Comparative Study of Different Machine Learning Methods for Skin Lesion Classification

Leah Restad and Tiffany Nakai

Motivation

- Skin cancer is most common type of cancer
- Early detection of skin cancer is crucial
- Lots of work in imaging, many architectures developed that perform well
    - Through competitions like ImageNet
- Various papers on using machine learning techniques for skin lesion classification

Goal: Compare machine learning methods on skin lesion classification

Content adapted from 02518 slides (lecture 22)

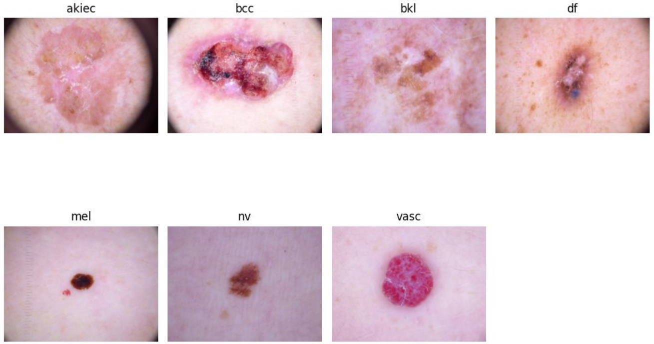
Data and Methods

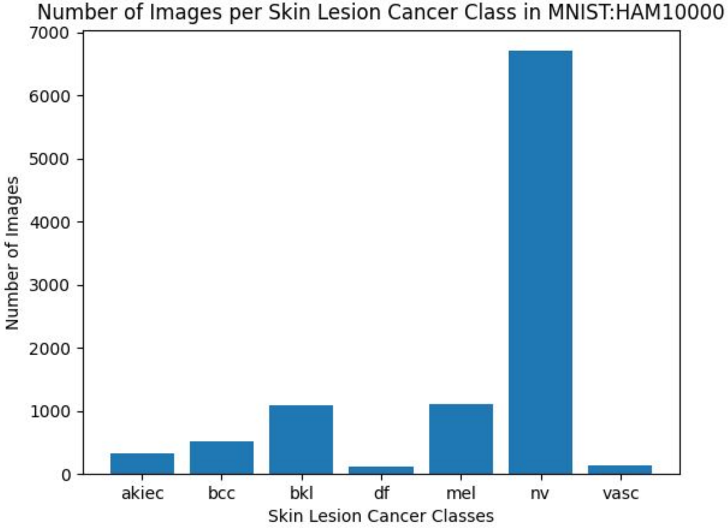
Dataset: MNIST:HAM10000

7 classes: akiec, bcc, bkl, df, mel, nv, vasc

Transfer Learning with 9 deep learning models

Used: PyTorch, sklearn, TensorFlow



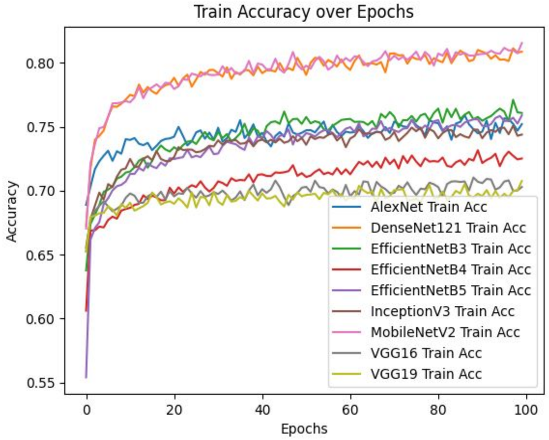


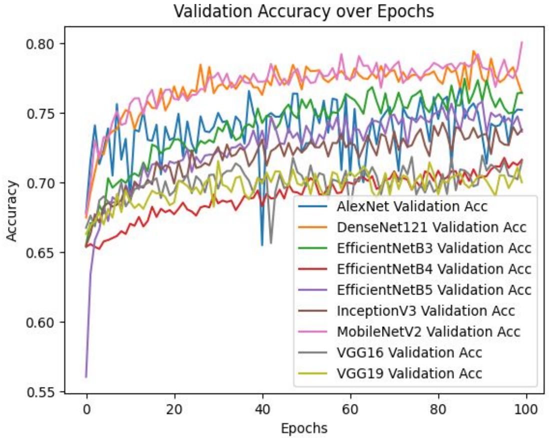
Result – Training Plots

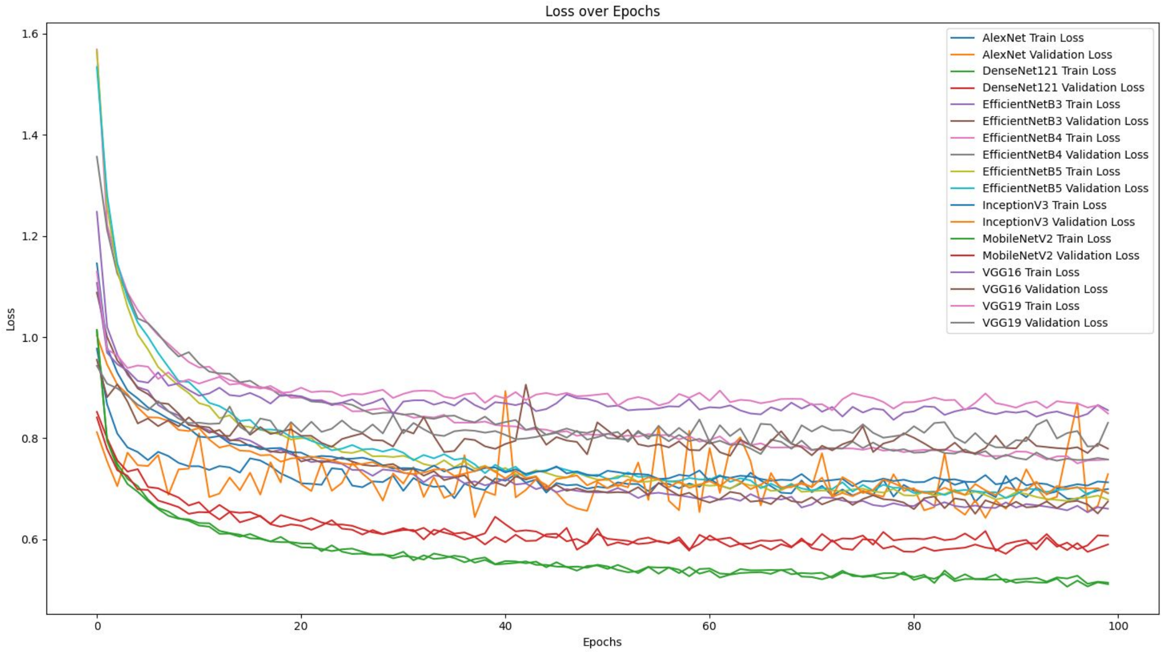
