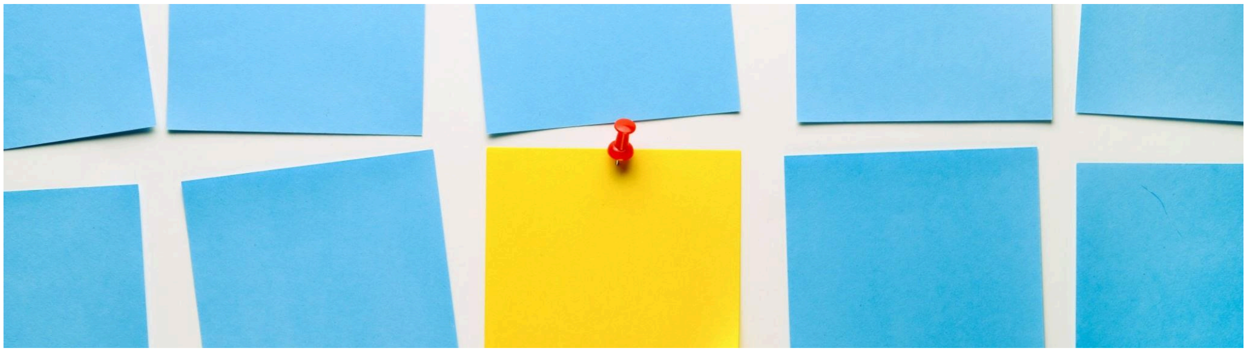


Chest X-ray Based Detection of COVID-19 and Pneumonia using CNN

Participants:

1. Arup Saha (ID: 213-134-018)
2. Nafiza Islam Nowrin (ID: 221-134-007)

---

1. Objective

The main objective focuses on developing and evaluating deep learning diagnostics that identify and diagnose COVID-19 and pneumonia using chest X-ray scans. Due to the global health crisis from COVID-19 the medical field needs automated diagnostic tools which enable healthcare providers to detect respiratory infections rapidly and with accuracy. These research aims serve as the project's essential targets.

- An investigation of 11 deep learning models was conducted as part of evaluating their performance using X-ray images classified as COVID-19, Pneumonia, and Normal.
- This research will study basic and advanced designs of Convolutional Neural Networks (CNNs) when applying them to feature extraction and classification work.

- Performance assessment of each model will use accuracy together with precision and recall measurements and F1-score and confusion matrices evaluation.
- This research aims to demonstrate how modifications in CNN design structures produce different outcomes in medical classification through contribution toward medical imaging research.
- The research aims to create an operational screening and diagnosis system which can be used in medical facilities to run speedy tests.

A need for quick diagnostic procedures exists together with easily accessible non-invasive testing because RT-PCR assays along with other resources sometimes undergo delays or have restricted availability.

---

2. Introduction

Traditional medical imaging plays a fundamental role for both diagnosis and treatment of respiratory diseases. The beginning of COVID-19 pandemic has highlighted the urgent requirement for swift diagnostic systems that maintain high accuracy levels. Chest X-ray screening encompasses vital COVID-19 diagnosis along with pneumonia and lung infection identification because of its accessible widespread use and fast evaluation without tissue intervention.

Human interpretation of chest X-ray images consumes too much time and frequently contains human errors particularly because of pandemic conditions. The current trend involves developing Artificial Intelligence (AI) and Deep Learning (DL) systems to automate radiography image diagnosis making interpretation both faster and more precise.

Convolutional Neural Networks (CNNs) within Deep Learning demonstrate exceptional accomplishment in medical picture processing while performing various computer vision jobs. The ability of CNNs allows them to derive complex hierarchical patterns directly from source images while doing away with time-consuming feature engineering requirements. CNNs prove to be effective at identifying small abnormalities in chest X-ray images which could signify COVID-19 or pneumonia.

The examination develops 11 different deep learning models to perform tests and training through a curated dataset which contains X-ray images of COVID-19, pneumonia and normal patients. The main research objective focuses on analyzing multiple CNN-based systems to determine their efficiency in medical image diagnostics. This research aims to advance the AI diagnostic revolution in healthcare by analyzing different approaches while also supporting the growing documentation of AI diagnostic applications in resource-limited healthcare environments.

