

Automatic Sample Decomposition & Detection of Proteinopathy: Using Colourimetry and Machine Learning

Phase 2 Delivery Report

Bradley Pearce, Peter Coetzee

Polygeist LTD, UK

March 2023

Commissioned by NHS AI Lab Skunkworks.

In Partnership with

PARKINSON'S^{UK}
CHANGE ATTITUDES.
FIND A CURE.
JOIN US.

**Imperial College
London**

ACE Accelerated
Capability
Environment

© 2023 Polygeist, some rights reserved.

Conditions Of Supply & Licence

This document was created and supplied by Polygeist LTD as a deliverable for the Accelerated Capability Environment C319.1 project, funded by NHS Transformation Directorate AI Lab Skunkworks. This document is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International licence (CC BY-ND 4.0); a copy of this licence is available here: <https://creativecommons.org/licenses/by-nd/4.0/legalcode>. Images presented here that are derivative works of biological imagery, are used with permission, and are copyright © 2022 Parkinson's UK, and © 2022 Imperial College London, all rights reserved. Parkinson's UK logo remains © 2022 Parkinson's UK. The Polygeist logo and title page graphic are © 2022 Polygeist LTD. All foreground intellectual property is transferred with some rights reserved by Polygeist to the Crown, and all background intellectual property remains that of the respective owners.

Executive Summary

The identification of Parkinson's Disease (PD) from post-mortem brain slices is time consuming for highly trained neuropathologists, often taking many hours per case. Accurate classification and stratification of PD is critical for confirmation that the brain donor suffered from PD, or in the case of healthy donors, does not have early PD pathology. This permits the use of donor brains in research studies to better understand the causes of PD and foster drug development.

Parkinson's UK Brain Bank, at Imperial College London, has produced a dataset containing digitised images of brain sections immunostained for the protein alpha-synuclein (α -syn), the pathological marker of PD; along with control cases from healthy donors, and those with comorbidities such as Alzheimer's disease, or Multiple System Atrophy. This dataset is much larger (over 400 cases), more consistent, and of higher quality (all have been stained with the same protocol and imaged within the same laboratory) than has been documented elsewhere in the literature; including those found in a meta-analysis study on detection of neurological disorders containing over 200 papers (Lima et al., 2022). In conjunction, those cases with comorbidities have also been immunostained for beta-amyloid (A β) and Tau proteins; which represent markers for a variety of amyloid and Tau pathologies, and are present to a lesser degree in ageing brains.

This project extends and expands on previous work detecting α -syn, and by extension Parkinson's Disease, in brain samples using machine learning. This project focused on developing the general sensor physics to automatically detect stained proteins (in a protein and sensing independent way), and then aimed for both classical and machine learning approaches to automatically detect PD and related proteinopathies using the developed techniques. That is, the project's objective was to design and develop a sensing and processing approach to automatically measure these three proteins (α -syn, A β and Tau) for the first time, and use those measurements to classify and differentiate samples into their ground truth classifications, provided by pathologists.

The project team, consisting of neuroscientists and subject matter experts from Imperial College London, NHS AI Lab Skunkworks, Parkinson's UK, and Polygeist have undertaken a 16 week project to examine the possibility of producing a Proof-of-Concept (PoC) tool, demonstrating the core technology using example cases from the laboratory. The PoC tool can automatically detect each protein, discriminating between healthy controls and pathology cases with performance beyond the state-of-art (ranging from 80-99% detection, with less than 1% false alarms), while at the same time being fully explainable to the clinical user, as the technologies developed in this project are based on physical measurements.

Volumetric estimation of target proteins has been demonstrated experimentally, using custom hardware, and hyperspectral imaging devices (see Wu et al., 2016). Here we demonstrate a method akin to this, protein density measurement, using standardised hardware already found in the laboratory.

The key tangible output of this project is an open-source PoC tool that demonstrates the technology for a single test microscope, released with documentation to allow others to build upon this work. However, the intangible benefit of this project is that for the first time we have demonstrated an approach that can, with configuration, be applied to any protein staining problem, and provide researchers and clinicians with a new measurement tool. This core technology is ready for in-the-lab validation and configuration, to demonstrate that it can be used in the pathologist's workflow. This report details the development project, alongside some of the implications, challenges, and opportunities for the future.

Document Navigation & Section Summary Note

The technical introduction in **Section 1** gives an overview of the relevant literature and the motivations for the project, and a top level hypothesis regarding Diaminobenzidine (DAB) detection; namely, that more DAB should be detected in samples where PD pathology has been observed by the pathologist than those control samples.

The fundamental methodology for DAB isolation and segmentation can be found in **General Methods (Sections 2 - 3)**. Further sections also cover the most effective way of classifying proteins (using the median density, or PDNet to classify patches of samples, or Regions of Interest (ROIs) in **Section 4**). It also covers, in detail, the procedure for converting brain images into density measurements, and the sensor physics and colourimetry unique to this project (see **2.2.5**).

The general discussion in **Section 7** covers the main discussion and conclusions, along with a summary of the core findings from each other section. This should give a rounded view of the project's outputs. **Section 7** provides a surface level analysis of those results presented in each of the results sections (see **Section 4.1-3, 5.1 and 6.2**). For further extensive analysis, end of sprint reviews are available through the ACE sharepoint, with extensive discussion of each work package.

Sections 5 and 6 are highly experimental work packages, looking at methods of classification and staging using highly compressed data. These work packages were undertaken to explore what data are necessary to effectively classify and satisfy pathology using whole sample and whole brain measurements.

Some information such as sample counts, and case details have been repeated in each methods section, to aid readability.

Contents

| | |
|---|-----------|
| Executive Summary | 1 |
| Document Navigation & Section Summary Note | 2 |
| Contents | 3 |
| 1 Technical Introduction | 5 |
| 2 General Methods | 9 |
| 2.1 Spectral Basis Decomposition & Estimation | 9 |
| 2.2 Spectral Decomposition Methods | 11 |
| 2.2.1 Section Summary | 11 |
| 2.2.2 Materials | 11 |
| 2.2.3 Design | 12 |
| 2.2.4 Automatic Background Removal | 12 |
| 2.2.5 Procedure | 13 |
| 2.2.6 Statistics | 14 |
| 2.3 Protein Configuration and Segmentation | 14 |
| 2.3.1 Materials | 15 |
| 3 General Machine Learning Methods | 16 |
| 3.1 Machine Learning Methodology Summary | 16 |
| 3.2 PDNet | 17 |
| 3.3 Data Segmentation & Training Methodology | 17 |
| 3.3.1 RGB ROI Utilisation | 17 |
| 3.3.2 Data Partitioning | 17 |
| 3.3.3 Preprocessing | 17 |
| 3.3.4 Judgement and Loss | 18 |
| 3.3.5 Validation and Case Based Classification | 18 |
| 4 Protein Specific Methods & Results | 19 |
| 4.1 α -Synuclein Segmentation & Classification Results | 19 |
| 4.2 Tau Segmentation & Classification Results | 20 |
| 4.3 A β Segmentation & Classification Results | 21 |
| 5 Braak Staging Data Analysis & Classification | 23 |
| 5.1 Methods and Results | 23 |
| 5.1.1 Materials | 23 |
| 5.1.2 Principal Component Analysis | 24 |
| 5.1.3 PCA Results | 25 |
| 5.1.4 Statistical Models and Classification Procedure | 26 |
| 6 Machine Learning on Density Maps | 28 |
| 6.1 Multidimensional Whole Sample Classification Summary | 28 |
| 6.2 Summarised Methods & Results | 29 |
| 6.2.1 Targeted Proteins and Differentiation | 29 |
| 6.2.2 CNN Design | 29 |

| | |
|--|-----------|
| 6.2.3 Preprocessing | 29 |
| 6.2.4 Judgement and Loss | 29 |
| 6.2.5 Procedure | 29 |
| 6.2.6 Results | 30 |
| 7 General Discussion | 31 |
| 7.1 Protein Segmentation | 31 |
| 7.2 Appearance Model Classification of ROIs | 32 |
| 7.3 Automated Braak Staging & Multidimensional Model | 33 |
| 7.4 Scope, General Applications & Conclusions | 34 |
| 7.5 Next Steps | 35 |
| 8 References | 36 |
| 9 Appendix A | 40 |

1 Technical Introduction

Accurate diagnosis of Parkinson's Disease (PD) is essential to furthering our understanding of the pathological causes of PD, which will ultimately lead to the development of more effective therapies. Post-mortem neuropathological diagnosis of PD is labour intensive, requiring manual microscopic examination of stained brain sections and adherence to a number of diagnostic protocols and staging systems. The predominant staging system, Braak staging (Braak et al., 2003) defines six stages, progressing from Stage I (the beginning of disease, in the dorsal motor nucleus of the vagus nerve), to Stage VI (the apex of the disease, within the neocortex). These protocols are the subject of active research, with variable inter-rater reliability between expert raters, when staging pathology ($0.4 < \text{Krippendorf's } \alpha^1 < 0.6$, for review and comparative analysis, see Attems et al. 2021). These protocols share in common the identification of Lewy Bodies (LBs), which are largely constituted by α -syn protein; which can be found at various sites within the nervous system including olfactory, neocortical and subcortical structures, and the brainstem (Attems et al., 2021; Alafuzoff et al., 2009; Braak et al., 2003; McKeith et al., 1996).

This α -syn protein can be stained via Immunohistochemistry (IHC), with commercially available antibodies and with Diaminobenzidine (DAB) detection (for a comparative study, see Crosier, et al., 2006). This allows visual identification of both intracytoplasmic and extracellular α -syn, and can be made using light microscopy (using Aperio scanners) at 20x objective (~200x visual) magnification (see Crosier, et al., 2006; Beach, et al., 2018). Due to the nature of the morphology of α -syn protein, this is a manual process conducted by pathologists often using both analogue and digital pathology methods (Williams, Ismail, Chakrabarty & Treanor, 2021); however, the presence of α -syn protein does provide binary classification of PD post-mortem, and provides stratification of severity via site presence such as in the dorsal motor nucleus of the vagus nerve or the frontal cortex (Attems et al., 2021).

PD pathology as well as other Neurodegenerative Disorders (NDD) such as Alzheimer's Disease (AD), Multiple System Atrophy (MSA), Pick Disease, Frontotemporal Dementia, and Corticobasal Degeneration (among others) are marked by the presence of other malformed or abnormally aggregated proteins (Soto & Estrada, 2008; Ross & Poirier, 2005); most notably for this current work A β and pathological Tau. The presence of excessively deposited and aggregated A β is a key pathological marker, and can be found in brains with AD pathology (Glenner & Wong, 1984; for review of mechanisms see Folch et al., 2018), in addition to typically ageing brains (Näslund et al., 1994); the toxicity of which affects several mechanisms such as mitochondrial dysfunction, cell membrane disruption, and synaptic toxicity, among others (for a review of the mechanism's biochemistry, see Rajasekhar, Chakrabarti, & Govindaraju, 2015). Whilst significant variance between brains exists neuropathologically, both A β and Tau Neurofibrillary Tangles (Tau, NFTs) are present in a majority of patients presenting with Parkinson's Disease Dementia (PDD), with frequency to

¹ Krippendorf's α measures the average level of agreement between raters, who are using a unit scale (like the level of disease present in a case) to score an observation. A high score indicates a high level of agreement.

warrant classifications of AD co-morbidity (for review see, Irwin, Lee & Trojanowski, 2013). Pathological Tau may play a role in mediating A β (Wang & Mandelkov, 2015), as well as presenting in various morphologies in typically ageing brains, and those diagnosed with, amongst other pathologies, progressive supranuclear palsy and parkinsonism (for review of Tauopathies, see Kovacs 2018). Whilst different experimental methodologies are employed in an attempt to elucidate the nature and interplay of these proteins to the stratification, classification and progression of each of these NDDs, a very similar protocol can be applied to stain and examine them visually. Each of α -syn, A β and Tau can be stained via IHC with DAB detection, resulting in different morphological appearance at magnification via light microscopy.

Previously, we have built an appearance model, a Convolutional Neural Network (CNN), to artificially stain and detect α -syn (see Pearce, et al., 2022). This approach utilised the difference in luminance between the stained areas of α -syn protein, relative to the rest of the brain tissue, to produce a rough protein mask, which was further refined by the *ideepcolour* (Zhang et al., 2017) generative network to ‘synthetically-stain’ α -syn. This produced highlighted regions in the original images with the same texture as the masked proteins. These preprocessed images (true positives and false positive controls) were then passed to a CNN for appearance-based binary classification of PD or Control. This technique yielded binary classification of $\sim 1000\mu\text{m}^2$ tissue patches with performance $F_1=.91$ (for details, see Pearce, 2022). Due to the similar protocol of staining A β and Tau, a similar procedure should be applicable for identification of each target protein; however, the aforementioned technique has a number of core weakness: firstly, the core identification of the target protein (α -syn) is produced by an opaque process (*ideepcolour*); secondly, the classification CNN is trained using imagery bound to the specific microscope that captured them. Moreover, in the case of α -syn, binary classification is available, but in the case of A β and Tau, morphological classification and a direct measurement of the amount of detected protein is desirable; this is due to the presence of the target proteins in a variety of pathologies and (in the case of Tau) typically ageing brains (control cases).

This project sought to develop an experimental technique that could be used to identify each of three target proteins: α -syn, A β and Tau. The aim was to demonstrate differential sensitivity for each protein against a control group using a general detection technique that applied to each protein; that is, the isolation of sensor response of the DAB stain that is used in the staining protocol of each protein. By producing a pixel-level DAB detection algorithm, the classification of the morphology of each protein as a function of pathology conditions is akin to shape detection at the image level. More explicitly, a perfect binary mask of DAB detection for each digital image yields the percentage of protein present and the shape of that mask identifies the protein. To achieve this we required the ability to specifically isolate the sensor response of the DAB stain apart from the sensor response contributions of other cellular bodies, brain tissue, amplification counter-stain and so forth.

