory distress is characterised by tachypnoea (i.e., a respiratory rate greater than 60 cycles/minute), chest retractions, nasal exhalation, and grunting. The non-inclusion criteria were as follows: major malformations (congenital diaphragmatic hernia, adenomatoid pulmonary malformation) and association with other respiratory pathologies, such as pneumonia.

In phase 2, we used deep learning, also known as machine learning in the domain of artificial intelligence (AI), we chose an approach based on the use of convolutional neural networks (CNN) and transfer learning via models (DENSNET121, VGG16, RESNET 50, InceptionV3, EfficientNetV2 and ConvNeXt), which are recognized for their ability to extract relevant features from complex images [30,31], making them particularly suitable for medical image analysis.

Our methodology involved several key steps. First, we assembled a dataset containing representative images of RDS as well as healthy images to serve as a reference. Next, we divided this dataset into Training Validation and Test sets to ensure an accurate evaluation of the models' performance. During the training phase, the models were exposed to many RDS images and their parameters were adjusted to minimize the loss function, thereby improving their ability to discriminate images with signs of RDS.

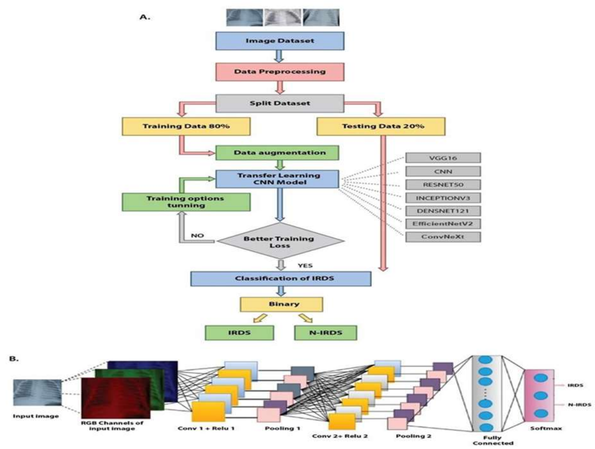

Figure 1. Workflow of CNN-Based Framework for Automated Diagnosis of Neonatal RDS.

(A) Processing steps: preprocessing, data splitting, transfer learning–based training, optimization, and binary classification.

(B) Illustrates the typical internal flow of a CNN model used in this study for binary classification of IRDS versus N-IRDS from neonatal chest radiographs.

3.2 Data Collection

Although this study is based on retrospective data extracted from patient records, it is important to emphasize that the University Hospital Center (CHU) of Tangier serves as the main referral hub for the entire northern region of Morocco. The CHU encompasses multiple healthcare facilities of varying levels, including provincial hospitals, specialized institutions, and university departments, spread across several provinces such as Tangier, Tetouan, and Al Hoceïma. As a result, the cases included in this study originate from diverse clinical units and hospital settings, reflecting a broad spectrum of neonatal profiles and offering a regionally representative dataset. This diversity strengthens the external validity of our findings and underscores a key regional healthcare challenge: the variability in diagnostic practices and the unequal access to specialized expertise in neonatal care. It is precisely within this context that the development of AI-based diagnostic tools becomes highly relevant. By offering reliable, standardized, and automated decision support, artificial intelligence has the potential to reduce inter-hospital disparities,

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

facilitate earlier diagnosis of respiratory distress syndrome, and enhance clinical decision-making even in under-resourced settings. Our work, therefore, reflects a dual commitment: to leverage real-world clinical data from across the region and to advance innovative AI solutions that promote equity and improve neonatal care outcomes. The dataset used in this work consists of neonatal chest X-ray images collected from two distinct sources:

1) Hospital records from the University Hospital of Tangier, specifically chest X-ray images of premature newborns admitted to the NICU. These images were accessed with appropriate ethical approval and were fully anonymized to ensure patient confidentiality.
2) Publicly available datasets, which were used in part to supplement the image database during the model pretraining or benchmarking phases. These datasets are de-identified and accessible for research use.

To ensure uniformity and optimize model performance, the entire dataset underwent standardized preprocessing steps as follows:

Rescaling: Images are resized to 224x224 pixels to match CNN input requirements.

Normalization: Pixel intensity values are scaled between 0 and 1 for uniform representation.

Data Augmentation: Techniques such as rotation, flipping, and contrast adjustments are applied to reduce overfitting.

Segmentation: Unnecessary artifacts are removed to enhance feature extraction.

Chest X-ray images undergo comprehensive preprocessing, including normalization and noise reduction, to standardize inputs and optimize quality, ensuring they are ready for further analysis. We included 1300 images, consisting of 536 RDS cases and 764 N-RDS cases. The dataset was then split, with 80% used for training the models and the remaining 20% reserved for the test set, which was used for confusion matrix evaluation. This division into a training set and a test set was a crucial step for the validation and evaluation of the model. To enhance the training set and improve the model's generalizability, various data augmentation techniques such as flipping, zooming, and rotation are applied. This adaptation enables the model's accuracy to be fine-tuned. Also, during the learning phase, the network gradually adjusts its weights and biases. This optimization enables the model to converge towards an optimal solution. Furthermore, assessing the need for retraining and optimizing training parameters are ongoing steps to ensure the model achieves and maintains a satisfactory performance level, capable of accurately identifying and classifying the conditions from the processed images.

3.3 AI Model Development:

This study evaluates the performance of six established convolutional neural network (CNN) architectures: DenseNet121, ResNet-50, VGG16, InceptionV3, EfficientNetV2, and ConvNeXt alongside a custom CNN model developed and trained from scratch on our specific dataset. While the core CNN architectures are not newly proposed, the novelty of our work lies in their comparative benchmarking and adaptation to a Moroccan neonatal clinical setting, which, to our knowledge, has not been previously studied for RDS detection using chest radiographs. The six pre-trained models were selected for their proven effectiveness in medical image analysis tasks, particularly in chest X-ray classification (COVID-19, pneumonia). However, these models have not yet been comparatively evaluated on neonatal respiratory distress syndrome (RDS), particularly in a low-resource, real-world North African context. This study fills that gap by providing a rigorous, side-by-side comparison using expert-verified diagnostic labels. In addition, we developed a custom CNN baseline model using a lightweight architecture with three convolutional layers, ReLU activation functions, max pooling, and fully connected layers with a SoftMax classifier [32]. This model was designed specifically for RDS detection with high recall, an essential metric in a clinical screening context where false negatives must be minimized. Unlike the pre-trained models that rely on transfer learning, our CNN was trained entirely from scratch using the hospital dataset, allowing us to evaluate its performance against more complex architectures and assess the value of localized training. The methodology is mathematically represented by the following equation:

$$f(x)=sign\ (2\sigma_{softmax}\ (W_{3\sigma relu}\ (W_1\, x+b_1) + b_3)\ -1) \quad (1)$$

Where "x" represents the input, "W" the weights of the neural network layers, "b" the biases, and "σ" the activation functions. Each model was optimized for binary classification (RDS vs. non-RDS), using a consistent dataset split (80% training, 20% test), and evaluated using the same preprocessing pipeline and performance metrics (accuracy, precision, recall, F1-score, AUC). The training process involved standard procedures, including loss minimization through backpropagation, data augmentation, and regularization to avoid overfitting. The originality of our work lies in the following key contributions:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

First comparative evaluation of state-of-the-art CNN models (including EfficientNetV2 and ConvNeXt) for RDS diagnosis in premature newborns using chest X-ray images from a real clinical setting in Morocco.