Accuracy was computed with F-score (harmonic mean of precision and recall)

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})}$$

image from google







MobileNetV2 and DenseNet have highest validation accuracy

Results – Model Comparison

Model	# Parameters Parameters info from pytorch	# Layers	Training Time	Best Validation Accuracy (F-score)	Test Accuracy (F-score)
AlexNet	61,100,840	8	444m 40s	76.9%	75.9%
DenseNet121	7,978,856	72	3789m 33s	79.4%	77.6%
EfficientNetB3	12,233,232	384 from TensorFlow - 4	2226m 35s	77.5%	76.1%
EfficientNetB4	19,341,616	474 from TensorFlow - 4	2824m 43s	71.8%	71.7%
EfficientNetB5	30,389,784	576 from TensorFlow - 4	3239m 30s	75.8%	74.6%
InceptionV3	27,161,264	48	3641m 22s	74.3%	72.9%
MobileNetV2	3,504,872	53	2457m 1s	80.1%	78.2%
VGG16	138,357,544	16	3154m 15s	71.2%	72.4%
VGG19	143,667,240	19	3421m 52s	71.5%	69.5%

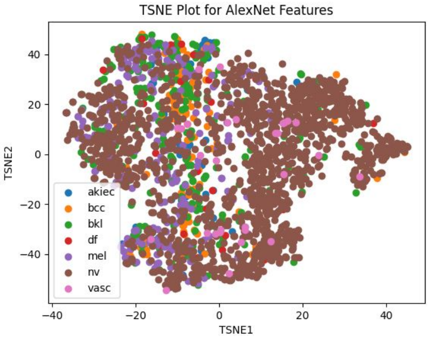
Weighted Precision/Recall accounts for class imbalance

