The impact of fundus autofluorescence on the management of age-related macular degeneration

Roger John Smyth

Doctor of Optometry

Aston University

March 2023

Aston University

The impact of fundus autofluorescence on the management of age-related macular degeneration

Roger John Smyth

Doctor of Optometry

2023

Background
Fundus autofluorescence (FAF) has been described as a topographical map of fluorophores that accumulate within the retinal pigment epithelium as a result of disease.
Study aims
To evaluate whether FAF offers information relevant to age-related macular degeneration over that gathered via colour fundus photography (CFP) and optical coherence tomography (OCT).
Methods
Ninety-three patients were imaged via CFP, OCT and FAF and the results analysed using Orange Data Mining artificial intelligence and SPSS software.
Results
Pupillary dilation makes a significant improvement to FAF image quality.
Nuclear sclerotic cataract of > 1.5 on the World Health Organisation scale indicates that there is ≃85% probability that the FAF image will not be of high quality. At > 1.9 there is ≃50% probability of the image not being clinically useful as defined by a novel grading scale.
Age was negatively associated with FAF comfort.
There is ≥ 90% probability of an abnormal FAF result for an eye with any of the following: > 50 small, > 40 intermediate, > 20 large drusen. Age > 92 years. > 30 packet years of smoking. Any pigmentary abnormalities. ≃80% for any reticular pseudodrusen (RPD).
FAF results can be predicted via CFP and OCT data using machine learning with informedness of up to 70.2% and area under the curve (AUC) of 0.903.
For transfer learning to be useful within primary care, image pre-processing is likely to be required.
Geographic atrophy and pigment epithelial detachments appear to be linked to a patchy FAF pattern. RPD are linked to a reticular FAF pattern.
Principle component analysis indicates that drusen were responsible for the greatest percentage of variability in this study's data (38.6%).
Conclusions
Clinical impact: FAF results can be predicted from CFP/OCT via machine learning with 70.2% informedness and AUC of 0.903. Drusen number/size were the most informative variables.

Key words: Fundus autofluorescence, age-related macular degeneration, colour fundus photography, optical coherence tomography.

To my family for supporting me in this project, and especially to my wife Helen, and my sons Piers and Jonty for their encouragement throughout.

Acknowledgements

My sincere thanks to Dr Hannah Bartlett, my academic supervisor, for assisting me throughout the project, and for her encouragement and expert opinion.

Also, many thanks to Dr Mark Dunne, my associate supervisor, for his help with the statistical analyses and his insight into research techniques.

List of abbreviations

AEN	Anterior embryonic nucleus
AREDS	Age-related eye disease study
AREDS2	Age-related eye disease study 2
ARM	Age-related maculopathy
ARMD	Age-related macular degeneration
ARMS2	Age-related maculopathy susceptibility 2 gene
ALIENOR study	Antioxydants, Lipides Essentiels, Nutrition et maladies OculaiRes
Anti-VEGF	Anti-vascular endothelial growth factor
AF	Autofluorescence
AI	Artificial intelligence
AUC	Area under the receiver operating curve
A2E	Bis-retinoid N-retinyl-N-retinylidene ethanolamine
BCVA	Best corrected visual acuity
BM/RPE	Bruch's membrane/Retinal pigment epithelium complex
BMI	Body mass index
CV	Cardiovascular
CACD	Central areolar choroidal dystrophy
CNV	Choroidal neovascularization
CFP	Colour fundus photography
CATT	Comparison of age-related macular degeneration treatment trials
CFH	Complement factor H
CML	Conventional machine learning
CNN	Convoluted neural network
COR	Cortical cataract
cSLO	Confocal scanning laser ophthalmoscope
DL	Deep learning
DOAJ	Directorate of Open Access Journals
EPR	Electronic patient record
EZ	Ellipsoid zone
ERM	Epiretinal membrane
ELM	External limiting membrane
FA	Fluorescein angiography
FAF	Fundus autofluorescence
FDA	US Food and Drug Administration
FLIO	Fluorescence lifetime imaging ophthalmoscopy

FTMH	Full thickness macular hole
GA	Geographic atrophy
HL	Hyperreflective loci
ICGA	Indocyanine green angiography
IHRF	Intra-retinal hyperreflective foci
IV	Intravitreal
LF	Lipofuscin
LH	Lamellar hole
LIPC gene	Hepatic lipase gene
MA	Macular atrophy
MeSH	Medical Subject Heading Descriptors
MC	Macular cysts
ML	Machine Learning
NIR	Near Infrared reflectance
NN	Neural network
NUC	Nuclear sclerotic cataract
OCT	Optical Coherence tomography
ODM	Orange data mining
ONL	Outer nuclear layer
OPL	Outer plexiform layer
PA	Pigmentary abnormalities
PC	Personal computer
PCA	Principal component analysis
PED	Pigment epithelial detachment
PEN	Posterior embryonic nucleus
PCV	Polypoidal choroidal vasculopathy
PIC	Punctate inner chorioretinopathy
PSC	Posterior subcapsular cataract
QA	Quantum Yield
qFAF	Quantitative Fundus Autofluorescence
RPD	Reticular pseudodrusen
RAP	Retinal angiomatous proliferation
RPE	Retinal pigment epithelium
RNFL	Retinal nerve fibre layer
ROC	Receiver operating curve
SD-OCT	Spectral domain optical coherence tomography

SIRE	Shallow irregular retinal pigment epithelium elevation
SSS	Simplified severity score
SVM	Small vector machine
SW-FAF	Short wavelength fundus autofluorescence
TL	Transfer learning
VMT	Vitreomacular Traction
VA	Visual acuity

List of Tables

List of Figures

Chapter 1

1.1 General introduction

1.1.1 ARMD prevalence

Age-related macular degeneration (ARMD) is a leading cause of sight loss in the industrialised world.(1) The condition can be challenging to assess, particularly with regard to identifying the limit of the retina affected.(2) Short wavelength fundus autofluorescence imaging technology (SW-FAF), from now on referred to simply as FAF, was first described the early 1970s(3) and offers enhanced image contrast between healthy and atrophic retina allowing greater accuracy for the measurement of retinal areas affected by ARMD.(4) FAF imaging has been described as a topographical mapping of lipofuscin (LF) in the retinal pigment epithelium (RPE) as well as other fluorophores that can accumulate with disease in the outer retinal layers.(5)

1.1.2 ARMD Pathogenesis

It has been suggested that geographic atrophy may occur as a result of localised RPE inflammation and death, whilst choriocapillaris loss leads to RPE ischaemia, causing release of vascular endothelial growth factor (VEGF) and choroidal neovascularisation. Histopathological research of ARMD has identified complement inhibitors (C3 and C5) within drusen. It has been suggested that aging leads to a reduction of adhesion of complement inhibitors to Bruch's membrane resulting in complement activation, which in turn leads to drusen formation, release of anaphylatoxins including C3a and C5a and subsequent tissue destruction.(6) Drusen also contain inflammatory proteins which promote a toxic environment, including apolipoprotein E, coagulant proteins, acute phase proteins and immunoglobulin G.(7) With regards to exudative ARMD, complement inhibitor inhibition in mouse models has also been shown to play a role in the reduction of choroidal neovascularisation.(8)

1.1.3 ARMD clinical features

Patients with intermediate (>63<125μm) to large drusen ≥125μm with or without retinal pigment epithelium (RPE) pigmentary abnormalities within the macula are considered to have early ARMD, whilst patients with geographic atrophy and neovascular (wet) ARMD are considered to have the late form of the disease. The characteristics of ARMD include

degenerative changes to the choroid, Bruch's membrane and the RPE. Age is considered an important risk factor for the development and progression of the condition, and degeneration is generally bilateral. Once an eye has developed the wet form of ARMD, the fellow eye is also at an increased risk of choroidal neovascularisation (CNV)(9).

1.1.4 ARMD risk factors

ARMD is a multifactorial disease whose onset is governed by both genetic and environmental factors. Risk factors for dry ARMD include genetic predisposition, with common alleles responsible for between 50-70% of heritability. Ocular risk factors for dry ARMD include reduced choriocapillaris density, drusen, reticular pseudodrusen, Bruch's membrane thickening, thinning of the photoreceptor layers, atrophy and other RPE changes, however, instead of being viewed as distinct entities, it may be more appropriate to consider these features as interrelated, with one defect within the retina leading to, or forming as a result of another, as each layer relies on the overall healthy functioning of the retina to remain patent.(10) Other risk factors for ARMD include increased age, presence of ARMD in the fellow eye, a family history of ARMD, smoking, systemic hypertension, a BMI ≥30kg/m², a diet low in omega 3 and 6, carotenoids, vitamins and minerals, a diet high in fat and lack of exercise.

1.1.5 ARMD grading scales

There are a number of grading scales for describing the stages of ARMD, including, but not exhaustively, the NICE system, the AREDS 9 and 4 step systems, the AREDS simplified severity scale, the Three continent system, the Clinical Age-Related Maculopathy Staging system (CARMS), the Sandberg system and the International ARM Epidemiological Study Group system. These systems are based on the identification of drusen size and type and on the presence of pigmentary abnormalities of the RPE, and in the case of the NICE and the International ARM systems, on the presence of atrophy of the RPE, serous PEDs, choroidal neovascular membranes, fibrous scars and other retinal abnormalities.(11) For this study the AREDS simplified severity scare was utilised as a collected variable for statistical analysis due its simplicity and numerically based scale.

1.1.6 Background and history of FAF imaging

FAF has the benefit of being a non-invasive technique, and can be delivered via a range of commercially available systems, including fundus camera-based, confocal scanning laser

ophthalmoscope (cSLO) based and ultra-widefield imaging based systems, with each modality offering its own merits and disadvantages.(12) Whilst this review will focus on the use of FAF imaging for the investigation of ARMD, FAF has a wide range of other clinical applications in ophthalmology for the evaluation of conditions including optic disc drusen, white dot syndromes, retinal dystrophies and retinal drug toxicities.

The human retina contains a variety of fluorophores, each with their own absorption and emission spectra. FAF works by detecting these fluorophores within the retina, which absorb incident light of various wavelengths, and then release the energy by emitting light of longer wavelengths.(13) Note that the blue excitation wavelength used in FAF may cause patient discomfort and even prove toxic to the retina, although this latter limitation has not been corroborated by formal studies.(12)

The first person to describe fluorescence was Sir John Frederick William Herschel (1792-1871), demonstrating the phenomenon with a glass of tonic water (containing the fluorophore quinine), sunlight and observation by the naked eye in 1845. Fluorescence was first illustrated diagrammatically with the Jablonski diagram, proposed first by Polish Physicist Aleksander Jablonski in 1933, to show how light energy is absorbed and emitted during fluorescence.(14) The absorption of energy raises an electron to a higher state, with the electron then losing some of its energy by colliding with other particles and falling to a marginally lower energy level (non-radiative transition), before returning to, or near to, its original, lower state, whilst in the process releasing energy as light of a longer wavelength. This shift towards longer wavelength light is known as the “Stokes Shift”, and was first described by the Irish Physicist Sir George Gabriel Stokes in 1852.(15)

Fluorescence is a type of photoluminescence event (photo meaning light, luminescence meaning the emission of light), which is different from another form of photoluminescence called phosphorescence which persists for a longer period following stimulation by light. Excited states in fluorescence are momentary, lasting approximately 10^{-8} seconds, and the molecular structure and the chemical environment dictate the intensity of the fluorescence. Sophisticated optical equipment is therefore required to detect and measure fluorescence(15), with the equation governing energy levels being as follows:

$S0 + hv = S1$

Where:

S0 is the ground state of the fluorophore

S1 is the first excited state

h is Planck's constant

v is the frequency of the exciting light

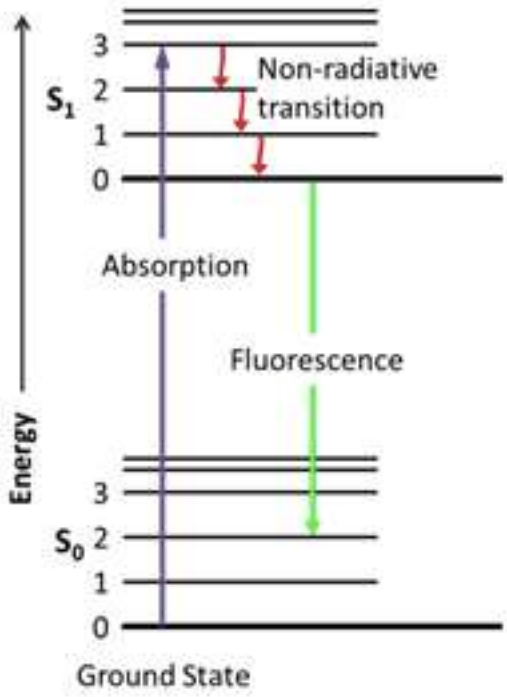

Figure 1.1 The Jablonski diagram

The Jablonski diagram (named after the Polish physicist Aleksander Jablonski, the "father" of fluorescence spectroscopy) showing how electrons are raised to higher energy states by specific wavelengths of light, before returning to, or close to, their original state whilst releasing light of a longer wavelengths than that used for excitation.(16)

The fluorescence “quantum yield” (QA or Φ) is the efficiency of the fluorescence process, and is given as the ratio of the emitted photons to the absorbed photos. If every photon absorbed results in an emitted photon, then the QA = 1. Substances with a QA as low as 0.10 have been considered to possess significant fluorescent properties, however, the major constituent of LF, Bis-retinoid N-retinyl-N-retinylidene ethanolamine (otherwise known as A2E) has been shown to possess a QA of only 0.003 ± 0.001.(17) This is important, as it means that the human retinal FAF signal obtained from LF is of a very low order magnitude. Below is an illustration of retinal layers and the direction of the exciting light and subsequent autofluorescence produced.

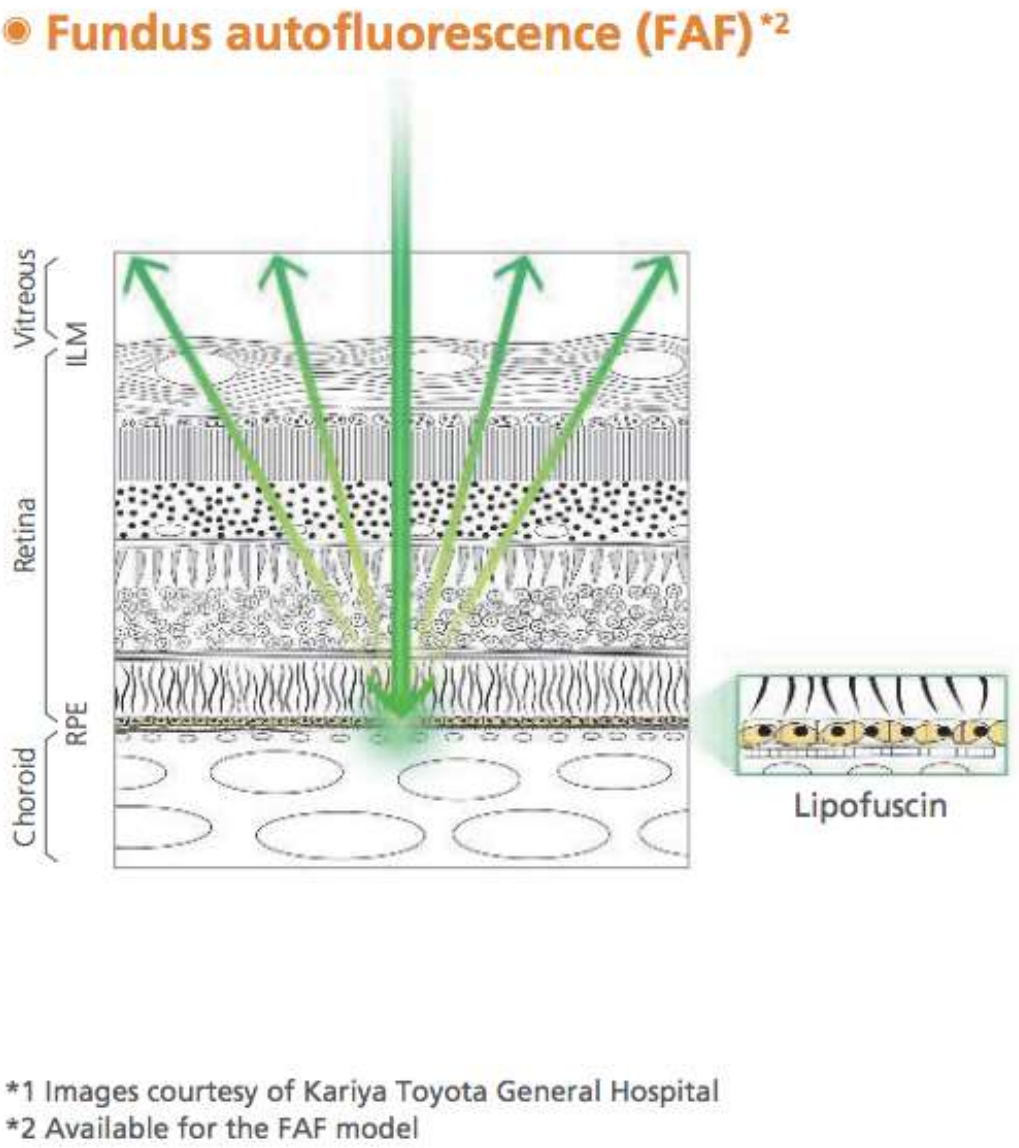

Figure 1.2 Image demonstrating fundus autofluorescence within the human retina
Image reproduced by kind permission of NIDEK from the RS-330 manual

It is important to note that structures anterior to the retina also naturally auto-fluoresce which can cause interference with FAF imaging. The cornea excites at 365-480nm, emitting at 620nm,(18) whilst the crystalline lens absorbs at 420-430nm, emitting at 520nm.(19) Cataracts compound this problem by increasing lens light absorption and scatter further still,(20) whilst glycation products in diabetic individuals can increase corneal autofluorescence.(21) These autofluorescence artefacts are a major limitation of FAF, as well as its low signal strength, which is up to 100 times less than the peak signal derived from fluorescein angiography (FA).(12)

Below is a table showing a variety of retinal diseases and the associated FAF imaging findings.(22)

Retinal disease type	FAF findings
Central serous chorioretinopathy	Hypo-fluorescence at site of leakage, with hyper-fluorescence related to precipitates
Stargardt's disease	Hypo-fluorescence of the entire macula surrounded by hyper-fluorescent flecks with peripapillary sparing
Angioid streaks	Hypo-fluorescence maps to the ruptures in Bruch's membrane
Plaquenil (Hydroxy-chloroquine) toxicity	"Bullseye" maculopathy, with a parafoveal ring of hyper-fluorescence
RPE tears	Well demarcated areas of hypo-fluorescence with adjacent hyper-fluorescence due to rolled up redundant RPE
Best's Macular Dystrophy	Vitelliform stage: well circumscribed homogenous hyper-fluorescence of the macula
Retinitis Pigmentosa	59% show hyper-fluorescent parafoveal (Robson-Holder) ring. 18% have an abnormal central hyper-fluorescence which extends to the fovea. 23% have neither pattern
Choroideremia	Bilateral symmetrical midperipheral zones of hypo-fluorescence due to RPE atrophy
White Dot Syndromes	Multifocal hyper-fluorescence
Punctate Inner Choroidopathy (PIC)	Hypo-fluorescent spots often with hyper-fluorescent margins
Fundus Albipunctatus	Severely attenuated background fluorescence
Deferoxamine-induced retinal toxicity	Minimal, focal, patchy or speckled fluorescent patterns.

