Topic: Develop Third Modeling Approach

Alice Tang and Yinda Chen

TF Lite Breast Cancer Detection Week 8: Third Model Approach

1. Model Approach

Based on the promising results with transfer learning and pre-trained model architectures last week, we opted to explore an EfficientNetV2B0 model for breast cancer detection in mammography images. We chose the EfficientNetV2B0 model this week since it offers a scalable and robust architecture that can balance accuracy with computational performance, making it suitable for our breast cancer detection task.

Additionally, EfficientNetV2B0 uses compound scaling, which helps with optimizing the depth, width, and resolution of the network at the same time. This architecture allows it to maintain accuracy while minimizing the computational load, which will be crucial in building our TFLite app. Similarly to last week, transfer learning was employed through loading pre-trained weights on the EfficientNet backbone, which uses previously learned features and quickens convergence for our specific dataset.

1.1 Complexity of the Modeling Approach

Our modeling approach involved a complex setup that carefully considered model expressiveness as well as computational efficiency. To improve the capabilities of our model, we decided to implement a layered approach through tuning various parameters. Ultimately, we aimed to optimize model generalizability with our parameters while minimizing overfitting, due to the delicate nature of medical imaging classification. The following are key components that added complexity to our model:

- **Custom Dropout and Trainable Layer Tuning:**
  - Multiple dropout values (0.1, 0.2, 0.3) were tested to address overfitting by reducing dependence on any single neuron during training. Naturally, this technique is crucial in medical imaging, where precise learning is key to capturing subtle patterns in data.
  - At the same time, we will test a variety of trainable layers (9, 20, 30) to balance between using pre-trained ImageNet weights and enabling the model to adapt to features specific to our dataset. We also selectively unfroze layers within EfficientNetV2B0, which encourages the model to retain foundational knowledge from its pre-training while learning the specific characteristics for breast cancer detection.
- **Adaptive Learning Rate:**
  - For this week, we used a ReduceLROnPlateau callback to dynamically reduce the learning rate once there are plateaus in validation loss. Using this strategy has helped with maintaining steady model improvement while avoiding getting stuck in local minima.

- **Evaluation and Model Selection:**
  - For each combination of dropout and trainable layers, we trained the model on our dataset and tracked the validation accuracy for each combination. This approach allowed us to compare different configurations more efficiently and to objectively select the best model parameters.
  - The results of each model were stored in a dictionary for quick reference and easy comparison of model performances, essentially helping us identify the best model configuration.

In total, running this notebook with this modeling approach takes around 2 hours. It is computationally intensive, but incredibly thorough since we can identify the optimal balance between complexity and performance.

2. Evaluating the Hyperparameters

Our hyperparameter evaluation process was designed to systematically test and compare multiple configurations, in turn, clearly exhibiting the most effective model. In order to achieve this, we evaluated the model's performance amongst a variety of combinations of dropout rates and trainable layer counts, all while iteratively adjusting the model to achieve the highest validation accuracy.

1. **Testing Procedure:** Each model configuration combined a particular dropout rate (0.1, 0.2, 0.3) with a designated number of trainable layers (9, 20, 30). For each configuration, we had to train the model from scratch, while using a standardized training dataset and validation set to ensure consistency. Though it took a couple of hours, we tested nine different model training runs which each targeting a balance between feature retention from ImageNet pre-training architecture and adaptability to our mammography dataset.
2. **Tracking Performance:** To track the performance of each model, we recorded key metrics like validation accuracy, validation loss, and training accuracy to assess each configuration on its generalization. We used validation accuracy as our main metric of determining good performance, since we believe it reflects the model's potential in a real-world medical imagery context. To mitigate the risk of overfitting, we implemented learning rate callbacks that would allow us to maximize each model's efficiency while preventing excessive training on a specific learning rate.
3. **Configuration Comparison:** After each model configuration, we stored the results in a structured dictionary. This allows us to compare and identify trends for each dropout and layer tuning configuration. Our final analysis compared each configuration's validation accuracy and drew a conclusion based on their stability across training epochs.