Digital sensors and photoreceptors suffer from sensor univariance (Rushton, 1972); that is, the microscope’s digital sensor produces the same sensor response for many photons of a wavelength that is not sensitive to, and for few photons of a wavelength that it is very

sensitive to. This means that it is possible to produce sensor metamer; spectra which have different power as a function of wavelength but produce the same tristimulus value (one vector of three values, one for each of its Red, Green and Blue (RGB) designated channels), by balancing the relative spectral power at each wavelength to preserve the same tristimulus value (Hunt, 1991; Wyszecki & Stiles, 1982). This means that the recovery of the spectral content with the tristimulus values alone is not solvable. This has been directly documented in the case of light microscopy; Cukierski, Qi and Foran (2009) demonstrate metamer in tissue of neutrophil as evidence for the need of multispectral microscope imagery.

Finlayson, et al (2014) demonstrated that metamer sets can be computed as a number of linear weighted combinations of spectral basis functions for a given receptor. That is, given a number of primary illuminations (such as spectra emitted from a number of Light Emitting Diodes), a given receptor response can be matched by a weighted sum of those basis functions. If we assume that the sample slide media can be treated as a colour filter, filtering the microscope's lightsource, then light arriving at the microscope sensor is a linear combination of filter functions (the media transmission functions) applied to the lightsource spectra. This assumption is a simplistic model, ignoring subsurface scattering, mutual illumination, and any interaction of surfaces within the media. However, if such assumptions hold, any sensor tristimulus value (pixel in the imagery) can be computed as a weighted sum of the sensor tristimulus values of each of the possible illuminations. That is to say, if all the light hitting the sensor is a mixture of primary sources that do not interact, then it should be possible to estimate each of them, given the sensor response.

In the case of the imagery produced by the IHC staining protocol, the basis functions are the DAB, Eosin and Hematoxylin reflectance functions, as well as the average healthy brain transmission function each combined with the illumination source. Each tristimulus then becomes a 6 element vector (one element for each stain, brain transmission and illumination basis function), representing the relative contribution of that component's tristimulus value to the sensed tristimulus value. A full description of this decomposition process can be found later in the General Methods section.

The DAB component for each pixel when taken together form a DAB image plane, which directly measures the DAB stain measured by the microscope, in addition to background noise. Our working prediction for this project was that the DAB image plane would have greater activation (higher weighting values) for images containing stained protein (such as α -syn) than those images from the control groups wherein no protein was present. In the case of Tau, where some protein was expected in the control groups, we predicted that the morphology, defined here as the shape of DAB activations within the plane, would differ between the groups in addition to the overall level of protein activation.

In this study, to test these predictions, the Parkinson's UK Brain Bank, at Imperial College London, provided samples of brain tissue (hereafter interchangeably referred to as slides²)

² Slides are typically referring to the entire tissue, media and image glass background. Samples are typically referring to the tissue on one or more slides.

from 401 cases (individual brains; including 301 PD cases). These samples were grouped by case and by pathological designation: containing PD pathology; containing pathology of MSA; containing both PD and AD pathology; and Controls (typically ageing brains). All slides were acquired and digitised using the same methodology, and are described in more detail herein. These slides were loaded and decomposed to produce DAB planes which were then segmented into subsection windows (as described by Signaevsky et al., 2022) to produce ROIs with high DAB activation. These ROIs were then evaluated using a variety of statistical methods outlined herein, to establish the differences between the case level conditions, as well as matched anatomical regions between each case. For example, relative DAB activation for slides containing stained α -syn were examined in the dorsal motor nucleus of the vagus nerve, in cases with associated PD diagnosis and compared to controls; to determine the threshold required for DAB activation to reliably act as a binary discriminator of PD.

To further investigate the morphological differences between the groups, original brightfield imagery (stored as red, green and blue component tristimulus values, or RGB imagery) as well as the DAB decomposed ROIs were used to train the PDNet CNN appearance model (for full details on this model, see Pearce et al., 2022). The model was used to perform binary classification at a patch level, to determine if initial filtering of samples to ROIs would aid feature selection and improve binary discrimination, stratification and group level differentiation. We predicted that if significant differences between the groups were apparent during the staining protocol (that is, proteins were stained and visible by the pathologist), then these differences would be both measurable by the raw decomposition statistics and evident by differential sensitivity in the appearance model.

Here we present the development, preprocessing, training, and validation procedure for automatically detecting and differentiating PD, PD+AD, and Tau pathologies, via post-hoc classification of synthetically stained patches. Full software source code including engineering assurance, documentation and ancillary project outputs can be found on the NHS Transformation Directorate public github repository at:

<https://github.com/nhsx/skunkworks-parkinsons-detection>.

2 General Methods

2.1 Spectral Basis Decomposition & Estimation

The microscope sensor response for any spectra is (as a function of wavelength) the combination of the sensor's sensitivity with the illumination spectra, and with the surface reflectance function of any objects within the sensor's field of view (that scatter light towards the sensor), and subsequently the transmission functions of any filter media (sample media) and Bayer filters covering the sensor array (see formula B in Figure 2.1; see Cukierski, et al. 2009, for review see Hunt, 1991).

The Aperio AT2 microscope in use at Imperial College's histopathology labs has a fixed light source illuminating the sensor. If we consider that the sensor exposure remains fixed, then at maximal power and unimpeded by slide media, the microscope illuminator produces the maximal sensor response. That maximal response is referred to here as the whitepoint tristimulus value (w); for any pixel in the scene it can be defined as $w = [R(\lambda)_1=1, R(\lambda)_2=1, R(\lambda)_3=1]$ for the microscope this translates to maximum RGB, see Figure 2.1 below for illustrative implications. Other sensor responses vary the tristimulus value (a variable response of the sensor per pixel) and are the result of either: some filter altering the amount of light as a function of wavelength from the lightsource hitting the sensor, reflected light from other objects, or sensor noise (see A in Figure 2.1). Given the known spectra of the lightsource, and sensor response functions, only the reflectance functions / filter functions of the media between the light source remain as unknowns ($F(\lambda)$ in Figure 2.1). Given a large set of possible surface reflectances and variable scene geometry (media refraction and subsurface scattering), the estimation of the spectral reflectance from the sensor response is still not possible; there are an infinite number of surface reflectances which can yield the same sensor response (known as metamerism; see Hunt, 1991).

It is, however, possible to estimate the surface reflectance and filtering functions as the number of possible reflectances are reduced (For explanation of unknown recovery see Darrodi, et al 2015). For example, if two linear filters, 'A' and 'B' are independent (with minimal spectral crossover), and 'A' filters only those wavelengths to which the sensor is sensitive in the red channel, with 'B' filtering only those wavelengths to which the sensor is sensitive in the blue components, then the observation of a low blue value at a pixel is an observation of filter 'B'. If filters can overlap spatially, a low blue and red component at a pixel is an observation of both 'A' and 'B' filters (ignoring geometric interference such as subsurface scattering).

Equally, consider a sample in the view of the microscope containing a dye that filters only short wavelengths of visible light (400-420 nm), and therefore only affects the sensor's blue channel component's spectral response, the amount of that blue response that is filtered is proportional to the amount of dye present. Now consider an unknown mixture of two dyes - if the spectral transmission function of those dyes is known, then provided that there is not significant overlap in their interaction with the illumination or sensor channels, the

contribution of each to the sensor response can be calculated as a linear combination of those basis functions that produce that sensor response (see Figure 2.1.C).

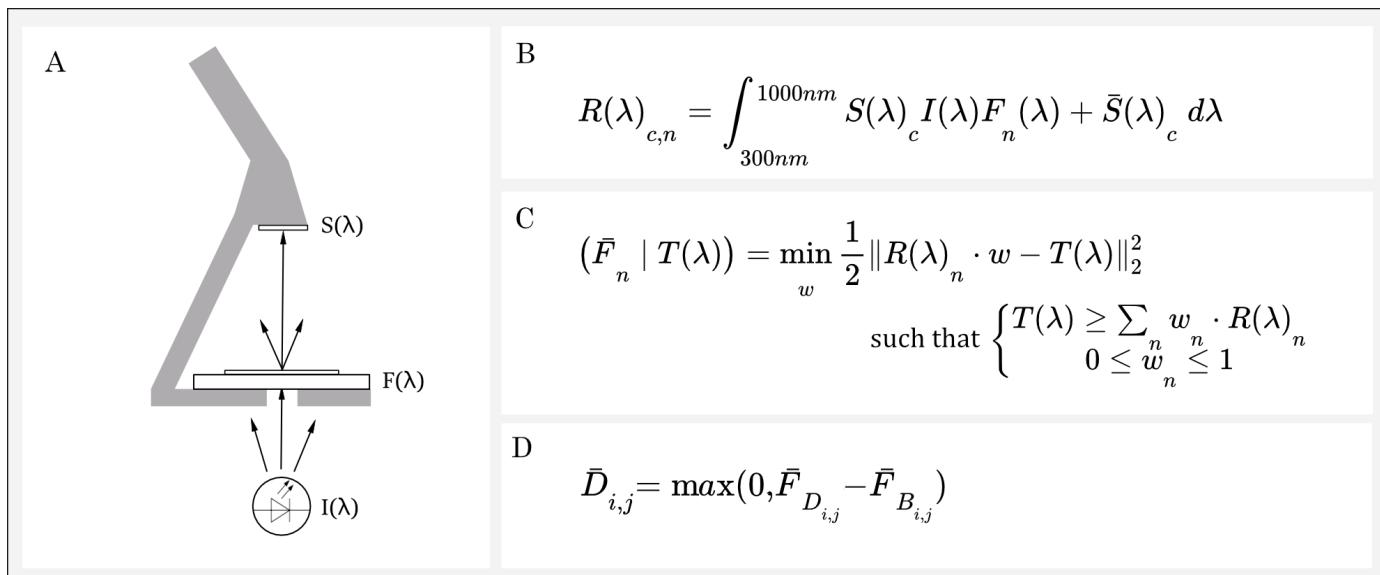


Figure 2.1. Schematic of microscope and key formulae for spectral estimation from basis functions

A. The light source, as a function of wavelength $I(\lambda)$ illuminates the sample media, which acts as a combined filter function $F(\lambda)$ and enters the microscope sensor, with combined sensor sensitivity and Bayer channel filters $S(\lambda)$; **B.** The response tristimulus R is a combination of illumination $I(\lambda)$ and the n filter functions $F(\lambda)$, and then combined with sensor sensitivities S for each channel c , with contribution of \bar{S} sensor noise components; **C.** The spectral estimate \bar{F} can be estimated given a pixel tristimulus value $T(\lambda)$ by linear optimisation, to find which linear combination of vectors of weight w for each possible reflectance tristimulus value $R(\lambda)$ could produce $T(\lambda)$ - with the constraint that each contribution is between 0 and 1. **D.** The density estimate for the DAB reflectance contribution \bar{D} at pixel i,j is the DAB channel spectral estimate (\bar{F}_D) minus the brain transmission estimate (\bar{F}_B) when the result is positive.

The IHC staining process and the microscope imaging colour space is optimised for the human observer. The microscope operates in the trichromatic human colour space, specifically optimised for the Commission internationale de l'éclairage (CIE) sRGB colour space (standardised RGB, see standard IEC 61966-2-1:1999 for details, IEC, 1999). The DAB stain produces a sensor response under the light microscope which is colourimetrically opposed to the background cellular filter response; that is, the DAB stain produces a light brown precipitate colouration (CIE 1931 2° Yxy = [x=.499, y=.417]) under natural viewing, with cellular structures stained with Hematoxylin counterstain producing a blue hue (CIE 1931 2° Yxy = [x=.276, y=.279]); the human perceptual difference between equiluminant signals is $\Delta E = 19.80$ (for context, the perceptual difference between equiluminant ANSI traffic light signal is $\sim \Delta E = 15.62$; see Lopez, et al, 2020). The DAB tristimulus values are opposing in colour space, and therefore metamers are likely to still be highly correlated to the target signal; that is, collisions in sensor space are still likely to cluster around the stain protein even though the univariance of the sensor makes direct spectral recovery impossible.

Recovery of the DAB stain is therefore similar to the example set out above, where we can treat the DAB filtering response as orthogonal to the transmission response of the stained

brain matter. By calculating the hypothetical isolated sensor response for each of the expected surface reflectances within the samples, DAB can be directly measured; this can be simplified as seen in Figure 2.1.D, where the DAB response minus the brain transmission response as estimated by linear optimisation against the hypothetical response values yields the expected density (as a function of sensor pixel coordinates).

2.2 Spectral Decomposition Methods

2.2.1 Section Summary

This methods section details the spectral decomposition and segmentation of α -syn protein in detail. It uses the methodology for identification of DAB stain via spectral basis estimation, and then ROI identification via segmentation of slides, at a given window sampling resolution. An assessment is then performed on the basis of a number of density statistics. Following these methods are the summary of configurations for the Tau and A β . For the sake of brevity, these methods are not duplicated, and are referred to as the 'segmentation procedure' throughout. Differences between the number of cases between protein and pathology conditions are detailed, along with the configuration parameters used for each segmentation procedure.

2.2.2 Materials

These experiments used digitised images, shared with the project group by the Parkinson's UK Brain Bank. Images were acquired using a Leica/Aperio T2 digital light microscope at 20x magnification. Images were stored in compressed Aperio (.svs) format, with each pixel width and height spanning .504 μ m, and a typical tissue area of ~6.6cm 2 per sample. Samples were taken from post-mortem sections of human tissue, ~7 μ m in thickness, from six anatomical sites, these can be found in Section 5, Table 5.2. Samples were fixed in formalin and sections were subsequently stained with an IHC protocol, targeting α -syn, Tau and A β proteins. Staining was intensified with DAB and a subsequent Hematoxylin counterstain; for detailed staining procedure see (Jakes, et al 1994).

The processing of α -syn samples utilised a total of 401 cases (totaling 401 slide samples), 301 of which had been classified as containing pathology associated with PD, with the remaining (100 cases, 100 slide samples) constituting controls.

All experiments were conducted *in silico*, using Python code running atop generally available workstation hardware (System); the System used an AMD Ryzen 9 5900X (12-core @ 3.75GHz), NVIDIA GeForce RTX 3090 (24GB VRAM), and 64GB of RAM. The system utilised Linux 6.1.2, CUDA 11.7 (r11), PyTorch 1.12.0, and NVIDIA APEX extensions for PyTorch (version a0f5f3a). Pytorch was installed with the ATLAS library for linear algebra enabled. All other configurations, and installation instructions can be found on the public GitHub repository at <https://github.com/nhsx/skunkworks-parkinsons-detection> (herein, Repository).