Table 1.1 AF findings in a variety of retinal diseases

1.1.7 Background to machine learning

1.1.7.1 Introduction

It was scholar John McCarthy, in 1956, who proposed the concept of AI in the "Dartmouth Workshop", stating that computers could behave in an apparently intelligent manner. Shortly after this, in 1959, machine learning (ML) was first referred to by Arthur Samuel(23) as a subset of AI, whereby computers detect patterns within data, "learning" from this information (i.e. creating algorithms that extract generalised principles from the data), before applying this knowledge to new data. These profound suggestions have only recently become

practically relevant with the advent of “Big Data”. Deep learning (DL) is an advanced form of ML that harnesses layers of neural networks (NN) that mimic the structure of the human brain. Examples of NNs include Microsoft’s Cortana, Amazon’s Alexa and Apple’s Siri. In healthcare, NNs have already been put to good use for disease recognition in fields as diverse as skin and lung oncology, cardiology and ophthalmology.(24)

1.1.7.2 Big Data

Several definitions have been proposed for Big Data, with IBM’s succinctly characterising it as any one or more of the “V” words: Volume, Variety and Velocity.

Volume refers to the large amounts of data that are constantly flowing in from a range of sources including the internet, as well as social media channels such as Twitter, Facebook and Instagram.

Variety concerns the different types of data, for example, structured data such as mobile phone location data and web browsing records. However, Variety also includes unstructured data, for example the data from blogs, audio and video.

Velocity involves the ever-increasing speed of data capture, meaning that decisions regarding the incoming data also need to keep pace with the information flow. These decisions may influence the next data selected to be captured, and therefore this element may add an additional dimension to the concept of data Velocity.(25)

Volume and Variety in ophthalmology could represent the large quantity and types of data being collected daily e.g. via digital retinal photography, OCT scans, corneal topographers, visual field plots and electronic patient records (EPR).

Examples of Velocity in ophthalmology could be the ever-increasing scan speed of OCTs, the burgeoning field of healthcare apps and the evolving complexity of EPR software systems.

Elsewhere, a fourth “V” has been proposed, namely Veracity. Veracity is a measure of the quality and accuracy of data and ideally should be complete, consistent, clean, current and compliant (26). Disappointingly, EPRs are unlikely to meet all of these demands in the majority of contemporary medical databases.

It is interesting to note that MeSH 2023 does not explicitly refer to Veracity, and describes Big Data as “extremely large amounts of data which require rapid and often complex computational analyses to reveal patterns, trends and associations, relating to various facets of human and non-human entities”.

AI allows users to perform complex Big Data analysis that would not be practical, or even possible for humans to calculate mathematically, and therefore Big Data and AI are

inextricably entwined (along with the integration of developments in data analysis and information technology).

Whilst Big Data has the potential to assist in clinical decision making, we must ensure that the algorithms produced are representative of real populations. Biases in patient selection could result in misdiagnoses and poor outcomes for those groups not represented. One solution to this problem is the process of data synthesis for boosting underrepresented data.

Other definitions of Big Data from recent literature:

Author	Definition
Boyd and Crawford (2012)(27)	A cultural, technological and scholarly phenomenon that rests on the interplay of technology, analysis and mythology
Chen et al (2012)(28)	The data sets and analytical techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis and visualisation techniques
Dumbill (2013)(29)	Data that exceeds the processing capacity of conventional database systems
Jacobs (2009)(30)	Data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time

Table 1.2 Other definitions of Big Data from literature

1.1.7.3 Conventional Machine Learning

Random forests and SVMs are the most common CMLs found so far in ophthalmological research.(24) Below Is a description of the most common CMLs:

1. Decision Trees

Decision trees are algorithmic models that look at the incoming information and split this data into a tree like structure. The tree “root” is at the top, with the initial root node representing the entire data set. Attributes are assessed and decisions are made to split the data at each bifurcation into smaller and smaller decision nodes until a leaf node is reached, which cannot be split down any further. The learning process is continually updated and based on feedback. The assessment of attributes is managed by attribute selection measures which can be based on a variety of techniques including Information Gain, Entropy, Gini Index and Gain Ratio.

2. Random Forests

Random forests (which are built from numerous decision trees) organise root and decision nodes randomly, utilising “bagging” to create the required predictions. Bagging involves using a separate training dataset and allowing the decision trees to generate predictions, which are then ranked with the best trees being chosen for the final predictions.

3. Support Vectors Machines

SVMs build a “hyperplane” which is a boundary separating the two classes as widely as possible, so that classification errors are kept to a minimum.

4. Bayesians Classifiers

Bayesians classifiers are a family of algorithms based on the principle that the value of a particular feature is independent of the value of any other feature. For example, if an orange is 10cms in diameter, orange in colour and has a pitted surface, each of these features contributes independently to the probability that the fruit is an orange, regardless of correlations that may exist between the various features. A classifier is then built which

constructs a probabilistic model of the features and uses that model to predict the classification of new data.

5. k-nearest neighbours

This technique is utilised for supervised machine learning. k-nearest neighbours works by calculating the distances between the new (as yet unclassified) data point and all the previously recorded data points, and then selects the specified number (k) of points which are closest to the new point. For classification, the most frequently occurring previous point determines the label given to the new point (for classification, i.e. in the case of a categorical target variable) or averages the labels (for regression, i.e. in the case of a numeric target variable).

6. k-means

k-means is an AI technique that may be utilised for unsupervised machine learning. It classifies new data into a number (k) of clusters based on similarities between the data points.

7. Linear Discriminant analysis

Linear determinant analysis is a generalisation of Fisher's linear discriminant (a statistical technique), which finds linear-combinations (predictive functions) of features that classify data.

8. Neural Networks

An NN is an AI method by which computers process information in a way inspired by the structure of the human brain. NNs feature a layered structure of nodes. Logistic regression is an example of a very simple NN with no hidden layers.(24)

1.1.7.4 Deep Learning

The first AI efforts employed human “knowledge engineers” who extracted domain specific features, using the AI for classification only. However, this was superseded by ML which was tasked with both feature classification *and* extraction tasks. This approach is especially useful in retinal imaging, as each patient now effectively presents a “Big Data” challenge, with the advent of ever improving OCT resolution and the subsequent finding of novel features and subclinical diagnoses.(23) The same is true for other healthcare fields where ML has so far proved successful, including pathology, radiology and dermatology.

Whilst ML is concerned with computers being able to think and act with relatively little human intervention, DL is concerned with computers processing data in a way that can be described as actually modelled on the neural structure of the human brain. DL algorithms can be described as “black boxes”, as they make connections that are too highly dimensional and convoluted to be understood or interpreted by human operators. In other words, DL algorithms employ greater complexity at the expense of interpretability. Examples of DL algorithms include convoluted neural networks (CNNs), deep kernel machines and deep recurrent NNs, with the most suitable for imaging data being CNNs. CNNs are composed of layers which carry out “convolution”, a mathematical process allowing an individual neuron to process data from its receptive subfield,(23) and outperform CMLs. A key advantage of DL over CMLs is that the performance continuously improves with the size of the training dataset, and this benefit has been realised within realistic time frames with the advent of ever more powerful computers.(23) However, whilst DL is preferable to CMLs for large annotated datasets when accuracy is the primary aim, CMLs still have a place when the annotated datasets are smaller or when transparency is paramount over maximising performance.(23)

Training DL models usually requires a large number of well-annotated library images specific to the field of study, which limits its usefulness to certain circumstances. For example, in 2016, DeepMind Health and Moorfields Eye Hospital NHS Foundation Trust partnered to create a training set using 14,884 OCT volume scans from 7621 patients.(31) This research produced encouraging results, matching or outperforming the eight experts involved in the study, when relying solely on the scans provided. Another study published in 2017 and based in South Korea, used a comparatively smaller training set from 399 participants based on retinal nerve fibre layer (RNFL) scans and visual field tests to detect glaucoma. Similarly, the results were impressive, offering a sensitivity of 0.983 and a specificity of 0.975.(32) In 2018, Ahn et al also published research involving DL and glaucoma, but only using CFPs

from just over 1000 images. Whilst the results did not quite match the performance of the South Korean study, they were still encouraging with the added convenience of using only one imaging technique.(33)

1.1.7.5 Transfer Learning

ML does not always need to be trained on a large repository of closely related images associated with a particular field. With a specific type of ML called "Transfer Learning" (TL), is possible to retrain an algorithm on a specific data set after initial training on a diverse population of images. This has led to image CNN classification architecture models being available in a ready to use format, for example, Google's Inception V3[6] which is trained on 1.2 million images from the ImageNet library, with other examples of these models including Alexnet and ResNet. ImageNet contains, as of 2016, over 14 million images(23) of real life objects including vehicles, animals, tools etc. In this way, TL takes ML from the study of one discipline and applies it to another quite different area of research.(34) This approach has been used to train classifiers after initial pre-training, to recognise features from, for example, OCT-B images, and need only a fraction of the specialised data for training that would otherwise have been required.(35, 36)

1.1.7.6 Deep Learning for enhancing OCT imaging

Speckle noise affects the quality of OCT images and previous attempts have been made to improve images via frame averaging techniques. These, however, lead to longer scan durations with a subsequent decline in patient comfort. A study published in 2018 reported that images captured with the Heidelberg Spectralis OCT had been processed with a DL algorithm trained with 2328 "clean" OCT B-scans coupled with the corresponding "noisy" B-scans. The network was able to clean 1552 B-scans of optic nerves with a processing time of 20ms.(37) In December 2021 Nidek® also launched a similar DL OCT denoising software in its NAVIS-EX image filing software. The Nidek Retinascan Duo RS-330 used in this study, however, used the previous version on NAVIS-EX without this additional feature as data collection concluded in February 2021.

1.1.7.7 Orange Data Mining

Scientists often lack the skills required to analyse images via computer technology, and therefore a user-friendly system for image analysis that could easily be employed by new users, after only a short period of training, would be of great benefit. Orange Data Mining (ODM), a free open source software package, offers both a user-friendly AI system and a TL image analytics "add-on". The user can create a workflow of widgets to process, model and visualise the data, with the facility to check the progress made after each step in the process

(by connecting a Data Table widget to any widget in the workflow). SqueezeNet is one of a number of programs that can be selected within the ODM embedding widget, which is also trained on the ImageNet library. SqueezeNet is a small deep CNN, which still delivers AlexNet-level accuracy with 50 times fewer parameters. AlexNet is a CNN that won the ImageNet Large Scale Recognition Challenge in 2012,(38) a competition that has run annually since 2010 involving the classification of 1.2 million natural images into 1000 categories.(39) CNNs models have recently reported to have reached a human level of ability for this specific image identification task.(40) Small deep CNNs have the advantage of being more amenable to use on hardware with limited memory, e.g. a personal computer (PC). SqueezeNet has the added advantage that the analysed images do not need to be uploaded onto a remote server (unlike Google's Inception V3[6]), but rather can be processed locally on the user's own PC.(41)

1.1.7.8 Building and evaluating Artificial Intelligence Models

The steps for building an AI model include the pre-processing of images, splitting the data into separate test and training sets, and finally evaluating the results. Pre-processing will normally involve noise reduction, normalising data, ensuring data from different sources is integrated and adjusted to a common scale, and extracting the features that are most relevant to the conditions being studied. A disadvantage is that during the initial pre-processing information will be lost. For example, haze in an ocular photograph or scan may indicate other pathology that the physician could find useful, e.g. a poor macular scan may indicate media opacities or haze due to other conditions such as uveitis. Also, the pre-processing will be enhancing the algorithm's performance for detecting a single disease, and so numerous pre-processing branches may be required for ML to match the ability of the human physician to detect a wide variety of pathology. To evaluate the results, recall (or sensitivity), specificity and area under the receiver operating characteristic curve (AUC) may be used. The receiver operating characteristic curve (ROC) is created by plotting the true positive rate on the vertical axis (tpr or sensitivity) against the false positive rate on the horizontal axis (fpr or (1 - specificity)). A perfect classifier will have a graph that hugs the top left-hand corner, with the worst following the straight line that runs from the bottom left to the top right of the graph. There are various methods to score the model's performance on the ROC graph, however the most commonly used is the AUC. AUCs of effective models give values between 0.5 and 1.0, with a higher value indicating a better model performance.(24) 0.5 represents a random guess, whereas 1.0 indicates 100% sensitivity and specificity. The

AUC may therefore be used to represent important diagnostic and prognostic outcomes for AI analysis endpoints.(23)

1.1.7.9 Supervised and unsupervised machine learning

Supervised and unsupervised learning are two forms of ML. Supervised learning is to train a model with data that has already been labelled by a human researcher, and tunes the influence of the inputs to optimise the predictions that the model is making. Human labelling can be challenging, and therefore supervised ML can also be “weakly supervised”, to deal with partly or unclearly labelled data. Unsupervised learning is training the model without human labelling, with the model independently searching for patterns within the data and thus creating mathematical models to describe the structure of the data.(23) Accuracy has so far proven to be superior from the application of supervised machine learning by most healthcare research projects.(24)

1.1.7.10 Artificial Intelligence in ophthalmology

How diseases are diagnosed depends to a large degree on the physician’s experience and knowledge, much of which is based on pattern recognition. Therefore, AI potentially has an important role to play in assisting less experienced clinicians to make clinical decisions by detecting patterns within treatment choices made by senior clinicians. Furthermore, EPRs, which essentially represent large data sets, could be examined by AI to detect patterns contained within them e.g. the patient history, signs, symptoms and clinical findings which would otherwise not be humanly possible to detect. AI has been used within ophthalmology for anterior and posterior eye disease recognition including keratoconus, cataract, diabetic retinopathy, glaucoma, ARMD, retinopathy of prematurity, retinal detachment, retinal vein occlusions, strabismus, ocular oncology and peri-orbital trauma, and its usefulness is likely to continue expanding.(42). In particular, AI has been shown to be capable of impressive levels of accuracy for detecting, classifying and quantifying GA from features observed via OCT imaging. However, when considering CFP, OCT and FAF data captured in day to day ophthalmology / optometry clinics, images may be of lower quality compared with the “clean” images that have been carefully chosen and manipulated specifically for use in research involving AI. As a consequence of this problem, many recent papers have focused on using AI to assess the quality of images in order to address this issue (43).