Introduction of a custom-built CNN baseline model trained from scratch, optimized for recall, to assess the trade-offs with transfer learning methods in low-resource environments.

Use of clinician-validated ground truth labels based on standardized neonatal diagnostic criteria, ensuring clinical relevance and robustness of the training data.

Context-specific analysis of model strengths and weaknesses, emphasizing metrics (recall and F1-score for RDS) most relevant to neonatal screening scenarios.

This hybrid approach, evaluating both transfer learning and locally trained models, offers actionable insights into how AI can be adapted to support clinical decision-making in neonatology, especially in under-resourced healthcare systems.

3.4 Ethical considerations:

The present study received approval from the Ethics Committee of the University Hospital of Tangier (CEHUT) under the reference number AC92JT/2024, and ethical approval was obtained from the relevant institutional review board. Furthermore, all patient data were anonymized, and informed consent was obtained where necessary. The study fully adheres to the ethical principles outlined in the Declaration of Helsinki.

3.5 Performance Evaluation:

The models are assessed using key performance metrics which are:

Precision: This is indicative of the proportion of positive predictions that are actually correct.

Recall (Sensitivity): which measures the model's ability to correctly identify positive cases among all the truly positive cases available.

F1-Score: Harmonic mean of precision and recall. It is a measure that combines precision and recall into a single metric.

Overall Accuracy: measure percentage of correctly classified images.

ROC-AUC Curve: measures the model’s ability to distinguish between RDS and non-RDS cases.

Statistical analysis was performed using ANOVA (Analysis of Variance) to compare group means and identify significant differences, with post-hoc tests for pairwise comparisons when necessary. Performance differences among models were evaluated using paired t-tests and ANOVA, with 95% confidence intervals reported for key metrics. This approach ensures the reproducibility and validity of the AI-based RDS detection findings.

4 RESULTS

This section presents the results of a structured comparative analysis of seven convolutional neural network (CNN) architectures for the automated diagnosis of neonatal respiratory distress syndrome (RDS) using chest X-ray images. Unlike previous studies that have applied CNNs primarily to COVID-19 or pneumonia detection, our work focuses specifically on the classification of RDS versus non-RDS in premature infants, using data annotated by neonatology experts in a real Moroccan clinical setting. This population and diagnostic context remain largely underrepresented in the AI medical imaging literature. In addition to benchmarking six well-known CNN architectures (InceptionV3, EfficientNetV2, ConvNeXt, ResNet-50, DenseNet121, and VGG16), we also evaluate a custom-built CNN model trained from scratch on our dataset. This model was developed to test the feasibility and performance of a lightweight architecture tailored for environments with limited computational resources and no access to large-scale pretrained weights. Despite its simplicity, the custom CNN showed strong sensitivity (recall of 95% for RDS), making it a viable screening tool in clinical scenarios where early detection is critical.

The analysis was structured in two stages: firstly, an individual reading of the learning curves (Accuracy and Loss) over 500 epochs for each model; secondly, a comparative summary via aggregated visualisations (Boxplots, curves grouped over 20 epochs). This two-pronged approach, both detailed and macroscopic, guarantees a robust assessment of the stability, convergence and generalisability of the models tested. Each model was trained and validated under the same conditions to ensure fairness and reproducibility: identical dataset split, identical image preprocessing pipeline, and the same evaluation metrics (accuracy, precision, recall, F1-score, and area under the ROC curve (AUC)). This allowed us to not only compare model performance but also analyze their strengths and weaknesses across clinically meaningful indicators, such as recall for RDS, where missed diagnoses can lead to severe consequences. Our results show that InceptionV3 outperformed all other models, achieving the highest accuracy (93%), the highest recall for the RDS class (96%), and the lowest loss,

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

indicating both robustness and clinical reliability. EfficientNetV2 and ConvNeXt followed closely, offering a strong balance between sensitivity and specificity. In contrast, models like VGG16 and DenseNet121, despite their historical relevance, struggled to match the performance of the more modern or customized architectures. Importantly, our custom CNN model, although simpler, delivered competitive results and notably high recall for RDS, underscoring its potential for deployment in resource-limited neonatal care units.

4.1 Model-by-Model Performance Analysis:

4.1.1 Custom CNN versus DenseNet-121

From figure 2, which illustrate the rapid convergence of the CNN and the increased stability of the DenseNet-121, our custom CNN model demonstrated rapid convergence, achieving stable performance with an overall accuracy of 89% and a recall of 95% for the RDS class. Although it exhibited a lower recall (77%) for the non-RDS class, its performance in identifying true RDS cases makes it particularly effective for early screening purposes. In contrast, DenseNet-121, while exhibiting a more stable training curve and better balance between the two classes, reached a lower RDS recall of only 70%. Despite its architectural depth and connectivity, DenseNet-121's lower sensitivity to pathological cases limits its reliability in high-risk clinical applications where false negatives are critical. This comparison highlights the added value of localized, task-specific training, as our simpler CNN (though less complex) was better suited to the dataset and the clinical priority of detecting RDS reliably.

4.1.2 VGG16 vs ResNet-50

Figure 3 show a comparison between a classic architecture (VGG16), which is more stable but performs less well, and a deep architecture with residual connections (ResNet-50), which is more sensitive but performs better on pathological cases. VGG16, a classical architecture with stacked convolutional layers, stabilized quickly but showed limited adaptability, particularly for RDS detection, with a recall of 74% and a relatively high number of false negatives. This makes it less suitable for neonatal diagnostic support, where missed detections have serious clinical implications. By contrast, ResNet-50, leveraging residual connections that allow for deeper learning without vanishing gradients, achieved a recall of 93% for RDS and an overall accuracy of 87%, showing better detection capability for complex radiographic patterns. However, its performance on non-RDS cases was lower (recall of 77%), indicating a trade-off in class balance. Overall, while ResNet-50 outperforms VGG16 in sensitivity to RDS features, its variability calls for more refined tuning for deployment.