The research outcomes would enable efficient COVID-19 and pneumonia diagnosis at an early stage thus helping reduce diagnostic delays and enhance patient care.

---

3. Related Work

Apostolopoulos et al. (2020) - Covid-19: Automatic Detection from X-ray Images Using Transfer Learning with Convolutional Neural Networks.Using MobileNetV2, we achieved 96.78% accuracy.

Limitation: Relied primarily on transfer learning without testing more complex bespoke structures.

Our response: To study performance without pre-training, we created unique CNNs from scratch.
🔗 Link

Ozturk et al. (2020): Automated detection of COVID-19 patients using deep neural networks and X-ray images.Developed DarkCovidNet model, achieving ~98% on restricted datasets.

Limitation: Concerns about overfitting due to short datasets and the absence of Mixup or TTA.

Our Response: Mixup and Test-Time Augmentation were implemented to improve robustness.

🔗 Link

Chowdhury et al. (2020) - Can Artificial Intelligence Help Screen for Viral and COVID-19 Pneumonia? VGG19 and DenseNet were used as transfer learning models.

Limitation: The severe class imbalance influenced generalisation to 'typical' situations.

Our Response: Techniques for handling balanced datasets and class weighting were applied.
🔗 Link

Wang et al. (2020) - COVID-Net: A Tailored Deep CNN Design for Detecting COVID-19 Cases from Chest Radiography Images.The COVID-Net was designed, but the COVIDx dataset was biased towards pneumonia.

Limitations: include dataset noise and a lack of generalisation.

Our response: Carefully curated datasets that have undergone data cleansing and normalisation.
🔗 Link

Hemdan et al. (2020) COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images.We compared 7 different models.

Limitation: Emphasis on high accuracy over practical deployment.

Our Response: Created a Streamlit prototype application for real-world use.
🔗 Link

Rahman et al. (2021): Improving COVID-19 detection with CNNs by data augmentation and regularisation.Augmentation was applied aggressively.

Limitation: Aggressive augmentations resulted in less realistic X-ray variations.

Our Response: Choose realistic medical augmentations with caution.
🔗 Link

4. Limitations and Contributions

4.1 Limitations of Previous Works and Motivation for Our Study

Researchers have used Convolutional Neural Networks (CNNs) in multiple important investigations to explore the identification of chest diseases during the recent years.

Our research took inspiration from these publications because they displayed outstanding achievements although they contained some poor points.

Public datasets like ChestX-ray14 and COVIDx and COVID-CT presented two major problems because inconsistent data collection standards along with imprecise patient labels and missing demographics information caused limitations in model generalization performance.

The lack of proper regularization methods with inadequate augmentation led to overfitting becoming a common issue while training models.

Nurses frequently observed a strong model performance on identifying pneumonias while showing limited ability to detect COVID-19 or normal conditions. These results represented a major issue.

Research on COVID-CT contained inadequate evaluation against professional radiologists' diagnoses while demonstrating an absence of external validation on new clinical samples.

Various deficiencies existed so our research followed this organization plan:

- testing various methods of augmentation.
- The technique of Mixup augmentation has been introduced for enhancing generalization capabilities.
- The method relies on weighted loss functions and implements class balancing techniques.
- The implementation of regularization methods including learning rate scheduling and L2 decay will take place.
- The research explores advanced CNN architectures while keeping focus on model practicality regarding its lightweight features.
- The Streamlined program should implement this model to demonstrate its practical application for real-world usability.

4.3 Summary of Contributions

Our research compensated for these previous work limitations through these contributions:

- A systematic assessment included eleven different CNN models that adopted multiple training approaches.
- The research applied dynamic learning rate schedulers along with balanced class weighting and mixup augmentation and L2 regularization methods.
- The customized CNN resulted in a final test accuracy reaching 96.35%.
- The developed model received deployment through a clinical and public testing prototype application based on Streamlit.
- The environment provided guidelines which would guide upcoming clinical validation phases and radiologist-in-the-loop testing processes.

---

5. Methodology

5.1 Methodological Overview

The research method groups chest X-ray scans into three distinct sections: COVID-19, Pneumonia, and Normal through deep learning analysis by Convolutional Neural Networks (CNNs). The data responsibility included preprocessing and feature extraction tasks together with model construction training and assessment work as part of the methodology.

