Insights From HAM10000 Classification Using Learned Compression

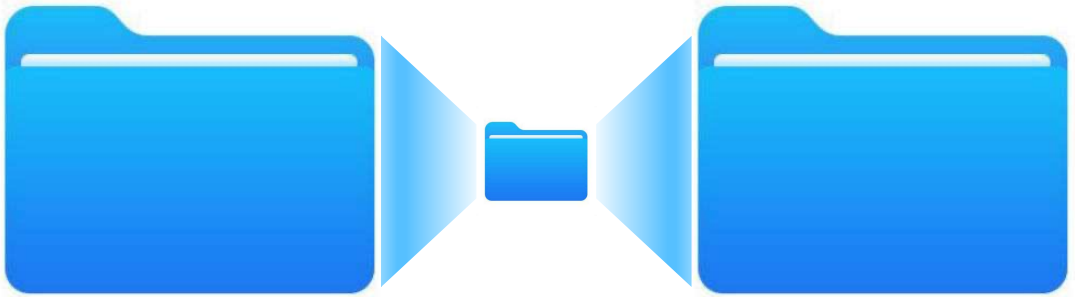
Yesh Doctor

BME548

Duke

Motivation

Imaging System



Compression/Digitization Pipeline

Melanoma

Hypothesis

- Learned compression can educate both imaging system design and image classification architectures

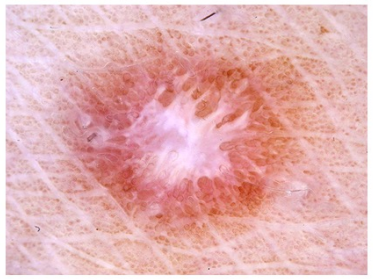
Duke

Dataset

Human Against Machine (HAM)-10000

Nevus

Dermatofibroma

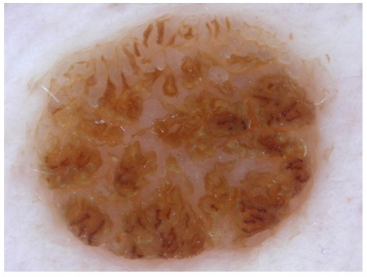


Melanoma

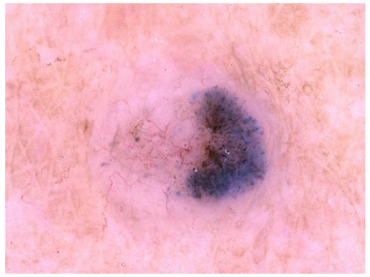
Vascular

Pigmented Bowen's

Pigmented Benign Keratoses



Basal Cell Carcinoma



10015 Dermatological Images

Duke

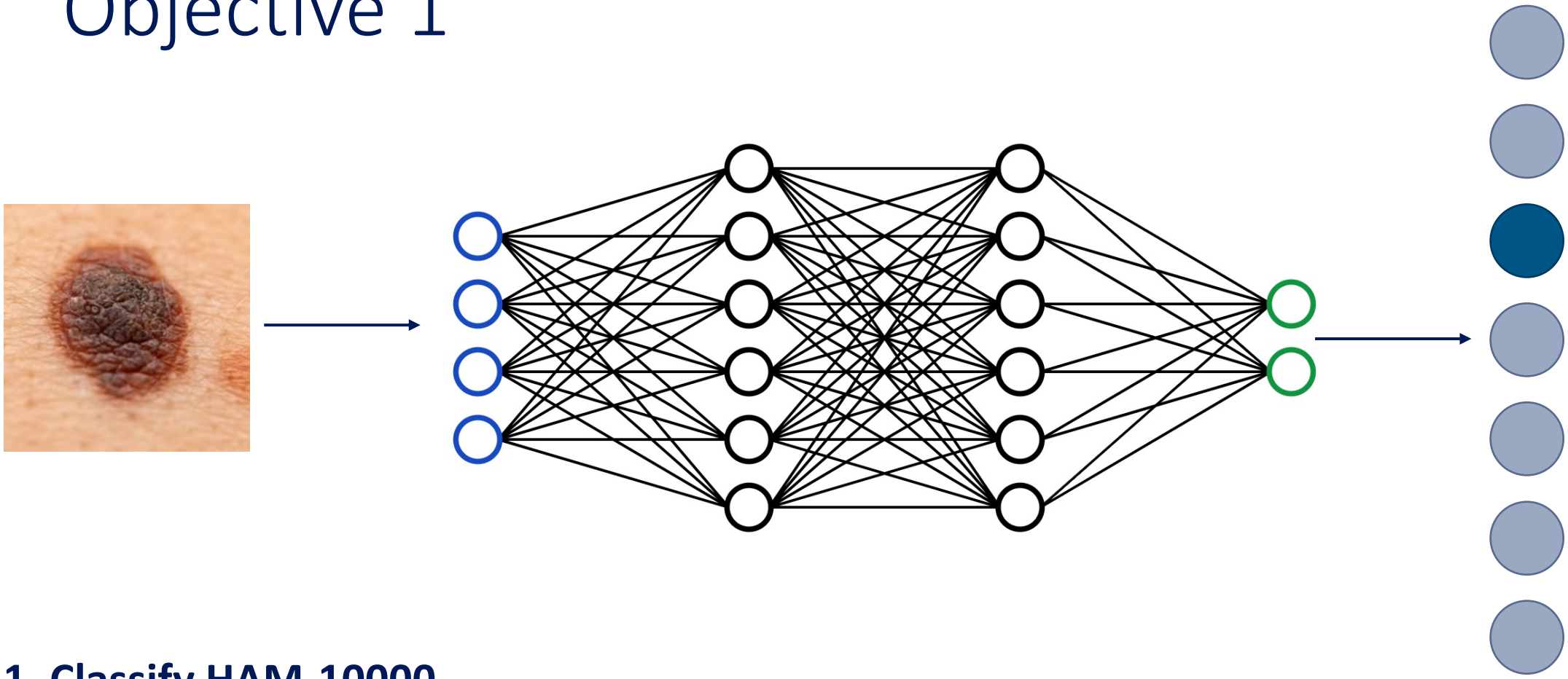
Dataset

Skin Pathology	*Number of Images*
Actinic keratosis (akiec)	327
Basal cell carcinoma (bcc)	514
Dermatofibroma (df)	115
Melanoma (mel)	1113
Nevus (nv)	6705
Pigmented benign keratosis (bkl)	1099
Vascular lesions (vasc)	142
Total	10015

- PRO: Representative of images that could be taken with a smartphone
- CON: Uneven label distribution

Duke

Objective 1



1. Classify HAM-10000

Duke

Objective 2

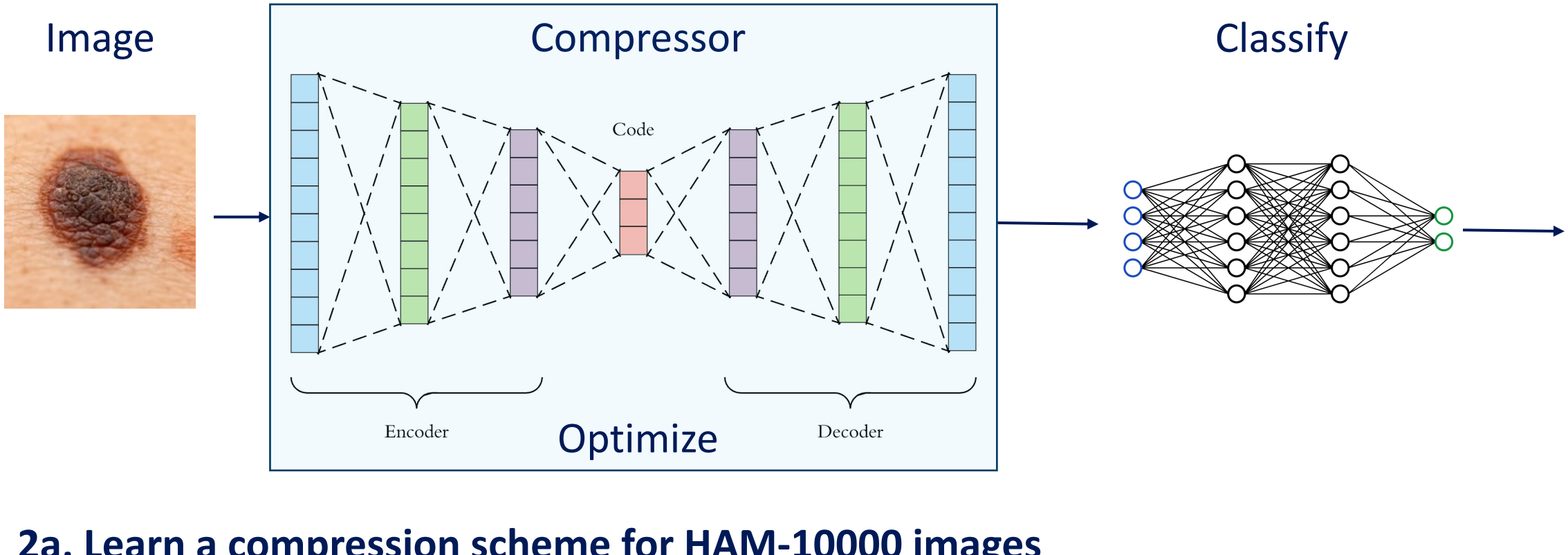
- Current standard image compression is done as JPEG/PNG (etc.)
- Image compression is moving towards Neural Image Compression **(NIC)**

