Exploring the Power of Transfer Learning on Medical Image Datasets

Ananya Jain
Indraprastha Institute of Information Technology Delhi
New Delhi, India
ananya19408@iiitd.ac.in

Kirthana Natarajan
Indraprastha Institute of Information Technology Delhi
New Delhi, India
kirthana19053@iiitd.ac.in

Prashasti Agarwal
Indraprastha Institute of Information Technology Delhi
New Delhi, India
prashasti19075@iiitd.ac.in

Pritish Wadhwa
Indraprastha Institute of Information Technology Delhi
New Delhi, India
pritish19440@iiitd.ac.in

Raghav Nakra
Indraprastha Institute of Information Technology Delhi
New Delhi, India
raghav19083@iiitd.ac.in

Shashank Dargar
Indraprastha Institute of Information Technology Delhi
New Delhi, India
shashank19107@iiitd.ac.in

Shounak Ghatak
Indraprastha Institute of Information Technology Delhi
New Delhi, India
shounak19109@iiitd.ac.in

KEYWORDS

Transfer Learning, Medical Datasets, resnet, vision transformer, VGG, MEDMNIST

ACM Reference Format:

Ananya Jain, Kirthana Natarajan, Prashasti Agarwal, Pritish Wadhwa, Raghav Nakra, Shashank Dargar, and Shounak Ghatak. 2022. Exploring the Power of Transfer Learning on Medical Image Datasets. In *Proceedings of Computing For Medicine (CM Report).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 ABSTRACT

Medical data often contains various forms of image data such as x-ray scans, MRI scans, and CT scans. Computer-aided diagnosis systems can benefit from making use of such image data. Our problem statement is to perform classification tasks on medical image datasets, specifically Retinal OCT and Blood cell microscope images. In recent times, the application of computational or machine intelligence in medical diagnosis is a new trend for large medical data applications. Most of the diagnosis techniques in medical field are systematized as intelligent data classification approaches.

2 INTRODUCTION

Medical data classification is a prime data mining problem being discussed about for a decade that has attracted several researchers around the world. Most classifiers are designed so as to learn from the data itself using a training process, because complete expert knowledge to determine classifier parameters is impracticable. Providing accurate and accessible diagnoses is a fundamental challenge for global healthcare systems. [4]

In recent years, artificial intelligence and machine learning have emerged as powerful tools for solving complex problems in diverse domains. In particular, machine learning assisted diagnosis promises to revolutionise healthcare by leveraging abundant patient data to provide precise and personalised diagnoses. With the rapid development of digital image acquisition and storage technologies, image understanding by computer programs has become an attractive and active topic in the machine learning field and in application-specific studies [1]

Computer vision (CV) has a rich history spanning decades of efforts to enable computers to perceive visual stimuli meaningfully. Machine perception spans a range of levels, from low-level tasks such as identifying edges, to high-level tasks such as understanding complete scenes.

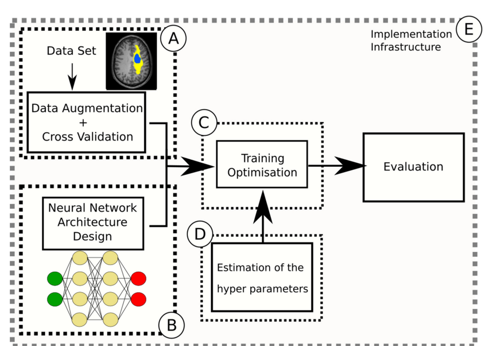

Figure 1: The different steps of a DL framework [3]

CM Report, Monsoon 2022 , IIITD

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Transfer learning was introduced to reduce the need for the annotation process by transferring the deep learning models with knowledge from a previous task and then by fine-tuning them on a relatively small dataset of the current task. [2]. So our aim was to exploit and apply transfer learning on 2 such medical datasets (BloodMNIST and OCTMNIST) using various models (Resnet, VGG, AlexNet, SqueezeNet, BEiT) to see how they perform relative to the benchmarks provided by Yang et al. [5].

