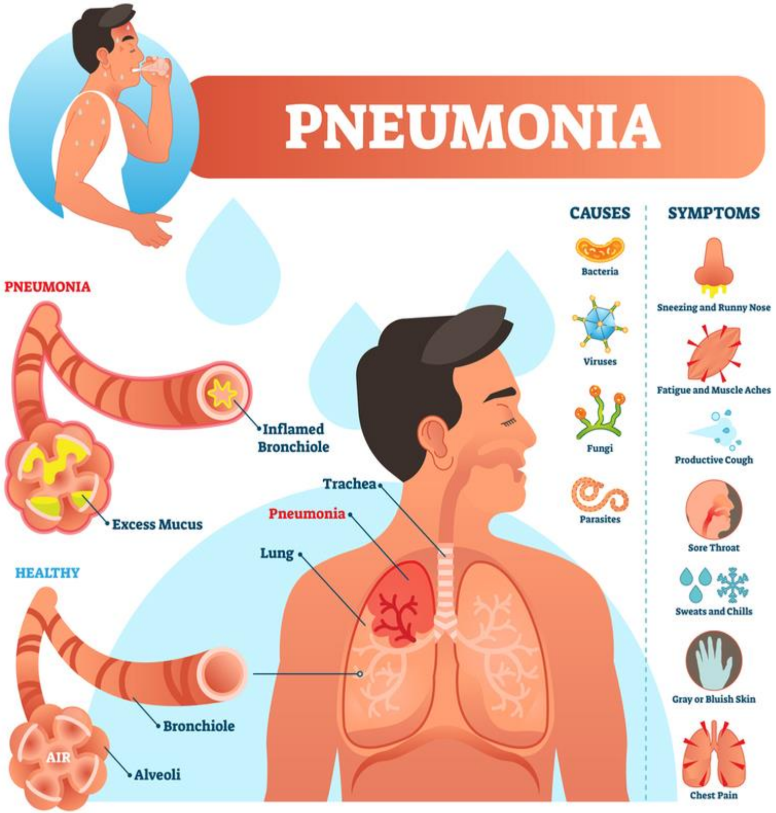
Automated Pneumonia Detection and Image Analysis Using Deep Learning and Convolutional Neural Networks (CNN) on Chest X-ray Images

GROUP MEMBERS:
KAVYA SRI PACHCHAVA, SESHA SAI RAMINENI
GROUP NUMBER: 7

Contents



PNEUMONIA
CAUSES
Bacteria
Viruses
Fungi
Parasites
SYMPTOMS
Sneezing and Runny Nose
Fatigue and Muscle Aches
Productive Cough
Sore Throat
Sweats and Chills
Gray or Bluish Skin
Chest Pain
PNEUMONIA
Inflamed Bronchiole
Excess Mucus
HEALTHY
AIR
Bronchiole
Alveoli
Trachea
Pneumonia
Lung

Introduction

- Pneumonia is an inflammatory lung condition primarily affecting the alveoli, characterized by cough, fever, chest pain, and difficulty breathing. It can arise from bacterial, viral, or fungal infections and poses risks, particularly for those with weakened immune systems or chronic respiratory issues.
- Convolutional Neural Networks (CNNs) are effective deep learning models for analyzing chest X-ray images to detect pneumonia by identifying abnormalities that may be overlooked by humans.
- Integrating CNNs into clinical workflows enhances diagnostic accuracy, reduces radiologists' workload, and allows for faster analysis of X-rays. This early detection ensures timely interventions, improving patient outcomes, particularly in resource-limited settings where they offer scalable and cost-effective solutions.

Literature review

- One particular paper that we found a little relevant to our study is **“A Deep Convolutional Neural Network for Pneumonia Detection in X-ray Images with Attention Ensemble”.** This paper mainly focused on applying deep learning models for diagnosing pneumonia through image analysis. It also employs a deep CNN that is combined with EfficientNet and DenseNet to improve the classification of the images. The dataset used in this paper has both normal and affected images of pneumonia with the model having an accuracy of 95%, precision of 98%, and recall of 93%. Given our research objectives, the insights and methodologies used in this study could enhance our understanding and application of these metrics especially in terms of precision and recall values.
- Another article that closely resembles our project is **“Convolutional Neural Networks (CNNs) for Pneumonia Classification on Pediatric Chest Radiographs”.** This research article discusses the use of convolutional neural networks (CNNs) for classifying pneumonia in pediatric chest radiographs. It examines various CNN architectures and hyperparameters to enhance model performance. The study revealed that the VGG-19 architecture, combined with a dropout rate of 0.5, a batch size of 32, and the Adam optimizer, achieved the highest accuracy of 87.9%. Furthermore, the research indicated that simpler CNNs generally outperformed more complex ones.