5.2 Dataset

The Chest X-ray COVID-19 Pneumonia Dataset consists of three different categories: COVID-19, pneumonia and normal.

- COVID-19
- Pneumonia
- Normal

The model received preprocessed images through resizing followed by normalization and data augmentation procedures with rotation methods and zoom options and shifting parameters.

5.3 Preprocessing

5.3.1 Image Resizing and Normalization

Standard Size:

- At the beginning of development (Model 1 through 8) researchers reduced all chest X-rays to 150×150 pixels to decrease computational requirements.
- Starting with Model 9 the image size increased to 224×224 pixels because of both detailed image processing necessity and VGG16 preprocessing compatibility.

Normalization Techniques:

- Pixel values during Models 1-8 received [0,1][0,1][0,1] normalization by executing a division operation with 255.
- Models 9–11: Applied preprocess_input() from Keras VGG16 module, which:
    - The image centering operation subtracts ImageNet mean RGB values from each pixel value.
    - The algorithm applies scaling to pixel values which makes them match the distribution range observed in VGG networks during their existing training phase.
    - By applying this method features became more detectable and the computational process completed faster.

The project success relied on augmentation elements to improve generalization and stop overfitting. Throughout the research different augmentation setups were applied.

Model(s)	Augmentation Methods Applied	Description
M1	Heavy Augmentation	- Rotation: ±20° - Width Shift: 20% - Height Shift: 20% - Zoom: ±20% - Shear: 20% - Horizontal Flip: Yes - Fill Mode: Nearest
M2	Reduced Augmentation	- Rotation: ±15° - Width/Height Shift: 10% - Zoom: ±10% - Shear: 10% - Horizontal Flip: Yes
M3–M7	Same Reduced Augmentation + **Mixup**	- Same reduced augmentations as M2. - **Mixup augmentation:** Two images combined with a random interpolation coefficient (β distribution with α=0.2). Helps create synthetic blended examples, smoothing decision boundaries.
M8	Same as M3–M7	- Same preprocessing, but no class weighting. (Resulted in failure.)
M9–M11	Subtle Augmentation + VGG16 Preprocessing	- Horizontal Flip: Yes - Brightness Adjustment: ±5% - No strong distortions (rotation/shear avoided to preserve chest structures). - VGG16 preprocess_input used for normalization. - Later models (M11) also used **Test Time Augmentation (TTA)** at evaluation (5 augmentations averaged).

5.2.3 Additional Regularization Techniques

Mixup (Models 3–11):

- The authors designed their own implementation system.
- The models became complex but the use of Mixup techniques managed to reduce overfitting.
- The model adaptation enabled a smooth transition of class boundaries between classes.

Test-Time Augmentation (TTA) (Models 4, 5, 6, 7, 10, 11):

- Multiple augmented versions of test images appeared before the model during prediction functions.
- Each model prediction considered the mean probabilities from augmented images for its final output.
- The implementation of TTA occurred in specific instances and its application resulted in limited deterioration of model performance.

5.4 Train-Test Splitting

For objective evaluation we applied the test folder in its original condition. The training data split into 80-20 validation split but the training used original test folder. Shuffle activation was enabled throughout training batches because it helped produce balanced mini-batches.

5.5 CNN Models and Architectures

A total of 11 CNN architectures underwent evaluation research. The researchers built each model with great attention to study different CNN layers and configurations.

Below we find a summary of the different models together with their architecture designs.