Neural Image Compression for Gigapixel Histopathology Image Analysis

David Tellez, Geert Litjens, Jeroen van der Laak, Francesco Ciompi

Duke

Objective 2



2a. Learn a compression scheme for HAM-10000 images

2b. See how effective this compression is by analysis using the classifier

Duke

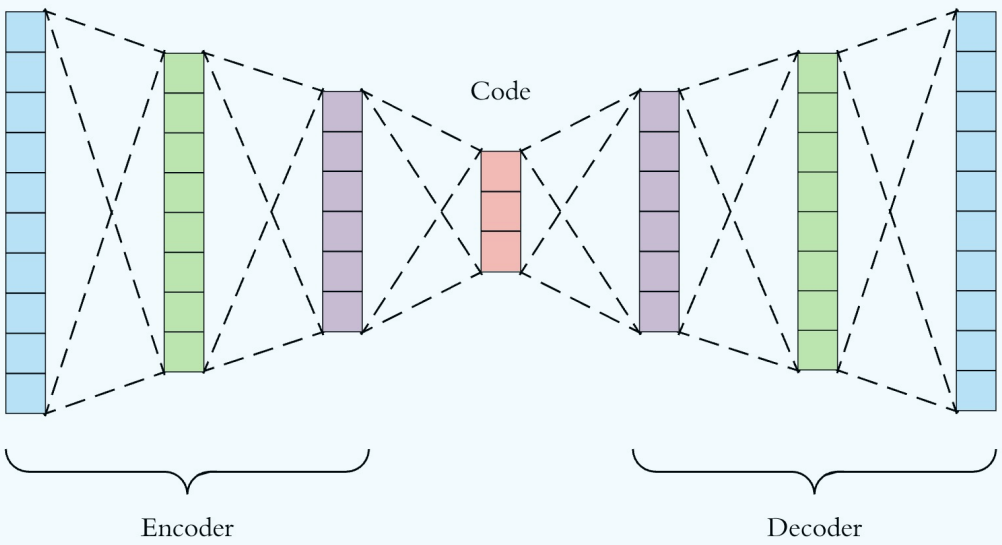
Objective 3

Score

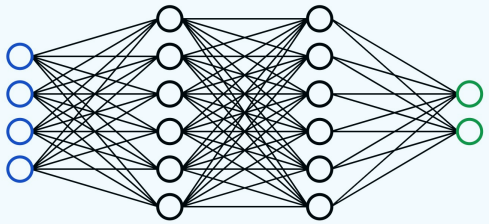
Optimize Together

Image

Compressor



Classify



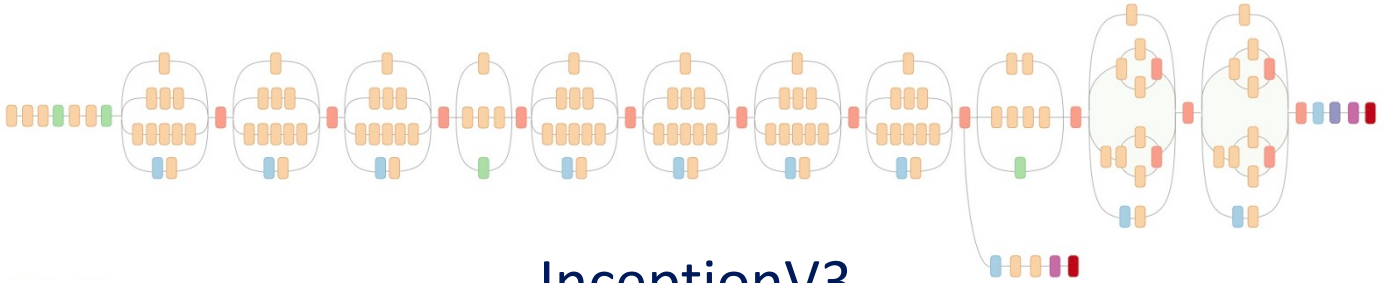
2a. Learn a compression scheme jointly with the classifier for HAM-10000 images

2b. After training, extract images and observe “optimal image” for classifier

Duke

Classifier Architecture

Image
(128,128,3)



InceptionV3
(Imagenet weight
initialization)

Dense
(128 Nodes)

Dense
(7 Nodes)

Class

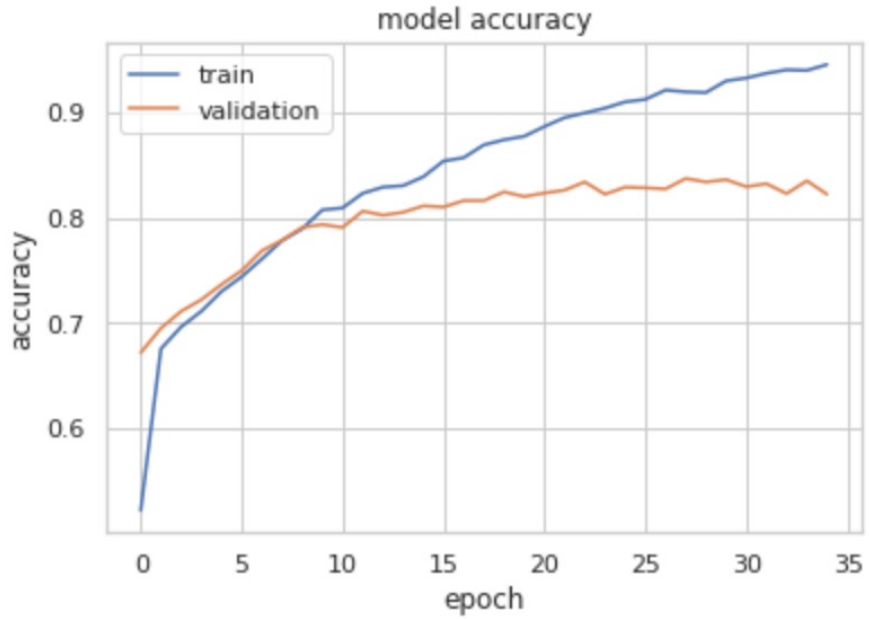
Dermatologist-level classification of skin cancer with deep neural networks

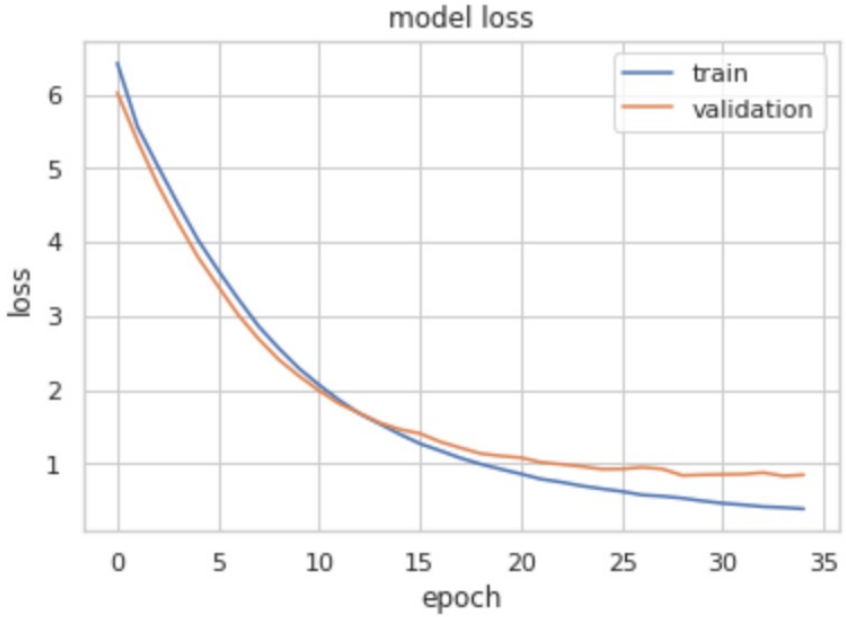
Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

Duke

Details

- 70/20/10 Train/Validation/Test Split
- Adam optimizer
- L2 regularization
- Data Augmentation
- Categorical Crossentropy Loss
- Normalize inputs to 0-1
- Trained for 35 Epochs





