DATA 622/MDCH 615 Final Project
Group 3
Eva Berepiki, Michelle-Ann Goremusandu, Claire Ali, Niloofar Mirzadzare

Pre-trained Model Exploration

An initial literature review was conducted to assess possible pre-trained models that could be used for image classification with the COVID-19 chest X-ray dataset. From this search, the models selected for further testing are as follows: VGG16 (Deep architecture and simple construction), ResNet50V2 (Deep architecture with solid gradient flows), Xception (Fully separable, convolutional architecture), and DenseNet169 (Improved gradient flow with feature reusability and reduced parameters) [1]. Each of these models was selected based on frequent usage in medical image classification literature and their respective strengths in transfer learning. We found a 2021 study on single-cell conventional pap smear imaging that used and compared 10 transfer learning models that assessed each model with accuracy, precision, recall, and F1 score as metrics [2]. All of our selected models, excluding VGG19, were evaluated in this paper and scored fairly high on all metrics. As such, we included all above transfer learning models in our initial testing phase.

Transfer Learning

To address the chest X-ray classification problem, we utilized transfer learning by incorporating the pre-trained DenseNet169 model from Keras' applications module. DenseNet169 was chosen for its dense connectivity between layers, which facilitates stronger feature propagation, efficient parameter usage, and improved generalization on medical imaging datasets. We excluded DenseNet's original classification layers, which are specific to ImageNet's 1000-class output. Then, we froze all layers of the base model to retain these learned weights and avoid disrupting the robust feature hierarchy already encoded in the model. This was particularly important to prevent overfitting to our dataset. We also constructed a custom classifier consisting of a layer, followed by multiple fully connected layers with ReLU activations, batch normalization, and dropout regularization. The final layer used softmax activation to produce class probabilities across the 3 targeted categories. The intermediate layers were progressively reduced in dimensionality, and each layer included L2 regularization to penalize large weights and reduce the risk of overfitting. Batch normalization was also included after each dense layer to stabilize training and to allow for faster convergence by mitigating internal covariate shift.
By combining the representational power of DenseNet169 with a tailored classifier head, our model effectively transferred general knowledge to a highly specific medical task. The approach enabled us to achieve high classification accuracy **(87.66%)** while ensuring stable training with minimal overfitting.

Final Model

Using the same base parameters for all pre-trained models, DenseNet169 yielded the highest accuracy values and was selected for further development and tuning. The process began with the definition of key hyperparameters, including 20 training epochs and the implementation of early stopping to prevent overfitting. Five models—Regular CNN, VGG16, Xception, ResNet50V2, and DenseNet169—were initially tested on the unprocessed CIFAR-100 dataset to establish a performance baseline.

After preprocessing, the dataset was passed through the same models, resulting in the following accuracies: Regular CNN (87.66%), VGG16 (69.87%), Xception (67.31%), ResNet50V2 (39.72%), and DenseNet169 (86.76%). Despite the superior performance of the Regular CNN, the objective was to explore transfer learning using pre-trained models.

DenseNet169 was selected for fine-tuning due to its relatively high baseline accuracy. Two tuning strategies were explored. The first involved modifying the architecture by introducing batch normalization layers and setting all dropout layers to 0.2. The second strategy included the addition of extra convolutional and pooling layers. The first configuration proved more effective, achieving an accuracy of 86.1%. Input size adjustments were necessary to accommodate the saved model architecture.These enhancements led to a final accuracy of 87.6%, which was considered the best-performing configuration

Hyperparameter Tuning

Hyperparameter tuning was performed manually to create a balance between model complexity and generalization performance. One of the primary hyperparameters tuned was the dropout rate. Dropout layers were placed after each dense layer to prevent co-adaptation of neurons. We selected moderate dropout values (0.3), which provided effective regularization while preserving model learning capacity. Another hyperparameter was the learning rate. We used the Adam optimizer with a learning rate of 1e-4, which is a commonly recommended starting point for fine-tuning pre-trained models. This value was selected to ensure stable and gradual updates to the newly added layers. We also incorporated L2 regularization in all dense layers to penalize large weight magnitudes, further reducing the risk of overfitting. The regularization strength (0.0001) was chosen to create a balance between over-penalizing and under-regularizing the network weights. To optimize the number of training epochs, we implemented early stopping with a patience of 15 epochs. This allowed the model to train as long as it continued to improve on the validation set while stopping training once overfitting began. A maximum of 30 epochs was set, but early stopping ensured we only trained as long as necessary. Together, these hyperparameters and architectural decisions formed a robust pipeline for training a high-performing, generalizable classification model.

Group Contributions

Leveraging our collective experience in transfer learning, this project was both engaging and technically challenging. The team's efforts were dedicated to model development, with each member contributing to specific support tasks. Initially, Eva and Michelle focused on data preprocessing, ensuring the dataset was properly formatted for training. Concurrently, Claire and Niloofar conducted research to identify suitable models for transfer learning. Four models were shortlisted to test initial performance. From this the best model was selected, then Eva and Niloofar proceeded with training while Claire also provided insights on hyperparameter optimization and developed the model architecture. Finally, Michelle worked on the explainability to visualize the model’s decision-making process. Throughout the project, the team engaged in collaborative discussions, addressing challenges as they came up. This structured approach ensured efficient use of both time and computational resources, optimizing the overall workflow.

References

[1] GeeksforGeeks, “Top PreTrained Models for Image Classification,” *GeeksforGeeks*, Jul. 03, 2024. https://www.geeksforgeeks.org/top-pre-trained-models-for-image-classification/

[2] M. A. Mohammed, F. Abdurahman, and Y. A. Ayalew, “Single-cell conventional pap smear image classification using pre-trained deep neural network architectures,” *BMC Biomedical Engineering*, vol. 3, no. 1, Jun. 2021, doi: https://doi.org/10.1186/s42490-021-00056-6.