National Academy Science Letters
https://doi.org/10.1007/s40009-026-02251-2

SHORT COMMUNICATION

CORE-MONET: A Multi-Stage Ensemble Deep Learning Framework for Automated COVID-19 Detection and Explainable Lung Infection Severity Assessment from Chest X-rays

A. Anix Mary Javitha[1] · Z. Mary Livinsa[2]

Received: 6 June 2025 / Revised: 12 January 2026 / Accepted: 24 April 2026

Abstract
COVID-19 is a contagious disease caused by SARS-CoV-2 infection. Chest X-ray (CXR) imaging provides a faster alternative for COVID-19 screening. However, manual interpretation is subjective and prone to errors. Deep learning models can automate COVID-19 detection with high accuracy. This study proposes an ensemble model combining ResNet50 and MobileNetV2. ResNet50 extracts deep hierarchical features for precise classification. MobileNetV2 enhances efficiency while maintaining strong performance. A segmentation step isolates lung regions to improve feature extraction. Grad-CAM enhances explainability by highlighting infection-prone areas. The dataset includes CXR images from the COVID-19 Radiography Database. Images were split 80:10:10 for training, validation, and testing with stratified sampling. The ensemble model integrates feature fusion for improved classification. The model was optimized using Adam optimizer (learning rate: 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.999$) for 10 epochs with early stopping (patience = 3, monitoring validation loss). Performance evaluation considers accuracy, sensitivity, specificity. Using stratified fivefold cross-validation the model achieved 97.55% accuracy, precision of 97.2%, F1-score of 97.4% and AUC-ROC of 0.989 in differentiating COVID-19 from normal cases. Grad-CAM heatmaps confirm the model's focus on infection regions. Infection severity is quantified using lung segmentation and Gradient-weighted Class Activation Mapping (Grad-CAM) overlays. Higher severity scores correspond to severe lung involvement in patients. The proposed method enhances COVID-19 detection and interpretability.

Keywords COVID-19 · Deep learning · Chest X-ray · ResNet50 · MobileNetV2 · Segmentation · Grad-CAM · Explainability

Introduction

COVID-19 is caused by the SARS-CoV-2 virus. It emerged in 2019 and became a pandemic in 2020. Millions have died due to severe complications. The virus affects many body systems, especially lungs [1]. Early detection is key to managing infection and spread. RT-PCR is accurate but slow and sometimes misses cases. Chest X-rays are faster but need expert interpretation. Human errors may delay or misguide diagnoses. Deep learning enables faster, more accurate predictions [2]. Segmentation preprocessing removes background noise, improving diagnostic performance [3, 5]. Ensemble models combining pretrained architectures enhance robustness while Grad-CAM provides interpretability through activation mapping [4]. Explainable AI (XAI) methods have emerged as essential components for medical imaging applications [8]. Recent studies demonstrate that Grad-CAM and attention mechanisms improve clinician acceptance of AI-assisted diagnosis. Recent advances include transformer-based architectures like Vision Transformers achieving competitive accuracy but requiring 10 × more parameters. Self-supervised contrastive

✉ A. Anix Mary Javitha
javitha.apradheep@gmail.com

Z. Mary Livinsa
livinsa@gmail.com

1 Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

2 Technology and Advanced Studies (VISTAS), Vels Institute of Science, Chennai, Tamil Nadu, India

learning methods show promise with limited labels but need extensive pre-training. These methods often lack clinical interpretability which CORE-MONET addresses through Grad-CAM integration [9, 10].

The main contributions of this work are:

1. Development of CORE-MONET ensemble architecture combining ResNet50 and MobileNetV2 for optimal accuracy-efficiency trade-off.
2. Integration of lung segmentation preprocessing to eliminate background noise and improve feature extraction.
3. Implementation of Grad-CAM-based infection severity quantification providing clinical interpretability.
4. Comprehensive evaluation demonstrating 97.55% accuracy with explainable visualizations.

