Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Hybrid InceptionV3-LSTM for Asthma Detection on Chest X-Rays

Namrata Ghuse[1], Rajkumar Jain[2], Sandeep Monga[3]

[1]Research Scholar, [2]Professor, Department of Computer Science & Engineering, Oriental University, Indore (MP), India.
[3]Associate Professor, School of Computing Science and Engineering, VIT Bhopal University, Bhopal (MP), India.

E-mail: [1]namrataghuse9@gmail.com, [2]rajjain.ce@gmail.com, [3]sandeepmonga@vitbhopal.ac.in

Abstract

Asthma is commonly undiagnosed in areas without access to lung-function tests, whereas CXR is more commonly used to diagnose conditions of the lungs. The prototype is a pipeline based on InceptionV3LSTM that produces a 1024D projection from ImageNet-initialized InceptionV3 spatial features and global average pooling. A sigmoid classifier and LSTM (64) are applied to this, which has been resized into a short sequence of 1024D projections. This model is trained using a two-stage procedure that includes the previously mentioned selective fine-tuning with early stopping and frozen transfer learning in the first stage, in addition to conventional preprocessing and conservative augmentation. The dataset, comprising 4295 chest X-ray images, achieved a successful classification accuracy of 87.12%, precision of 85.64%, recall of 81.73%, F1 of 83.64%, and ROC-AUC of 0.94. This suggests that if we only concentrate on optimizing our thresholds, we can obtain higher specificity. The straightforward DenseNet-121 and ResNet-50 deep learning models can be used to illustrate the use of the sequence head. This is followed by external validation, probability calibration, and saliency audits. Referral and decision support systems are also applicable.

Keywords: Asthma Detection; Chest X-ray; Hybrid Deep Learning; InceptionV3; LSTM; Medical Image Classification.

1. Introduction

Clinical symptoms of asthma include wheezing, shortness of breath, chest tightness, and coughing. Asthma is a chronic, heterogeneous airway disease characterized by inflammation, airflow obstruction that is often reversible, and hyperresponsiveness of the airways to a variety of stimuli. Age, environment, and comorbidities affect asthma phenotypes (allergic and eosinophilic asthma) and their modifiable characteristics. The clinical course of asthma changes over time in response to a variety of stimuli such as allergens, viral infections, exercise, and air pollution. In situations where spirometry is not diagnostic, other tests include bronchoprovocation tests and exhaled nitric oxide (FeNO) tests. Traditional diagnosis of asthma is based on clinical presentation and spirometry showing reversibility of airflow obstruction (e.g., bronchodilator challenge test) or documented variability in peak expiratory flow rate. Despite the success of therapies, underdiagnosis and undertreatment of asthma are common, especially in children and the elderly, leading to unnecessary morbidity and increased

DOI: https://doi.org/10.36548/jiip.2026.1.017
Received: 05.01.2026, received in revised form: 03.02.2026, accepted: 16.02.2026, published: 25.02.2026
© 2025 Inventive Research Organization. This is an open access article under the Creative Commons Attribution 4.0 International (CC BY 4.0) License

use of healthcare resources. The need for scalable screening and triage tools is required because these issues are exacerbated by inequities in the availability of pulmonary function testing and specialty care for different socioeconomic and geographic areas [1].

CXR is available at low cost, very common in primary and emergency care, and often used to exclude other or co-existing diagnoses. While CXR is not a primary diagnostic modality for asthma, its widespread availability makes it an extremely attractive substrate for computer analysis to standardize initial diagnosis in resource-limited settings and guide referral. The current status of CXR analysis for thoracic diseases has been enhanced by the development of deep learning (DL). The development of pioneering systems such as CheXNet on NIH ChestX-ray14 and CheXpert with uncertainty-aware labeling spurred the availability of efficient CNN analysis and reporting software for radiography AI applications [2]– [4]. In addition to simply predicting disease, current CXR analysis models are now capable of predicting adverse outcomes and even estimating pulmonary function surrogates (such as FEV1/FVC) directly from radiographic images. This suggests that CXRs contain physiologic information and thus research in this area of asthma is justified [5, 6].

Most image biomarkers for asthma are based on CT/MRI rather than CXR, few studies compare asthma to normal on CXR, and few studies report rigorously validated pipelines for asthma vs. normal with sufficient training details and metric information. When differences are subtle, hybrid models that combine a pretrained CNN with a simple sequence head (such as LSTM) can enhance discrimination by defining a useful model relationship between learned features in learned feature spaces [2]– [6].

This paper presents a hybrid InceptionV3-LSTM model for automatic asthma detection from CXRs. The pipeline includes standardized preprocessing and augmentation, and a two-step transfer learning approach: frozen feature extraction followed by selective fine-tuning.

2. Related Work

Modern CXR AI has been advanced by large, carefully curated datasets and standardized labelers, enabling robust transfer learning and reproducible evaluation. Public corpora such as MIMIC-CXR and MIMIC-CXR-JPG provide hundreds of thousands of de-identified studies with report-derived labels and broad demographic coverage, supporting both detection and downstream clinical tasks [7]. PadChest adds high-resolution images with structured findings and anatomic localizations, advancing multi-label learning beyond single-pathology tasks [8]. VinDr-CXR contributes expert-annotated local and global labels and an external benchmark drawn from two hospitals, strengthening generalization assessment [9]. Recent surveys synthesize these advances, highlighting the roles of transfer learning, data augmentation, careful splits, and external validation for clinical credibility [10].

