Data Improvement

Extract Image Features

- **Color Histogram**: Extracts a histogram representing the distribution of colors within an image to capture color characteristics.
- **Hu Moments**: Extracts shape-based features that are invariant to image transformations like translation, rotation, and scale.
- **Haralick Texture**: Captures texture properties using the gray-level co-occurrence matrix (GLCM) to provide metrics such as contrast, dissimilarity, and homogeneity.

PCA (Principal Component Analysis) Features

- **Dimensionality Reduction**: Applied PCA to transform the image data into lower-dimensional space while retaining 95% of the variance. This helped to reduce computational complexity and highlight key data patterns for improved model learning.

Image Data Augmentation

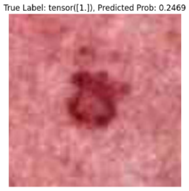
- **Random Horizontal Flip**: Introduced randomness by flipping images horizontally to help the model generalize better and reduce overfitting.
- **Random Rotation**: Applied rotation within a specified degree range to make the model more robust to different orientations.
- **Color Jitter**: Added variations in brightness, contrast, saturation, and hue to simulate different lighting conditions and enhance the diversity of the training data.

Error analysis

A total of 26 images with the true label of Class 1 were misclassified. Upon closer examination, the predicted probabilities for these Class 1 images are notably low and fall significantly below the classification threshold of 0.5. This indicates that the model is struggling to confidently identify these images as Class 1, possibly due to a lack of distinct features in these cases or a general bias towards Class 0.

True Label: tensor([1.]), Predicted Prob: 0.2860

True Label: tensor([1.]), Predicted Prob: 0.2469

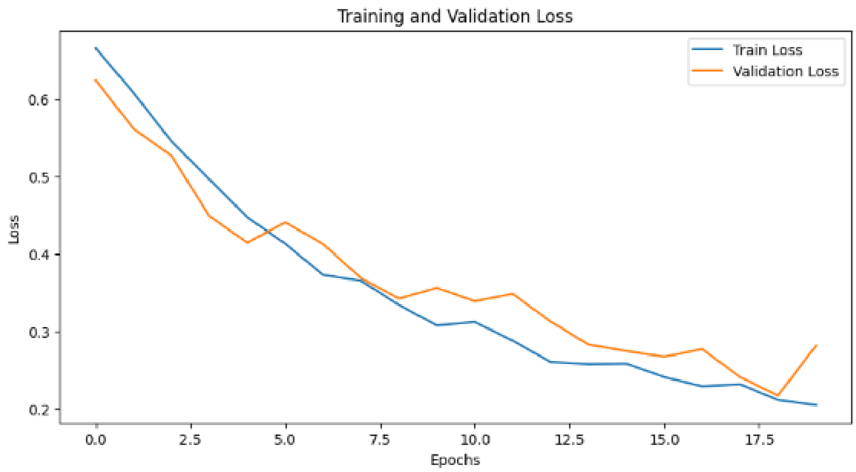


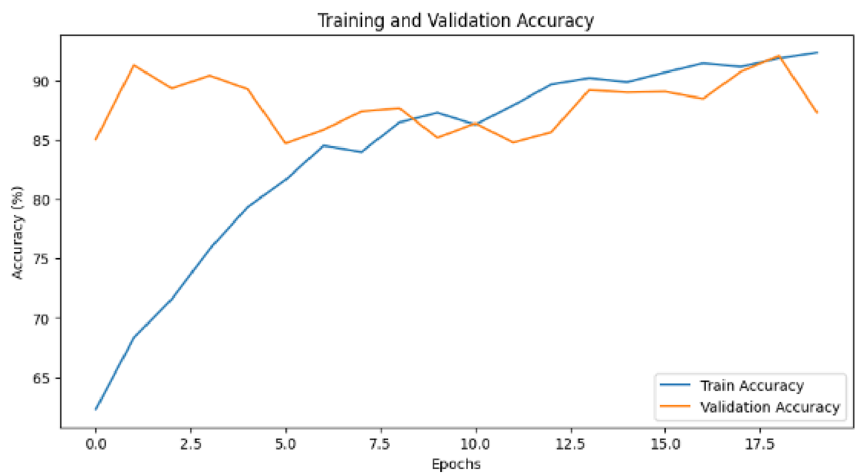
True Label: tensor([1.]), Predicted Prob: 0.3641