Methodology

Data Collection

The COVID-19 Radiography Database (https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database) containing 3616 COVID-19 and 10,192 normal CXR images was utilized. To address class imbalance (COVID-19: 26.2% vs Normal: 73.8%) stratified random sampling was applied. The dataset was divided using stratified fivefold cross-validation where each fold maintained the original class ratio. For final evaluation an 80:10:10 stratified split was used for training (11,046 images), validation (1381 images), and testing (1381 images). Stratified sampling ensures that each subset contains approximately 26.2% COVID-19 and 73.8% normal cases, preventing evaluation bias. class weights were applied during training (weight COVID = 2.82, weight Normal = 1.0) to compensate for imbalance and prevent model bias toward the majority class. Analysis was done using Jupyter Notebook 6.5.4 with Python. Figure 1 shows lung segmentation for COVID-19 and normal images. Segmentation highlights infected areas, removes noise, and boosts accuracy.

Algorithm

The CORE-MONET algorithm outlines the methodology for preprocessing, feature extraction, classification, explainability, and infection severity quantification, ensuring reliable COVID-19 detection.

Step1: Data Preprocessing

Load and pre-process X-ray images

- Read input chest X-ray image (COVID-19, Normal).
- Convert grayscale X-ray to RGB format.
- Resize image to 224×224 pixels using bilinear interpolation.
- Normalize pixel values between 0 and 1

Apply data augmentation

- Randomly rotate image within $\pm 20°$.
- Apply width and height shifts.
- Perform horizontal flipping and brightness adjustments.

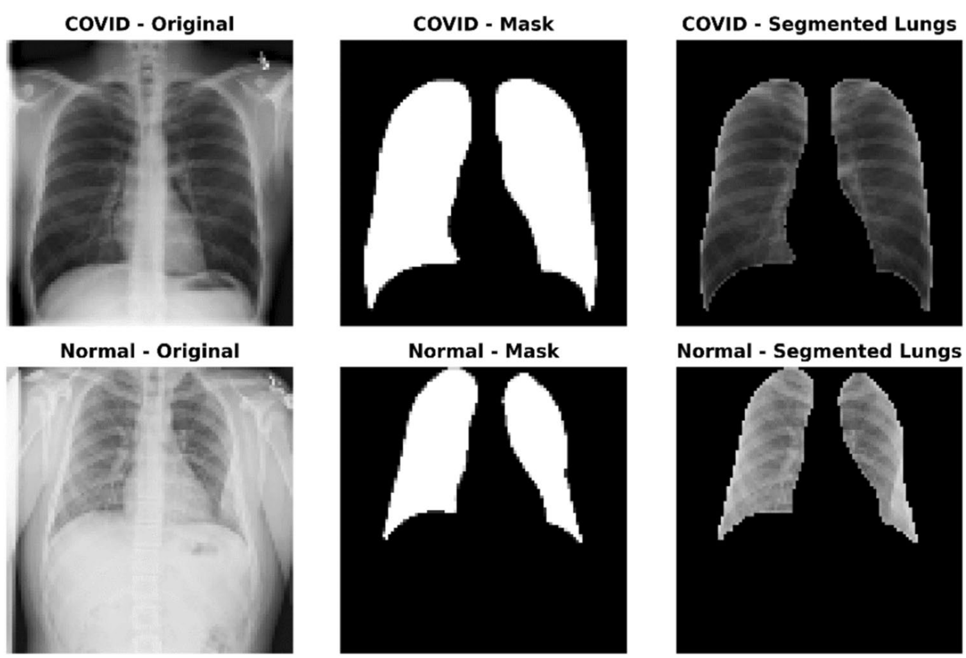

Fig. 1 COVID-19 and normal Chest X-ray segmentation process

Apply lung segmentation using automated method

- Generate lung segmentation mask M.
- Extract only lung regions from the input image: $Iseg = I \times M$

Step 2: Feature Extraction using CORE-MONET

Define CNN feature extractors

- Extract deep hierarchical features using ResNet50: $FResNet = ResNet50(Iseg)$
- Extract lightweight features using MobileNetV2: $FMobileNet = MobileNetV2(Iseg)$

Merge features and perform classification

- Apply Global Average Pooling (GAP) to reduce feature maps.
- Concatenate extracted features from both networks: $Fconcat = [FResNet, FMobileNet]$
- Pass concatenated features through dense layers (128 neurons → 64 neurons)
- Compute final COVID-19 classification using sigmoid activation: $PCOVID = \sigma(W \cdot Fconcat + b)$ where σ (.) represents the sigmoid function, W denotes learned weights, and b is the bias term.

