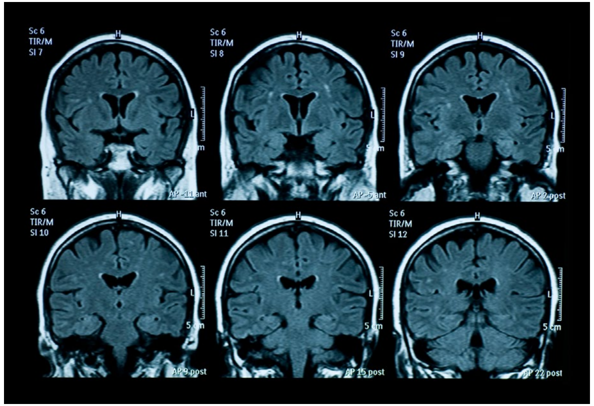
Sensitivity Assessment of a Two-Step Method in Skin Image Identification

Brain Fodale and Grant Himmelmann

Motivation

Motivation

- **Image Processing** is a growing field of interest in the scientific and medical communities
- Medical Applications: Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Skin Identification
- Our Focus: Evaluate three skin identification methods and employ them on melanoma pictures to help identify areas of skin cancer



IOWA

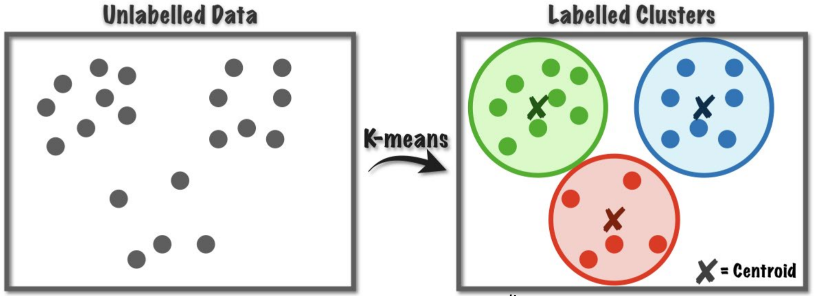
Methods

Goals

- Classify image pixels as skin/not skin
  - Perform clustering on image training data
  - Fit model to each cluster
  - Use fit to perform classification on test data
- Identify best classification method for our problem
  - Methods: GLM (Logit and Probit), FF-NN
  - Data taken from *GLMs Applied for Skin Identification in Image Processing*: Basterrech et al. (2015)
- Extend developed model to melanoma images

Clustering

- **Clustering**: Process of dividing a dataset into several similar groups
- K-Means Clustering Algorithm
  - Step 1: Divide items into K clusters
  - Step 2: Calculate the centroid of each cluster
  - Step 3: Perform reassignments if necessary
  - Step 4: Repeat
- Choosing K: Elbow Method



medium.com

Logistic Regression

- **Logistic Regression** is a process of modeling the probability of a binary outcome given an input variable
- Link function: Logit

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

Probit Regression

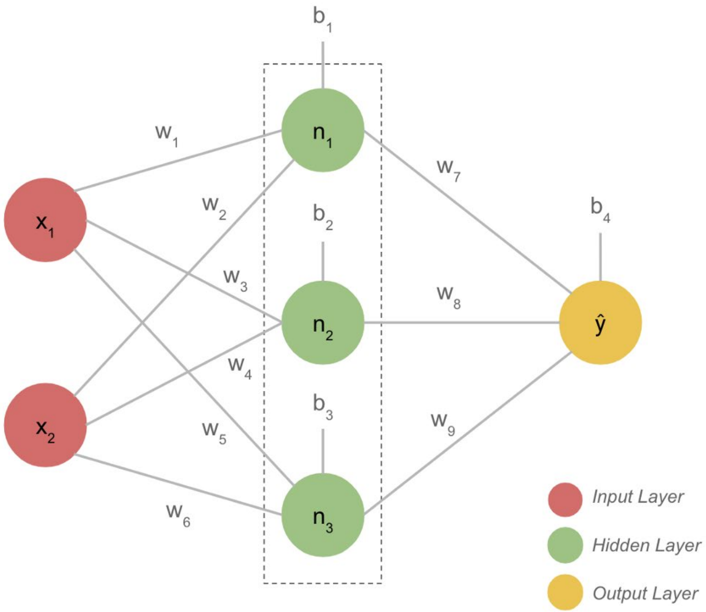
- **Probit Regression** is another member of the GLM family
  - Link function: Inverse Cumulative Normal (Probit)
- Tail Behavior: Approaches 0 and 1 slightly quicker than Logistic Regression

$$\phi^{-1}(p) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

IOWA

Neural Networks (FF-NN)