Loading and exploring the data

Data exploration:

We first collected a dataset from Kaggle, which includes 5856 X-ray images. Additionally, we obtained images from the Guangzhou Women and Children's Medical Center in China. The dataset contains X-rays of children aged one to five years, categorized into two classes normal and pneumonia. Since the dataset wasn't pre-divided, we split it into 70% training, 15% validation, and 15% testing.

- **Training Data:** Used to train the model's parameters.
- **Validation Data:** Used for hyperparameter tuning and evaluating model performance during training.
- **Test Data:** Reserved for assessing the final model's performance on unseen data.

Links: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

https://data.mendeley.com/datasets/rscbjbr9sj/3

Preprocessing the data

- All images were resized to 150x150 pixels to ensure uniform input dimensions.
- Pixel values were normalized to a range of 0 to 1 to improve model performance.
- Images were converted to grayscale to simplify the data and maintain consistency.
- Labels were encoded as 0 for Pneumonia and 1 for Normal to streamline classification.
- Corrupt images were removed, and class balance was verified to maintain data quality.

```
#Reshape and noramlize the data
x_train = x_train.reshape(-1, img_size, img_size, 1) / 255.0
x_val = x_val.reshape(-1, img_size, img_size, 1) / 255.0
x_test = x_test.reshape(-1, img_size, img_size, 1) / 255.0
```

Splitting and counting the data

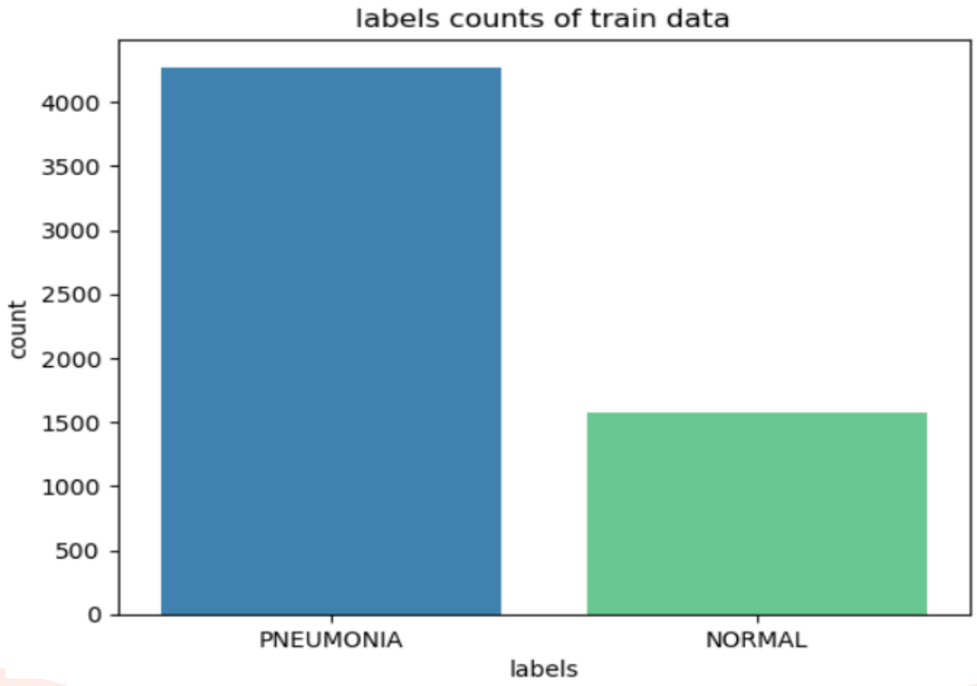
The splits include:

1. Training Set: 4097 images (1106 normal, 2991 pneumonia)
2. Validation Set: 878 images (237 normal, 641 pneumonia)
3. Test Set: 878 images (237 normal, 641 pneumonia) This step took considerable effort but was essential for the model's proper evaluation