True Label: tensor([1.]), Predicted Prob: 0.4236

Models Performance

Model 1: Performance of Week 9 Winning Model with Validation Data





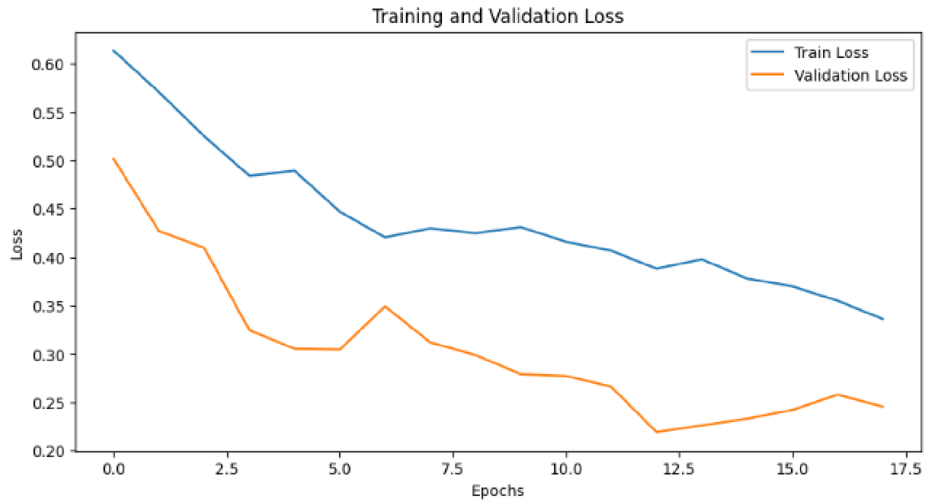
Classification Report:

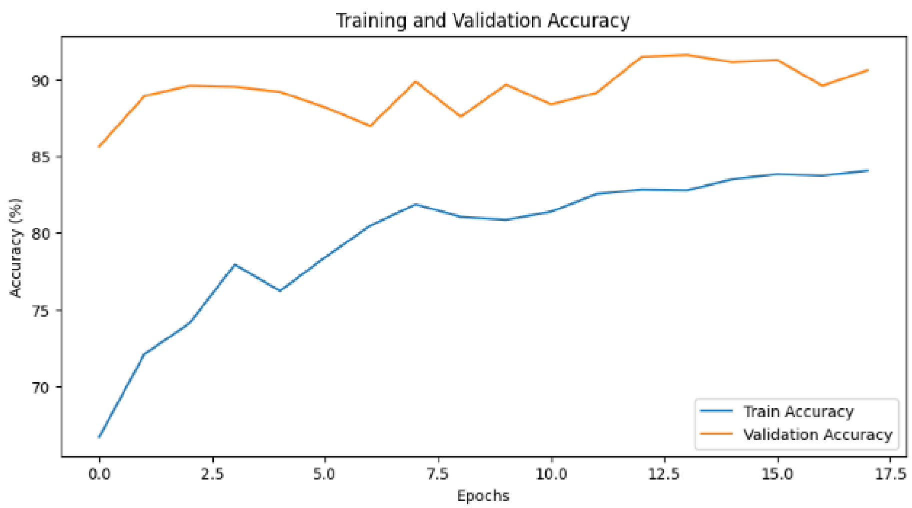
	precision	recall	f1-score	support
Class 0	0.98	0.88	0.93	1431
Class 1	0.18	0.64	0.29	59
accuracy			0.87	1490
macro avg	0.58	0.76	0.61	1490
weighted avg	0.95	0.87	0.90	1490

