Machine Learning Classification Using Transfer Learning and X-Ray Images

Felipe Castano Gonzalez[1], Carlos Arteaga de Anda[2], Nadeem Bishtawi[2], Grace Kennedy[2]

[1]*Department of Electrical and Software Engineering, University of Calgary, Calgary, Alberta, Canada*
[2]*Faculty of Graduate Studies, University of Calgary, Calgary, Alberta, Canada*

***Abstract*— This project explores the use of transfer learning for multi-class classification of chest X-ray images into normal, viral pneumonia, and bacterial pneumonia. Leveraging the publicly available CoronaHack dataset, we evaluated several state-of-the-art pre-trained convolutional neural networks including ResNet, DenseNet, EfficientNet, and Xception. Each model was fine-tuned and assessed using performance metrics such as accuracy, precision, recall, F1-score, and AUC. EfficientNetV2-S achieved the highest results, with a test accuracy of 92.79% and a micro-AUC of 0.9780. Grad-CAM visualizations confirmed the model's interpretability by highlighting clinically relevant regions in the images. Hyperparameter tuning and data augmentation techniques further enhanced generalization. All necessary information are fully documented and accessible in the project repository available at this GitHub link.**

Keywords*— Machine Learning, Transfer Learning, Chest X-ray images, Image Processing.

I. Introduction

Machine learning (ML) continues to revolutionize the healthcare landscape, offering data-driven tools to enhance diagnostics, personalize treatment, and improve outcomes [1]. A key enabler of this progress is transfer learning (TL), which allows models trained on large-scale datasets to be fine-tuned for medical tasks with limited data, significantly improving performance in clinical imaging contexts [2]. In particular, ML has shown promise for the early detection of respiratory diseases, including chronic obstructive pulmonary disease and pneumonia, by analyzing patterns in chest radiographs and other clinical data [3]. Among ML strategies, TL using deep neural networks has emerged as a powerful method for identifying pulmonary abnormalities from chest X-rays, facilitating the classification of various types of pneumonia, including viral and bacterial [4]. These approaches are especially valuable in resource-constrained settings where rapid and accurate triage is essential. In this study, we propose a TL-based framework to classify chest X-ray images as normal, viral pneumonia, or bacterial pneumonia, leveraging the publicly available Coronahack Chest X-Ray dataset [5]. Our goal is to support early diagnosis and clinical decision-making through an efficient and reproducible image-based model, in line with recent advancements such as Vision Transformers for pneumonia detection [6].

II. Background

TL has emerged as a reliable method for pulmonary disease detection from chest X-ray images, particularly using the CoronaHack dataset released on Kaggle. Several pre-trained convolutional neural networks (CNNs) have been successfully adapted to this task.

DenseNet-based architectures have shown exceptional performance, with DenseNet201 achieving up to 90% accuracy when fine-tuned on CoronaHack images [7]. Xception has also been effectively employed; Jain et al. used it in a three-class classification setup and reported 97% accuracy [8]. AlexNet, known for its efficient depth and width, was applied by Asif et al., achieving approximately 95.66% accuracy in distinguishing COVID-19, pneumonia, and normal cases [9].

EfficientNet-B0, due to its scalable design, has been applied successfully. M. abdullah et al. reported high precision (96%) in binary classification on enhanced CoronaHack data [10]. MobileNet was integrated into an ensemble network by Wang et al., yielding 85% accuracy for three-class classification [11].

ResNet50 has also been adapted using ImageNet weights and additional pretraining on CheXNet. Bolhassani et al. demonstrated its robustness under class imbalance using CoronaHack data [12]. VGG19, though older, continues to perform competitively. Apostolopoulos et al. reported 98.75% accuracy in binary classification and 93.5% in a three-class setup [13].

These studies support the viability of TL with various CNN architectures on the CoronaHack dataset for automated pulmonary disease classification.

III. Trasnfer Learning Testing Models

To identify the most suitable architecture for classifying chest X-ray images into normal, viral pneumonia, and bacterial pneumonia, we experimented with several pre-trained CNNs known for their strong performance in medical imaging.