Duke

Classifier Performance

Table 1: Classifier Metrics

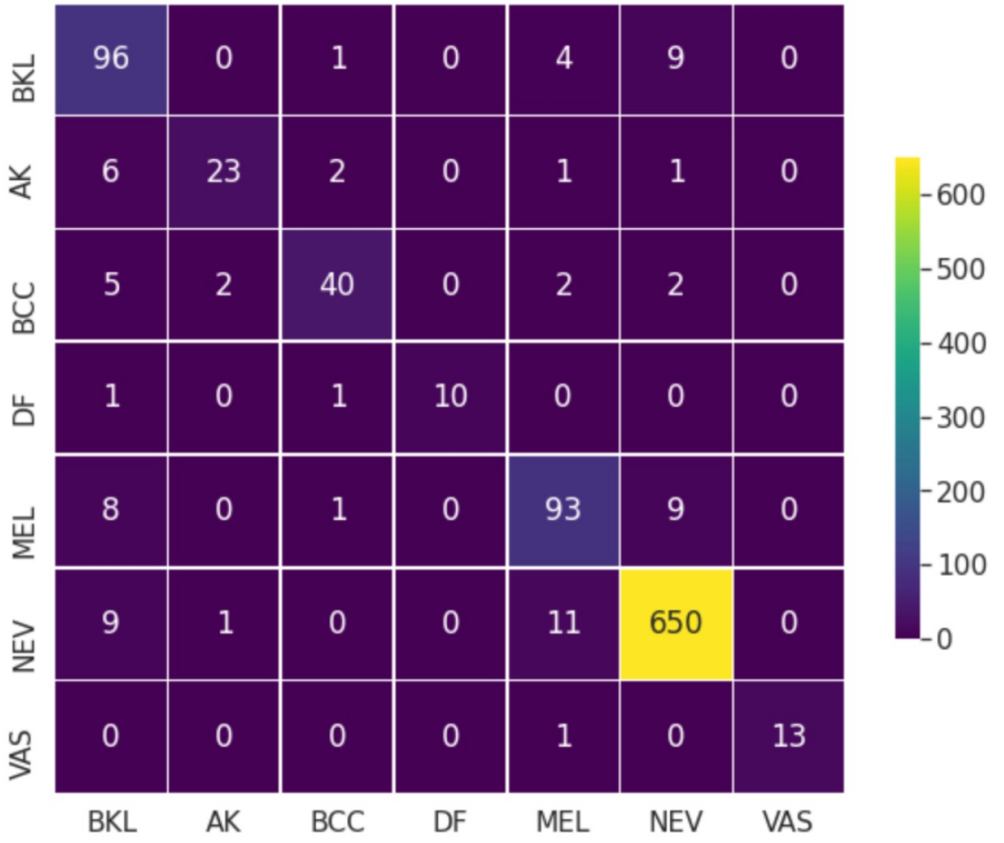
Data Set	Loss	Accuracy
Training	0.3819	0.9506
Validation	0.8479	0.8225
Test	0.473367	0.923154

Duke

Classifier Performance

Table 2: Classifier Dataset Precision, Recall, F1-Score, and Support

Skin Lesion	Precision	Recall	F1-Score	Support
BKL	0.61	0.69	0.65	110
AK	0.67	0.42	0.52	33
BCC	0.89	0.65	0.75	51
DF	0.70	0.58	0.64.	12
MEL	0.55	0.56	0.56	111
NEV	0.91	0.92	0.92	671
VAS	0.85	0.79	0.81	14
Accuracy			0.82	1002
Macro Average	0.74	0.66	0.69	1002
Weight Average	0.82	0.82	0.82	1002

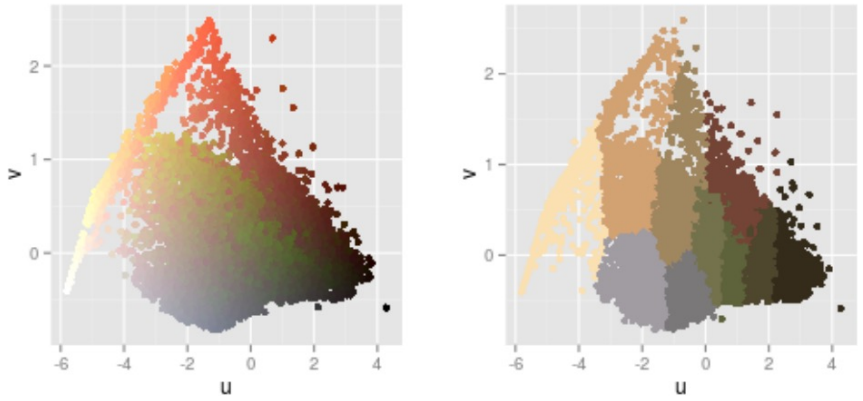


Duke

Good at NEV, Bad at MEL, BKL

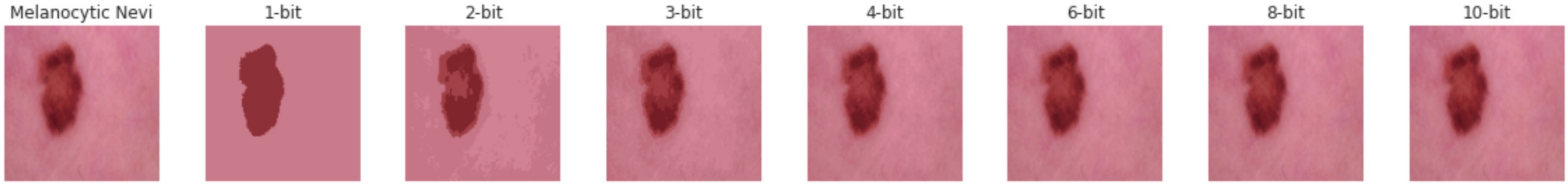
Compression Method

- Naïve, first attempt: K-Means Clustering



Results

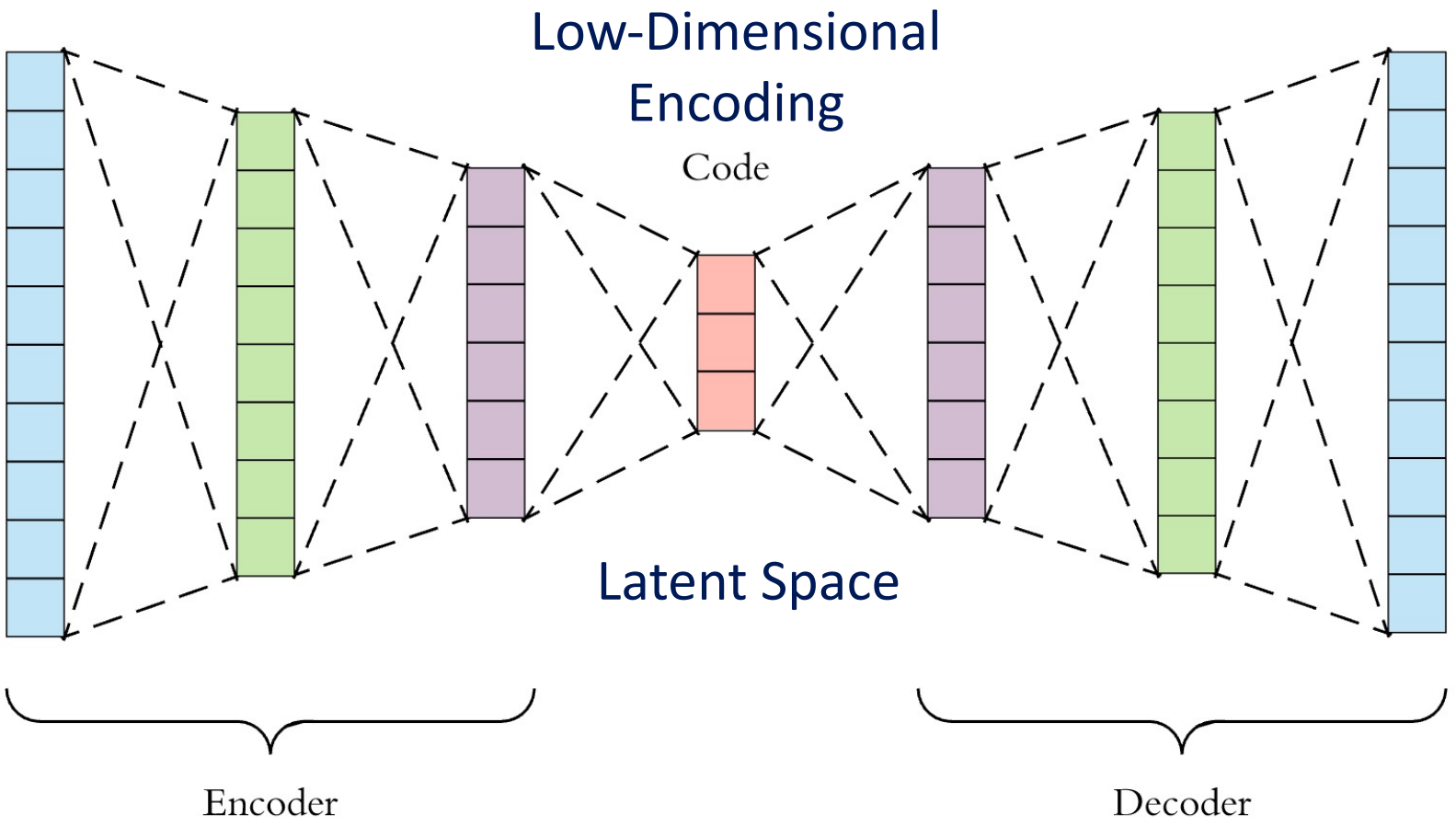
- Key Issue: Not jointly trainable with the classifier.