4. **Selection of Optimal Hyperparameters:** Ultimately, our best model was selected based on the highest validation accuracy, which considers both the generalization ability and stability of the model. By analyzing each combination's recorded metrics, we could make a decision on the specific dropout and training layer settings that yielded the best results for our breast cancer detection task.

By taking a structured approach to hyperparameter evaluation, we were able to systematically refine the model, thus allowing us to find a balance between interpretability, performance, and generalizability.

3. Model Performance Metrics

As mentioned in the previous section, these were the model performance metrics we chose for this week's model approach:

1. **Training Accuracy:** This metric indicates the percentage of correct predictions the model makes on the training dataset, throwing light on how effectively it is learning patterns from the data.
2. **Validation Accuracy:** This measures the percentage of correct predictions on the validation dataset, acting as a substitute for how the model is likely to perform on unseen data. It helps with assessing the model's ability to generalize.
3. **Training Loss:** This indicates the model's performance on the training dataset, where lower values indicate better performance. It demonstrates the difference between the model's predictions and the actual values.
4. **Validation Loss:** Similarly to training loss, this metric reveals how well the model is performing on the validation dataset. It is crucial for understanding if the model is overfitting or underfitting.

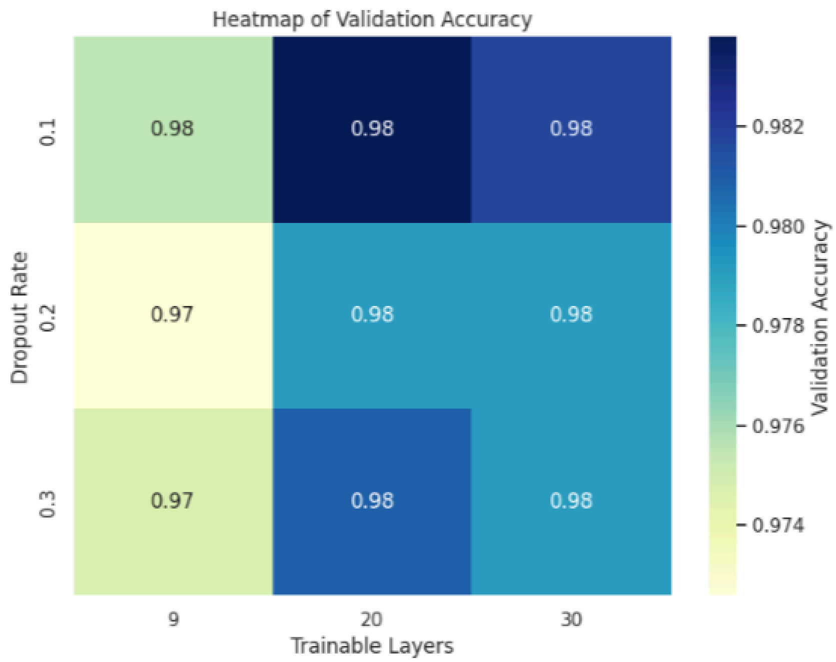
4. Analysis of the Training and Validation Metrics Across Variations

Given that we trained 9 models this week, we will primarily focus on the top three performers. Ranked from third to first, these models have configurations of 0.3 dropout rate with 20 trainable layers, a 0.1 dropout rate with 30 layers, and a 0.1 dropout rate with 20 layers. Before delving into further analysis, we'll present the final validation accuracy of each of the 9 models.

4.1 Comparison Table of Each Model

Dropout	Trainable Layers	Validation Accuracy
0.1	9	0.975538
0.1	20	0.983790
0.1	30	0.981727
0.2	9	0.972591
0.2	20	0.979075
0.2	30	0.979075
0.3	9	0.974654
0.3	20	0.980843
0.3	30	0.979075

4.2 Heatmap of Validation Accuracy for Each Model



As seen from the validation accuracies, the three selected models performed best, which showcases the benefit of fine-tuning the dropout rate and trainable layers. The model with a 0.1 dropout rate and 20 trainable layers achieved the highest validation accuracy at 0.9838, followed closely by the configurations of 0.1 dropout with 30 layers (0.9817) and 0.3 dropout with 20 layers (0.9808).