We selected ResNet101 and ResNet152 due to their deep residual architectures, which help mitigate vanishing gradient problems and have been previously successful in chest radiograph analysis. EfficientNetV2-S, EfficientNetV2-M, and EfficientNetB0 were included for their compound scaling strategies, which balance model depth, width, and resolution to optimize both accuracy and computational efficiency. DenseNet201 and DenseNet121 were evaluated because of their dense connectivity, which promotes feature reuse, particularly useful for subtle visual differences in pathology. We tested Xception based on its top-tier performance on the Keras leaderboard and reports from Kaggle users achieving up to 97% accuracy on similar tasks. InceptionV3 was chosen following evidence of robust performance in binary pneumonia classification. To assess speed-accuracy trade-offs, NASNetMobile was included, allowing faster experimentation cycles. Finally, VGG19 was tested as a benchmark model due to its widespread use and historically strong results in chest X-ray classification tasks.

Each model was selected to explore a diverse set of architectural strategies and compare their effectiveness in TL applied to the CoronaHack dataset (*Table I*).

Model	Accuracy
EfficientNet_V2S	0.9279
ResNet101	0.8846
Densenet201	0.8782
EfficientNet_V2M	0.8766
Xception	0.8622

Dense121	0.8654
ResNet152(v1)	0.8333
EfficientNetB0	0.8397
InceptionV3	0.8189
VGG19	0.8141
ResNet152(v2)	0.7332

Table I. Tested models for medical image classification

IV. TRASNFER LEARNING APPROACH

All pre-trained models were initially frozen to preserve their learned representations. A custom classification head comprising Dropout, Dense layers with ReLU activation, and a final softmax layer was added to enable three-class prediction. After training the head, upper layers of the backbone were selectively unfrozen for fine-tuning at a reduced learning rate. Architectures like DenseNet121 and Xception included additional normalization layers and alternative activations such as swish. For InceptionV3, EfficientNetB0, and NASNetMobile, three Dropout layers were used for regularization. Class weights were applied during training to address class imbalance.

V. SELECTED BEST MODEL

The selected best model was EfficientNetV2-S, chosen based on its superior performance across multiple evaluation metrics on the test set. It achieved the highest validation accuracy (0.9279), along with strong micro-averaged scores for AUC (0.9780), F1-score (0.9279), precision (0.9279), and recall (0.9279), outperforming all other architectures tested. The model's reliability was further supported by visual explanations using Grad-CAM, which confirmed that its predictions were grounded in clinically relevant regions of the chest X-rays. This combination of quantitative performance and qualitative interpretability led to its selection as the final model.

VI. HYPERPARAMETER TUNING

Hyperparameter tuning was conducted to optimize model generalization and stability during training. We focused on three key parameters: learning rate, dropout rate, and hidden layer size. A combination of manual search and experimental validation was employed to identify values that yielded high performance without overfitting. The final configuration used for the EfficientNetV2-S model included a learning rate of 6.24×10^{-5}, a dropout rate of 0.411, and a hidden layer size of 256 units. These values were selected after iterative testing, where we monitored validation accuracy and loss across multiple epochs.

The optimizer AdamW was selected due to its improved handling of weight decay, which contributes to better convergence and regularization. Training was carried out for 20 epochs with a batch size of 32, and all runs were seeded for reproducibility using fixed seeds for NumPy, PyTorch, and CUDA backends.

In addition, data augmentation techniques such as random cropping, horizontal flipping, rotation, and color jitter were incorporated into the training pipeline to enhance generalization and robustness. These augmentations were especially important given the modest size and slight imbalance of the CoronaHack dataset.

VII. TEAM MEMBER'S CONTRIBUTIONS

To ensure maximum variation in model experimentation and improve our chances of identifying the best-performing architecture, each team member independently implemented and tested different pre-trained models. Throughout the project, we maintained regular communication, shared findings, and collaborated to interpret results and refine model performance based on collective insights.

