

PulmoScan AI: Automated Chest Radiograph Analysis for Pneumonia Detection and Clinical Decision Support

Asst. prof .Sushma S
Dept.of Computer Science &
Engineering
Mysuru Royal Institute Of Technology

Siddaraju R
Dept.of Computer Science &
Engineering, Mysuru Royal Institute
Of Technology

Stany Stephen Raj J
Dept.of Computer Science &
Engineering
Mysuru Royal Institute Of Technology

Shiva Prasad
Dept.of Computer Science & Engineering
Mysuru Royal Institute Of Technology

Pavan Kumar S
Dept.of Computer Science & Engineering
Mysuru Royal Institute Of Technology

Abstract—Manual screening of chest radiographs is time-consuming, and many automated systems return a bare class label without indicating how reliable it is or which image regions drove it. This paper presents PulmoScan AI, a web-based clinical decision-support system for preliminary pneumonia risk assessment from frontal chest X-rays. A pretrained DenseNet-121 extracts a 1024-dimensional feature vector. A class-weighted Linear Support Vector Machine (SVM) classifies it, and Platt-scaled probability calibration converts the margin into a posterior that is mapped to LOW, MEDIUM and HIGH risk tiers. A Grad-CAM-style activation map, derived from the penultimate feature maps and the classifier weights, gives visual explanation with hemithorax localization. The system also provides JWT-secured access with administrator two-factor authentication, MongoDB GridFS storage and structured PDF reporting. The model was trained on the NIH ChestX-ray14 dataset (112,120 radiographs, 30,805 patients) with a patient-level 80/20 split. On the held-out test set of 22,784 images it achieved 88.42% accuracy, 85.23% sensitivity, 88.46% specificity and 86.85% balanced accuracy. Precision was low at 8.92% (F1 = 16.14%) because pneumonia accounts for only 1.28% of images. The results indicate that pairing deep features with a calibrated linear classifier is a viable, low-cost triage aid. Precision under extreme class imbalance and the lack of external validation remain the main limitations.

Index Terms—Pneumonia detection, chest X-ray, DenseNet-121, support vector machine, probability calibration, Grad-CAM, clinical decision support.

I. INTRODUCTION

Pneumonia is an infection of the lung parenchyma that remains a leading cause of morbidity and mortality worldwide. Delayed recognition raises the risk of complications such as respiratory failure and sepsis. The chest radiograph is the most widely available imaging modality for suspected pneumonia, but its interpretation depends on trained radiologists. In high-volume or resource-limited settings, reading delays and inter-reader variability can affect the timeliness of care. Computer-aided analysis of

chest radiographs has therefore attracted sustained research interest.

Since the release of large labelled corpora such as ChestX-ray14 [1], convolutional neural networks (CNNs) have been widely applied to thoracic disease recognition, and CheXNet [2] reported radiologist-level pneumonia detection with a DenseNet-121 backbone [3]. A practical route to reuse these models is transfer learning, in which a pretrained CNN acts as a fixed feature extractor and a lightweight classical classifier is trained on the extracted representations. This decouples representation learning from decision making and makes classifier experimentation inexpensive.

Despite this progress, several gaps persist. First, many studies report only a class label or an uncalibrated score, although clinical users need a probability that can be interpreted and thresholded into actionable tiers. SVM decision values are margins, not probabilities, and are rarely calibrated in the surveyed work. Second, black-box predictions without spatial explanation limit trust and auditability. Third, most published pipelines stop at model evaluation and do not provide the surrounding workflow of authenticated upload, record management, explanation and reporting that a deployable decision-support tool requires. Fourth, evaluation is often conducted on balanced, curated subsets, which hides the precision problem that arises at realistic prevalence.

This paper addresses these gaps with PulmoScan AI. The principal contributions are:

- A hybrid pipeline in which a pretrained DenseNet-121 supplies 1024-dimensional descriptors to a class-weighted Linear SVM, evaluated on the full ChestX-ray14 corpus under a patient-level split with no patient overlap.
- A Platt-based probability calibration stage (5-fold sigmoid cross-validation) that converts SVM margins

into posteriors and maps them to LOW (<0.35), MEDIUM (0.35 to <0.65) and HIGH (≥ 0.65) risk tiers.

- An explainability module that produces Grad-CAM-style heatmaps and overlays with patient-right, patient-left or bilateral localization, and an "unavailable" state when activation evidence is unreliable.
- An end-to-end web application (React and Flask) with JWT authentication, role-based access, administrator 2FA, MongoDB Atlas/GridFS storage, patient history and automated PDF reporting.
- A candid evaluation on an extremely imbalanced test set (1.31% positives), reporting sensitivity, specificity, precision, F1, balanced accuracy, MCC, likelihood ratios and prevalence-dependent predictive values.