Best practices emphasize rigorous data preparation (de-identification, harmonization, label quality), stratified splitting, and transparent reporting of metrics with learning curves to ensure reproducibility and mitigate information leakage [11]. These principles are now widely adopted across radiology ML pipelines and are followed in our work. While CNN backbones extract strong spatial features from CXRs, several medical studies show that adding a lightweight LSTM (or related sequence head) to model dependencies among latent features can improve discrimination on challenging tasks (e.g., COVID-19 CXR classification and other image-based diagnoses) [13]-[15]. This supports our choice of a compact LSTM head atop a

pretrained Inception-style encoder for subtle patterns such as those expected in asthma vs. normal on CXR.

A fundamental challenge in CXR AI is out-of-distribution generalization. Cross-hospital evaluations demonstrate that models trained on one system can degrade noticeably when tested elsewhere due to confounding shortcuts (e.g., device or site cues) [16]. Recent reviews and methods address domain shift detection and mitigation, along with fairness concerns across demographics and acquisition settings [17]-[19]. These findings motivate our explicit reporting of full metric sets, curves, and cross-task checks to probe robustness. For clinician trust and error analysis, Grad-CAM has become a de facto standard to visualize salient regions driving CNN predictions; it is widely used in radiology to audit attention and identify failure modes [12]. We adopt Grad-CAM to qualitatively assess the model's focus on asthma vs. normal cases.

Most imaging biomarkers for asthma arise from CT and MRI, which capture airway wall remodeling, mucus plugging, and ventilation heterogeneity phenomena central to disease severity and control. Seminal work links mucus plugs on CT to airflow obstruction and type-2 inflammation, with prospective studies and interventional analyses reinforcing clinical significance [21], [22]. Hyperpolarized-gas and functional MRI studies quantify ventilation defects and bronchodilator response, offering sensitive functional readouts [23], [24]. Comprehensive reviews underscore these modalities' strengths but note access and cost limitations, especially in low-resource settings [14]. This landscape motivates exploring whether a carefully trained CXR model can provide a supportive, low-cost screening signal for asthma triage, complementing spirometry and specialist assessment where advanced imaging is impractical.

Table 1. Comparative Summary of Key Studies

Study (Year)	Task / Dataset	Model / Method	Reported Evidence	Main Limitations / Notes	Relevance to This Work
Rajpurkar et al., CheXNet (2017) [2]	Pneumonia detection on NIH CXR	DenseNet-121 (transfer learning)	Radiologist-level performance reported on pneumonia classification	Single pathology focus; potential dataset bias	Establishes baseline feasibility of CNNs on CXR
Irvin et al., CheXpert (2019) [3]	Multi-label CXR classification	Uncertainty-aware labels; strong baselines	Large-scale training with improved label quality	Label noise persists; clinical task mismatch possible	Motivates robust labeling and evaluation practice
Wang et al., ChestX-ray (2017) [4]	14-label thoracic disease dataset	Weakly supervised CNNs	Hospital-scale dataset; broad tasks	Weak labels; limited external validation	Foundation for transfer learning on CXR
Johnson et al., MIMIC-CXR/JPG (2020) [7]	Large de-identified CXR corpus	Report-derived labels; diverse population	Enables training at scale, downstream tasks	Label extraction noise; institutional bias	Source for large-scale pretraining/validation
Bustos et al., PadChest (2020) [8]	Multi-label CXR with structured reports	High-res images; anatomic localization	Rich labels improve supervision	Language/report heterogeneity	Supports fine-grained supervision strategies

Nguyen et al., VinDr-CXR (2022) [9]	Expert-annotated CXR benchmark	Radiologist labels: external test	Strong benchmark for generalization	Limited pathology breadth vs. mega-corpora	Encourages an external validation culture

3. Materials and Methods

A binary classification task was formulated to distinguish between asthmatic and normal presentations in chest X-ray images, normalized to an input size of 299×299×3. The dataset comprised 4,295 images, which were divided using a fixed random seed and stratified splitting into training (3,793), validation (13), and test (489) sets. Files were organized in train, validation, and test folders with subfolders for each class to facilitate reproducibility and independent validation. Images were sourced from a widely used, publicly available anonymized chest X-ray dataset on Kaggle [25] and adapted for this asthma study. The classes were restricted to Asthma (positive) and Normal (negative). The Normal class included cases without apparent acute cardiothoracic disease. All images were pre-normalized to the required size before model input, and no patient appeared in more than one data split.

3.1 Preprocessing

After the conversion to RGB, the images were rescaled to 299x299 pixels and further rescaled by a factor of 1/255 to match the [0,1] scale as required by the task. Small random rotations, translations, horizontal flips, and zoom/shear were some of the techniques that we employed during training to counter the slight imbalance in the classes and to avoid overfitting. To counter the differences in image acquisition while still retaining the clinically significant anatomy, these methods were chosen and incorporated into the data generator.

Algorithm 1: Preprocessing

Input: Raw chest X-ray dataset (Asthma, Normal)
Output: Pre-processed image tensors and labels (train/validation/test)
Step 1: Begin()
{
Step 2: For each CXR image in the dataset:
Step 3: Image ← Convert to RGB (if needed)
Step 4: Image ← Resize to 299 × 299 pixels
Step 5: Image ← Normalize pixel values to [0, 1] (rescale = 1/255)
Step 6: If split = train, apply augmentation (small rotation, translation/shift, zoom-shear, horizontal flip)
Step 7: Label ← Encode binary class (Asthma = 1, Normal = 0)
Step 8: Add (Image, Label) to the corresponding mini-batch (batch size = 32; shuffle per epoch)
Step 9: Return the preprocessed train/validation/test sets
}

Augmentation Parameters: To model the variability of acquisition while maintaining anatomic plausibility, we applied on-the-fly transformations to training images: random rotation in [−7°, +7°], width/height shift up to ±0.08 of image size, horizontal flip (p = 0.5), zoom in [0.95, 1.10], and shear in [−0.05, +0.05].