1.1.7.11 Discussion

The fast-moving field of Artificial Intelligence (AI) has shown promise in the field of healthcare, with AI being harnessed to assist less experienced clinicians to make decisions based on treatment choices made by senior colleagues, as well as analysing large EPR data sets including clinical findings, symptoms, and history. Machine Learning (ML) is a branch of AI through which computers can learn from these large repositories of information and apply the algorithms created to novel data. Deep Learning (DL) is an advanced form of ML, which is modelled on the human brain's complex network of neural layers which can create "black box" diagnostic tools. DL has already produced encouraging results in ophthalmology including studies in the fields of medical retina and glaucoma, however, ophthalmology graders must appreciate the inherent risks, errors, biases and reliability of these "black box" tools to utilise them correctly for the routine management of ocular disease.(23) Transfer Learning (TL) is a technique that has great potential by eliminating the need for large training data sets based specifically on the subject area, with the machine learners instead pre-trained on large libraries of images unrelated to the field of study before final, and relatively lean training on the subject area, with subsequent economy of resources. TL is therefore an exciting opportunity for optometrists, enabling the use of off-the-peg software packages retrained on ocular images to efficiently identify specific pathologies across a wide variety of ophthalmological disciplines.

1.2 Literature Review: The role of fundus autofluorescence in the detection and monitoring of age-related macular degeneration.

1.2.1 Purpose, aims, recent findings and summary of literature review

Purpose/Aims: Fundus autofluorescence (FAF) was first described in the 1970s,(3) however, its value in the detection and monitoring of age-related macular degeneration (ARMD) is still unclear.(44) A review of the available literature was conducted to identify the clinical value of fundus autofluorescence in ARMD.

Recent findings: FAF can offer improved detection and repeatability when measuring geographic atrophy (GA) area compared to colour fundus photography (CFP), and is also superior to CFP for detecting small areas of GA(2) and halo effects surrounding areas of GA.(10) Halo effects have been found to be indicative of further likely expansion of GA (45) Enface spectral domain optical coherence tomography (SD-OCT) may offer an alternative to FAF imaging for the detection of halo effects. FAF has also been reported to be superior for reticular pseudodrusen (RPD) detection compared to SD-OCT and CFP,(46) and therefore FAF may have a role in ARMD prognosis given the associations between RPD and the progression to advanced ARMD(47) .

Summary: FAF is superior to CFP for the detection of small areas of GA, quantifying areas of GA,(48) detection of halo effects surrounding areas of GA(10) and is superior to both SD-OCT and CFP for the detection of RPD.(49) FAF may therefore have a role in the prognosis of ARMD, given the association of these findings with progression to advanced disease.(50) Enface SD-OCT could provide an alternative to FAF for the detection of halo effects surrounding areas of GA.

1.2.2 Key Words

1. Fundus autofluorescence
2. Age-related macular degeneration
3. Colour fundus photography
4. Optical coherence tomography

1.2.3 Methods of Literature Search

Relevant articles were identified that reported on FAF in relation to ARMD published in peer reviewed journals through a multi-staged approach. In the first stage, a computerised search of the Web of Science database, the SCOPUS database, the PubMed database (National Library of Medicine), the Embase database, the Cochrane database and the Directorate of Open Access Journals (DOAJ) was performed to identify all relevant articles published between 1970-74 (depending on the database) and December 2022. Terms and words used for the search included “fundus autofluorescence”, “age-related macular degeneration”, “colour fundus photography” “optical coherence tomography” and MeSH descriptors for these terms, as well as common abbreviations and variations. Searches were limited to papers based on “adult humans” and articles published in all languages were included, provided an English translation could be obtained.

In the second stage, abstracts of the articles found via the search strategy were examined to identify papers that included studies involving the use of FAF imaging for the detection, evaluation and monitoring of ARMD, and/or those that compared FAF to other retinal imaging modalities including colour fundus photography (CFP) and optical coherence tomography (OCT). In the third stage, full articles were reviewed (including their bibliographies using the same guidelines) and relevant information was incorporated into the manuscript.

1.2.4 Inclusion criteria (PICO items listed below)

Population: Adult humans over the age of 50 years. Having ARMD was not an essential inclusion criterion i.e. the study also included healthy participants.
Interventions: Retinal fundus autofluorescence imaging, colour fundus photography, optical coherence tomography.
Comparisons: Retinal fundus autofluorescence in relation to colour fundus photography and/or optical coherence tomography, for the imaging of eyes with and without age-related macular degeneration.
Outcomes: To include information on the detection, evaluation and monitoring of age-related macular degeneration.
To include: Peer-reviewed articles. Access to full articles was not strictly required, and the search also included conference abstracts.
Articles published in any language provided an English translation is available.

1.2.5 Exclusion criteria

Non-peer reviewed articles.
Articles not judged to be clinically relevant.
Studies not involving adult human subjects.
No English translation available.

1.2.6 Databases used

Web of Science, SCOPUS, PubMed, Embase, Cochrane database, and the Directory of Open Access Journals

Web of Science - general science-based database
SCOPUS – general science-based database
PubMed - alternative database which also uses Medical Subject Headings (MeSH) for searches
Embase – database covering a wide range of scientific, medical and healthcare disciplines (all journal articles available by reference to the Medline database are also available through Embase).
Cochrane - an internationally renowned and trusted source for searches
DOAJ – general database of open access journals

1.2.7 Key Words

1. Fundus autofluorescence
2. Age-related macular degeneration
3. Colour fundus photography
4. Optical coherence tomography

1.2.8 Abbreviations of Key Words used

1. FAF, AF
2. ARMD, ARM
3. CFP
4. OCT

Medical Subject Heading (MeSH) descriptors are available for Key Words 1, 2 and 4 with the results detailed below. Boolean operators were selected to encompass all MeSH descriptors for these terms, with exception of Macular Dystrophy as detailed below.

1.2.9 Medical Subject Heading Descriptors for Key Words (Exact Term Match), searched on 15.12.2022.

For 1.

Autofluorescence Imaging
Fluorescence Imaging
Fundus Autofluorescence Imaging

For 2.

Age-Related Macular Degeneration
Age-Related Maculopathies
Age-Related Maculopathy
Macular Degeneration, Age-Related
Macular Dystrophy – **not relevant for this project on ARMD**
Maculopathies, Age-Related
Maculopathy
Maculopathy, Age-Related

For 3.

Not featured in MeSH December 2022

For 4.

OCT Tomography
Optical Coherence Tomography

Search strategy table with Boolean descriptors for Key Words

Search date 15.12.2022	**OR**	**OR**
Concept 1 Fundus NEAR/5 *fluorescence Notes: to capture the words in either order in close proximity	FAF	AF
AND		
Concept 2 Age-related NEAR/5 macul* degeneration Notes: to capture words in either order, in close proximity, and to include macular and maculopathy	ARMD	AMD
AND		
Concept 3 Colo$r fundus photography Notes: to capture both US and UK spellings of “colour”	CFP	
AND		
Concept 4 Optical coherence tomography	OCT	
AND		

Table 1.3 Database search strategy table.

Results 1.3

Number of results for each database searched on 15.12.2022:

Number of results combined searches	WOS (Topic - title, abstract and keywords)	SCOPUS (Article, abstract and keywords)	PubMed (Article, title and keywords)	Embase (all fields)	Cochrane (title, abstract and keywords)	DOAJ (all fields)
Concept 1	74589	90	100962	179602	74	484
Concept 1 and 2	946	27	289	552	27	82
Concept 1,2 and 3	104	0	17	53	0	0
Concept 1,2,3 and 4	85	0	13	45	0	0
Total number of papers	(85+27+13+45+27+82)	(27) only using concepts 1 and 2	13	45	(27) only using concepts 1 and 2	(82) only using concept 1 and 2
Total papers with duplicates (108) removed	279 minus 108 duplicates = 171					

Table 1.4 Results of database searches.

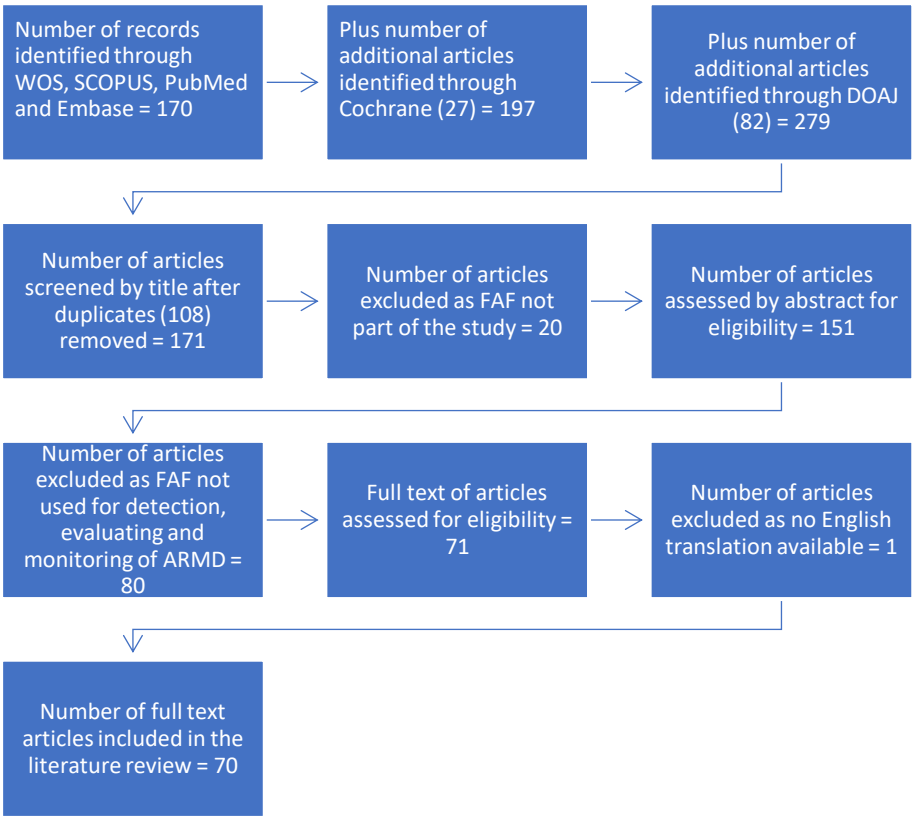

Figure 1.3 Flowchart showing how retrieved articles were screened

1.3.1 The role of lipofuscin in fundus autofluorescence imaging

Over a typical human lifespan, three billion photoreceptor outer segments are phagocytosed within the retina by the structure known as the RPE.(51) The RPE is a single layer of uniform-sized, polygonal cells that separate the choroid from the neurosensory retina.(51) Its job is to phagocytose, digest and breakdown (by means of lysosomes) the pigmented outer segments (tips) of photoreceptors on a daily basis, allowing for renewal leading to the maintenance of photoreceptor excitability.(51) A small fraction of these products of the visual cycle (LF and melanolipofuscin) is not chemically suitable for the digestion and therefore accumulates in the lysosomes of the RPE.(52) The majority of FAF is produced by dominant fluorophores within these LF granules located in the cells of the RPE,(52) and excessive quantities of LF is thought to portend the degeneration of photoreceptors, as well as the onset of new areas of retinal GA and expansion of existing atrophy.(53) Drusen are thought to be caused by vast accumulations of LF, and as ARMD progresses, drusen may wane and eventually disappear whilst atrophy of the RPE develops.(51) When utilising classic blue light excitation, FAF may be considered to represent a topographical map of LF within the RPE.(54) LF has a peak excitation wavelength of 470nm and emits yellow-green light peaking at 600-610nm.(55)

FAF associated with LF is minimal at the fovea, and rises to a maximum level at 7-8 degrees from the fovea before falling again gradually towards the periphery. This autofluorescence is not distributed evenly, and is highest approximately 12 degrees temporally and superiorly and is lower in the inferior nasal region, where it is highest at 7 to 8 degrees from the foveal centre.(56) This distribution roughly maps to the distribution of rod cells. The low foveal FAF signal may be due to cone cells not producing such a high level of autofluorescence due to a slower rate of LF production, and also that melanin and macular pigment at the foveal centre may also absorb the excitation light.

LF is produced in the membranes of photoreceptor segments from vitamin A aldehyde reactions, and then deposits in the lysosomal compartments of RPE cells with its major components related to A2E.(10) A2E has phototoxic effects, detergent qualities, inhibits the lysozyme proton pump (leading to a rise in lysosomal pH), inhibits lysosomal enzymes, and reduces the efficacy of the phagocytosis of waste material.(57),(58),(59),(60) Reactions of cell components with A2E result in glycation products that can lead to inflammation.(10) A2E

has an excitation wavelength of 430-450nm, emitting light at 560-575nm(61) and its toxic effects are thought to be caused by its ability to generate reactive oxygen species when irradiated by blue light (62, 63). However, it has been suggested that A2E is less damaging than its precursor, all-trans-retinal, and therefore the conversion to A2E may actually have a protective effect on the retina.(64)

As well as occurring in complex retinal diseases and genetic disorders of the eye, for example, Stargardt's and Best's disease,(65) excessive LF accumulation can occur with normal aging and may occupy up to 33% of the free space within RPE cells in individuals over 70 years of age.(66)

LF is therefore a naturally occurring substance in the RPE,(67) and accumulates in this layer of the retina increasingly as a result of a number of processes specifically related to aging.(67) These aging changes have been identified as:

1. A reduction in the efficiency of intracellular lysosomes whose roles include the breakdown of LF.(67)
2. Reduction in the rates of autophagy. This is where components within the cells are transferred to lysosomes for chemical breakdown, and a slowing of this process results in an accumulation of LF.(67)
3. A failure of the breakdown of aging mitochondria, causing an increase in the concentration of reactive oxygen species and an acidic pH shift within cells, culminating in cellular stress.(67)

These three processes lead to an incomplete degradation of phagocytosed outer segments of photoreceptors in the post-miotic RPE, with the subsequent accumulation of LF.(5),(51)

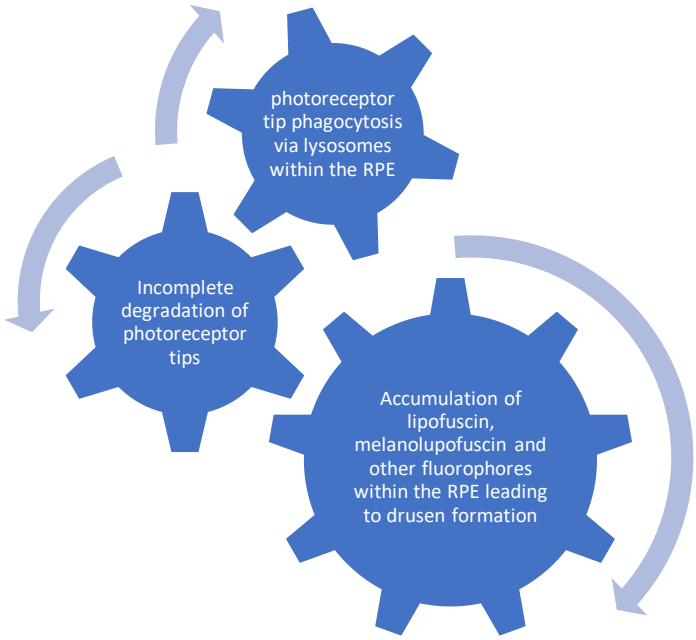

Figure 1.4 Illustration of the retinal processes leading to drusen formation

Note that drusen and pigmentary abnormalities do not map precisely to FAF abnormalities, which appears to indicate that FAF offers the clinician information over and above that offered by other ophthalmic imaging methods including CFP, SD-OCT and fluorescein angiography (FA).(68, 69) Furthermore, drusen do not produce a uniform abnormal FAF

signal, but rather this can vary from a normal, increased or decreased FAF intensity, presumably due to the variable composition of drusen which may include fluorophores other than LF, as well as variable changes to the overlying RPE.

Using alternative excitation wavelengths can detect other fluorophores present in the retina e.g. using near-infrared FAF (NIR-FAF) (excitation 790nm and peak detection >800nm) can detect melanin (peak emission 787nm).(70) Melanin protects the eye by absorbing/blocking visible light and ultraviolet wavelengths as well as having an antioxidant role within the retina.(71, 72) Note, however, that NIR-FAF, as delivered via the Heidelberg Retinal Angiograph cSLO (operating in Indocyanine Green mode), generates a signal which is approximately 60-100 times weaker than short wavelength FAF.