Spectral emission, absorption, and transmission data were digitised for computer use, by experimenters manually sampling rasterised digital sources. Discover ChromoMap DAB (produced by Roche Diagnostics) which was utilised in the staining of samples, was

digitised from both the datasheet supplied by the manufacturer and matched to the spectral reflectances provided by Cukierski et al., (2009), as was the Eosin response function. The AT2 microscope utilised the LXA7-PW57 lightsource, and Dalsa Piranha PC-30 camera; the illumination and RGB response functions and filters were digitised from Olson (2013). The average spectral brain transmission was digitised and inverted from absorption of healthy human brain grey-matter tissue from Shapey et al. (2022). All spectral sensitivity and absorption functions were interpolated to 1nm resolution and clipped to the sensitivity range of the microscope and illumination source (between 300nm to 1100nm); these density plots and associated tristimulus values can be found in Appendix A.

2.2.3 Design

The pathological classification of a sample as either a PD or control case constituted an independent variable partitioning those samples into groups. These classifications were determined through brain-level analysis; by pathologists, during the induction protocol at the Brain Bank, and based on the pathologist's observation of that brain containing PD pathology. This classification process and these data have been described previously (Attems et al., 2021; Pearce et al., 2022). This study examined case, sample and region of interest (ROI) level statistics for each group. For the segmentation of samples into ROIs there were two independent variables: the integration window size over which DAB activation density was measured (a dependent variable), and the threshold applied to that DAB density over which that window was defined as a ROI. We analysed digitised images of samples using 1034 μm non-overlapping windows (512px² using 4x downsampling with each pixel constituting $\sim 2\mu\text{m}^2$ of tissue) and considered regions with a mean DAB density of 0.125% (32,768 pixels with DAB activation) as a ROI. The total DAB density of each ROI constituted the dependent variable that was measured and thresholded to discriminate between samples, and cases in each group. The null hypothesis was that no significant difference between the decomposition of each sample would be observed as a function of pathological classification; if the DAB density estimate yielded a greater response in the pathology group (PD classification) then both a higher density as a function of spatial frequency would be measured and more ROIs would be segmented.

2.2.4 Automatic Background Removal

Some Tau and A β samples had undergone a different coverslip protocol, resulting in a manual exposure being applied to the slide background. To control the background chromaticity and ensure that regions were not erroneously segmented, a sampling based background normalisation procedure was applied to all samples before any further processing was performed. This involved sampling the top 100 rows of pixels in each samples (containing just the background illumination) and setting all pixels in the image whose RGB values were within 5% of the mean of that region to the illumination chromaticity whitepoint (w).

2.2.5 Procedure

Each sample was loaded into computer memory as an RGB digitised image at 16-bits per channel; slides were downsampled (4x) to $\sim 2\mu\text{m}$ per pixel using bilinear interpolation. The

least-squares fit between each pixel's RGB triplet and the sensor response (an RGB triplet) for each transmission material and lightsource was computed (see Finlayson, et al, 2014); to produce a 6 component vector for that pixel corresponding to the best linear combination of basis responses that could produce that sensor response, with each component loading positively onto one function. Each component plane then resembled a monochrome image pertaining to weights (loading on one spectral basis function) as a function of spatial frequency (image coordinate); one of which was the DAB plane, representing DAB activation as a function of illumination filtering and microscope sensitivity. To isolate the DAB stain amplification corresponding to the α -syn target protein, the brain transmission plane (representing a neutral density filter) was subtracted from the DAB plane (see Figure 2.1.D above). DAB pixels were then identified by thresholding at 30% of the maximal DAB signal (the DAB threshold). This threshold was chosen by prima facie observations of DAB segmentation in an example case (see 2.3 for details).

The processed DAB plane was then traversed in $\sim 1034\mu\text{m}$ (512px) non-overlapping windows, and the signal density was computed by taking the mean of the binary mask within that window, producing a density map array for the image with each cell in that map denoting the average DAB activation density over that $\sim 1034\mu\text{m}$ region; see Figure 2.2 below for schematic of the density estimation procedure. A threshold of 0.125% α -syn protein (DAB activation) was selected to define ROIs, chosen in the same manner as detailed in section 2.3. The window coordinates for those windows that met this threshold were used to produce 512px² images for both the raw RGB pixels, and spectral estimate planes for that region. Density maps were then analysed by computing the overall median density, and the median density of ROIs (those over a 0.125% threshold, see statistics section).

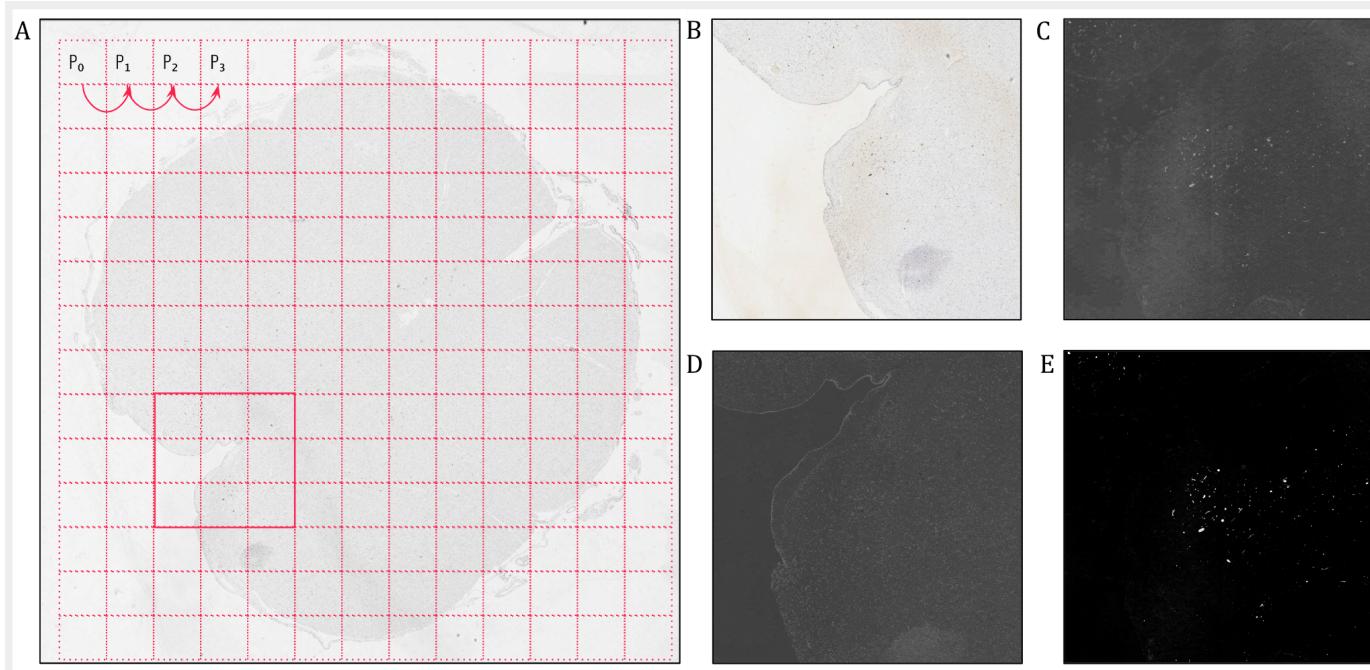


Figure 2.2. Procedure of ROI identification using a tumbling window, and example estimation planes.

A. Low-resolution rasterisation of an RGB slide sample, with P_0-P_3 denoting tumbling window over which regions were processed, the region highlighted in solid pink is expanded in **B**. Magnification of the pink region as shown in A, the RGB image in sRGB colour space; **C**. The spectral estimate \bar{F}_D (DAB) plane of the same region, where high values of DAB are light values and low values are dark; **D**. The spectral estimate \bar{F}_B , the brain transmission plane, with low values dark and high values light (note, the background is dark due to that region loading on the illumination plane, not present); **E**. The DAB density estimate \bar{D} , the difference between \bar{F}_D and \bar{F}_B , most of the background and shared variance is removed, with only DAB and some pigmented cells remaining.

2.2.6 Statistics

Histograms of the median density maps were computed to assess the distribution of median DAB activation at the group level. A threshold sweep was performed on the median densities at an ROI and case level (median of median ROIs) to produce a simple binary classification of PD or Control based on DAB activation. A simple sign test was used to demonstrate the effect size at the group level, and the Kolmogorov-Smirnov test (for computation discussion see Justel, *et al*, 1997) was used as a further hypothesis test to estimate the risk of type I error. Further post hoc analyses were performed and are documented later.

2.3 Protein Configuration and Segmentation

Each segmentation procedure utilised an independent configuration targeting one protein: either α -syn, Tau or $A\beta$. This configuration was selected during discussions in the weekly project working group meetings, and was selected via examination of prima facie results for cases: PD1011 (α -syn, Slide 17, Dorsal Motor Nucleus of the Vagus Nerve) and PD1116 (Tau and $A\beta$, Slide 10, Posterior Hippocampus). The configurations used for each segmentation pass are shown in Figure 2.3 below, along with the example cases used for the prima facie examination.

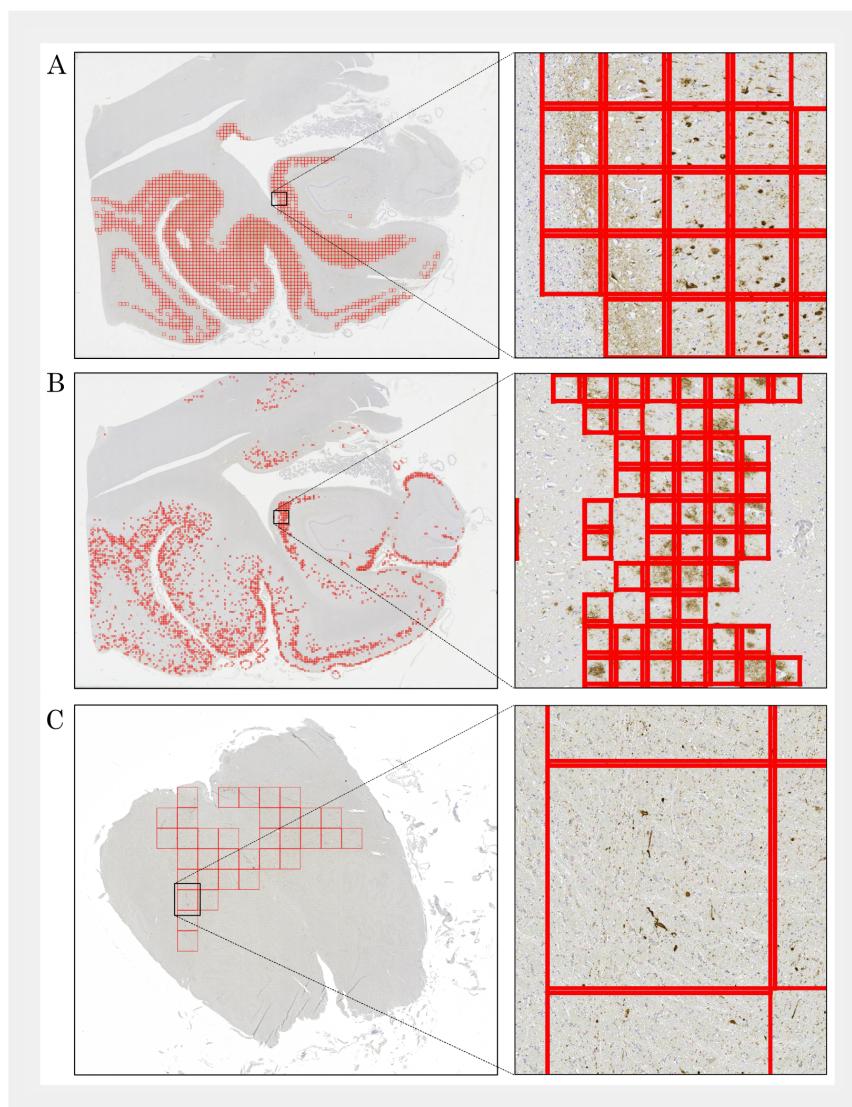


Figure 2.3. Protein specific spectral decomposition and ROI segmentation parameters.

- A.** Case PD1116, slide index 10, (Posterior Hippocampus), using Tau segmentation configuration in **D**, cutout shows magnified region with ROIs segmented in red;
- B.** Case PD1116 (as in **A**), using $\text{A}\beta$ segmentation configuration in **E** cutout shows a magnified region with ROIs segmented in red;
- C.** Case PD1011, slide index 17 (Dorsal Motor Nucleus of the Vagus, DMNoV), using α -syn segmentation configuration in **F**, cutout shows magnified region with ROIs segmented in red, white background can be seen where automatic background removal has been applied;
- D.** The Tau configuration parameters, including the processing window size used to designate ROIs, the DAB threshold used for determining DAB pixels and the density threshold used for designating an ROI;
- E.** The $\text{A}\beta$ configuration parameters, including those same parameters as described in **D**.
- F.** The α -syn configuration parameters, including those same parameters as described in **D**.

2.3.1 Materials

Tau

The processing of Tau samples utilised a total of 72 cases (totaling 627 slide samples). Of these cases, 4 were classified by pathologists as containing both AD and PD pathology, 37 as containing PD pathology, with the remaining 31 constituting controls.

A β

The processing of A β samples utilised a total of 87 cases (totaling 707 slide samples). Of these cases, 20 were classified by pathologists as containing both AD and PD pathology, 45 as containing PD pathology, with the remaining 22 constituting controls.

3 General Machine Learning Methods

3.1 Machine Learning Methodology Summary

The segmentation process produces ROIs and density statistics, and is sufficient for proving the physical relationship between the presence of DAB stain and the presence of PD, AD, or MSA in a patient. Given some error, we can infer this statistically, and a simple sign test (Test condition - Control condition yields a positive result) can provide a given confidence required to reject the null hypothesis that there is no difference between the groups.

However, nuance in morphology, shape and frequency by location is lost during this process, and we cannot learn about features in our dataset that we are not specifically measuring; this is due to statistical analysis of samples and ROIs reducing the sample to a single test statistic, or a vector of those statistics for each case. These methods look at using the segmented ROIs (those regions identified as containing DAB stain) to train appearance models for classification. This will allow us to explore other features in the segmented data that can be used to describe our respective cohorts.

