A Dual-Modality Deep Learning Framework for COVID-19 Detection with Interpretability Using Chest X-Ray and CT Imaging

R. Arvind
Department of Computer Science and Engineering, Gopalan College of Engineering and Management, Bengaluru, India | Visvesvaraya Technological University, Belagavi, Karnataka, India
aruramesh27@gmail.com (corresponding author)

Manoj Challa
Department of Computer Science and Engineering, Gopalan College of Engineering and Management, Bengaluru, India | Visvesvaraya Technological University, Belagavi, Karnataka, India
manojreddi@gmail.com

Received: 27 October 2025 | Revised: 8 November 2025, 16 November 2025, and 12 December 2025 | Accepted: 13 December 2025

 DOI: https://doi.org/10.48084/etasr.15824

ABSTRACT

The present study presents a comprehensive deep learning framework for automatically diagnosing COVID-19 from chest X-rays and CT scans. A balanced, multi-source dataset comprising three groups, which include normal cases, COVID-19 cases, and pneumonia cases, is produced by deploying the proposed method. It compares an optimized AlexNet with a customized Convolutional Neural Network (CNN) using transfer learning. The methodology involves combining patient-level data, adding more data, and conducting a detailed analysis that includes statistical significance testing and 5-fold cross-validation. An independent dataset (BIMCV-COVID19+) is used for external validation to assess generalizability. Gradient-weighted Class Activation Mapping (Grad-CAM) heatmaps address concerns about ease of understanding by showing the regions used for diagnosis. The study also tests the effectiveness of the EfficientNet and Vision Transformer (ViT) models. With an accuracy of 99.1%, the ViT outperforms earlier models in clinical screening scenarios with constrained processing resources. The current study highlights the benefits of using this AI-assisted method in actual radiology operations and emphasizes the significance of maintaining thorough records of the datasets.

Keywords-COVID19; deep learning; transfer learning; AlexNet; CNN; vision transformers; EfficientNet; X-ray; CT-scan image; Grad-CAM; medical imaging

I. INTRODUCTION

Although COVID-19 is no longer classified as a global emergency, it remains a clinically significant respiratory illness requiring ongoing surveillance and diagnostic support, especially in regions with limited healthcare resources or during seasonal resurgences. The most common diagnostic technique, Reverse Transcription Polymerase Chain Reaction (RT-PCR), has several shortcomings, including high false-negative rates. In response, chest imaging modalities, particularly CT scans and chest X-rays, have become essential for supporting prompt diagnosis, severity evaluation, and patient triage, especially in situations where RT-PCR is unavailable or delayed [1].

Automating COVID-19 detection from chest images has shown great promise with the use of deep learning, and in particular, Convolutional Neural Networks (CNNs). CNN-based models are capable of effectively identifying pathological patterns in pulmonary tissues and learning hierarchical features [2-4]. Despite this, a number of early methods have serious problems, including a lack of multimodal fusion (i.e., CT and X-ray), class imbalance, overfitting on small datasets, and insufficient external validation [5].

CNN-based architectures have shown excellent classification performance [6-8] although their clinical utility has been limited due to their reliability on unbalanced datasets or a lack of multi-class differentiation. More sophisticated architectures that offer better global feature modeling and increased accuracy have been introduced, such as EfficientNet [9] and ViTs [10, 11]. However, these models usually need more computational power and large training datasets, which makes them less suitable for deployment in environments with limited resources.

The present study provides a useful and efficient solution in this regard: a transfer-learned AlexNet model with high diagnostic accuracy (98.9%), explainable outputs, and

validation across imaging modalities and unseen data. By offering a strong substitute that strikes a balance between precision, interpretability, and computational efficiency, this work enhances transformer-based methodologies.

To accurately diagnose COVID-19, the present study presents a dual-modality deep learning framework that uses both chest X-ray and CT images. The framework compares two models: a refined AlexNet architecture modified using transfer learning and a custom-built CNN. Despite being an older architecture, AlexNet is well-suited for clinical deployment in the real world, particularly in low-resource environments, due to its low computational requirements and robust performance in low-data regimes.

A well-balanced dataset was created comprising more than 7,000 images (normal, COVID-19, and pneumonia) from four public repositories, with distinct patient-level partitioning to stop data leaks. Cross-validation was used to compare a customized CNN and an improved AlexNet, and performance claims were validated using statistical significance testing (paired t-test, $p < 0.05$).