Additionally, rhodopsin is a human visual pigment contained within rod outer segments and can lead to a reduction of the FAF signal by absorbing the incident excitation wavelengths.(73) It is important to note that the “bleaching effect” of rhodopsin, after exposure to light, can increase FAF by up to 30% compared to a dark-adapted eye.(74)

Whilst there appears to be a consensus that hyper-autofluorescence, especially of the focal increased variety,(67) may be a precursor for the onset and progression of dry ARMD, there appears to be controversy over whether an accumulation of LF can also predict the onset of wet ARMD.(75) Some authors have reported that an increase in FAF was an uncommon finding in the fellow eyes of patients exhibiting choroidal neovascularisation (CNV),(76) however, other have stated that there may indeed be a link between retinal autofluorescence and the risk of progression to wet ARMD.(77, 78) Furthermore, a specific type of hypo-autofluorescence, of the reticular variety, has been implicated in widespread RPE inflammation, and may be a sensitive marker for the likelihood of future choroidal neovascularisation.(79)

1.3.2 Specific retinal anomalies and fundus autofluorescence

1.3.2.1 Growth in geographic atrophy

No therapy is currently available to reverse the effects of geographic atrophy (GA), and therefore all interventions are aimed at stopping or at least reducing the rate of progression.(80) Standard CFP is not an ideal modality for the differentiation of GA from its precursors, i.e. drusen and depigmentation, which have a similar appearance to GA, with the borders of atrophy being particularly difficult to discern in lightly pigmented eyes.(2)

FAF can offer improved detection and repeatability when measuring overall GA area compared to CFP, and is also superior for detecting small areas of GA.(2) The most likely reason for CFP being of limited value in measuring GA area is poor image contrast, and at present FAF is considered the gold standard for monitoring GA growth due to the striking reduction in FAF signal detected in a zone of GA.(1, 80) FAF, along with SD-OCT, are therefore considered to be the most powerful modalities for the diagnosis and monitoring of ARMD,(80, 81) as well as measuring the effectiveness of any treatment.(82, 83) Furthermore, several authors have described how FAF can be used to predict future growth of retinal GA area.(84, 85) This can be done in two ways:

1. By accurately measuring the area of retina affected, as the initial lesion size has been shown to be significantly related to the prognosis and annual rate of VA decline ($p<0.01$).(83) The key initial area of GA appears to be 2.6 mm^2 (approximately equal to one optic disc area), with lesions equal to or greater than this size having significantly higher annual reductions of LogMAR best corrected visual acuity (BCVA), as well as a faster annual rate of growth in the area of retina affected.(83)
2. FAF can be considered a topographical map of LF in the RPE, and an increase in the FAF signal surrounding GA in a halo configuration has been implicated as a risk factor for the progression,(10, 86-89) and the increased rate of progression of GA.(10, 80, 90)

The specific pattern of FAF observed in proximity to the GA is also important,(2, 83) as significantly lower rates of progression have been demonstrated between those with focal or no FAF pattern, compared to the higher rates seen with a banded or diffuse pattern.(10) An alternative to FAF for the purpose of examining GA is enface OCT imaging, which shows areas of photoreceptor outer segment disruption that extends beyond the limits of the

manifest GA, similar to the halo effect detected with FAF, however, the field of view with enface OCT is limited compared to FAF. It has also been suggested that enface OCT has the advantage of not exposing the eye to potentially damaging blue light, and as SD-OCT is currently the modality of choice for studies into clinical trials involving GA anatomy, enface OCT would appear to be a natural choice for the imaging of areas of potential GA expansion.(91) Enface OCT imaging used in conjunction with OCT B-scanning has also been shown to be reliable in identifying and localising areas of hyperpigmentation identified via CFP.(92)

Examining the results of research quantitatively, GA growth has been measured with CFP to be on average 1.45mm^2±0.06mm^2 per year. Compare this to FAF at 1.43mm^2±0.06mm^2 per year, with FAF tending to detect GA earlier. However, over longer periods the methods give comparable results.(93) Similarly, the age-related eye disease study 2 (AREDS2) study (imaging 2202 patients) noted that agreement between CFP and FAF started out at 42.9% but increased to 80.9% over 5 years,(93) however, at every visit the area of GA measured by FAF was larger than that detected by CFP.(93) These findings can potentially be explained by the observation that hypo-autofluorescence occurs not just in areas of manifest GA, but changes preceding GA, i.e. nascent GA (identified via SD-OCT), and hence FAF appears to be detecting GA and the next sliver of retina primed to become atrophic,(93) whereas CFP measures only the frank areas of retinal GA. Mean differences between graders were found to be comparable between CFP and FAF at 0.02mm^2 (CI 1.76 to 1.8), however, intergrader variability was greater for CFP. This finding is most likely explainable by virtue of FAF imaging providing better contrast than CFP at the borders of GA as mentioned earlier.(93)

Studies into OCT changes seen in ARMD and their relation to FAF alterations indicate that an area of decreased FAF next to an atrophic lesion represents an area where changes in retinal architecture have occurred, and this makes a case for these areas to be included in the overall measurement of GA present in longitudinal studies, where GA is taken as a surrogate end point for loss of visual function.(82) This finding could have important ramifications for future studies into GA growth.

1.3.2.2 Reticular pseudodrusen

A significant section of this review focuses on reticular pseudodrusen (RPD). This was considered necessary due to the large number of references identified in the literature review search that highlighted RPD as a specific retinal anomaly with close links to ARMD, as well as their being more readily visible with FAF compared to other imaging technologies.(85) The detection of RPD could be important for understanding the process of ARMD onset, progression and prognosis,(94) however, the precise mechanism of RPD formation, and also their significance, is not fully understood. Associations have been suggested between smoking and diet and the presence of RPD, and as a consequence FAF could have an important future role in counselling patients on lifestyle advice to avoid future sight loss from ARMD.

1.3.2.2.1 Reticular pseudodrusen history and description

Pseudodrusen were described for the first time by Minoun et al in 1990, as a variant of drusenoid changes whose visibility is enhanced when viewed under blue light.(95) In 1991, Klein et al first coined the term “reticular drusen”, and described these as soft, indistinct, and occurring in networks of interlacing ribbons.(95) Reticular pseudodrusen (RPD) is the two terms combined, with Arnold et al (in 1995) the first to use this term, also making a link between these features and a higher risk of choroidal neovascular membrane development.(95) RPD develop along the superior-temporal arcades and/or perifoveally, superior and nasal to the macula and around the optic nerve head.(46, 94) The CFP appearance is of yellow networks of interlacing lesions, round to oval in shape and 125-250μm in diameter. However, CFP is a poor way to visualise RPD,(46, 96) with FAF being a superior modality. FAF shows ill-defined hypoautofluorescent lesions against a background of mild hyperautofluorescence,(97, 98) discrete round to oval in shape,(99) in a target configuration.(46, 50, 94, 95) On SD-OCT B-scan, RPD have been described as distinct round or triangular, granular hyperreflective deposits situated *above* the RPE, between the RPE and the inner and outer segments of the photoreceptors, sometimes breaching this boundary in more advanced cases.(46) In some cases the tops are rounded resembling a “haystack”. A 2016 study defined RPD as the presence of five or more of these lesions arranged in a network.(50)

The Antioxydants, Lipides Essentiels, Nutrition et maladies OculaiRes (ALIENOR) study carried out between 2011 and 2012 reported an RPD prevalence of 13.5% (15.6% in women and 10.2% in men), increasing to almost 50% over 85 years of age. RPD were found in

4.6% of eyes with no ARMD, 34.6% with atrophic ARMD and 8.1% in neovascular ARMD.(46) RPD were considered to be present if detected by two of the following: CFP, FAF, near infrared reflectance (NIR) and SD-OCT. The prevalence measured in older studies that utilised CFP for the detection of RPD was considerably lower, e.g. the Beaver Dam study (1991) and the Blue Mountains study (1995) recorded the incidence of RPD over 75 years of age at 6.6%, and 4.9% respectively.(100) Therefore there is evidence of an historic underestimation of the prevalence of RPD. Note that although RPD have not been found to contain markers for photoreceptors, RPE and Muller cells, photoreceptors are negatively affected around these deposits, having a reduced visual function.(94)

1.3.2.2.2 Theory of reticular pseudodrusen formation

A vascular theory of RPD formation was first proposed by Arnold et al in 1995, who demonstrated that a significant loss of the middle choroidal layer of small blood vessels coupled with an increased separation between choroidal veins was associated with RPD, concluding that choroidal stromal fibrosis and loss of choroidal vasculature resulted in RPD formation.(98, 101) Subsequently, Querques et al proposed that disruption in the RPE structure due to atrophy and fibrosis of the choroid beneath may cause photoreceptor outer segments to accumulate above the RPE and lead to RPD formation.(98, 102) Further evidence of a link between RPD and choroidal changes comes from a study finding that RPD were not detected by means of SD-OCT and FAF in central areolar choroidal dystrophy (CACD), however, they were found to be present in 52.6% of early ARMD, and 100% of the eyes examined with advanced ARMD.(103) The authors postulated that this is due to CACD being a genetic disorder causing direct photoreceptor damage with subsequent loss of the RPE, rather than the complex multifactorial disease processes occurring in ARMD, with sub-RPE alterations leading to photoreceptor loss, RPE disruption and secondary angiogenic processes affecting the choroidal and retinal vasculature.(103)

1.3.2.2.3 Reticular pseudodrusen and their association with rods cells

Rods cells are most abundant in a horizontal ellipsoid configuration, with maximum density at the limits of the macula, in a similar distribution to the areas where RPD are normally detected. Furthermore, RPD have also been described as being functionally and topographically associated with rods, and may be responsible for a profound reduction in dark adaptation and retinal sensitivity.(96) It therefore appears that the rod system may be especially vulnerable to RPD formation, and this could be explained by virtue of the relatively

low rod/RPE cell ratio in rod dominated retinal regions. (104) This may help to explain why in GA, patches of atrophy initially appear in the parafoveal region, proceed to enlarge, then coalesce, with the fovea (featuring a low density of rod cells) usually remaining spared until late in the disease process.(94) Note that delayed rod dark adaption has also been identified as the best surrogate endpoint for early ARMD,(81) with the rod recovery slope being the best predictor of the CFP ARMD grade and FAF classification.(105)

1.3.2.2.4 Reticular pseudodrusen and their link to advanced ARMD

RPD have been linked to macular atrophy (MA), especially in the inferior macular region with a 6.1 times increased likelihood for the advanced form of MA in the presence of RPD.(49) Other research found the prevalence of RPD to be 52% and 23% in patients with and without ARMD respectively, concluding that patients are 3.4 times more likely to have ARMD if RPD are present.(94) Furthermore, both the Beaver Dam and Blue Mountains studies reported that over 5 years, patients with RPD were 4-6 times more likely to progress to late ARMD than those without RPD but having other early signs of ARMD.(79, 106-108) RPD have been linked to ARMD progression resulting from geographic atrophy,(96) however, their link to wet ARMD is controversial, where many studies have shown a positive association and others have not.(101, 107, 109).

1.3.2.2.5 Reticular pseudodrusen and their association with wet ARMD

Regarding conversion to wet ARMD, there was no link found between RPD and CNV formation from the ALIENOR study, with only 8.1% of eyes with RPD converting from dry to wet ARMD.(46) Conversely, other studies *have* shown a link between RPD and CNV formation. One study following the fellow eye in patients with CNV over 5 years found that RPD were positively associated with wet ARMD.(97) This was supported by the finding that RPD are an independent risk factor for the 5 year progression rate in eyes whose fellow has CNV,(94) and RPD have also been found to be associated with worse visual function from the early stages of ARMD, with a higher likelihood for the formation of both dry and wet subtypes.(80)

1.3.2.2.6 Detection of reticular pseudodrusen

RPD increase with age reaching a prevalence of approximately 50% in the over 85s.(10) FAF, NIR and SD-OCT are reported as being superior for RPD detection compared to CFP,

FA and indocyanine green angiography (ICGA), with NIR being the most sensitive of all modalities. (46) One author has found that a patchy FAF pattern is a common finding in RPD,(110) with others describing the appearance of RPD as a “target” (46, 50, 94, 95, 98, 100) and with a distinct pattern of isoautofluorescence surrounded by an area of reduced FAF signal,(46, 100) which may be of “any size”.(94) Sensitivities for RPD for the various modalities have been found to be: NIR (93%), FAF (92%), OCT (74%), red-free (RF) (33%), and CFP (29%),(97) with another research group revealing similar findings, albeit with a better sensitivity for CFP at 42%, and a fractionally lower sensitivity for FAF of 89%.(50) The latter study also reported that those participants whose RPD were detected with both CFP and FAF were significantly younger than those whose RPD were detected with only one imaging technique. With the heavy reliance on CFP in previous studies, the authors conclude that RPD in the general population is significantly underestimated.(50)
RPD are therefore poorly visualised via CFP, but seen readily with FAF and other imaging modalities, (95) and therefore it has been recommended that at least two modalities are used for the diagnosis of RPD. Their presence wanes in very advanced wet ARMD, as RPD appear to fade in the vicinity of neovascular membranes.(95) In conclusion, there are no clear recommendations on the most effective way to view or image RPD, but a multimodal approach appears to be the overall consensus.(94, 100)

1.3.2.2.7 Reticular pseudodrusen and genetics

Genetic susceptibility to RPD has been linked to individuals carrying minor allelic variations in the ARMD, complement factor H gene (CFH), and hepatic lipase (LIPC) genes. Interestingly, lipophilic statin medication has been reported as associated with a lower incidence of RPD.(50) Also, a greater frequency of the age-related maculopathy susceptibility 2 (ARMS2) allele was found in subjects with bilateral large drusen, and RPD in either eye, and this association with large drusen may implicate RPD as a risk factor for the progression to advanced ARMD.(50) The authors recommend that OCT, NIR and FAF imaging are important for the phenotyping of patients with intermediate ARMD to enable better counselling, particularly through the detection of RPD.(50) Other possible risk factors suggested for the formation of RPD are female gender, increased age, high body mass index (BMI), reduced choroidal thickness, low education, cardiovascular factors and smoking.(100)

1.3.2.2.8 Reticular pseudodrusen and geographic atrophy growth

RPD have been found to be highly correlated with GA, and may be considered an early manifestation of GA itself.(98) Also, RPD may portend the expansion of GA into unaffected areas of the retina, and may be associated with a higher GA growth rate.(47) Specifically, unilobular GA expands at a significantly lower rate compared to the multilobular GA, and RPD have been shown to be more prevalent in the latter type of GA than the former.(98) Furthermore, quantitative analysis showed the mean GA progression rate for all eyes to be 0.8mm²±0.6mm² per year), with unilobular GA expanding at 0.3mm² per year and multilobular at 0.9mm² per year (statistically significant difference to the 5% level, $p = 0.02$), with RPD associated with 28.6% of unilobular lesions compared to 97.0% of multi-lobular.(98) Overall, 74.2% of patients with and 41.7% without RPD showed subsequent GA expansion.(98) RPD were also reported to be specifically and independently associated with late ARMD, especially when these advanced changes resulted from GA and/or retinal angiomatous proliferation (RAP) lesions.(96, 98)

Below are some examples, taken during this study, of reticular pseudodrusen viewed under different imaging modalities

Figure 1.5 Reticular pseudodrusen viewed via colour fundus photography, appearing as series of yellow dots surrounding the fovea, indicated by the yellow arrows

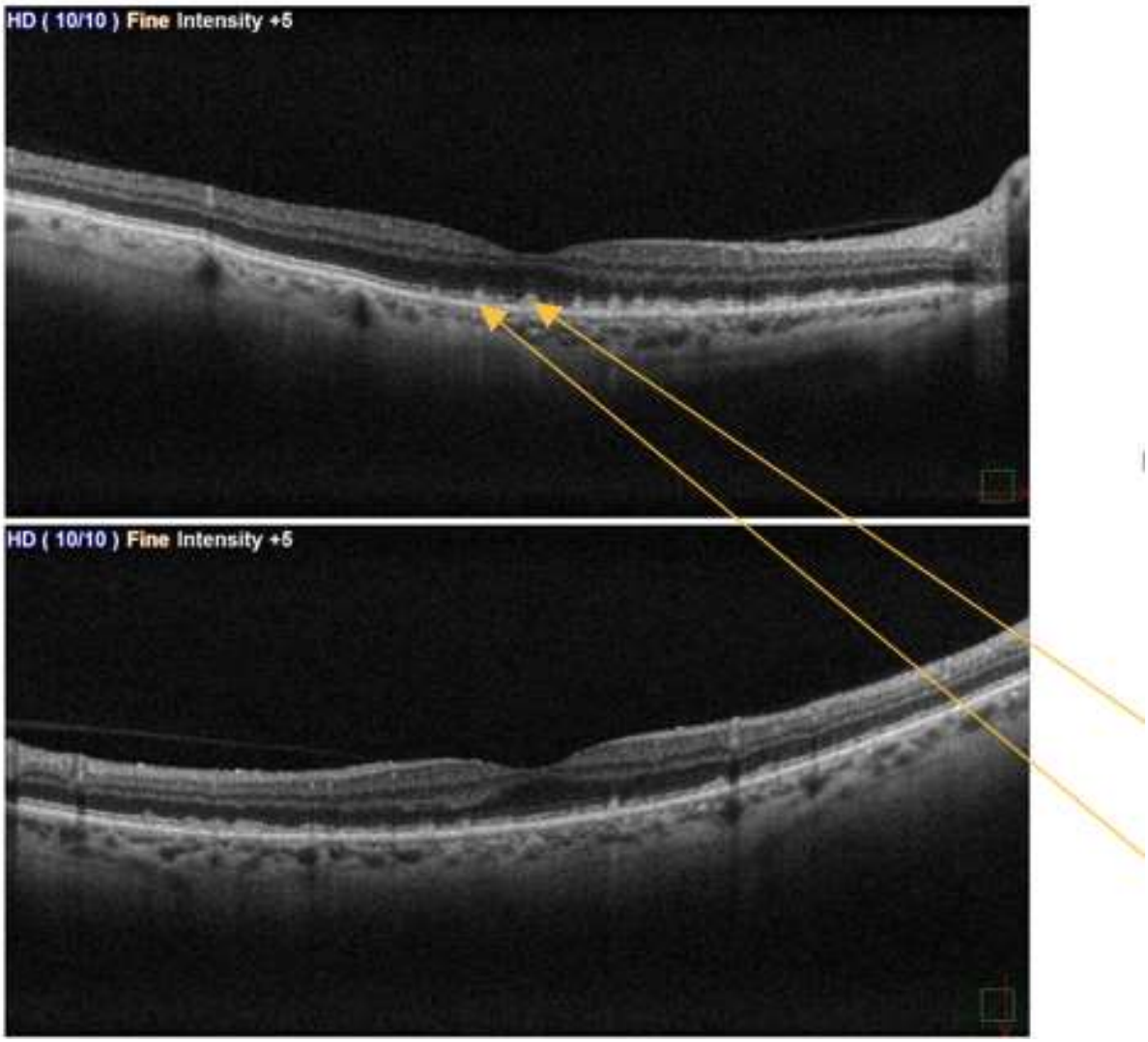

Figure 1.6 Reticular pseudodrusen from the same eye as in Figure 1.5, viewed via OCT, appearing as “haystack” shaped lesions situated above the retinal pigment epithelium, indicated by the yellow arrows