The primary assumption is that either the body density, and therefore the visual appearance of the proteins within ROI imagery, will be visually differentiable between cohorts; or, that the morphology as expressed by shape and colouration will vary both within and between cohorts (for example, between Braak stages of PD cases, or PD cases and controls). For a schematic of the procedure outlined in this section, see Figure 3.1.

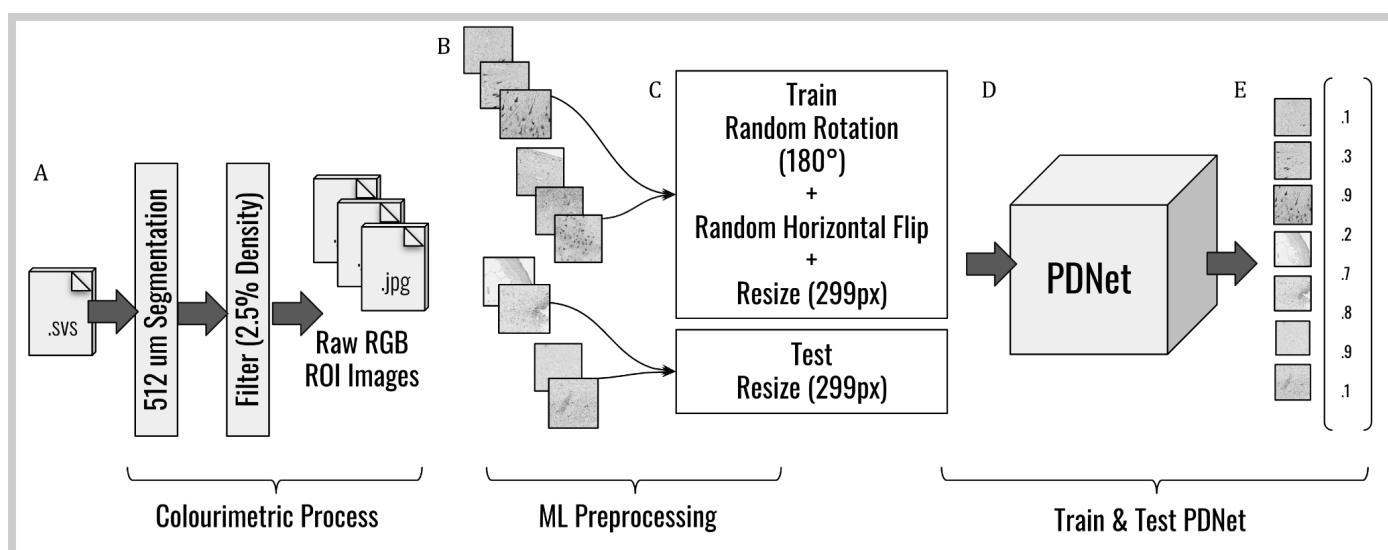


Figure 3.1. Training and classification procedure using PDNet and Colourimetric Segmentation.

A. SVS slides are loaded and undergo the segmentation procedure outlined previously, producing ROIs which are carved out of the raw RGB imagery and saved to disk; **B.** ROIs are partitioned into cases and cases are partitioned into training and test sets; **C.** Training set undergoes augmentations such as rotation and flips

as well as resizing to the network input size, whereas the test set only undergoes resizing; **D**. The PDNet takes a batch of these ROIs and produces outputs ; **E**. Outputs of PDNet which are tied to each ROI and represent a confidence score in that ROI belonging to the test condition (pathology).

3.2 PDNet

PDNet is an adaptation of EfficientNet, using a number of converging linear layers to produce a vector of the appropriate classification length; for example, for two groups, a two element vector is produced where each element represents a confidence that a given input belongs to that group. For a full description of PDNet, see the previous description in Pearce et al., (2022).

Here we utilised PDNet to provide appearance based classification between control and test conditions for each of the three protein groups. PDNet was trained on each protein in isolation (for example ROIs from PD group stained α -syn vs controls).

3.3 Data Segmentation & Training Methodology

3.3.1 RGB ROI Utilisation

To avoid data reduction and inadvertently adding bias to the sampling, the raw RGB pixels from those ROIs that were identified during the segmentation process were used during training and inference. The segmentation procedure was utilised to find the regions in both control and test conditions, with those regions found in the control regions representing either false positives (in the case of α -syn) or differential morphology (in the case of Tau). The regions were then sliced out of the RGB image in memory and stored as a separate ROI RGB image with case, slide and ROI annotations.

3.3.2 Data Partitioning

ROIs were first separated by case, such that any single case could either appear in the training set or the test set. Data was weighted towards training, with 60% of the cases in the dataset reserved for training with the remainder used for testing. Cases were not uniform in their ROI count, and therefore a second flattened training procedure was conducted for verification of case contribution, whereby ROIs were split only by their cohort (test or control) before being further split into training and test sets.

3.3.3 Preprocessing

ROIs were reshaped before being input into PDNet using the PyTorch transforms library. For both training and validation, ROIs were resized to 299px², while during training an additional set of transforms were applied; this included a random rotation of up to 180°, and a random horizontal flip (50% chance). In cases where the segmentation procedure produced ROIs larger than the input size, this was a downscaling procedure (α -syn = 3.45x) otherwise ROIs were upscaled(Tau = 1.16x, A β =4.67x).

3.3.4 Judgement and Loss

PDNet was trained on single shot patch discrimination; that is, the task was to minimise the per-ROI classification loss (regardless of case splits). Each training run used balanced sets ($N(\text{Control}) = N(\text{Test}) = \frac{1}{2}B$, where B is batch size), such that the task was abstracted away from learning the frequency of pathology regions and towards absolute discrimination of features within each ROI. Loss was computed as binary cross entropy with logit loss, as described by Keren, et al (2018). We used a modified stochastic gradient descent (SGD) optimisation function similar to that described by Gupta, Ramanath, Shi, & Keerthi (2021), where balanced sets and a 50% threshold on confidence provided binary classification between control and test conditions. Learning rate and momentum were held as constant hyperparameters ($m=0.9$, $lr=0.0125$) during this study, as the purpose was to demonstrate differentiation, rather than optimise for absolute accuracy. Training was performed over 500 epochs, with the best overall loss on the training set being accepted as the best model candidate, providing the basis for the results presented in the following sections.

3.3.5 Validation and Case Based Classification

The test split was used to produce ROI-level statistics, where a confidence threshold was swept instead of producing a binary decision for each output. Then, each confidence score was partitioned by case, to produce a confidence vector (c) of m elements long for each of the n cases. Each confidence vector was then reduced to a case confidence (K) by taking the median confidence score $K_{i-1} = \text{median}(c_{i-1})$. The K vector, representing a confidence for each case, was then swept in the same way as before to produce case level discrimination statistics.

4 Protein Specific Methods & Results

4.1 α -Syn Segmentation & Classification Results

For the dorsal motor nucleus of the vagus nerve, the median (η) DAB activation response for each 1032.2 μm window (512px sensor region) in the PD pathology group was $\eta_{\text{DAB}}=2.84\%$, with a 95th percentile of 9.84%, in contrast to the control condition with median activation of $\eta_{\text{DAB}}\leq0.01\%$ and 95th percentile of 1.03%. A median of $\eta(\text{ROI})=36$ ROIs (those windows above 0.125% activation) were detected across all samples in the PD pathology group, in contrast to $\eta(\text{ROI})=4$ for the control group.

For ‘simple’ case-level binary classification, the use of median DAB activation with threshold $\theta(\eta_A)=2.28\%$ yielded 57.14% true positive detection and 2.00% false positives (precision: 96.62%, specificity: 98.00%). Binary discrimination using the median confidence outputs from PDNet for 1032.2 μm ROIs of each sample yielded 86.13% true positive detection and <.01% false positives (precision: >99.99%, specificity > 99.99%), with median confidence threshold $\eta\theta(\text{ROIs})=99.99\%$; the receiver operating characteristic curve can be seen in Figure 4.1.C. Signal detection statistics based on the best F1 score can be seen in Table 4.2, with a comparison with human raters and against previous results.

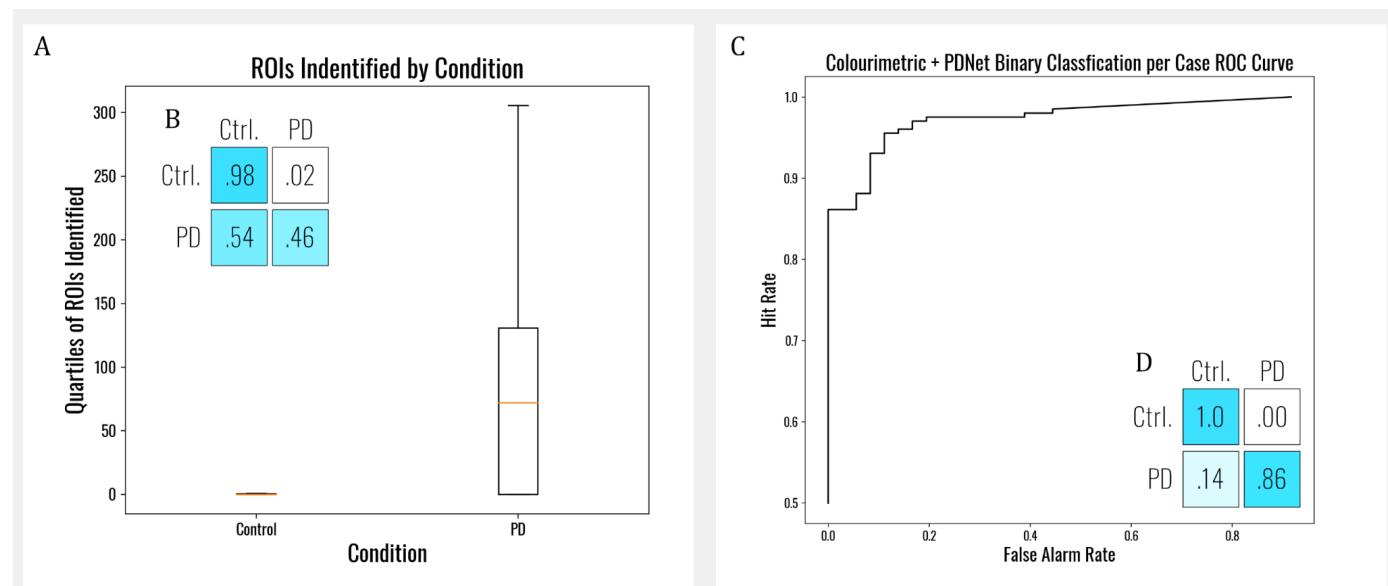


Figure 4.1. Segmentation and classification performance profiles for α -syn in the dorsal motor nucleus of the vagus nerve.

A. Median and interquartile ranges of the ROIs identified per case, per condition; **B.** Confusion matrix based on classification by ROI count, with threshold set at 80 ROIs for PD classification; **C.** Receiver operating characteristic curve for case level classification using median scores of ROIs processed by PDNet; **D.** Confusion matrix based on classification of median PDNet score.

Table 4.2. PDNet performance characteristics at a case level with colourimetric segmentation, synthetic staining (from Pearce, 2022) and comparisons for clinicians, (both on Braak 1+ Staging).

| System | Sensitivity ³ $tp / (tp + fn)$ | Specificity ⁴ $tn / (tn + fp)$ | Precision ⁵ $tp / (tp + fp)$ | F1 ⁶ $2 * P * R / (P + R)$ | Accuracy ⁷ $(tp + tn) / N$ |
|---|--|--|--|--|--|
| PDNet + Colourimetric | 0.92 _{.01} | 0.92 _{.01} | 0.92 _{.01} | 0.92 _{.01} | 0.92 _{.01} |
| PDNet + Synthetic Staining | 0.93 _{.01} | 0.88 _{.01} | 0.88 _{.01} | 0.91 _{.01} | 0.90 _{.01} |
| Expert Raters Signaevsky et al. (2022) | 0.59 _{.16} | 0.88 _{.30} | 0.94 _{.13} | 0.71 _{.12} | 0.66 _{.13} |

4.2 Tau Segmentation & Classification Results

The median DAB activation for all ~516.1 μ m windows (256px sensor region) of all slides, in all conditions, (both ROIs and non-ROIs) was $\eta_{DAB} \leq 0.01\%$. The median number of ROIs detected for the control condition was $\eta(\text{ROI})=183$ (7.3%) for the control condition, $\eta(\text{ROI})=79$ (3.1%) for the PD condition and $\eta(\text{ROI})=180$, (7.1%) for the AD+PD condition. The MSA condition yielded ROIs=22, ($\leq 0.05\%$). A Chi-squared test between all counts and slide conditions reviewed a significant difference between the ROI segmentation count and condition X^2 (27, $N = 123.4K$) = 37,817.41, $p = .\leq 0.001$. The interquartile ranges of ROI detection counts for each condition and slide index can be seen in Figure 4.3 below. Extensive analyses between the interaction effects of DAB density and ROI frequency were not performed, however some differences were identified; for example, the median DAB density for the control ROIs for slide 12 was $\eta_{DAB}=64.63\%$ compared to the AD+PD group at $\eta_{DAB}=35.11\%$, however the median number of ROIs identified for the control group was $\eta(\text{ROI})=3.5$ (0.02%), compared to the AD+PD group of $\eta(\text{ROI})=270.0$ (10.62%) .

As Tau some positive detections were present in PD, PD+AD and control conditions, PDNet was utilised to differentiate between the morphology of Tau pathology versus the control condition. The PD and AD+PD conditions were collapsed into a *tauopathies* condition and ROIs from this group and the control group were partitioned to train and test PDNet (discriminating between morphology as opposed to true positive vs false positive segmentations). For binary discrimination of Tau pathology between the conditions using the median confidence outputs from PDNet for 516.1 μ m ROIs of each sample yielded 86.10% true positive detection and <4.9% misclassification (false positive) rate; see Figure 4.3 and Table 4.4 below.

³ Sensitivity is the proportion of the positive cases (PD) that were detected by the algorithm (true positives / true positives + false negatives).

⁴ Specificity is the proportion of the control cases that were detected as control cases, against those that were falsely detected as positive cases (true negative / true negative + false positive).

