# Exploring Transfer Learning on Pneumonia Detection Using MobileNetV2 Model and the Future of Transfer Learning on Medical Images

Zhenghui Chen and Malhar Kute

## Abstract

The capability of neural networks to perform transfer learning is a crucial element of what makes deep learning so powerful. Being able to pre-train and fine-tune a model accelerates progress in the AI field. This paper utilizes Google's MobileNetV2 [1], which is trained on the ImageNet database [2], to make predictions about X-ray images of pneumonia and tell us if a given image is bacterial or viral pneumonia, or free of pneumonia. We performed fine-tuning on the top 14 layers of the base model, resulting in 1,043,843 trainable parameters out of 2,261,827 total. We ultimately obtained a test accuracy of 91.4% on our test set, consisting of pneumonia X-ray images from two different datasets, which greatly exceeds human performance. Our results show that transfer learning is a powerful tool that can be applied to image-based medical diagnosis, and can potentially generalize to other types of diseases and conditions found in X-ray images (fractures, infections, cancer, etc.).

## Introduction

Pneumonia is one of the world's leading causes of death for young children, with around 18% of deaths in children under 5 years old attributed to this disease. In addition, 2.5 million people died from pneumonia in 2019 with ⅔ of those victims being older than 5 [3]. Treatment of this disease varies depending on what type of pneumonia it is - viral or bacterial. We want to design a model that will be able to detect pneumonia from chest X-ray images, a common diagnostic tool for pneumonia, and classify them into either viral, bacterial, or normal (non-pneumatic) to assist clinicians in rapidly and accurately deciding the next steps for treatment. To fulfill this, our model takes in chest X-ray images and outputs whether the image contains bacterial pneumonia, viral pneumonia, or no pneumonia. Additionally, our model aims to show the effectiveness of transfer learning as a general approach to solving such problems.

## Related Work

Utilizing deep learning to classify medical images, specifically pneumonia in our case, has become more prevalent within the past decade due to advancements in ML algorithms and techniques leading to an abundant source of related literature. A study by Hofmeister et al. [4], aims to validate the accuracy of deep-learning diagnosis of pneumonia against emergency clinicians and certified radiologists. They employed transfer learning using the EfficientNet-B4 architecture, pre-trained on the ImageNet database, and achieved a model that was more accurate than diagnoses from emergency clinicians and on par with certified radiologists (63.5% accuracy for emergency clinicians, 72.5% for certified radiologists, and 71.8% for the model). Irvin et al. [5] used a similar approach with the DenseNet121 architecture. Both studies use gradient-weighted class activation mapping, a technique that essentially finds the region where the model "looks" to make their decision. This technology can assist radiologists, by helping point out areas to look at that could make diagnoses more efficient and accurate. A third study by Gabruseva et al. [6] uses the RetinaNet model as their base and found that data augmentation (rotations up to 6 degrees, shifts, scales, shears, horizontal flips, blurring, etc.) helped improve their accuracy, yet they do not report accuracy values. We note that there does not appear to be a consensus in the literature on which existing models are best for transfer learning. We also note that existing models have somewhat low accuracies, though this is in line with human accuracy.

**Dataset**

We found two datasets [7-8] from the online site Kaggle which had multiple datasets of normal and pneumatic chest X-ray images. Between the two datasets, we obtained 15064 labeled X-ray images for bacterial and viral pneumonia, as well as non-pneumatic scans, comprising 38.3%, 29.5%, and 32.2% of the data, respectively. Our data was randomly split, with 70% for the training set, 10% for the validation set, and 20% for the test set. We opted for a larger test set, since our early attempts had very high variance, with our test accuracy lagging behind the training and validation accuracy. As such, we wanted to ensure we could obtain accurate testing accuracies, so we could properly evaluate our model to avoid overfitting. The images were preprocessed using standard TensorFlow features, which include resizing the images and rescaling the pixel values between -1 and 1, which matches the data formatting of our base model. No data augmentation was performed, as X-ray scans are generally all black and white and oriented the same way. Although studies show that X-ray scans are frequently misinterpreted by medical professionals, we make the assumption that all examples are correctly labeled.