Average Color Deviation	36.05	22.03	14.00	10.11	6.00	3.95	2.35

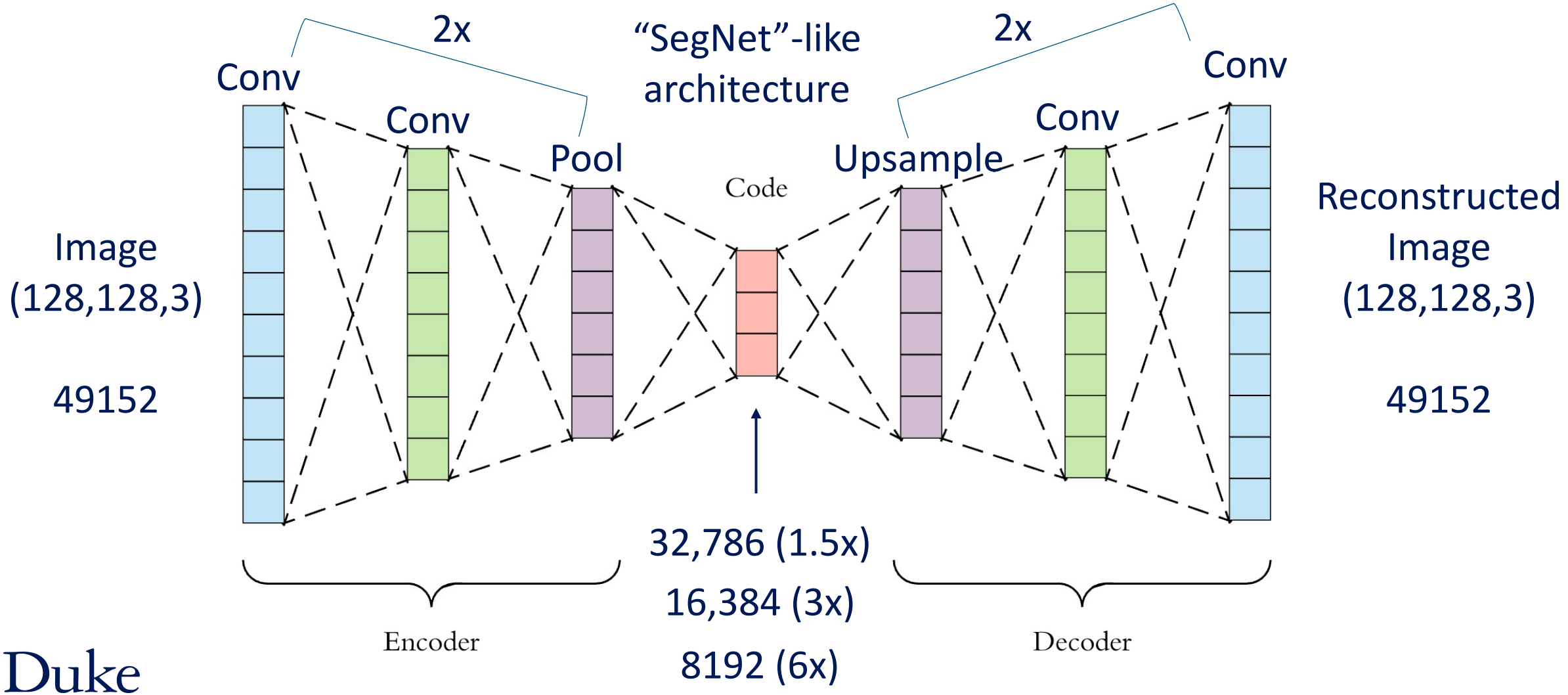
Duke

Autoencoder



Duke

Autoencoder Architecture



Duke

Training Details

- Train for 50 epochs
- Mean Absolute Error loss

Data Set	Loss	Accuracy
Training	0.0149	-
Validation	0.0149	-
Test	1.1768	0.7538

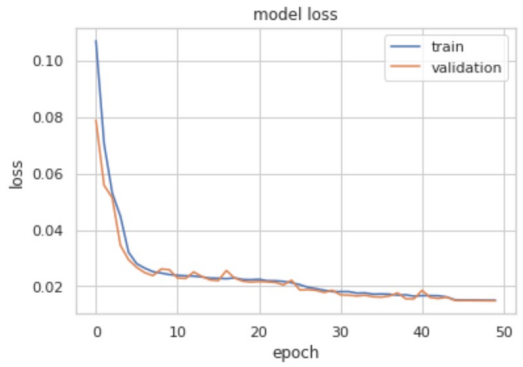
Table 4: 6X Autoencoder

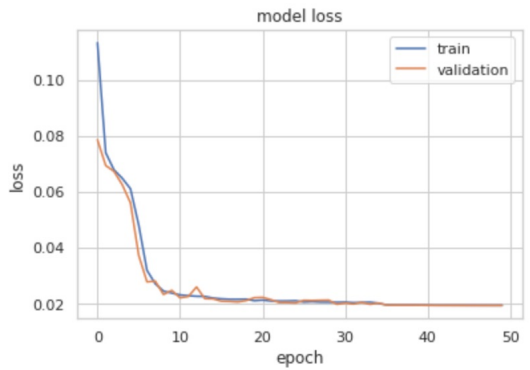
Data Set	Loss	Accuracy
Training	0.0194	-
Validation	0.0194	-
Test	1.2854	0.7315

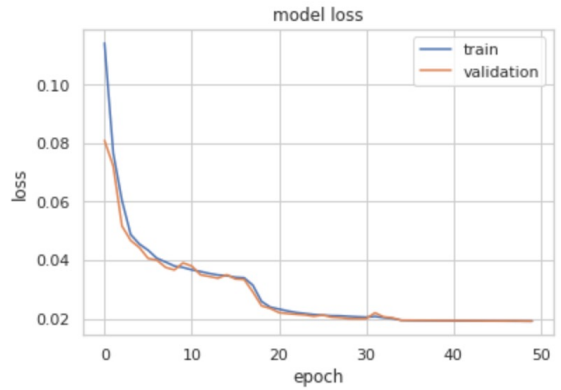
Table 5: 3X Autoencoder

Data Set	Loss	Accuracy
Training	0.0193	-
Validation	0.0193	-
Test	1.3058	0.7196

Table 6: 1.5X Autoencoder

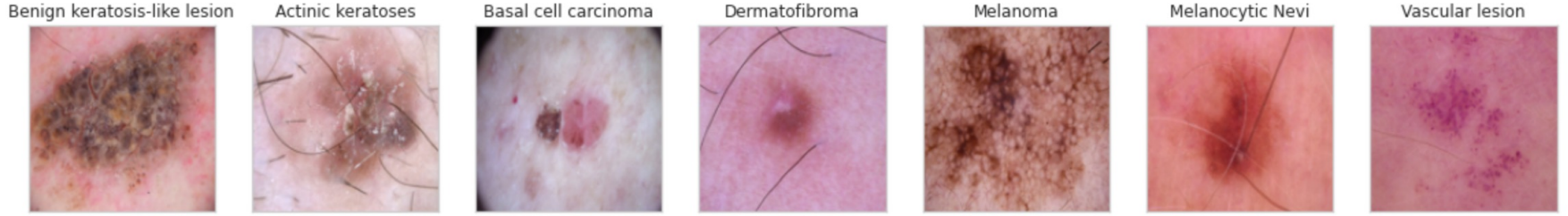
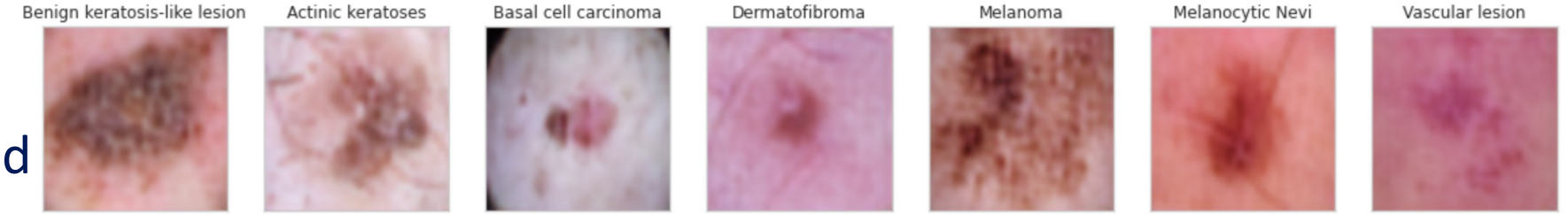
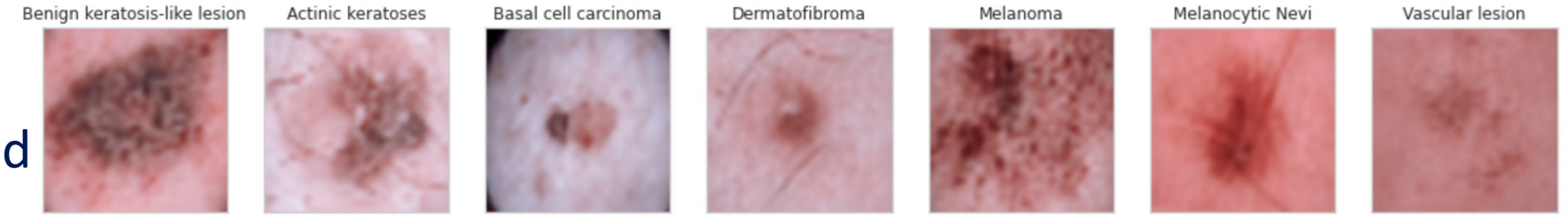
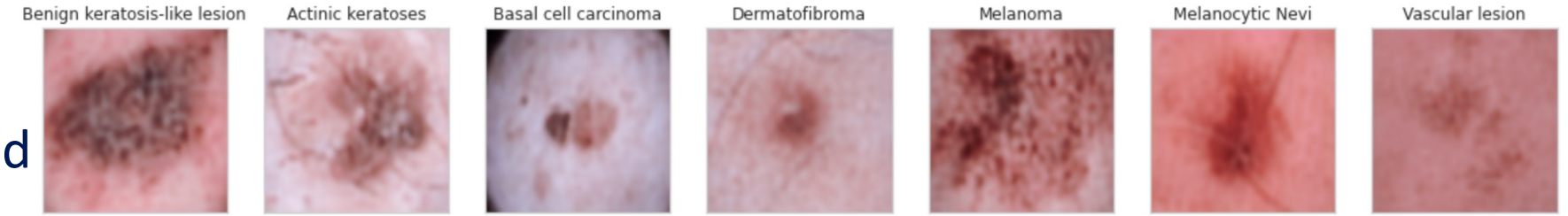