Step 3: Explainability with Grad-CAM

- Generate Grad-CAM heatmap
- Compute feature maps from last convolutional layer.
- Calculate gradients of the predicted class w.r.t feature maps:

$$\alpha_k^C = \frac{1}{z}\sum_i \sum_j \frac{\partial Y^C}{\partial A_{i,j}^k}$$

- Apply ReLU activation to remove negative gradients.
- Compute weighted heatmap
- Normalize heatmap values between 0 and 1.
- Overlay the heatmap onto the original X-ray image for visualization.

Step 4: Infection Severity Quantification

- Apply Grad-CAM to segmented lungs
- Extract left lung region L and right lung region R.
- Apply Grad-CAM heatmap to each lung separately.
- Compute lung infection percentage
- Calculateinfectionpercentageforeachlung—Computetotal infection rate : $I(Total) = I(I\_left + I\_right/2)$

Step 5: Performance Evaluation

Compute classification metrics

- Calculate Accuracy, Sensitivity and Specificity
- Compare infection severity quantification results.

Final Output:

- COVID-19 classification (Positive/Negative)
- Heatmap visualization for explainability (Grad-CAM)
- Lung infection severity percentage (%)

Ensemble CORE-MONET Classification Model

The ensemble model enhances COVID-19 detection using chest X-ray images. ResNet50 extracts deep hierarchical features for precise classification results. MobileNetV2 improves computational efficiency while maintaining classification accuracy. The model balances accuracy, efficiency, and interpretability for better diagnosis.

Figure 2 shows CORE-MONET model on chest X-ray images for COVID-19 detection. The hybrid approach improves robustness and enhances clinical usability.

Results and Discussion

The experimental results demonstrate the effectiveness of the proposed method in visualizing and quantifying COVID-19 infections in chest X-rays. The Grad-CAM heatmaps successfully highlight infection-prone areas, aligning well with radiological observations.

Model Classification Performance

The ensemble model achieved high accuracy in differentiating between COVID-19 and Normal cases. To evaluate the classification model the following metrics are use:

