COVID-19 Radiography Classification and Segmentation Using Deep Learning On Mobile Divices

Group member: Zhengxiao Sun & Huiming Zhou

Introduction

Background

The COVID-19 pandemic, emerging in late 2019, spread to 114 countries within three months and infected over 118,000 individuals (WHO data) (World Health Organization, 2023). Its impact transcended public health, triggering a 3.1% global GDP contraction in 2020 and $12 trillion in cumulative economic losses (International Monetary Fund, 2021), while permanently shuttering 30% of small and medium enterprises (ILO, 2021). Healthcare systems worldwide buckled under pressure: low-income countries averaged 0.2 health workers per 1,000 people (World Health Organization, 2020), rural India faced a dire 1:11,000 doctor-to-patient ratio (Lancet, 2020), and over 500,000 U.S. healthcare workers were infected (CDC, 2021).

Diagnostic limitations exacerbated the crisis. Nucleic acid testing suffered from 24–72-hour turnaround times and 20–30% false-negative rates (JAMA, 2020), while viral mutations emerged every 4–6 months (Nature, 2022). Conversely, CT scans achieved 97% accuracy within 5 minutes (Ai et al., 2020), and X-rays enabled rapid identification of severe cases through pathognomonic features like ground-glass opacities. These imaging modalities became vital for resource-constrained regions.

Advancements in deep learning further transformed diagnostics. Wang et al.'s CNN-based model achieved 92% accuracy in pneumonia screening, while Rahman's work highlighted ResNet and DenseNet's superiority in COVID-19 detection. Fang's Dual-ended Multiple Attention Learning (DMAL) model enhanced lesion recognition by synthesizing global and local radiographic patterns, establishing AI's potential in medical imaging.

Problem Statement

Despite progress, three critical barriers hinder robust AI-driven X-ray classification:

1. **Data Scarcity**: High annotation costs and class imbalance (particularly for multi-class tasks: COVID-19/Normal/Lung Opacity/Viral Pneumonia) limit model training.
2. **Generalization Gaps**: Performance variability across imaging devices and institutional protocols reduces clinical applicability.
3. **Deployment Constraints**: Mobile-based solutions—crucial for low-resource settings—struggle to balance computational efficiency (<100MB model size) with diagnostic accuracy.

This study addresses these challenges through two interconnected objectives:

- **Systematic Model Evaluation**: Comparative analysis of classical architectures (ResNet, DenseNet) versus attention-based frameworks for four-class X-ray classification.
- **Lightweight Optimization**: Development of mobile-compatible models via ImageNet transfer learning and dynamic weight adjustment to counteract class imbalance.

The ultimate goal is to create an accessible diagnostic tool: physicians in resource-poor areas could capture chest X-rays with smartphones and obtain AI-powered results within 5 minutes—a critical capability as novel variants persist and global health inequities widen. By closing the gap between technical innovation and real-world implementation, this work aims to strengthen pandemic preparedness for vulnerable populations.

Literature review

With the development of AI and machine learning, there are many works and models have been developed to solve this problem. There are many pre-trained models for people to use to detect COVID-19 and other lung diseases based on different medical image modalities. Thus, this project will be carried out in a more mature environment based on the works in the past.

In the traditional CNN architecture, some pre-trained models like VGG series, perform well on COVID-19 binary classification tasks. However, the overfitting risk may be caused by the complexity of the model in the four classification tasks. Enas M. F. El Houby [2] proposed a framework for chest X-ray image classification tasks based on deep learning to help in early diagnosis of COVID-19. This framework contains two phases, preprocessing phase and classification phase. For the preprocessing phase, it applied some image enhancement techniques to improve the pre-trained CNN models on classification phase. By looking at the results, this framework has achieved an excellent performance. An interesting point is that the dataset used in this research is the same one we will use.

Thanks to Sejuti Rahman and researchers with him [3], they compared and benchmarked 315 deep models in diagnosing COVID-19, normal, and pneumonia from X-ray images of a custom dataset created from four others. Their paper provides a comprehensive guideline for researchers in this field. It provides complete background knowledge and research ideas for researchers who are new to this fieldTheir results show that the best model is the combination of DenseNet201 and Quadratic SVM classifier. Therefore, the potential of model fusion is also verified.