Duke

Autoencoder Results

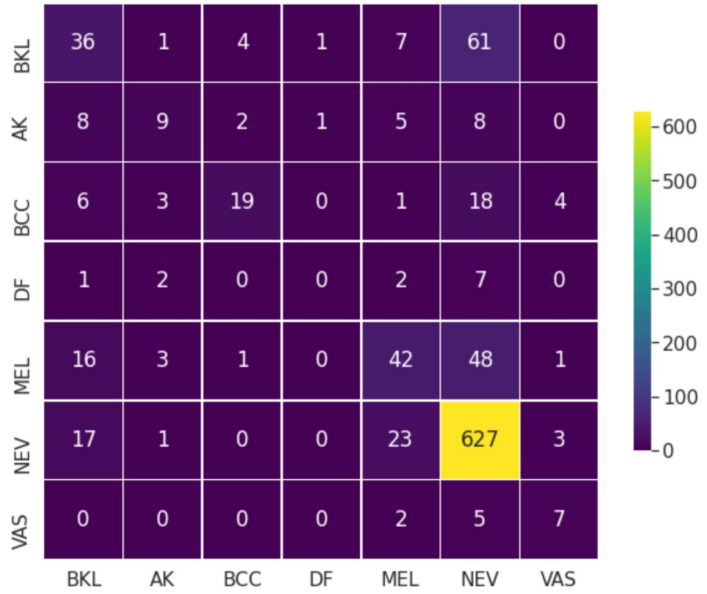
		Test Accuracy
Original		
6X Compressed		0.7538
3X Compressed		0.7315
1.5X Compressed		0.7196

Autoencoder Results – 6X

Table 10: 6X Autoencoder + Pretrained Classifier Precision, Recall, F1-Score, and Support

Skin Lesion	Precision	Recall	F1-Score	Support
BKL	0.78	0.26	0.39	110
AK	0.57	0.24	0.34	33
BCC	0.44	0.78	0.57	51
DF	0.00	0.00	0.00	12
MEL	0.52	0.32	0.39	111
NEV	0.81	0.96	0.88	671
VAS	1.00	0.07	0.13	14
Accuracy			0.76	1002
Macro Average	0.59	0.38	0.39	1002
Weight Average	0.74	0.76	0.72	1002

Figure 7: 6X Autoencoder + Pretrained Classifier Classifier Confusion Matrix



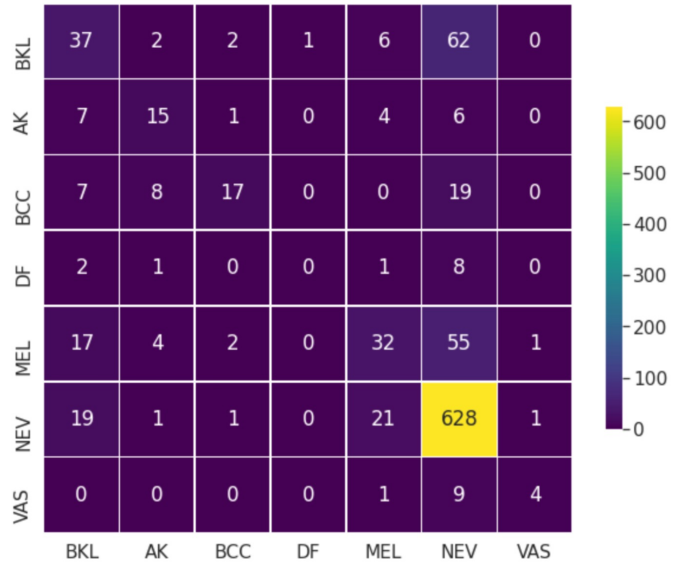
Duke

Autoencoder Results -3X

Table 11: 3X Autoencoder + Pretrained Classifier Precision, Recall, F1-Score, and Support

Skin Lesion	Precision	Recall	F1-Score	Support
BKL	0.60	0.60	0.60	110
AK	0.67	0.48	0.56	33
BCC	0.63	0.78	0.70	51
DF	0.00	0.00	0.00	12
MEL	0.56	0.48	0.52	111
NEV	0.91	0.94	0.92	671
VAS	1.00	0.93	0.96	14
Accuracy			0.82	1002
Macro Average	0.62	0.60	0.61	1002
Weight Average	0.80	0.82	0.81	1002

Figure 8: 3X Autoencoder + Pretrained Classifier Classifier Confusion Matrix



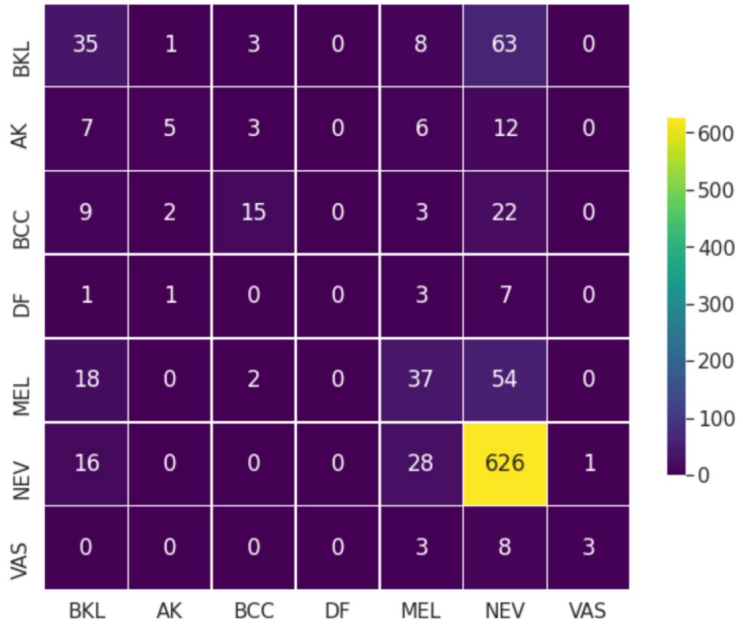
Duke

Autoencoder Results -1.5X

Table 12: 1.5X Autoencoder + Pretrained Classifier Precision, Recall, F1-Score, and Support

Skin Lesion	Precision	Recall	F1-Score	Support
BKL	0.60	0.51	0.55	110
AK	0.52	0.33	0.41	33
BCC	0.56	0.61	0.58	51
DF	0.00	0.00	0.00	12
MEL	0.53	0.44	0.48	111
NEV	0.86	0.94	0.90	671
VAS	0.70	0.50	0.58	14
Accuracy			0.78	1002
Macro Average	0.54	0.48	0.50	1002
Weight Average	0.76	0.78	0.77	1002

Figure 9: 1.5X Autoencoder + Pretrained Classifier Confusion Matrix



Duke

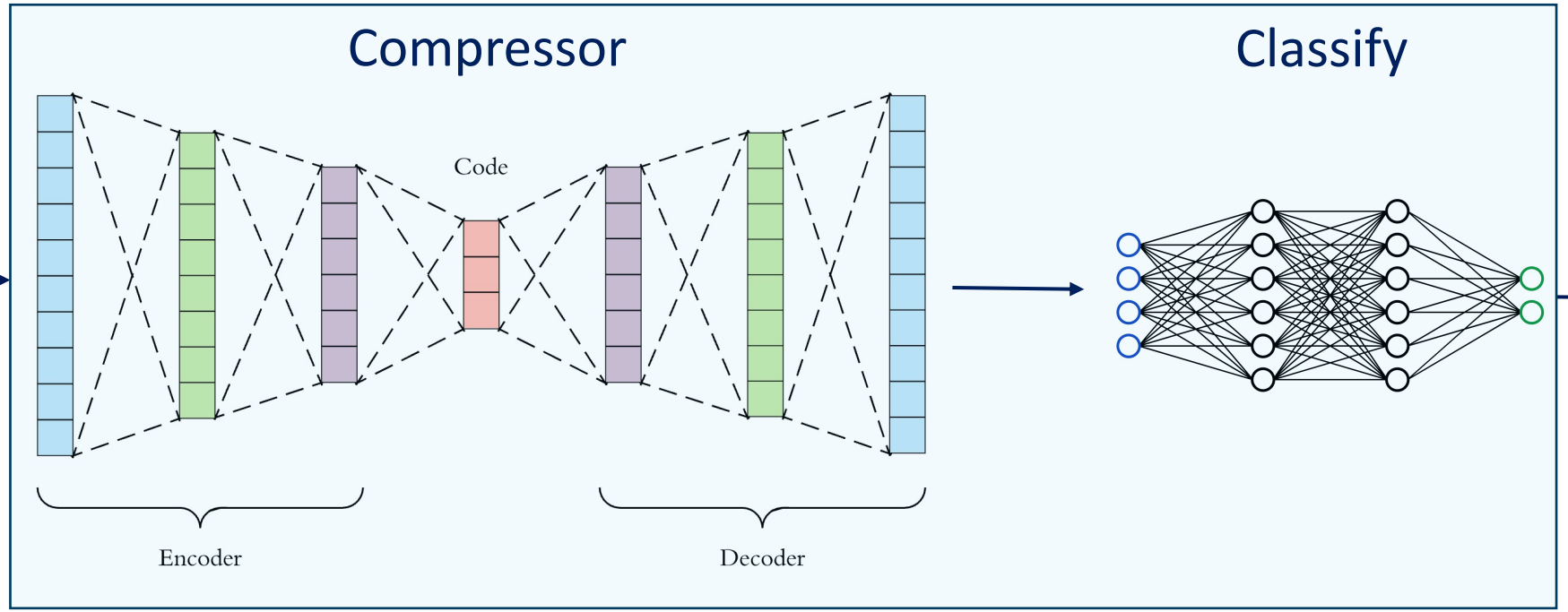
Autoencoder Results Summary

- All three autoencoders had similar accuracies, with the 6X being the most accurate and 1.5X being the least accurate on test set
- All three models did the best at classifying NEV
- All three models were unable to properly classify DF

