Newcastle University

INTERPRETABLE AI FOR PROTEIN EXPRESSION IMAGES AND ASSOCIATED CLINICAL METADATA

Sushanth Shivpura Ramesh
MSc Data Science
C1044105

Supervisors

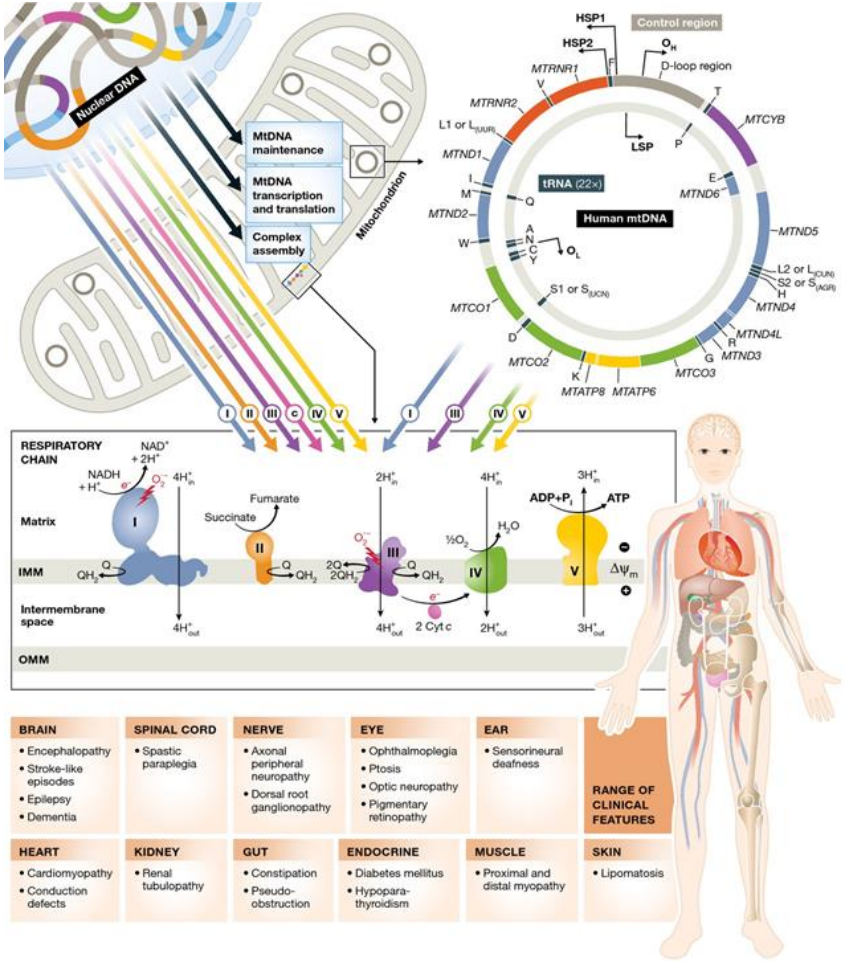
Stephen McGough
Atif Khan

AIM

To use Interpretable and explainable artificial intelligence approaches such as saliency maps and Neural Disentanglement to comprehend the underlying pathology of mitochondrial diseases with the features extracted by deep learning models.