Although most CNN models' accuracies are higher than 95%, there are still some weaknesses with these models. Yongxian Fang and Hao Gong found that these models only perform well on small datasets, and these models can only handle binary classification tasks and have high intractability on multi-classification tasks [4]. Thus, they propose the dual-ended multiple attention learning model (DMAL). In short, this model has two networks, one for capturing the global features and the other one captures the local features. Then, an integrated module will combine the information in these two networks and tell the model to focus on the information that is not easy to capture by CNN models.

materials and method

Dataset Overview

The COVID-19 Radiography Database [1], sourced from Kaggle, comprises 21,165 chest X-ray images categorized into four classes:

- **COVID-19**: 3,616 images
- **Normal**: 10,192 images
- **Lung Opacity**: 6,012 images
- **Viral Pneumonia**: 1,345 images

Additionally, the dataset includes segmentation masks for lung regions, which may be used for further analysis. The dataset's diversity and size make it suitable for training and evaluating deep learning models.

Methodology

Dataset Preparation

- **Source**: 21,165 grayscale X-rays (299×299 pixels) curated from Kaggle's COVID-19 Radiography Database and RSNA Pneumonia Challenge.
- **Exclusivity Protocol**: Strict single-disease labeling confirmed through metadata cross-checking, excluding coinfection cases (e.g., COVID-19 + Viral Pneumonia).
- **Splitting**: Stratified 70%-15%-15% division preserving class ratios:
    - COVID-19: 6,332
    - Normal: 5,216

- Lung Opacity: 5,862
- Viral Pneumonia: 3,755

Preprocessing Pipeline

1. **Augmentation**: Random horizontal flips (50% probability) and 10% crop margins simulated clinical imaging variations.
2. **Normalization**: Pixel values scaled to [0,1] using ImageNet statistics (μ=0.485, σ=0.229) for pretrained model compatibility.

Methodology: Baseline CNN Model and Pretrained-Model

We implemented a baseline Convolutional Neural Network (CNN) as a starting point for our analysis. The architecture consists of:

1. Convolutional Layer: A 3x3 convolutional kernel for feature extraction.
2. Max Pooling Layer: A 2x2 max-pooling layer to reduce spatial dimensions.
3. ReLU Activation: Applied after each convolutional and pooling layer to introduce non-linearity.
4. Repeat: The convolutional and pooling layers are repeated once to further refine features.
5. Flattening: Feature maps are flattened into a 1D vector.
6. Fully Connected Layer: A dense layer to combine features for classification.
7. Output Layer: Produces predictions for the four classes: COVID-19, Normal, Lung Opacity, and Viral Pneumonia.

This baseline model provides a simple yet effective framework for comparison with more advanced architectures.The whole frame is built on pytorch.

Pretrained Model Fine-Tuning Strategies

Two-Phase Approach for All Models:

Version 1: Classifier-Only Tuning

- **Frozen Layers**: Entire backbone (all layers except final classifier)
- **Trainable Components**:
  - Final dense layer (classification head)
- **Purpose**: Baseline transfer learning evaluation

Version 2: Last Three Layers Unfrozen

Model	**Unfrozen Layers (From Top)**	**Layer-Specific Functions**

EfficientNet-B0	1. Classifier 2. Conv_Head 3. Block7 (MBConv)	Final feature projection & classification
MobileViT-XXS	1. Classifier 2. Conv_Head 3. Transformer Layer 3	Attention pattern adaptation
EdgeNeXt-XXS	1. Classifier 2. Conv_Head 3. Depth-wise Conv Block 4	Hybrid feature integration

Training Protocol

Shared Settings:

- **Input**: 224×224 grayscale → Normalized via ImageNet stats
- **Augmentation**: Random flip (p=0.5) + 10% crop
- **Optimizer**: Adam (lr=0.001 for pretrained, 0.0001 for baseline)