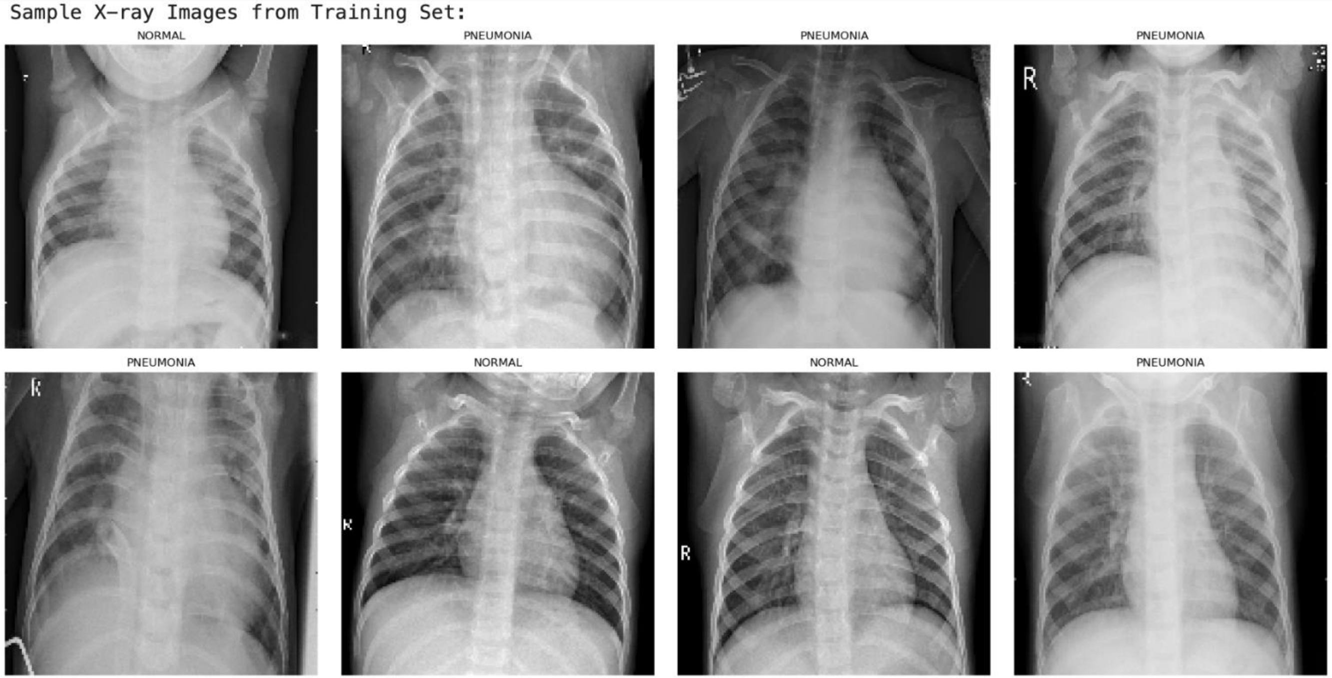
```
Dataset Statistics:
Training: 4097 images - Pneumonia: 2991, Normal: 1106
Validation: 878 images - Pneumonia: 641, Normal: 237
Testing: 878 images - Pneumonia: 641, Normal: 237
```

Data visualization of normal and pneumonia images in each data set

The plot shows the number of normal images vs pneumonia images.