4.1.3 EfficientV2 and CONVNeXt

As demonstrated in figure 4, the two modern architectures show rapid convergence and high stability. ConvNeXt shows a slight superiority in sensitivity for the RDS class, while EfficientNetV2 offers a good overall balance between the two classes.

Both EfficientNetV2 and ConvNeXt showed excellent generalization, with accuracies of 90% and 88%, respectively. EfficientNetV2 achieved a balanced performance with a recall of 92% for RDS and 82% for non-RDS, showing robustness in both classes. ConvNeXt, on the other hand, favored RDS detection slightly more (F1-score: 90% for RDS) at the cost of reduced specificity (non-RDS recall: 80%). These models reflect the benefit of modern architectural improvements (compound scaling in EfficientNetV2 and transformer-inspired design in ConvNeXt), which support efficient yet accurate learning on medical imaging datasets. Their performance makes them strong candidates for real-time deployment in clinical tools.

4.1.4 InceptionV3: reference model for RDS classification

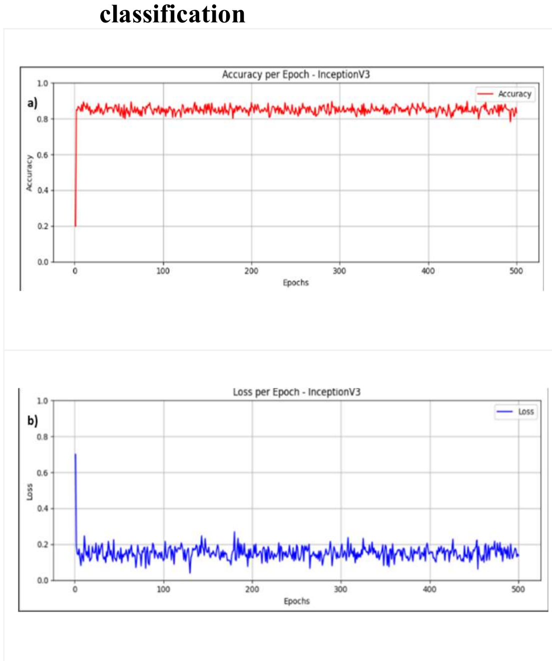

Figure 5: Learning curves for the InceptionV3 model:(a) Accuracy evolution over 500 epochs, (b) loss function evolution

InceptionV3 stood out as the best-performing architecture across all metrics. It achieved: overall accuracy of 93%, recall of 96% for the RDS class, and F1 scores of 94% (RDS) and 90% (N-RDS). The area under the ROC curve (AUC) was 0.94, and the loss function converged to a low value of 0.085, reflecting excellent optimisation and generalisation capabilities. In terms of learning, the accuracy curve (Figure 5. a) shows a rapid improvement from the first epochs, followed by a stable plateau around 88-90%, with no major oscillations. This behaviour suggests rapid convergence and significant stability, with no signs of overlearning. Similarly, the loss curve (Figure 5.b) decreases rapidly to stabilise at a constant low level, confirming a controlled and effective training phase. The InceptionV3 multi-scale feature extraction capability enabled it to handle variability in chest radiograph presentation, especially in premature infants. The low false-negative rate (only 5 cases) confirms its reliability as a reference model for future clinical AI applications in neonatology.

4.2 Performance Summary and Statistical Significance:

The overall results of all models are summarized in tables 1 and 2, and their statistical robustness is confirmed via ANOVA testing, which revealed a highly significant difference in model accuracies ($F = 568.88$, $p < 10^{-124}$). The table 1 compares the performance of all models across RDS and non-RDS classes, using precision, recall, and F1-score.

Table 1: Class-wise performance summary (RDS / N-RDS)

Model	Accuracy	Precision (RDS/ N-RDS)	Recall (RDS/ N-RDS)	F1-Score (RDS/ N-RDS)
CNN	89%	88% / 90%	95% / 77%	91% / 83%
DenseNet121	83%	83% / 84%	70% / 91%	76% / 87%
VGG16	81%	78% / 83%	74% / 86%	76% / 84%
ResNet-50	87%	86% / 88%	93% / 77%	89% / 82%
InceptionV3	93%	92% / 94%	96% / 87%	94% / 90%
EfficientNetV2	90%	89% / 91%	92% / 88%	90% / 89%
ConvNeXt	86%	87% / 82%	72% / 92%	75% / 88%

InceptionV3 and EfficientNetV2 show the best balance across metrics, with high sensitivity (recall) and low false-negative rates for RDS. The custom CNN demonstrates strong recall for RDS, despite limited architectural complexity. This validates the performance gap observed between classical models and more modern or customized architectures. Furthermore, ROC curves, confusion matrices, and training loss graphs consistently reinforced that InceptionV3, EfficientNetV2, and our custom CNN provided the most clinically relevant performance especially in terms of recall for RDS, which is critical for safe early diagnosis. Table 2 presents overall metrics for each model: accuracy, loss, AUC, and average scores. InceptionV3 achieves the best global performance across all indicators. EfficientNetV2 and the custom CNN show strong robustness and generalizability. Models like DenseNet121 and VGG16 underperform, particularly in sensitivity for RDS.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Overall Model Scores of the tested CNN models

Model	Loss	Accuracy	Precision	Rappel	AUC	Score-F1
InceptionV3	0.085	0.93	0.92	0.96	0.94	0.94
EfficientNetV2	0.105	0.9	0.89	0.92	0.92	0.9
CNN	0.145	0.89	0.88	0.95	0.89	0.91
ConvNeXt	0.125	0.86	0.87	0.72	0.9	0.81
ResNet-50	0.165	0.87	0.86	0.93	0.86	0.89
DenseNet121	0.185	0.83	0.83	0.7	0.8	0.81
VGG16	0.205	0.81	0.78	0.74	0.78	0.8

Before interpreting overall model performance, it was necessary to first evaluate training quality through analysis of loss curves, which confirmed stable convergence without overfitting. Once learning stability was established, a statistical comparison using boxplots of Accuracy, AUC, and Loss was conducted for InceptionV3 and six other CNN models as shown in figure 7.