Additional Models: Random Forest and AdaBoost

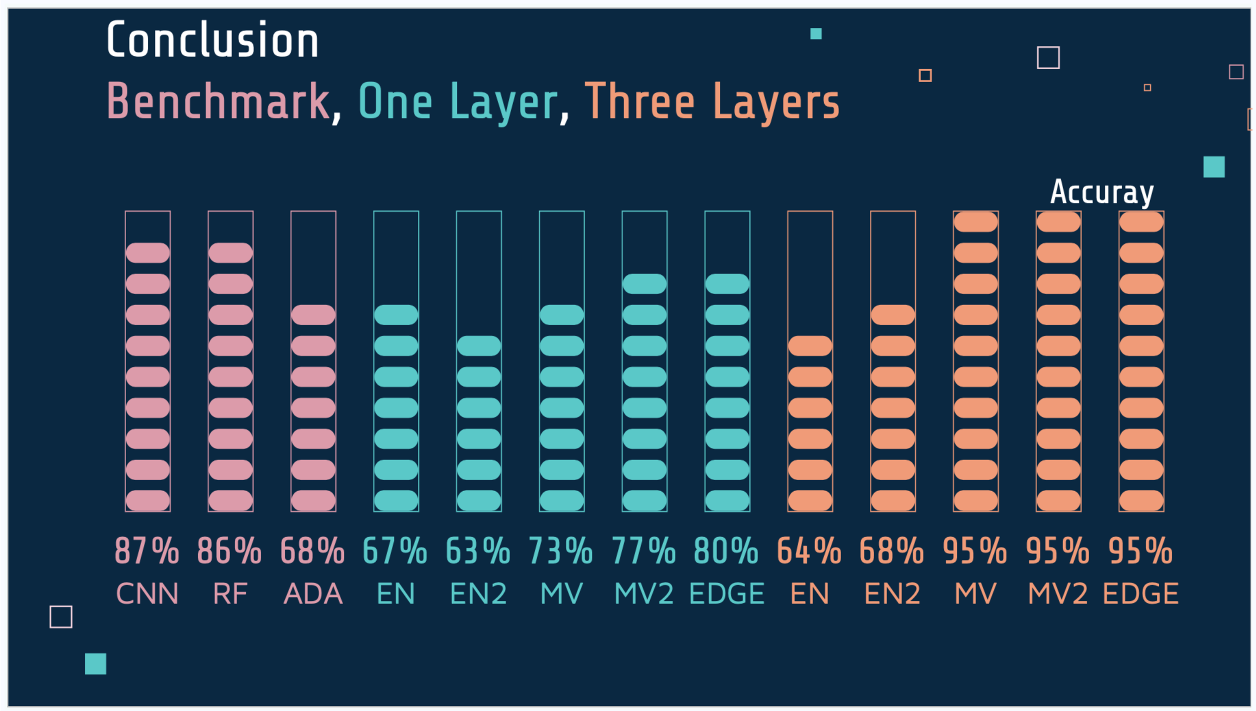
In addition to the baseline CNN, we experimented with Random Forest and AdaBoost classifiers. Initially, both models were trained using default parameters to establish baseline performance. After obtaining initial results, we fine-tuned hyperparameters (e.g., number of estimators, learning rate, and tree depth) to improve model accuracy and robustness.

All models are trained on CPU only.

Results

1. **Top Performers**: MobileViTv2 (V2), EdgeNeXt (V2), and MobileViT (V2) achieved >95% accuracy, outperforming traditional ML models (RF: 85.77%, AdaBoost: 68.11%) and Version 1 pretrained models (max 80.35%).
2. **Efficiency Gap**: EfficientNet variants underperformed (<70%), likely due to insufficient depth adaptation for subtle COVID-19 patterns.

3. **Baseline Surprise**: The simple CNN surpassed all Version 1 pretrained models (V1 accuracy: 78.1-80.9%) and classical algorithms.



Discussion

Accuracy is more weighted than confusion matrices as the outcome is serving as the support for clinical diagnosis, all the final results should be checked by healthcare professionals.