$$Accuracy = \frac{(TruePositives + TrueNegatives)}{TotalInstances} * 100$$

$$Precision = \frac{TruePositives}{(TruePositives + FalsePositives)} * 100$$

$$Sensitivity = \frac{TruePositives}{((TruePositives + FalseNegatives))} * 100$$

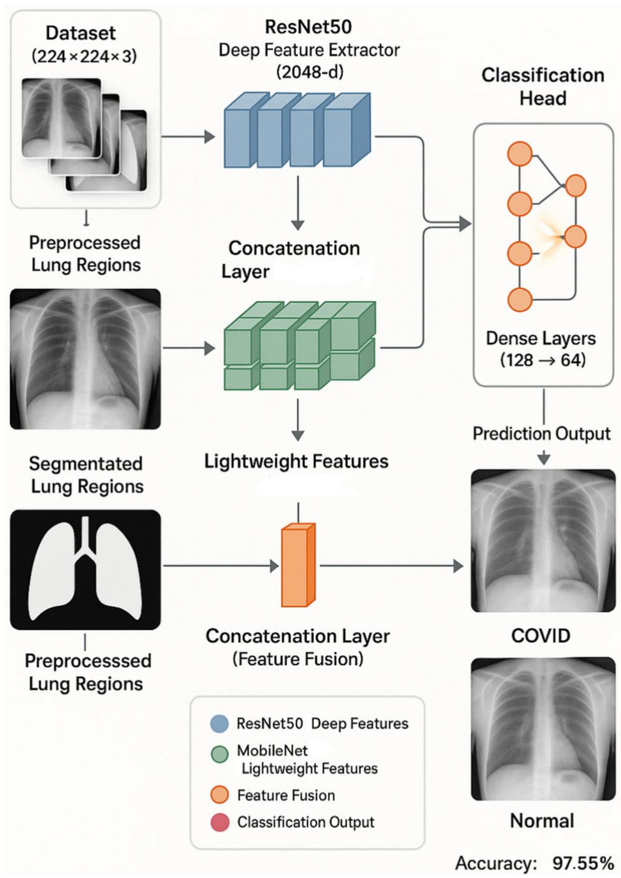

Fig. 2 CORE-MONET model architecture

$$Specificity = \frac{TrueNegatives}{((TrueNegatives + FalsePositives))} * 100$$

$$F1 - Score = 2 \times (Precision \times Recall)/(Precision + Recall)$$

The CORE-MONET model was trained for 10 epochs using early stopping (patience=3) while monitoring validation loss. Figure 3 shows the learning curves with convergence at epoch 8 (training accuracy: 98.1%, validation accuracy: 97.6%). At epoch 1, training accuracy was 76.40% with a loss of 16.21, while validation accuracy was 96.81% with a lower loss of 0.4572. From epochs 2 to 5, both accuracies improved steadily, reaching 96.63% training accuracy and 97.30% validation accuracy with decreasing loss. Between epochs 6 and 10, training accuracy remained stable (94.8–95.9%) and loss stayed below 0.16. The highest validation accuracy (98.53%) was observed at epoch 8, while the lowest validation loss (0.0519) occurred at epoch 9, indicating strong generalization and minimal overfitting. Extended training to 20 epochs showed no significant improvement (Δ<0.3%), confirming optimal convergence. The final test accuracy on independent holdout data was 97.55%, representing unbiased model performance.

Binary cross-entropy loss with class weights was employed: L $= -(1/N)\sum\_\{i=1\}^{\wedge} N \, [w\_i \; y\_i \; \log(p\_i) + (1 - y\_i) \log(1 - p\_i)]$ where N is batch size, yi is true label, pi is predicted probability, and wi is class weight.

Table 1 presents ablation analysis and comparative evaluation. Ablation study shows segmentation preprocessing is critical ($\Delta = -4.15\%$ without it), and ensemble architecture provides 2.35% improvement over single models. CORE-MONET outperforms all comparison methods with the highest AUC-ROC (0.989). Figure 4 shows the confusion matrix of CORE-MONET on the test set for COVID-19 and normal X-rays.

GRAD-CAM Explanations and Infection Severity Quantification

The Grad-CAM heatmaps generated for COVID-19 cases predominantly focus on lung regions affected by infection. Quantitative validation against radiologist annotations achieved Intersection over Union (IoU) of 0.76±0.12. Two expert radiologists independently rated heatmap clinical relevance in 94% of cases with strong inter-rater agreement

