As we have demonstrated before, the results of the initial models for both text and image data are not reliable to use. Therefore, we decided to condense the target from six categories to two categories, by combining the BCC, MEL, and SCC into skin cancer, and another three into skin disease.

This change not only addresses the data imbalance problems but also provides more data for training models and improving their performances. Due to this change, I switched the activation function from SoftMax to Sigmoid, which treats each output independently, making it more appropriate for the binary classification task.

As a result, we can simply observe the performance of the binary classification outperforms the six classifications. For text data, the LGBM excels and improves its result from 43% to 93%. Random Forest and XGBoost also demonstrated strong performance with 92% accuracy. The stacking model also achieved 91% accuracy. For image data, the SAM model performed best with 38% accuracy. However, for binary classification, the VGG model took the lead with 72% accuracy. The image stacking model also showed improvement in the binary classification task compared to the six classifications.

Due to minimal performance variation among text models, a stacking model was chosen for meta-stacking to ensure consistent and reliable performance. With the obvious differences from the performance of the image mode, the best-performing model was selected for meta-stacking. At last, we stacked SAM and text stacking models for six classifications. The VGG and text stacking models for binary classification.

Finally, we summarized all the results in this table. Since we’ve discussed this before, I’ll just skip this section.