The remainder of the paper is organized as follows. Section II reviews related work. Section III describes the methodology. Section IV presents experiments and discussion. Section V concludes and outlines future work.

II. RELATED WORK

A. Deep learning on large chest radiograph corpora. Wang et al. [1] introduced ChestX-ray8/14, whose NLP-derived labels enabled large-scale supervised training and weakly supervised localization. Rajpurkar et al. [2] trained a 121-layer DenseNet [3] on this data and reported performance comparable to radiologists on pneumonia detection. DenseNet's dense connectivity encourages feature reuse and yields compact, discriminative penultimate representations, which motivates its use here as a frozen feature extractor.

B. Hybrid CNN and classical classifier pipelines. A large body of recent work combines CNN features with SVM or Random Forest classifiers. Meindiawan and Muljono [13] paired MobileNetV2 features with an SVM, Sharma et al. [15] used EfficientNet features, Kumar et al. [16] applied transfer learning with an SVM, and Zhang et al. [14] used a multi-scale CNN with an SVM to capture lesions at several scales. Araújo et al. [11] compared several CNN backbones with SVM and Random Forest classifiers. These studies show that deep features and a shallow classifier are a competitive design, but the surveyed descriptions focus on classification and do not describe an integrated reporting workflow.

C. Feature fusion and optimization. Akshay et al. [9] fused VGG16 and ResNet features with PCA before SVM/RF classification. Hashmi et al. [10] combined ResNet50 features with handcrafted DWT and GLCM texture descriptors. Vijayakumar et al. [12] applied the Harris-Hawks optimizer to select deep features, and Chen et al. [17] fused deep features prior to SVM classification. These approaches can improve representation quality but add pipeline complexity and hyper-parameters, and their evaluation is typically on binary, curated pneumonia datasets.

D. Explainability and calibration. Grad-CAM [4] and CAM [8] localize class-discriminative regions and are the de facto tools for visual justification of CNN decisions. Separately, Platt scaling [7] converts SVM margins into probabilities. Both ingredients are well established, yet their joint use with tiered risk output and a deployable clinical workflow is seldom reported for pneumonia detection.

E. Summary and positioning. The literature demonstrates that CNN-plus-SVM designs are effective, but three aspects remain under-addressed. These are calibrated risk stratification, spatial explanation tied to the classifier itself, and an integrated secure workflow from upload to report. PulmoScan AI fills this gap by combining DenseNet-121 features, a calibrated Linear SVM, Grad-CAM-style localization and a complete web-based decision-support stack, and by evaluating it on the full, highly imbalanced ChestX-ray14 distribution rather than a balanced subset.

III. PROPOSED METHODOLOGY

A. System architecture. PulmoScan AI follows a five-stage pipeline (Fig. 1): (1) authenticated image upload through a React interface; (2) validation and preprocessing in a Flask REST backend; (3) feature extraction with DenseNet-121 and classification with a Linear SVM; (4) probability calibration, risk tiering and explanation; and (5) result display, PDF report generation and persistent storage. Original radiographs, heatmaps, overlays and PDF reports are stored in MongoDB GridFS, while analysis metadata and user records reside in MongoDB Atlas. Access is controlled by bcrypt-hashed credentials, JWT tokens, role-based permissions, CAPTCHA at registration and OTP-based two-factor authentication for administrators.

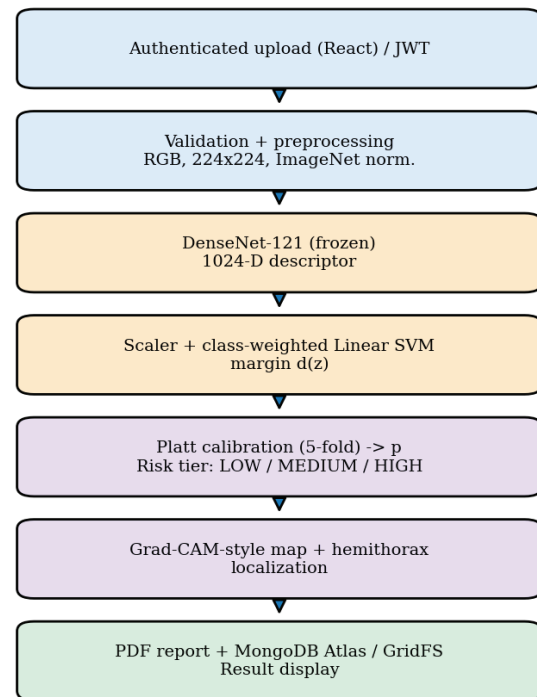


Fig. 1. System block diagram of the PulmoScan AI pipeline.