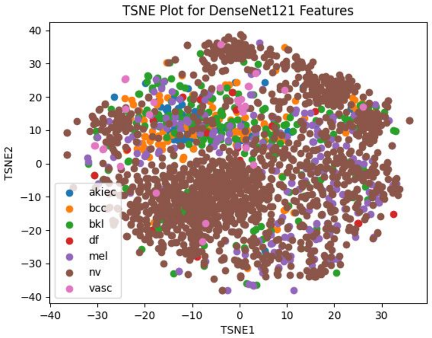
Results – Precision Recall

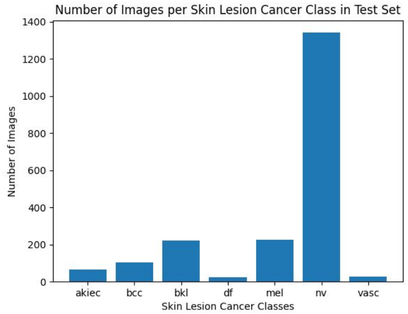
Model	Precision	Recall	Weighted Precision	Weighted Recall
AlexNet	58.2	50.9	74.5	76.5
DenseNet121	64.7	53.2	77.2	78.7
EfficientNetB3	60.9	44.7	73.6	76.1
EfficientNetB4	41.8	27.6	67.6	73.3
EfficientNetB5	66.5	41.1	72.2	75.1
InceptionV3	62.1	36.3	70.5	73.9
MobileNetV2	64.4	56.0	78.0	79.4
VGG16	52.1	37.7	69.8	73.5
VGG19	46.4	33.7	67.2	71.0

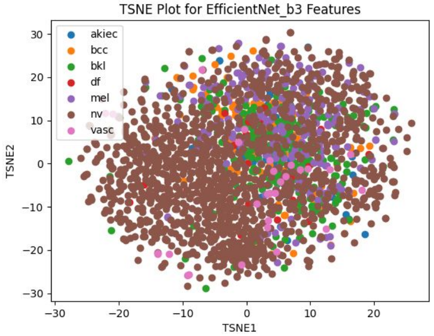
Results - t-SNE Feature Plots

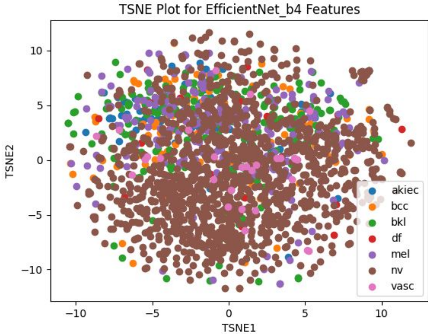
Removed last classification layer and ran t-SNE on features
MobileNetV2 and DenseNet121 had the ‘best’ grouping of classes (also had highest test accuracy)