Essentially, we can infer from these results that pairing a lower dropout rate with more trainable layers improves generalization of the model, which captures important mammography features while avoiding overfitting.

4.2 Training and Validation Accuracy of Each Model

We'll analyze the table of each of these three models now.

0.3 Dropout Rate with 20 Layers

Epoch	Accuracy	Loss	Val_Accuracy	Val_Loss	Learning_Rate	Time_per_step
1	0.6252	0.889	0.7754	0.4859	1.00E-04	34ms
2	0.7483	0.5817	0.8603	0.3264	1.00E-04	26ms
3	0.7918	0.4665	0.8895	0.2661	1.00E-04	26ms
4	0.8409	0.3712	0.9128	0.2272	1.00E-04	26ms
5	0.8722	0.3071	0.9246	0.1929	1.00E-04	26ms
6	0.8865	0.2763	0.9328	0.1937	1.00E-04	27ms
7	0.9023	0.2465	0.9446	0.1431	1.00E-04	27ms
8	0.9195	0.208	0.9505	0.1455	1.00E-04	27ms
9	0.9232	0.1976	0.9567	0.1312	1.00E-04	27ms
10	0.9241	0.1902	0.9576	0.1175	1.00E-04	27ms
11	0.9322	0.1747	0.9658	0.0999	1.00E-04	27ms

12	**0.9435**	**0.1486**	**0.9664**	**0.1056**	**1.00E-04**	**27ms**
13	**0.944**	**0.1452**	**0.9688**	**0.0914**	**1.00E-04**	**27ms**
14	**0.9461**	**0.1374**	**0.9632**	**0.115**	**1.00E-04**	**27ms**
15	**0.9503**	**0.1293**	**0.9643**	**0.1102**	**5.00E-05**	**27ms**
16	**0.9581**	**0.1171**	**0.9752**	**0.085**	**5.00E-05**	**27ms**
17	**0.9658**	**0.0914**	**0.9747**	**0.0898**	**5.00E-05**	**27ms**
18	**0.9699**	**0.0844**	**0.9732**	**0.0876**	**2.50E-05**	**27ms**
19	**0.9617**	**0.105**	**0.9785**	**0.0745**	**2.50E-05**	**26ms**
20	**0.9701**	**0.0851**	**0.9803**	**0.07**	**2.50E-05**	**26ms**
21	**0.9703**	**0.0805**	**0.98**	**0.071**	**2.50E-05**	**26ms**
22	**0.9719**	**0.0774**	**0.9794**	**0.0749**	**1.25E-05**	**26ms**
23	**0.9745**	**0.0704**	**0.9785**	**0.0762**	**1.25E-05**	**26ms**

24	**0.9761**	**0.0648**	**0.9794**	**0.0719**	**6.25E-06**	**26ms**
25	**0.9781**	**0.0629**	**0.9808**	**0.0698**	**6.25E-06**	**27ms**

The 0.3 dropout rate and 20-layer model showed strong performance over the 25 epochs, with a steady improvement in both accuracy and loss.

- **Training Metrics:** Accuracy rose from 0.6252 to 0.9781, and loss declined from 0.889 to 0.0629, showing the model is learning effectively.
- **Validation Metrics:** Validation accuracy showed consistent improvement, reaching 0.9808, whereas validation loss dropped significantly and stabilized around the 19th epoch, implying strong generalization ability.
- **Learning Rate:** The adaptive reductions implemented helped with gradual improvements and reduced the risk of overfitting.

Overall, it's quite a strong and balanced model. We'll move onto analyzing the next.