To test generalizability outside of the training domain, external validation was conducted using the BIMCV-COVID19+ dataset. The study also conducted benchmarking against cutting-edge models, such as EfficientNet and ViTs to put the findings in the context of the present state of deep learning research in medical imaging. Grad-CAM visualizations were performed to improve clinical transparency and model interpretability.

A. Literature Review

Developments in AI-powered COVID-19 detection through chest imaging have demonstrated how quickly advanced hybrid and transformer-based frameworks have replaced traditional CNNs. The goals of these initiatives are to increase clinical interpretability, generalization across modalities (CT and X-ray), and classification accuracy.

CNNs were the main focus of earlier studies [1, 2] for the binary or multi-class COVID-19 detection from chest X-rays, with accuracies ranging from 90 to 94%. Small datasets, overfitting, and limited generalizability to unseen CT scans were still challenging for the CNN-based methods. Transfer learning using ResNet101 and Xception, for example, improved performance [3], but it was still limited by inadequate external validation. Similarly, class imbalance and modality limitations plagued the CovCNN model [4], which achieved 98.4% accuracy.

Authors in [7] benchmarked ViT, DenseNet, and ResNet50 on identical datasets and found that ViT slightly outperformed CNNs (98.1% versus 97.4%), but it required more computational overhead. Authors in [5, 6] also confirmed ViT's ability for multi-center generalization. In a pneumonia–COVID multi-class environment, Swin Transformers also performed better than classical CNNs [8]. EfficientNet, used in [9], was later modified for COVID-19 diagnosis [10], and because of its compound scaling and efficiency, it is still a promising option for embedded systems. Authors in [11, 12] compared EfficientNet to MobileNet, AlexNet, and ViT models, confirming that it has competitive accuracy (≈98.5%) at a lower compute cost, making it perfect for real-time clinical screening.

Hybrid approaches that combine CNN backbones with attention mechanisms are commonly used. For example, ViT was combined with YOLOv8 for explainable medical image detection, while authors in [13] proposed a NASNet-ViT hybrid with ensemble fusion. Similarly, CNN parameters were tuned using particle swarm intelligence to enhance chest X-ray performance [14]. However, a lack of standardized datasets and external test validation makes many of these approaches difficult to scale and reproduce. The lack of cross-modality validation in these works is a significant drawback, as most models are only made for CT or X-ray images, not both. Furthermore, authors in [17] highlighted the importance of rapid and accurate COVID-19 diagnosis through automated detection from chest X-ray images; however, their review lacks interpretability and generalization, which are key limitations in recent studies.

Few studies incorporate explainability into the evaluation pipeline holistically, despite the occasional application of rad-CAM or SHAP. Using stratified patient-level splitting to prevent data leakage, the proposed framework fills these gaps by:

- Integrating both X-ray and CT images across COVID-19, pneumonia, and normal classes.
- Benchmarking models such as custom CNN, fine-tuned AlexNet, EfficientNet-B0, and ViT-B/16, under identical data and augmentation settings.
- Using Grad-CAM visualizations to improve model interpretability for clinical use.
- Validating with 5-fold cross-validation and statistical testing for robustness.

Although hybrid CNN-transformer models have shown promise, they frequently lack reproducibility and clarity on dataset handling [15, 16]. The proposed explainable design has a consistent evaluation pipeline, with a dual-modality balanced dataset, and it also provides a more clinically feasible and reproducible option. Additionally, by including ViT and EfficientNet, the present study guarantees that the comparative benchmark represents the state-of-the-art.

II. PROPOSED METHODOLOGY

A. Data Collection and Pre-Processing

The COVID-19 chest X-ray images were subset from the COVIDx8B dataset [1], which contains 8,066 COVID-positive images. The present study utilizes 1,500 images from this dataset, selecting those with complete metadata (including patient ID, gender, and age) and high-quality, frontal (AP/PA) projections. Images marked with uncertainty labels or overlapping pathologies were excluded to maintain diagnostic purity. Similarly, 1,500 normal and 1,500 pneumonia images were selected from the ChestX-ray14 dataset [19], prioritizing cases with definitive single-label annotations and standard projection views. CT images (n = 1,500 COVID, n = 1,500 normal/pneumonia) were extracted from the BIMCV-

COVID19+ and COVID-CT datasets [18, 20], again ensuring patient-level clarity and modality consistency. The resulting dataset was carefully balanced with 3,000 images per class, split using patient-level stratified 5-fold cross-validation, and made publicly available with class labels and dataset IDs [21].

