Medical Imaging Diagnosis Using EfficientNetV2

Jungyu Lee

Matheus Vinicius Ferreira Figueiredo Teixeira

Pratheepan Gunaratnam

COMP-258: Neural Networks

Abstract

This paper explores

Keywords: Neural Network.

Medical Imaging Diagnosis Using EfficientNetV2

Medical imaging plays a critical role in diagnosing a vast array of conditions, from common treatable diseases to complex disorders. However, the process is heavily reliant on the availability of skilled radiologists, which can lead to significant bottlenecks, particularly in underserved areas. Furthermore, the manual interpretation of these images is susceptible to errors, compounding the challenge of ensuring reliable and timely diagnostics. Addressing these challenges is essential for improving health outcomes and the efficiency of healthcare resource utilization.

EfficientNetV2, introduced by Tan and Le (2021), is a state-of-the-art convolutional neural network that improves upon previous architectures with its efficiency in training and inference speeds. This model’s performance is characterized by its ability to maintain high accuracy while using fewer computational resources, which makes it an attractive option for medical imaging applications where timely and accurate interpretations are critical.

What performance have you achieved? The adoption of EfficientNetV2 in medical imaging could revolutionize the field by enhancing the accuracy of diagnostics while reducing dependency on scarce radiological expertise. Preliminary tests show that this model can outperform traditional models used in medical image analysis, demonstrating enhanced capabilities in detecting anomalies with faster inference speeds, thereby expediting the diagnostic process and potentially increasing accessibility to high-quality diagnostic services.

# Related Work

(0.5 Page) Has the problem been solved in the past? Generally, the answer is Yes. Select up to 5 References of those who have solved the problem. How did they solve it? What performance did they achieve?

## EfficientNetV2: Smaller Models and Faster Training

## Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning

# Methodology

## Data Description and Source

The dataset used in this project is a subset specifically designed for the classification of pneumonia from chest X-rays, typically referred to as the "ChestXray-2017" dataset. This dataset is comprised of 5,856 X-ray images derived from pediatric patients from the Guangzhou Women and Children’s Medical Center, Guangzhou. Each image is labeled as either "Normal" or "Pneumonia," with a total distribution of 1,583 normal cases and 4,274 pneumonia cases. This binary classification setup simplifies the task, focusing solely on the presence or absence of pneumonia.

## Model Architecture

The architecture chosen for this study is EfficientNetV2S, influenced by the work done on similar problems in the field, such as Rajpurkar et al.'s CheXNet and Kermany et al.'s model. EfficientNetV2S was selected due to its scalability and efficiency, which allows for high accuracy even with limited computational resources. It’s particularly suited for the task due to its depth and width scalability, which can be finely adjusted to fit the size and complexity of the medical imaging dataset.

## Data Preprocessing

Rescaling: Each image was rescaled so that pixel values are normalized to a range of 0 to 1. This normalization helps in managing model training dynamics, improving numerical stability.

Augmentation: To increase the robustness of the model and to simulate a variety of imaging conditions, data augmentation techniques such as rotation (up to 40 degrees), width and height shift (up to 20%), shear transformations, zoom (up to 20%), and horizontal flipping were applied. These augmentations help prevent overfitting and ensure that the model generalizes well to new, unseen data.

Train-Validation-Test Split: The dataset was divided into training, validation, and test sets. The training set included 4,710 images, the validation set contained 522 images, and the test set comprised 624 images. This split was designed to provide a comprehensive evaluation of the model across different subsets of data, ensuring that the model’s performance metrics are robust and reliable. Parameter Tuning

## Parameter Tuning and Performance Tracking

Early Stopping: To avoid overfitting, an early stopping mechanism was utilized. This mechanism monitors the validation loss and stops the training process if the validation loss does not improve for five consecutive epochs, restoring the best model weights observed during training.

Optimizer and Learning Rates: The model was compiled with the Adam optimizer, starting with a learning rate of 0.001 for initial training and reduced to 1e-5 during the fine-tuning phase after the base model layers were unfrozen.

Model Evaluation: Performance was tracked using accuracy and loss metrics. These metrics were plotted after each training epoch for both training and validation sets to visualize the model’s learning progress. Model performance on the test set was critically evaluated using accuracy, precision, recall, F1-score, and ROC-AUC scores to assess its clinical applicability.

## Accuracy – Loss Plots

### Initial Training (Frozen)

A graph of a line graph

Description automatically generated with medium confidence

### Fine tunning

A graph of different colored lines

Description automatically generated with medium confidence

# Evaluation

## **Evaluation Criteria and Rationale**

The evaluation of the pneumonia detection model was centered around key performance metrics tailored to clinical significance and predictive accuracy. These include:

Accuracy: Reflects the overall effectiveness of the model across all classifications.

Precision and Recall: Particularly important in medical diagnostics where it is crucial to minimize false negatives (failing to detect pneumonia) and manage false positives (incorrectly diagnosing pneumonia).

F1-Score: Balances precision and recall and is especially useful in scenarios where class imbalance might affect the performance metrics.

ROC-AUC Score: Indicates the model's ability to discriminate between the classes at various threshold levels, which is vital for adjusting sensitivity in clinical settings.

## Focus on Reducing False Negatives

For this medical application, reducing false negatives was prioritized to ensure no cases of pneumonia go undetected, potentially leading to severe health consequences if not treated promptly. Therefore, a high recall rate is desirable, even if it results in a higher number of false positives, which, while not ideal, are less dangerous because they would lead to further diagnostic testing rather than missed treatments.

## Performance Overview

The model's performance was analyzed through its progression over two main phases: initial training and fine-tuning, followed by testing on a separate test set.

A collage of x-ray images of a child's chest

Description automatically generated

## Model Training and Validation Performance

|  |  |
| --- | --- |
| Metric | Test Set |
| Accuracy | 85.74% |
| Precision | 88% |
| Recall | 90% |
| F1-Score | 89% |
| ROC-AUC | 0.905 |

## Confusion Matrix

|  |  |  |
| --- | --- | --- |
|  | Predicted Normal | Predicted Pneumonia |
| Actual Normal | 184 (True Negative) | 50 (False Positive) |
| Actual Pneumonia | 39 (False Negative) | 351 (True Positive) |

## ROC Curve

A graph of a curve

Description automatically generated

# Limitations

(0.25 Page) Mention the drawbacks of using your model. Under what inputs you think it will generate unexpected results? Are you planning to extend the project in the future?

# Conclusion

(0.25) In brief, reiterate what you mentioned in the Introduction. What is the problem you aimed to solve? why is it important to solve? How far did you solve the problem? what is the idea you used to solve? What performance have you achieved?

References

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M. K., Pei, J., Ting, M. Y. L., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., ... Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5), 1122–1131. <https://doi.org/10.1016/j.cell.2018.02.010>

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv preprint arXiv:1711.05225. <https://arxiv.org/abs/1711.05225>

Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller models and faster training. arXiv. <https://arxiv.org/abs/2104.00298>