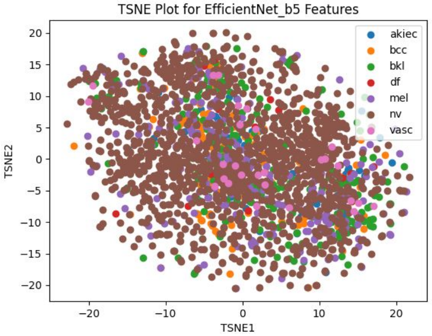


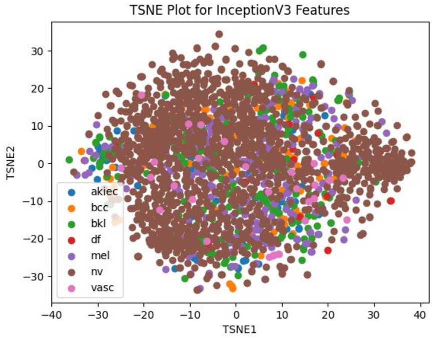


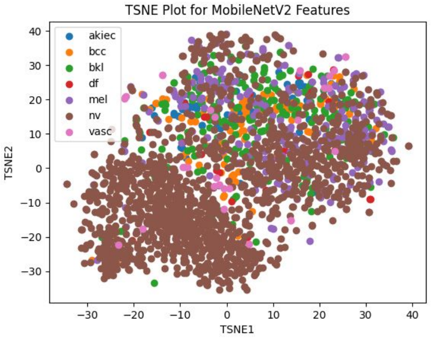


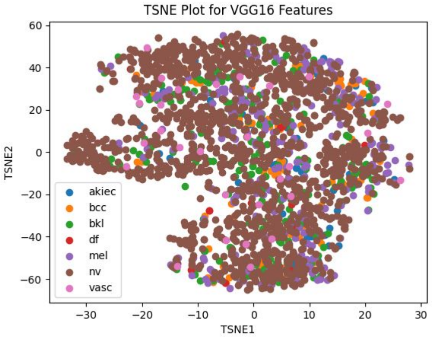


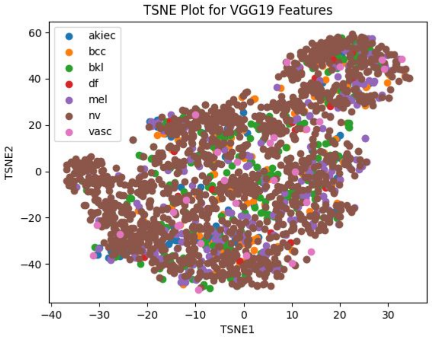






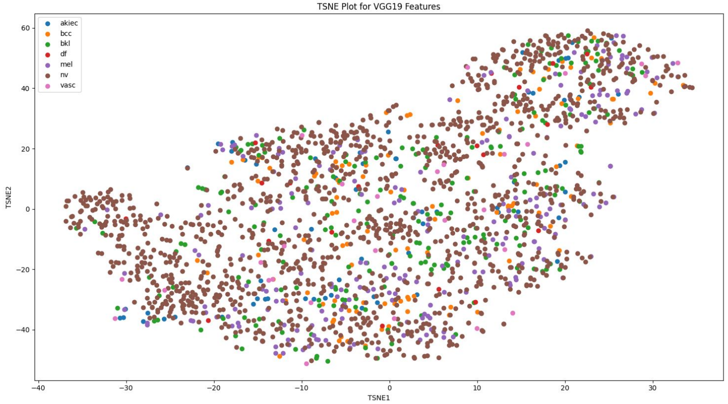


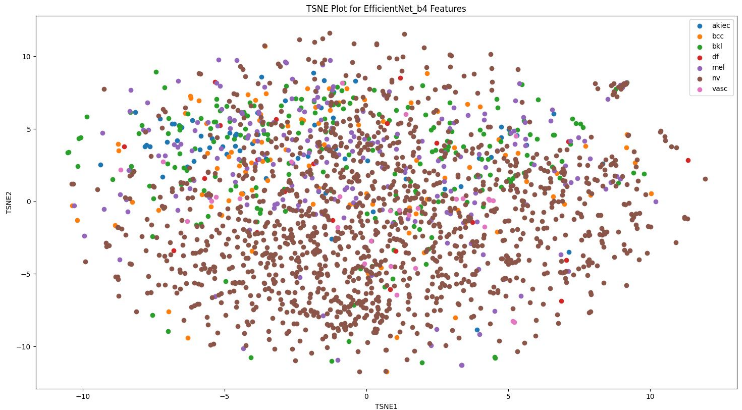




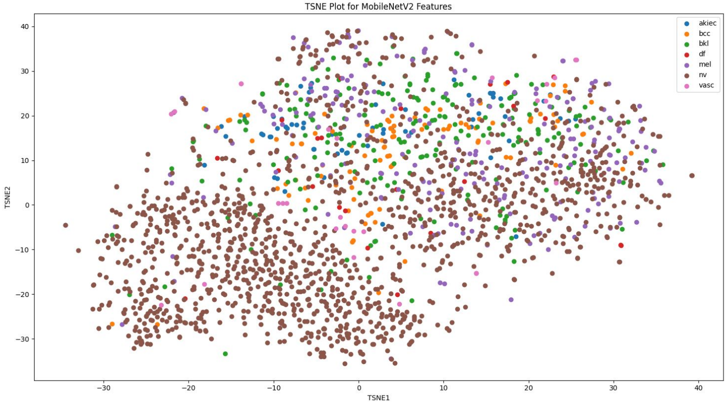
Results - t-SNE Feature Plots – closer look