- **Neural Networks** are a relatively new development in the machine learning field
- Structure:
  - Nodes
  - Hidden Layer
  - Connections/Weights



IOWA

Fitting the Neural Network

- Three input variables
- Two output nodes
- Fit models with 4, 8, 16, and 32 nodes in one hidden layer
- Optimized with decay parameter to prevent overfitting
- Used MachineShop package in R (B. J. Smith)

Hand Image Reconstruction

RGB Decomposition

	V1	V2	V3	V4
1	160	122	163	1
2	166	128	169	1
3	172	134	175	1
4	175	137	178	1
5	178	140	181	1

	V1	V2	V3	V4
1	108	88	64	0
2	101	81	57	0
3	113	93	69	0
4	121	101	77	0
5	120	100	76	0

Training and Testing

Training

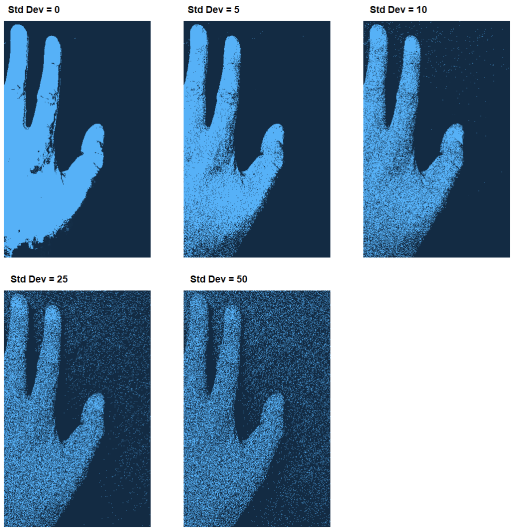
Testing

Output

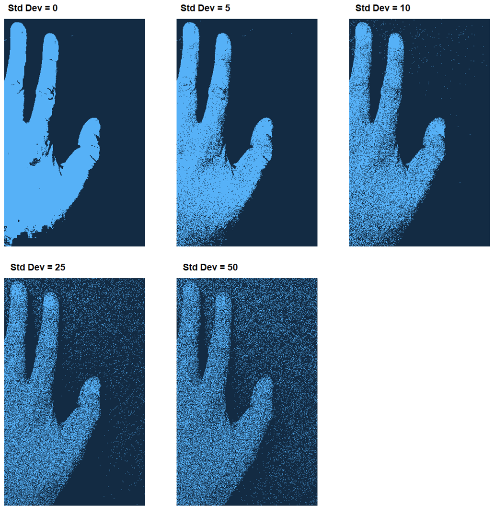
Inducing Noise

- Why add noise?
  - Increase model utility: model can better handle blurry/non-perfect images
  - Test the limit of the model
- Methods:
  - Gaussian noise
  - Raw image modification