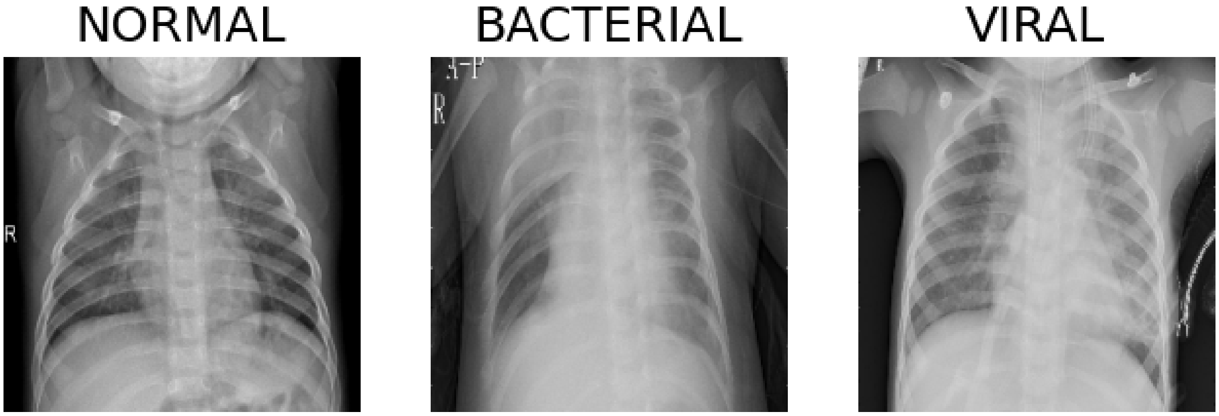

Figure 1. Examples from our dataset showing the three classes of X-Ray images.

**Methods**

To output the prediction of bacterial, viral, or no pneumonia from chest X-ray images, we used transfer learning to build a convolutional neural network (CNN) model with a softmax output layer. For our base model, we used Google's MobileNet V2 architecture [1], which is pre-trained on images from the ImageNet database [2].

A convolutional neural network is a neural network that specializes in computer vision allowing computers to make decisions based on images. At the very base level, images are made up of pixel values which is what allows CNNs to work. Instead of viewing an image as the colors represented by pixel values like humans do, CNNs look at an image for what they are, a matrix of pixel values. This allows for mathematical manipulation by the computer to find certain features of an image and ultimately build up what they "see" to make a decision. The way this is done is via "kernels" or "filters", which vary for each neuron of the neural network, and look at subsections of an image to build a feature map that stores features of an image. These feature maps can detect features that build in complexity as you get deeper into the hidden layers meaning they go from detecting features like edges to complex patterns that only humans might recognize. Additionally, as you progress through these "convolutional" layers, the feature maps go through ReLU activations and Max Pooling activations which introduce nonlinearity to the model, allowing for more complex patterns to be learned, and reduce the dimensionality of the feature map, allowing for more efficient learning. In the end, CNNs still require Dense layer(s) to make the final decision regarding an image based on all the features learned via the convolutional layers.

We decided to use transfer learning on MobileNetV2, which is also a CNN since the neurons of the convolutional layer would already be initialized to learn patterns rather than starting from scratch. This makes use of the power of transfer learning as it allows us to create a model for a specialized task without requiring large amounts of data. Since the layers of MobileNetV2 were trained on a database containing ~14 million labeled images, each neuron is a lot more "mature" compared to neurons on any model built from scratch on only 15064 labeled images. The common practice of transfer learning is to keep the base model largely unchanged, only doing small fine-tuning after a predictive layer is fit to the dataset.

The output of the base model feeds into a single 3-neuron softmax predictive layer. This layer has linear activation, but the cross-entropy loss is computed from "logits", i.e. exp(z), where z is the output value, as shown in Equations 1-2. We interpret the neuron with the largest value as the prediction from the model, where each neuron corresponds to a label of no pneumonia, bacterial pneumonia, or viral pneumonia.

$$\hat{y_k} = \text{softmax}(z_1, \ldots, z_k) = \frac{e^{z_k}}{\sum_k e^{z_k}} \qquad \text{(Eq. 1)}$$

$$\text{Loss} = -\sum_{i=1}^{k} y_i \cdot \log(\hat{y}_i) \qquad \text{(Eq. 2)}$$

**Experiments, Results, and Discussion**

Our approach to building this model largely follows the workflow outlined in the TensorFlow tutorial on transfer learning [9]. The tutorial outlines a simple method of using an existing image classification model and adapting it for a more specific purpose.