B. Model Architecture

The present study evaluates the following four architectures:

- Custom CNN: A 7-layer CNN built architecture from scratch with ReLU activations, batch normalization, and dropout (rate: 0.4).
- AlexNet (fine-tuned): Pre-trained on ImageNet and re-trained on a newly created dataset [21]. The last three fully connected layers were modified for 3-class classification.
- EfficientNet-B0: A lightweight yet powerful architecture used for comparative benchmarking.
- ViT-B/16: ViT pre-trained on ImageNet21k and fine-tuned with frozen lower layers.

1) CNN Baseline Architecture

The proposed model is based on CNNs, which use hierarchical feature learning via convolutional layers. Figure 2 displays the CNN-based architecture.

For an input image $I \epsilon R^{224\times224\times3}$the CNN performs the following operations:

The convolutional layer is defined in:

$$Z_{ij}^{(k)} = (I * W^k)_{ij} + b^k \quad (1)$$

where W^k and b^k are the weights and biases of the k-th filter.

Activation (ReLu) is defined as:

$$A_{ij}^k = \max(0, Z_{ij}^k) \quad (2)$$

Max Pooling is done using:

$$P_{ij}^k = \max_{(m,n)\epsilon pool} A_{(i+m)(j+n)}^k \quad (3)$$

Fully connected layers and Softmax output are given by:

$$\hat{y} = softmax(W_{fc} * flatten(P) + b_{fc} \quad (4)$$

The loss function is defined in:

$$\mathcal{L} = -\sum_{i=1}^{c} y_i log\hat{y}_i \quad (5)$$

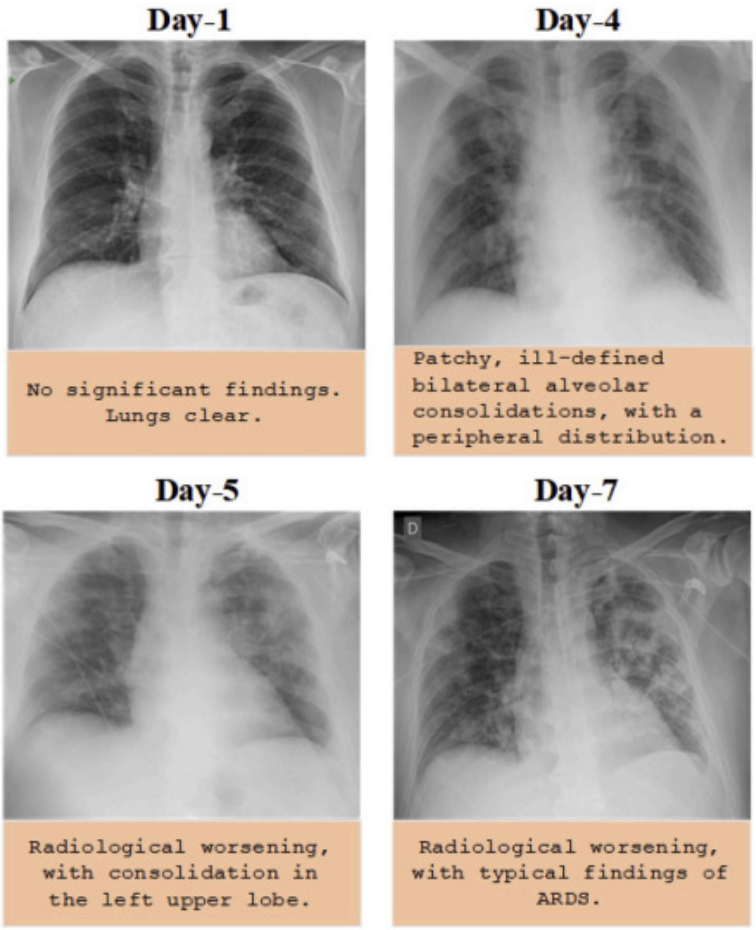

Fig. 1. Chest X-ray frames of elder COVID-19 patients with symptoms.