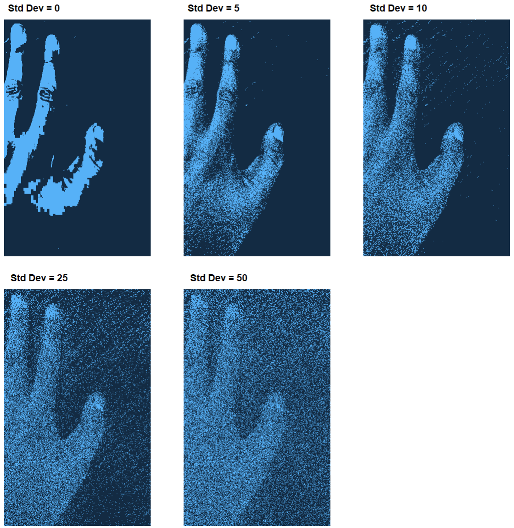
Gaussian Noise, K=32



Logit

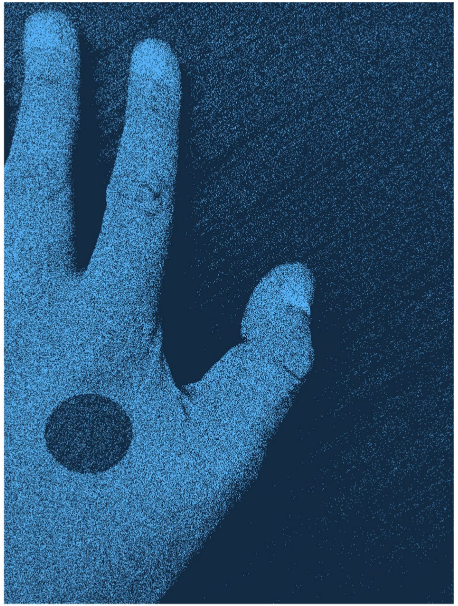


Probit



NN

Raw Image Modification



Performance

Binary Classification Outcomes

- In a binary classification we have four kind of outcomes that may happen during the estimation:
  - True Positive
  - True Negative
  - False Positive
  - False Negative

Metrics

- Sensitivity
- Specificity
- Precision
- Accuracy
- Brier Score (BS)

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{N}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Accuracy = \frac{TP + TN}{P + N}$$

$$BS = \frac{1}{n}\sum_{t=1}^{n}(f_t - o_t)^2$$

IOWA

Assessment

- 10-fold Cross-Validation used to calculate metrics
  - Metrics calculated based on dataset from Basterech paper

Table 1: Performance metrics for classfiers

Method	Parameter	Sensitivity	Specificity	Precision	Accuracy	Brier
Logit	4	0.9996264	0.9629605	0.8760729	0.9705701	0.0316073
	8	0.9998230	0.9722757	0.9043543	0.9779929	0.0193164
	16	0.9994101	0.9829401	0.9390858	0.9863583	0.0117286
	32	0.9964411	0.9951956	0.9819545	0.9954541	0.0039514
Probit	4	0.6893737	0.9616833	0.8152927	0.9051689	0.0953172
	8	0.9997641	0.9710347	0.9004418	0.9769972	0.0207742
	16	0.9996461	0.9811842	0.9332148	0.9850158	0.0126652
	32	0.9952810	0.9942636	0.9785368	0.9944748	0.0045514
FF-NNet	4	0.9971883	0.9971472	0.9892255	0.9971558	0.0024950
	8	0.9979158	0.9979969	0.9924014	0.9979801	0.0018819
	16	0.9983877	0.9910861	0.9720810	0.9926016	0.0059103
	32	0.9988006	0.9982235	0.9932563	0.9983432	0.0014647

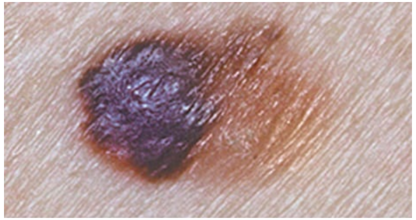
IOWA

Application

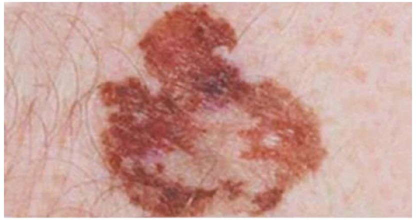
Melanoma

- Melanoma is a type of skin cancer that develops from pigment-producing cells
  - An estimated 7,650 people will die of melanoma in 2022, making it the 9th most deadly cancer type
- Exposure to UV radiation from sunlight and tanning beds increases your risk of developing melanoma
- Symptoms:
  - Changes in existing moles
  - The development of new pigmented or growths on the skin