⁵ Precision is the proportion of the correctly detected positive cases against those that were falsely detected as positive cases (true positive / true positive + false positive).

⁶ Weighted F score is a ratio of precision and recall which combines these statistics to produce a single metric, see Chinchor (1992). Here F1 is used.

⁷ Accuracy measures the proportion of true positive and true negative cases against the total number of judgements (true positive + true negative / number of judgements). Accuracy does not consider the relative density of positive cases in the overall set.

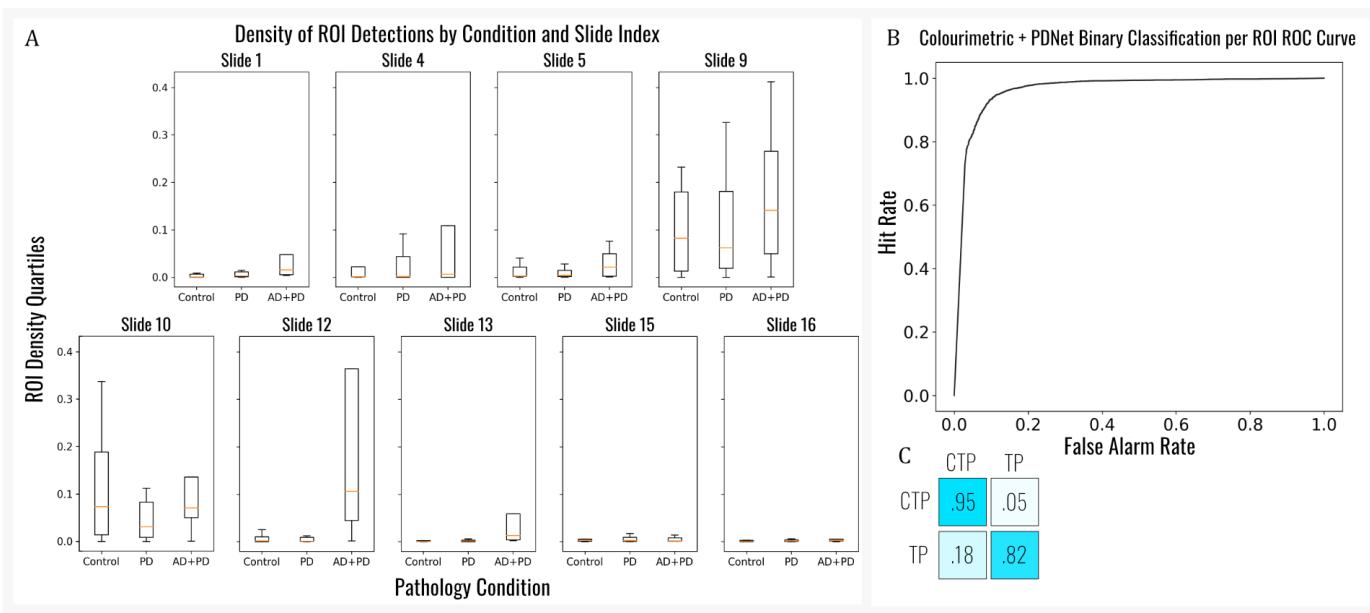


Figure 4.3. Segmentation and classification performance profiles for Tau across each slide index.

A. Median and interquartile ranges of the ROI density identified per case, per condition, and slide indices; **B.** Receiver operating characteristic curve for ROI level classification of Tau morphology using confidence scores produced from PDNet; **C.** Confusion matrix based on classification of Tau ROIs between control Tau pathology and Tau pathology conditions.

Table 4.4. PDNet performance characteristics for morphology detection at an ROI level for Tau pathology.

| System | Sensitivity $\frac{tp}{tp+fn}$ | Specificity $\frac{tn}{tn+fp}$ | Precision $\frac{tp}{tp+fp}$ | F1 $\frac{2*P*R}{P+R}$ | Accuracy $\frac{(tp+tn)}{N}$ |
|-----------------------|--------------------------------|--------------------------------|------------------------------|------------------------|------------------------------|
| PDNet + Colourimetric | 0.86 _{±0.01} | 0.94 _{±0.01} | 0.93 _{±0.01} | 0.89 _{±0.01} | 0.90 _{±0.01} |

4.3 A β Segmentation & Classification Results

The median DAB activation was consistent across each slide index, with the median DAB activation for ~129 μ m ROIs significantly higher in the PD+AD group ($\eta_{DAB}=25.27\%$ compared to PD and control ($\eta_{DAB}=18.09\%$ and $\eta_{DAB}=17.07\%$ respectively); the median density across all windows (both ROI and non-ROI) was $\eta_{DAB}\leq0.01\%$. The 95th percentile DAB activation also followed this trend, with DAB=95.63% for PD+AD versus DAB=54.71% for PD alone, and DAB=45.02% for control. The interquartile ranges for ROI density, collapsed by slide index, can be seen in Figure 4.4. Due to a smaller window size, a higher frequency of windows were identified than in previous experiments, with the control condition yielding $\eta_{(ROI)}=404$ (1.03%), PD condition $\eta_{(ROI)}=2935$ (6.68%) and PD+AD condition $\eta_{(ROI)}=3120$ (7.48%); demonstrating similar frequency of ROIs for the PD+AD and PD groups but with 1.75x greater DAB density for the 95th percentile.

Due to the smaller window size, significantly more confusable (false alarm) ROIs were segmented. While it was not possible to examine these thoroughly prima facie, control and PD+AD ROIs were utilised for classification by PDNet. Due to the high variability in ROI counts per case, two training and testing procedures were conducted; one where cases were split between training and testing, and one where cases were collapsed and

partitioned between conditions only. These runs can be seen in Figure 4.5. Binary discrimination of A β pathology ROIs between control and PD+AD using confidence outputs from PDNet yielded 89.60% true positive detection and 2.10% false alarms; see Figure 4.5 and Table 4.6 below.

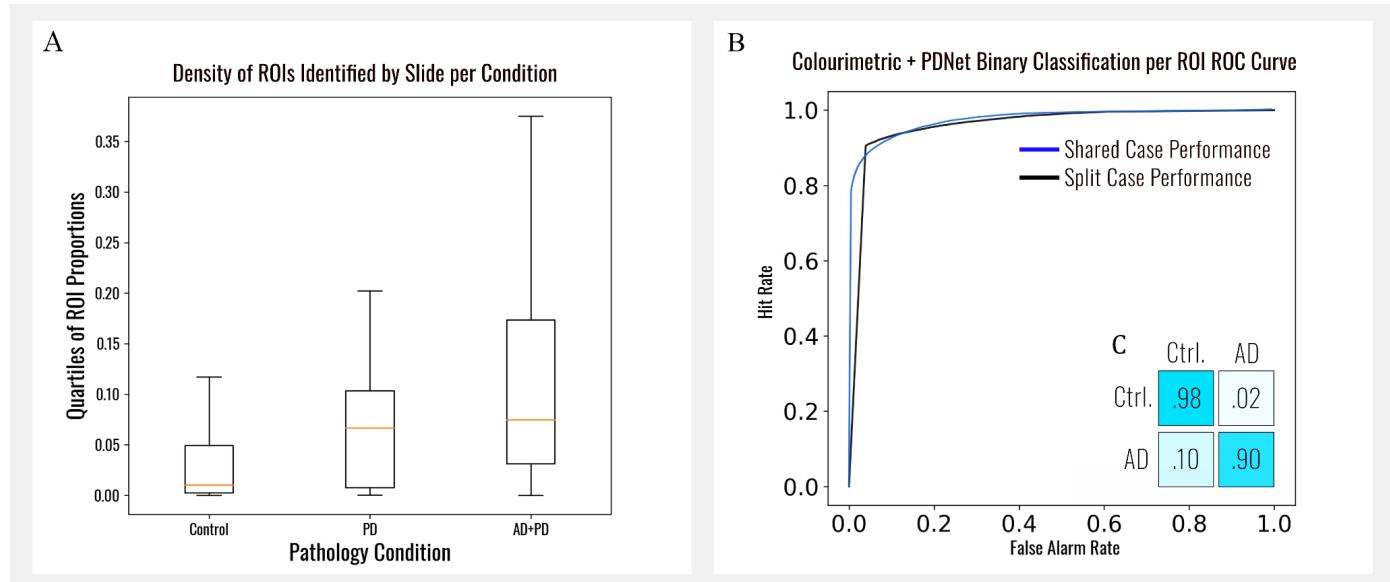


Figure 4.5. Segmentation and classification performance profiles for A β across conditions, collapsed by slide index.

A. Median and interquartile ranges of the ROI density per sample/slide, per condition; **B.** Receiver operating characteristic curve for ROI level classification using confidence scores produced by PDNet. The shared case condition shows validation performance for the condition where ROIs from the same case could span test and control sets (same case, independent ROI) and the split case condition preserves case level isolation between test and control; **D.** Confusion matrix based on classification of ROIs using PDNet score for the split case condition.

Table 4.6. PDNet performance characteristics for classification at an ROI level for A β pathology.

| System | Sensitivity $tp / (tp + fn)$ | Specificity $tn / (tn + fp)$ | Precision $tp / (tp + fp)$ | F1 $2 * P * R / (P + R)$ | Accuracy $(tp + tn) / N$ |
|-----------------------|------------------------------|------------------------------|----------------------------|--------------------------|--------------------------|
| PDNet + Colourimetric | 0.93 \pm 0.01 | 0.93 \pm 0.01 | 0.93 \pm 0.01 | 0.93 \pm 0.01 | 0.93 \pm 0.01 |

5 Braak Staging Data Analysis & Classification

This section investigates the automatic classification of brain samples across a case into their designated Braak stages. Varying staging systems exist, and there is active research into both the number of, and the qualifying criteria for, each stage (Attems, et al., 2021). Each case in the dataset was assigned a Braak stage by pathologists at Imperial College London.

As the presence of α -syn is diagnostic, a reliable ROI detection using the segmentation procedure and PDNet is sufficient to provide accurate staging (Pearce et al., 2022). For example, an observation of an ROI in the frontal cortex should indicate Braak 6. However, the relationship between the overall density measurement, and nuances such as diminishing presence of α -syn in the DMNoV as PD progresses, is lost with this methodology.

Due to there being seven classes of data (control and six Braak stages), partitioning of the available data into training and test sets was not feasible; only a limited sample size (less than 15 in some classes) was available during the project timeframes. However, as the segmentation algorithm was so sensitive to α -syn detection, and as clinicians use these identifications for manual staging currently, it was expected that whole brain knowledge applied to the segmentation procedure should be sufficient to provide accurate staging.

Density measurements were collapsed into median statistics for each case, and those data were analysed statistically to understand if that information was predictive (using principal component analysis, PCA). A number of simple classifiers (decision trees and random forest) were trained and evaluated to explore how the clinical staging characteristics were arrived at from first principles (for example, using slide 17 as diagnostic for PD vs control). This classification regime could then be propagated back to the segmentation algorithm to provide automatic Braak classification, as samples became available.

Due to the ability to simultaneously provide the statistical analysis and build the classifier structure, we utilised JASP to conduct these analyses; using density measurements produced by the segmentation procedure.

5.1 Methods and Results

5.1.1 Materials

These analyses were performed in JASP version 0.16.3. The same materials as described in the General Methods section were utilised to produce density maps for the control and PD conditions using the segmentation procedure. Those density maps each constituted an array of DAB densities, for each $\sim 1032\mu\text{m}$ window of a sample. Each of the samples/slides corresponded to one of six anatomical regions as described in Table 5.2. Braak staging data was provided by the research team at the Parkinson's UK Brain Bank, such that each case

was designated between 0 and 6, with 0 being a control case and 1 to 6 being the classification of Braak 1 to 6. The processing of α -syn samples utilised a total of 401 cases (totaling 4079 slide samples), 303 of which had been classified as containing pathology associated with PD, with the remaining 98 cases constituting controls. The breakdown of cases by stage can be seen in Table 5.2.

Table 5.2 Slide index, associated anatomical region, Braak stage when pathology is present, and cases classified as that stage.

| Slide Site Index | Anatomical Region | Associated Braak Stage | Number of PD Group Cases at Stage |
|------------------|-----------------------------------|------------------------|-----------------------------------|
| 1 | Frontal Cortex | 6 | 124 |
| 2 | Cingulate Cortex | 5 | 46 |
| 4 | Temporal Cortex | 4 | 96 |
| 15 | Substantia Nigra | 3 | 35 |
| 16 | Locus Coeruleus | 2 | 1 |
| 17 | Dorsal Motor Nucleus of the Vagus | 1 | 1 |

5.1.2 Principal Component Analysis

As there was an imbalance in the frequency of cases for each Braak stage, and a low case count in Braak stages 1, 2, 3 and 5, a principal component analysis (PCA) was conducted before any model training. This constituted removing the two cases at Braak 1 and 2, and then computing the 95th percentile DAB density of each density map; producing a six element vector for each case. Those cases were then passed to the JASP software package's PCA routine, and a number of factors were accepted whose eigenvalues were above 1. This procedure, along with the histograms for the median densities by slide index can be seen in Figure 5.3 below. As can be seen in the histograms, the dorsal motor nucleus of the vagus nerve is diagnostic for differentiating Parkinson's disease.