B. Validation and preprocessing. Uploads are checked for supported file type and minimum size, and invalid files are rejected. Valid images are converted to three-channel RGB and resized to 224×224 pixels. Pixel intensities are scaled to $[0,1]$ and normalized with ImageNet statistics:

$$\tilde{x}c = (xc - \mu c) / \sigma c \quad (1)$$

with $\mu = (0.485, 0.456, 0.406)$ and $\sigma = (0.229, 0.224, 0.225)$.

C. Deep feature extraction. The classification head of an ImageNet-pretrained DenseNet-121 is removed. The final $7 \times 7 \times 1024$ feature maps A are global-average-pooled to give the descriptor

$$f = (1/49) \sum_{i,j} A(i,j) \in \mathbb{R}^{1024} \quad (2)$$

The backbone is used frozen, so no CNN fine-tuning is performed and only the classifier stage is trained.

D. Class-weighted Linear SVM. Descriptors are standardized, $z = (f - m) \oslash s$, using training-set statistics. A Linear SVM (LinearSVC) then minimizes the regularized squared-hinge loss:

$$\min \frac{1}{2} \|w\|^2 + C \sum_i c(y_i) \cdot \max(0, 1 - y_i(w^T z_i + b))^2 \quad (3)$$

where $y \in \{-1, +1\}$ (pneumonia = +1) and the balanced class weight is $c(y) = N / (2N_y)$. Because pneumonia is only 1.28% of the corpus, this weighting prevents the optimizer from collapsing to the majority class. The decision value is $d(z) = w^T z + b$, a signed margin and not a probability.

E. Probability calibration. To obtain interpretable posteriors, $d(z)$ is passed through a sigmoid (Platt) mapping fitted by 5-fold cross-validation (CalibratedClassifierCV):

$$p = P(y=1 | d) = 1 / (1 + \exp(A \cdot d + B)) \quad (4)$$

Parameters A and B are estimated by maximum likelihood on out-of-fold decision values, and the final probability is averaged across folds.

F. Risk stratification. The calibrated probability is mapped to a tier:

Risk(p) = LOW if $p < 0.35$; MEDIUM if $0.35 \leq p < 0.65$; HIGH if $p \geq 0.65$ (5)

The interface displays the tier, the calibrated score, the SVM margin distance and a physician recommendation, together with an explicit statement that the output is a model estimate and not a diagnosis.

G. Explainability and localization. For a global-average-pooled representation followed by a linear classifier, the gradient of the class score with respect to the k -th activation map is constant across spatial positions and proportional to the classifier weight. Grad-CAM weights [4] therefore reduce to the (scaled) SVM weights. The class-activation map is

$$M(i,j) = \text{ReLU}(\sum_k \tilde{w}_k \cdot A_k(i,j)), \quad \tilde{w}_k = w_k / s_k \quad (6)$$

The map is upsampled to 224×224 , normalized to $[0,1]$, colour-coded and blended with the original radiograph as a semi-transparent overlay with adjustable opacity. Activation mass is then compared between the patient-right and patient-left hemithoracic halves to label the finding as right, left or bilateral. When activation is diffuse or weak, localization is reported as unavailable. The overlay is an interpretive aid, not a segmentation or radiological finding.

H. Training and evaluation protocol. The dataset was partitioned 80/20 at the patient level so that no patient contributes images to both partitions. Feature extraction, scaling, SVM training and calibration used only the training portion, and metrics were computed on the untouched test partition. Performance was summarized using:

$$\text{Sens} = \text{TP} / (\text{TP} + \text{FN}), \quad \text{Spec} = \text{TN} / (\text{TN} + \text{FP}) \quad (7)$$

$$\text{Prec} = \text{TP} / (\text{TP} + \text{FP}), \quad \text{F1} = 2 \cdot \text{Prec} \cdot \text{Sens} / (\text{Prec} + \text{Sens}) \quad (8)$$

$$\text{BA} = (\text{Sens} + \text{Spec}) / 2 \quad (9)$$

Because of the extreme imbalance, we additionally report the Matthews correlation coefficient (MCC), likelihood ratios and prevalence-adjusted predictive values.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental setup. The system was implemented in Python 3.10+ using PyTorch and Torchvision [6] for DenseNet-121, scikit-learn [5] for scaling, LinearSVC and calibration, and OpenCV/Pillow for image handling. The backend uses Flask and PyMongo, the frontend uses React 18, Vite and Tailwind CSS, and PDF reports are generated with ReportLab. Development and testing were carried out on Windows 10/11 and Ubuntu 20.04 workstations (Intel Core i5/i7-class CPU, 8 to 16 GB RAM, with an optional CUDA-capable NVIDIA GPU for faster feature extraction).