Figure 1.7 Reticular pseudodrusen from the same eye as in Figures 1.5 and 1.6,viewed via FAF, appearing as hyporeflective lesions surrounded by an area of relative hyperreflectivity, indicated by the yellow arrows

1.3.2.3 Polypoidal choroidal vasculopathy

Correctly identifying polypoidal choroidal vasculopathy (PCV) is of crucial importance to ophthalmologists, as 25% of patients with wet ARMD can be non-responsive to treatment with anti-vascular endothelial growth factor (anti-VEGF) intraocular injections (demonstrating either recalcitrance and/or tachyphylaxis).(111) Between 50% and 90% of these non-responsive cases have PCV, most often misdiagnosed as typical wet ARMD. Compared to other forms of neovascular ARMD, PCV has a more favourable long-term prognosis, with a lower incidence of eventual progression to GA.(49)

PCV is a neovascular ARMD subtype characterised by a type 1 aneurysmal choroidal neovascular lesion (when macular neovascularisation vessels from the choriocapillaris grow into the sub-RPE space and are characterised by the presence of polyp-like dilations)(112), producing an orange subretinal nodule on CFP, and specific features on SD-OCT. These OCT findings include tall but narrow (thumb-like) pigment epithelial detachments (PEDs), notched PEDs (or the “sleeping snowman”), the “bubble sign”, (a subretinal circular anomaly featuring a hyper-reflective border and a hypo-reflective core) and a depression of Bruch’s membrane under a serosanguinous PED; if 2 out of 5 of these CFP and SD-OCT findings are observed, a sensitivity of 0.88 and specificity of 0.92 can be achieved along with a predictive accuracy from the area under the receiver operating characteristic curve (AUC) of 0.90.(112) However, ICGA remains the gold standard test for PCV, but this test has the disadvantage of being an invasive technique(112) (and contraindicated in patients with allergies to iodine-based dyes(113)). In PCV, FAF imaging has been described (specifically in the polypoid region) as detecting a hyperautofluorescent ring with a central granular hypoautofluorescence (with the abnormal vascular network hypofluorescing) but delivers a lower level of sensitivity of 0.67 and specificity of 0.50 compared to combined CFP and SD-OCT.(112) Therefore FAF does not provide an improvement in the accuracy of diagnosis of PCV, but can still provide some additional imaging features that may be clinically useful.(112) A mottled FAF pattern has also been described, i.e. a diffuse area of irregular autofluorescence, found to be present in 31.8% of eyes with PCV.(1) This may be due, however, to mottled FAF patterns being linked to RPE depigmentation that are found in 76.5% of eyes with neovascular ARMD.(114) RPE abnormalities with drusen have also been noted as being present prior to the development of PCV.(1)

Another associated retinal finding with PCV is cuticular drusen. These were first described by Gass et al in 1977. They are dynamic, with periods of absorption and coalescence, and may be associated with neovascular ARMD.(115) They are sub-RPE deposits visible as small yellow spots distributed in various patterns throughout the fundus on CFP and appear with FAF imaging as small hypoautofluorescent dots.(115) They are considered to be the result of central RPE erosion with these triangular protrusions stemming from the RPE base layer, and may be responsible for the choroidal hyper-transmission signal seen via OCT.(116) Cuticular drusen are also readily detectable with FA,(115) resulting in the classic “starry sky” pattern.(116) In summary, FAF is an inferior technique when compared to a combination of CFP, OCT and FA for the detection of PCV, however, FAF may help to add extra information and therefore assist in diagnosis via the detection of specific FAF patterns and cuticular drusen.

1.3.2.4 Retinal angiomatous proliferation

Retinal angiomatous proliferation is type 3 neovascular form of wet ARMD whereby an angiomatous proliferation originates, somewhat controversially, from within the retina, invades the outer retina reaching the subretinal space, and in some cases results in an eventual communication with the choroidal circulation. This is the opposite of the classic (type 2) course of wet ARMD whereby choroidal neovascularisation (CNV) breaks through the RPE to invade the neurosensory retina, eventually communicating with the retinal blood vessels to form an anastomosis between the retinal and choroidal circulation. RAP and classic CNV can therefore be easily confused.(117) FAF patterns may have a role in identifying RAP lesions,(1) as demonstrated by a study examining wet ARMD, with RAP lesions the most likely to demonstrate abnormal FAF patterns at 85.7%, next was intermediate typical wet ARMD at 54.1%, and lastly PCV at 36.4%.(1) A granular FAF pattern was identified in 42.9% of eyes with RAP lesions, with this pattern more prevalent in RAP lesions than in the other forms of wet ARMD.(1) As discussed earlier, a link has also been found between the presence of RPD and RAP lesions, (96, 110) with RPD being independently and specifically associated with advanced ARMD featuring RAP lesions(94) and GA.(110) A further FAF pattern identified as having a strong correlation with RAP lesions is the RPD ribbon dominant type, found to be significantly more common in RAP lesions (69%), and with GA (78.6%), when compared to typical ARMD.(99) More evidence for this link with ribbon patterns of RPD and RAP come from another study which reviewed 321 eyes with a new diagnosis of neovascular ARMD. In the group with both wet ARMD *and* RPD, the RPD ribbon FAF pattern was more prevalent in RAP at 69.2%, with this pattern less common in wet ARMD without RPD and PCV at 40% and 16.7% respectively.(106) In terms of visual function, an ability to detect, diagnose and treat RAP early could be particularly advantageous, delivering better visual outcomes, as shown in the Comparison of Age-Related Macular Degeneration Treatment Trials (CATT) study, which also highlighted that RAP is a baseline risk factor for future progression to GA.(118)

1.3.3 Associations between SD-OCT findings and FAF results

Whilst there is no clear link between retinal changes observed on SD-OCT imaging and the results from FAF, close relationships have been suggested, with changes in FAF intensity tending to be associated with advanced changes in the outer layers of the neurosensory retina in ARMD.(82) These outer retinal layer changes include outer nuclear layer (ONL) thinning, disruption of the external limiting membrane (ELM), ellipsoid zone (EZ) disruption, Bruch's membrane/Retinal pigment epithelium (BM/RPE) complex disruption, hyperreflective loci (HL) and an increased choroidal hyper-transmission signal.(82) Of these changes, ONL thinning, ELM disruption, EZ changes and an increased choroidal hyper-transmission signal tend to be associated with a *decreased* FAF signal.(119) The most common SD-OCT findings associated with an *increased* FAF signal are HL, (82) whilst visual acuity (VA) has been found to be most closely related to the integrity of the ELM and the EZ (120), and therefore a reduction in VA is most likely to be associated with a hypoautofluorescence on FAF imaging. The table below summarises how SD-OCT findings relate to increased and decreased FAF signals and VA.

SD-OCT changes detected	**Changes associated with an increased FAF signal**	**Changes associated with a decreased FAF signal**	**Changes closely associated with a reduction in VA**
ONL thinning		Y	
ELM disruption		Y	Y
EZ changes		Y	Y
BM/RPE complex disruption		Y	
HL	Y		
Increased choroidal hyper-transmission signal		Y	

Table 1.5 Summarising how SD-OCT findings are associated with increased and decreased FAF signals and VA.

A limitation of FAF is that it tends to have as relatively high intergrader discrepancy compared to the observation of other retinal features, one study finding this to be 12.3% (for comparison the same study reported the intergrader discrepancy for the interface of the photoreceptor layer was 11.0%, external limiting membrane 9.6%, RPE 9.6%, focal hyperreflectivity 6.8% outer nuclear layer 11.0% and choroidal hyperreflectivity 1.4%). SD-OCT grading can also have its challenges, with changes to the ELM, EZ and RPE particularly difficult to grade, with ONL changes, choroidal hyper-transmission and HL easier

to assess.(82) It has been reported that drusen associated with a normal retinal architecture do not have a decreased FAF signal, and overall, a good general rule is that when a normal FAF signal is detected, only relatively minor retinal alterations are likely to be found on SD-OCT.(82)

1.3.4 Three instruments available for fundus autofluorescence imaging

There are three main types of commercially available instrumentation capable of FAF imaging. One based on a modified fundus camera, one based on a confocal scanning laser ophthalmoscope (cSLO), and one based on ultra-widefield imaging technology (also a type cSLO based system).

Both the fundus camera and cSLO based systems utilise a confocal aperture for the filtering of backscatter from outside the plane of focus.(119) In cSLO FAF, superior contrast and spatial resolution are obtained compared to the fundus camera, by means of a raster scanning laser,(121) which suppresses light from planes anterior and posterior to the plane of interest.(4)

However, cSLO FAF utilises blue light excitation (λ = 488nm), (with the exception of Ultra-widefield cSLO FAF – more details below) and with macular pigment containing xanthophylls tending to block this wavelength, the result can be a decreased FAF signal at the macula.(10, 93, 119, 122) This artefact with cSLO FAF can mimic involvement of the macula in patients with GA, and can therefore lead to an overestimation of the total area of GA present.(4) Fundus camera-based FAF on the other hand, by employing a red-shifting filter, uses an excitation illumination of green-orange light (λ = 510-610nm) which is less affected by the blocking effect of xanthophylls, giving a brighter macular signal.(119)

cSLO FAF is commonly provided in clinical trials by the Heidelberg retina angiogram in fundus autofluorescence mode, taking eight images which are then averaged to reduce noise.(122) This averaging makes cSLO FAF superior to fundus camera-based FAF for the detecting of smaller signals which may be present, for example, surrounding areas of GA in a halo configuration.(122) A longer period of steady fixation by the subject is, however, required when compared to fundus camera-based FAF (which captures the FAF image in a single flash of light) causing a higher failure rate for the capture of acceptable images.(122) This problem with cSLO FAF was borne out by a study investigating cSLO and fundus camera FAF reliability which found that 80 out of the total of 292 eyes did not produce an acceptable image. 76 of these failures were with the cSLO FAF averaging system, whilst the remaining four failed with both systems.

The cSLO FAF's averaging system does, however, have the advantage of being able to function at a significantly lower illumination power, with only a small central part of the

patient's pupil used for input light, with the rest of the pupil used for collection. Fundus camera-based FAF is inferior in this respect as it uses annular illumination with only the centre of the pupil involved in collection. This may make cSLO FAF more comfortable for patients with fewer immediate after-effects from the less intense flash, however the longer fixation period required with cSLO could also have a negative effect on patient comfort.(121) A further reason for a higher failure rate with cSLO FAF may be that the relatively short wavelength is more prone to the light scattering effects of cataract, and in particular nuclear sclerotic lens opacities, compared to fundus camera-based FAF whose longer wavelength mitigates this particular artefact.(122)

Another difference in the functionality between the fundus camera and cSLO modalities is that latter FAF modality detects only direct light, whereas fundus camera-based FAF records indirect, scattered light, which can emanate from all tissue levels.(10) This variation in light scatter may explain why several pathologies have been shown to differ in characteristics between the two FAF modalities, including fibrovascular membranes and RPD, with cSLO FAF demonstrating a brighter signal than fundus camera FAF.(119) Both instruments, however, perform similarly for GA, most likely due to the absence of back scatter from this type of lesion.(122) Scattered light may also have the disadvantage of masking subtle hyperautofluorescence, as well as producing the phenomena called "pseudo-autofluorescence" where a false FAF signal is detected from structures outside of the retinal plane.(123) However, one study has reported that the difference in scattered light between cSLO and fundus camera FAF is statistically insignificant.(122) A potential advantage of this scattered light detected via fundus camera-based FAF is that deeper retinal structures, including choroidal blood vessels, fluoresce only with modified fundus camera-based FAF, due to the deeper penetration of the RPE with the longer wavelength used.(122) This could offer extra information on the choroidal circulation that may not possible to obtain via cSLO FAF. Finally, with cSLO FAF, integration into conventional fundus cameras is not possible as with fundus camera-based FAF,(119) making cSLO FAF relatively more expensive to integrate into clinical practice.(122)

1.3.5 Colour fundus autofluorescence and macular pigment density

In most cases, images from both cSLO and fundus camera-based FAF are monochromatic. Some cSLO systems e.g. the Heidelberg Spectralis have the option of multicolour reflectance-based imaging, via blue, green and infrared wavelengths. Recently a so-called quantitative colour FAF imaging device (from EIDON, CenterVue, Padua, Italy), based on light-emitting diode technology has become available which has an excitation wavelength of 450nm, with two different emission spectra of 510-560nm (green) and 560-700nm (red),

which have the potential to identify minor fluorophores that might otherwise be masked by the dominant LF (124). Quantitative colour FAF has even shown potential to differentiate between active and inactive macular neovascularisation, but this technology is, at the present time, only in its infancy and will require more research to corroborate these results.(125)

There has also been much interest in recent years in macular pigment density (MPOD), with changes in MPOD playing a potential role in retinal disease processes (126, 127). cSLO FAF systems have been used to quantitively measure MPOD (composed of lutein, zeaxanthin and meso-zeaxanthin) which act as antioxidants and filter blue light to protect the retina.(128, 129) For example, the Heidelberg Spectralis HRA2 cSLO uses two excitation wavelengths with barrier filters to create autofluorescence, calculating the MPOD from the difference between the recorded signals.

cSLO FAF	**Fundus camera-based FAF**
Superior resolution and contrast due to the averaging system employed	Quicker and therefore more comfortable for the patient as taken via a single flash, however the flash is more intense than that delivered via cSLO.
Macular xanthophylls reduce the FAF signal	Not as prone to xanthophyll artefacts as cSLO FAF
Can be taken through a small pupil, and can be ultra-wide	Uses annular illumination so larger pupil size required
Greater motion artefacts due to averaging system	Fewer motion artefacts
Relatively expensive equipment costs	Relatively cheap equipment costs
Uses relatively short wavelengths for excitation (488nm). (Optos uses 532nm and 635nm).	Uses relatively long wavelengths for excitation (510-610nm).

Table 1.6 The main differences between cSLO and Fundus camera-based FAF systems.

Peripheral ultra-widefield fundus autofluorescence can enable examination of the metabolic activity of the retina/RPE in areas which lie beyond the reach of conventional SD-OCT and fundus cameras.(67) The Optomap Ultra-Widefield ® system by Optos ® uses a combination of cSLO technology with an ellipsoid mirror to capture images up to a field of view covering 200º (delivering approximately an 82% view of the retinal area).(130) The Optos ® system uses lasers with excitation wavelengths of 532nm (green light for visualising the RPE and for performing FAF) and 635nm (red light for visualising the choroid) with an emission filter of > 540nm,(74). By using this longer wavelength of light Optos ® benefits from a reduced absorption by macular pigment in a similar way to fundus camera based FAF. Optos® automontage is an additional feature that allows the user to take images from five different perspectives, on central, up, down, left and right gaze, offering a 220º field of view equating to approximately 97% of the retinal area. The drawbacks/limitations of Optos®

include: the colour images not being “true” colour (but rather created from lasers and therefore being “pseudocolour”) and peripheral magnification/distortion and lid artefacts.(130)

Examples of cSLO devices capable of FAF	Examples of Fundus Camera-Based systems capable of FAF
Heidelberg Spectralis	Nidek Retinascan Duo RS-330
Nidek Mirante	Topcon Maestro
Optovue OCT	Kowa series
Optos: Daytona/California/Monaco/Silverstone	Canon CR series

Table 1.7 Devices capable of delivering FAF by means of cSLO and fundus camera-based systems.