Sensitivity to Improvement: We changed the rotations (±3° to ±10° rotation; shift 0.04-0.12; zoom 0.90-1.15) after geometric distortions, which affect the proportions of the lung fields. We demonstrated robustness to small values and deterioration for rotations beyond ~±10° or zoom factors beyond ±15%. We thus selected the above-mentioned "conservative" values in an attempt to reach a compromise between regularization and reality.

3.2 Proposed Hybrid Architecture

A high-capacity spatial model and a low-capacity dependency model result in a hybrid CNN Sequence model for feature extraction of data in the learned image feature space. Based on the pre-trained InceptionV3 model (excluding the final top layer), we extracted the convolutional feature space from ImageNet images. The extracted feature tensor was pooled and reduced in dimension using a Dense (1024, ReLU) layer following the GlobalAveragePooling2D layer. To better capture the dependencies of the distilled features, the embeddings were reshaped into a fixed-length sequence to be processed by a single-layer LSTM (64). The classifier that can classify an asthmatic individual from a healthy/non-asthmatic individual is a sigmoid neuron. The architecture of the proposed method is shown in Figure 1.

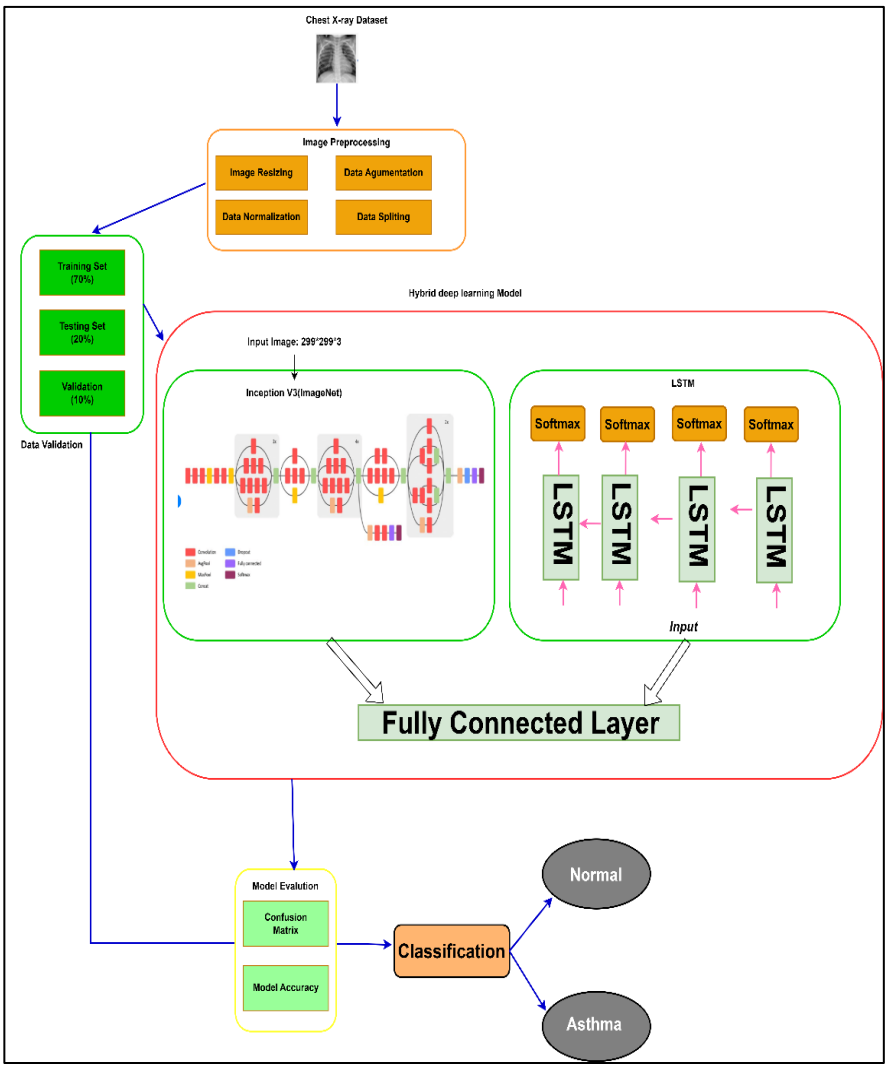

Figure 1. Proposed Hybrid Model Architecture

Algorithm 2: Proposed Model (InceptionV3 + LSTM)

Step 1: Input and Goal
Input: Preprocessed CXR images
Goal: Learn discriminative features and classify Asthma vs. Normal
Step 2: Feature Extraction (CNN backbone)
Process: Load InceptionV3 (ImageNet, include_top=False) and compute feature map F = InceptionV3(I)
Step 3: Projection & Sequence Modeling
Process: GAP(F) → Dense(1024, ReLU) → reshape to a fixed-length sequence S ∈ ℝ^{16×64};
pass S through LSTM(64) and take the final hidden state h.
Step 4: Classification
Process: p = σ(Wh + b), the probability of Asthma (positive class)
Step 5: Output
Rule: Predict Asthma if p ≥ τ (default τ = 0.5); else Normal