We trained our model over 30 total epochs. For the first 10, we kept the base model frozen and only trained on a single predictive output layer. We used a batch size of 32 images, the sparse cross-entropy loss function (Eq. 1-2), and the Adam optimizer with a learning rate of 0.0001. We use accuracy, defined as the percentage of predictions that match the label, as our primary metric. After these initial 10 steps, we had achieved accuracy in the training, validation, and test sets of 81.2%, 82.3%, and 83.1%, respectively. These results, while not very impressive in accuracy, indicate no signs of overfitting.

To fine-tune our model further, we unfroze the top 14 layers of the base model, resulting in 1,043,843 trainable parameters out of 2,261,827. We then train the model for 20 additional epochs, with a learning rate of 0.00001. This smaller learning rate helps ensure that we do not alter the base model too greatly. This fine-tuning step resulted in training, validation, and test accuracies of 96.2%, 91.5%, and 91.4%, respectively. The training and validation curves can be seen in Fig. 1. The training error is ~5% higher than the validation and test accuracy, indicating a slight overfit, but we decided that this difference was acceptable, given that all three accuracies were significantly higher than the human accuracy of 72.5%. This was a significant improvement over our initial attempt to train such a model, with more trainable parameters and fewer training examples, which had an over 10% gap between the training and test set accuracies.

To further evaluate how our model performs, we can study trends in the confusion matrix (Tab. 1). We note that our model has more false positives than false negatives – this is generally the direction we want this model to lean, as false negatives can have much more severe consequences in the context of medical diagnoses. We also note that our model struggles to differentiate between viral pneumonia and no pneumonia. This may be due to some visual similarity between the two categories, though it is difficult to confirm this since we are not medical professionals.