3 DATASET

Our project uses the following datasets: **OCTMNIST** and **BLOODMNIST**. Both of these datasts are taken from MEDMNIST v2 [5], a large-scale MNIST-like collection of standardized biomedical images, including 12 datasets for 2D and 6 datasets for 3D. All images are pre-processed into **28 x 28 (2D) or 28 x 28 x 28 (3D)** with the corresponding classification labels, so that no background knowledge is required for users. Covering primary data modalities in biomedical images, MedMNIST v2 is designed to perform classification on lightweight 2D and 3D images with various data scales (from 100 to 100,000) and diverse tasks (binary/multi-class, ordinal regression and multi-label).

(1) BloodMNIST: The BloodMNIST is based on a dataset of individual normal cells, captured from individuals without infection, hematologic or oncologic disease and free of any pharmacologic treatment at the moment of blood collection. It contains a total of **17,092 images** and is organized into **8 classes** .
- We split the source dataset with a ratio of 7:1:2 into training, validation and test set.
- The source images with resolution 3×360×363 pixels are center-cropped into 3×200×200, and then resized into 3×28×28.
- Classes in the BloodMNIST dataset and their counts (See table 1):

Table 1: Table showing BloodMNIST classes and counts

Class Label	Label	Count
0	basophil	852
1	eosinophil	2181
2	erythroblast	1085
3	immature granulocytes	2026
4	lymphocyte	849
5	monocyte	993
6	neutrophil	2330
7	platelet	1643

(2) The OCTMNIST is based on a prior dataset of 109,309 valid optical coherence tomography (OCT) images for retinal diseases. The dataset is comprised of 4 diagnosis categories, leading to a multi-class classification task. We split the source training set with a ratio of 9:1 into training and validation set, and use its source validation set as the test set. The source images are gray-scale, and their sizes are (3841,536)×(277512). We center-crop the images and resize them into 1×28×28. (See table 2)

Table 2: Table showing OCTMNIST classes and counts

Class Label	Label	Count
0	choroidal neovascularization	33484
1	diabetic macular edema	10213
2	drusen	7754
3	normal	46026

4 METHODOLOGY

We approached the problem systematically, we began with the visualisation of the datasets to understand them more and then proceeded on application of transfer learning on these datasets using various models discussed later.

4.1 Exploratory Data Analysis

4.1.1 ***Class Count.*** Each class has different representation in the dataset. To understand the imbalance in the dataset, we visualize the counts of each class.See figure 4 and 3.

