Interpreting Deep Learning: ISIC Melanoma

Matteo Russo, Dale Lee, Andrew Griffin

Department of Computer Science, Princeton University

Motivation and Goal

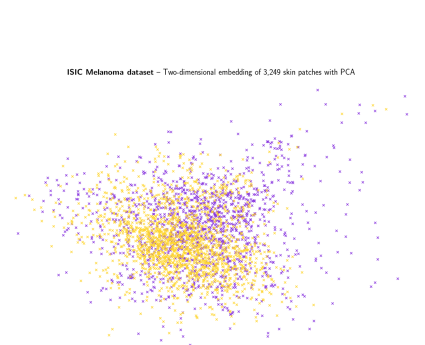
Machine Learning powered systems have revolutionized our data-governed world:

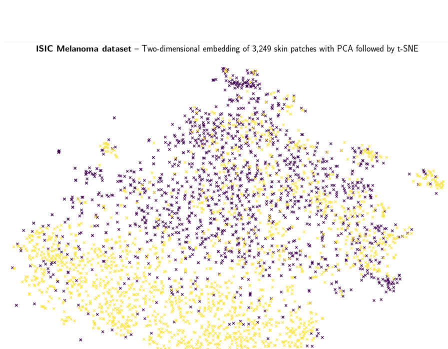
- **Motivation**: Can an algorithm save a life? In the field of dermatology, an early melanoma diagnosis can make the difference between life and death.
- **Goal**: The goal of this project is to design a strong and interpretable classifier for melanoma visual recognition.

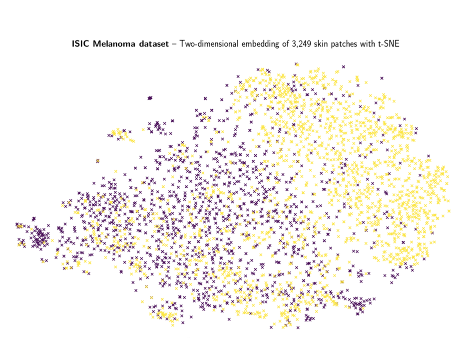
Related Work

Neural Networks are widely used in image classification—in particular, convolutional neural networks have been found to reduce computation time while reducing the chance of overfitting and providing improved results on smaller datasets. Convolutional Neural Networks in the biomedical imaging sphere have gained widespread success in classifying and segmenting images. A recent study published in Nature by Stanford researchers found that a CNN was able to outperform board-certified dermatologists when using a pre-trained GoogleNet Inception v3 CNN.

Data Visualization


PCA


PCA followed by t-SNE


t-SNE

Limitations and Future Work

Limitations:

1. The dataset is small and thus, we could achieve higher performances with more data.
2. Despite steps towards interpretability, there is a need to craft new methods for meaningful interpretation.

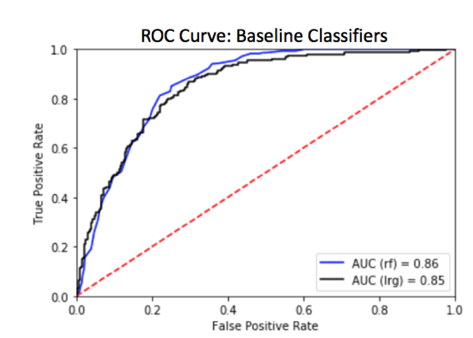
Future Work:

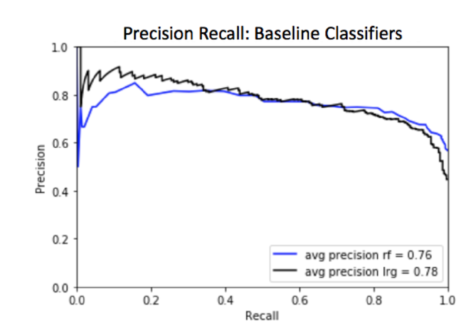
1. Expand analysis to larger imbalanced dataset.
2. Construct network architecture for higher resolution images.