B. Dataset. Experiments use the NIH ChestX-ray14 corpus [1], with pneumonia as the positive class and all other images (including other thoracic findings and "no finding") as the negative class. Table I summarizes the distribution. Train-partition counts are derived by subtracting the test counts from the totals.

TABLE I. DATASET DISTRIBUTION

Subset	Images	Pneumonia	Non-pneumonia
Full corpus (30,805 patients)	112,120	1,431 (1.28%)	110,689
Training ($\approx 80\%$)	89,336	1,133	88,203
Test ($\approx 20\%$)	22,784	298 (1.31%)	22,486

C. Quantitative evaluation. Fig. 2 and Table II show the confusion matrix and Table III the derived metrics on the held-out test set.

Actual	Pneumonia	TP = 254 (85.2% of row) FN = 44 (14.8% of row)
	Non-pneumonia	FP = 2,595 (11.5% of row) TN = 19,891 (88.5% of row)
		Pneumonia Non-pneumonia Predicted

Fig. 2. Confusion matrix on the 22,784-image test set (row-normalized percentages).

TABLE II. CONFUSION MATRIX ON THE HELD-OUT TEST SET

****	Predicted pneumonia	Predicted non-pneumonia
Actual pneumonia	TP = 254	FN = 44
Actual non-pneumonia	FP = 2,595	TN = 19,891

TABLE III. CLASSIFICATION PERFORMANCE

Metric	Value
Accuracy	88.42%
Sensitivity (recall)	85.23%
Specificity	88.46%
Precision (PPV)	8.92%
F1-score	16.14%
Balanced accuracy	86.85%
ROC-AUC	0.7087
PR-AUC	0.5857
MCC (derived)	0.253
Negative predictive value (derived)	99.78%
LR+ / LR- (derived)	7.39 / 0.167

The classifier correctly classified 20,145 of 22,784 test images. The normalized confusion matrix shows 88% of non-pneumonia and 85% of pneumonia images classified correctly. Only 44 of 298 pneumonia cases (14.8%) were missed, which is the clinically more costly error in a screening setting. The negative predictive value of 99.78% means a LOW/negative output is highly reassuring at this prevalence, and LR- = 0.167 indicates a substantial reduction in post-test odds of disease for a negative result.

D. Effect of prevalence on precision. The low precision (8.92%) is a direct consequence of the operating point and the 1.31% prevalence, not merely of the classifier. With specificity of 88.46%, roughly 11.5% of the 22,486 negatives (2,595 images) are flagged, which outnumbers the true positives by about ten to one. Under the assumption of constant sensitivity and specificity, Bayes' rule gives an expected PPV of approximately 28% at 5% prevalence and 45% at 10% prevalence. The system is therefore better suited to enriched clinical populations, such as symptomatic patients, than to unselected population screening. Accuracy alone (88.42%) is misleading here, since a trivial classifier that predicts every image as negative would score 98.7%, which is why sensitivity, balanced accuracy and MCC are the more informative indicators.

E. Comparison with existing work. The surveyed studies [9]–[17] were evaluated on different datasets, class balances and protocols, most on curated binary pneumonia versus normal sets. Direct numerical comparison would therefore be invalid, and we compare design capabilities instead (Table IV).

TABLE IV. CAPABILITY COMPARISON WITH RELATED CNN+SVM SYSTEMS

Work	Backbone / classifier	Risk tiers	Visual expl.	Reporting workflow
[13]	MobileNetV2 / SVM	NR	NR	NR
[15]	EfficientNet / SVM	NR	NR	NR
[16]	Pretrained CNN / SVM	NR	NR	NR
[17]	Fused CNN+PCA / SVM	NR	NR	NR
Ours	DenseNet-121 / calibrated Linear SVM	LOW / MED / HIGH	Grad-CAM-style, hemithorax	Secure web app, history, PDF

NR = not reported in the surveyed description.

The distinguishing contribution is therefore the integration of calibration, explanation and workflow around a

competitive feature-classifier core. Relative to CheXNet [2], which fine-tunes the entire network, our frozen-backbone design trades some potential accuracy for much cheaper retraining, since only the linear classifier and calibrator need refitting.