Nadeem developed the modular training pipeline for EfficientNet, ResNet, and DenseNet architectures. He also implemented the classify_group_X.py script, which integrated Grad-CAM visualizations for model explainability and supported final test-time evaluation. **Carlos** built the training pipeline for DenseNet121 and Xception, completing the entire workflow including the classification of five test images and evaluation of Grad-CAM outputs. **Grace** implemented pipelines for InceptionV3, EfficientNetB0, and NASNetMobile, enabling comparative analysis across lightweight and high-capacity architectures. **Felipe** developed models based on VGG19 and ResNet152, informed by scientific literature to closely replicate hyperparameter configurations and network structures shown to perform well in chest X-ray classification tasks.

Although all team members achieved high-performing models with accuracies ranging between 80% and 90%, Nadeem's EfficientNetV2-S model consistently outperformed others across validation and test metrics. As a result, his model was selected for final submission. All models and output scripts were uploaded to our public repository, available at this GitHub link.

VIII. REFERENCES

[1] R. Awasthi *et al.*, "Artificial Intelligence in Healthcare: 2023 Year in Review".

[2] P. Kora *et al.*, "Transfer learning techniques for medical image analysis: A review," *Biocybern. Biomed. Eng.*, vol. 42, no. 1, pp. 79–107, Jan. 2022, doi: 10.1016/j.bbe.2021.11.004.

[3] X. Shen and H. Liu, "Using machine learning for early detection of chronic obstructive pulmonary disease: a narrative review," *Respir. Res.*, vol. 25, no. 1, p. 336, Sep. 2024, doi: 10.1186/s12931-024-02960-6.

[4] N. C. Kundur, B. C. Anil, P. M. Dhulavvagol, R. Ganiger, and B. Ramadoss, "Pneumonia Detection in Chest X-Rays using Transfer Learning and TPUs," *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 5, pp. 11878–11883, Oct. 2023, doi: 10.48084/etasr.6335.

[5] "CoronaHack -Chest X-Ray-Dataset," Kaggle. [Online].

[6] S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek, and S. Selvarajan, "Efficient pneumonia detection using Vision Transformers on chest X-rays," *Sci. Rep.*, vol. 14, no. 1, p. 2487, Jan. 2024, doi: 10.1038/s41598-024-52703-2.

[7] M. Turkoglu, "COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble," *Appl. Intell.*, vol. 51, no. 3, pp. 1213–1226, Mar. 2021, doi: 10.1007/s10489-020-01888-w.

[8] R. Jain, M. Gupta, S. Taneja, and D. J. Hemanth, "Deep learning based detection and analysis of COVID-19 on chest X-ray images," *Appl. Intell.*, vol. 51, no. 3, pp. 1690–1700, Mar. 2021, doi: 10.1007/s10489-020-01902-1.

[9] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *Appl. Intell.*, vol. 51, no. 2, pp. 854–864, Feb. 2021, doi: 10.1007/s10489-020-01829-7.

[10] M. Abdullah, F. B. Abrha, B. Kedir, and T. Tamirat Tagesse, "A Hybrid Deep Learning CNN model for COVID-19 detection from chest X-rays," *Heliyon*, vol. 10, no. 5, p. e26938, Mar. 2024, doi: 10.1016/j.heliyon.2024.e26938.

[11] M. Chakraborty, S. V. Dhavale, and J. Ingole, "Corona-Nidaan: lightweight deep convolutional neural network for chest X-Ray based COVID-19 infection detection," *Appl. Intell.*, vol. 51, no. 5, pp. 3026–3043, May 2021, doi: 10.1007/s10489-020-01978-9.

[12] Z. Mousavi, N. Shahini, S. Sheykhivand, S. Mojtahedi, and A. Arshadi, "COVID-19 detection using chest X-ray images based on a developed deep neural network," *SLAS Technol.*, vol. 27, no. 1, pp. 63–75, Feb. 2022, doi: 10.1016/j.slast.2021.10.011.

[13] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, Jun. 2020, doi: 10.1007/s13246-020-00865-4.