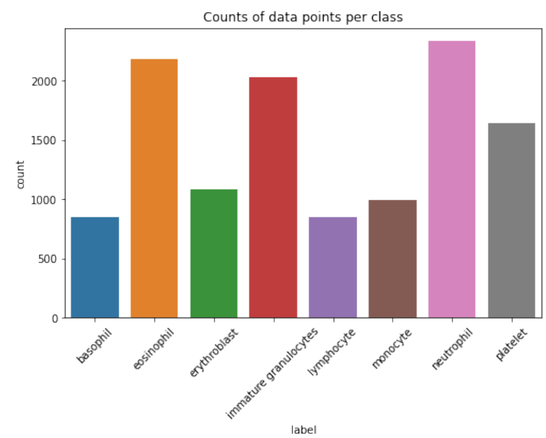

Figure 3: BloodMNIST Class Count

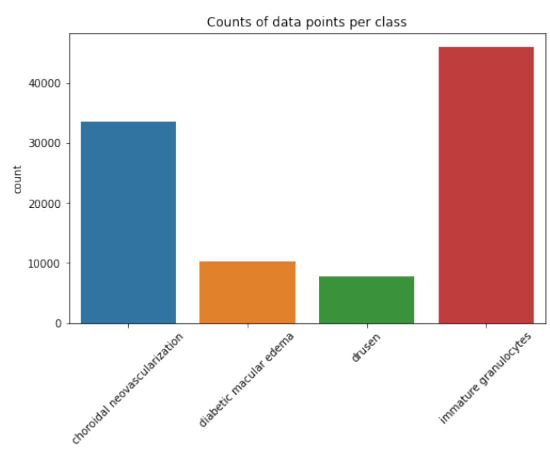

Figure 4: OCTMNIST Class Count

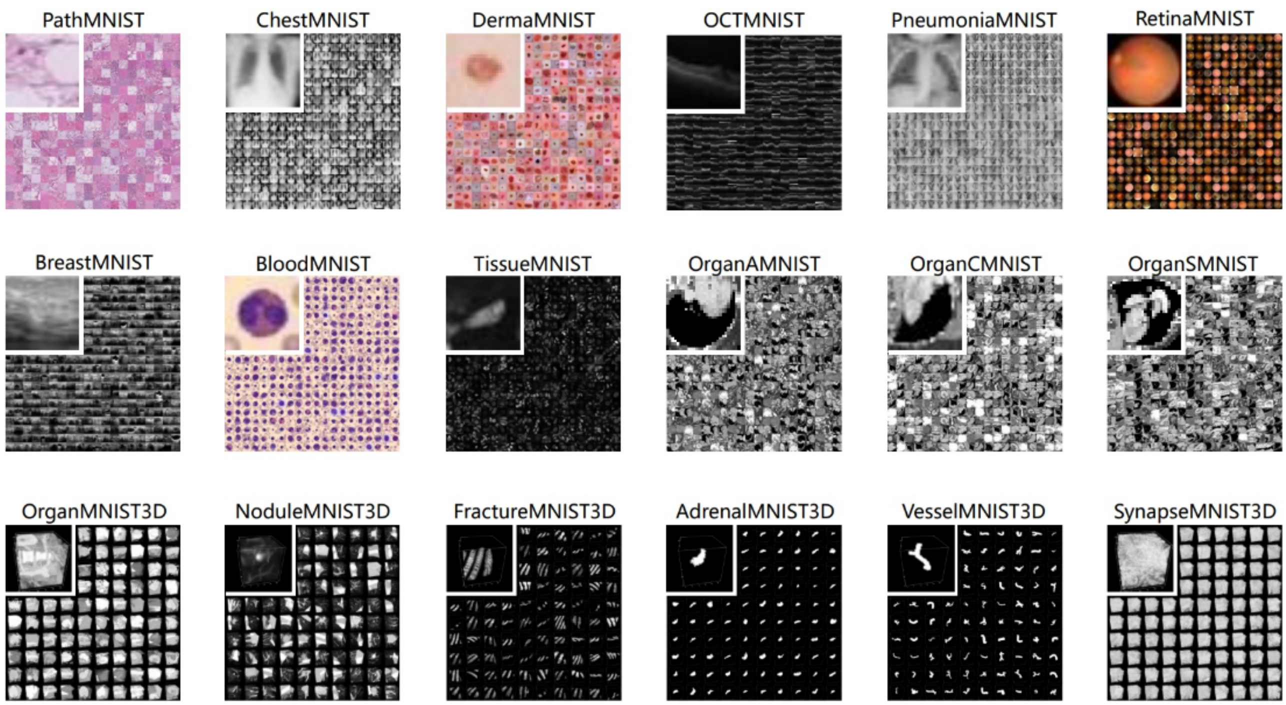

Figure 2: Variants available in the Medmnist Dataset

4.1.2 ***PCA***. Principal Component Analysis is used to transform high dimensional data to lower dimensions and can be used to visualize high dimensional data. See figure 5, 6, 7 and 8

Figure 5: BloodMNIST 2D PCA

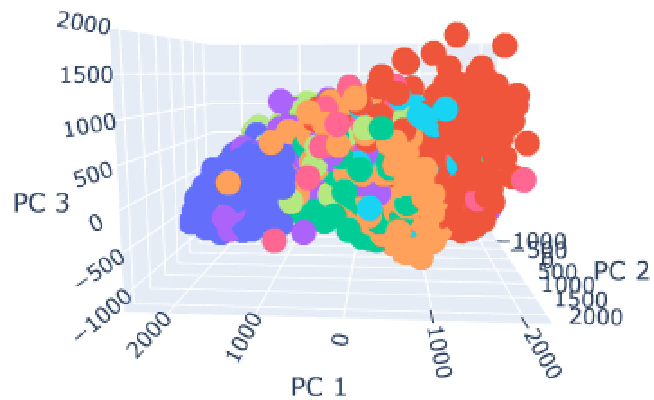

Figure 6: BloodMNIST 3D PCA

4.1.3 ***t-SNE***. t-SNE (t-Distributed Stochastic Neighbor Embedding) is a technique used to visualize high-dimensional datasets. It creates a mapping of the data points in lower dimensions by considering the local relationships between the points, which makes it a better choice to account for non-linear structure in the data as compared to PCA. See figure 9, 10, 11 and 12.