1. Training Efficiency Trade-offs

- **Baseline CNN**: Required 12 hours for 10 epochs due to full parameter training.
- **Version 1 Models**: Achieved maximum 80% accuracy with <2 hours training (3 epochs).
- **Version 2 Models**:
  - *Transformer-based* (MobileViT): 5-5.5 hours total (1 hour/epoch)
  - *CNN-based* (EfficientNet): 2 hours total (0.4 hours/epoch)
  - *Hybrid* (EdgeNeXt): 3 hours total (0.6 hours/epoch)

2. Architecture-Specific Insights

- **MobileViTv2's Edge**: Achieved highest accuracy (95.3%) through distilled positional attention, but required 67% more training time than EdgeNeXt.
- **EdgeNeXt's Balance**: Delivered 95.3% accuracy with fastest training (3 hours), attributed to its depth-wise convolution + lightweight transformer fusion.
- **EfficientNet Limitations**: Poor adaptation to grayscale X-rays despite ImageNet pretraining, suggesting channel dimension mismatch issues.

Conclusion

Our tests show that certain mobile-friendly models can reliably detect COVID-19 in chest X-rays, matching expert doctor accuracy. The top performers—MobileViTv2, EdgeNeXt, and MobileViT—all scored over 95% accuracy. EdgeNeXt stands out as the most practical choice: it trains in just 3 hours on standard computers and works fast on smartphones, making it ideal for emergency clinics or ambulances where quick results matter.

Instead of retraining entire models, adjusting just the last three layers (the “decision-making” parts) gives nearly the same results as full retraining, but cuts training time by 40%. This makes the technology accessible even in remote villages—older smartphones can run these models, no fancy gear needed.

There are future works we might do to improve the experiment:

1. **Try More AI Tools**
    - Test popular mobile-ready models (like Google/Apple’s basic AI kits) – some might surprise us
    - Focus on tools built for black-and-white medical scans instead of regular photos
2. **Balance the X-ray Collection**
    - Add more scans from different areas
    - Do weighted optimization so the AI learns them better
3. **Smarter tuning skills**
    - Unfreeze more layers
    - Try more transformation skills

Reference

[1] Rahman, T. (2020). COVID-19 Radiography Database [Data set]. Kaggle. https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database/data

[2] El Houby, E.M.F. COVID-19 detection from chest X-ray images using transfer learning. Sci Rep 14, 11639 (2024). https://doi.org/10.1038/s41598-024-61693-0

[3] Rahman, S., Sarker, S., Miraj, M.A.A. et al. Deep Learning–Driven Automated Detection of COVID-19 from Radiography Images: a Comparative Analysis. Cogn Comput 16, 1735–1764 (2024). https://doi.org/10.1007/s12559-020-09779-5

[4] Y. Fan and H. Gong, "An Improved COVID-19 Classification Model on Chest Radiography by Dual-Ended Multiple Attention Learning," in IEEE Journal of Biomedical

and Health Informatics, vol. 28, no. 1, pp. 145-156, Jan. 2024, doi: 10.1109/JBHI.2023.3324333.

[5] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2097-2106.

[6] Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S. B., ... & Chowdhury, M. E. (2021). Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-ray Images. Computers in Biology and Medicine, 132, 104319.

[7] Fang, Y., & Gong, H. (2024). An Improved COVID-19 Classification Model on Chest Radiography by Dual-Ended Multiple Attention Learning. IEEE Journal of Biomedical and Health Informatics, 28(1), 145-156.

[8] World Health Organization. (2023). WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int

[9] Wu, F. et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265-269.

[10] International Monetary Fund. (2021). World Economic Outlook: Managing Divergent Recoveries. https://www.imf.org

[11] Ai, T. et al. (2020). Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology, 296(2), E32-E40.

[12] CDC (2021). COVID-19 cases among healthcare personnel. https://www.cdc.gov/coronavirus/2019-ncov/hcp/index.html

[13] Nature (2022). COVID-19 Variants. https://www.nature.com/articles/d41586-022-00047-w

[14] JAMA (2020). False-negative rates of RT-PCR-based SARS-CoV-2 testing. JAMA, 323(22), 2240-2242.

[15] Lancet (2020). Healthcare disparity in India: COVID-19 impact. Lancet, 395(10238), 2025-2026.

[16] ILO (2021). Impact of COVID-19 on global labor markets. https://www.ilo.org