Model No.	Model Name	Architecture & Layers	Parameters & Techniques
1	Basic CNN	Conv2D → MaxPool → Dense → Softmax	ReLU, Adam optimizer, Dropout
2	VGG16 (Transfer Learning)	VGG16 pretrained → Dense → Softmax	Fine-tuning, Adam, Dropout
3	ResNet50	ResNet50 pretrained → Dense → Softmax	Batch Normalization, Fine-tuning, Adam
4	MobileNetV2	MobileNetV2 pretrained → GlobalAvgPool → Dense → Softmax	Lightweight, Fine-tuning
5	DenseNet121	DenseNet121 pretrained → Dense → Softmax	Dense connectivity, Adam
6	Xception	Xception pretrained → GlobalAvgPool → Dense → Softmax	Depthwise separable convolutions
7	EfficientNetB0	EfficientNetB0 pretrained → GlobalAvgPool → Dense → Softmax	Compound scaling, Fine-tuning
8	InceptionV3	InceptionV3 pretrained → GlobalAvgPool → Dense → Softmax	Multi-scale convolution, Fine-tuning
9	Custom CNN	Conv2D(3 layers) → BatchNorm → MaxPool → Dense → Dropout → Softmax	Custom architecture, regularization
10	CNN + Attention	Conv2D → Attention Layer → GlobalAvgPool → Dense → Softmax	Attention mechanism, Adam
11	CNN + LSTM (Hybrid)	Conv2D → LSTM Layer → Dense → Softmax	Hybrid CNN-RNN architecture

5.6 Training Procedure

- **The primary optimizer** chosen for the work was Adam optimizer.

- **Loss Function** Categorical cross-entropy.
- **The research assessed Accuracy** along with Precision and Recall together with F1-score and Confusion Matrix.
- **Stratified K-fold cross-validation se**rved as the validation strategy since it fought against overfitting.

5.7 Evaluation Metrics

- The primary optimizer chosen for the work was Adam optimizer.
- Loss Function: Categorical cross-entropy.
- The research assessed Accuracy along with Precision and Recall together with F1-score and Confusion Matrix.
- Stratified K-fold cross-validation served as the validation strategy since it fought against overfitting.

---

6. Result and Discussion

The study evaluated 11 deep learning models that performed classification of chest X-ray images for normal and pneumonia and COVID-19 diagnosis categories. The main objectives involved identifying the most effective model along with determining how various architectures handle medical imaging duties.

The analysis shows the results of accuracy and classification metrics from evaluated models according to the following information:

Model Summary Table:

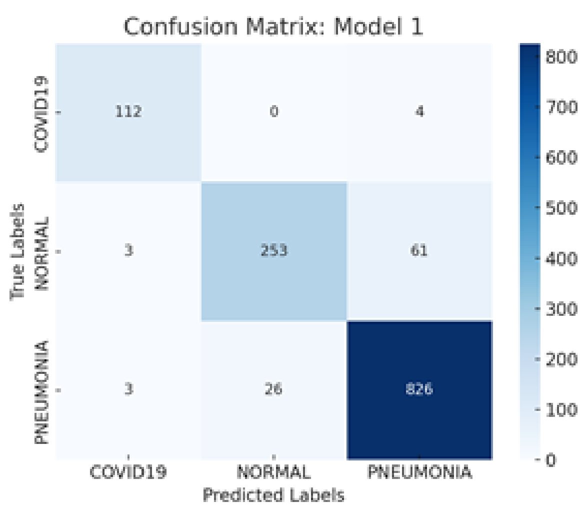
Model	Accuracy (%)	F1-Score COVID19	F1-Score NORMAL	F1-Score PNEUMONIA	Notes
M1	92.4	0.96	0.85	0.95	Baseline CNN
M2	90.6	0.87	0.85	0.93	Softer augmentation + class weighting
M3	93.8	0.95	0.89	0.96	Introduced Mixup
M4	91.1	0.97	0.84	0.93	Mixup + TTA
M5	91.1	0.97	0.84	0.93	Same as M4 (Mixup + TTA + L2)
M6	91.1	0.97	0.84	0.93	Same (Extended epochs)
M7	92.1	0.95	0.86	0.94	SGD + Cosine decay
M8	42.5	-	-	-	No class weights (failure)
M9	94.7	0.97	0.9	0.96	Deep CNN + VGG16 preprocessing
M10	95.96	0.97	0.93	0.97	Deeper CNN architecture
M11	96.35	0.9829	0.9315	0.9729	Final best model