- **Training and Validation Loss**: The training loss steadily decreases, with validation loss generally following the trend but with some fluctuations. However, it appears that the training loss is marginally lower than the validation loss, suggesting good learning without significant overfitting.
- **Training and Validation Accuracy**: Training accuracy continues to increase over epochs, converging with validation accuracy, which also rises and stabilizes. This indicates that the model is well-generalized on both training and validation datasets.
- **Classification Report**:

- Class 0: Precision (0.98), Recall (0.88), F1-score (0.93)
- Class 1: Precision (0.18), Recall (0.64), F1-score (0.29)

Model 2: Performance of Week 9 Winning Model with Updated Validation Data






```
Classification Report:
              precision    recall  f1-score   support

     Class 0       0.98      0.92      0.95      1431
     Class 1       0.24      0.64      0.35        59

    accuracy                           0.91      1490
   macro avg       0.61      0.78      0.65      1490
weighted avg       0.95      0.91      0.93      1490
```

- **Training and Validation Loss**: The training loss also decreases but with less improvement compared to validation loss. Validation loss here remains significantly lower than training loss, which could indicate underfitting.
- **Training and Validation Accuracy**: Validation accuracy stabilizes at a high value early in the training, while training accuracy lags behind, suggesting the model may not be learning effectively on the training data.

- **Classification Report**:
    - Class 0: Precision (0.98), Recall (0.92), F1-score (0.95)
    - Class 1: Precision (0.24), Recall (0.64), F1-score (0.35)

Analysis

1. **Recall for Class 1**: Both models achieve a recall of 0.64 for Class 1, indicating they are equally capable of identifying positive instances.
2. **F1-score for Class 1**: Model 2 has a slightly higher F1-score for Class 1 (0.35 vs. 0.29 in Model 1), suggesting it better balances precision and recall for minority class.
3. Accuracy and Loss Trends:
    - Model 1 shows better training-validation alignment in both loss and accuracy, suggesting it might be less prone to overfitting.
    - Model 2, with lower validation loss than training loss, indicates possible underfitting.

Conclusion

Model with the original train data and validation data seems to perform more consistently and aligns better with expected trends for well-generalized learning, making it more reliable across datasets. Therefore, Model 1 is the recommended model to deploy due to its stable learning and closer alignment between training and validation metrics.

Test Data Performance

```
The partial auroc of the final model on the test image is 0.13625176183538829
              precision    recall  f1-score   support

     Class 0       0.98      0.92      0.95      1431
     Class 1       0.24      0.61      0.35        59

    accuracy                           0.91      1490
   macro avg       0.61      0.77      0.65      1490
weighted avg       0.95      0.91      0.93      1490
```

Based on the test dataset metrics, the model achieved a recall rate of 0.61 for Class 1 and a pAUC-aboveTPR of 0.136, both of which are higher than the validation set metrics. Given the constraints in environment and computing power, these results are satisfactory. However, performance could likely be improved with tailored data preprocessing and normalization specifically suited for EfficientNet. Despite these positive results, I believe the model still isn't optimal for this prediction task, as achieving higher performance would likely require more data for Class 1. This would help the model better capture the patterns necessary to increase recall and overall accuracy in the target class.

Validation, and test performance metrics of the winning model

	validation	test
recall	0.64	0.61
pAUC-aboveTPR	0.1034	0.136

The recall rate decreased from 0.64 on the validation set to 0.61 on the test set, while the pAUC-aboveTPR increased from 0.1034 on the validation set to 0.136 on the test set.

Insights based on validation, and training errors

Overall, Model 1 appears to be a well-balanced model with minimal overfitting and good generalization. However, the imbalance between Class 0 and Class 1 performance suggests that further improvements, such as adjusting the decision threshold or employing class-balancing techniques, might enhance the model's ability to capture the minority class more accurately. The training and validation trends indicate that the model is ready for deployment, but ongoing monitoring and potentially minor adjustments may improve minority class performance in production.