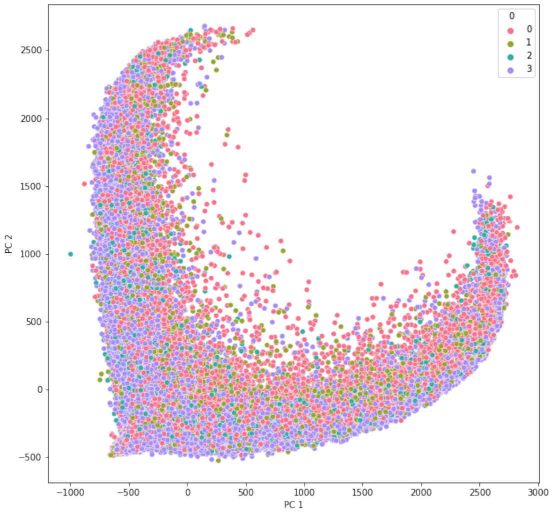

Figure 7: OCTMNIST 2D PCA

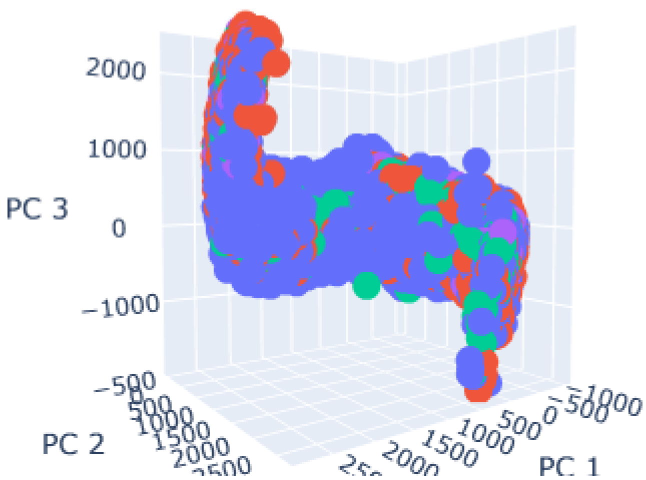

Figure 8: OCTMNIST 3D PCA

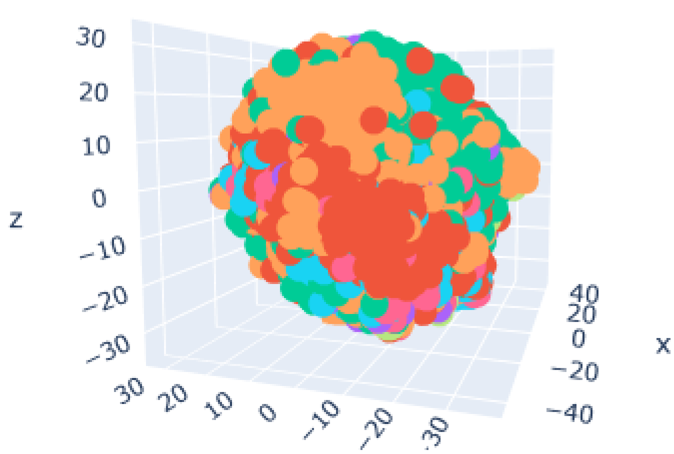

Figure 10: BloodMNIST 3D t-SNE

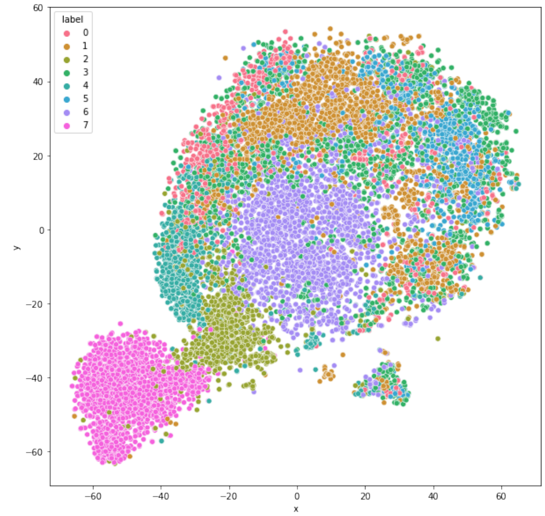

Figure 9: BloodMNIST 2D t-SNE

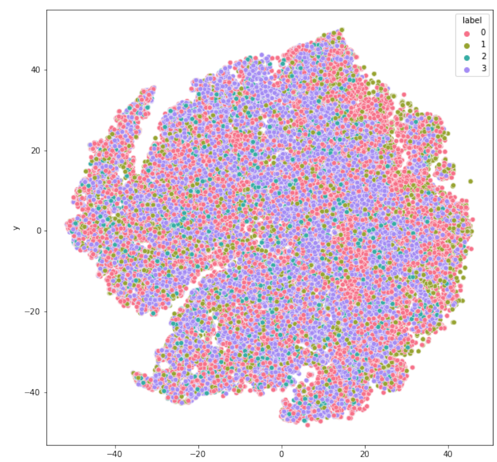

Figure 11: OCTMNIST 2D t-SNE

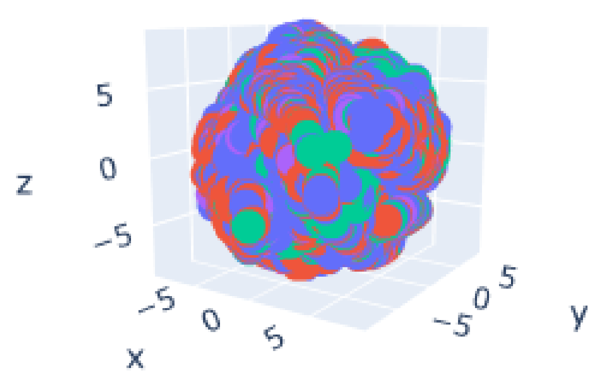

Figure 12: OCTMNIST 3D t-SNE

4.1.4 ***Mean Pixel***. Calculating the mean of the pixel values for each class in the data. See figure 13 and 14

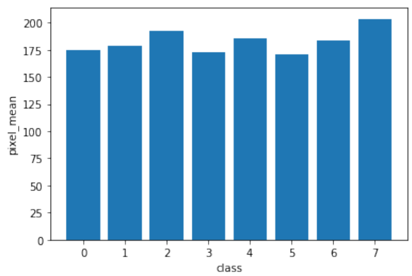

Figure 13: BloodMNIST Mean Pixel

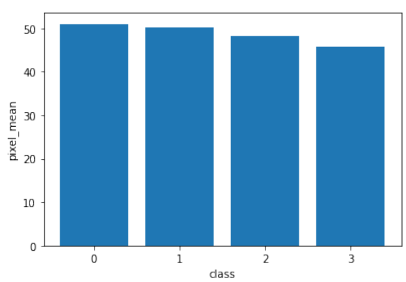