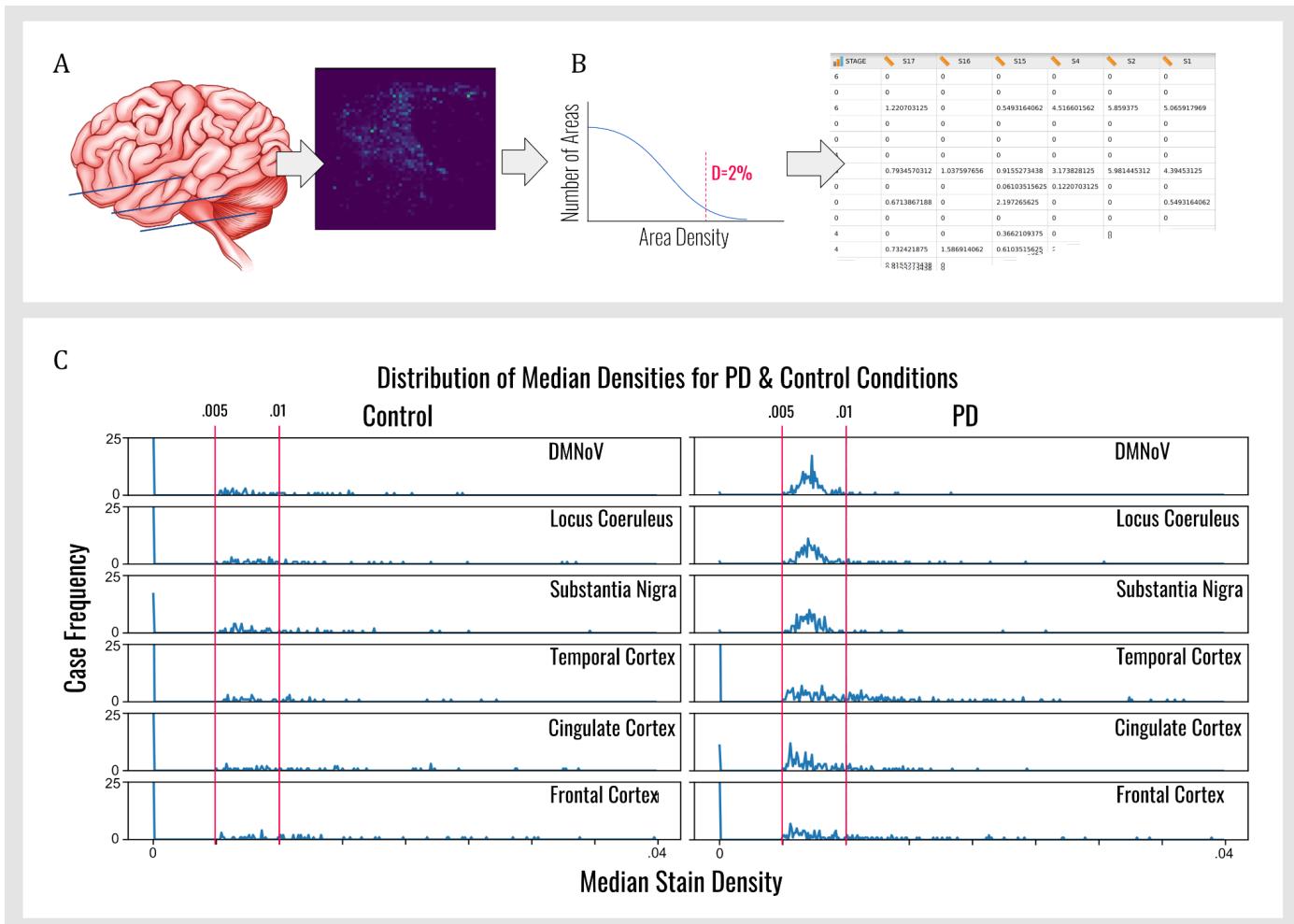


Figure 5.3. Segmentation and Sample Level Density Measurements.

A. Illustration of sample, and density maps generated by the segmentation procedure; **B.** Illustration of the 95th percentile sampling of density distribution, and producing table of densities per case; **C.** The histograms of median densities for the control and PD conditions (all stages) by anatomical location, markers indicate integration region over which reliable classification can be made.

5.1.3 PCA Results

Even though anatomically the frontal cortex is diagnostic of Braak 6, due to the data imbalance, it does not contribute as a factor; instead, the temporal and cingulate cortex are diagnostic in these data. These principal components highlight three main groups: control, minor and major pathology. Cases were collapsed into these groups by using the following criteria: the control group consisted of all cases at stage 0; the minor group consisted of all cases in stages 1 through 5 and the major group consisted of Braak stage 6 cases.

Table 5.4 Slide index, associated anatomical region, and principal component loadings as well as statistical uniqueness of each anatomical region.

| Slide Index | Region | PC1 | PC2 | Uniqueness |
|-------------|------------------|-------|-------|------------|
| 17 | DMNoV | 0.959 | | 0.119 |
| 15 | Substantia Nigra | 0.595 | | 0.595 |
| 16 | Locus Coeruleus | 0.488 | | 0.772 |
| 4 | Temporal Cortex | | 0.652 | 0.600 |
| 2 | Cingulate Cortex | | 0.436 | 0.816 |
| 1 | Frontal Cortex | | | 0.975 |

5.1.4 Statistical Models and Classification Procedure

Firstly, the random forest algorithm (Breiman, 2001) was utilised within JASP to take the 6 element vector as inputs (the median density for each of the six anatomical regions for that case) and produce a classification between either class 0 (control), or 3-6 (Braak 3 - 6). There were 33 trees utilised with 2 features per split. Out-of-the-bag accuracy was chosen as the optimisation constraint; with 256 cases reserved for training and the remaining 143 used for validation, sampled randomly.

Validation accuracy was 60%, with poor performance for Braak 3 (detected with random chance at low false alarm rates, see Figure 5.5) and comparable performance to PDNet for binary discrimination for control cases (~80% for >1% false alarms).

To assess the validity of new classifications, a decision tree was constructed using the algorithm described by James, et al(2013); also conducted using the JASP statistical software. During training, a branching condition of 18 cases was specified as the minimum number of successful classifications to spawn a new node. Data was split into 241 cases sampled randomly for training, and 160 for testing (401 total). The slide median density scores for each of the six slides were used as inputs, and the three classifications (control, minor and major) were used as output labels. The decision tree achieved a test accuracy of 70.6%, requiring only two nodes. The confusion matrix, and decision nodes and weightings can be seen in Figure 5.5 below.

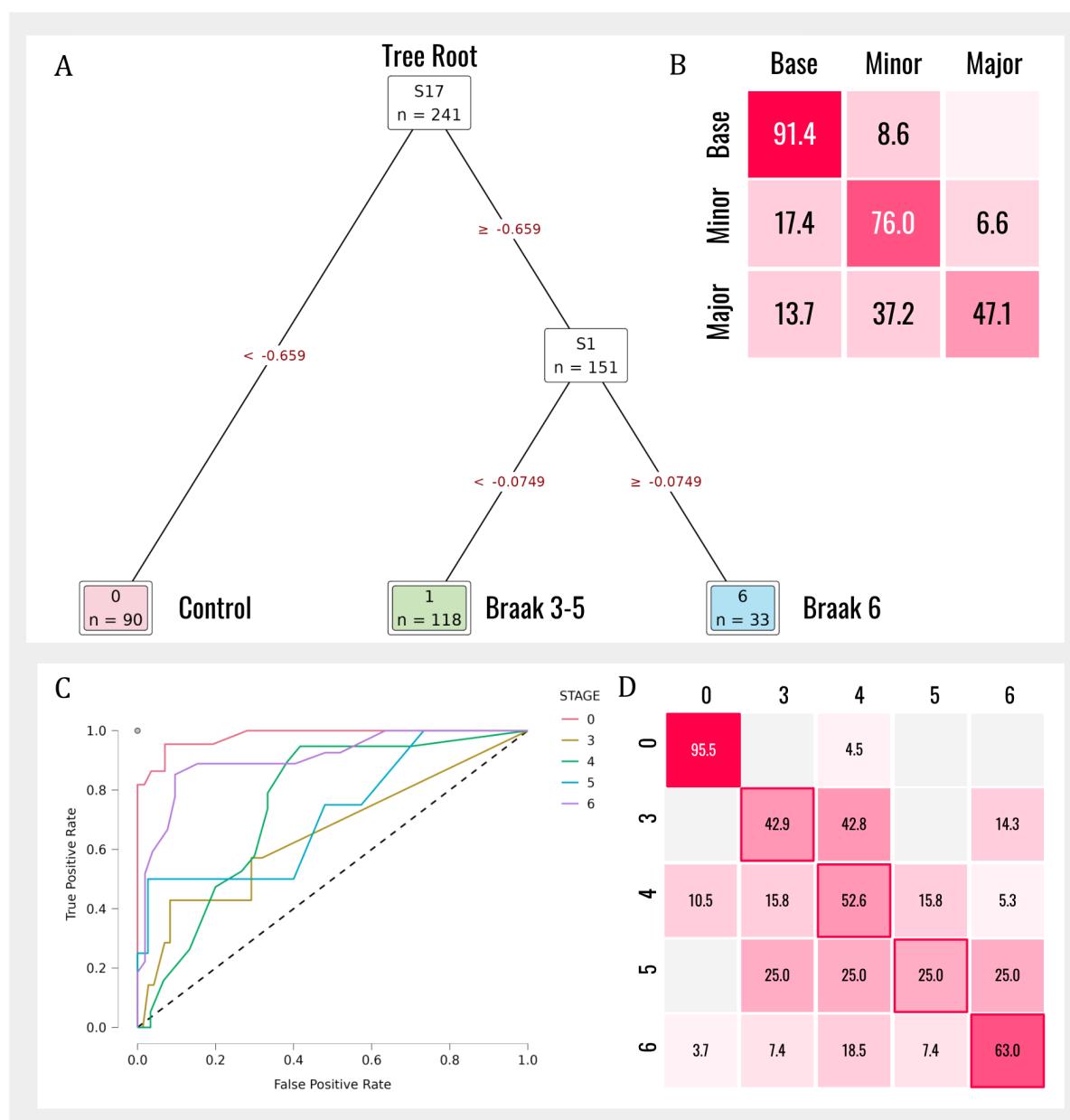


Figure 3.5. Random forest, decision tree and confusion matrices for classification of median density characteristics

A. Decision tree utilising the median density input vector for each case, and reducing that down to either the control, Braak 3-5 or Braak 6; **B.** Confusion matrix for the decision tree; **C.** Receiver operating characteristic curve for random forest algorithm, for each of the Braak stages, using the median density input vector **D.** Confusion matrix based Braak stage classification using the random forest algorithm.

6 Machine Learning on Density Maps

6.1 Multidimensional Whole Sample Classification Summary

A parallel methodology was employed to investigate the use of both single-slide and whole-case density maps for pathology classification. This included, for a given slide index (for example, slide 17 for PD cases vs control), loading the density maps into memory and reshaping them to a uniform size (32px^2) before attempting classification using a small CNN, see Figure 6.1 below. This process was designed to assess the midpoint between simple density statistics and a fully fledged appearance model, and investigate to what extent structural information about pathology progression propagated to the sample level. Over 99.99% of all information from the original samples were discarded for this experiment, using just the raw density maps for each window region, for each case.

The core assumption was that this approach should produce at least as good performance as the basic statistics used to process the density maps, albeit with lower fidelity as some data was used for training, as the CNN expression is a superset of those simple statistical methods. Due to a much reduced training and test set, because of case level classification (from tens of thousands of ROIs, to 87 maps where 87 is the number of cases), performance was not expected to be comparable between the training and test sets (due to likely overfitting). However, the purpose of this set of experiments was to explore the encoded information within the density maps, such that an informed understanding of the data profile and relative classification power of various techniques could be made.

In the previous work we established that a simple ‘whole slide’ measurement such as median or 95th percentile density was sufficient to provide differentiation of both pathology and Braak staging. This work hoped to extend that capability by developing kernels that could span multiple density maps.

The null hypothesis for this work was simply that if the appropriate information was not present consistently within the density maps for each cohort and stratified within each cohort as a function of disease progression, then training would not be possible. This is caveated with the possibility that any commonly utilised model would have enough entropy in its parameters to store the entire dataset; therefore, only validation accuracy must be considered when testing these hypotheses.

The findings of this work package were indeed that classification was possible with reasonable validation accuracy, even with a small number of cases. While it is an important finding that such classification is possible with such little information (often less than 100 data points per case), this line of enquiry shows significantly less sensitivity and specificity than both the simpler models advanced in the Braak staging analyses, or the PDNet performance. Therefore, these experiments are summarised here, but are shortened for brevity.

6.2 Summarised Methods & Results

6.2.1 Targeted Proteins and Differentiation

α -syn

Data was partitioned into PD and control groupings and further partitioned into training and test groups as previously. A total of 110 cases were utilised for training and 291 for testing (243 PD cases, and 48 controls).

$A\beta$

Data was partitioned into PD+AD and control groupings and further partitioned into training and test groups as previously. A total of 50 cases were utilised for training and 50 for testing (7 PD+AD cases, and 43 controls).

6.2.2 CNN Design

A small CNN based on 2D convolutional kernels, with max pooling layers, was utilised. To downsample to a vector of elements equal to the number of classes, a flattening layer was utilised to create a $1 \times N$ array containing the penultimate embedding before a dual fully connected layer was utilised. For the full network diagram see the project repository.

6.2.3 Preprocessing

Density maps were loaded and rescaled to $32px^2$ and downsampled from 32-bit floating point to 16-bit for speed of training and inference. No data augmentations or perturbations were applied to the dataset. Data could be loaded as a 2d array with each map concatenated along the right edge (see A in Figure 6.1), or loaded as $C \times U \times V$ arrays, where each sample was locked with each other in the UV dimensions.

6.2.4 Judgement and Loss

Outputs were treated as a PDF for each of the classes (two for pathology vs control). The dot product of this PDF and the classification label array was taken to get a floating point stage classification, which was then rounded to produce a computed label. Loss was computed by the mean squared error between the computed label (class) and the pathological classification. The Adam optimiser was utilised for gradient descent (see Kingma & Ba, 2014).

6.2.5 Procedure

The broad methodology was similar to that in previous experiments with PDNet. Density maps were loaded into memory, and reshaped and concatenated as described above. These data were then partitioned into test and training sets and passed to the network over 1000 epochs using the PyTorch Dataset class. Once training had completed, the test dataset was used as a simple hypothesis test, with proportionally more classifications for the pathology condition expected in the pathology group. The judgement procedure outlined above was utilised to classify each case in the test set.

6.2.6 Results

Both the α -syn and A β networks trained successfully. Training accuracy for α -syn (PD) pathology vs controls was 99.43%, with validation accuracy of 89.20% correct. Training for the A β (AD+PD) vs controls yielded 95.45% accuracy, with validation accuracy lower at 71.99%. For confusion matrices see Figure 6.1.

More than 99.99% of the data encoded in the original images for this experiment was discarded. The average case consisted of 9.9GB (file size) of pixel data, with density maps reduced to 0.0008GB in size.