Duke

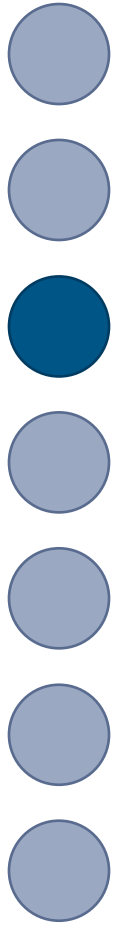
Joint Training Architecture

Image
(128,128,3)



Optimize Together

Score

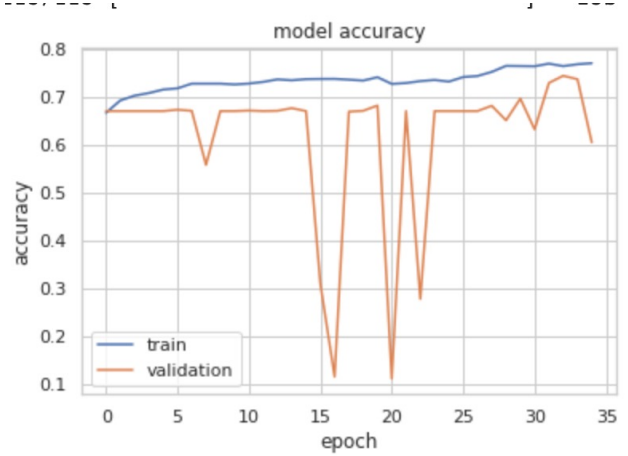


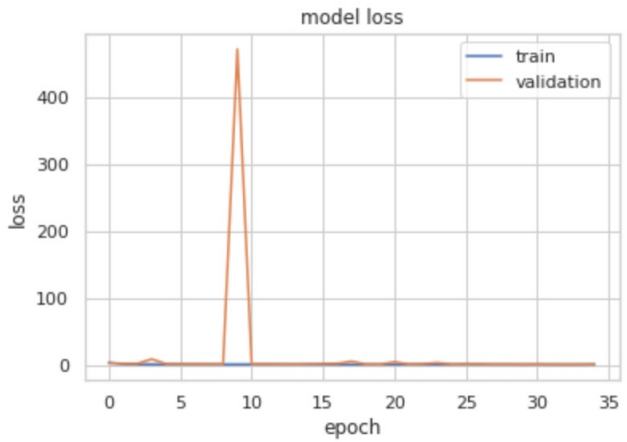
Duke

Training Details

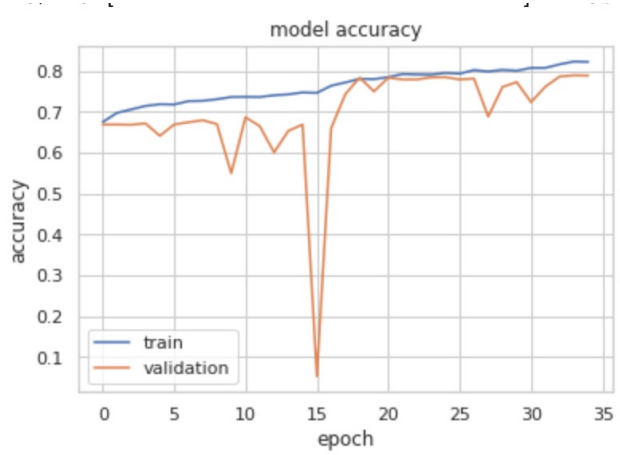
- 70/20/10 Train/Validation/Test Split
- Adam optimizer
- L2 regularization
- Data Augmentation
- Categorical Crossentropy Loss
- Normalize inputs to 0-1
- Trained for 35 Epochs
- Checkpointing due to unstable validation accuracy