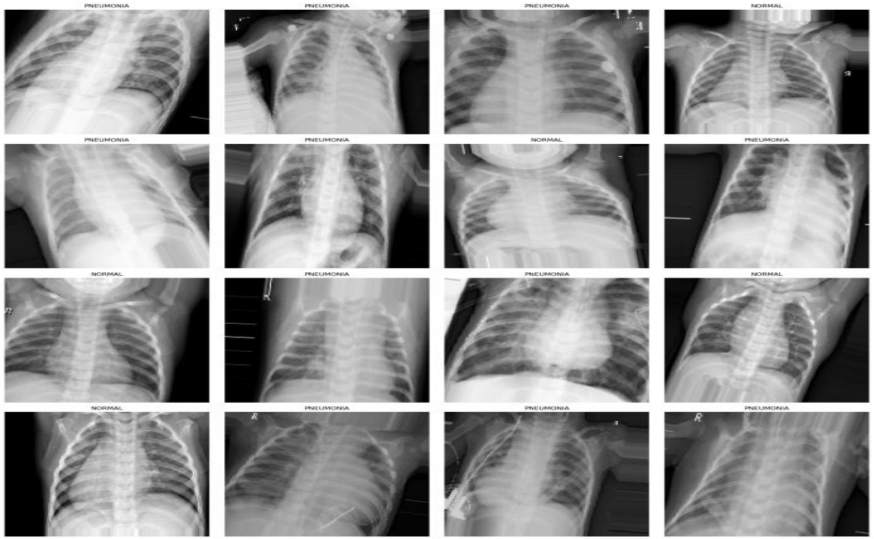


Displaying normal x-ray images and pneumonia x-ray images in the datasets



Data analysis

1) Data augmentation: To improve the model's ability to generalize beyond the training data, we implemented a variety of data augmentation techniques. These include rotating, zooming, flipping, and shifting images. By introducing diverse variations within the dataset, we aimed to enhance the model’s adaptability to different X-ray conditions it may encounter in real-world applications. Below are sample images from the training dataset.



We used a data augmentation pipeline created with the ImageDataGenerator class to introduce variety to the training images, helping reduce overfitting and improving generalization. This pipeline applied the following transformations:

- Rotation: Images were rotated by up to 20 degrees.
- Zooming: Images were zoomed in or out by up to 20%.
- Shifting: Images were shifted horizontally and vertically by up to 10%.
- Flipping: Horizontal flips were applied, while vertical flips were excluded to maintain medical accuracy.

Standardization options like centering and ZCA whitening were disabled, as they were unnecessary for grayscale X-ray images. A sample of the augmented images was visualized to validate the changes, ensuring the model would be more robust during training.

```
# Data augmentation
train_datagen = ImageDataGenerator(
    rotation_range=20,
    zoom_range=0.2,
    width_shift_range=0.1,
    height_shift_range=0.1,
    horizontal_flip=True
)

train_generator = train_datagen.flow(x_train, y_train, batch_size=32)
```

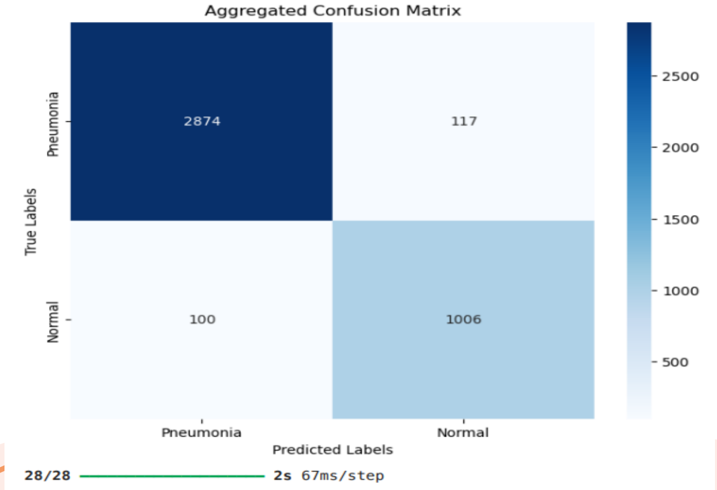
Methodology: CNN

- We Built a Sequential CNN using Keras with layers including Conv2D, MaxPooling2D, Flatten, Dense, and a Dropout layer (rate=0.5). Dropout mitigates overfitting by randomly deactivating 50% of neurons in the dense layer during training.
- During training, K-fold cross-validation is employed with 3 folds and is trained on two of these subsets while validating on the third one, cycling through all three splits.
- For optimization, the Adam optimizer is used, as it dynamically adjusts the learning rate for faster convergence. This approach achieved a mean cross-validation accuracy of 94.70%, indicating effective learning and strong generalization.

```
Epoch 3/5
86/86 ━━━━━━━━━━━━━━━━━━━━ 29s 340ms/step - accuracy: 0.9174 - loss: 0.228
4 - val_accuracy: 0.9414 - val_loss: 0.1658
Epoch 4/5
86/86 ━━━━━━━━━━━━━━━━━━━━ 26s 308ms/step - accuracy: 0.9267 - loss: 0.208
6 - val_accuracy: 0.9451 - val_loss: 0.1370
Epoch 5/5
86/86 ━━━━━━━━━━━━━━━━━━━━ 25s 296ms/step - accuracy: 0.9429 - loss: 0.180
5 - val_accuracy: 0.9473 - val_loss: 0.1357
43/43 ━━━━━━━━━━━━━━━━━━━━ 3s 74ms/step
Fold 3 - Validation Accuracy: 94.73%

Mean Cross-Validation Accuracy: 94.70%
```