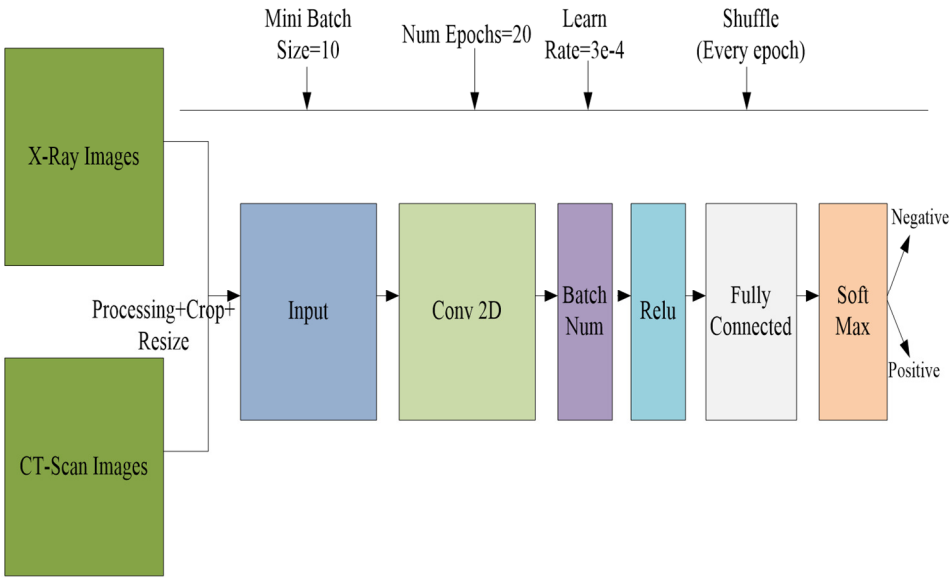

Fig. 2. Custom CNN-based architecture.

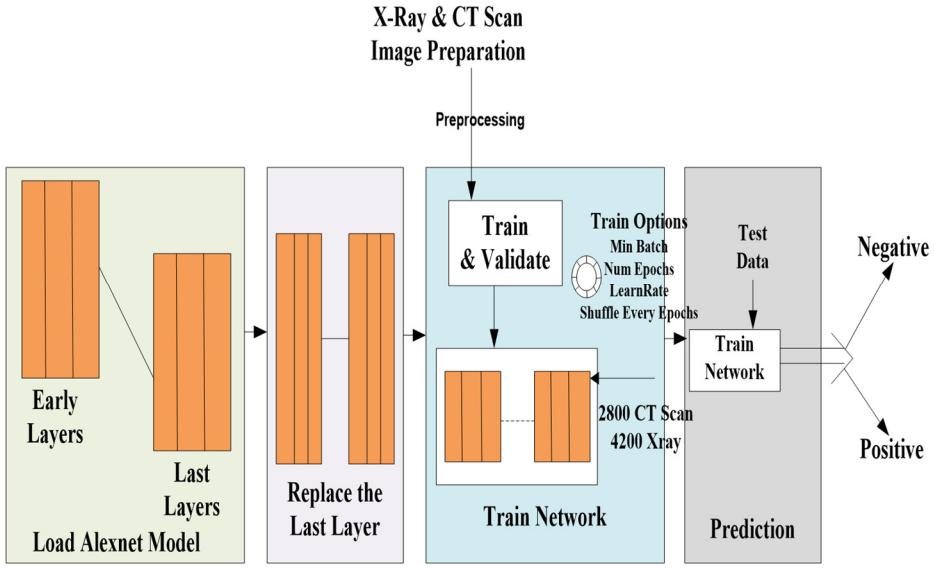

Fig. 3. Fine-tuned Alex-Net network.

2) Fine-Tuned AlexNet Model

To classify COVID-19 instances, an existing AlexNet model was modified from its original training on the ImageNet dataset. To implement transfer learning, specialized classification layers were used in place of the last completely linked ones. As observed in Figure 3, the model was fine-tuned to improve its diagnostic capabilities using a prepared dataset.

Some important changes made to the AlexNet model include:

- Adding a new fully connected layer to accommodate three-class categorization instead of the previous one.
- Preventing overfitting by modifying the dropout rates.
- Establishing a constant training rate of 3×10^{-4} by optimizing the learning process.

3) EfficientNet-B0

To balance network depth, width, and resolution, the CNN architecture EfficientNet-B0 was optimized using compound scaling. This produces a highly accurate model that utilizes significantly fewer parameters than conventional CNNs. EfficientNet employs squeeze-and-excitation optimized mobile inverted bottleneck (MBConv) blocks, as illustrated in Figure 4.

4) Vision Transformer (ViT-B/16)

ViT-B/16 is a transformer-based image classification model, and its architecture is portrayed in Figure 5. It splits the

image into 16×16 patches, embeds them into linear tokens, and then passes them through a number of self-attention layers, in contrast to CNNs that use convolutional filters.

