Machine Learning for Medical Imaging

Margie Vollkommer
Advisors: Dr. Spanias, Vivek Sivaraman Narayanaswamy
8/5/2022 – Final Presentation
ASU SENSIP REU

Ira A. Fulton Schools of Engineering
ARIZONA STATE UNIVERSITY
REU Site sponsored in part by NSF I/UCRC Award 1540040

Pre-Training at ASU:

- Matlab/Python Introduction
- Machine Learning Introduction
    - Clustering/K Means
    - Classification + Confusion Matrices
    - Regression and Prediction
    - Neural Networks
    - PCA

Statement of Problem:

- Potentially useful application of machine learning is the diagnosis of ailments when provided with an X-ray or CAT scan
- Not enough accurately labeled datasets [1]
- Many existing datasets are not diverse, resulting in datasets that possess an implicit bias [1]
  - Additionally, many have a hidden stratification, causing a model to only perform well on the most common varieties of a given condition [2]

Abstract **– Machine learning has emerged as a potentially life-changing tool for medical diagnosis. The aim of this project was to investigate the potential methods that might best lend themselves to this task. The eventual goal will be to improve the accuracy of ailment detection using machine learning methods trained on labeled data consisting of medical scans.**

Index Terms **—Machine Learning, Medical Imaging, MedMNIST**

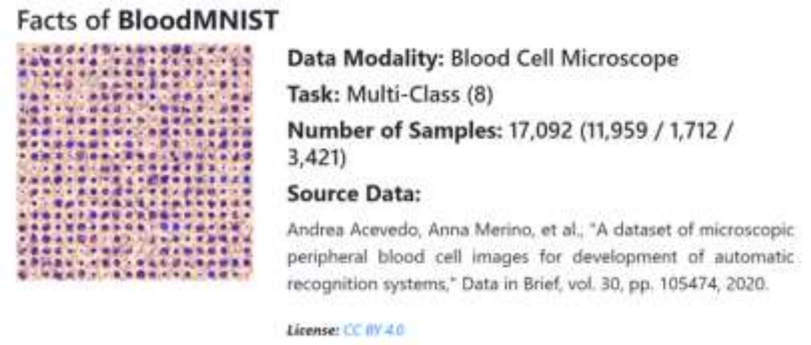
Proposed Solution:

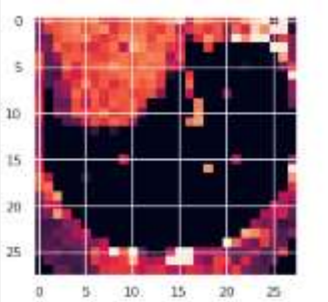
- MedMNIST v2 set was utilized in this project
  - Open-source dataset with thousands of 2D and 3D medical scans
  - Aimed to achieve accuracy of 90% or greater
  - Current accuracy is between 0.531 and 0.998, depending on the specific dataset
- Investigated different machine learning techniques that would be best suited to medical detection

Figure 1: Scans of human organs for use in medical classification through machine learning. (https://medmnist.com/)

Machine Learning Techniques Used:

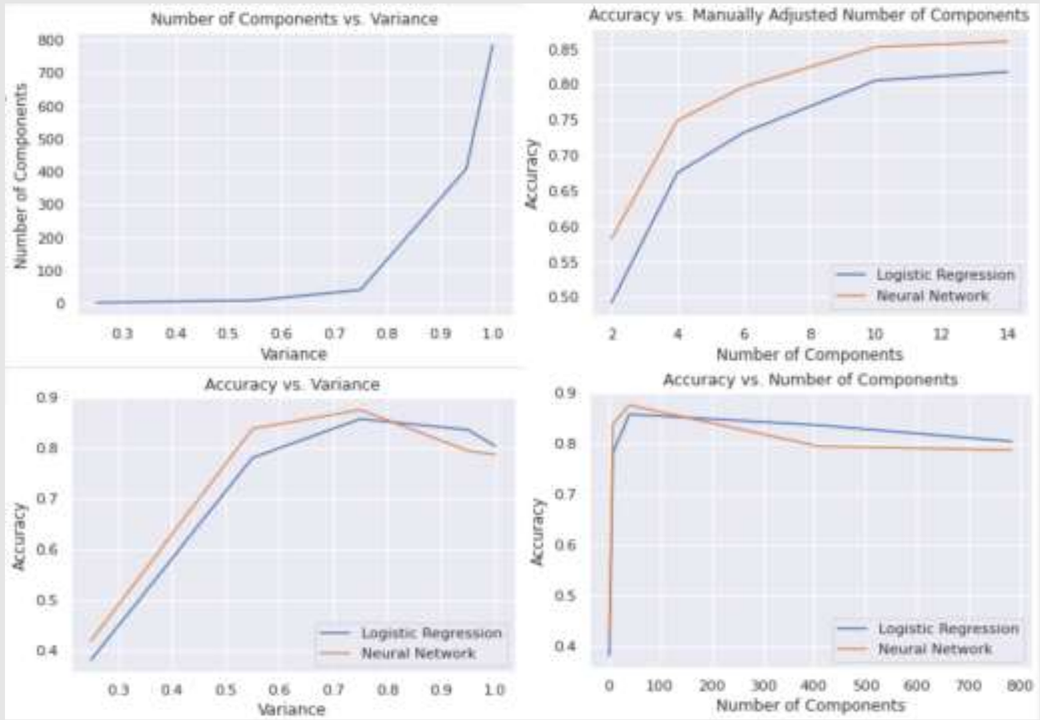
- Neural Network on Organ A
- Logistic Regression on Organ A
- Principal Component Analysis on OrganAMNIST dataset
- Convolutional Neural Network on BloodMNIST Dataset
- Principal Component Analysis on BloodMNIST Dataset