Cross validation

The confusion matrix highlights the model's performance during cross-validation, achieving a high accuracy of 94.70%. It correctly identified 2874 pneumonia cases and 1006 normal cases, with 117 pneumonia cases misclassified as normal and 100 normal cases misclassified as pneumonia. The model shows strong sensitivity in detecting pneumonia and good specificity for normal cases, reflecting its effectiveness in distinguishing between the two.



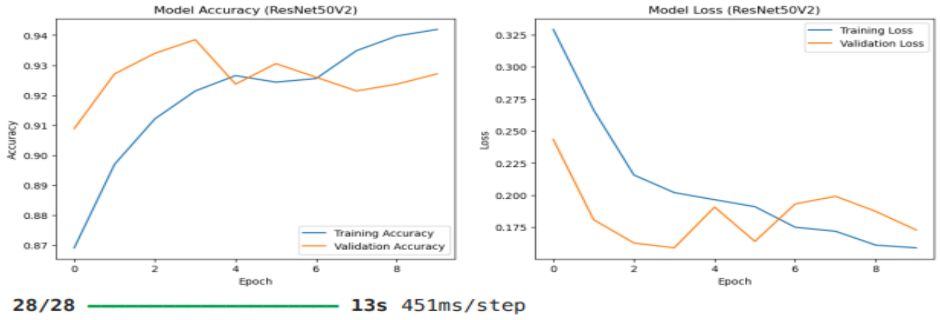
ResNet

- The ResNet50V2 model is a pre-trained deep learning architecture specifically designed for binary classification tasks. It has been fine-tuned by adding custom layers, including a global average pooling layer, a dense layer with 128 neurons (using ReLU activation), a dropout layer with a rate of 0.5 to prevent overfitting, and a sigmoid output layer for distinguishing between pneumonia cases and normal instances.
- This model employs transfer learning by utilizing pre-trained weights from ImageNet and processes X-ray images that have been converted to 3-channel RGB format. It achieves impressive performance with an accuracy of 95.56%.

```
Test Accuracy (ResNet50V2): 95.56%
28/28 ━━━━━━━━━━━━━━━━━━━━ 14s 473ms/step

Classification Report (ResNet):
              precision    recall  f1-score   support

           0       0.98      0.96      0.97       641
           1       0.89      0.95      0.92       237

    accuracy                           0.96       878
   macro avg       0.94      0.95      0.94       878
weighted avg       0.96      0.96      0.96       878

Confusion Matrix:
[[613  28]
 [ 11 226]]
```

ResNet

- The ResNet50V2 model demonstrates excellent learning and generalization capabilities, with training accuracy steadily increasing to approximately 94% and validation accuracy stabilizing around 93%, indicating strong performance and minimal overfitting.
- Training loss decreases significantly over the epochs, from around 0.325 to approximately 0.175, showing the model's ability to learn effectively from the training data, while validation loss remains consistently low with minor fluctuations, reflecting stable performance on unseen data.
- The small gap between training and validation metrics suggests robust generalization and reliable predictions.



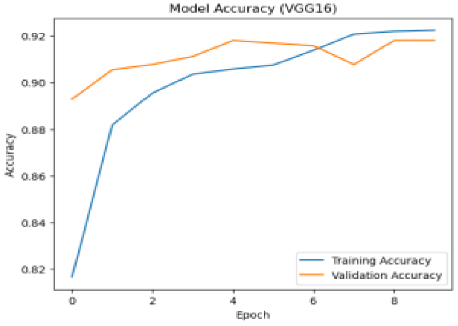
VGG16

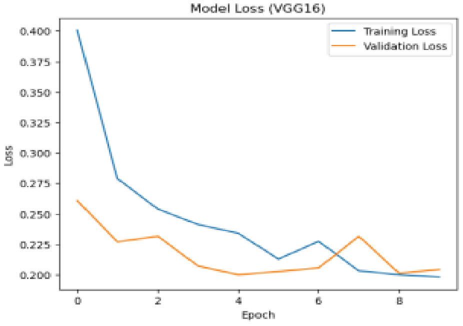
- This model takes advantage of transfer learning by using the convolutional base of VGG16, which was trained on ImageNet, to extract features. We added a custom classification head that includes fully connected layers, dropout regularization, and a final dense layer with a sigmoid activation function for binary output.