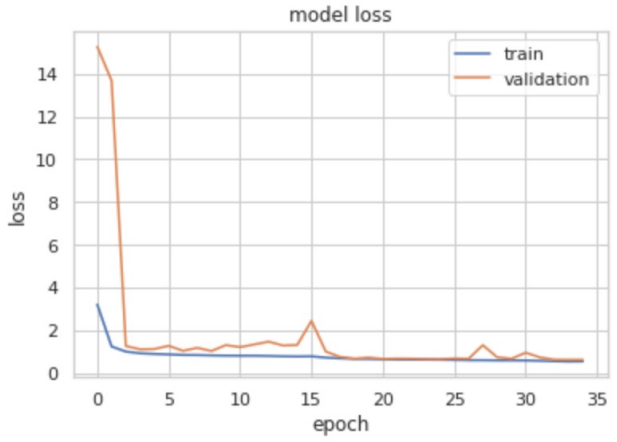
Training Details



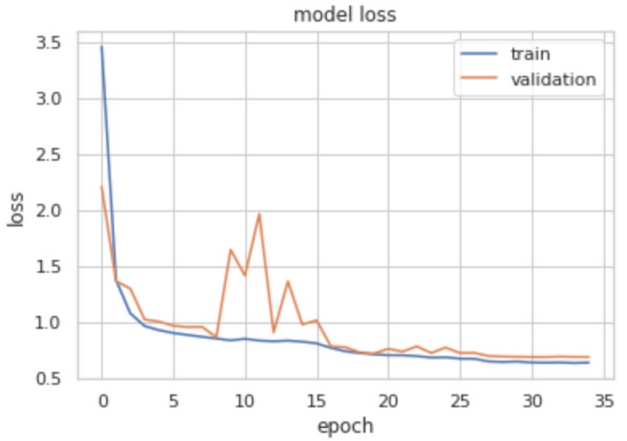


1.5X





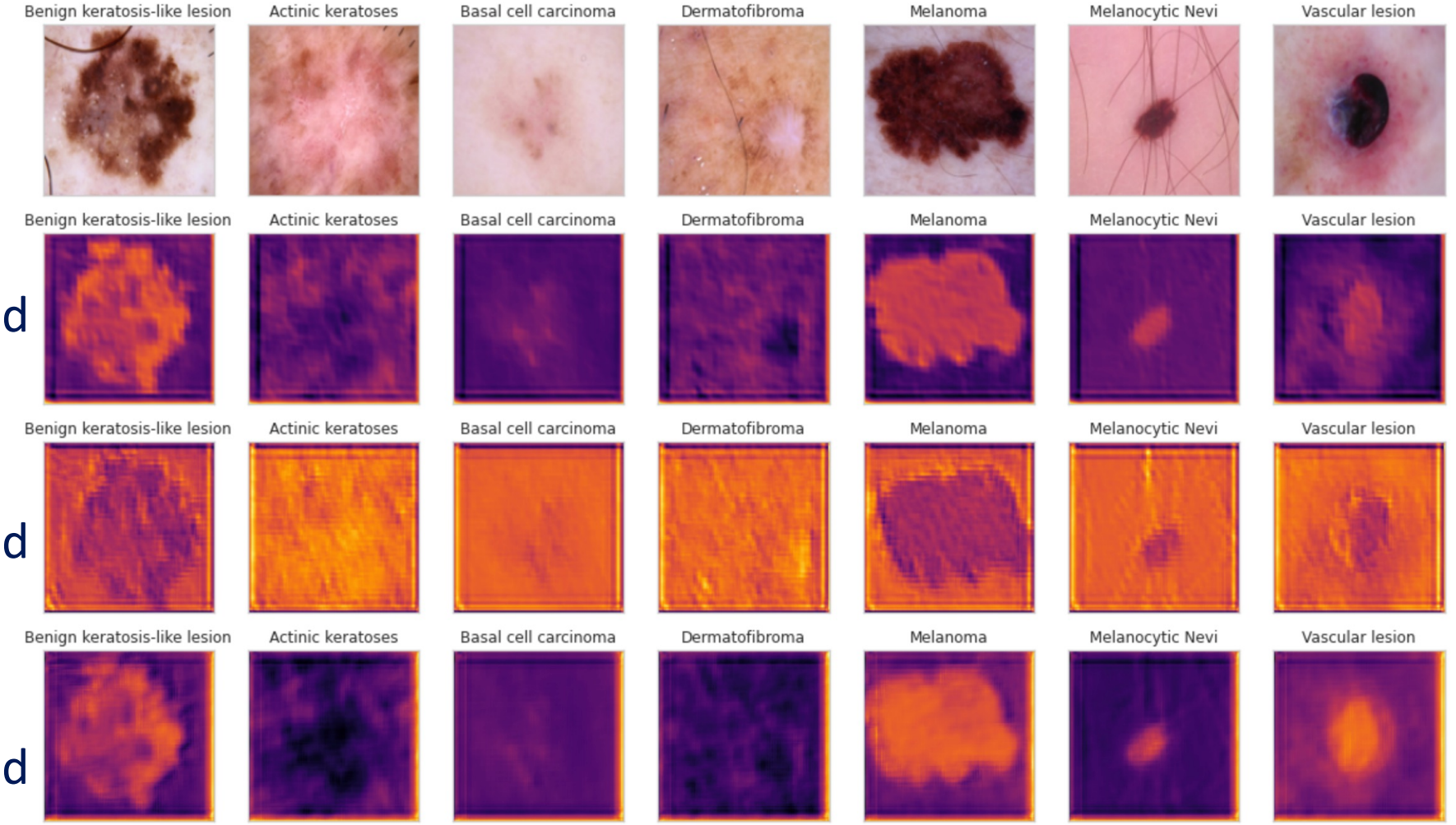
3X



6X

Duke

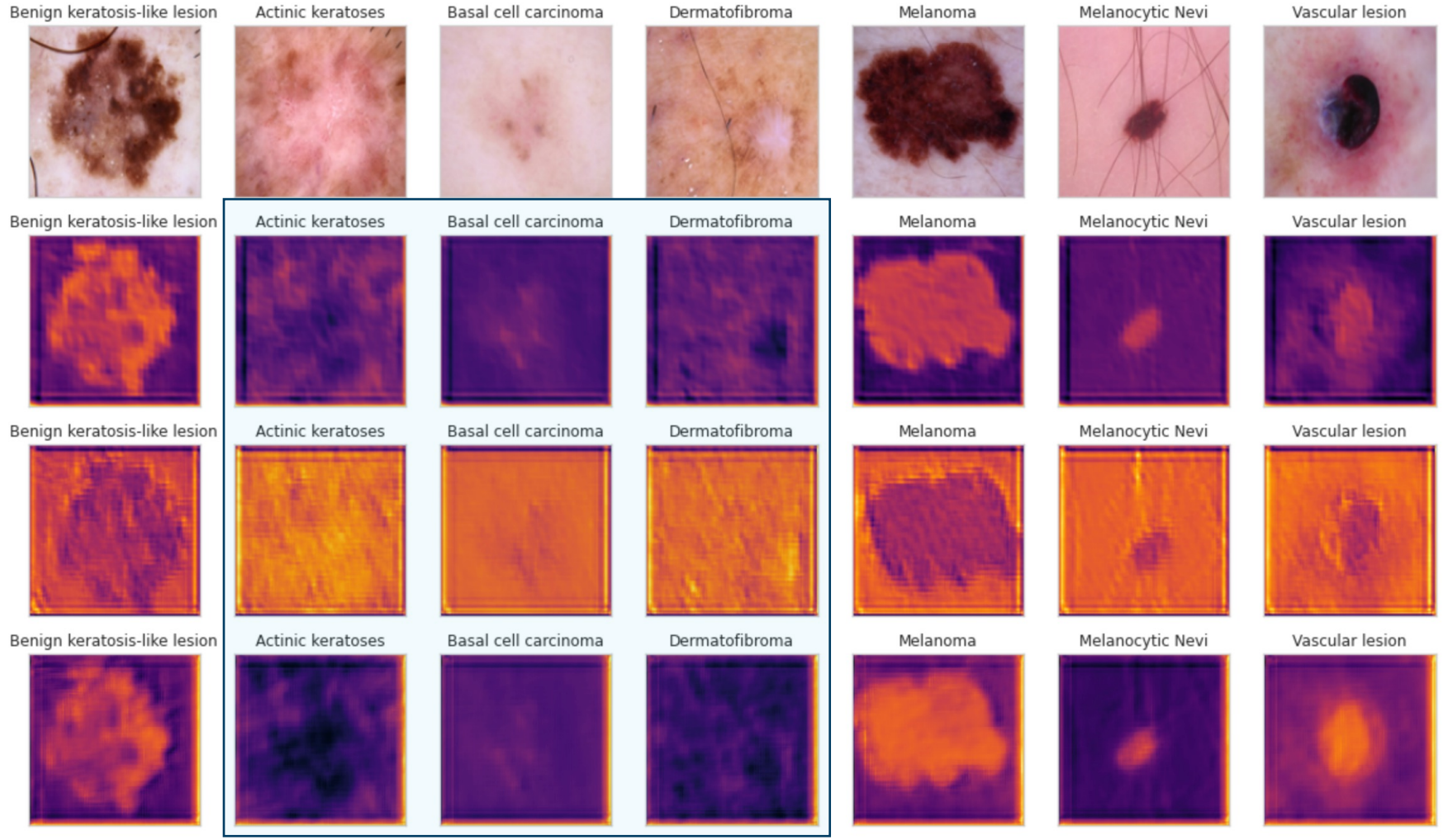
Joint Training Results



	Test Accuracy
Original	
6X Compressed	0.7555
3X Compressed	0.8183
1.5X Compressed	0.7814

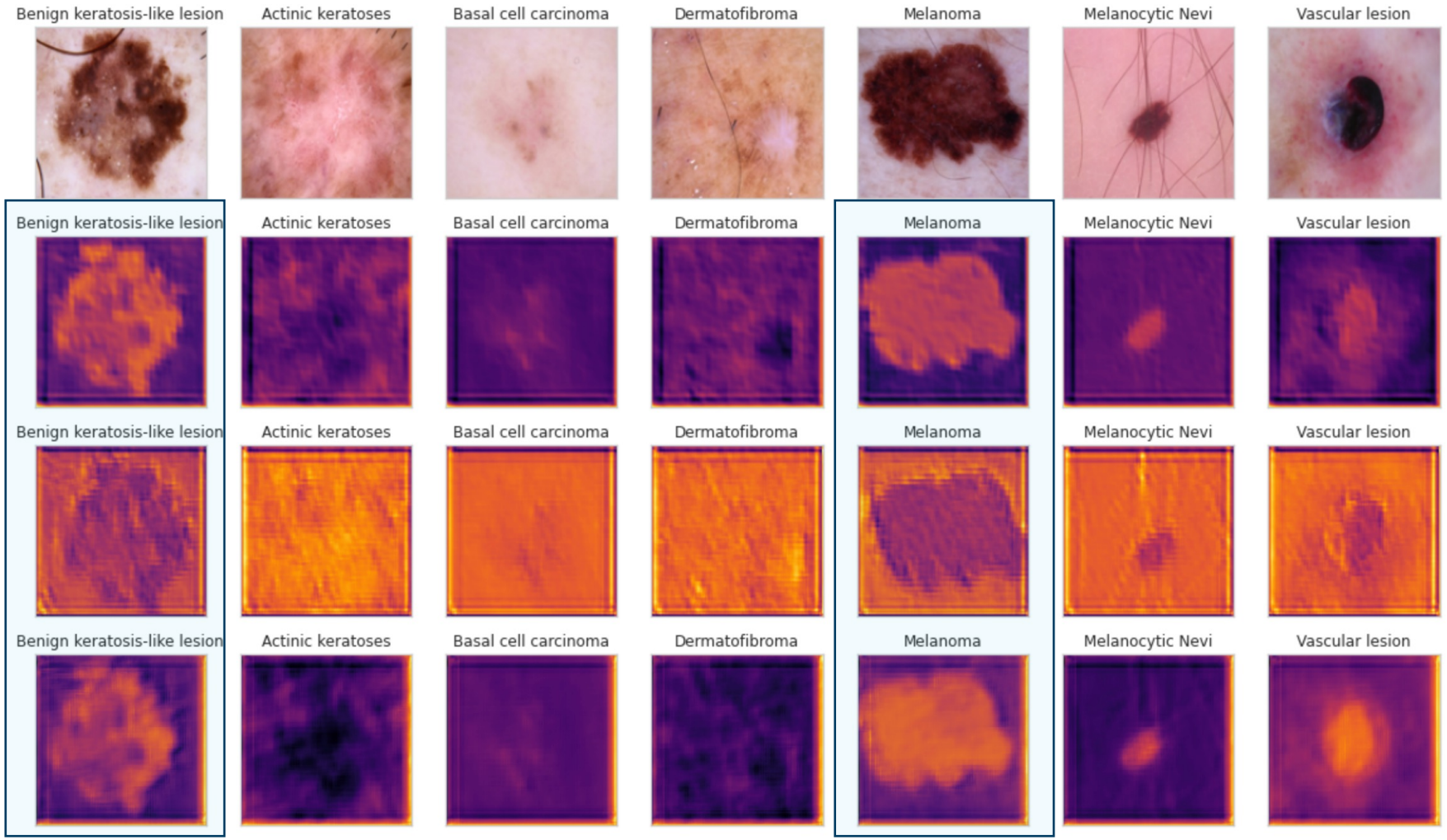
Contrast is critical for classification!

Joint Training Results



DF, BCC, AK have poor contrast, may suggest why model is confused.