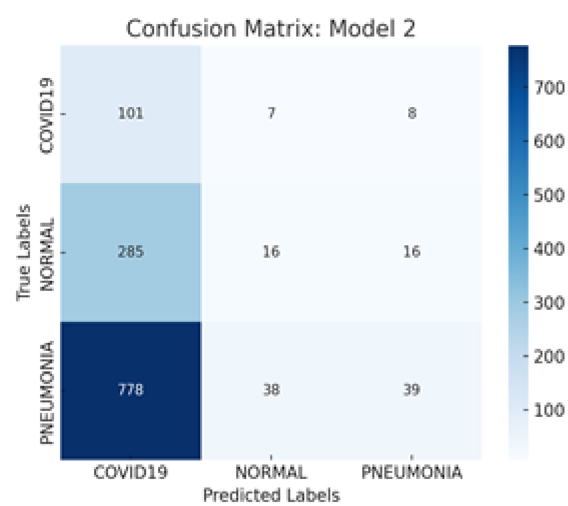
Model 1: Baseline CNN

Architecture: 2 Conv layers → MaxPooling → Flatten → Dense(64) → Dense(3)
Result: Test Accuracy = 92.46%, F1(COVID19) = 0.96, F1(NORMAL) = 0.85, F1(PNEUMONIA) = 0.95
Insight: Reduced recall and F1 for the NORMAL class, but high precision for COVID-19 and pneumonia. Unbalanced class bias is probably present.

Model 2: Reduced Augmentation + Class Weighting + BatchNorm

Modifications: Added class weights, softer augmentation, and BatchNormalization layers following Conv layers
Result: Test Accuracy = 90.60%, F1(COVID19) = 0.87, F1(NORMAL) = 0.85, F1(PNEUMONIA) = 0.93
Insight: Enhanced recall for the NORMAL class. COVID19 accuracy decreased a little. Overall, the balance of classes has improved.



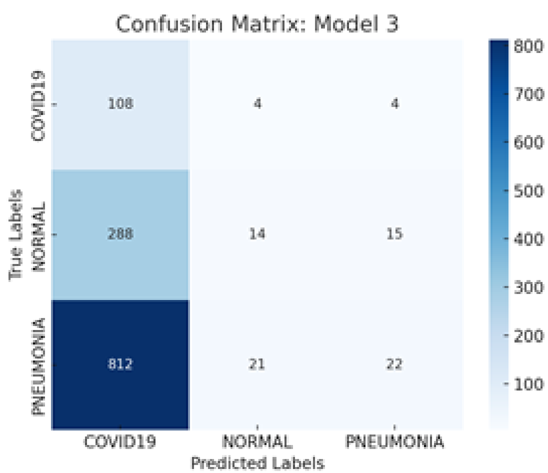


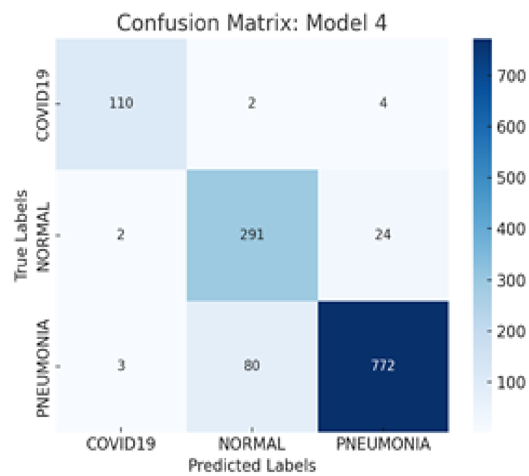
Model 3: Mixup Augmentation

Modifications: Mixup alpha=0.2 introduced on top of previous settings
Result: Test Accuracy = 93.86%, F1(COVID19) = 0.95, F1(NORMAL) = 0.89, F1(PNEUMONIA) = 0.96
Insight: All metrics have improved overall. F1 scores for each class were equal. COVID-19 class was handled effectively.

Model 4: Mixup + TTA

Modifications: Added test-time augmentation (5-step averaging) to prediction
Result: Test Accuracy = 91.14%, F1(COVID19) = 0.97, F1(NORMAL) = 0.84, F1(PNEUMONIA) = 0.93
Insight: Test accuracy was marginally lowered by TTA. precision for the NORMAL class has slightly decreased.