Mitochondrial Diseases

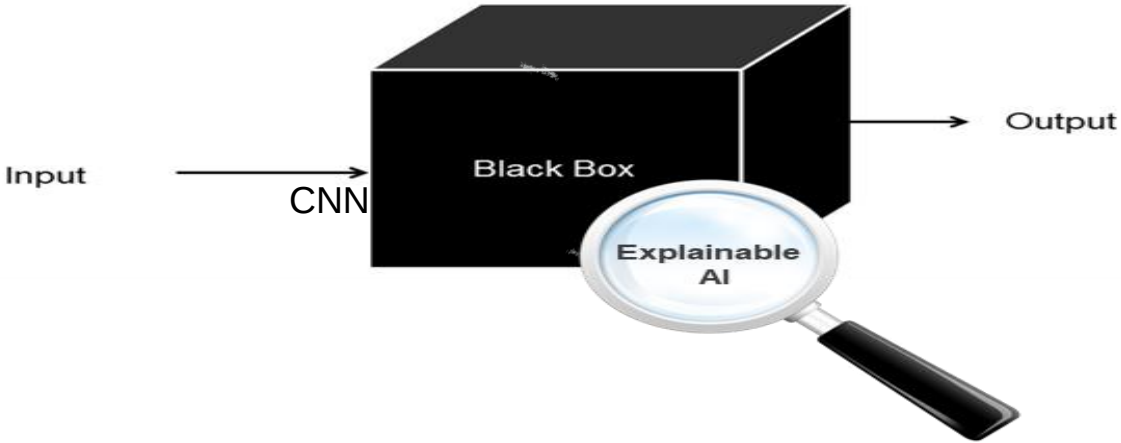
- Untreatable and affect 1 in 5,000
- Manifest as a result of mutations in genes that encode mitochondria
- Highly heterogeneous and adversely affect high energy demanding cells e.g. skeletal muscle cells and neurons
- Studied by finding relationship between genome, OXPHOS proteins & disease symptoms



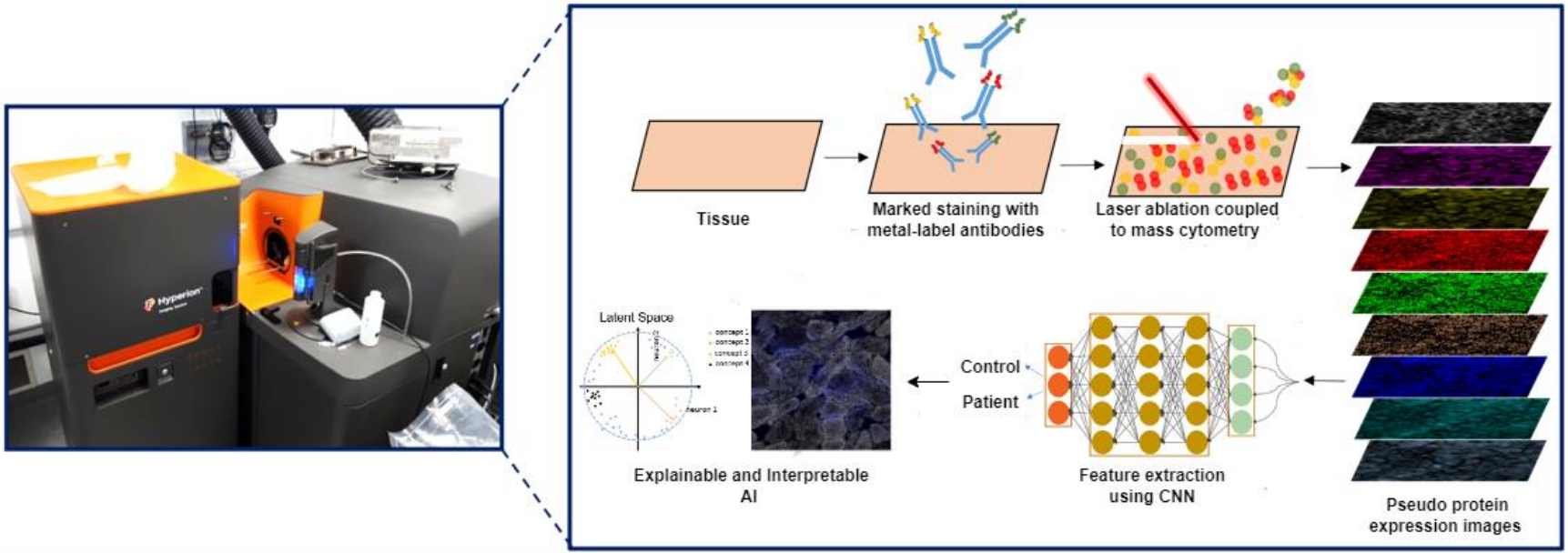
Credit: EMBO Molecular Medicine

Why Explainability and Interpretability ?