Joint Training Results



Duke

BKL and MEL have similar contrast, theory for confusion matrix results

Joint Training Results

Data Set	Loss	Accuracy
Training	0.6933	0.7699
Validation	0.7685	0.7432
Test	0.7198	0.7555

Table 7: 6X Joint Model

Data Set	Loss	Accuracy
Training	0.5469	0.8261
Validation	0.6289	0.7898
Test	0.5478	0.8183

Table 8: 3X Joint Model

Data Set	Loss	Accuracy
Training	0.6291	0.7869
Validation	0.6870	0.7621
Test	0.6098	0.7814

Table 9: 1.5X Joint Model

3X Encoder Architecture had the best performance

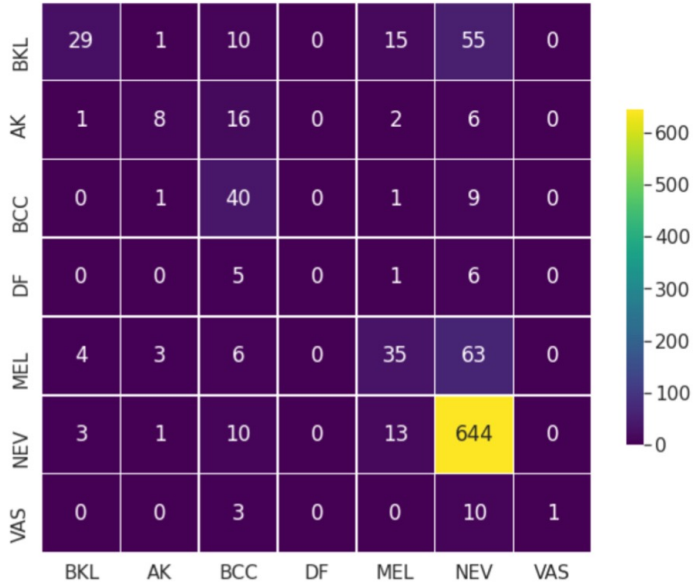
Duke

Joint Training Results – 6X

Table 13: 6X Compression Jointly Trained Model Precision, Recall, F1-Score, and Support

Skin Lesion	Precision	Recall	F1-Score	Support
BKL	0.78	0.26	0.39	110
AK	0.57	0.24	0.34	33
BCC	0.44	0.78	0.57	51
DF	0.00	0.00	0.00	12
MEL	0.52	0.32	0.39	111
NEV	0.81	0.96	0.88	671
VAS	1.00	0.07	0.13	14
Accuracy			0.76	1002
Macro Average	0.59	0.38	0.39	1002
Weight Average	0.74	0.76	0.72	1002

Figure 10: 6X Compression Jointly Trained Model Classifier Confusion Matrix



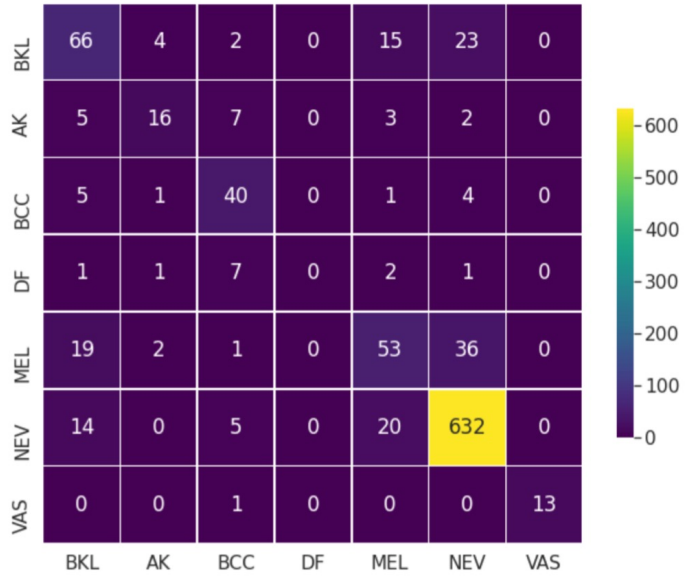
Duke

Joint Training Results – 3X

Table 14: 3X Compression Jointly Trained Model Precision, Recall, F1-Score, and Support

Skin Lesion	Precision	Recall	F1-Score	Support
BKL	0.60	0.60	0.60	110
AK	0.67	0.48	0.56	33
BCC	0.63	0.78	0.70	51
DF	0.00	0.00	0.00	12
MEL	0.56	0.48	0.52	111
NEV	0.91	0.94	0.92	671
VAS	1.00	0.93	0.96	14
Accuracy			0.82	1002
Macro Average	0.62	0.60	0.61	1002
Weight Average	0.80	0.82	0.81	1002

Figure 11: 3X Compression Jointly Trained Model Classifier Confusion Matrix



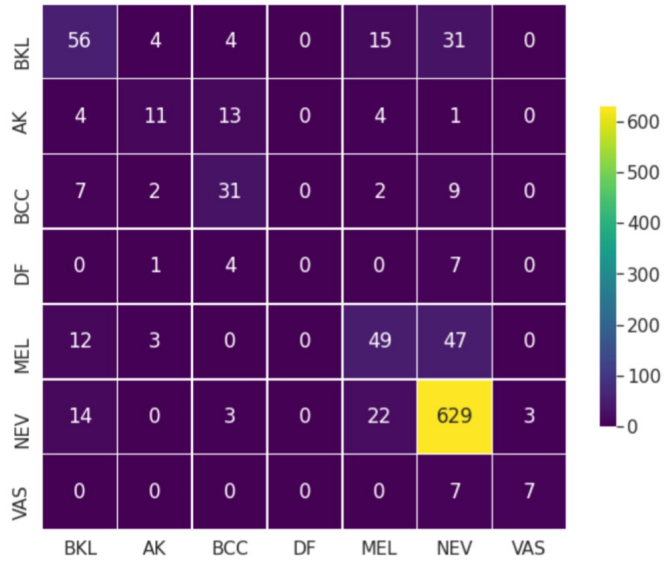
Duke

Joint Training Results -1.5X

Table 15: 1.5X Compression Jointly Trained Model Precision, Recall, F1-Score, and Support

Skin Lesion	Precision	Recall	F1-Score	Support
BKL	0.60	0.51	0.55	110
AK	0.52	0.33	0.41	33
BCC	0.56	0.61	0.58	51
DF	0.00	0.00	0.00	12
MEL	0.53	0.44	0.48	111
NEV	0.86	0.94	0.90	671
VAS	0.70	0.50	0.58	14
Accuracy			0.78	1002
Macro Average	0.54	0.48	0.50	1002
Weight Average	0.76	0.78	0.77	1002

Figure 12: 1.5X Compression Jointly Trained Model Classifier Confusion Matrix



Duke

Joint Training Results Summary

- 3X model had **superior average precision and recall** to 1.5X and 6X models.
  - Follows trend seen with autoencoder training
- 3X Model **was able to distinguish vascular lesions from nevi**, other two models were not
- 3X Model was **superior at classifying BKL**
- All three models **confused between BCC, AK, DF**
- 3X Model test accuracy **10% less accurate than classifier alone** (82%)
- All three models performed the **best on nevus classification**

Duke

Takeaways

- **Image background contrast is critical for HAM-10000 classification**
  - Imaging scheme: highlight contrast between lesion and background
  - Compression scheme: possible to only encode boundaries between lesion and background
- **Latent space size may not directly be correlated to compression effectiveness**
- **Autoencoder effectiveness may not be directly correlated to classification effectiveness**

Duke

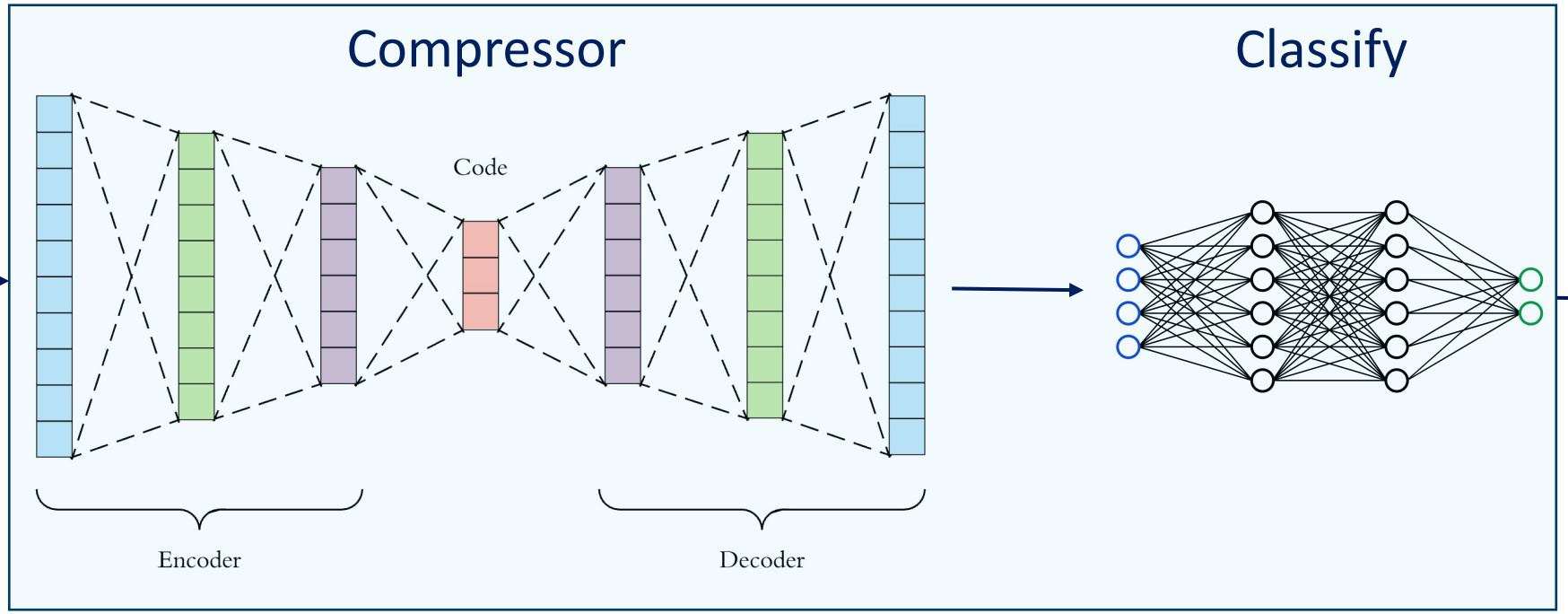
Future Steps

- Train physical layer that serves as an edge detection filter, add to front of joint model and observe performance.
- Address unbalanced dataset issue
- Test different underlying classification architectures

Thank you!

- Any questions?

Image
(128,128,3)



Score

Optimize Together

Duke