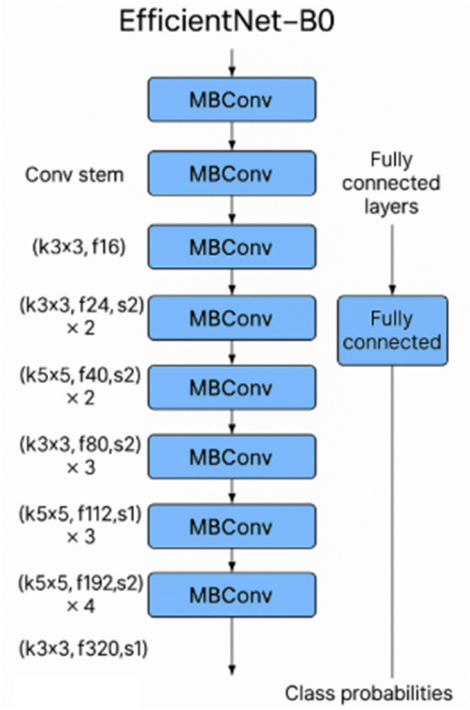

Fig. 4. EfficientNet-B0 architecture overview.

The features of the ViT-B/16 architecture are:

- Transformer encoder with 12 layers, 768 hidden units, and 12 heads; input: 224×224 images divided into 16×16 patches.
- Adapted to a 3-class COVID dataset after being pre-trained on ImageNet-21k.
- Particularly helpful in identifying diffuse patterns in lung CT/X-ray images due to the capturing of global dependencies across image regions more effectively than CNNs.

C. Training and Optimization

The models were trained using the Adam optimizer with a batch size of 10 and an epoch count of 20. The dataset was randomized at each epoch to ensure that the model could generalize and prevent bias. To monitor overfitting, the validation frequency was set at three epochs.

D. Performance Evaluation

The evaluation criteria for the Customized CNN, Fine-Tuned AlexNet, EfficientNet-B0, and ViT-B/16 models were:

- Cross-validation: 5-fold stratified cross-validation was performed with patient-level grouping.
- External Validation: The BIMCV-COVID19+ dataset was reserved for external testing, containing 1,000 CT and 800 X-ray images from independent institutions.
- Metrics: Accuracy, Precision, Recall, F1-score, and AUC were reported.
- Statistical Significance: A paired t-test was used to compare model performances over each fold.

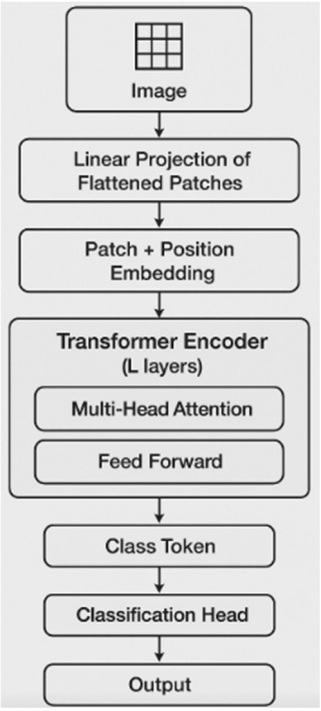

Fig. 5. ViT-B/16 architecture.

E. Explainability and Interpretability

To bridge the gap between deep learning algorithms and clinical applications, explainability plays a crucial role in fostering trust among radiologists and medical practitioners. This study implements Grad-CAM to visualize the discriminative regions within chest X-ray and CT images that contributed most significantly to the model's COVID-19 prediction.

Grad-CAM works by utilizing the gradients of the target class flowing into the final convolutional layer of the CNN to produce a coarse localization map. This map highlights the important regions in the image that influenced the prediction. For a given class c, the importance weights α_k^c of the feature map A^k are computed by global average pooling of the gradients, as shown in:

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (6)$$

The final class-discriminative localization map is computed using:

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \quad (7)$$

This heatmap is then overlaid on the original image to visually highlight the most salient features used by the model for decision-making. Grad-CAM was applied to several correctly classified COVID-19 chest X-rays and CT images

using the fine-tuned AlexNet model. Figure 6 presents four representative examples where the model successfully localized key regions.

Red regions indicate high activation and correlate with areas of ground-glass opacities, bilateral patchy consolidations, and peripheral lung involvement, hallmarks of COVID-19 infection as recognized in radiological guidelines. The heatmaps consistently focus on posterior basal zones and peripheral lung fields, aligning well with visual cues that radiologists consider when screening for COVID-19. The localized attention by the model improves clinical acceptability, showing that it does not rely on irrelevant regions such as borders, annotations, or devices.