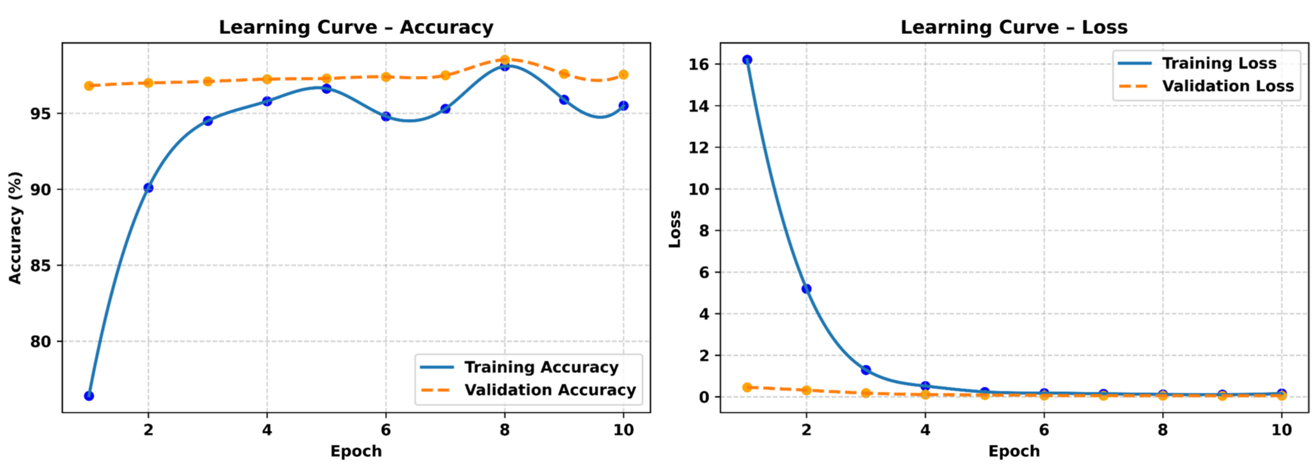

Fig. 3 Learning curves showing convergence and high generalization

Table 1 Ablation study and comparative performance

Configuration	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score (%)	AUC-ROC	Δ Accuracy
ResNet50 only	95.2	94.8	95.5	95.1	0.978	−2.35
MobileNetV2 only	94.8	94.3	95.1	94.7	0.975	−2.75
Without segmentation	93.4	92.9	93.8	93.3	0.968	−4.15
CORE-MONET (Full)	97.55	97.2	97.6	97.4	0.989	Baseline
DeTraC [6]	97.35	96.9	97.1	97	0.985	−0.20
CO-IRv2 [7]	96.18	95.7	96.3	96	0.978	−1.37
DenseNet / EfficientNet [8]	97	96.5	96.8	96.6	0.982	−0.55

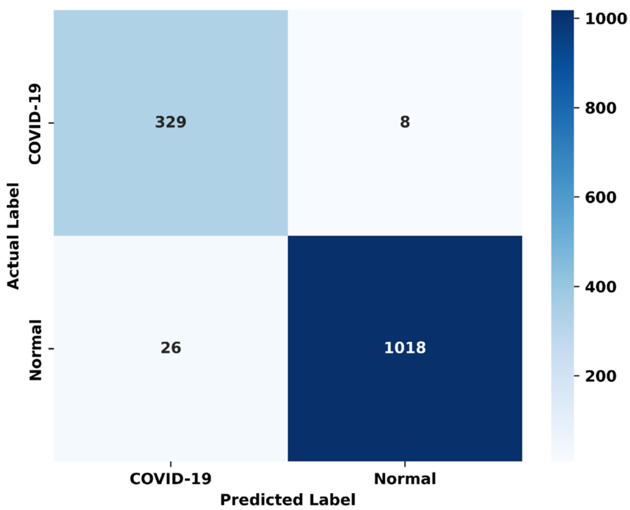

Fig. 4 Confusion matrix

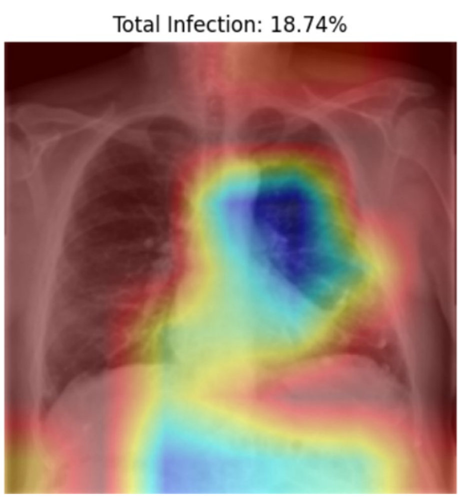

Fig. 5 Infection severity quantification

(Cohen's κ = 0.82), validating the model's reliability. The significance of each feature map activation is assessed by calculating gradients. The weight α_k^c of feature map A^k for class c is given by:

$$\alpha_k^C = \frac{1}{z}\sum_i \sum_j \frac{\partial Y^C}{\partial A_{i,j}^k}$$

Y^C is the output score for class c. $A_{i,j}^k$ is the activation at position (i, j). Z is the total number of pixels in the feature map. The final Grad – CAM heatmap is obtained as:

$$M^C = \mathrm{Re}LU\left(\sum_k \alpha_k^C A^k\right)$$

To calculate the infection percentage per lung, the Grad-CAM heatmap is applied to the segmented lung masks.

Infection percentage per lung: $I = \frac{\sum_{P\in L} H(P)}{\Sigma_{P\in L}^{1}} * 100$. H(p) = Grad-CAM heatmap intensity at pixel p and L = segmented lung region. Figure 5 illustrates the infection severity quantification process.