Figure 14: OCTMNIST Mean Pixel

4.1.5 ***Pixel Data***. Looking at the distribution of the pixels for each class in the data. Plot the histogram for each class (all the images pixels values count). See figure 15 and 16.

4.1.6 ***Montage***. Visualization of the images in the data. See Figure 18 and 17

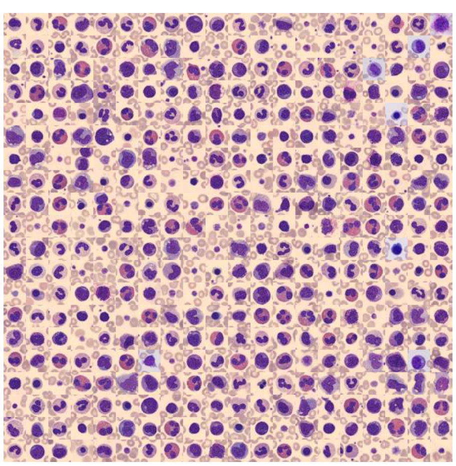

Figure 17: BloodMNIST Montage

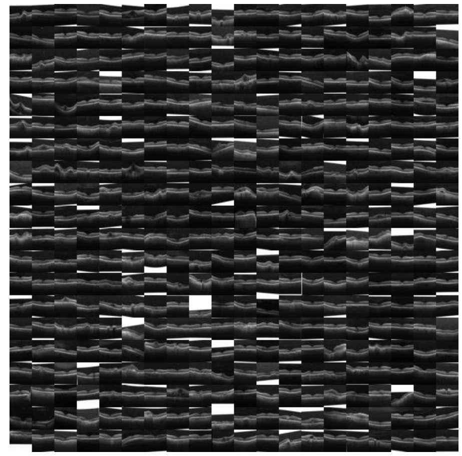

Figure 18: OCTMNIST Montage

4.2 Preprocessing

The images were originally 28x28. The images were resized to 224x224 for Alexnet and 227x227 for VGG. Since OCTMNIST images are black and white, the channels were replicated thrice to allow models to operate.