Results:

- Significant decrease in accuracy for both logistic regression and neural network models at a variance of ~0.6 or ~100 components
- OrganAMNIST Neural Network accuracy on 34,581 images: 97.09%
- BloodMNIST Convolutional Neural Network Accuracy: 65.73%

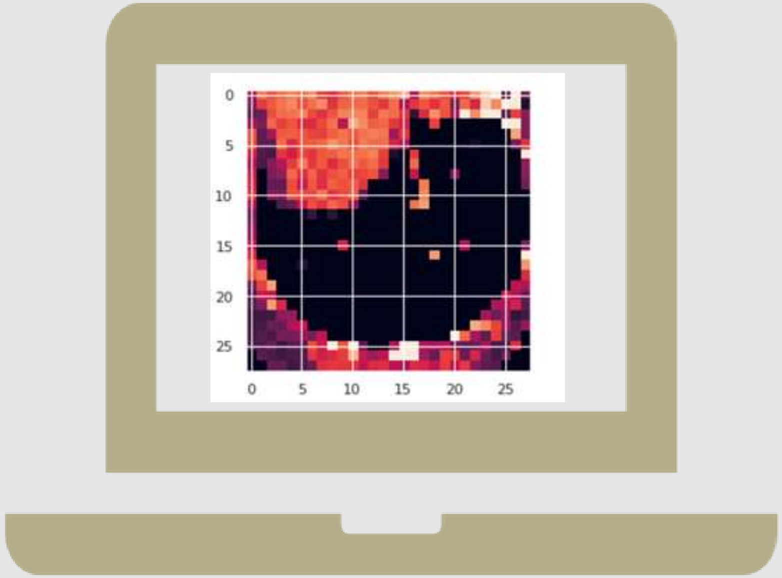


Conclusions:

- PCA can be used to reduce the variance of some medical images of organs to ~0.6 and still maintain accurate results
  - Will result in faster-running algorithms
- A neural network with a hidden layer of size 500 can produce an accuracy of upwards of 97% on some medical data (images of organs)
- A convolutional neural network can produce an accuracy of 65% on some medical data (images of blood)

Next Steps:

- Improve upon the Convolutional Neural Network accuracy
- Investigate other machine learning techniques that might be useful
- Investigate other medical image datasets



Reflection:

- Learned a lot about machine learning techniques
- Improved my coding skills
- Improved my presentation skills
- Made new friends!
- Chipped my two front teeth while moving from summer housing back to Tooker

References:

[1] Barragán-Montero, Ana, Umair Javaid, Gilmer Valdés, Dan Nguyen, Paul Desbordes, Benoit Macq, Siri Willems et al. "Artificial intelligence and machine learning for medical imaging: A technology review." Physica Medica 83 (2021): 242-256.

[2] Oakden-Rayner, Luke, Jared Dunnmon, Gustavo Carneiro, and Christopher Ré. "Hidden stratification causes clinically meaningful failures in machine learning for medical imaging." In Proceedings of the ACM conference on health, inference, and learning, pp. 151-159. 2020.

[3] Yang, Jiancheng, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. "Medmnist v2: A large-scale lightweight benchmark for 2d and 3d biomedical image classification." arXiv preprint arXiv:2110.14795 (2021).