- High-stake decision-making settings such as health care, as in our case requires the model to be interrogated
- Networks are very complex and sophisticated
- No clear explanation



Protein expression data: Image mass cytometry (IMC)

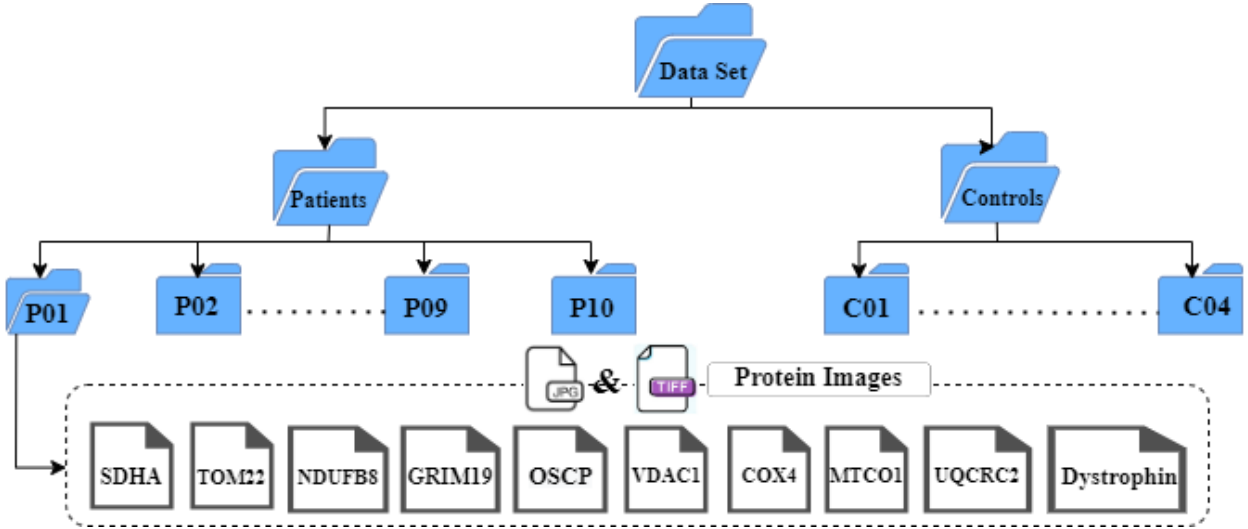


IMC Process A) Hyperion imaging mass cytometer B) Diagrammatic representation of IMC experiment and associated analysis workflow

Credit: Warren et al.,(2020) Nature Communications . doi: 10.1038/s41598-020-70885-3

Dataset

- Images from Image Mass Cytometry (IMC)
- Patient : 10 Samples
- Control : 4 Samples



Multi-channel protein images (IMC)

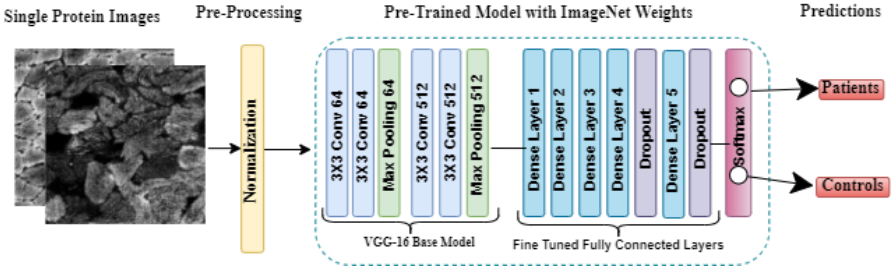
Data Preprocessing

- Training data : Single Protein image TOM22
- Image is divided into sub images to increase the training data.
- Normalized Images before training
- Split into Train, Test, and Validation sets 80:10:10 ratio

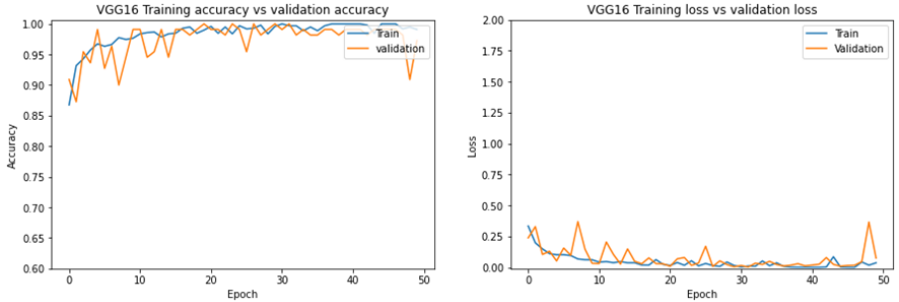
970 images ------ > Training Set
120 images -------> Validation Set
124 images --------> Test Set