0.1 Dropout Rate with 30 Layers

Epoch	Accuracy	Loss	Val_Accuracy	Val_Loss	Learning_Rate	Time_per_step
1	**0.6325**	**0.831**	**0.8199**	**0.3965**	**1.00E-04**	**34ms**
2	**0.7929**	**0.4669**	**0.8848**	**0.2699**	**1.00E-04**	**29ms**
3	**0.8463**	**0.3641**	**0.9187**	**0.2079**	**1.00E-04**	**28ms**
4	**0.8869**	**0.2796**	**0.9322**	**0.1798**	**1.00E-04**	**29ms**
5	**0.9029**	**0.2375**	**0.9372**	**0.1746**	**1.00E-04**	**29ms**

6	**0.9156**	**0.2191**	**0.952**	**0.1365**	**1.00E-04**	**29ms**
7	**0.9299**	**0.1854**	**0.9561**	**0.1252**	**1.00E-04**	**28ms**
8	**0.9319**	**0.1789**	**0.9596**	**0.1194**	**1.00E-04**	**28ms**
9	**0.9345**	**0.1679**	**0.9629**	**0.1064**	**1.00E-04**	**28ms**
10	**0.9478**	**0.1387**	**0.9649**	**0.1108**	**1.00E-04**	**28ms**
11	**0.9497**	**0.139**	**0.9655**	**0.1137**	**5.00E-05**	**28ms**
12	**0.9565**	**0.1175**	**0.9729**	**0.084**	**5.00E-05**	**28ms**
13	**0.9621**	**0.1048**	**0.9738**	**0.0792**	**5.00E-05**	**28ms**
14	**0.9642**	**0.0936**	**0.9752**	**0.0794**	**5.00E-05**	**28ms**
15	**0.9696**	**0.0856**	**0.9788**	**0.0744**	**5.00E-05**	**28ms**
16	**0.9673**	**0.0866**	**0.9776**	**0.0792**	**5.00E-05**	**28ms**
17	**0.9685**	**0.0846**	**0.9729**	**0.084**	**2.50E-05**	**28ms**

18	**0.9726**	**0.0722**	**0.9794**	**0.0683**	**2.50E-05**	**28ms**
19	**0.9792**	**0.057**	**0.9773**	**0.0746**	**2.50E-05**	**28ms**
20	**0.9782**	**0.0598**	**0.9817**	**0.0704**	**1.25E-05**	**28ms**
21	**0.9766**	**0.0664**	**0.98**	**0.0713**	**1.25E-05**	**28ms**
22	**0.9815**	**0.0511**	**0.9817**	**0.0686**	**6.25E-06**	**28ms**
23	**0.9809**	**0.0493**	**0.9805**	**0.066**	**6.25E-06**	**28ms**
24	**0.9784**	**0.058**	**0.9797**	**0.0691**	**6.25E-06**	**29ms**
25	**0.9861**	**0.0445**	**0.9811**	**0.0643**	**6.25E-06**	**29ms**

For the 0.1 dropout rate and 30-layer model, there was a steady improvement in performance across each of the 25 epochs, with significant metrics:

- **Training Metrics:** Accuracy grew from 0.6325 to 0.9861, and loss decreased from 0.831 to 0.0445, indicating effective model training.
- **Validation Metrics:** Validation accuracy reached 0.9817, with validation loss decreasing, indicating strong generalization and minimal overfitting.
- **Learning Rate:** The adaptive reductions implemented particularly helped in later epochs and facilitated an environment for the model to continuously improve.

This model configuration balanced the increased layers (30) with a low dropout rate, ultimately providing high accuracy and stability. We'll move onto our last model.