Peripheral FAF has recently been investigated specifically with regards to macular degeneration. It has been suggested that hypoxia and ischaemia in the peripheral retina have an important role to play in the development of wet ARMD.(131, 132) Furthermore, peripheral autofluorescence changes have been found in 39.6% of eyes with ARMD, and 28.9% of healthy eyes, with hypoautofluorescence being the most common abnormality observed.(133) One study suggests that older age, female gender, and neovascular ARMD have been found to be associated with peripheral FAF abnormalities,(114) whilst another found that patients with ARMD, older patients, and those with poor VAs were more likely to demonstrate peripheral anomalies with FAF.(134) Research in Japan has also noted that more abnormal peripheral FAF patterns were associated with typical AMD, PCV and RAP subgroups compared to controls, and the authors concluded that future treatment strategies for ARMD may be based on peripheral FAF findings.(1) Therefore, ARMD can be considered as not purely a condition affecting the macula, but rather one that may affect the entire retina, including the far periphery.(133)

1.3.6 Quantitative FAF

Quantitative FAF (qAF) has been described as the process of measuring the intensity of FAF after excitation with short wavelength light. The method discussed here involves the Heidelberg Spectralis HRA2, utilising a cSLO with 488nm excitation and 500nm to 680nm emission spectra. qAF is calculated using published algorithms(135) that use data such as the laser power, the instrument’s “zero” signal, the detector sensitivity, subject refractive error and ocular media status.(135) Greyscale values are then assigned to pixels by reference to a standard AF signal loaded within the instrument.(136) Detector sensitivity is adjusted to ensure that FAF capture is within the dynamic range of the detector (to avoid, for

example, overexposure) and to guarantee a linear relationship between the FAF signal emitted and the recorded signal.(137) No normalisation is carried out on the data to increase image contrast as is usually the case for FAF. This is because normalisation would mean that comparisons of grey levels between images of the same or different eyes could no longer be made.(137) Magnification of the fundus image due to axial length and/or corneal curvature variations must also be accounted for by application of a scaling factor,(135) however, this scaling allowance will not hold well for eyes that have undergone refractive laser surgery.

Work has been done to calculate the correction factor for subjects without cataract to account for the natural changes in ocular media transparency that occur with age,(138) however, research is yet to be done to allow for similar, but more extreme attenuations of the FAF signal caused by cataract. For pseudophakic eyes, the specific transmission factors for particular lens implant materials/designs is accounted for within the instrument. As with traditional FAF, macular pigment is also an issue in qAF, as variations exist between individual subjects, however, at eccentricities of more than 7°, absorption by these carotenoid macular pigments (lutein and zeaxanthin) is negligible.(139) Therefore, for qAF measured beyond 7-9° there is little effect on the signal collected.(137) Other steps should be taken as with any ophthalmic imaging to achieve the maximum image quality, e.g. optimum patient positioning, good pharmacologically induced pupillary dilation and exclusion of artefacts e.g. those induced by lashes or a poor tear film quality.(137) Software for the HRA2 has been designed (HEYEX) to capture qAF in three concentric rings, each split into eight segments of the fundus image, with the figure qFAF8 referring to the mean qAF obtained, with the possibility to also select an area of interest.(135) qAF has been shown to be greater in older individuals, those of white ethnicity, smokers and females, and is maximal supero-temporally.(140) A study investigating early and intermediate ARMD and qAF found that the measurements of FAF were not statistically different (to the 5% level) in eyes with cuticular and/or soft drusen compared to healthy age-adjusted eyes, however, the levels were significantly lower in eyes with reticular pseudodrusen.(141)

So far qAF has not, however, been adopted widely by clinicians in practice, most probably due to the need for internal reference features within instruments and additional software.(137)

1.3.7 Practical uses of fundus autofluorescence

1.3.7.1 Ability to detect disease

In 2012, a community-based study in the USA, (reported as the first prospective study of its kind), successfully imaged all subjects with a fundus camera-based FAF system, and found that there was a 29% increase in disease detection with FAF.(48) It is important to note that all of the retinal images in the study were graded by an onsite medical director. In 89% of the patients, FAF improved identification and characterisation of pathology compared to CFP alone. It was suggested that this improvement may be due to FAF's superiority to CFP for detecting subtle disturbances of the RPE. The study concluded that a significant improvement in diagnostic accuracy may therefore be possible by using imaging modalities other than CFP, and future referral algorithms may benefit from the addition of FAF data.(48) This study, however, only included one patient with ARMD, and therefore its relevance to the current study may be minimal regarding the impact of FAF imaging for this specific condition. The results from the American study are also in conflict a study conducted in Australia in 2018, which reported that the total detection level for any macular pathology by a cohort of community optometrists using CFP alone was 94%. The study found that CFP provided 61% accuracy for ARMD in all cases examined, which included other pathologies. For each additional imaging modality i.e. NIR, FAF and SD-OCT, presented in that order, a small additional increase in diagnostic accuracy of 1% was observed, with a concurrent increase in false positives.(142) However, it should be noted that this difference in diagnostic accuracy between these two studies could be explained by the differing experience of the graders.

1.3.7.2 The Drawbacks of FAF

The 2012 USA study suggests several practical drawbacks of FAF imaging. Additional clinic time was noted as particularly relevant as although none of the subjects had pupillary dilation carried out, 2-3 minutes were left between the image capture for the two eyes which could prove time hungry in a busy clinical practice. Other disadvantages are the cost of the extra equipment, training in image analysis, lack of specific protocols, and the influence of media opacities on image quality.(48) The need for extra training was also a drawback highlighted by the 2018 Australian study, which reported a lack of improvement in diagnostic accuracy as well as an increase in false positives when additional information from modalities uncommonly used by community optometrists were made available.(142) Technical drawbacks of the various imaging modalities were considered in a French study from 2015 on the evaluation of GA, which succinctly states the key differences between CFP, NIR, FAF and SD-OCT as being down to two artefacts; macular pigment and low contrast. CFP and NIR suffer from low contrast; conversely FAF has excellent contrast, and

it is this property that is most likely responsible for the superior intergrader measurement of GA as reported in previous studies.(143-146) However, FAF can present problems with macular pigment leading to this imaging modality not correctly identifying foveal sparing. SD-OCT was identified as providing better tracking of atrophic retinal lesions compared to FAF and CFP, and also had the best inter and intra-grader agreement for foveal sparing.(4) This illustrates why choosing the best modality for analysing a particular retinal condition may be complex, and a multimodal approach is justified in clinical practice. It is important to note, however, that in the French study all images were assessed by a professor in medical retina, and two ophthalmology medical retina specialists.

1.3.7.3 Training required

In the 2018 Australian study, SD-OCT was found to be the most preferred imaging method as identified by 75% of the grading optometrists. However, FAF was the most preferred modality when retinal pathology other than ARMD was suspected. If an eye was incorrectly diagnosed by means of CFP alone, multi-modal imaging was of most benefit in correcting these errors and facilitating the correct diagnosis. The study therefore shows that additional imaging modalities can help clinicians to make more informed decisions regarding retinal disease, however, the effect is small for the cohort of practitioners, who were typically non-therapeutically trained optometrists with a BSc qualification working in community practice.(142) The evidence from the French and American papers indicate that improved clinician training could enhance outcomes for patients when utilising advanced imaging modalities in practice, as FAF imaging had a greater impact on ARMD diagnosis when utilised by medically trained clinicians in these studies.(4, 48) Community optometrists have been found to be strongest in diagnosing early and intermediate ARMD, but poor when tested with cases of advanced ARMD, often mistaking this condition for other retinal diseases.(142) This could reflect their experience in practice which would involve many more cases of early forms of ARMD. Again, this adds weight to the argument that clinicians' performance could be improved by extra training in the use of a variety of retinal imaging techniques, especially for more advanced disease.

1.3.8 Supplementation

ARMD is a common and important cause of loss of visual function in elderly populations throughout the World, and no therapy can regenerate the photoreceptors and RPE.(80) This is the reason why all current medical interventions are aimed at slowing the progression of

atrophy,(80) which makes early diagnosis and a preventative strategy paramount in the treatment of ARMD.

A recent study investigating retinal sensitivity in relation to the use of nutritional supplementation found that a reduced or abnormal FAF signal portends a reduction in retinal sensitivity, within 1 degree of the lesion, that occurs after the 3rd year.(147) A speckled FAF pattern, by year 5, had the lowest retinal sensitivity, with linear, patchy and focal plaque-like patterns all reducing sensitivity to a lesser degree. It was observed that an initial lace-like pattern on FAF resulted in an *improvement* in retinal sensitivity over time. This lace-like pattern of FAF may be related to RPE hyperplasia, or another transient condition involving retinal stress.(147) Retinal sensitivity also generally increases as a function of the distance from the abnormal FAF signal, with this association found to be statistically significant.(147) A previous study indicated that supplements may be an important factor in retinal sensitivity recovery,(110) and therefore an abnormal FAF pattern could be interpreted as an indication for the commencement of nutritional supplementation in certain patients. FAF imaging could also be used as a means of monitoring whether dietary supplements are improving the health and biological functioning of the retina. The level of supplement intake required to make a statistically significant improvement in retinal sensitivity, even in healthy eyes, has been reported as a daily 6mg dose of lutein.(148), however further studies, with improved designs (including randomisation) may be required to investigate the healing and/or prophylactic properties of dietary supplementation.(148) FAF and SD-OCT are considered the most useful modalities for the diagnosis and subsequent follow-up of patients with dry ARMD, including monitoring the rate of GA progression.(80) This could be important with respect to the nutritional advice given to patients, as the first Age-Related Eye Disease (AREDS) study concluded that patients at high risk of progression were the most likely to benefit from supplementation.(80) The AREDS study created four categories of ARMD: 1 to 4, with 4 being the most advanced form.

Category 1	None or few small drusen (<63µm).
Category 2 (early ARMD)	Any of the following: multiple small drusen, few intermediate drusen (63-124µm), RPE pigmentary abnormalities.
Category 3 (intermediate ARMD)	Any of the following: extensive intermediate drusen, at least 1 large drusen (≥125µm), GA not involving the foveal centre.
Category 4 (advanced or "late" ARMD)	GA involving the foveal centre and/or any features of neovascular ARMD.

Table 1.8 AREDS ARMD classification categories 1 to 4.

The first AREDS study found that the largest benefit from dietary supplements was found in categories 3 and 4, who received a combination of 500mg of Vitamin C, 400 international units of Vitamin E, 15mg of beta-carotene (for non-smokers only), 80mg of zinc oxide and 2 mg of cupric oxide (copper; to avoid anaemia with the high zinc intake) leading to a 25% reduction in progression.(149) The later AREDS 2 study tested the addition of either omega 3 fatty acids or 10mg lutein plus 2mg of zeaxanthin for their ability to further reduce ARMD progression rates, however, after 5 years of follow-up, no additional benefit was found.(90) However, the AREDS 2 study recommended a supplement combination to include lutein and zeaxanthin as this represented a safe and effective alternative to beta-carotene (reported increased risk of lung cancer from beta-carotene in this cohort).(150) It is important to note that while supplements significantly reduced the risks of the progression of ARMD, this was mainly due to the prevention of conversion to neovascular ARMD, with no impact on the incidence of GA.(10)

AREDS1	**AREDS2**
500mg of Vitamin C	500mg of Vitamin C
400 International Units of Vitamin E (or 268mg)	400 International Units of Vitamin E (or 268mg)
80g of zinc oxide	80g of zinc oxide
2mg of cupric oxide	2mg of cupric oxide
15mg of beta-carotene	10mg lutein
	2mg of zeaxanthin

Table 1.9 Supplements recommended by AREDS1 and AREDS2 studies.

1.3.9 Lipids

The outer blood retinal barrier is created by the RPE (formed by the tight junctions between the cells of the RPE), with the inner blood retina barrier created by the retinal blood vessels

themselves (formed by the tight junctions between the retinal capillary endothelial cells). Lipids appear to provide an insight into to how the retina functions with regards to the blood retinal barrier. Drusen have been described as extracellular deposits of lipids and membranous debris between the inner collagenous layer of Bruch's membrane and the basal lamina of the RPE,(110, 151) i.e. they are not shielded from the choroidal vascular circulation by the outer blood-retinal barrier. Conversely, RPD are located extracellularly between the photoreceptors and the RPE, i.e. behind the outer blood-retinal barrier and are therefore shielded from the choroidal vascular circulation.(94) Mass spectrometry, x-ray microanalysis and histochemistry show that lipids are a major component of drusen, which includes both esterified and un-esterified cholesterol.(151) Other research has demonstrated that the composition of RPD and drusen differ, with the former containing lower quantities of lipid compared to the latter, and this difference could be a result of RPD being relatively shielded from systemic factors and instead being influenced to a greater degree by local metabolic dysfunction, including impaired lipid cycling. This lipid cycling deficit may result in protein spill into the subretinal space, leading to photoreceptor damage and ultimately GA.(94, 98) In addition to this, large bilateral drusen have been found to be linked to atherosclerosis and hypercholesterolemia, whereas RPD have not,(50) so it appears that drusen may form as a result of systemic lipid-related factors, whereas RPD are more a result of local lipid cycling deficits as already mentioned, however the picture is far from clear. Furthermore, the variants of the hepatic lipase (LIPC) gene, which is responsible for encoding for hepatic triglyceride lipase expression in the liver, retinal cells and subretinal space, have been identified as risk factors for RPD,(100) although this link has also been disputed.(152) To add to the confusion, other studies even suggest that LIPC gene variants are linked to a *decreased* risk of ARMD.(153, 154)

It has been reported that there may be a protective role in ARMD from lipophilic statins, with a lower incidence of RPD in patients receiving this type of medication.(100) Conversely, other studies have found no such reduction in the risk of ARMD with lipophilic medication.(155-157) It would appear, therefore, that further research with larger samples is required to provide more evidence surrounding ARMD and the use of lipophilic statins.(100)

1.3.9.1 Detecting lipids in the retina via fundus autofluorescence

The "onion sign", a lesion which is associated with chronic exudation associated with type 1 neovascularisation in ARMD contains lipid, collagen and fibrin, and tends to occur within a vascularised PED.(158) Utilising CFP, the onion sign is visualised as glistening patches within yellow-grey deposits, and with FAF mild hypoautofluorescence in observed.(158)

When confluent, these lesions form into plaques in the outer plexiform layer (OPL), corresponding to hard exudates on SD-OCT imaging, which represent lipid rich deposits. The “onion” layers visible on SD-OCT B-scan are created as a result of intermittent periods of exudation, in the same manner as the formation of sedimentary rocks.(158) It would therefore appear that lipids are best observed via SD-OCT with FAF providing a supporting, confirmatory role.

1.3.10 The role of computers/algorithms in fundus autofluorescence

One study has proposed that it may be possible to use computerised algorithms to improve how FAF signals are interpreted, and reduce intergrader discrepancies.(159) These programmes enhance dark images (reducing the need for pupillary dilation), remove the effect of interfering blood vessels, eliminate outliers and minimise the dispersion of results. However, even after computer analysis, variability in subjective interpretations of FAF images still exists, although significantly reduced. The study aimed to simplify the grading of images, and reducing the number of steps required to create a full analysis, but the authors found that a combination of methods was still the most effective method, and more efficient than using a single tool.(159) Another study also reported that computerised analysis of FAF can aid the detection and interpretation of FAF images specifically with regard to GA area, and that the intergrader agreement was significantly improved over manual interpretation methods.(160) In summary, the role of computers in analysing FAF images is rapidly progressing, particularly with the advent of the greater use of AI, however, more research is required before computers can be relied upon to match or exceed the performance of human graders.

1.3.11 Discussion

FAF may be thought of as providing a topographical map of fluorophores within the retina, and plays an important role in analysing the macula, however, a multimodal approach is recommended for the analysis of retinal changes associated with ARMD. There are three commercially available instuments capable of FAF imaging with cSLO having advantages over fundus camera based FAF of superior contrast, better resolution, and enhanced low signal detection, however it suffers from greater foveal artefacts due to light absorption by macular pigments. There is debate over whether cSLO or fundus camera based FAF is affected most by artefacts related to cataract formation. There has been much interest recently in Ultrawide FAF which may play a future role in the management of ARMD as the importance of peripheral retinal health in macular degeneration becomes more apparent.

FAF is an excellent imaging modality for the detection of small areas of GA, measurement of GA size and the detection of halo effects surrounding areas of GA which may indicate progessive lesions. En face OCT also may have a future role in determining similar features to FAF with research in this area ongoing. Along with NIR, FAF is one of the best imaging modalities for the detection of RPD, with these lesions having strong links to progressive ARMD. FAF may also have a role in determining retinal sensitivity, and also in determining the need for, and the monitoring of the effectiveness of nutritional supplementation for ARMD.

Conclusion.

FAF imaging is likely to become a more widespread imaging modality in the fields of primary, secondary and research optometry/ophthalmology, with the understanding of its uses and relevance continuing to expand.

1.3.12 Further Research Opportunities

Questions that the following Chapters two to nine will attempt to answer:

1. Is pharmacological pupillary dilation required to obtain clinically useful FAF images?
2. How do cataracts affect FAF image quality?
3. Is FAF imaging clinically acceptable to patients in terms of visual/ocular comfort?
4. Can artificial intelligence, machine learning and deep learning use patient history and the results from CFP and OCT to predict FAF imaging results?
5. Can transfer learning be used to identify patterns within FAF images?
6. What clinical features are related to specific FAF patterns?
7. Do relationships exist between the clinical variables collected in this study?

Limitations of Chapter one: Only six databases were utilised for the search strategy, and an English translation was could not be obtained for one article identified as relevant.

In conclusion, Chapter one has highlighted that FAF offers extra information, which although challenging to interpret, may add relevant data for patient profiling, diagnosis, prognosis and treatment which could have a practical role in enhancing the care offered to patients by clinicians.