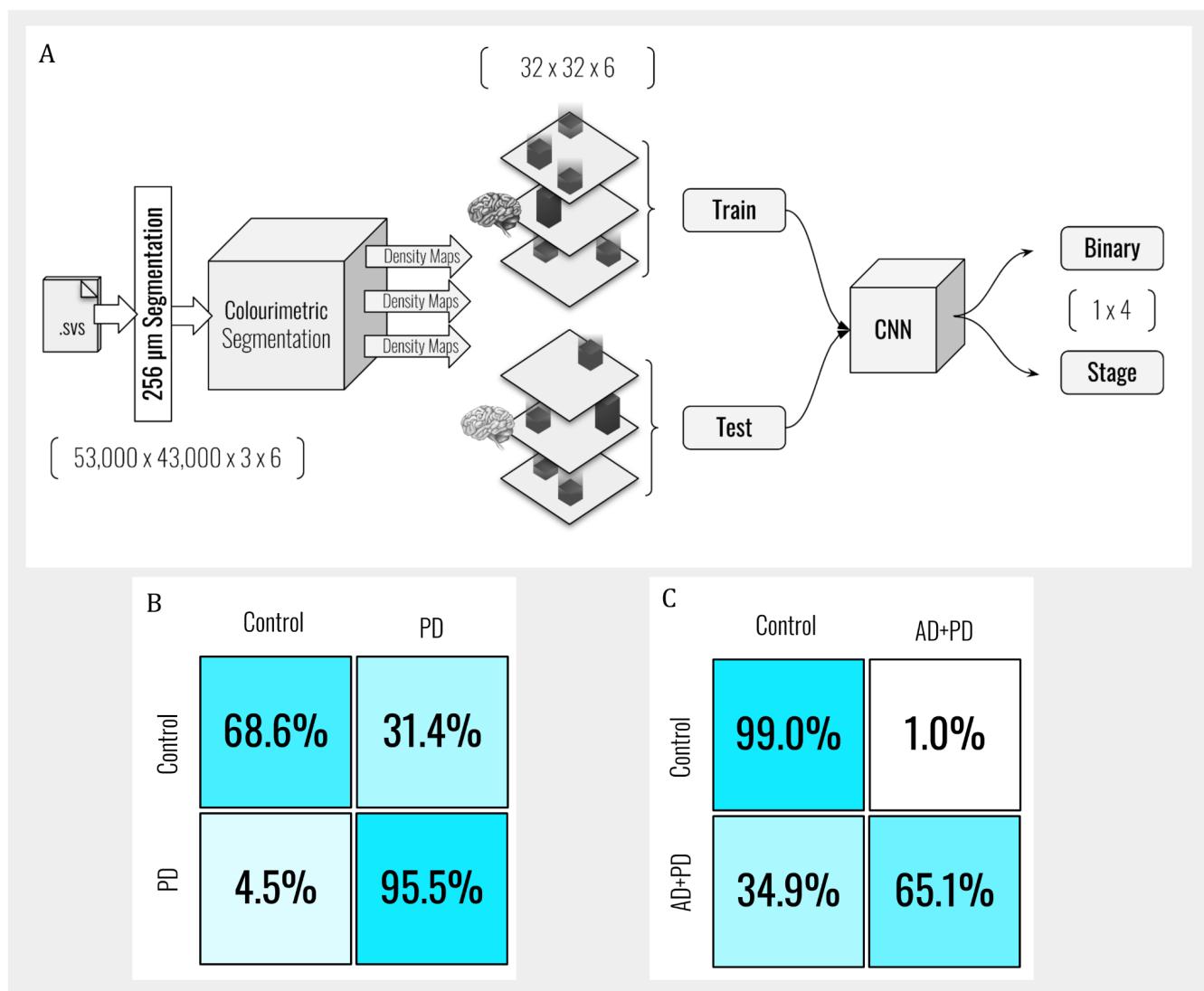


Figure 6.1. Flow and CNN Diagram for multisample classification of Braak Staging.

A. Density maps are segmented by the segmentation procedure and then density maps are generated for each case and concatenated together to produce a set of case tensors. These are split into training and test sets which are then passed to the CNN for training and inference; **B.** Confusion matrix for PD vs Control conditions using the α -synuclein density maps; **C.** Confusion matrix for the AD+PD vs Control conditions using the A β density maps.

7 General Discussion

7.1 Protein Segmentation

These experiments present strong evidence for the spectral decomposition correctly differentiating the DAB signal. Firstly, the density maps produced from the segmentation procedure contained reliably differentiable signals between cohorts. When targeting α -syn, where the presence of protein is expected in the PD group but not the control group, protein was reliably identified in that group, with only minor erroneous detections in the control group. Moreover, the concentration and anatomical location of the segmentations follows the expected disease progression. Moreover, when targeting different proteins (Tau and A β), with differing morphology, the decomposition technique continues to perform correctly, with only configuration parameter changes. For example, A β is reliably identified at a greater density in the PD+AD group.

The decomposition routine was designed without knowledge of spectral qualities of the microscope sensor or the DAB staining transmission functions; there is no evidence to suggest that this technique is restricted to this microscope, stain or indeed targeting protein. This automatic protein segmentation, and density measurement technique should generalise to all non-confusable DAB staining protocols.

During this study, configuration parameters were chosen on the basis of *prima facie* region identification efficacy on a single case. Therefore, there is a high likelihood that the optimal parameters for segmentation have not yet been determined, and more cases should be examined to identify them. There is a high likelihood that the current segmentation procedure only requires those configuration parameters to be chosen for successful segmentation across all DAB stained imagery. Moreover, the current segmentation procedure isolates eosin, but does not use this segmentation information other than isolating eosin as a component plane. It should be the case that complex counterstain protocols could utilise this segmentation technology to measure counterstain efficacy, by expanding the computed density with additional terms (additional decomposition planes).

The challenges of spectral recovery from brightfield imaging outlined by Cukierski (2009), namely that metamerism is present in pathological imagery and that colour is non-neighbourhood preserving, appears to be of negligible effect to the segmentation procedure utilised here. While sensor metamerism necessitates some confusion pixels, especially by those cellular bodies that have a similar pigmentation to the DAB stain, those pixels do not impede macro (slide level) or even ROI measurement.

For simple thresholding of the DAB decomposition for the α -syn samples, ~57% of all cases could be reliably identified with ~2% false alarms. While this accuracy is substantially less than that described later (in the PDNet specific discussion), it does demonstrate a majority of cases are correctly identified with an unoptimised configuration.

A likely key contribution to the efficacy of the segmentation procedure is the simultaneous decomposition of each basis function, including the background cellular matter and the brightfield itself. Bini et al., (2011) used spectral analysis to simulate the histology-like appearance of human skin cells. Indeed, their methodology follows the initial calculations found in Appendix A, of spectrally modelling the hematoxylin response using matrix transformations applied to the RGB values. The results presented here demonstrate that the additional step of simultaneously solving for the contributions of each inverted RGB value significantly improves decomposition. These data also provide evidence that for samples such as these, existing brightfield imagery can produce similar results to elaborate experimental hardware. Wu, et al (2016) use linear estimation and spectral decomposition to perform volumetric microscopy, using bespoke hyperspectral imaging hardware. Here we demonstrate that measurement akin to volumetric estimation (that is, stain density) can be achieved using standardised hardware already found in the laboratory, when the filter functions and their interaction with the lightsource are known.

Framing the measurement of DAB as a linear fitting problem drastically reduces the processing time required to identify the protein. No attempt was made to profile or optimise the research software; however, as the problem reduces to linear decomposition of 8-bit triplets, the BLAS library can compute the DAB decomposition of the full resolution samples in microseconds on desktop workstations. This potentially widens the application of the technology, as demanding compute hardware is not required; with an optimised production system likely able to return visual results to the experimenter in real time.

7.2 Appearance Model Classification of ROIs

The PDNet appearance model yielded accuracy over 90% for the discrimination of single ROIs for each of the three proteins. Moreover, this was consistent between conditions where binary discrimination was expected (α -syn) and where differences were largely morphological (Tau). Case level performance was significantly improved over the previous experiments, with a 4% increase in specificity, and precision approaching human observers (2% difference), while being significantly more sensitive.

PDNet performance also held across ROIs of varying sizes. The A β ROIs were 8x smaller than the α -syn, and yet PDNet provided comparable performance. This suggests that there is significant scope to examine the window configuration during the segmentation procedure, to maximise the opportunity of detecting each protein.

Greater classification accuracy was gained at lower false alarm rates by collapsing ROIs by case. This is likely due to the disbalance of detections between cases, resulting in disproportionate signal moving out of the training set, when a number of cases with a large frequency of ROIs are sampled for the validation set. This suggests that a larger cohort should be utilised when training PDNet for any diagnostic system, to ensure the appropriate statistical power and a representative model. This aside, overfitting was not evident during training, with a representative loss function for both training and validation sets.

A key behaviour of all three models (one for each protein) was very high reliability at classification of the control condition; this is observable in the confusion matrices presented, with only 14% of PD cases missed and no control cases misclassified (false alarms). Moreover, each ROI was not inspected visually by the experimenters, and therefore it is not clear at what rate false segmentations occurred in the PD cases. These false alarms should be uniform between PD and the control cases if there is no interaction between the protein and other confusables; however, we have observed that confusable pigmented cells are less prevalent in the PD cases than in the controls. Understanding this detection profile will require further experiments, and will likely be most successful by examining the outputs of the segmentation procedure, as it is transparent and explainable in its filters.

This project has developed a single, unified model for the detection of DAB stained proteins in neuropathology brightfield imagery. This technology is automatic, in that regions of interest can be algorithmically identified without human interaction. It is explainable, in that each stage of the decomposition procedure is based on sensor physics; and, unlike our previous work (Pearce, 2022), the classification process operates on unmodified raw pixel content; the segmentation procedure identifies ROIs, but it does not synthetically stain or otherwise alter those data.

7.3 Automated Braak Staging & Multidimensional Model

The initial hypothesis, that Braak staging could be reduced to the binary presence or absence of DAB detection at each anatomical region, was largely substantiated. Reducing the raw slides to DAB density maps, and then further to single measures of DAB density represented a 22,688:1 compression ratio for the density maps, and a 907,500,000:1 compression ratio for the 95th percentile density measurement. At this level of compression, the random forest algorithm was able to detect 85.7% of PD cases at a 4.5% false alarm rate, with variable performance at each Braak stage (between 63% - 25% accuracy).

When reducing the categories for staging, a simple decision tree identified slide 17 (the dorsal motor nucleus of the vagus nerve) as the key diagnostic slide for differentiating the control cohort, and slide 1 (the frontal cortex) as differentiating minor (Braak 3) from major (Braak 6) cases.

These experiments demonstrate that standard machine learning algorithms can encode pathological definitions, and that the segmentation procedure can provide diagnostic data at an extreme compression ratio. They also demonstrate ROI, sample and case level differentiation by leveraging the segmentation procedure developed in this work, with median DAB density distribution differences observable at the cohort level.

The multidimensional model approach of classifying the density maps for each case suffered from the same data challenges as the more simple machine learning techniques; however there was evidence of overfitting between the test and control conditions.

Therefore, no further analysis was undertaken on these results, as the classification accuracy was poorer than both the decision tree and random forest classifiers. This also suggests that the median or 95th percentile is a reliable way of weeding out erroneous density regions in the density maps. This approach should be reconsidered should further advancements be made on the segmentation procedure, or should substantially more data become available.

Due to time constraints and unbalanced data sets, we were unable to perform rigorous statistical analysis of the classification power of these techniques. However, we have demonstrated their viability, and further research can be undertaken to tailor both the approach and dependent variables that are most useful to the pathologist. More specifically, it is difficult to determine whether greater variability in the Braak 5 region was due to greater variability in the cases, the difficulty they present to pathologists during staging, or their interaction with the segmentation procedure. However, these are possible research questions for future work. Moreover, these experiments demonstrate that a machine learning algorithm can be applied to the case level statistics (collections of statistics containing multiple samples) and provide diagnostically relevant classifications. This supplements the PDNet approach, which utilises single ROIs from a single sample.

7.4 Scope, General Applications & Conclusions

The purpose of this project was to de-risk the hypothetical colourimetric segmentation, and detection, of stained proteins from neuropathology imagery. This was approached by developing a general segmentation algorithm (the segmentation procedure), and then testing its viability on each of three proteins α -syn, Tau and A β .

These experiments have demonstrated successful identification of each protein at an ROI, sample and case level. This translates to predictive models of pathology and effective cohort separation and classification. Within an individual sample, protein can be identified using the full resolution imagery (RGB brightfield microscope imagery ROIs) and PDNet, or a window-level density measure from the segmentation routine. For a given case, just the median or top (95th) percentile protein measurements are sufficient to provide state-of-the-art binary classification. Moreover, for Tau, which is present across multiple cohorts with varying amounts and morphology, both the density differences and those morphological differences were observable across multiple techniques.

We have demonstrated multiple paths to explore novel research questions in the laboratory that could have significant impact on pathology research. We have also demonstrated how manual classification can be alleviated with automatic protein detection that could be implemented on the microscope hardware, allowing real time feedback to the pathologist. During this project we have also automatically identified over 85% of all the PD and PD+AD cases (over 400) using multiple protein markers, in the time taken for a laboratory examination of a single case, using one protein marker.

The current state of the technology is PoC, requiring lab viability testing, and robustness added to the technique for the variability in laboratory hardware, operating conditions, protocol variation, and staining manufacturer. However, we were able to confirm that the proposed technique yielded viable results under arbitrary conditions, and without extensive analysis of each protein marker.

We conclude that the use of colourimetry and spectral decomposition for sample segmentation is an effective method for automatic identification, segmentation and measurement of DAB stained proteins. Moreover, a variety of machine learning approaches can be employed to provide high accuracy pathological classification.

7.5 Next Steps

The ability to reliably measure any DAB stain pathology, given knowledge of the laboratory equipment utilised in capturing the imagery, has broad applicability in research and diagnostics. Polygeist have outlined the viability, exploitability pathway, and potential for commercialisation to other pathologies, via these technologies, in full, in “Deployment & Exploitation of Automatic Pathology Scanning Technology at Imperial College London” as a commercially sensitive proposal to the stakeholders.

