Interpretable AI For Protein Expression Images and Associated Clinical Metadata

SUSHANTH SHIVPURA RAMESH
School of Computing, Newcastle University

INTRODUCTION

- Mitochondrial diseases are genetically inherited disorders caused due to mitochondrial dysfunction and currently, there is no cure for mitochondrial diseases which affects 1 in 5000 people[1,2].
- Bio medical scientists at Wellcome Center for Mitochondrial Research (WCMR) have identified 10 protein targets from all five by oxidative phosphorylation (OXPHOS) complexes using skeletal muscle tissues of patients with genetically characterized mitochondrial disease[3].
- The pseudo images from Image Mass Cytometry (IMC) is then segmented, quantified using various statistical methods, and analyzed to understand relationship amongst various protein expression levels and disease pathology. This method is not very useful due to high dimensionality of data and limitation of statistical methods with high throughput image data[3].
- In this project we aim to use Interpretable and Explainable Artificial Intelligence approaches to help us understand the underlying pathology of mitochondrial diseases with the features extracted by the CNN network from the pseudo protein expression images.

METHODOLOGY

- To fine tune and use existing pre-trained models VGG-16 and ResNet-101v2 to classify the single protein images into patient and control classes and apply class saliency Map to visualize and understand the basics of prediction and disease pathology.
- To build and train a multi-channel model to classify 10-channel protein image into patient and control classes.

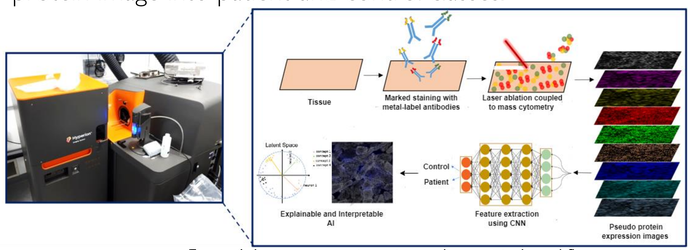

Figure 1: Image mass cytometry and associated workflow

- Pseudo protein expression image data provided by WCMR which is categorized into patient (P)and control(C)[3].

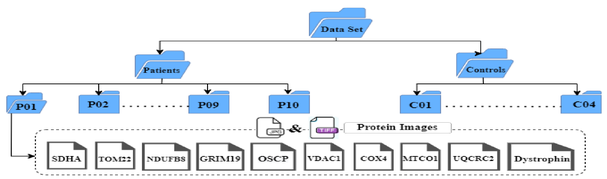

Figure 2: Data structure representing 10-protein expression images in each patient and control data

RESULTS

➢ Single Protein Image Classification

- Pre-trained VGG-16 and ResNet-101v2 with ImageNet weights and fine tuned Fully connected layers with Dense layers and dropout's as shown in the architecture below are trained on TOMM22 single protein images
- Performance metrics for both the models are summarized in the table below.
- VGG-16 has the best Accuracy and F1-scores with 95% and 0.94 Respectively

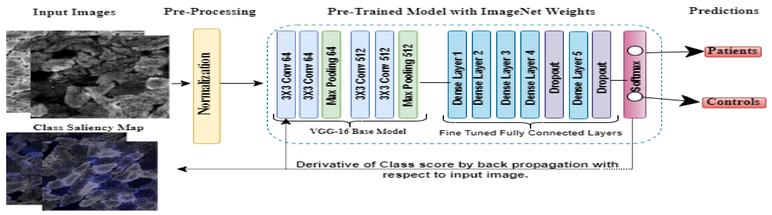

Figure 3: Fine tuned VGG-16 architecture for single protein image classification with class saliency implementation

Model	Precession	Recall	F1-Score	Accuracy
VGG-16	0.96	0.95	0.94	0.95
ResNet-101v2	0.87	0.86	0.87	0.89

Table1: Performance metrics for single protein image classification models

➢ Class Saliency Map

- Class saliency map is one of the post hoc explainable methods implemented to highlight the pixels used by the network to classify the images.
- Gradient of the SoftMax output with respect to the input of VGG-16 model is calculated to produce the saliency as shown in figure 3
- Brighter the pixel colour, the higher the pixel contributes to the prediction, which helps in understanding the disease pathology.