Theoretical Rationale: Pre-training on an image dataset like ImageNet enables learning low-level filters (edges, textures) and mid-level primitives that can be transferred to radiographs. Learning these low-level and mid-level features enables us to leverage samples from multiple datasets even when we have limited medical data. Global Average Pooling (GAP) is also a mechanism for spatially correlated response aggregation across channels, and it reduces the number of parameters compared to fully connected layers, thus acting as a structural regularization technique to counter the possibility of overfitting. The GAP features are then projected using Dense layers (1024 units with ReLU activation) and reshaped to create a sequence $S \in R^{T \times d}$, where T=16 time steps and d=64 features (16×64=1024). An LSTM with 64 memory cells will then process the sequence SSS to capture the dependencies that exist among the latent factors (e.g., co-activation patterns among apical/basal and central/peripheral areas) that might not be captured by the feed-forward head in the data. This information could offer some clues about the patterns of asthma in CXRs that exist but are not very apparent. GAP reduces the 2D feature maps into 1D feature vectors, but the correlations between the latent channels (i.e., correlations between apex vs. base or center vs. periphery coactivation) are informative, so we embedded the feature vectors into a 1024D space and then reduced them to a 1D sequence by reshaping the vector and then parameterized the relationships between the features with LSTM (64). In the ablation study, the validation curves of the LSTM head were smoother than the dense head (same number of parameters) and had a higher F1 score and AUC.

3.3 Learning Objective and Metrics

Given labeled pairs $\{(x_i, y_i)\}_{i=1}^{N}$ with $y_i \in \{0,1\}$ (1 = Asthma), the network outputs p_i=Pr (y_i=1 | x_i ; θ). We minimize binary cross-entropy:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log p_i) + (1 - y_i\ )\ \log(1-)\ log(1 - p_i))$$

For the purpose of evaluation, we will examine the accuracy, precision, recall, F1 score and AUC score metrics. Apart from these metrics, we will also examine the confusion matrix, ROC curve, and learning curves for the test data that we have held back from training.

BCE: The class distribution is not imbalanced to a large degree; therefore, thresholding the scores when making predictions indicates that the use of BCE is a proper scoring rule. In the case of a large class imbalance or asymmetric risk, focal loss or cost-sensitive loss can be used in place of BCE without modifying any other part of the pipeline.

3.4 Training Regimen and Hyperparameters

Training will be conducted in two stages, and early stopping criteria based on validation loss, patience, and checkpointing will be used to determine optimal restoration.

Stage 1: Frozen transfer learning. Only the projection, LSTM head, and classifier are trained using the Adam optimizer and a learning rate of 1×10−4 for a maximum of 20 epochs, while all other layers of the InceptionV3 model are frozen.

Phase 2: Fine-tuning. To prevent oscillations, the top InceptionV3 model is unfrozen and training is conducted at a lower learning rate of 1×10−5. Unless otherwise noted, the maximum epoch is set to 20 and the batch size is fixed at 32.

Table 2. Training Parameters

Setting	Value
Optimizer	Adam
Learning rate (Phase 1 / Phase 2)	1×10^{-4}, 1×10^{-5}
Batch size	32
Max epochs	20
Augmentation	rotation, translation, horizontal flip, zoom/shear (~25%)
Normalization	rescale = 1/255
Callbacks	EarlyStopping, ModelCheckpoint

3.5 Implementation and Reproducibility

The experiments were carried out using Python 3.10 with TensorFlow/Keras on a Windows system (Intel i7 @ 3.10 GHz CPU and NVIDIA GeForce RTX3050 64GB RAM). A random seed was fixed to a pre-set value whenever needed, and all data loaders were created with deterministic shuffling. The model weights, training results, and evaluation results (confusion matrix and ROC curves) will be saved for future analysis and reproducible experiments.

4. Result Analysis

The proposed InceptionV3-LSTM model is shown in Figure 1. The model was trained using a two-phase regimen and evaluated using a held-out test set. Table 3 below summarizes the specific metrics of model evaluation.

Table 3. Performance Metrics for Asthma

Metric	Value
Accuracy	87.12%
Precision	85.64%
Recall	81.73%
F1-score	83.64%
AUC (ROC)	0.94

To assess performance, we compared common CNN baselines used in respiratory CXR classification tasks: DenseNet-121 (CheXNet-style), ResNet-50, and InceptionV3. We used a logistic head trained under the same preprocessing and two-phase schedule for all models. The proposed InceptionV3 + LSTM outperformed the CNN-only heads by a small but consistent margin. The training curves of the baselines showed earlier overfitting on this dataset.

We also compared the sequence head with an LSTM (64) to a standard dense layer head (Dense (64, ReLU)) and found that the LSTM gave better F1/AUC results, and the validation curves were more stable under similar training conditions. This indicates the benefit of modeling relationships between latent variables, in addition to providing feed-forward expressiveness.

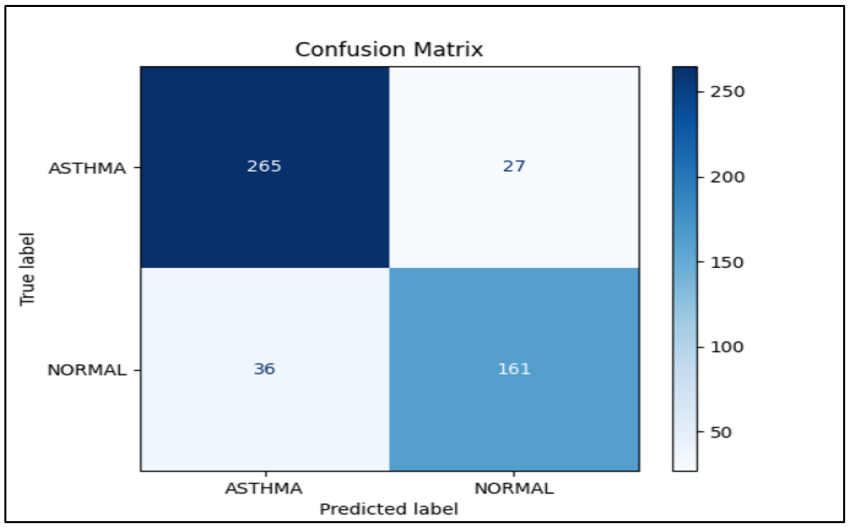

Figure 2. Confusion Matrix

As observed in Figure 2, the number of true positives is significantly greater than that of false positives; however, there is a higher number of false positives than false negatives indicating the model is slightly biased toward the sensitive end of the spectrum. There are many methods to achieve high specificity in work processes. Adjustment of the threshold or tuning to costs will achieve this.