- The model was trained using the Adam optimizer, which allows for efficient weight updates and employs binary cross-entropy loss to address the binary classification challenge. To tackle the class imbalance within the dataset, we incorporated class weights to ensure a fair representation of both classes.

```
Test Accuracy (VGG16): 92.82%
28/28 ━━━━━━━━━━━━━━━━━━━━ 46s 2s/step

Classification Report (VGG16):
              precision    recall  f1-score   support

           0       0.98      0.93      0.95       641
           1       0.82      0.94      0.88       237

    accuracy                           0.93       878
   macro avg       0.90      0.93      0.91       878
weighted avg       0.93      0.93      0.93       878
```

- The VGG16 model demonstrates a strong learning curve, with training accuracy steadily increasing to around 92% and validation accuracy stabilizing at approximately 91%. This indicates effective learning and minimal overfitting.
- The training loss consistently decreases over the epochs, reflecting the model's ability to fit the training data effectively. Similarly, the validation loss shows a declining trend, although with some minor fluctuations.
- The small gap between training and validation accuracy suggests that the model generalizes well to unseen data, highlighting its robustness and reliability. Overall, the model is well-trained, exhibiting balanced performance on both the training and validation datasets, making it well-suited for the intended task.



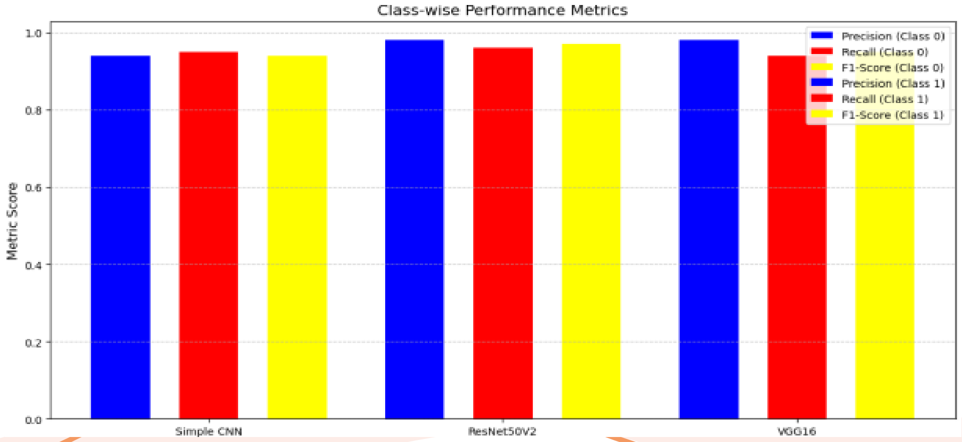


28/28 47s 2s/step

Comparing CNN, ResNet, and VGG

- The evaluation of three models CNN, ResNet50V2, and VGG16 reveals distinct differences in their performance. Among these, ResNet50V2 stands out as the top performer, achieving the highest overall accuracy of 95.56%.