0.1 Dropout Rate with 20 Layers

Epoch	Accuracy	Loss	Val Accuracy	Val Loss	Learning Rate	Time/Step
1	0.6305	0.8211	0.7769	0.4744	1.00E-04	31ms
2	0.7797	0.4872	0.8562	0.3335	1.00E-04	26ms
3	0.8322	0.3794	0.8933	0.2574	1.00E-04	26ms
4	0.8656	0.3183	0.9089	0.2308	1.00E-04	26ms
5	0.8952	0.2635	0.9254	0.1921	1.00E-04	26ms
6	0.903	0.2381	0.9375	0.1643	1.00E-04	26ms
7	0.9124	0.2202	0.9416	0.1553	1.00E-04	26ms
8	0.9266	0.1946	0.9555	0.1269	1.00E-04	26ms
9	0.9251	0.1849	0.954	0.1336	1.00E-04	26ms
10	0.9354	0.1683	0.9661	0.1085	1.00E-04	26ms
11	0.9412	0.1548	0.9623	0.1124	1.00E-04	26ms

12	**0.9408**	**0.1484**	**0.9658**	**0.1038**	**1.00E-04**	**26ms**
13	**0.9481**	**0.1435**	**0.9643**	**0.1058**	**1.00E-04**	**26ms**
14	**0.9512**	**0.1254**	**0.972**	**0.0811**	**1.00E-04**	**26ms**
15	**0.9535**	**0.1227**	**0.9696**	**0.0863**	**1.00E-04**	**26ms**
16	**0.9575**	**0.1163**	**0.9702**	**0.0979**	**5.00E-05**	**26ms**
17	**0.9565**	**0.1132**	**0.9788**	**0.0706**	**5.00E-05**	**26ms**
18	**0.9712**	**0.0789**	**0.9776**	**0.0711**	**5.00E-05**	**26ms**
19	**0.9642**	**0.0905**	**0.9773**	**0.0732**	**2.50E-05**	**26ms**
20	**0.9709**	**0.0805**	**0.9794**	**0.0633**	**2.50E-05**	**27ms**
21	**0.9735**	**0.0775**	**0.9797**	**0.0661**	**2.50E-05**	**27ms**
22	**0.974**	**0.0685**	**0.9823**	**0.0648**	**1.25E-05**	**26ms**
23	**0.975**	**0.0624**	**0.982**	**0.062**	**1.25E-05**	**26ms**

24	**0.9772**	**0.0611**	**0.9826**	**0.0608**	**1.25E-05**	**26ms**
25	**0.9793**	**0.0536**	**0.9838**	**0.061**	**1.25E-05**	**26ms**

Lastly for the 0.1 dropout rate and 20-layer model, the performance metrics revealed promising results:

- **Training Metrics:** Accuracy grew from 0.6305 to 0.9793, and loss decreased from 0.8211 to 0.0536, showing solid training progression.
- **Validation Metrics:** Validation accuracy peaked at 0.9838, with validation loss also exhibiting a downward trend, showing strong model generalization and less overfitting.
- **Learning Rate:** The gradual reduction of the learning rate aided with model performance in later epochs.

This configuration effectively balanced the 20-layer model architecture and lower dropout rate, where it achieved high accuracy and reliable generalization.

5. Model of the Week: Best Performance

Among each of the models tested, the **0.1 Dropout Rate with 20 Layers** model shined brighter than the rest. With the strongest validation accuracy of 98.38% and low final validation loss of 0.061, the model achieved the highest accuracy and best generalization.

Here is the rationale for our choice:

- **Steady Improvement:** The model exhibited consistent progress across the 25 epochs, with the training accuracy and validation accuracy steadily increasing while keeping the validation loss low, implying minimal overfitting.
- **Optimized Complexity:** Though the model had a 20-layer structure, combined in conjunction with 0.1 dropout rate, struck an ideal balance between depth and regularization, capturing the complex patterns effectively without any sort of unnecessary complexity.
- **Learning Rate Adjustments:** The gradual reduction in learning rate enabled the model to refine its accuracy over time, where it was able to reach an optimal point without sacrificing speed or efficiency.

Overall, with high accuracy, stability, and efficiency, the **0.1 Dropout Rate with 20 Layers** model is our chosen model of the week.