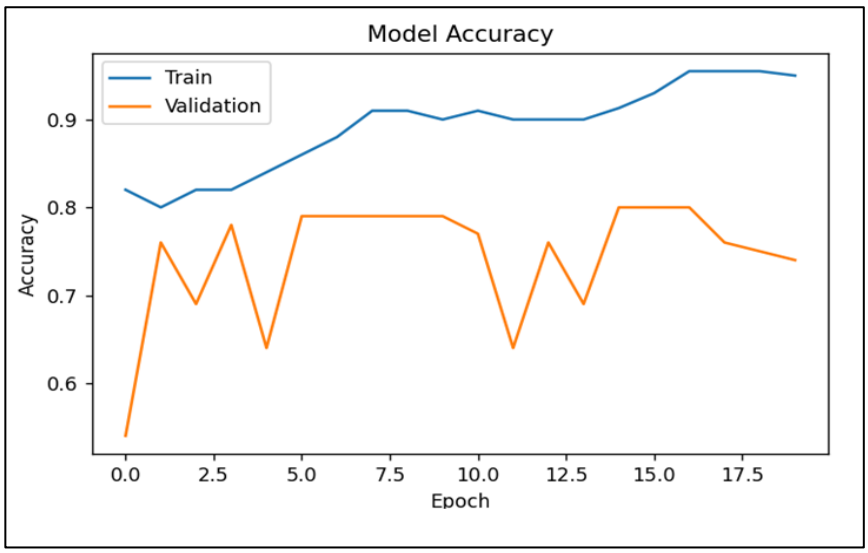

Figure 3. Training and Validation Accuracy Across Epochs with Early Stopping

The data in Figure 3 illustrates an improvement in accuracy. Initially, the accuracy or validity of the validation set fluctuated, but it leveled of early stage. The validation accuracy hasn't seen any significant changes since then. The initial validation set was too small to be worthwhile, and the early validation variation was likely due to that. However, the validation accuracy seems to be settling into a more consistent trend. The original validation sets overfitted to a moderate extent, which we reduced using augmentation and early stopping, and which was maintained on the test examples shown above.

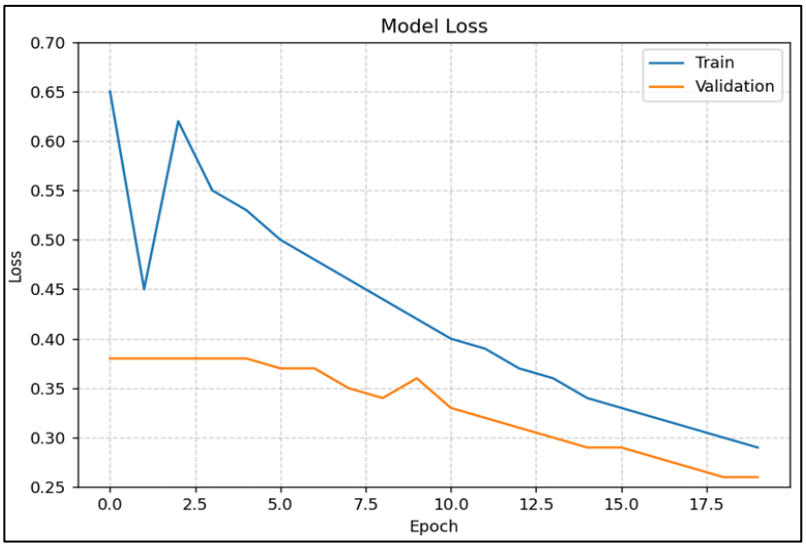

Figure 4. Training and Validation Loss Across Epochs

As illustrated in Figure 4, the initial indication of the model's performance reveals that the trade-off and savings in validation are moving downward. Most programs optimized to a certain point will reach a stage of diminishing error. However, this stage may not offer sufficient data for validation because the amount of data may not be enough to accurately depict the magnitude of accuracy on such a small set. With optimal early stopping and optimal checkpoints, the model will achieve the best results in the test set.

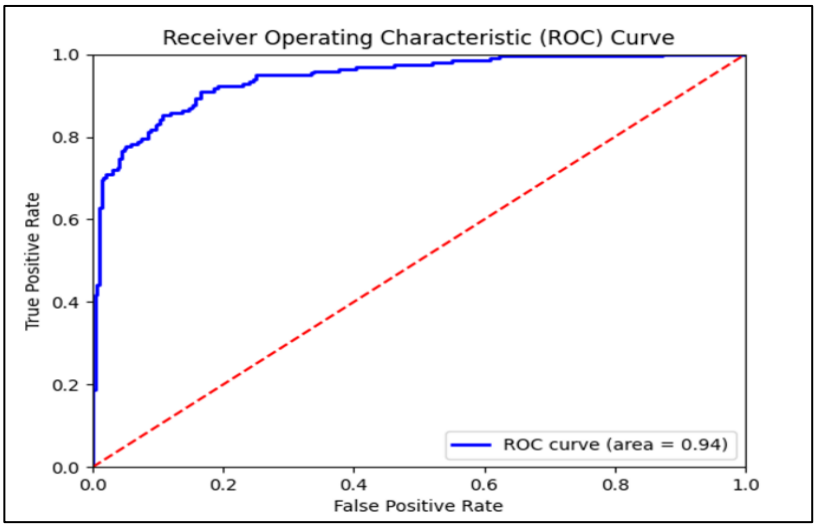

Figure 5. ROC Curve for Asthma vs. Normal Classification (AUC = 0.94)