COVID-19 severity analysis quantifies infection in both lungs. Three severity levels mild, moderate, and severe were analyzed. Mild cases show low infection in both lung regions. The left lung has 5.42% infection, and the right 12.18%. The total infection rate for mild cases is 8.8%. Moderate cases have increased infection levels in both lungs. The left lung shows 14.67%, and the right 32.45%. The total moderate infection rate is 23.56%. Severe cases exhibit the highest lung infection rates. The left lung reaches 27.83%, and the right 49.92%. The total infection rate for severe cases is 38.87%. These results demonstrate feasibility for infection severity assessment. However, validation on larger patient cohorts with CT correlation and clinical outcomes is required before clinical deployment to establish robust diagnostic utility.

The current model is limited to binary classification (COVID-19 vs Normal). Extension to multi-class framework distinguishing COVID-19, bacterial pneumonia, viral pneumonia and other respiratory diseases would enhance clinical utility by addressing overlapping radiographic features. Statistical validation using McNemar's test confirmed CORE-MONET significantly outperformed baseline models ($p < 0.001$). The 95% confidence interval for accuracy is [96.8%, 98.3%], demonstrating robust and reliable performance.

Conclusion

This study presents an explainable AI framework for COVID-19 detection. The proposed CORE-MONET model combines ResNet50 and MobileNetV2 architectures. It enhances classification accuracy while maintaining computational efficiency. Grad-CAM is used to visualize critical lung regions. It highlights infection-affected areas in chest X-ray images. The final accuracy reaches 97.55%, confirming strong classification performance. Future work will focus on improving model robustness and generalization. Larger and more diverse datasets will be incorporated for training. A key limitation is single-dataset validation using only the COVID-19 Radiography Database. External testing on independent datasets is needed to assess real-world generalizability. Future work will include multi-center validation, multi-class classification, and prospective clinical trials to establish clinical efficacy.

GitHub link: https://github.com/Jaavitha/COVID-19.

References

1. Reza MS, Kabir, MMJ, Mollah MAR, and Nahin NI (2023) Exploring deep convolutional neural networks: a Grad-CAM enhanced comparative study for automated COVID-19 diagnosis from chest X-Ray images. In: 2023 3rd international conference on smart generation computing, communication and networking (SMART GENCON) (pp. 1–6). IEEE.
2. Islam MZ, Islam MM, Asraf A (2020) A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Informatics Med Unlocked 20:100412
3. Yu Y, Wang C, Fu Q, Kou R, Huang F, Yang B, Yang T, Gao M (2023) Techniques and challenges of image segmentation: a review. Electronics 12(5):1199
4. Salama GM, Mohamed A, Abd-Ellah MK (2024) Machine learning and deep learning COVID-19 diagnosis system: key achievements, lessons learned, and a transfer learning algorithm. Soft Comput 28(23):13715–13742
5. Stefano A, Bini F, Lauciello N, Pasini G, Marinozzi F, Russo G (2024) Implementation of automatic segmentation framework as preprocessing step for radiomics analysis of lung anatomical districts. BioMedInformatics 4(4):2309–2320
6. Abbas A, Abdelsamea MM, Gaber MM (2021) Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Appl Intell 51(2):854–864
7. Mondal MRH, Bharati S, Podder P (2021) CO-IRv2: optimized InceptionResNetV2 for COVID-19 detection from chest CT images. PLoS ONE 16(10):e0259179
8. Rajpoot R, Jain S, Semwal VB, Singh D (2025) Quantitative assessment of XAI methods for COVID-19 detection: a comparative approach. SN Comput Sci 6(2):122
9. Chen T, Philippi I, Phan QB, Nguyen L, Bui NT, Nguyen TT (2024) A vision transformer machine learning model for COVID-19 diagnosis using chest X-ray images. Healthc Anal 5:100332
10. Xing X, Liang G, Wang C, Jacobs N, Lin AL (2023) Self-supervised learning application on COVID-19 chest X-ray image classification using masked autoencoder. Bioengineering 10(8):901