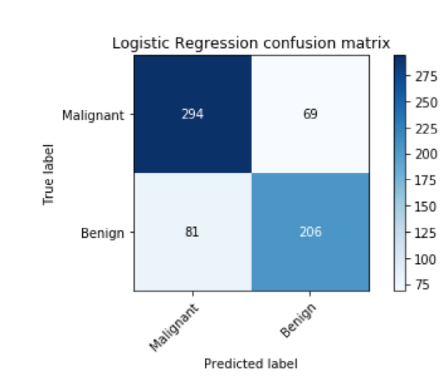
Baselines

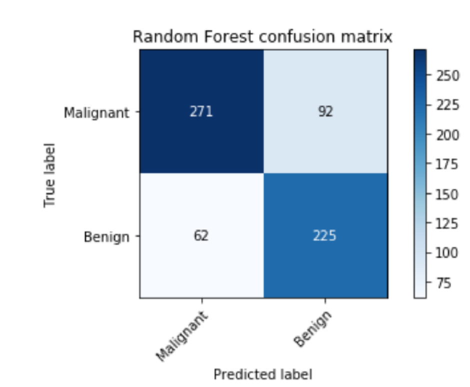
Through Grid-CV, we found by training the classofiers on 224×224 images

- **Random Forest** with a max tree depth of 5 and number of estimator trees of 20.
- **Logistic Regression** with ℓ_2 regularization.


ROC curve for LR and RF

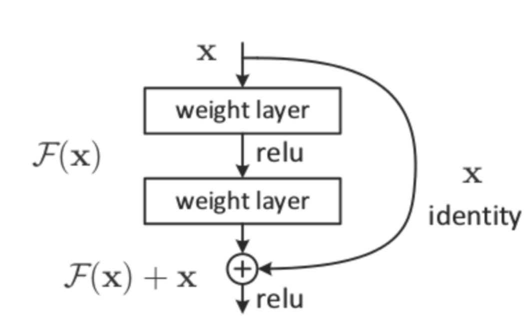

Precision-Recall curve for LR and RF


Logistic Regression Confusion Matrix

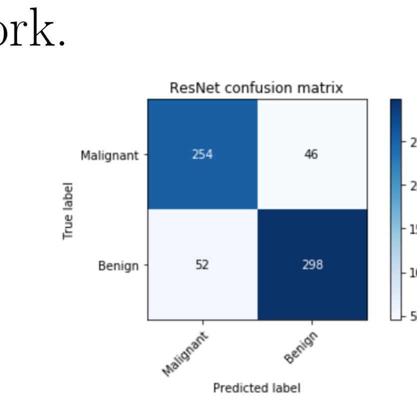

Random Forest Confusion Matrix

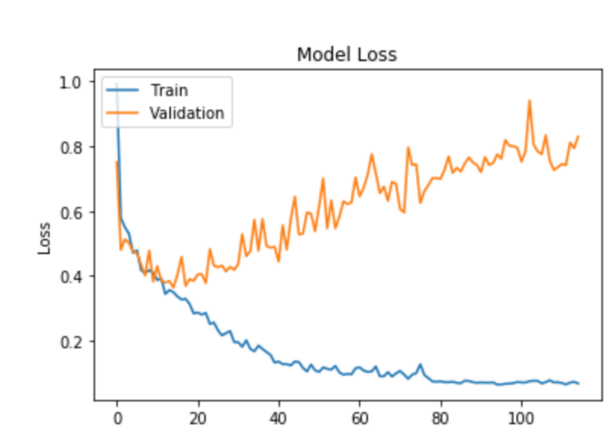
Deep Residual Network

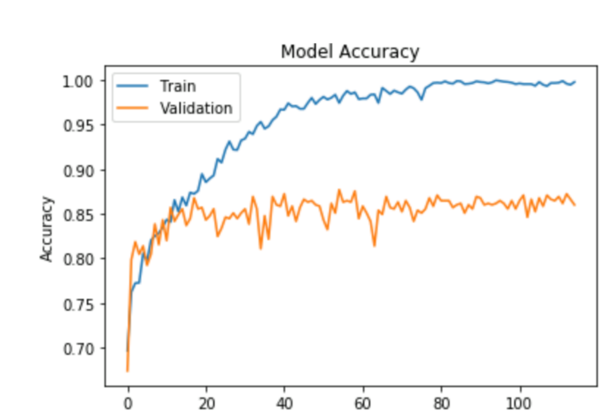
A considerable improvement in performance came through a ResNet with 6 joined convolutional, activation and batch normalzation layers, 3 maximum pooling and dropout layers, and 1 dense layer (MLP) at the end of the convolutional network.