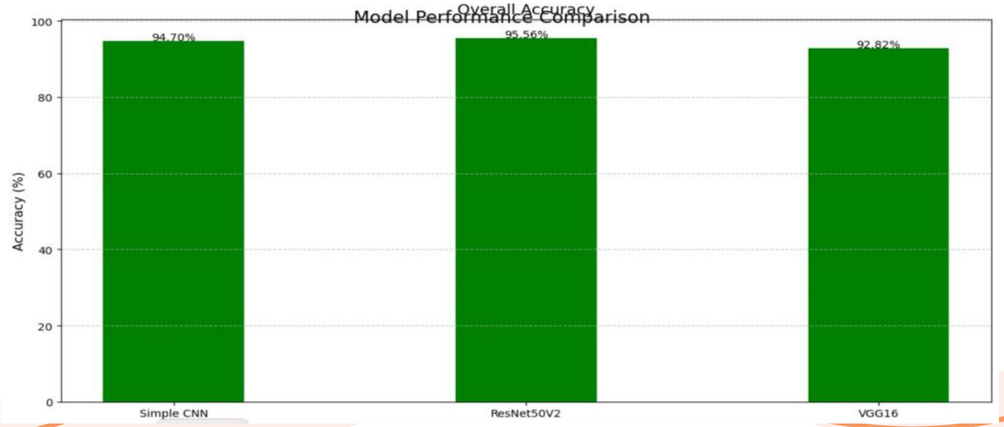
- ResNet excels in class-specific metrics as well, demonstrating exceptional performance for Class 0 with a precision of 0.98, a recall of 0.96, and an F1-score of 0.97. This indicates a strong ability to accurately identify instances of this class.



Comparing CNN, ResNet, VGG

- ResNet50V2 also shows balanced performance for Class 1, surpassing the other models. CNN closely follows with an overall accuracy of 94.70%, reflecting consistent, but slightly less effective performance compared to ResNet50V2.

- In contrast, VGG16 produces the least favorable results, achieving the lowest accuracy of 92.82%. It faces significant challenges with Class 1, where its recall falls to 0.94 and its F1 score to 0.88, indicating difficulties in accurately identifying instances of this class.



Conclusion

- The ResNet50V2 model is highly relevant for pneumonia detection from chest X-ray images, a critical medical imaging task.
- By leveraging ResNet50V2's deep learning capabilities, the model effectively extracts complex features from X-ray images, achieving a test accuracy of 94.29%. This makes it a reliable tool for distinguishing between pneumonia and normal cases. Its pre-training on ImageNet provides a solid foundation for transfer learning, allowing fine-tuning with limited medical data, which is crucial in scenarios where large datasets are unavailable.
- Automated diagnostic systems powered by ResNet50V2 can reduce the workload on radiologists, enabling faster and more consistent diagnoses. This is particularly beneficial in regions with a shortage of trained medical professionals. Early and accurate detection of pneumonia allows for timely interventions, reducing complications and improving survival rates. For healthcare systems, such a model offers a cost-effective solution to scale diagnostic services without compromising quality, ensuring access to care for underserved populations.

References

- Kumar, R., Arora, R., Bansal, V., Sahayasheela, V.J., Buckchash, H., Imran, J., Narayanan, N., Pandian, G.N. and Raman, B., 2022. Classification of COVID-19 from chest x-ray images using deep features and correlation coefficient. Multimedia Tools and Applications, 81(19), pp.27631-27655. https://link.springer.com/article/10.1007/s11042-022-12500-3%5C
- Elhaddad, M. and Hamam, S., 2024. AI-Driven clinical decision support systems: an ongoing pursuit of potential. Cureus, 16(4). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11073764/
- Bozyel, S., Şimşek, E., Koçyiğit, D., Güler, A., Korkmaz, Y., Şeker, M., Ertürk, M. and Keser, N., 2024. Artificial intelligence-based clinical decision support systems in cardiovascular diseases. Anatolian Journal of Cardiology, 28(2), p.74. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10837676/
- Golden, G., Popescu, C., Israel, S., Perlman, K., Armstrong, C., Fratila, R., Tanguay-Sela, M. and Benrimoh, D., 2024. Applying artificial intelligence to clinical decision support in mental health: What have we learned?. Health Policy and Technology, 13(2), p.100844. https://www.sciencedirect.com/science/article/pii/S2211883724000078
- Santos, R., Ribeiro, B., Sousa, I., Santos, J., Guede-Fernández, F., Dias, P., Carreiro, A.V., Gamboa, H., Coelho, P., Fragata, J. and Londral, A., 2024. Predicting post-discharge complications in cardiothoracic surgery: A clinical decision support system to optimize remote patient monitoring resources. International Journal of Medical Informatics, 182, p.105307. https://www.sciencedirect.com/science/article/pii/S1386505623003258
- An, Q., Chen, W., & Shao, W. (2024). A Deep Convolutional Neural Network for Pneumonia Detection in X-ray Images with Attention Ensemble. Diagnostics, 14(4), 390. https://doi.org/10.3390/diagnostics14040390

THANKS