These visualizations revealed clear differences in performance distributions, with InceptionV3 showing low variance and high central tendency. To validate these observations, a one-way ANOVA was performed, confirming that the differences in accuracy between models were statistically significant ($p < 10^{-124}$), thus rejecting the null hypothesis of equal performance. Consequently, this analysis supports the conclusion that InceptionV3 significantly outperforms models like VGG16 in this clinical classification task.

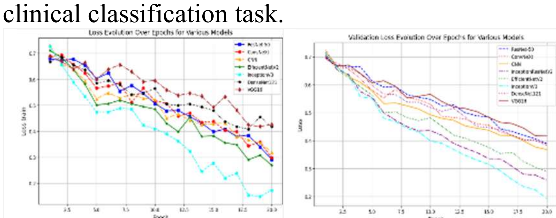

Figure 8: Learning and Validation Loss Curves: (a) loss learning and (b) loss validation analysis for Inception v3 compared to six other advanced CNN models.

Figure 8 illustrates the training loss evolution over 20 epochs for seven CNN models, highlighting their learning dynamics and convergence behavior. Overall, all models show a gradual decrease in loss, but with significant variation in speed and stability. InceptionV3 stands out with rapid, stable convergence and the lowest final loss (<0.2), indicating excellent generalization. EfficientNetV2 also performs well, with steady decline and minimal fluctuation, while the custom CNN shows acceptable convergence at a slower pace. In contrast, ResNet-50 and ConvNeXt exhibit moderate oscillations and higher final loss (0.35–0.4), suggesting room for optimization. DenseNet121 and VGG16 perform the worst, with high and unstable loss curves (>0.5), indicating poor adaptation to the task. These findings confirm that newer or well-tuned models like InceptionV3 and EfficientNetV2 offer better learning efficiency than older or overly complex architectures, regardless of their parameter count. In summary, InceptionV3 emerged as the most clinically viable model due to its combination of high recall, precision, and training stability. Our custom CNN, though simpler in architecture, proved highly sensitive to RDS and well-suited for low-resource environments. Meanwhile, EfficientNetV2 and ConvNeXt offered strong performance with efficient computation, making them attractive for scalable deployment. In contrast, VGG16 and DenseNet121, despite their historical relevance, underperformed on our dataset and are not recommended without significant retraining or adaptation.

5 DISCUSSION

Artificial intelligence (AI) continues to transform neonatal care, particularly in diagnosis, monitoring, and personalized treatment planning. While existing studies have applied CNNs to various respiratory conditions such as pneumonia, COVID-19, or BPD, few have directly addressed infant RDS using localized, real-world clinical datasets. To contextualize our findings, Table 3 presents a comparative overview of the proposed AI-based models and existing diagnostic methods for neonatal respiratory distress syndrome (RDS) as reported in the literature, emphasizing their accuracy, sensitivity, specificity, and real-world applicability. Clinical methods such as lung ultrasound (LUS) interpreted by experts (Liu et al. [26], Ahmad et al. [27], Vergine et al. [28], Hiles et al. [20]) show high diagnostic performance but remain limited by operator dependency and scalability. AI-based models from the literature, including hybrid CNNs and deep learning architectures (e.g., Ceylan et al. [24], Yildirim et al. [25], Jang et al. [15]), demonstrate strong results, though often lack external validation or are trained on narrow populations. The models proposed in this study, particularly InceptionV3 and EfficientNetV2, outperform others with high accuracy (93% and

90%, respectively) and sensitivity (96% and 92%), making them suitable for early screening. The custom CNN, while simpler, achieves strong recall (95%) and decent overall accuracy (89%), highlighting its value in resource-limited settings. In contrast, older models like DenseNet121 and VGG16 underperform in sensitivity, reinforcing the importance of architecture choice and dataset specificity. Overall, the table underscores the clinical relevance of modern CNNs trained on localized data and the trade-offs between model complexity, performance, and deployability.

This study fills that gap by evaluating both state-of-the-art CNNs and a task-specific, lightweight model on a Moroccan dataset curated and labeled by a neonatology team. The comparative framework not only highlights the strengths of advanced architectures like InceptionV3 and EfficientNetV2, but also illustrates the clinical utility of simpler models when well-trained on targeted data. Importantly, we show that accuracy alone is not sufficient in a medical context. Metrics such as recall for RDS, low false-negative rates, and robust learning curves are far more indicative of a model's practical reliability. InceptionV3 consistently excelled on these fronts. Conversely, older architectures like VGG16 and DenseNet121 underperformed, underscoring the need for careful model selection based on clinical priorities not popularity or complexity. Furthermore, the development and strong performance of the custom CNN model demonstrate the feasibility of building effective solutions without reliance on large pretrained models, particularly in settings where computational resources or internet access are limited. This work emphasizes the importance of ethical data use, local validation, and tailored architecture choices when deploying AI in neonatology. Convolutional Neural Networks (CNNs), especially modern variants like InceptionV3, EfficientNetV2, and ConvNeXt, are increasingly used in medical imaging due to their ability to automatically extract hierarchical features from radiological images without requiring handcrafted inputs. In the diagnosis of neonatal respiratory distress syndrome (RDS), these models can detect subtle signs of distress on chest X-rays, often imperceptible to the human eye, particularly in early or borderline cases. Their architectural diversity in terms of depth, width, and feature representation allows them to balance accuracy, inference speed, and computational efficiency depending on clinical needs. Importantly, once trained, CNNs can provide results within seconds, offering near-instant diagnostic support, which significantly outpaces manual radiologist interpretation. This efficiency helps accelerate clinical decision-making and alleviate bottlenecks in neonatal intensive care units (NICUs), especially in high-volume or under-resourced settings. Furthermore, CNN deployment reduces reliance on specialized personnel, streamlines triage, and facilitates better resource allocation. These models can also be embedded into portable or cloud-based platforms, supporting remote diagnosis and telemedicine services. Consequently, for patients in rural or underserved areas, CNN-based systems can reduce transport needs, lower costs, and shorten treatment delays, thereby improving access to timely care. To encourage the broader adoption of such systems, it is crucial to emphasize not just diagnostic accuracy but also their practical utility, scalability, and safety. Although some literature reports near-perfect results, these are often obtained under tightly controlled conditions and rely on costly, expert-dependent, or invasive procedures. In contrast, CNNs applied to non-contact, low-cost modalities like thermal imaging offer radiation-free, real-time diagnostics suitable for continuous bedside monitoring. With minimal human input, these models are ideal for early screening and decision support, particularly in resource-constrained environments. Rather than replacing established gold standards, CNN-based tools are designed to complement them by enhancing diagnostic reach, consistency, and speed. Our proposed model fits this vision by providing a reliable, high-sensitivity screening solution that can identify at-risk neonates early, thus enabling prompt intervention and more efficient use of medical resources. Several limitations must be acknowledged. First, the dataset size, though clinically validated, remains modest compared to large international repositories. Second, the retrospective design restricted control over variability in acquisition protocols and image quality. Third, only imaging data were considered, without integration of multimodal clinical variables such as blood gas values or patient demographics. Finally, the lack of external validation across other hospitals and healthcare contexts limits the generalizability of our findings. Future research should incorporate these clinical variables and validate the models in multicenter, prospective studies involving larger and more diverse populations. Despite these limitations, the novelty of this study lies in its comparative evaluation of CNNs specifically tailored to neonatal RDS in a North African context. Unlike previous research focusing on other respiratory diseases, this work introduces a lightweight CNN that combines computational efficiency with diagnostic sensitivity, making it particularly suitable for low-resource healthcare environments.