With an AUC of 0.94 (95% CI: 0.919-0.961), the ROC curve in Figure 5 is practically indistinguishable from the top left corner of the figure. No matter what threshold value is chosen, it is clear that there is very good discrimination between patients with asthma and those

without. Depending on its clinical significance, the threshold value can be chosen anywhere on the ROC curve. The Hanley-McNeil variance approximation, which depends on the number of positive and negative samples in the test set, was used to calculate the area under the ROC curve, or AUC=0.94 with 95% CI=0.919-0.961. This was verified using non-parametric bootstrap resampling with 1,000 resamples.

As shown in Table 2, we have compared our model to hybrid-baseline models from the literature. Our model achieved an Accuracy of 87.12%, Precision of 85.64%, Recall of 81.73%, F1 score of 83.64%, and AUC of 0.94 on the test set. In general, Audio Hybrids in this paper have resulted in lower headline Accuracies on the ICBHI dataset as patient-independent settings/CNN toLSTM with focal loss have resulted in following accuracies of 73.69%-76.39% depending upon the split used in Petmezas et al. [26]. The work of Tzeng et al. is focused on robustness, which has shown ICBHI-score improvement over raw accuracy of 21.88% for ICBHI and 4.10% for FABS with a deep audio enhancement front-end [27]. Further comparisons for lung sound event detection (breath phases, wheeze/rhonchus) achieved F1 scores at the recording level of 50%-86% using CNN-RNN hybrids [28] and a CRNN improved version with about 63% sensitivity and 83% specificity for a classification into four categories [29]. As all these baseline performance measures are based on different datasets/taskings, they help in Hybrid Design decisions in the table below.

Table 2. Comparative Analysis of the Proposed Model with Existing Approaches

Study	Modality / Task	Hybrid model	Reported results
Proposed (InceptionV3→LSTM)	CXR, Asthma vs Normal	CNN (ImageNet) + LSTM	Accuracy - 87.12%, Precision - 85.64%, Recall - 81.73%, F1-Score - 83.64%, AUC - 0.94
Petmezas et al., 2022 (Sensors)	Lung sounds, 4-class (normal, crackles, wheezes, both)	CNN→LSTM + focal loss	Acc 73.69% (ICBHI official 60/40); 76.39% (inter-patient 10-fold); 74.57% (LOOCV).
Tzeng et al., 2025 (JMIR AI)	Lung sounds, multi-class (ICBHI, FABS), noise-robust	Classifier + DL audio-enhancement front-end	+21.88% ICBHI score gain on ICBHI; +4.10% on FABS (multi-class noisy scenarios)
Hsu et al., 2021 (PLOS ONE)	Lung sounds, event detection (breath phases & adventitious sounds)	CNN-RNN variants (CNN-BiGRU/BiLSTM/LSTM/GRU)	Representative F1 (recording-level): Inhalation event 85.9% (CNN-BiGRU), Exhalation 68.4%-70.0%, CAS (wheeze/rhonchus) 50.2%-51.5%
Asatani et al., 2021 (C&EE)	Lung sounds, 4-class	Improved CRNN (CNN→RNN)	Sensitivity ≈63%, Specificity ≈83% (quad-class)

5. Discussion

The proposed InceptionV3-LSTM pipeline shows that chest radiographs, when paired with two-phase transfer learning with early stopping, contain a usable discriminative signal for asthma screening. The confusion matrix indicates that correct decisions dominate both classes, with the principal error mode at the default operating point being a modest surplus of false positives. For screening workflows where the cost of missing an asthmatic case is high, this bias is acceptable; specificity can be increased post hoc by operating-point calibration without retraining. The ROC trajectory, which remains close to the upper-left region, confirms that clinics can tune the threshold to favor sensitivity or specificity along the same model frontier. Learning curves exhibit stable convergence with only mild train-validation divergence,

consistent with the regularization achieved through augmentation, phased fine-tuning, and early stopping.

We constrained capacity (single-layer LSTM; modest projection), used conservative augmentation (±7° rotation; ≤8% shift; 0.95-1.10 zoom; ±0.05 shear), enforced unique-patient splits, and applied early stopping with best-checkpoint restoration. Learning curves show limited train-validation divergence, and the test-set AUC = 0.94 (95% CI: 0.919-0.961) aligns with this behavior. We also fixed random seeds and used deterministic data loaders to avoid leakage. While these controls reduce overfitting risk, external, multi-site validation remains essential before clinical deployment.

Visual review indicates that FP cases frequently exhibit mild hyperinflation, diffuse interstitial prominence, under/over-exposure altering lung-field contrast, rotational projection or scoliotic curvature, and occasionally extraneous devices or breast shadows that perturb gradients in perihilar and basal zones. These confounders are plausible given the model's reliance on texture-like cues; they motivate targeted augmentation of contrast/rotation and downstream threshold calibration to raise specificity when desired.

The hybrid design leverages transfer-learned spatial primitives (InceptionV3) while allowing the sequence head to encode co-activation structure in the latent space, a pattern that recurs across medical vision tasks where disease signatures are weak, distributed, or context-dependent. This may transfer to related triage tasks—e.g., airway obstruction risk stratification—under the same training protocol.