Chapter 2 Study Methods

2.1 Geographical setting

All the images used in this study were taken personally by the Primary Researcher, community and hospital optometrist Roger Smyth (RS), in a community high street optometry practice based in the market town of Beverley in the County of the East Riding of Yorkshire, in the United Kingdom. The East Riding of Yorkshire is one of the most monoethnic areas in the United Kingdom, with 97.4% of the inhabitants being recorded as "white" in the most recent 2021 national census.(161) Genetically speaking, Hull and East Yorkshire have a strong Scandanavian/Viking influence, anecdotally illustrated by the high prevalence of pseudoexfoliative syndrome in the population (from verbal discussions with local ophthalmologists specialising in glaucoma). East Yorkshire was a prime site for invaders in the ninth century due to the fertile land and nearby river Humber, which gave access to the river Ouze leading to the Viking capital of York (Jorvik).

2.2 Inclusion criteria

Patients were selected on the basis of being over 50 years of age. This age criterion was selected based on the design of a study conducted into the prevalence of ARMD over three continents including Europe, America and Australia published in 2013. The three age ranges selected for this Three Continent ARMD Consortium study were: In the European cohort > 55 years of age, in the American cohort 43-84 years of age, and in the Australian cohort > 49 years of age.(162) The average lowest age of inclusion for these three studies is 49 years. Also, according to earlier research by The International ARM (age-related maculopathy) Epidemiological Study Group, ARM is defined as a degenerative disorder in people > 50 years of age.(11) Therefore, > 50 years of age was selected as an appropriate criterion for inclusion in this study.

2.3 Exclusion criteria

Since this study was aimed at eliciting whether FAF is useful for identifying retinal characteristics associated with ARMD, patients with macular disease other than ARMD were excluded from the study. Patients with previous known central/branch retinal vein or artery occlusions and/or those with diabetic retinopathy were excluded from the study, however, simply being diabetic alone was not a reason for exclusion. Other non-macular exclusion criteria were any ocular surgery occurring within three months of the study starting, any previous vitreoretinal surgery, choroidal neovascular membranes occurring for reasons other than ARMD, previous retinal photocoagulation in either eye and hereditary retinal disorders. Patients with epiretinal membranes (ERMs), vitreomacular traction (VMT), lamellar holes (LHs), full thickness macular holes (FTMHs) and macular cysts (MCs) were not excluded from the study, as these have been found to be relatively common in patients over the age of 50 years and are therefore often found in conjunction with ARMD. A recent epidemiological study of vitreoretinal interface abnormalities found, using SD-OCT in a cohort of patients from 63-102 years of age, that ERMs had a prevalence of 34.1%, VMT 1.6%, LHs 3.6%, FTHMs 0.4% and MCs 5.6%. The study concluded that the prevalence of MCs, ERMs and VMT all increased with age, whilst LHs were not age associated.(162) Patients in the current study were initially screened for narrow iridocorneal angles by assessment of the anterior chamber depth and by use of the Van Herrick slit lamp examination technique, both temporally and nasally, and those patients considered to be at risk of an acute angle closure event were excluded and referred to secondary care, if appropriate, according to the latest “primary angle closure suspect plus (PACS+)” criteria, (163). Gonioscopy was not carried out on participants prior to dilation, but rather a Van Herrick grade of 0, 1 or 2 was considered to pose a significant risk of acute angle closure on dilation, based on findings from a previous study that states that Van Herrick grade 2, or a relation between the corneal thickness and the peripheral anterior chamber depth of <1/2 – ¼, signifies that angle closure is “possible”.(164) Patients were initially asked if they suffered from photosensitive epilepsy, or photosensitive migraines. If they belonged to either group, they were also excluded from the study.

Inclusion criteria	Exclusion criteria	Non-excluded conditions
Over 50 years of age	Known macular disease other than ARMD	Diabetes
	Previous CRVO/BRVO/CRAO/BRAO	Epiretinal membrane
	Diabetic retinopathy	Vitreomacular traction
	Narrow iridocorneal angles	Lamellar holes
	Photosensitive epilepsy	Full thickness macular holes
	Photosensitive migraine	Macular cysts
	Any ocular surgery occurring within three months of the study starting	Non-photosensitive epilepsy
	Any previous vitreoretinal surgery	Non-photosensitive migraine
	CNVs occurring for reasons other than ARMD	
	Previous retinal photocoagulation in either eye	
	Hereditary retinal disorders	

Table 2.1 Inclusion and exclusion criteria from the study

2.4 Patient Information

Patients were recruited when they presented for an eye examination at the community optometry practice. All consecutive patients who satisfied the inclusion and exclusion criteria were invited to participate in the study. Subjects were provided with information in the form of a detailed patient information sheet and were also given an appropriate amount of time to decide if they wished to participate in the study. For some patients this took the form of having their eyes examined and then first selecting spectacle frames before finally having the extra images taken, pre and post dilation. For others this meant returning to the practice on another day specifically for the extra tests, or to return for other services on a dual-purpose visit as necessary. Subjects were also given the opportunity to withdraw from the study at any time, and were informed that they could ask for their collected data to be deleted. Data was anonymised, and to assist with this process patients were issued with a unique patient number for the purposes of the project.

2.5 Ethical approval

Ethical approval was sought and granted from Aston University's Research Ethics Committee before the project began. REC REF #1604 was granted on 5th February 2020.

2.6 Instrumentation

Figure 2.1 The Nidek Retinascan Duo RS-330 (FAF model)
(reproduced by kind permission from NIDEK)

The Nidek Retinascan Duo RS-330 (FAF model), hereafter referred to as the RS-330, was used to capture all the ocular images for this project.

The RS-330 is a combined OCT and 12-megapixel 45-degree colour fundus camera which is also capable of modified fundus camera-based FAF. FAF is a non-invasive imaging

method that detects the presence of LF (within the RPE cells), without the need for contrast dye, allowing its distribution to be mapped. For the OCT imaging in this study, a wide area retina map scan of 12mm by 9mm was analysed via a 9mm by 9mm macular normative database preloaded within the instrument, as shown in Figure 2.2 below. For the FAF imaging, the area of image capture was the same as the field for the colour fundus photography, i.e. 45 degrees.

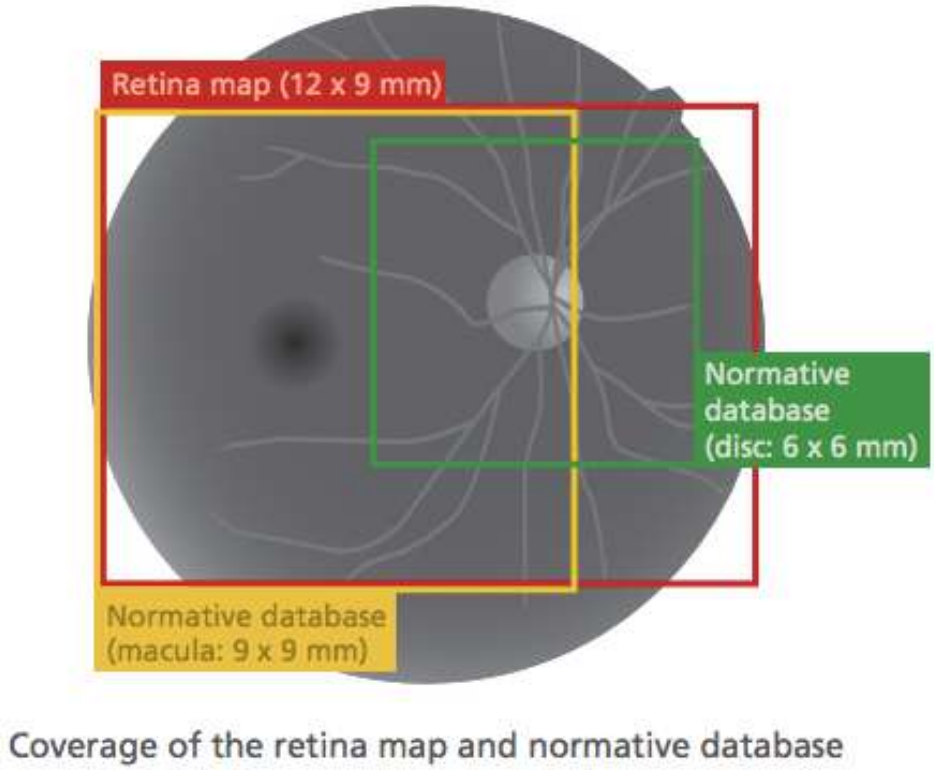

Figure 2.2 Coverage of the retina map and the normative database
(Reproduced by kind permission of NIDEK from the RS-330 manual)

NIDEK® 3-D auto-tracking allows rapid and user-friendly image capturing. OCT sensitivity can be selected, with a choice of either higher definition or higher speed. The “Ultrafine” setting, with a relatively high definition image capture, requires a longer acquisition time and performs 13,250A scans per second. The “Regular” setting benefits from a shorter image capture time at the expense of a lower definition at 53,000A scans per second, whilst the “Fine” setting is a compromise between speed and definition at 26,500A scans per second. For the purposes of this project the “Fine” setting was utilised for all image acquisition as a trade-off between speed of image acquisition and image definition.

The image enhancement function allows for adjustment of image brightness for advanced image quality, and this function was utilised throughout the project when considered necessary for better image capture.

The RS-330 features a joystick which allows for both lateral and vertical adjustment of the unit. A canthus marker on the head rest support enables initial gross alignment of the patient's head position utilising a motorised chin rest prior to fine eye position tuning via rotation of the joystick. Automatic small pupil adjustment optimises the fundus camera settings if a small pupil diameter (< 2mm) is detected. A slider, situated on the side of the unit, allows for compensation of high refractive errors. The unit also features both automatic and manual modes for OCT image acquisition, and during this project the automatic mode was selected by default. The clinician aligns the central green target on the OCT screen with the patient's pupil, prior to automatic image capture.

There are several OCT scans that may be performed with the RS-330 which include; macular cross (a series of horizontal and vertical B-scan cross-sections), macular map (or volumetric) scan and macular radial (a series of clock face cross-sectional scans). The macular map was selected as the most useful scan for the purposes of this study, as it provides a cube of data centred on the posterior pole over a 9mm by 12mm field, allowing the clinician to slowly "scroll through" an entire cross-sectional image of the macular area, also known as an "edge to edge" scan.

2.7 Data Collection Process

As the scans were taken as a part of a general eye health screening test, the combination scan feature of the RS-330 was utilised which allows the operator to carry out a series of scans on each patient which are automatically selected in a specific order. The series selected were as follows (all performed prior to the standard refraction): macular map, disc map, macular cross, CFP, and, for Chapter three only, FAF was performed prior to pharmacologically induced pupillary dilation.

Finally, FAF was performed with pupil dilation after the standard refraction for both the study on pupil dilation in Chapter three, and for the subsequent parts of the study in Chapters four to nine. If a scan was captured that was not considered to be of adequate quality, the “retry” function was used to capture a fresh image until an acceptable image was achieved. Once all images for a patient were collected, the images were saved onto the hard disc of the PC used to run the RS-330’s software. The RS-330 utilises NAVIS-Ex software, an image filing software that enables instrument networking. A weekly back-up of all study images was performed on a removable disc and kept off site outside business hours. Images were anonymised and the disc was kept in a locked cupboard.

2.8 System used in this study for the classification of age-related macular degeneration

For patient demographic purposes, the ARMD classification system suggested by the International ARM Epidemiologic Study Group was utilised in this study. This system was also adopted by Bindewald et al for their study on FAF pattern classification.(165) Early and late ARMD are classified as shown in the table below. Note that the for this classification system patients must be >50 years of age and the macula is defined as a 6000μm (20°) circle centred on the fovea however, for the current study a larger macular area defined by the circle whose diameter is defined by the vertical line drawn between the innermost temporal vascular arcades was used. Previous research has shown that this line in healthy eyes approximates to 9473±1974μm(33±7°)(166, 167)

2.9 The classification system used during this study for FAF images

In 2005, a classification system was proposed by Bindewald et al to place FAF results from patients with early ARMD into one of eight distinct categories. These were; normal, minimal change, focal increased, patchy, linear, lacelike, reticular and speckled.(165) A precise description of each phenotypic pattern in given below, with examples taken from the current study.

Normal: A homogenous background FAF with a gradual decrease in the inner macula toward the foveola due to the masking effect of the yellow macular pigment.

Figure 2.3 Example of a normal FAF image taken from this study.

Minimal change: Very limited irregular increase and decrease of background FAF without a clear pattern.

Figure 2.4 Example of a minimal change FAF image taken from this study.

Focal increased: Defined as having at least one area (< 200μm in diameter) of significantly increased FAF which is much brighter than the surrounding background's FAF signal. The borders are well defined, and the difference in FAF between the brighter area and its surrounding is not gradual. The brighter area may or may not demonstrate a darker halo surround.

Figure 2.5 Example of a focal increased FAF image taken from this study.

Patchy: Characterised by presence of one large area (> 200μm in diameter) of markedly increased FAF. The borders tend to be less well defined than the Focal increased pattern.

Figure 2.6 Example of a patchy FAF image taken from this study.

Linear: This pattern features at least one linear area of markedly increased FAF. The borders of these areas are usually well defined and the difference in FAF between the brighter area and its surrounding is not gradual. These linear patterns usually map to hyperpigmented lines on CFP.
These were no images identified as having a linear FAF pattern in this study.

Lacelike: Multiple branching linear structures of increased FAF in a lace-like pattern which may map to hyperpigmentation on CFP. The borders may be difficult to define.

Figure 2.7 Example of a lacelike FAF image taken from this study.

Reticular: Multiple small areas (< 200μm) of decreased FAF whose borders can be indistinct. This pattern tends to occur in the macular region, but also superotemporally in the retina. The pattern may or may not map to numerous small soft or hard drusen, or pigmentary abnormalities detected on CFP.

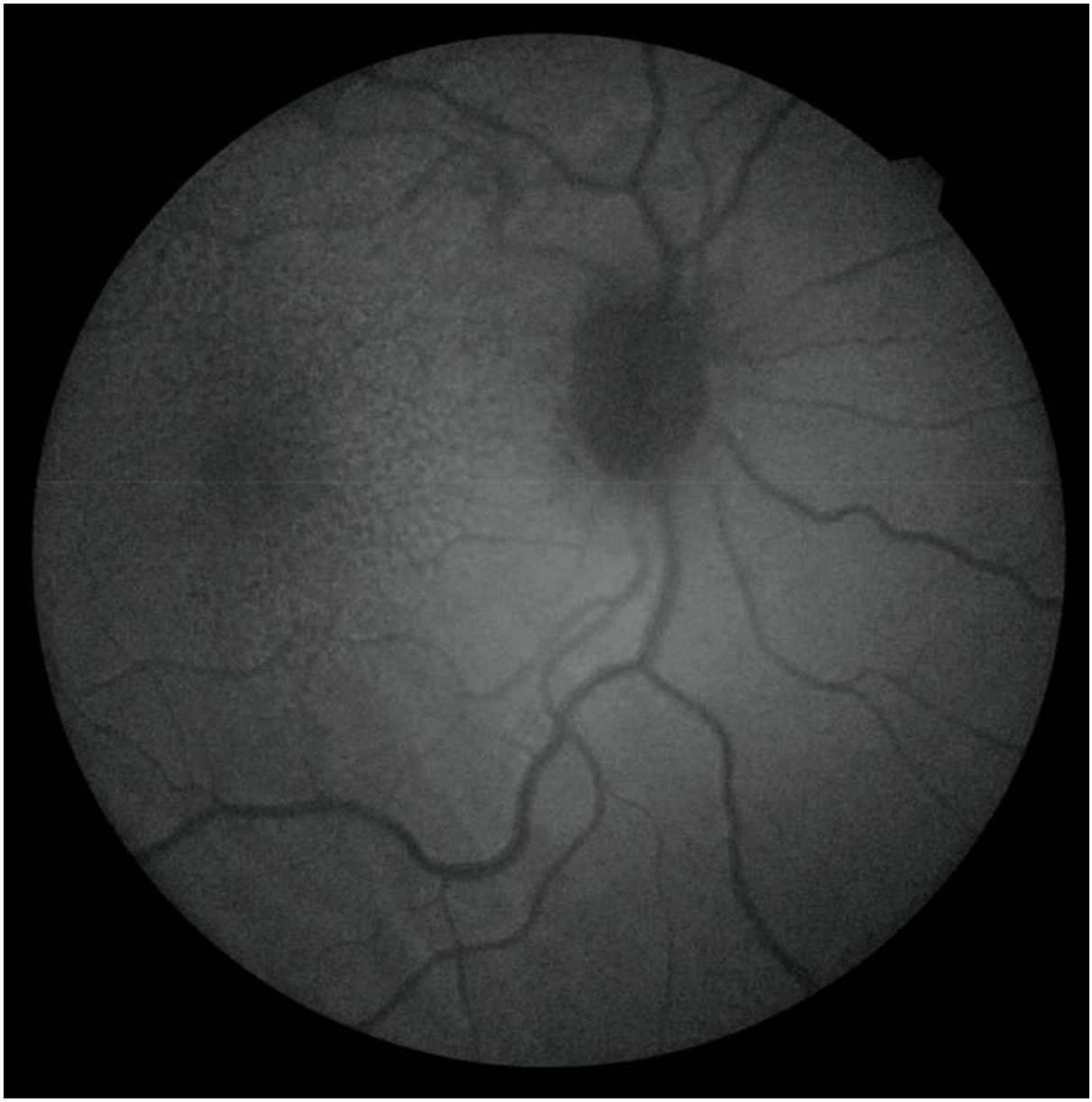

Figure 2.8 Example of a reticular FAF image taken from this study.