6 CONCLUSION AND PERSPECTIVE

This study presented the first comparative evaluation of seven CNN architectures, including InceptionV3, EfficientNetV2,ConvNeXt,ResNet-50, DenseNet121, VGG16, and a custom lightweight CNN, for the automated diagnosis of neonatal RDS using Moroccan clinical data. Among these, InceptionV3 achieved the best overall performance with 93% accuracy and 96% recall for RDS, while the custom CNN demonstrated strong recall (95%) despite its simplicity, underscoring its value for low-resource settings. Beyond reporting metrics, our conclusions critically assess the implications of these findings. InceptionV3's superior recall indicates its reliability for minimizing false negatives, which is a priority in neonatal care. EfficientNetV2 and ConvNeXt also showed robust generalization, offering alternatives where computational efficiency is required. By contrast, older architectures like DenseNet121 and VGG16 underperformed, reinforcing the importance of model selection beyond historical popularity. Several open research issues remain. Larger multicenter datasets are needed to improve generalization. Future work should integrate multimodal data, combining imaging with clinical variables, and investigate explainable AI approaches to strengthen clinician trust. Emerging architectures such as Vision Transformers (ViTs) may further enhance diagnostic performance and interpretability. Ultimately, this study provides a foundation for developing AI-powered decision support systems that are both clinically relevant and context-sensitive. The originality of our study lies in providing the first CNN benchmarking for neonatal RDS in Morocco, combined with the development of a lightweight CNN trained from scratch, which showed promising results for low-resource healthcare systems. This work offers clinicians in low-resource settings a practical AI-based screening tool to support early RDS diagnosis and timely intervention.

7 AUTHOR CONTRIBUTIONS

S. Sekkat: conceptualization, methodology, manuscript drafting.
M.Zouine: clinical data curation and validation.
Z.Cheker: data preprocessing and software development.
O.Fangachi: statistical analysis and visualization.
K.Sammoud: literature review and critical revision.
S.Chakkor: supervision and project administration.
A.Oulmaati: clinical supervision and validation.
A.Najdi: overall supervision, funding acquisition, final manuscript review.

REFERENCES:

[1] Grappone L, Messina F, Hyaline membrane disease or respiratory distress syndrome? A new approach for an old disease, Journal of Pediatric and Neonatal Individualized Medicine (JPNIM), 2014, 3(2):e030263-e030263.

[2] Farrell PM, Avery ME, Hyaline membrane disease, Am Rev Respir Dis, 1975, 111(5):657-688.

[3] Yadav S, Lee B, Neonatal Respiratory Distress Syndrome, In: StatPearls, StatPearls Publishing, 2025.

[4] Sweet DG, Carnielli VP, Greisen G, et al., European Consensus Guidelines on the Management of Respiratory Distress Syndrome: 2022 Update, Neonatology, 2023, 120(1):3-23.

[5] Sullivan BA, Beam K, Vesoulis ZA, et al., Transforming neonatal care with artificial intelligence: challenges, ethical consideration, and opportunities, J Perinatol, 2024, 44(1):1-11.

[6] Doreswamy N, Horstmanshof L, Human Decision-making in an Artificial Intelligence-Driven Future in Health: Protocol for Comparative Analysis and Simulation, JMIR Res Protoc, 2022, 11(12):e42353.

[7] Sullivan BA, Beam K, Vesoulis ZA, et al., Transforming neonatal care with artificial intelligence: challenges, ethical consideration, and opportunities, J Perinatol, 2024, 44(1):1-11.

[8] Shazia A, Xuan TZ, Chuah JH, Usman J, Qian P, Lai KW, A comparative study of multiple neural network for detection of COVID-19 on chest X-ray, EURASIP J Adv Signal Process, 2021, 2021(1):50.

[9] Ghoshal N, Anas M, Roy SS, Chest X-Ray Image Classification of Pneumonia Disease Using EfficientNet and InceptionV3, In: Roy SS, Hsu CH, Kagita V, eds. Deep Learning Applications in Image Analysis, Springer Nature, 2023:173-186.

[10] Dhombres F, Bonnard J, Bailly K, Maurice P, Papageorghiou AT, Jouannic JM, Contributions

of Artificial Intelligence Reported in Obstetrics and Gynecology Journals: Systematic Review, J Med Internet Res, 2022, 24(4):e35465.

[11] Gravina M, Gragnaniello D, Verdoliva L, et al., Deep learning in the ultrasound evaluation of neonatal respiratory status, Published online October 31, 2020. doi:10.48550/arXiv.2011.00337.

[12] Bassiouny R, Mohamed A, Umapathy K, Khan N, An Interpretable Object Detection-Based Model For The Diagnosis Of Neonatal Lung Diseases Using Ultrasound Images, Annu Int Conf IEEE Eng Med Biol Soc, 2021, 2021:3029-3034.

[13] Danaoui K, Lahmini W, Bourrous M, Respiratory Distress in the Newborn: Experience of Pediatric Emergency Department, Open Access Library Journal, 2023, 10(3):1-7.

[14] Jang W, Choi YS, Kim JY, et al., Artificial Intelligence-Driven Respiratory Distress Syndrome Prediction for Very Low Birth Weight Infants: Korean Multicenter Prospective Cohort Study, J Med Internet Res, 2023, 25:e47612.

[15] Neshat M, Ahmed M, Askari H, Thilakaratne M, Mirjalili S. Hybrid Inception Architecture with Residual Connection: Fine-tuned Inception-ResNet Deep Learning Model for Lung Inflammation Diagnosis from Chest Radiographs. Published online October 5, 2023. doi:10.48550/arXiv.2310.02591

[16] Prasetyo SY, Prabowo AS, Santy, Pepita P, Optimizing Pneumonia Classification on Pediatric Chest X-Ray Images Using ConvNeXt Network, In: 2024 International Conference on Information Management and Technology (ICIMTech), 2024:369-374.