4.3 Transfer Learning

Reusing a model that has already been trained on a different problem is known as transfer learning in machine learning. With transfer learning, a computer can use its understanding of one activity to better generalise about another. For instance, you may leverage the skills a classifier learned to identify drinks while training it to predict whether an image contains food.

In our project we use the following pre-trained models to perform our transfer learning task on our datasets. The following models were used:

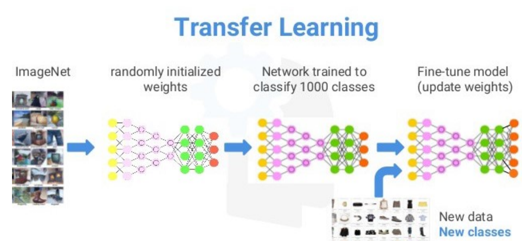

Figure 19: Visual Representation of the concept of transfer learning

4.3.1 ***AlexNet***. AlexNet is the name of a convolutional neural network (CNN) architecture. AlexNet competed in the ImageNet Large Scale Visual Recognition Challenge on September 30, 2012. AlexNet contained eight layers; the first five were convolutional layers, some of them followed by max-pooling layers, and the last three were fully connected layers. It used the non-saturating ReLU activation function, which showed improved training performance over tanh and sigmoid.

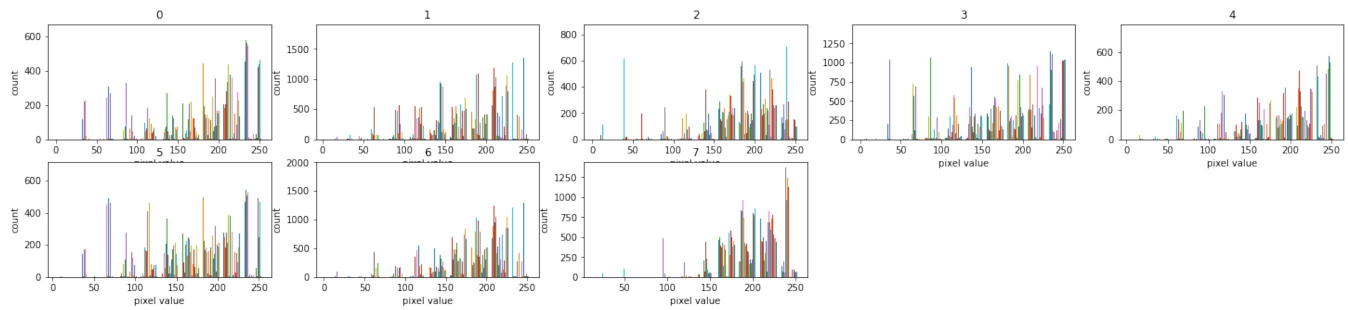

Figure 15: BloodMNIST Pixels

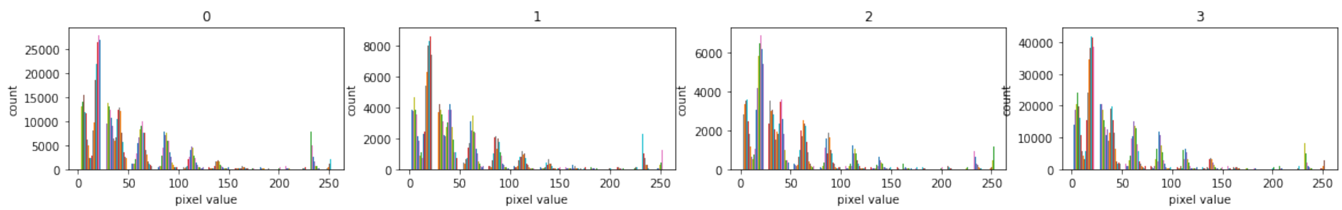

Figure 16: OCTMNIST Pixels

4.3.2 *Beit*.