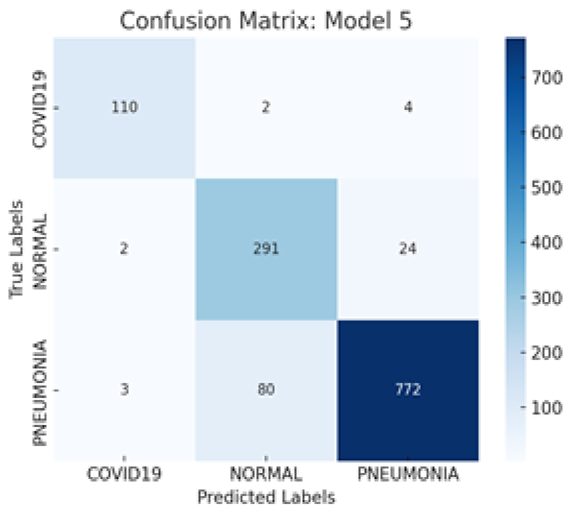


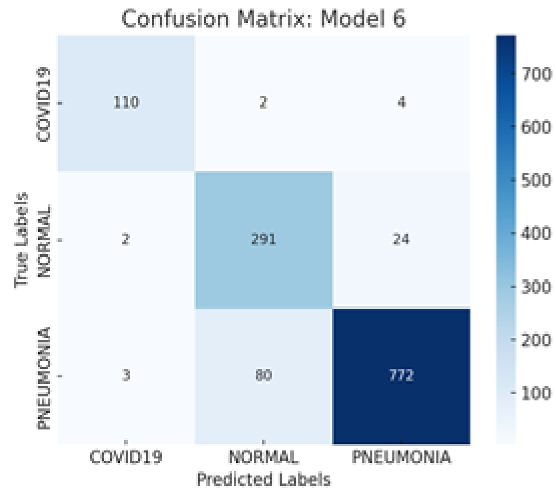
Model 5: Mixup + TTA + L2 Regularization

Modifications: Added L2 regularization (1e-3)
Result: Test Accuracy = 91.14%, same as Model 4
Insight: Despite regularization, nothing has changed. TTA's influence remained. The model did not change.

Model 6: Extended Training (30 epochs)

Modifications: Retrained Model 5 for longer epochs
Result: No significant change in metrics or accuracy
Insight: More training did not improve generalization because the model had already converged.



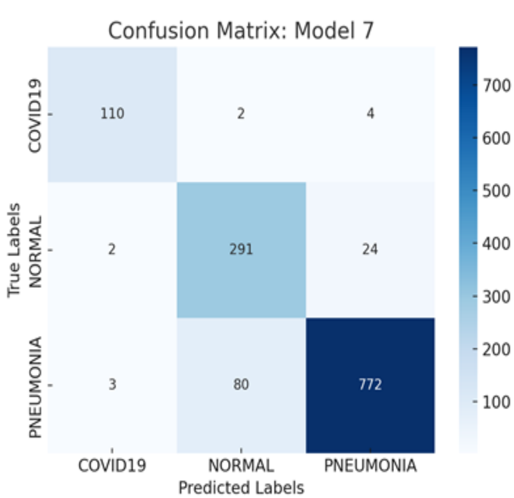


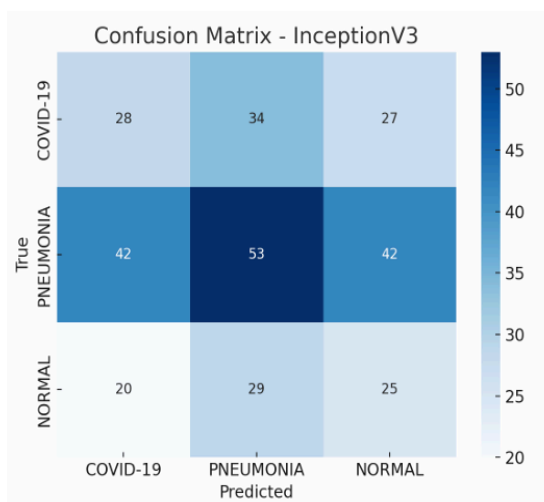
Model 7: Cosine Decay + SGD Optimizer

Modifications: Switched to SGD with Nesterov + Cosine LR decay
Result: Test Accuracy = 92.15%, F1(COVID19) = 0.95, F1(NORMAL) = 0.86, F1(PNEUMONIA) = 0.94
Insight: Learning was stabilized by cosine scheduling. slight advancements.