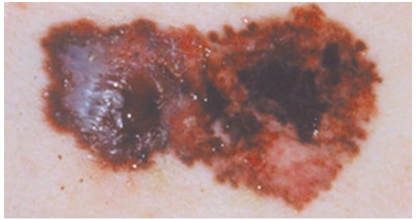
Data



Asymmetry



Border Irregularity

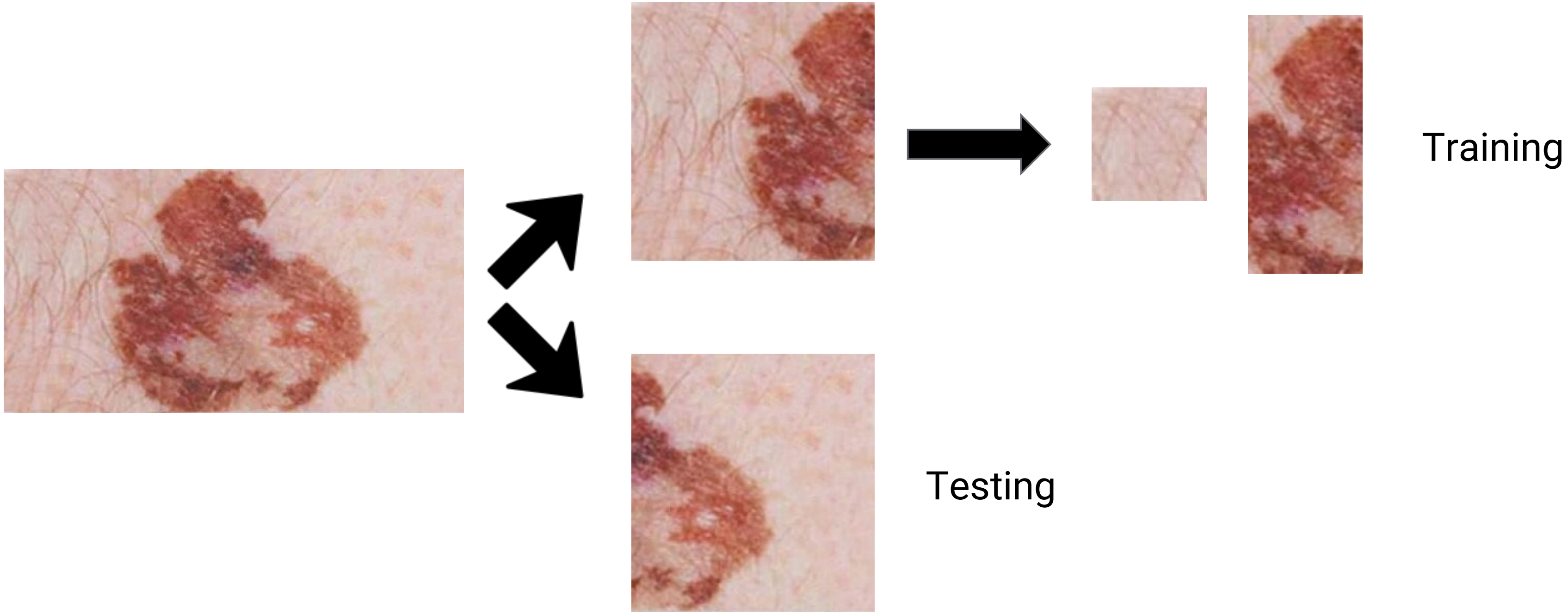


Color Changes

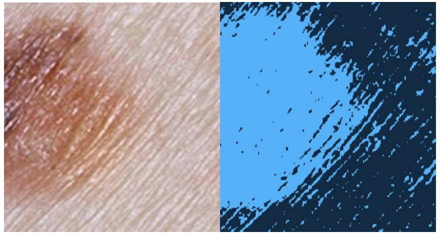
Diameter

Evolving

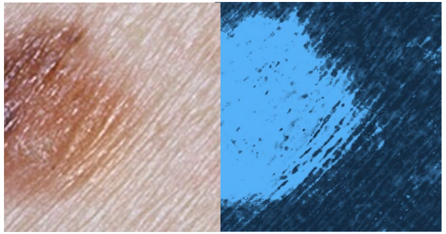
Method



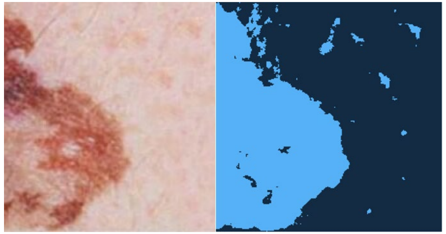
Results



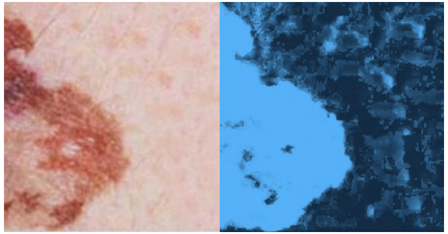
Asymmetry - NN



Asymmetry - Probit



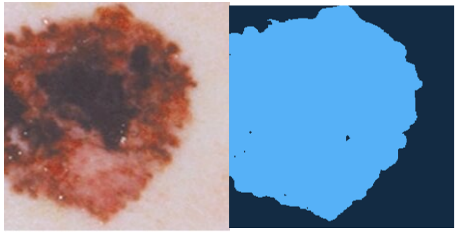
Border Irregularity - NN



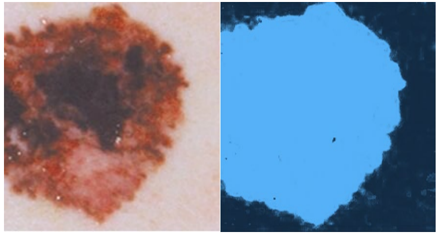
Border Irregularity - Probit

IOWA

Results

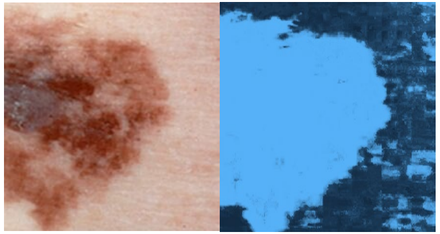


Color Changes - NN



Color Changes - Probit

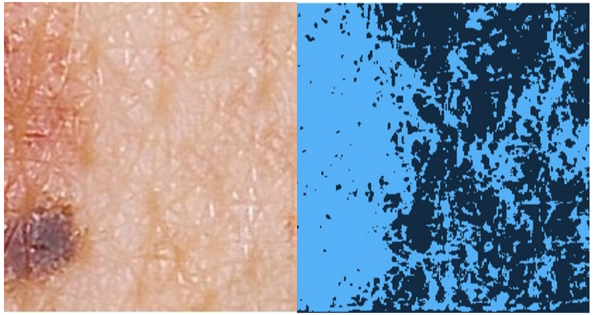
Diameter - NN