[17] Rashid M, Ramakrishnan M, Chandran VP, et al., Artificial intelligence in acute respiratory distress syndrome: A systematic review, Artif Intell Med, 2022, 131:102361.

[18] Sullivan BA, Beam K, Vesoulis ZA, et al., Transforming neonatal care with artificial intelligence: challenges, ethical consideration, and opportunities, J Perinatol, 2024, 44(1):1-11.

[19] Tran TK, Tran MC, Joseph A, Phan PA, Grau V, Farmery AD, A systematic review of machine learning models for management, prediction and classification of ARDS, Respiratory Research, 2024, 25(1):232.

[20] Hiles M, Culpan A-M, Watts C, Munyombwe T, Wolstenhulme S, Neonatal respiratory distress syndrome: Chest X-ray or lung ultrasound? A systematic review, Ultrasound, 2017, 25(2):80–91.

[21] Khan IU, Aslam N, A Deep-Learning-Based Framework for Automated Diagnosis of COVID-19 Using X-ray Images, Information, 2020, 11(9):419.

[22] Visuña L, Yang D, Garcia-Blas J, Carretero J, Computer-aided diagnostic for classifying chest X-ray images using deep ensemble learning, BMC Medical Imaging, 2022, 22(1):178.

[23] Manokaran J, Zabihollahy F, Hamilton-Wright A, Ukwatta E, Detection of COVID-19 from chest x-ray images using transfer learning, J Med Imaging (Bellingham), 2021, 8(Suppl 1):017503.

[24] Erdoğan Yıldırım A, Canayaz M, A novel deep learning-based approach for prediction of neonatal respiratory disorders from chest X-ray images, Journal of Applied Biomedicine, 2023, 43(4):635–655.

[25] Yildirim AE, Canayaz M, A novel deep learning-based approach for prediction of neonatal respiratory disorders from chest X-ray images, Biocybernetics and Biomedical Engineering, 2023, 43(4):635–655.

[26] Liu J, Cao H-Y, Wang H-W, Kong X-Y, The role of lung ultrasound in diagnosis of respiratory distress syndrome in newborn infants, Iranian Journal of Pediatrics, 2015, 25(1):e323.

[27] Ahmad ST, Parihar A, Bhat A, Wani NA, Lung ultrasound as modality in diagnosis of respiratory distress syndrome in newborn infants, International Journal of Contemporary Pediatrics, 2020, 7(12):2301–2306.

[28] Agerbirk N, Olsen CE, Heimes C, Christensen S, Bak S, Hauser TP, Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer Barbarea vulgaris, Phytochemistry, 2015, 115:130-142.

[29] Arslan MM, Yang X, Zhao N, Guan L, Cui T, Haider D, NDL-Net: A hybrid deep learning framework for diagnosing Neonatal Respiratory Distress Syndrome from chest X-rays, IEEE Open Journal of Engineering in Medicine and Biology, 2025, 6:407–412.

[30] Mukesh Kumar C, Sree Sankar J, Comparative Analysis of Convolutional Neural Networks for Brain Tumor Detection: A Study of VGG16, ResNet, Inception, and DenseNet Models, In: 2024 3rd International Conference on Applied

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Artificial Intelligence and Computing (ICAAIC), 2024:41-46.

[31] Krizhevsky A, Sutskever I, Hinton GE, ImageNet Classification with Deep Convolutional Neural Networks, In: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2012, Vol 25.

[32] Bishop C, Pattern Recognition and Machine Learning, Journal of Electronic Imaging, 2006, 16:140-155.

[33] Ervural S, Ceylan M, Convolutional neural networks-based approach to detect neonatal respiratory system anomalies with limited thermal image, Traitement du Signal, 2021, 38(2):437–442.

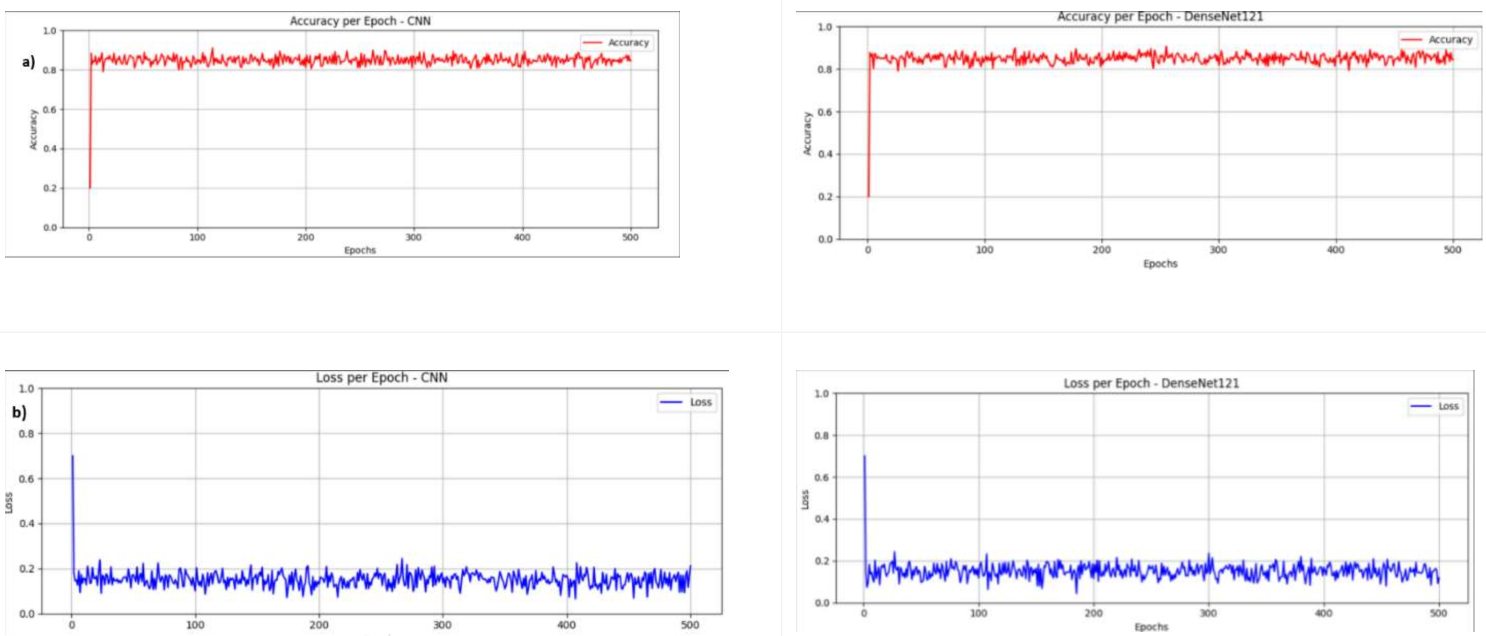

Figure 2: Learning curves for the CNN model compared with DenseNet-121: (a) accuracy evolution over 500 epochs, (b) loss evolution