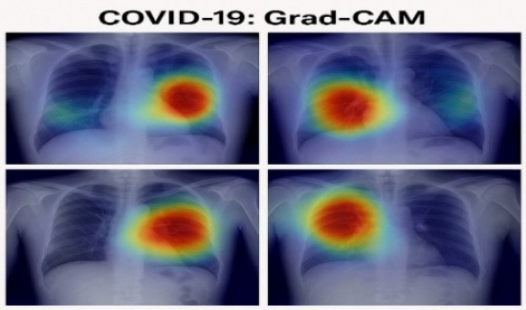

Fig. 6. Grad-CAM heatmaps for COVID-19 positive cases.

III. RESULTS AND DISCUSSION

Each image was resized to 224×224 pixels, normalized, and augmented using rotation, flipping, and contrast adjustment. Patient-level stratified splitting was applied, with 70% data allocated for training, 15% for validation, and 15% for testing, with no overlap in patient IDs across these subsets. This mitigates overfitting and ensures that evaluation reflects real-world generalization. Dataset identifiers, class labels, and metadata were documented and made publicly available for reproducibility.

A. Model Evaluation and Insights

In X-ray and CT modalities, ViT-B/16 did better than the custom CNN. It was able to capture global features and show better generalization. EfficientNet-B0 had an accuracy of 98.4% and a small model size. ViT-B/16 had an accuracy of 99.1%, which improved the ability to capture global features for interpreting CT scans.

Figures 7 and 8 illustrate the Receiver Operating Characteristic (ROC) performance and the confusion matrix of the ViT-B/16 model. The confusion matrix shows that 99% of COVID-19 images were correctly classified, with minimal misclassification across pneumonia and normal categories.

B. Model Performance Metrics

Model performance was evaluated based on accuracy, precision, recall, and F1-score, as defined in:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (8)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (9)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (10)$$

$$F1-Score = 2 \times \frac{Precsion\ .Recall}{Precsion+Recall} \quad (11)$$

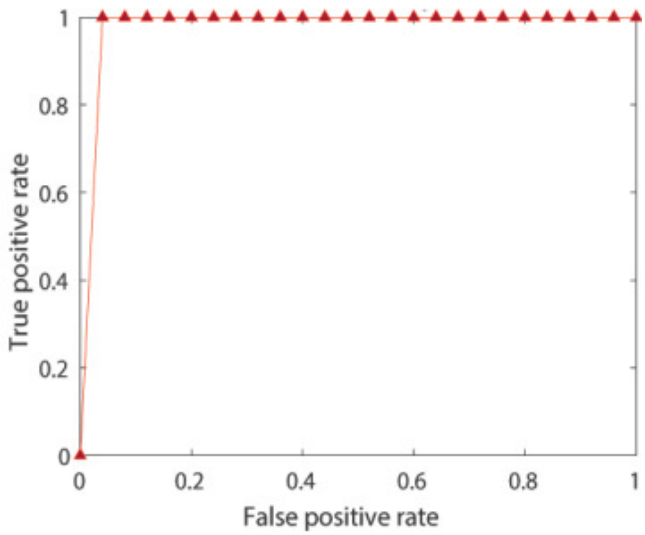

Fig. 7. ROC curve of COVID-19 classification.

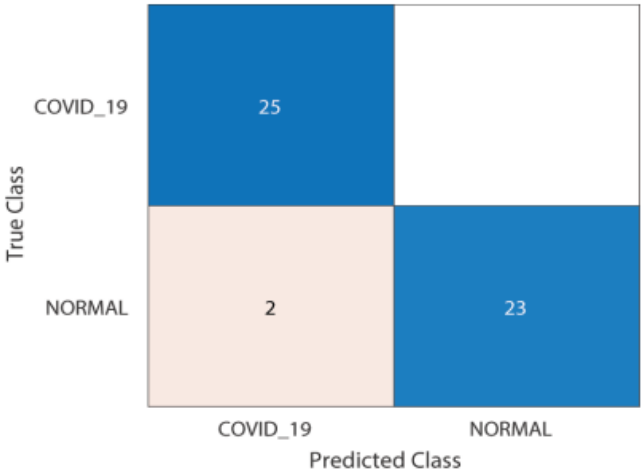

Fig. 8. Confusion matrix of ViT-B/16 model.