Single Protien Image Classification

- CNN with Transfer Learning – ImageNet Weights
- VGG16 and ResNet101v2 Models
- Best Model – VGG16



VGG16 architecture for single protein image classification



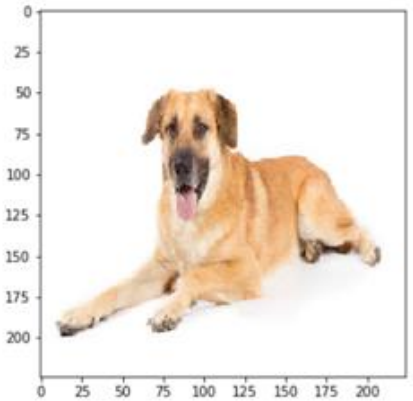
Accuracy and Loss curves for VGG16 single-channel model

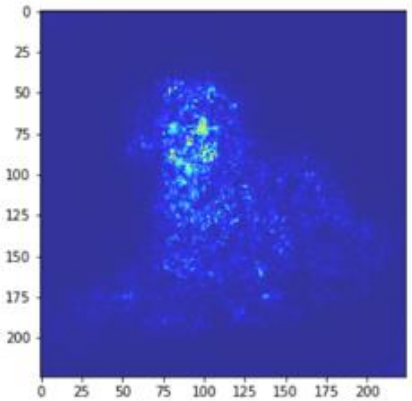
Model	Precision	Recall	F1-Score	Accuracy
VGG16	**0.95**	**0.94**	**0.94**	**0.95**
ResNet101 V2	0.91	0.89	0.90	0.92

Model performance parameters for single protein image classification models

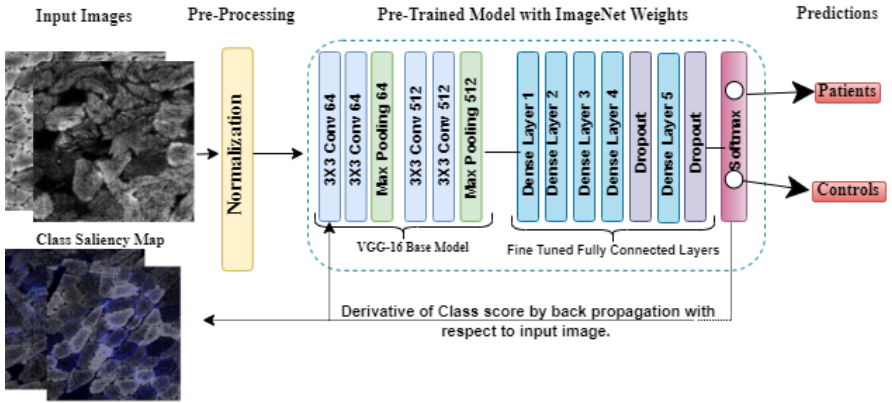
Class Saliency Map

- Uses gradient calculation to assign an importance score to an individual feature which reflects the influence on the model prediction
- Saliency maps helps us visualize and identify the features used by the network to train the model