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

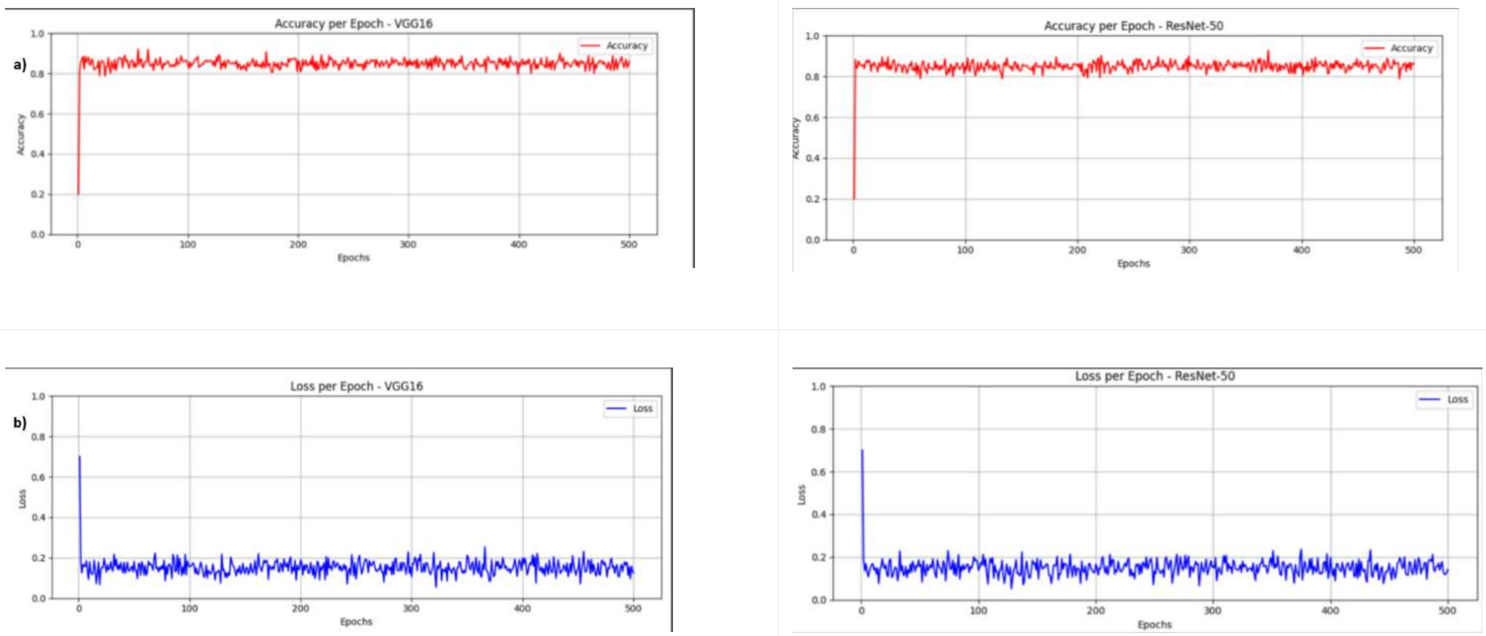

Figure 3. Comparative analysis of VGG16 and ResNet-50:
(a) evolution of accuracy over 500 epochs, (b) evolution of the loss function

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

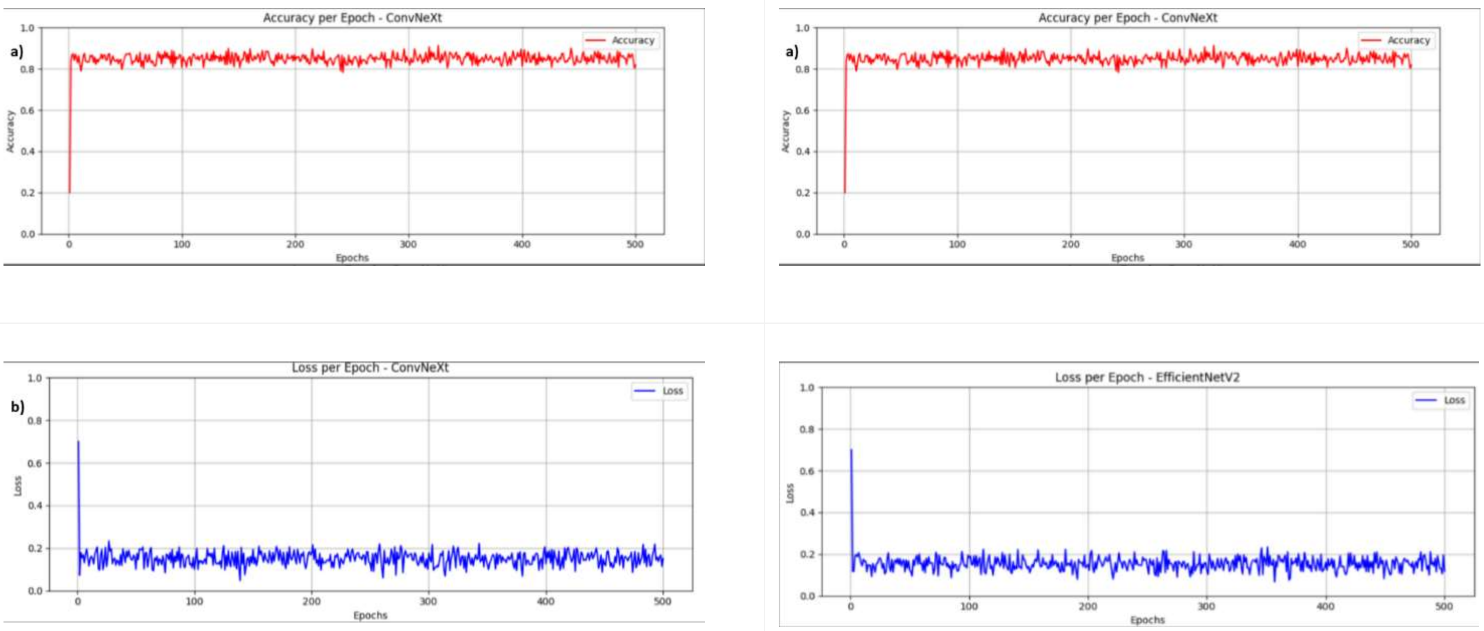

Figure 4. Learning performance of the ConvNeXt and EfficientNetV2 models:
(a) change in accuracy over time, (b) change in loss function