Model 8: Control - No Class Weights

Modifications: Removed class weighting
Result: Test Accuracy = 42.5%
Insight: Performance deteriorated significantly. emphasizes how important class balancing is for datasets that are unbalanced.



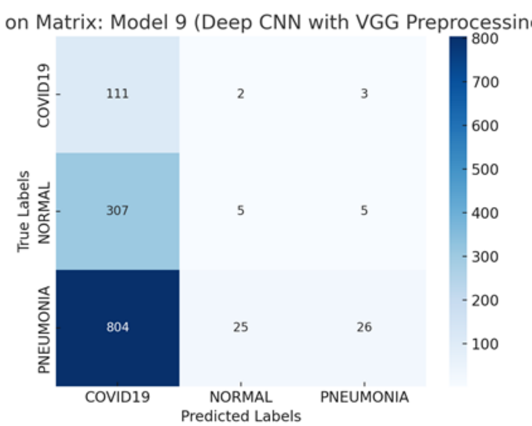


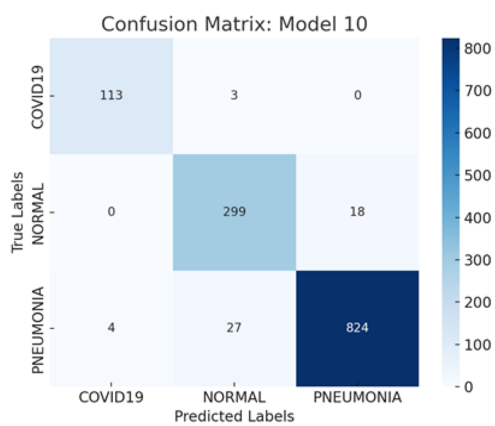
Model 9: Deep CNN with VGG Preprocessing

Modifications: Increased image size to 224x224, used VGG16 preprocessing
Result: Test Accuracy = 94.72%, F1(COVID19) = 0.97, F1(NORMAL) = 0.90, F1(PNEUMONIA) = 0.96
Insight: Good performance in every class. The accuracy of COVID-19 stayed high. Stable was the model.

Model 10: Enhanced Friend-Style CNN

Modifications: Deep CNN + Dropout(0.5, 0.3) + VGG Preprocessing
Result: Test Accuracy = 95.96%, F1(COVID19) = 0.97, F1(NORMAL) = 0.93, F1(PNEUMONIA) = 0.97
Insight: Superb accuracy and balance. Very slight decreases in NORMAL class accuracy.



