Machine Learning for Medical Imaging

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





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



The AM Vew load Radius Tool	140	
Televel corners 127 ×	the the district form	1000
de chemic de deserve à l'Alem Servera Tapané l'ant des lagres Acesses	n 🕹 python"	
2000 gr	An informal introduction to Python	
And Ray South Pagewong junt nan g Tarine	Pytyre is a paper programming language that is volumly, finding and to increase to one in of specified partners, and supported for land a strong developer community and more the Republic (Wern support all however) of the Approx. Including well approximation, well assess the Hardware and a second partners and the second partners and the Approx. In a Language second partner was an extention consumption of partners and partners and the second partners and partners a strong second partners and support consumption of partners and partners and the second partners and partners as a strong second partners and support consumption of partners and partners and the second partners and partners are as a strong partners and partners	
	Name of the second s	
	This waithin up to a upplier contact version of the property productor backet, with prove other and executively. Because all the color is involve color rate, proceeding that have each code rate of the prime of appendic Pyther executive window.	
	This introduction is written for Partice 3.4.7 Set is generally applicable to other Pathon 3.4 versions.	
	The propert material is Copyright (2) 2001 2018 Pathoa Sufficient Property in	

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:

- $\circ\,$ 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

Facts of BloodMNIST -----Data Modality: Blood Cell Microscope Task: Multi-Class (8) Number of Samples: 17,092 (11,959 / 1,712 / 3,421) Source Data: Andrea Acevedo, Anna Merino, et al., "A dataset of microscopic peripheral blood cell images for development of automatic recognition systems," Data in Brief, vol. 30, pp. 105474, 2020. License: CC IIV 4.0 [] class ConvNet(nn.Module): # edited 2 line atm def __init__(self): super(ConvNet, self).__init__() self.conv1 = nn.Conv2d(2, 6, 5) # 1st self.pool = nn.MaxPool2d(2, 2) # kern self.conv2 = nn.Conv2d(6, 16, 5) # 2n self.fc1 = nn.Linear(16 * 5 * 5, 120) self.fc2 = nn.Linear(120, 84) # next self.fc3 = nn.Linear(84, 11) # final def forward(self, x): # -> n, 3, 32, 32 x = self.pool(F.relu(self.conv1(x))) x = self.pool(F.relu(self.conv2(x))) 5 10 15 20 x = x.view(-1, 16 * 5 * 5)x = F.relu(self.fc1(x)) x = F.relu(self.fc2(x)) x = self.fc3(x)return x model = ConvNet() #.to(device)

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:

 \circ 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).