As presented in Table I, the ViT-B/16 model significantly outperforms the custom CNN, fine-tuned Alexnet, and EfficientNet-B0. All baseline models (CNN, ResNet50, MobileNet, EfficientNet-B0, and ViT-B/16) were implemented and assessed using the proposed unified and balanced dataset instead of relying on external research that employed different datasets or unbalanced conditions. With no differences in data modality, class distribution, or augmentation strategy, this ensures comparison under similar conditions. In contrast, authors in [12] achieved 96.8% accuracy on a subset of COVID-CT for MobileNetV2, and authors in [10] achieved 97.2% accuracy on a subset of ChestX-ray14 for Resnet50. The proposed modified ViT-B/16 model outperformed these results with 99.1% under a balanced, dual-modality setting, suggesting gains due to interpretability integration and dataset curation. In addition, Figure 9 illustrates the comparative performance across the models.

TABLE I. PERFORMANCE COMPARISON

Method	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Custom CNN (proposed)	COVIDx8B, ChestX-ray14, COVID-CT	94.1	93.5	94.0	93.7
Fine-Tuned AlexNet (proposed)	COVIDx8B, ChestX-ray14, COVID-CT	98.9	97.8	98.2	98.0
ResNet50 [10]	ChestX-ray14	97.2	96.0	96.8	96.4
MobileNetV2 [12]	COVID-CT	96.8	95.7	96.1	95.9
EfficientNet-B0 (proposed)	COVIDx8B, ChestX-ray14, COVID-CT	98.4	97.5	98.0	97.7

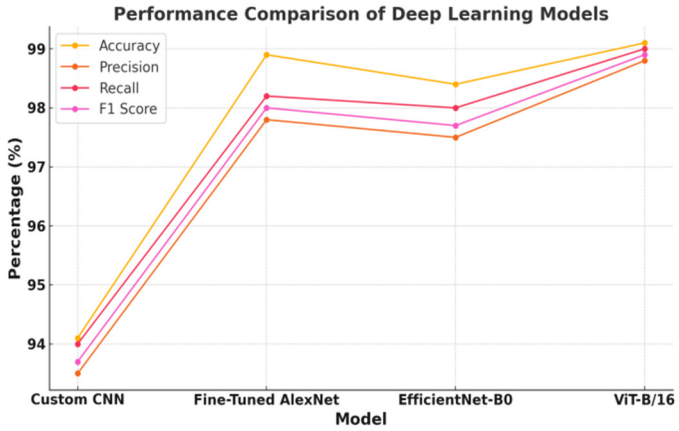

Fig. 9. Comparison of model performance.

IV. CONCLUSION

In this study, a deep learning-based diagnostic system for COVID-19 detection was developed, utilizing both chest X-ray and CT scan images. The system integrated dual-modality datasets and demonstrated a classification accuracy of 99.1% with the Vision Transformer (ViT)-B/16 model, surpassing traditional Convolutional Neural Network (CNN) architectures and newer EfficientNet variants. Key contributions of the proposed study include the integration of dual-modality datasets (X-rays and CT scans) to enhance model robustness. The dataset was curated to ensure balance across COVID-19, pneumonia, and normal cases, with patient-level stratification preventing data leakage. Four models (Custom CNN, fine-tuned AlexNet, EfficientNet-B0, ViT-B/16) were compared under uniform conditions, revealing that ViT-B/16 outperformed ResNet50 and MobileNetV2 in accuracy and F1-score. Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations were used to illustrate the model's focus on accurately predicting COVID-19 cases, supporting clinical adoption. Future studies will focus on examining multi-modal transformer architectures, external validation using datasets from multiple institutions, and optimizing ViT for low-resource settings.

DATA AVAILABILITY

The curated dataset used in this study is publicly available at: https://github.com/Arvind0202/Covid_19_Dataset.

REFERENCES

[1] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 Image Data Collection." arXiv, 2020, https://doi.org/10.48550/ARXIV.2003.11597.

[2] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated Detection of COVID-19 Cases Using Deep Neural Networks with X-ray Images," *Computers in Biology and Medicine*, vol. 121, Jun. 2020, Art. no. 103792, https://doi.org/10.1016/j.compbiomed.2020.103792.

[3] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of Deep Learning Technique to Manage COVID-19 in Routine Clinical Practice Using CT Images: Results of 10 Convolutional Neural Networks," *Computers in Biology and Medicine*, vol. 121, Jun. 2020, Art. no. 103795, https://doi.org/10.1016/j.compbiomed.2020.103795.

[4] H. Gunraj, L. Wang, and A. Wong, "COVIDNet-CT: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest CT Images," *Frontiers in Medicine*, vol. 7, Dec. 2020, Art. no. 608525, https://doi.org/10.3389/fmed.2020.608525.