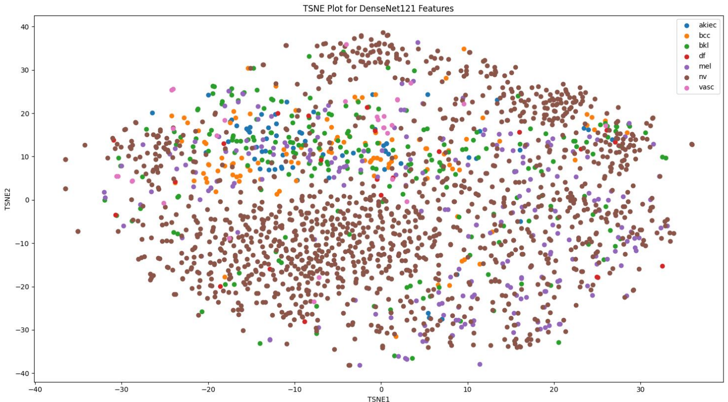
Did worse on test set





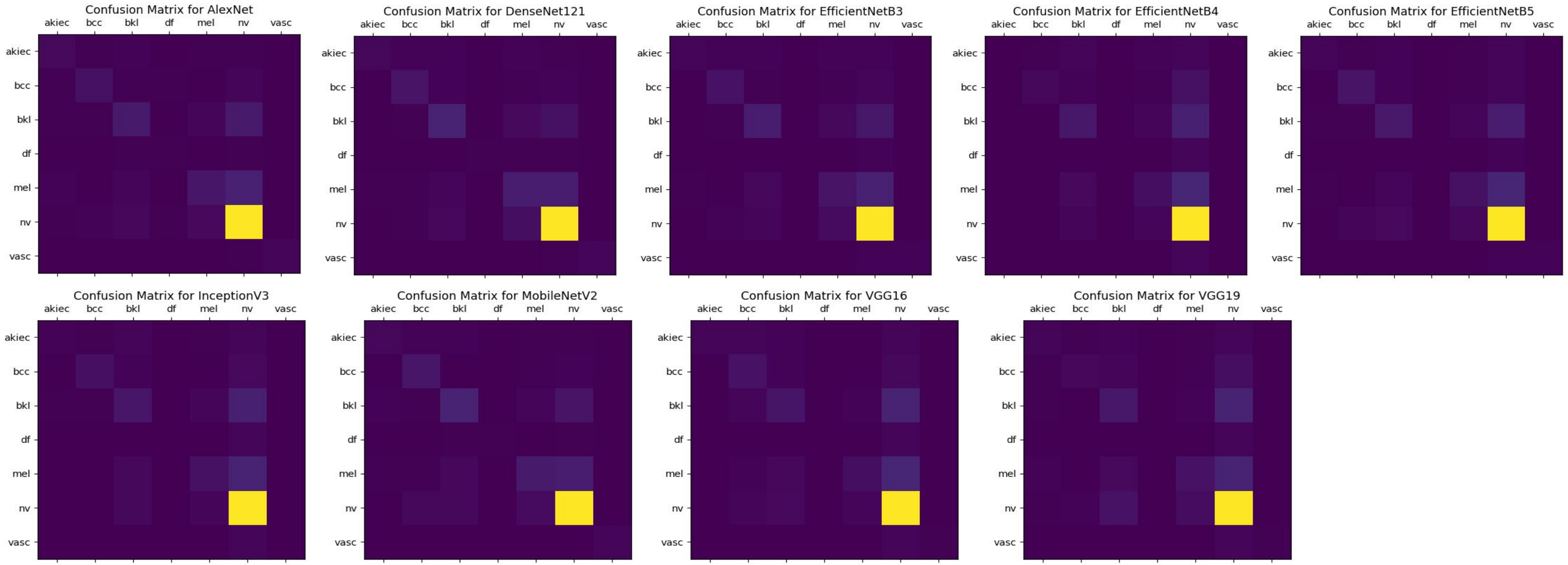
Did better on test set





Results – Confusion Matrices

Class imbalance issue: confusion matrices show that all models classify only the nv, melanocytic nevus, class well. Poorly classifies all other classes.



Scale: (low -> high), (purple -> yellow)

Conclusions + Next Steps

- MobileNetV2 and DenseNet performed best among all 9 models trained
- Overall, the models were heavily skewed towards the class with the most number of images (nv).

Would like to try:

- Try to figure out how to make models less affected by the class imbalance
- Run more models
    - had AWS issues so was limited by local computation power
- Run models with metadata incorporated
    - Models with additional metadata provided with HAM10000 dataset seemed to before better in some papers
- Run more comparative analyses

References

https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142#:~:text=If%20you%20count%20the%20total,below%20and%20the%20stem%20above. – used to figure out number of layers in efficientNet

https://github.com/lukemelas/EfficientNet-PyTorch#about-efficientnet-pytorch – says pytorch implementation is consistent with tensorflow – so used tensorflow to figure out number of layers in efficientNet

https://www.frontiersin.org/articles/10.3389/fonc.2022.931141/full

https://www.frontiersin.org/articles/10.3389/fmed.2021.644327/full

https://arxiv.org/pdf/2207.12715.pdf

https://www.sciencedirect.com/science/article/pii/S1746809421008235

https://arxiv.org/pdf/1802.00752v2.pdf

https://www.mdpi.com/2076-3417/12/12/5990

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9295694/pdf/nihms-1800396.pdf

https://arxiv.org/pdf/2010.05351.pdf

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0217293

https://ejmcm.com/article_7276.html

https://pure.strath.ac.uk/ws/portalfiles/portal/134360852/Aburaed_etal_ICSPIS_2021_Skin_cancer_classification_model_based_on_VGG19.pdf

https://www.researchgate.net/publication/325117430_Classification_of_skin_cancer_images_using_TensorFlow_and_inception_v3

https://arxiv.org/pdf/2110.04632.pdf