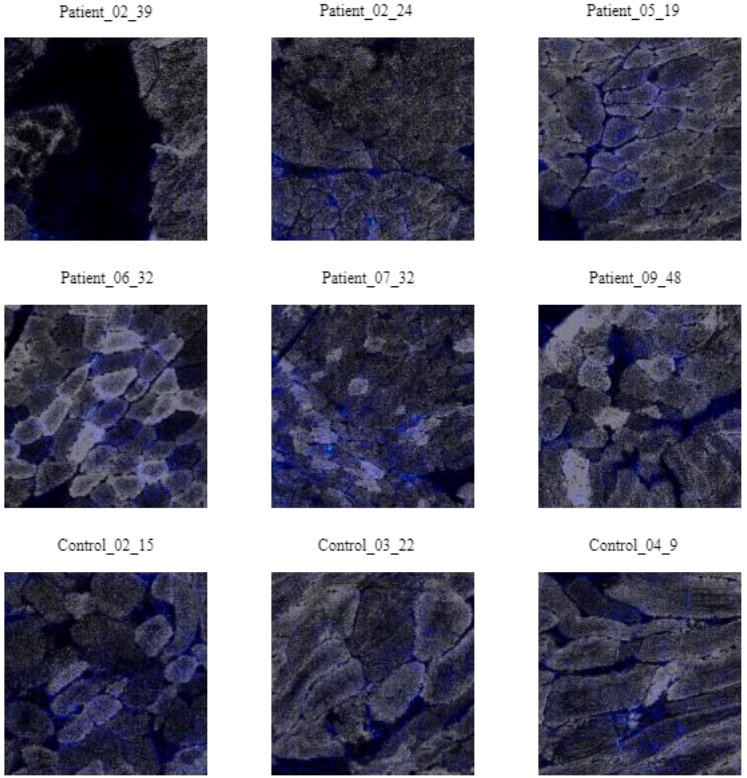




Saliency Map Architecture and Sample Output



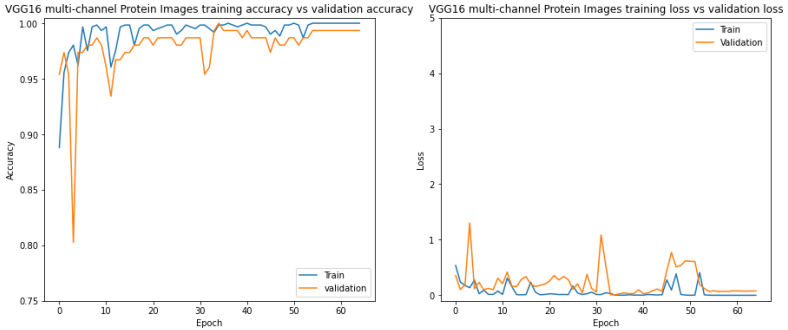
Class saliency Map implementation on VGG-16 model



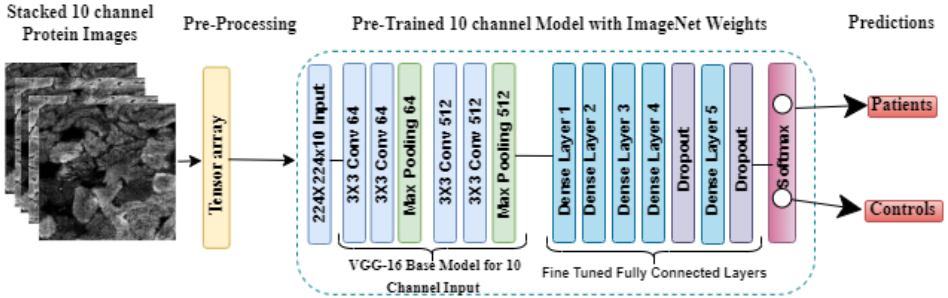
Class Saliency Map implementation on TOM22 single protein images

Mutli-Channel Protien Image Classification

- CNN models are modified to accomdate for 10-channel Protien Images with Transfer Learning – ImageNet Weights
- VGG16 and VGG19 Models
- Best Model – VGG16



Accuracy and Loss curves for VGG16 Multi-channel model



VGG16 architecture for multi-channel protein image classification

Model	Precision	Recall	F1-Score	Accuracy
VGG16	**0.96**	**0.96**	**0.95**	**0.98**
VGG19	0.94	0.91	0.92	0.93

Model performance parameters for multi-channel protein image classification models

Results and Future Work Recommendations

- 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

- In future we could try to use meta data to explore more interpretable methods such as Interpretable Image Recognition with Hierarchical Prototypes (HPnet)

- Further work can include the implementation of saliency maps on the 10-channel image classification output to visualize the features exacted in each channel of the image and explore more interpretable methods such as Layer-wise relevance propagation (LRP), Interpretable Image Recognition with Hierarchical Prototypes (HPnet).

THANK YOU !!!!!