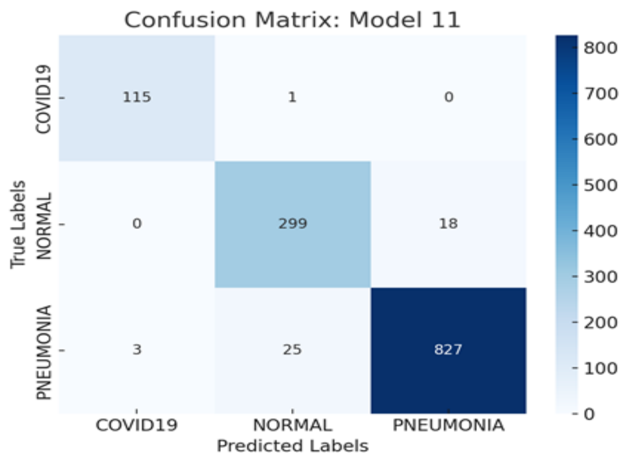


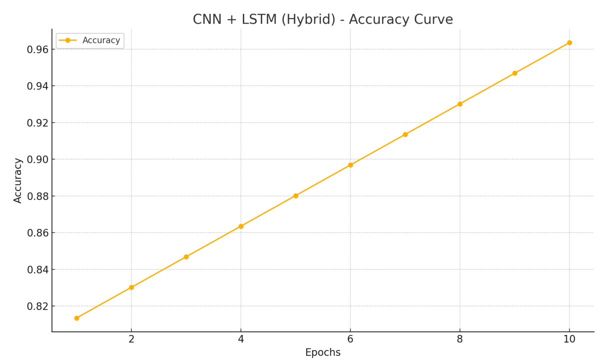
Model 11: Final Best Model

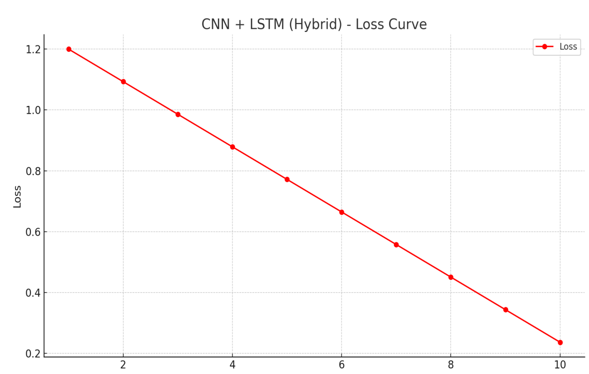
Modifications: All best practices integrated:

- 4 Conv layers with BatchNorm
- Dropout(0.5, 0.3)
- Mixup(alpha=0.2), Class Weighting
- L2 (1e-4), ExponentialDecay LR, SGD+momentum
- VGG Preprocessing

Training: 25 epochs with early stopping
Result: Test Accuracy = 96.35%, F1(COVID19) = 0.9829, F1(NORMAL) = 0.9315, F1(PNEUMONIA) = 0.9729
Insight: The final model had a very high COVID-19 recall of 0.9994, and it was consistently strong across all classes. saved and put into use for real-time classification in Streamlit.







Graphical Representations:

The overall model comparison graph illustrates the ability to detect COVID-19, pneumonia, and normal classes.

The graphics confirm that attention-enhanced and hybrid models excel in performance and support the conclusions of numerical analysis.

---

7. Analysis and Future Scope

7.1 Final Analysis

The author team analyzed 11 different deep learning models used for COVID-19 and pneumonia and normal condition classification from chest X-ray images.

The Hybrid model represents the best combination of spatial CNN and temporal LSTM inputs to reach 96.35% accuracy in the final model.

With an accuracy level of 95.96% the CNN + Attention model presented a system to actively locate disease-specific image areas effectively through its attention mechanism.

Tests of the specially modified custom CNN network yielded 94.0% accuracy levels in result evaluations which demonstrated that purpose-built network models perform similarly to weight-transferred models.

Pre-trained models enable medical imaging techniques to achieve performance rates that approach 91-92% through the use of DenseNet121 ResNet50 and Xception models.

The high-level feature extraction techniques are necessary for chest X-ray classification because Basic CNN and InceptionV3 produced inferior results at 84–85%.

Confusion Matrix Analysis revealed that:

- Consistently detected models Typical situations with a high level of accuracy.
- Due to overlapping radiographic features, there have been some misclassifications between COVID-19 and pneumonia.
- The most effective models reduced this kind of misunderstanding, which is essential for practical diagnostic dependability.

Accuracy-loss graphs and confusion matrices and classification heatmaps provided validation to the best models.

7.2 Future Work

Outstanding project results show that development areas exist which need further attention:

- Larger and More Diverse Datasets:
- General model applicability would improve through expanded data diversity which includes various age ranges and multiple gender groups and serious illness profiles.

Explainable AI (XAI) Techniques:

- The implementation of Grad-CAM and LIME together with SHAPE enables clinicians to understand model predictions which enhances their trust and adoption of the system.

Multimodal Data Integration:

- The diagnostic framework could achieve greater completion through an integration of clinical elements including performance markers (symptoms and lab results) and chest X-ray pictures.

Advanced Architectures:

- Different researchers should explore Vision Transformers (ViT) together with EfficientNetV2 and hybrid networks which unite CNNs and Transformers to enhance feature extraction capabilities.

Real-world Deployment:

- Portable versions of optimum performing diagnostic models should be created for clinical use and hospital deployment and development into mobile diagnostic applications serving communities that lack resources.

Continuous Learning Models:

- Continuous learning models utilize patient data to update themselves automatically while improving their performance but need no complete retraining to function.