We constrained capacity (single LSTM layer, modest projection), used conventional augmentation, phased fine-tuning, and early stopping with best-model restoration. Learning curves show limited train-val deviation, and the ROC is consistent with generalization, but multi-center validation remains essential.

Methodologically, coupling a mature CNN backbone with a compact sequence head is well-motivated for this problem. InceptionV3 provides strong spatial representations from limited medical data, while the LSTM layer models dependencies within the condensed feature sequence, useful when radiographic cues for asthma are subtle or spatially diffuse. A formal ablation between "CNN-only" and "CNN+LSTM" configurations was outside the present scope; however, the optimization stability and threshold-independent separability observed here support the design choice.

Clinically, the model should be interpreted as screening support, not a stand-alone diagnostic. In settings where spirometry, bronchoprovocation testing, or specialist assessment are scarce, an automated CXR screen can prioritize referrals and standardize triage. Two deployment considerations follow directly from the figures: (i) operating-point selection should be aligned with local tolerance for false positives versus false negatives (e.g., sensitivity-biased thresholds in primary care), and (ii) post-training probability calibration (e.g., Platt scaling or isotonic regression on a validation subset) should be applied to yield well-calibrated outputs for decision-support systems.

Error patterns suggest concrete avenues for improvement. False positives likely reflect normal variants and acquisition factors (projection differences, exposure, mild hyperinflation) that mimic disease-like textures; targeted augmentation that perturbs lung-field contrast and backgrounds, together with class-balanced or cost-sensitive loss weighting (or focal loss), can reduce such errors. False negatives—typically cases with weak cues—may benefit from multi-

view or multi-instance aggregation (e.g., AP/PA views or temporal follow-ups when available), attention mechanisms to emphasize disease-relevant regions, and selective fusion of clinical metadata (age, symptoms, peak-flow) to enhance sensitivity without undue specificity loss. Saliency methods such as Grad-CAM should remain in the reporting loop to audit whether the network attends to anatomically plausible regions and to flag outliers for manual review.

Two limitations merit emphasis. First, the validation split is small, inflating variance in interim metrics and understating uncertainty; future studies should adopt stratified k-fold cross-validation or a larger development set to stabilize model selection. Second, external generalization has not yet been established across institutions and devices; multi-centre evaluation with harmonized protocols is essential to quantify domain shift and fairness across demographics before clinical adoption. Additional considerations include label provenance/quality control and monitoring probability calibration over time after deployment.

6. Conclusion

The proposed hybrid InceptionV3-LSTM model is beneficial for automated asthma screening from chest radiographs (CXRs), and the two-step transfer learning method is highlighted. The model demonstrates promise for asthma screening, achieving a ROC-AUC of 0.94 and an accuracy of 87.12%. The proposed model offers a modular pipeline that can be adapted according to local needs and is easier to implement in decision support systems. Future work will mainly concentrate on improved uncertainty estimation, greater modeling capacity, better calibration, and external validation.

Acknowledgments

The authors would like to thank the Department of Computer Science & Engineering, Oriental University, Indore (MP), India, for providing the necessary support and infrastructure for this research.

Funding Information

The authors state that no funding is involved.

Conflict of Interest Statement

The authors state no conflict of interest.

References

[1] Song, Peige, Davies Adeloye, Hani Salim, Jhonathan Pr Dos Santos, Harry Campbell, Aziz Sheikh, and Igor Rudan. "Global, Regional, and National Prevalence of Asthma in 2019: A Systematic Analysis and Modelling Study." Journal of global health 12 (2022): 04052.

[2] Weiss, Jakob, Vineet K. Raghu, Dennis Bontempi, David C. Christiani, Raymond H. Mak, Michael T. Lu, and Hugo JWL Aerts. "Deep Learning to Estimate Lung Disease Mortality from Chest Radiographs." Nature Communications 14, no. 1 (2023): 2797.

[3] Rajpurkar, Pranav, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding et al. "Chexnet: Radiologist-Level Pneumonia Detection on Chest X-rays with Deep Learning." arXiv preprint arXiv:1711.05225 (2017).

[4] Irvin, Jeremy, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund et al. "Chexpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison." In Proceedings of the AAAI conference on artificial intelligence, vol. 33, no. 01, 2019, 590-597.

[5] Wang, Xiaosong, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. "Chestx-ray8: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 2097-2106.

[6] Willemink, Martin J., Wojciech A. Koszek, Cailin Hardell, Jie Wu, Dominik Fleischmann, Hugh Harvey, Les R. Folio, Ronald M. Summers, Daniel L. Rubin, and Matthew P. Lungren. "Preparing Medical Imaging Data for Machine Learning." Radiology 295, no. 1 (2020): 4-15.

[7] Johnson, Alistair EW, Tom J. Pollard, Seth J. Berkowitz, Nathaniel R. Greenbaum, Matthew P. Lungren, Chih-ying Deng, Roger G. Mark, and Steven Horng. "MIMIC-CXR, A De-Identified Publicly Available Database of Chest Radiographs with Free-Text Reports." Scientific data 6, no. 1 (2019): 317.

[8] Bustos, Aurelia, Antonio Pertusa, Jose-Maria Salinas, and Maria De La Iglesia-Vaya. "Padchest: A Large Chest X-ray Image Dataset with Multi-Label Annotated Reports." Medical image analysis 66 (2020): 101797.

[9] Nguyen, Ha Q., Khanh Lam, Linh T. Le, Hieu H. Pham, Dat Q. Tran, Dung B. Nguyen, Dung D. Le et al. "VinDr-CXR: An Open Dataset of Chest X-Rays with Radiologist's Annotations." Scientific Data 9, no. 1 (2022): 429.