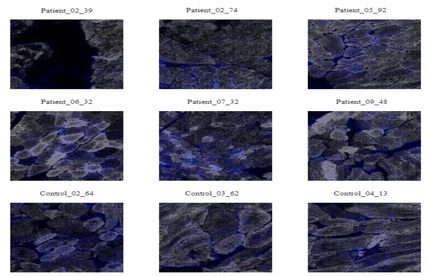

Figure 4: Saliency Map produced for both Patient and control images of TOM22 protein expression images

➢ Stacked Multi-Channel Image Classification

- Pre-trained VGG-16 and VGG-19 models are modified to train with 10-channel stacked protein images as shown in figure 5 .
- The First three channels use ImageNet weights, and the last 7 channels use the average of the ImageNet weights used in the first three channels
- VGG-16 has better performance than VGG-19 model and performance metrics for both the models are as shown in the table 2 below

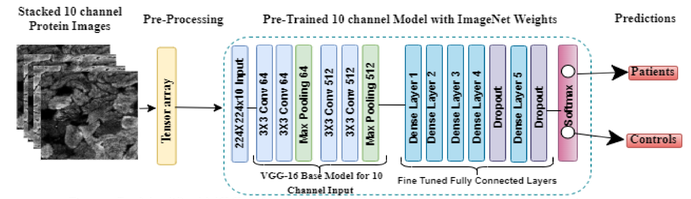

Figure 5: Modified VGG-16 model for 10–channel protein expression images

Model	Precession	Recall	F1-Score	Accuracy
VGG-16	0.96	0.96	0.95	0.98
VGG-19	0.94	0.91	0.92	0.93

Table2: Performance metrics for 10-channel protein image classification model

DISCUSSION

- Fine tuned VGG-16 has a very good model performance for both single protein image classification as well as stacked 10-channel image classification model.
- If the bio-medical scientists at WMRC can confirm that the model is looking into the right areas with the saliency map images, then we can conclude that deep learning with interpretability techniques might help us understand the underlying pathology of mitochondrial diseases from protein expression images.
- One of the limitation identified was, there is a need to collect more control samples, as there is data imbalance while training the model.
- In future we could try to use meta data to explore more interpretable methods such as Interpretable Image Recognition with Hierarchical Prototypes [4]

[1]Grainne S. Gorman, Andrew M. Schaefer, Yi Ng, Nicholas Gomez, Emma L. Blakely, Charlotte L. Alston, Catherine Feeney, Rita Horvath, Patrick Yu-Wai-Man, Patrick F. Chinnery, Robert W. Taylor, Douglass M.Turnbull, and Robert McFarland. Prevalence of nuclear and mitochondrial dna mutations related to adult mitochondrial disease. Annals of Neurology, 77(5):753–759, March 2015. doi: 10.1002/ana.24362.
imaging mass cytometry. Scientific Reports, 10(1), September 2020. doi: 10.1038/s41598-020-70885-3.
[2] Salvatore DiMauro. Mitochondrial diseases. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1658(1-2):80–88, July 2004.
doi: 10.1016/j.bbabio.2004.03.014.
[3] Charlotte Warren, David McDonald, Roderick Capaldi, David Deehan, Robert W. Taylor, Andrew Filby, Doug M. Turnbull, Conor Lawless, and Amy E. Vincent. Decoding mitochondrial heterogeneity in single muscle fibres by imaging mass cytometry. Scientific Reports, 10(1), September 2020. doi: 10.1038/s41598-020-70885-3.
[4] Peter Hase, Chaofan Chen, Oscar Li, and Cynthia Rudin. Interpretable image recognition with hierarchical prototypes, 2019. https://arxiv.org/abs/1906.10651

S.Shivpura-Ramesh2@newcastle.ac.uk C1044105