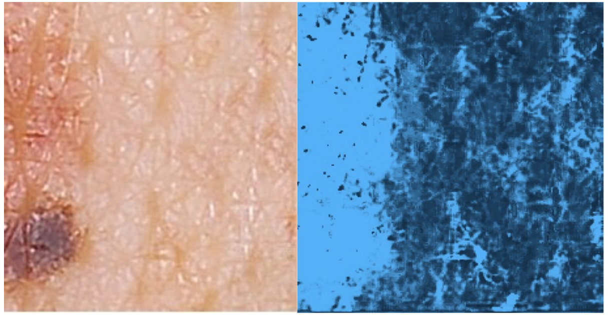
Diameter - Probit

Limitations

- Neural Network fit
- Recognizing evolving melanoma
- Detecting melanoma in non-caucasian individuals

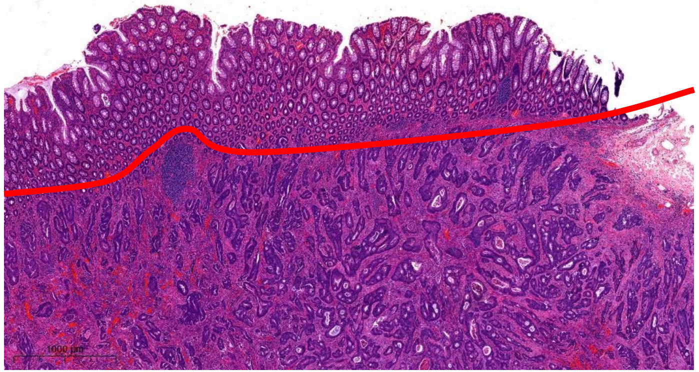


Evolving - NN

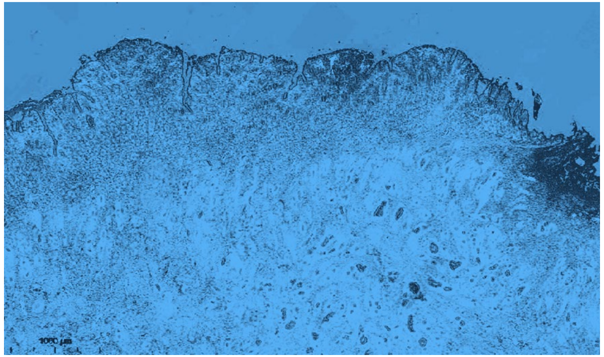


Evolving - Probit

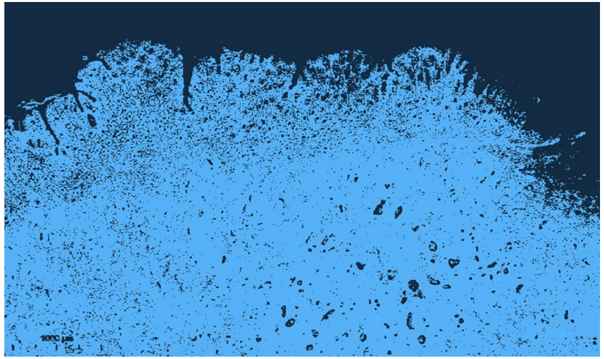
Colon Cancer



Probit



NN



Conclusion

- Probit and Neural Network methods succeeded in identifying areas of skin cancer
- Future Suggestions:
  - Tune Neural Network Model
  - Optimize image detecting threshold
  - Analyze noise effects on metrics
  - Increase the amount of training data used in detecting melanoma
  - Make significant adjustments to handle other types of cancer