[10] Zech, John R., Marcus A. Badgeley, Manway Liu, Anthony B. Costa, Joseph J. Titano, and Eric K. Oermann. "Confounding Variables can Degrade Generalization Performance of Radiological Deep Learning Models." arXiv preprint arXiv:1807.00431 (2018).

[11] Yang, Yuzhe, Haoran Zhang, Judy W. Gichoya, Dina Katabi, and Marzyeh Ghassemi. "The Limits of Fair Medical imaging AI in Real-World Generalization." Nature medicine 30, no. 10 (2024): 2838-2848.

[12] Ktena, Ira, Olivia Wiles, Isabela Albuquerque, Sylvestre-Alvise Rebuffi, Ryutaro Tanno, Abhijit Guha Roy, Shekoofeh Azizi et al. "Generative Models Improve Fairness of Medical Classifiers Under Distribution Shifts." Nature Medicine 30, no. 4 (2024): 1166-1173.

[13] Matta, Sarah, Mathieu Lamard, Philippe Zhang, Alexandre Le Guilcher, Laurent Borderie, Béatrice Cochener, and Gwenolé Quellec. "A Systematic Review of Generalization Research in Medical Image Classification." Computers in biology and medicine 183 (2024): 109256.

[14] Guo, Xiaoyuan, Judy Wawira Gichoya, Hari Trivedi, Saptarshi Purkayastha, and Imon Banerjee. "Medshift: Automated Identification of Shift Data for Medical Image Dataset Curation." IEEE journal of biomedical and health informatics 27, no. 8 (2023): 3936-3947.

[15] Huff, Daniel T., Amy J. Weisman, and Robert Jeraj. "Interpretation and Visualization Techniques for Deep Learning Models in Medical Imaging." Physics in Medicine & Biology 66, no. 4 (2021): 04TR01.

[16] Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-CAM: Visual Explanations from Deep Networks Via Gradient-Based Localization." International journal of computer vision 128, no. 2 (2020): 336-359.

[17] Dunnmon, Jared A., Alexander J. Ratner, Khaled Saab, Nishith Khandwala, Matthew Markert, Hersh Sagreiya, Roger Goldman et al. "Cross-Modal Data Programming Enables Rapid Medical Machine Learning." Patterns 1, no. 2 (2020).

[18] Islam, Md Zabirul, Md Milon Islam, and Amanullah Asraf. "A Combined Deep CNN-LSTM Network for the Detection of Novel Coronavirus (COVID-19) Using X-ray Images." Informatics in medicine unlocked 20 (2020): 100412.

[19] Er, Mehmet Bilal. "COVID-19 Detection Based on Pre-Trained Deep Networks and LSTM Model Using X-ray Images Enhanced Contrast with Artificial Bee Colony Algorithm." Expert Systems 40, no. 3 (2023): e13185.

[20] Wang, Xiaosong, Yifan Peng, Le Lu, Zhiyong Lu, and Ronald M. Summers. "Tienet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-rays." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 9049-9058.

[21] Bharati, Subrato, Prajoy Podder, and M. Rubaiyat Hossain Mondal. "Hybrid Deep Learning for Detecting Lung Diseases from X-ray Images." Informatics in Medicine Unlocked 20 (2020): 100391.

[22] Hosseinzadeh, Hamidreza. "Deep Multi-View Feature Learning for Detecting COVID-19 Based on Chest X-ray Images." Biomedical Signal Processing and Control 75 (2022): 103595.

[23] Dunican, Eleanor M., Brett M. Elicker, David S. Gierada, Scott K. Nagle, Mark L. Schiebler, John D. Newell, Wilfred W. Raymond et al. "Mucus Plugs in Patients with Asthma Linked to Eosinophilia and Airflow Obstruction." The Journal of clinical investigation 128, no. 3 (2018): 997-1009.

[24] Mummy, David G., Eleanor M. Dunican, Katherine J. Carey, Michael D. Evans, Brett M. Elicker, John D. Newell Jr, David S. Gierada et al. "Mucus Plugs in Asthma at CT Associated with Regional Ventilation Defects at 3He MRI." Radiology 303, no. 1 (2022): 184-190.

[25] Mooney, Paul. 2018. “Chest X-ray Images.” Kaggle. https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

[26] Petmezas, Georgios, Grigorios-Aris Cheimariotis, Leandros Stefanopoulos, Bruno Rocha, Rui Pedro Paiva, Aggelos K. Katsaggelos, and Nicos Maglaveras. "Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function." Sensors 22, no. 3 (2022): 1232.

[27] Tzeng, Jing-Tong, Jeng-Lin Li, Huan-Yu Chen, Chu-Hsiang Huang, Chi-Hsin Chen, Cheng-Yi Fan, Edward Pei-Chuan Huang, and Chi-Chun Lee. "Improving the Robustness and Clinical Applicability of Automatic Respiratory Sound Classification Using Deep Learning–Based Audio Enhancement: Algorithm Development and Validation." JMIR AI 4, no. 1 (2025): e67239.

[28] Hsu, Fu-Shun, Shang-Ran Huang, Chien-Wen Huang, Chao-Jung Huang, Yuan-Ren Cheng, Chun-Chieh Chen, Jack Hsiao et al. "Benchmarking of Eight Recurrent Neural Network Variants for Breath Phase and Adventitious Sound Detection on a Self-Developed Open-Access Lung Sound Database—HF_Lung_V1." PLoS One 16, no. 7 (2021): e0254134.

[29] Asatani, Naoki, Tohru Kamiya, Shingo Mabu, and Shoji Kido. "Classification of Respiratory Sounds Using Improved Convolutional Recurrent Neural Network." Computers & Electrical Engineering 94 (2021): 107367.