The BEiT model was proposed in BEiT: BERT Pre-Training of Image Transformers. Inspired by BERT, BEiT is the first paper that makes self-supervised pre-training of Vision Transformers (ViTs) outperform supervised pre-training. Rather than pre-training the model to predict the class of an image , BEiT models are pre-trained to predict visual tokens from the codebook of OpenAI's DALL-E model given masked patches.

4.3.3 *VGG*.

VGG stands for Visual Geometry Group; it is a standard deep Convolutional Neural Network (CNN) architecture with multiple layers. The "deep" refers to the number of layers with VGG-16 or VGG-19 consisting of 16 and 19 convolutional layers. The VGG architecture is the basis of ground-breaking object recognition models. Developed as a deep neural network, the VGGNet also surpasses baselines on many tasks and datasets beyond ImageNet. Moreover, it is now still one of the most popular image recognition architectures.

4.3.4 *Inception*.

Inception v3 is a convolutional neural network for assisting in image analysis and object detection, and got its start as a module for Googlenet. It is the third edition of Google's Inception Convolutional Neural Network, originally introduced during the ImageNet Recognition Challenge. The design of Inceptionv3 was intended to allow deeper networks while also keeping the number of parameters from growing too large: it has "under 25 million parameters", compared against 60 million for AlexNet.

4.3.5 *Resnet*.

A residual neural network (ResNet) is an artificial neural network (ANN). It is a gateless or open-gated variant of the HighwayNet, the first working very deep feedforward neural network with hundreds of layers, much deeper than previous neural networks.

4.3.6 *SqueezeNet*.

SqueezeNet is a convolutional neural network that is 18 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.

5 RESULTS

The above discussed models where fine tuned on the OCTMNIST and Blood MNIST datasets. We performed hyperparameter tuning (number of epochs, learning rate) for each model to obtain the best performance for each (exact parameters can be found in the submitted codes). The obtained results are as follows: See table 3

Table 3: Results on BLOODMNIST and OCTMNIST datasets for different models

Model	Specification	Blood MNIST	OCT MNIST
AlexNet		0.90	0.72
BEiT		**0.9664**	**0.818**
Inception		0.95	0.73
VGG	VGG-16	0.73	0.67
	VGG-19	0.87	0.72
ResNet	Resnet-18	0.80	0.69
	Resnet-34	0.92	0.72
	Resnet-50	0.92	0.64
	Resnet-101	0.93	0.64
	Resnet-152	0.82	0.68
SqueezeNet		0.73	0.58

6 CONCLUSION

In this project, we explored the power of transfer learning on medical image datasets using the BloodMNIST and OCTMNIST datasets from the MEDMNISTv2 database. Visualizing the data showed us that the data was not clearly separable, possibly requiring complex deep learning models to achieve good results. We applied several models with different specifications including AlexNet, BEiT, Inception, VGG, ResNet and SqueezeNet by fine-tuning them on our chosen datasets. We found the BEiT model (BERT Pre-Training of Image Transformers) to perform best on both the datasets, giving accuracy of 0.9664 for BloodMNIST and 0.818 for OCTMNIST. This beats the baseline performance obtained by Yang et al.[5] on OCTMNIST (ACC: 0.776) and matches that of BloodMNIST (ACC: 0.966). Thus, we can see that transfer learning can be a powerful tool in the classification of medical image datasets.

REFERENCES

[1] Thomas Davenport and Ravi Kalakota. 2019. The potential for artificial intelligence in healthcare. *Future Healthcare Journal* 6, 2 (June 2019), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

[2] Andre Esteva, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher. 2021. Deep learning-enabled medical computer vision. *npj Digital Medicine* 4, 1 (Jan. 2021). https://doi.org/10.1038/s41746-020-00376-2

[3] Félix Renard, Soulaimane Guedria, Noel De Palma, and Nicolas Vuillerme. 2020. Variability and reproducibility in deep learning for medical image segmentation. *Scientific Reports* 10, 1 (Aug. 2020). https://doi.org/10.1038/s41598-020-69920-0

[4] C. V. Subbulakshmi and S. N. Deepa. 2015. Medical Dataset Classification: A Machine Learning Paradigm Integrating Particle Swarm Optimization with Extreme Learning Machine Classifier. *The Scientific World Journal* 2015 (2015), 1–12. https://doi.org/10.1155/2015/418060

[5] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. 2021. MedMNIST v2: A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification. *arXiv preprint arXiv:2110.14795* (2021).