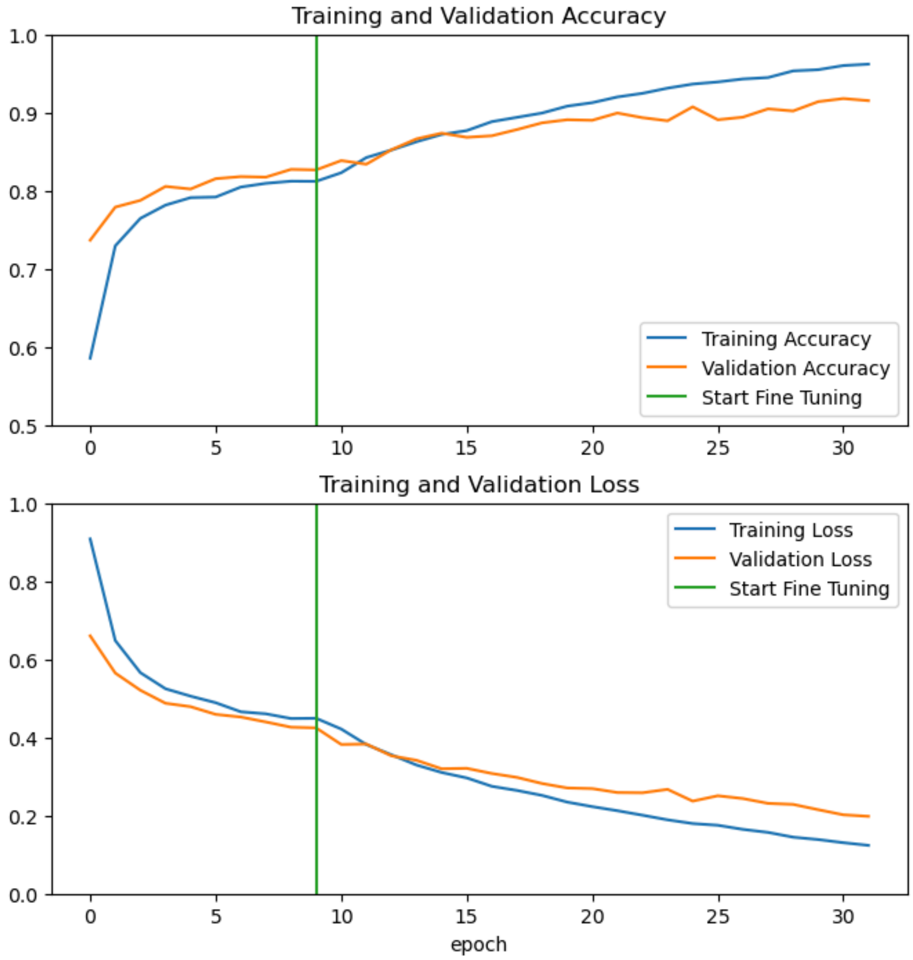

Figure 2: Training and validation curves for accuracy and loss during training. The red vertical line indicates the point where the top layers of the base model are unfrozen for fine-tuning.

Table 1. Confusion matrix of our model's performance on the test set

| | Predicts normal | Predicts bacterial | Predicts viral |
|---|---|---|---|
| Actual normal | 985 | 9 | 146 |
| Actual bacterial | 2 | 987 | 3 |
| Actual viral | 84 | 13 | 779 |

**Conclusions and Future Work**

Our results show that transfer learning can be used to adapt image classification models for the more specific and specialized task of identifying bacterial and viral pneumonia from X-ray images. Our model's accuracy requires relatively little data and training and achieves performance greatly exceeding those of medical professionals. This model could potentially serve as a valuable aid in pneumonia diagnosis. The greater implication of this report is that similar X-ray or image-based diagnosis problems can be modeled by the same transfer learning approach. The key benefit of large convolutional neural networks like MobileNet V2 is precisely its generalizability. It is unlikely that the original dataset contained many, if any, chest X-ray images, but its ability to recognize patterns and shapes allows for a highly specific model to be formed quite easily.

This is not to say that our model does not have room for improvement. As we previously noted, our model struggles to identify viral pneumonia. If we were to continue improving our model, our first approach would be to fine-tune the model on just the normal and viral pneumonia images that the model incorrectly labels, as well as collect more data to target this issue. This may exacerbate the existing slight overfitting, so regularizers may need to be employed to mitigate this. We could also evaluate the performance of different base models to see if MobileNet V2 is truly the best choice. A further extension of this project could also include applying this approach to other diseases, like cancers or infections, to show more definitively that this model is indeed widely generalizable.

## References

[1] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

[2] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*, 211-252.

[3] Popovsky, E. Y., & Florin, T. A. (2022). Community-acquired pneumonia in childhood. *Encyclopedia of Respiratory Medicine*, 119.

[4] Hofmeister, J., Garin, N., Montet, X., Scheffler, M., Platon, A., Poletti, P. A., ... & Prendki, V. (2024). Validating the accuracy of deep learning for the diagnosis of pneumonia on chest x-ray against a robust multimodal reference diagnosis: a post hoc analysis of two prospective studies. *European Radiology Experimental*, *8*(1), 20.

[5] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., ... & Ng, A. Y. (2019, July). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 590-597).

[6] Gabruseva, T., Poplavskiy, D., & Kalinin, A. (2020). Deep learning for automatic pneumonia detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 350-351).

[7] Sait, U., Lal, K. G., Prajapati, S., Bhaumik, R., Kumar, T., Sanjana, S., & Bhalla, K. (2020). Curated dataset for COVID-19 posterior-anterior chest radiography images (X-Rays). *Mendeley Data*, *1*, 1.

[8] Kermany, D., Zhang, K., & Goldbaum, M. (2018). Large dataset of labeled optical coherence tomography (oct) and chest x-ray images. *Mendeley Data*, *3*(10.17632).

[9] “Transfer learning and fine-tuning:  Tensorflow Core,” TensorFlow, https://www.tensorflow.org/tutorials/images/transfer_learning (accessed Mar. 18, 2024).