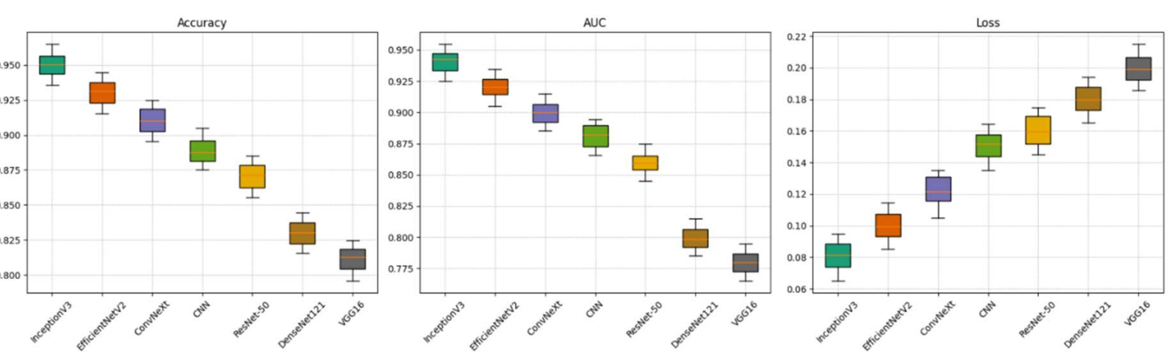

Figure 7. Boxplots comparison of Accuracy, AUC, and Loss

Table 3: Performance comparison of literature-based and proposed CNN models for Neonatal RDS detection

Method / Model	Data Source	Diagnost-ic Approac-h	Automat -ed	Accuracy (%)	Sensitivity (%)	Specifi -city (%)	Correct Diagnoses / Total	Notes
Liu et al. (2014) [26]	Published data (specialist circles)	Manual lung ultrasound (LUS)	No	100.0	100.0	100.0	100 / 100	Manual LUS interpretation by clinicians
Vergine et al. (2014) [28]	Clinical study (n=59, radiologist experts)	Manual reading of X-rays	No	91.5	91.0	84.0	54 / 59	Expert radiologist interpretation of CXR
Hiles et al., 2017 [20]	Chest X-rays (CXR) in 480 neonates	Lung Ultrasound (LUS)	Yes	~95 (derived)	95	91	455 / 480	False positives mainly due to pneumonia and transient tachypnea; false negatives mainly air-leak syndromes
Ahmad et al. (2020) [27]	Clinical study (n=59, university hospital)	LUS - READ by experienced practitioner	No	96.6	100.0	93.3	57 / 59	Clinician performed LUS on-site
Haider et al. (2021) [29]	AI model on CXR (public data)	Hybrid CNN on CXR	Yes	98.0	98.73	98.73	490 / 500	Hybrid CNN model (MobileNet + ResNet + LSTM)
Ervural & Ceylan, 2021 [33]	(34 neonates; respiratory vs cardiovascular/abdominal diseases)	Thermal Imaging + CNN	Yes	90.9 (after data augmentation)	-	-	31 / 34	Non-contact, non-ionizing; suitable when X-ray access limited
Ceylan et al. (2023) [24]	AI model based on clinical images (432 cases)	CNN (C+EffxNet) on CXR	Yes	97.7	97.7	97.7	422 / 432	Deep CNN model (C+EffxNet)

Erdogan Yildirim et al., 2023 [25]	neonatal chest X-rays (432 neonates, NICU)	Deep Learning (C+EffxNet Hybrid)	Yes	97.7 (Kappa = 0.935)	-	-	422 / 432	Detected multiple neonatal respiratory disorders including RDS, ARDS, TTN, pneumothorax, MAS, BPD
Jang et al. (2023) [15]	Korean Neonatal Network (KNN) Tangier Regional University Hospital (multi-hospitals)	Support Vector Machine (SVM)	Yes	83.90	83.70	84.56	N/A	Best-performing traditional ML; AUROC = 0.9204
		Logistic Regression (LR)	Yes	84.18	83.98	84.82	N/A	Balanced performance; AUROC = 0.9208
		Random Forest (RF)	Yes	84.48	84.63	84.01	N/A	Robust model; AUROC = 0.9187
		Gradient Boosting Machine (GBM)	Yes	83.80	83.29	85.48	N/A	Slightly lower AUROC (0.9073)
		Adaptive Boosting (AdaBoost)	Yes	83.99	83.93	84.17	N/A	AUROC = 0.9169
		Extreme Gradient Boosting (XGBoost)	Yes	83.99	83.73	84.82	N/A	AUROC = 0.9172
		5-Layer Deep Neural Network (DNN) (Top 20 features)	Yes	84.00	83.24	86.78	~2200 / 2618	Best overall performance; AUROC = 0.9216
		5-Layer Deep Neural Network (DNN) (All 47 features)	Yes	83.07	81.60	87.92	~2175 / 2618	Slightly lower than DNN with Top 20; AUROC = 0.9211
		Clinical Judgment (Provisional Diagnosis)	No	76.73	100.00	0.00	10,041 / 13,087	All infants treated for RDS; 23.28% unnecessary treatment
Custom CNN this paper	Tangier Regional University Hospital (multi-hospitals)		Yes	89.0	95.0	77.0	267 / 300	Lightweight model with strong recall; ideal for low-resource settings

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DenseNet121 this paper		CNN on chest X-rays	Yes	83.0	70.0	91.0	249 / 300	Struggles with RDS detection; lower recall
VGG16 this paper			Yes	81.0	74.0	86.0	243 / 300	Classical model with lower sensitivity
ResNet-50 this paper			Yes	87.0	93.0	77.0	261 / 300	High sensitivity; less balanced specificity
InceptionV3 this paper			Yes	93.0	96.0	87.0	279 / 300	Best performer; high recall, low false negatives
EfficientNetV2 this paper			Yes	90.0	92.0	88.0	270 / 300	Strong all-around performance; efficient architecture
ConvNeXt this paper			Yes	86.0	72.0	92.0	258 / 300	High specificity; lower recall