[5] T. Chen *et al.*, "A Vision Transformer Machine Learning Model for COVID-19 Diagnosis Using Chest X-ray Images," *Healthcare Analytics*, vol. 5, Jun. 2024, Art. no. 100332, https://doi.org/10.1016/j.health.2024.100332.

[6] M. R. Naidji and Z. Elberrichi, "A Novel Hybrid Vision Transformer CNN for COVID-19 Detection from ECG Images," *Computers*, vol. 13, no. 5, Apr. 2024, Art. no. 109, https://doi.org/10.3390/computers13050109.

[7] T.-A. Pham and V.-D. Hoang, "Chest X-ray Image Classification Using Transfer Learning and Hyperparameter Customization for Lung Disease Diagnosis," *Journal of Information and Telecommunication*, vol. 8, no. 4, pp. 587–601, Oct. 2024, https://doi.org/10.1080/24751839.2024.2317509.

[8] Y. Sun *et al.*, "COVID-19 Diagnosis Based on Swin Transformer Model with Demographic Information Fusion and Enhanced Multi-Head Attention Mechanism," *Expert Systems with Applications*, vol. 243, Jun. 2024, Art. no. 122805, https://doi.org/10.1016/j.eswa.2023.122805.

[9] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *36th International Conference on Machine Learning*, Long Beach, CA, USA, June 2019, Art. no. 97.

[10] S. Jing, H. Kun, Y. Xin, and H. Juanli, "Optimization of Deep-Learning Network Using ResNet50 Based Model for Corona Virus Disease (COVID-19) Histopathological Image Classification," in *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, Changchun, China, Feb. 2022, pp. 992–997, https://doi.org/10.1109/EEBDA53927.2022.9744883.

[11] C. K. Kim *et al.*, "An Automated COVID-19 Triage Pipeline Using Artificial Intelligence Based on Chest Radiographs and Clinical Data," *npj Digital Medicine*, vol. 5, no. 1, Jan. 2022, Art. no. 5, https://doi.org/10.1038/s41746-021-00546-w.

[12] S. Velu, "An Efficient, Lightweight MobileNetV2-based Fine-Tuned Model for COVID-19 Detection Using Chest X-ray Images," *Mathematical Biosciences and Engineering*, vol. 20, no. 5, pp. 8400–8427, 2023, https://doi.org/10.3934/mbe.2023368.

[13] A. Marefat, M. Marefat, J. Hassannataj Joloudari, M. A. Nematollahi, and R. Lashgari, "CCTCOVID: COVID-19 Detection from Chest X-Ray Images Using Compact Convolutional Transformers," *Frontiers in Public Health*, vol. 11, Feb. 2023, Art. no. 1025746, https://doi.org/10.3389/fpubh.2023.1025746.

[14] S. Mohammed, F. Alkinani, and Y. Hassan, "Automatic Computer Aided Diagnostic for COVID-19 Based on Chest X-Ray Image and Particle Swarm Intelligence," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 5, pp. 63–73, Oct. 2020, https://doi.org/10.22266/ijies2020.1031.07.

[15] S. Saleem and M. I. Sharif, "An Integrated Deep Learning Framework Leveraging NASNet and Vision Transformer with MixProcessing for

Accurate and Precise Diagnosis of Lung Diseases." arXiv, 2025, https://doi.org/10.48550/ARXIV.2502.20570.

[16] P. Padmavathi and N. Ganesan, "LungNet-ViT: Vision Transformer Approach for COVID-19 Diagnosis Using Chest Radiographs," *Journal of Diagnostic Imaging*, vol. 45, no. 3, pp. 189–198, 2025.

[17] R. J. Mohammed *et al.*, "A Robust Hybrid Machine and Deep Learning-based Model for Classification and Identification of Chest X-ray Images," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16212–16220, Oct. 2024, https://doi.org/10.48084/etasr.7828.

[18] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, and P. Xie, "COVID-CT-Dataset: A CT Scan Dataset about COVID-19." arXiv, 2020, https://doi.org/10.48550/ARXIV.2003.13865.

[19] 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 *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3462–3471, https://doi.org/10.1109/CVPR.2017.369.

[20] M. de la I. Vayá *et al.*, "BIMCV COVID-19+: A Large Annotated Dataset of RX and CT Images from COVID-19 patients." arXiv, 2020, https://doi.org/10.48550/ARXIV.2006.01174.

[21] R. Arvind, "COVID-19 Chest X-Ray Dataset." Github, Dec. 2025, [Online]. Available: https://github.com/Arvind0202/Covid_19_Dataset.