ResNet Learning Block


ResNet Confusion Matrix


ResNet Train and Validation Loss

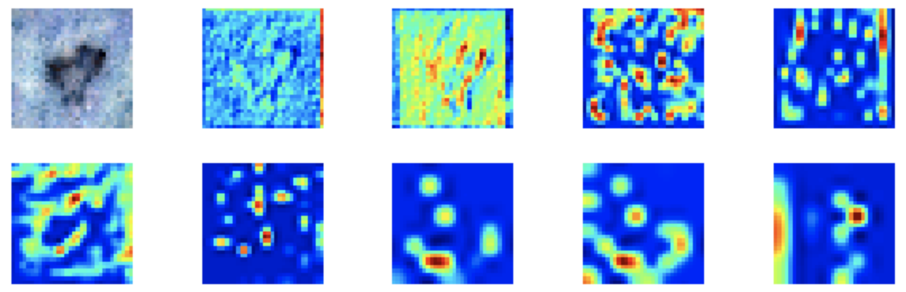

ResNet Train and Valid. Accuracy

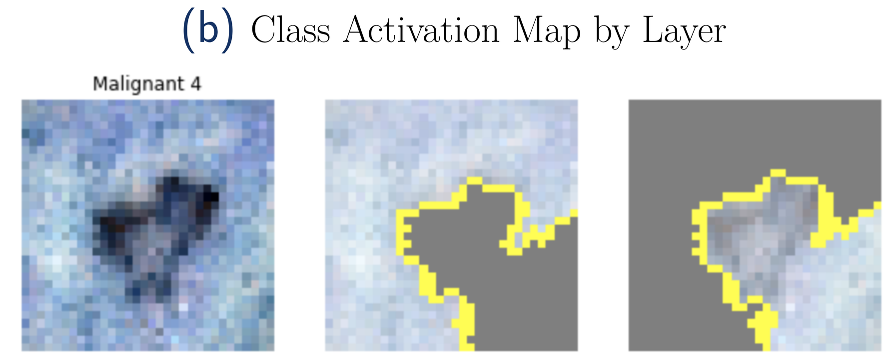
Interpretability Methodologies

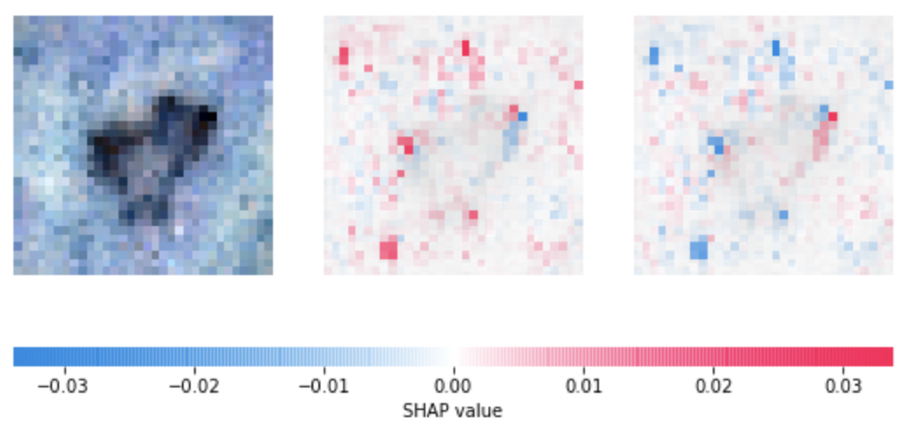
1. **CAM**: Backpropagate Activation Maps of loss in layers
2. **LIME**: $\lambda(\boldsymbol{x}) = arg\min_{g \in \mathcal{G}} \mathcal{L}(g, f, \pi_{\boldsymbol{x}}) + \Omega(g)$
3. **SHAP**: $\phi_i = \Sigma_{\mathcal{S} \subseteq \mathcal{F} \setminus \{i\}} \frac{|\mathcal{S}|!(|\mathcal{F}|-|\mathcal{S}|-1)!}{|\mathcal{F}|!}[f_{\mathcal{S}\cup\{i\}} - f_{\mathcal{S}}]$

Interpretability


(a) Melanoma


(b) Class Activation Map by Layer


(c) LIME Visualization


(d) SHAP Visualization

Summary of Results

Classifier	Accuracy	Precision	FPR	**FNR**
Logistic Regression	0.769	0.785	0.106	**0.125**
Random Forest	0.763	0.764	0.144	**0.092**
ResNet	0.872	0.764	0.080	**0.071**

Results from classifiers on test set with 80%-20% train-test split