Acknowledgments

- Dr. Gideon Zamba, ISIB Director and Project Mentor
- Logan Harris and Eliezer Santos-Leon, PhD Students
- Eldon Sorensen, Lab Teaching Assistant
- Terry Kirk, ISIB Coordinator
- ISIB Program sponsored by the National Heart Lung and Blood Institute (NHLBI), grant # HL-147231

NIH National Heart, Lung, and Blood Institute

References

Basterrech, S., Mesa, A., & Dinh, N.-T. (2015). Generalized linear models applied for skin identification in image processing. *Advances in Intelligent Systems and Computing*, 97–107. https://doi.org/10.1007/978-3-319-21206-7_9

Edgar, T. W., & Manz, D. O. (2017). Logistic regression. *Logistic Regression* - An Overview | ScienceDirect Topics. Retrieved July 16, 2022, from https://www.sciencedirect.com/topics/computer-science/logistic-regression

Faraway, J. J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models* (2nd ed.). CRC Press.

Johnson, R. A., & Wichern, D. W. (2002). Applied Multivariate Statistical Analysis (5th ed.). Pearson.

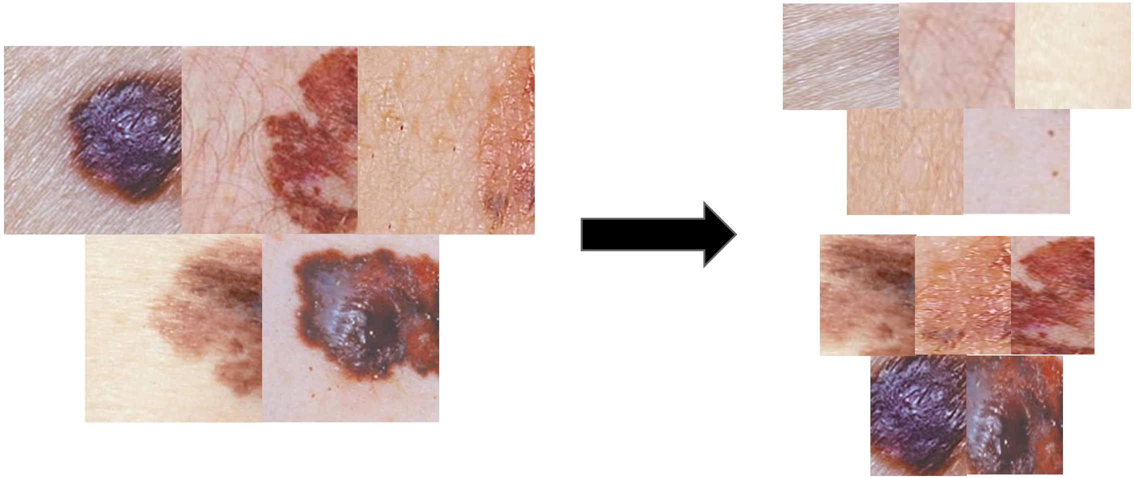
Marghoob, A. A., Halpern, A. C., & Reiter, O. (2022, January). *Melanoma*. The Skin Cancer Foundation. Retrieved July 12, 2022, from https://www.skincancer.org/skin-cancer-information/melanoma/#:~:text=Melanoma%20is%20a%20serious%20form,treated%20at%20an%20early%20stage.

Mayo Foundation for Medical Education and Research. (2021, March 10). *Slide show: Melanoma pictures to help identify skin cancer*. Mayo Clinic. Retrieved July 12, 2022, from https://www.mayoclinic.org/diseases-conditions/melanoma/multimedia/melanoma/sls-20076095?s=1

Stern, H. S. (1996). Neural Networks in Applied Statistics. *Technometrics*, 38(3), 220. https://doi.org/10.2307/1270604

Questions?

Training Set



IOWA