Chest X-Ray Classification: Final Report

DATA 622: Machine Learning for Health Data Science

Emma Bogner, Meghana Kompally, Serena Sun, Jason Wong (Group 4)

Selection of Pre-Trained Models

We chose a total of six pre-trained models to try for the chest X-ray classification problem: VGG16, ResNet50, InceptionV3, MobileNetV3, DenseNet121, and EfficientNetB7. Originally, these specific models were chosen because they are well documented and have been previously shown to have good performance in several image classification tasks due to their popularity (Chen et al., 2021). In addition, all of the chosen models are available through the Keras package, allowing for easy implementation. Overall, we expected the VGG16, ResNet50, and InceptionV3 models to show particularly good performance, as their success in pneumonia chest X-ray image classification has been reported in the literature (Jiang, Liu, Shao & Huang, 2021; Mujahid et al., 2022; Ikechukwu, Murali, Deepu & Shivamurthy, 2021). After some rough work involving initial training and minor adjustments, we found that InceptionV3 gave the best overall performance (i.e., test and validation accuracy) while still providing relatively quick runtimes for model training. Based on this, we further fine-tuned the InceptionV3 model to produce the final image classification model presented here.

Transfer Learning Approach

The top layers of the pre-trained InceptionV3 base model with ImageNet weights were first replaced with a global average pooling layer and a dropout layer, both of which are helpful in preventing overfitting, followed by an output layer with three nodes corresponding to our given multi-class classification task with three categories. This model architecture was mainly selected since all attempts to add additional layers resulted in a considerable decrease in initial accuracy.

For the first round of model training, all of the base model layers were frozen and only the new top layers were trained. In addition, the model was compiled with the RMSprop optimizer to take advantage of its adaptive learning rate over the long training period. After this first round of training, the initial accuracy of the model (around 33%, representing random guessing) increased to approximately 60%. An additional round of model training was then carried out—again with the base layers frozen, only training the top layers—using the SGD optimizer at a low learning rate to further fine-tune the model weights by making only small adjustments. Following this round of training, model accuracy increased to around 66%. We next attempted to unfreeze additional layers from the base InceptionV3 model (anywhere from three layers to all of the layers) and carry out additional model training; however, this always resulted in a decrease in model accuracy, with values generally falling between 50 and 60%. Based on this, we decided to only train the new top layers, as described above.

Hyperparameter Tuning

After settling on the model architecture and training process, we then fine-tuned the hyperparameters of our chosen model to further increase its performance. The hyperparameters we focused on included the dropout rate (for the dropout layer), number of neurons (for the dense layer), learning rate (for the second training step), patience and 'min delta' (for early stopping callbacks), batch size and number of epochs (for both training steps). In general, it was found that a lower dropout rate (0.1) and a greater number of neurons (1024) resulted in higher accuracy. For both training epochs, a moderate patience value (4) and a small min_delta value (0.0001) resulted in the most desirable early stopping behaviour. For the first round of training, a large number of epochs (25) in combination with the early stopping parameters above gave the best initial accuracy. For the second round of training with the SGD optimizer, a learning rate of 0.0001 and a smaller number of epochs (15) with the above early stopping parameters provided the best accuracy for the fine-tuning of weights. Finally, a moderate batch size (128) was found to give the best accuracy without affecting run times too severely. Following hyperparameter tuning, the accuracy of our model increased from approximately 66% to around 73–77%.

Group Contributions

Name	**Contributions**
Emma Bogner	Image preprocessing, initial model training (ResNet50 and InceptionV3)
Jason Wong	Initial model training (ResNet50, VGG16, MobileNetV3), explainability images
Meghana Kompally	Initial model training (EfficientNetB2, EfficientNetB3, EfficientNetB7, InceptionV3), hyperparameter tuning
Serena Sun	Model research, initial model training (VGG16, ResNet50, DenseNet121)

References

Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*, *13*(22), 4712.

Jiang, Z. P., Liu, Y. Y., Shao, Z. E., & Huang, K. W. (2021). An improved VGG16 model for pneumonia image classification. *Applied Sciences*, *11*(23), 11185.

Mujahid, M., Rustam, F., Álvarez, R., Luis Vidal Mazón, J., Díez, I. D. L. T., & Ashraf, I. (2022). Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network. *Diagnostics*, *12*(5), 1280.

Ikechukwu, A. V., Murali, S., Deepu, R., & Shivamurthy, R. C. (2021). ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. *Global Transitions Proceedings*, *2*(2), 375-381.