8 References

- Alafuzoff, I., Ince, P. G., Arzberger, T., Al-Sarraj, S., Bell, J., Bodi, I., ... & Kretzschmar, H. (2009). Staging/typing of Lewy body related α -synuclein pathology: a study of the BrainNet Europe Consortium. *Acta neuropathologica*, 117(6), 635-652.
- Attems, J., Toledo, J. B., Walker, L., Gelpi, E., Gentleman, S., Halliday, G., ... & McKeith, I. G. (2021). Neuropathological consensus criteria for the evaluation of Lewy pathology in post-mortem brains: a multi-centre study. *Acta neuropathologica*, 141(2), 159-172.
- Bini, J., Spain, J., Nehal, K., Hazelwood, V., DiMarzio, C., & Rajadhyaksha, M. (2011). Confocal mosaicing microscopy of human skin ex vivo: spectral analysis for digital staining to simulate histology-like appearance. *Journal of biomedical optics*, 16(7), 076008-076008.
- Braak, H., Del Tredici, K., Rüb, U., De Vos, R. A., Steur, E. N. J., & Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of aging*, 24(2), 197-211.
- Breiman, Leo. (2001). Random forests. *Machine Learning*. Springer, 45(1), 5-32
- Carrillo-Mora, P., Luna, R., & Colín-Barenque, L. (2014). Amyloid beta: multiple mechanisms of toxicity and only some protective effects?. *Oxidative medicine and cellular longevity*, 2014.
- Chinchor, N. (1992). MUC-4 evaluation metrics in Proc. of the Fourth Message Understanding Conference 22–29.
- Cukierski, W. J., Qi, X., & Foran, D. J. (2009). MOVING BEYOND COLOR: THE CASE FOR MULTISPECTRAL IMAGING IN BRIGHTFIELD PATHOLOGY. *Proceedings. IEEE International Symposium on Biomedical Imaging*, 5193251, 1111-1114.
<https://doi.org/10.1109/ISBI.2009.5193251>
- Darrodi, M. M., Finlayson, G., Goodman, T., & Mackiewicz, M. (2015). Reference data set for camera spectral sensitivity estimation. *JOSA A*, 32(3), 381-391.
- Finlayson, G., Mackiewicz, M., Hurlbert, A., Pearce, B., & Crichton, S. (2014). On calculating metamer sets for spectrally tunable LED illuminators. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 31(7), 1577-1587.
<https://doi.org/10.1364/JOSAA.31.001577>
- Folch, J., Ettcheto, M., Petrov, D., Abad, S., Pedrós, I., Marin, M., ... & Camins, A. (2018). Review of the advances in treatment for Alzheimer disease: strategies for combating β -amyloid protein. *Neurología (English Edition)*, 33(1), 47-58.
- Glenner, G. G., & Wong, C. W. (1984). Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. *Biochemical and biophysical research communications*, 122(3), 1131-1135.

- Gupta, A., Ramanath, R., Shi, J., & Keerthi, S. S. (2021). Adam vs. SGD: Closing the generalization gap on image classification. In OPT2021: 13th Annual Workshop on Optimization for Machine Learning.
- Hunt, R. W. G. (1991). Measuring colour. Ellis Horwood series in applied science and industrial technology (2nd ed., p. 313 p.). New York: E. Horwood.
- IEC, I. (1999). 61966-2-1: 1999 Multimedia systems and equipment-Colour measurement and management-Part 2-1: Colour management-Default RGB colour space-sRGB. CIE: Vienna, Austria.
- Irwin, D. J., Lee, V. M. Y., & Trojanowski, J. Q. (2013). Parkinson's disease dementia: convergence of α -synuclein, tau and amyloid- β pathologies. *Nature Reviews Neuroscience*, 14(9), 626-636.
- Jakes, R., Spillantini, M. G., & Goedert, M. (1994). Identification of two distinct synucleins from human brain. *FEBS letters*, 345(1), 27-32.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer New York.
- Justel, A., Peña, D., & Zamar, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & probability letters*, 35(3), 251-259.
- Keren, G., Sabato, S., & Schuller, B. (2018, November). Fast single-class classification and the principle of logit separation. In 2018 IEEE International Conference on Data Mining (ICDM) (pp. 227-236). IEEE.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kovacs G. G. (2018). Tauopathies. *Handbook of clinical neurology*, 145, 355–368.
<https://doi.org/10.1016/B978-0-12-802395-2.00025-0>
- Lima, A. A., Mridha, M. F., Das, S. C., Kabir, M. M., Islam, M. R., & Watanobe, Y. (2022). A Comprehensive Survey on the Detection, Classification, and Challenges of Neurological Disorders. *Biology*, 11(3), 469.
- Lopez, C., Swan, J., & Schmidtler, J. (2020). Effect of Perceived Contrast Enhancing Lens Technology on Traffic Signal Detection for Color-Deficient Individuals. In Advances in Human Aspects of Transportation: Proceedings of the AHFE 2020 Virtual Conference on Human Aspects of Transportation, July 16-20, 2020, USA (pp. 333-339). Springer International Publishing.
- McKeith, I. G., Galasko, D., Kosaka, K., Perry, E. K., Dickson, D. W., Hansen, L. A., ... & Perry, R. H. (1996). Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. *Neurology*, 47(5), 1113-1124.

- Näslund, J., Schierhorn, A., Hellman, U., Lannfelt, L., Roses, A. D., Tjernberg, L. O., ... & Greengard, P. (1994). Relative abundance of Alzheimer A beta amyloid peptide variants in Alzheimer disease and normal aging. *Proceedings of the National Academy of Sciences*, 91(18), 8378-8382.
- Olson, A. H. (2013). Calibration of Leica Scanscope AT2. ICC Medical Imaging Working Group. Retrieved from <https://www.color.org/groups/medical/Olson.pdf>, 2 March 2023.
- Pearce, B., Coetzee, P., Rowland, D., Linfoot, S., Dexter, D. T., Gveric, D., & Gentleman, S. (2022). Automatic Sample Segmentation & Detection of Parkinson's Disease Using Synthetic Staining & Deep Learning. *bioRxiv*, 2022-08.
- Rajasekhar, K., Chakrabarti, M., & Govindaraju, T. (2015). Function and toxicity of amyloid beta and recent therapeutic interventions targeting amyloid beta in Alzheimer's disease. *Chemical communications*, 51(70), 13434-13450.
- Ross, C. A., & Poirier, M. A. (2005). What is the role of protein aggregation in neurodegeneration?. *Nature reviews Molecular cell biology*, 6(11), 891-898.
- Rushton, W. A. H. (1972). Review Lecture. Pigments and signals in colour vision. *Journal of Physiology*, 220(3), 1P–31P. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1331666/>
- Shapey, J., Xie, Y., Nabavi, E., Ebner, M., Saeed, S. R., Kitchen, N., ... & Vercauteren, T. (2022). Optical properties of human brain and tumour tissue: An ex vivo study spanning the visible range to beyond the second near-infrared window. *Journal of biophotonics*, 15(4), e202100072.
- Soto, C., & Estrada, L. D. (2008). Protein misfolding and neurodegeneration. *Archives of neurology*, 65(2), 184-189.
- Wang, Y., Mandelkow, E. Tau in physiology and pathology. *Nat Rev Neurosci* 17, 22–35 (2016). <https://doi.org/10.1038/nrn.2015.1>
- Williams, B. J., Ismail, A., Chakrabarty, A., & Treanor, D. (2021). Clinical digital neuropathology: experience and observations from a departmental digital pathology training programme, validation and deployment. *Journal of clinical pathology*, 74(7), 456–461. <https://doi.org/10.1136/jclinpath-2019-206343>
- Wu, J., Xiong, B., Lin, X., He, J., Suo, J., & Dai, Q. (2016). Snapshot hyperspectral volumetric microscopy. *Scientific reports*, 6(1), 1-10.
- Wyszecki, G., & Stiles, W. S. (1982). Color science : concepts and methods, quantitative data and formulae. The Wiley series in pure and applied optics (2nd ed., p. xv, 950 p.). New York: Wiley. Retrieved from <http://www.loc.gov/catdir/bios/wiley041/82002794.html>

Zhang, R., Zhu, J. Y., Isola, P., Geng, X., Lin, A. S., Yu, T., & Efros, A. A. (2017). *Real-time user-guided image colorization with learned deep priors*. arXiv preprint arXiv:1705.02999.

Zhang, Y., Wu, KM., Yang, L. et al. Tauopathies: new perspectives and challenges. Mol Neurodegeneration 17, 28 (2022). <https://doi.org/10.1186/s13024-022-00533-z>

9 Appendix A

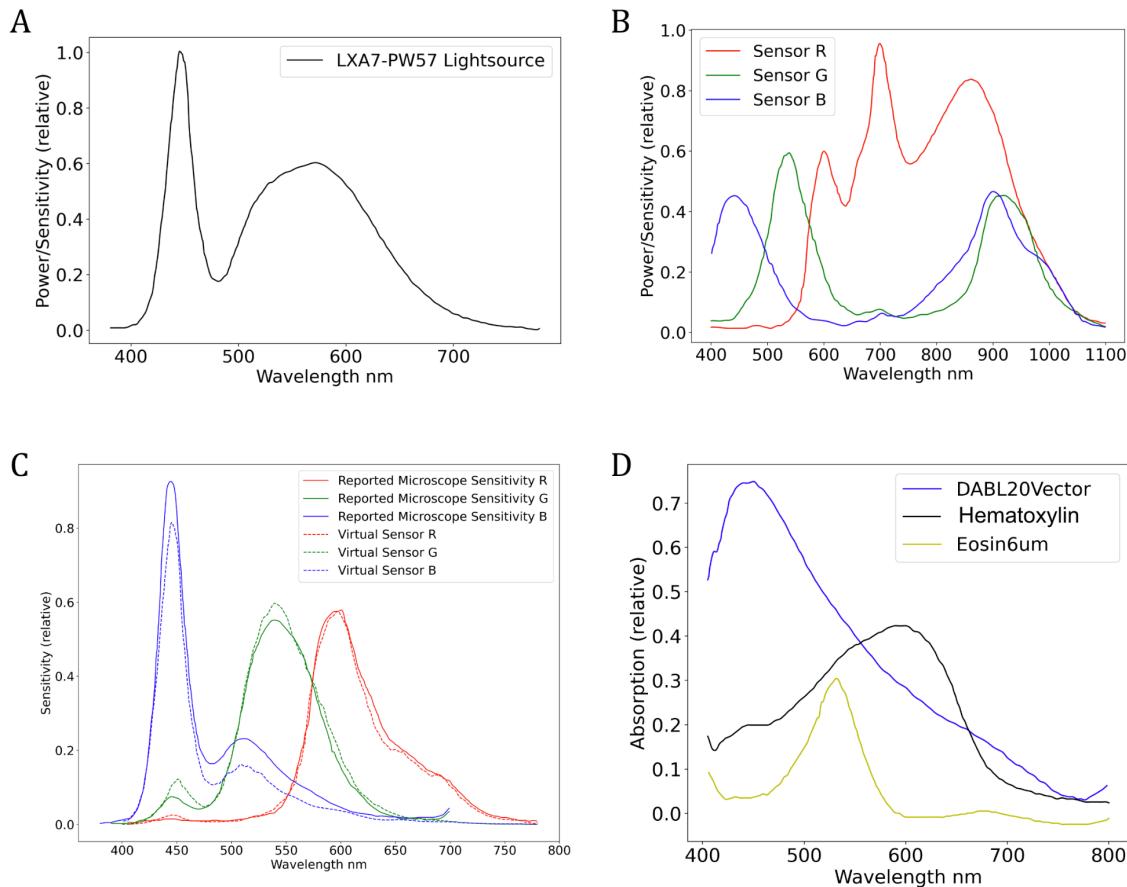


Figure A.1. Spectral power distributions, absorption, sensitivity and transmission functions used in microscope modelling.

A. LXA7 lightsource used in the Leica microscope; **B.** Filtered sensor sensitivity functions for the Leica microscope; **C.** Modelled sensor response for the lightsource and the sensor sensitivities against the reported calibration data from the manufacturer; **D.** Stain transmission spectra, as digitised from Cukierski.

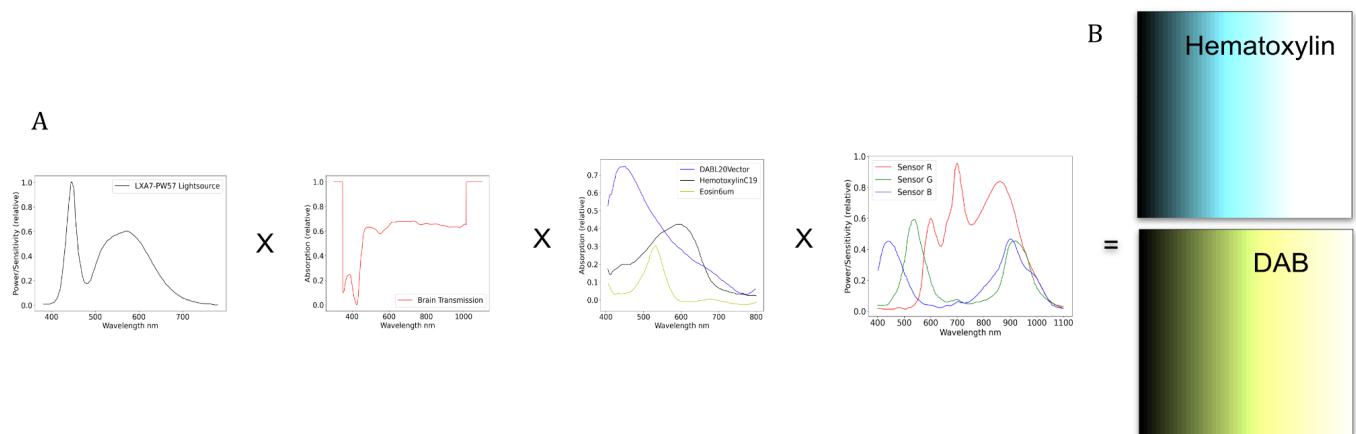


Figure A.2. Modelling example to produce the RGB response from spectral absorption of the media.

A. Combination of functions as found in Figure A.1 to produce sensor response; **B.** Modelled sensor response for the media rasterised as RGB values sweeping with lightsource intensity from completely filtered to bleached.