F. Qualitative results and system behaviour. Functional testing covered valid pneumonia and non-pneumonia uploads, rejection of unsupported or undersized files, authenticated PDF download, and denial of unauthorized or non-administrator access. All seven test cases passed. In representative outputs, an image with calibrated probability 63.4% was labelled MEDIUM risk with right-hemithorax localization, whereas images at 30.1% and 24.2% were labelled LOW risk with clear-field findings. The administrator dashboard aggregates risk-tier counts, per-patient histories and PDF access.

G. Efficiency and real-time viability. Because the backbone is frozen and the classifier is linear, inference consists of one DenseNet-121 forward pass, one matrix-vector product and a sigmoid, which is suitable for commodity CPUs with an optional GPU. Explanation reuses the feature maps already computed and adds negligible cost. Formal latency and throughput benchmarking has not yet been performed and is planned.

H. Limitations and edge cases. (i) Precision is low under the extreme imbalance (Section IV-D), and the operating threshold was not optimized for a target precision. (ii) ChestX-ray14 labels are text-mined from radiology reports and contain noise, so reported performance is bounded by label quality. (iii) No external validation was conducted, and performance on other hospitals, scanners and pediatric populations is unknown. (iv) The hemithorax localization has not been quantitatively validated against expert annotations. (v) The reported ROC-AUC (0.7087) is considerably lower than the operating-point sensitivity and specificity would normally imply, and warrants re-examination of how the curve was computed. (vi) The system is a decision-support aid and not a diagnostic device.

V. CONCLUSION AND FUTURE SCOPE

This paper presented PulmoScan AI, an explainable pneumonia risk-assessment system that couples a frozen DenseNet-121 feature extractor with a class-weighted, Platt-calibrated Linear SVM. The calibrated posterior is mapped to LOW, MEDIUM and HIGH tiers and accompanied by a Grad-CAM-style hemithorax localization, all delivered through a secure web application with persistent storage and automated PDF reporting. On a patient-level held-out test set of 22,784 ChestX-ray14 images, the system reached 88.42% accuracy, 85.23% sensitivity, 88.46% specificity and 86.85% balanced accuracy. A negative predictive value of 99.78% supports its use as a rule-out aid. The 8.92% precision confirms that the system's positive alerts require clinician confirmation, particularly in low-prevalence populations.

Future work will proceed in five directions. First, external validation on multi-centre and pediatric datasets, with prospective clinical evaluation, is required before any clinical use. Second, imbalance handling can be improved through threshold optimization against a target precision, cost-sensitive learning, hard-negative mining and fine-tuning of

the backbone, and the RF and KNN comparators considered in our design should be benchmarked quantitatively. Third, the explainability module should be validated against radiologist annotations and extended with complementary methods. Fourth, systematic latency, throughput and calibration-quality analysis (reliability diagrams, expected calibration error) should be added. Fifth, deployment can be extended to cloud and mobile platforms, with multilingual reporting and integration with hospital information systems (HL7/FHIR/DICOM).

REFERENCES

- [1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2097–2106.
- [2] P. Rajpurkar *et al.*, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.
- [3] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE CVPR*, 2017, pp. 4700–4708.
- [4] R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.
- [5] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [6] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2019.
- [7] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers*. Cambridge, MA, USA: MIT Press, 1999, pp. 61–74.
- [8] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE CVPR*, 2016, pp. 2921–2929.
- [9] S. Akshay *et al.*, "Hybrid deep learning and machine learning framework for automated pneumonia detection in chest X-ray images," 2025. [Add venue/DOI]
- [10] M. F. Hashmi *et al.*, "Hybrid deep learning and handcrafted feature fusion for pneumonia detection in chest X-rays," 2025. [Add venue/DOI]
- [11] G. S. Araújo *et al.*, "Hybrid approaches for pneumonia detection in X-rays: Combining CNNs and ML classifiers," 2025. [Add venue/DOI]
- [12] K. Vijayakumar *et al.*, "Classification of chest radiographs into healthy/pneumonia using Harris-Hawks algorithm optimized deep-features," 2025. [Add venue/DOI]
- [13] E. P. A. Meindiawan and Muljono, "Application of MobileNetV2 and SVM combination for enhanced accuracy in pneumonia classification," 2024. [Add venue/DOI]
- [14] Z. Zhang *et al.*, "Multi-scale CNN with SVM for pneumonia detection in chest X-ray images," 2025. [Add venue/DOI]
- [15] R. Sharma *et al.*, "EfficientNet-based feature extraction with SVM for pneumonia classification," 2025. [Add venue/DOI]
- [16] P. Kumar *et al.*, "Chest X-ray pneumonia detection using transfer learning and SVM," 2026. [Add venue/DOI]
- [17] L. Chen *et al.*, "Deep feature fusion and SVM for accurate pneumonia detection from chest X-rays," 2026. [Add venue/DOI]