Speckled: A mixture of hypo and hyper-autofluorescent FAF abnormalities covering a large area which may extend beyond the macular area to cover the entire posterior fundus. These small areas may be punctate or linear. They may map to hypo and hyperpigmentation and/or multiple sub-confluent and confluent drusen.

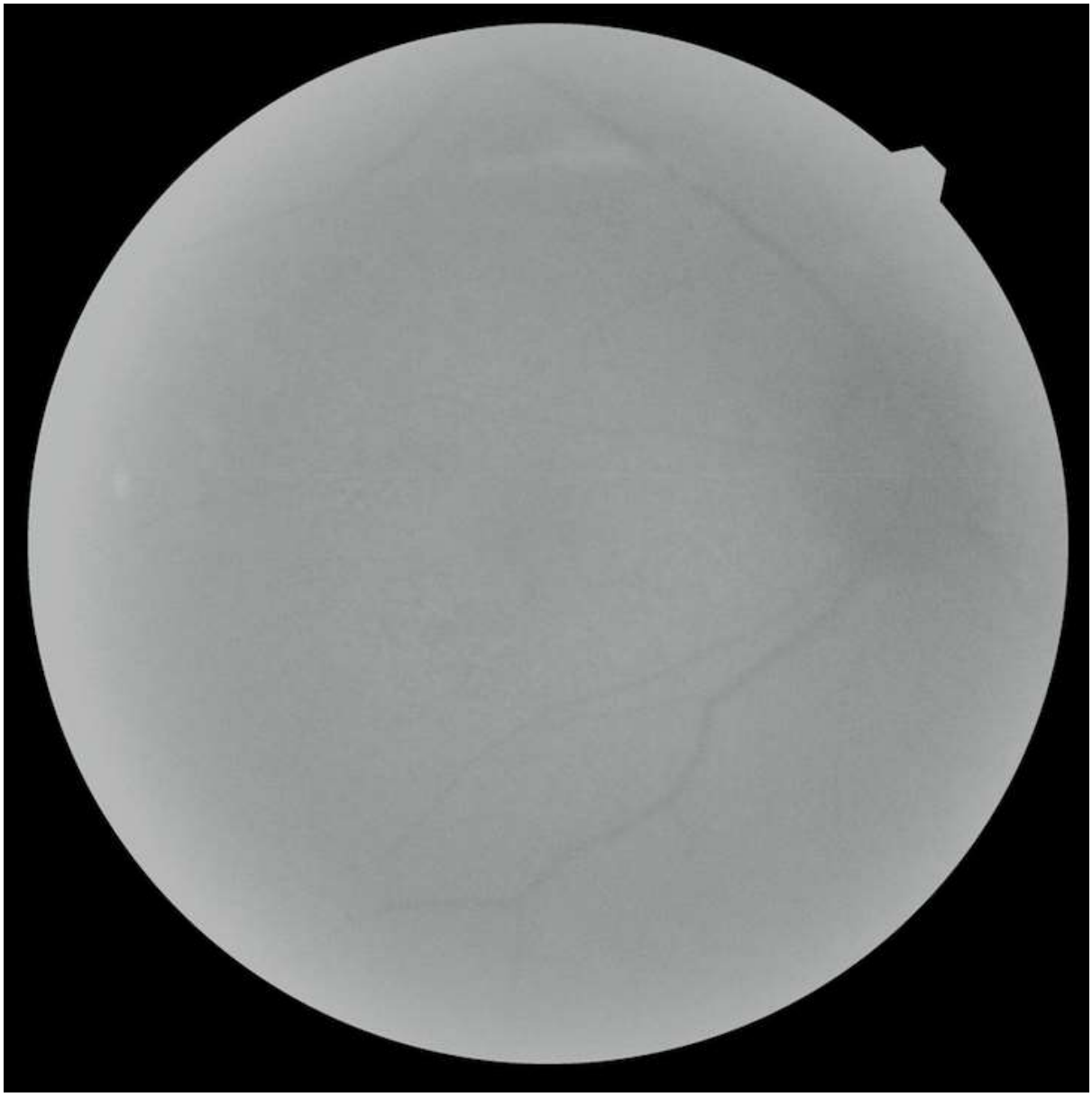

Figure 2.9 Example of a speckled FAF image taken from this study.

2.10 Orange Data Mining widgets used throughout this study explained

Microsoft Excel spreadsheets of variables were uploaded into the ODM software program and analysed via “widgets”, which form the components of this particular AI software. These widgets allowed analysis of the data from a variety of perspectives including: 1. Comparisons of associated frequencies between clinical findings (the variables), 2. Ranking of clinical findings as predictors of abnormal FAF classifications 3. Confusion matrices, 4. Predictive nomograms 5. Principle component analysis and 6. Image analysis.

ODM widgets used in this study:

1. File
2. Distributions
3. Boxplot widget
4. Data Sampler
5. Select Columns
6. Preprocess
7. Test and Score (with associated model learners: Naïve Bayes, Logistic Regression, k nearest neighbours (kNN), Random Forest, Tree, Neural Network and small vector machine (SVM))
8. Scatter Plot
9. Confusion Matrix
10. Rank
11. Principal component analysis (PCA)
12. Nomogram
13. Receiver operating characteristic (ROC) analysis
14. Import images and Import embedding widgets
15. Distances widget
16. Hierarchical clustering widget
17. Multidimensional scaling (MDS) widget

1. File Widget

The File widget is used to load data into the ODM canvas, as well as enabling the definition of class type and meta attributes. Orange is compatible with Microsoft Excel files or Google Sheets documents, or any comma or tab-delimited file. Attribute names are placed at the top of columns.

2. Distribution widget

This widget displays value distributions for a single variable, and allows exploration of how variables relate to one another. The widget features “fitted distribution” which fits selected distribution curves to the plot. The fittings are: Normal, Beta, Gamma, Rayleigh, Pareto, Exponential, and Kernel Density.

In the columns section, “split by” displays value distributions for instances of a particular class. “Stack columns” displays one column per bin, coloured by proportions of class values.

“Show probabilities” shows probabilities of class values at the selected variable, and finally “Show cumulative distribution” cumulatively stacks frequencies.

In this study, “split by” was used to explore the relationships between the variables and whether the cases demonstrated a normal or abnormal FAF, as this function produces a graph that is both easy to interpret and uncluttered, along with the “Normal” fitted distribution curve for simplicity. In the Columns section, “show probabilities” was selected.

3. Box plot widget

This widget allows examination of attribute value distributions. It also enables checking for outliers and/or duplicated values. In Figure 2.3 below, one FAF image graded as “minimal change” was labelled as “abnormal” in error (on the third row from the top), enabling correction. This is how the output is displayed when comparing two categorical variables.

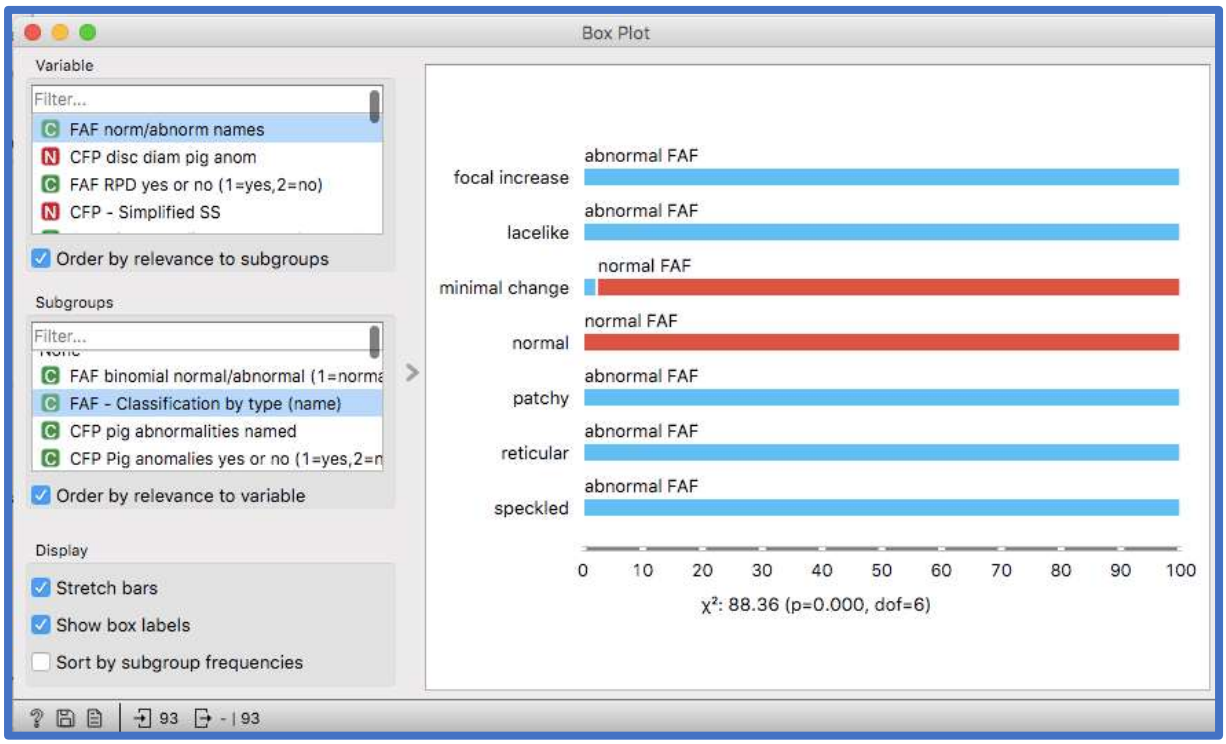

Figure 2.10 Orange Data Mining Boxplot widget input error screening

The chosen variable to be examined is first selected (in the upper left box), and if “Order by relevance of subgroups” is ticked, this presents the variables by order of their Chi² or ANOVA value with the most relevant at the top (ODM selects the most appropriate test).

Choosing a subgroup (in the lower left box) displays the Boxplot for this specific subgroup, and if “Order by relevance to variable” is ticked, this orders the subgroups by their Chi² or ANOVA value (again, as with the upper box, ODM selects the most appropriate test).

Below in Figure 2.11 is another example taken from ODM below, showing how the output is displayed for comparison of one numeric and one categorical variable. When “Annotate” in the “Display” box is ticked, the median is displayed below the line, and the mean (with standard deviation) above the line. Ticking the “compare means” box compares the means of the selected subgroups and gives the student’s t test “p” value.

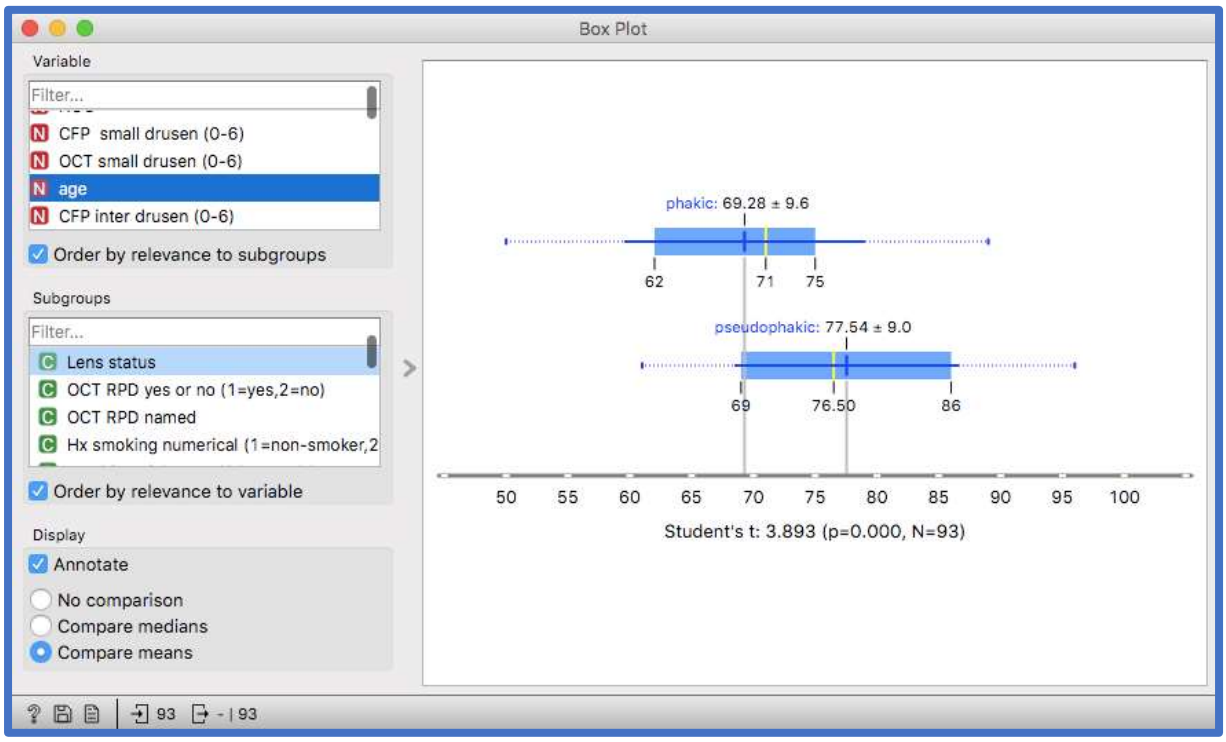

Figure 2.11 Orange Data Mining Boxplot of lens status and age.

The dark blue vertical line represents the mean, with the dark blue horizonal line indicating the standard deviation. The central light blue shaded area illustrates the values between the first (25%) and third (75%) quartile. The horizontal dashed blue indicates the range. Finally, the yellow vertical line represents the median.

When the same categorical attribute is selected as both the variable and subgroup, as shown in Figure 2.12 below, the bars represent the number of instances with each particular attribute value, i.e. 65 normal and 28 abnormal FAF images.

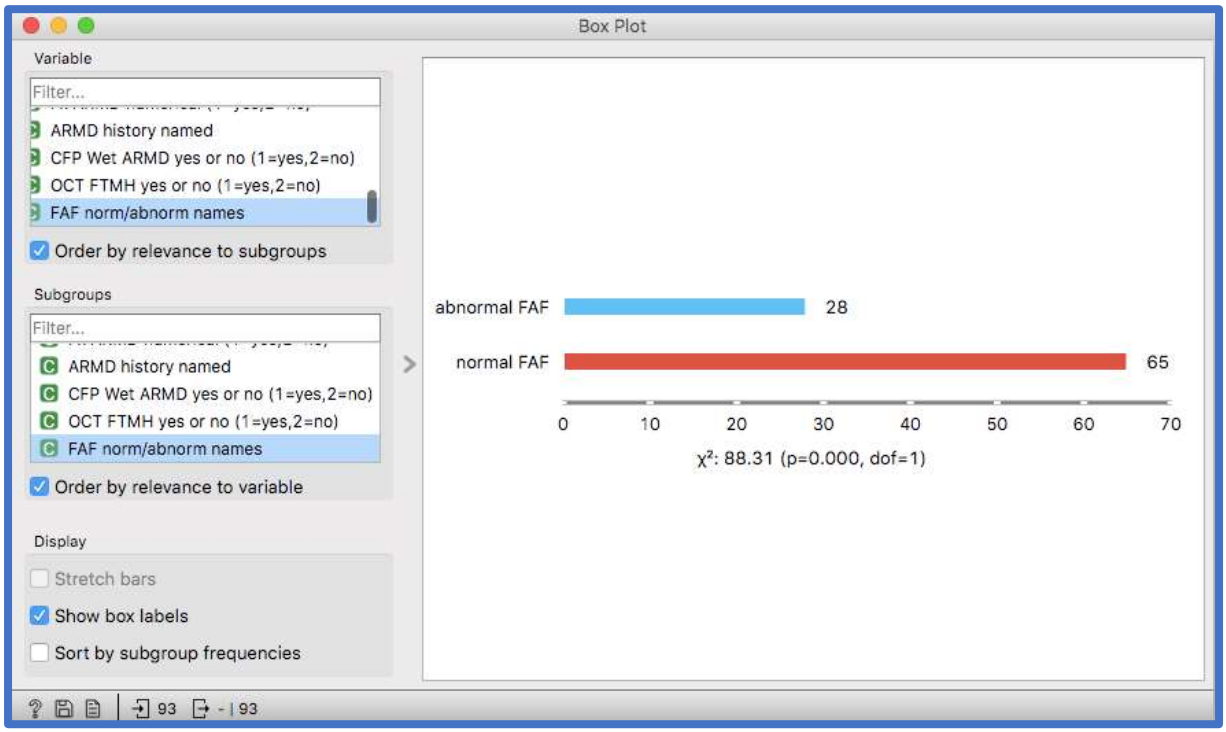

Figure 2.12 Orange Data Mining Boxplot of FAF classifications

Below in Figure 2.13 is an illustration of stretch bars and box labels. Stretch bars show the relative values of data instances; if stretch bars are not selected, absolute values are given. Ticking “sort by subgroup frequencies” simply sorts the subgroups by their descending frequency.

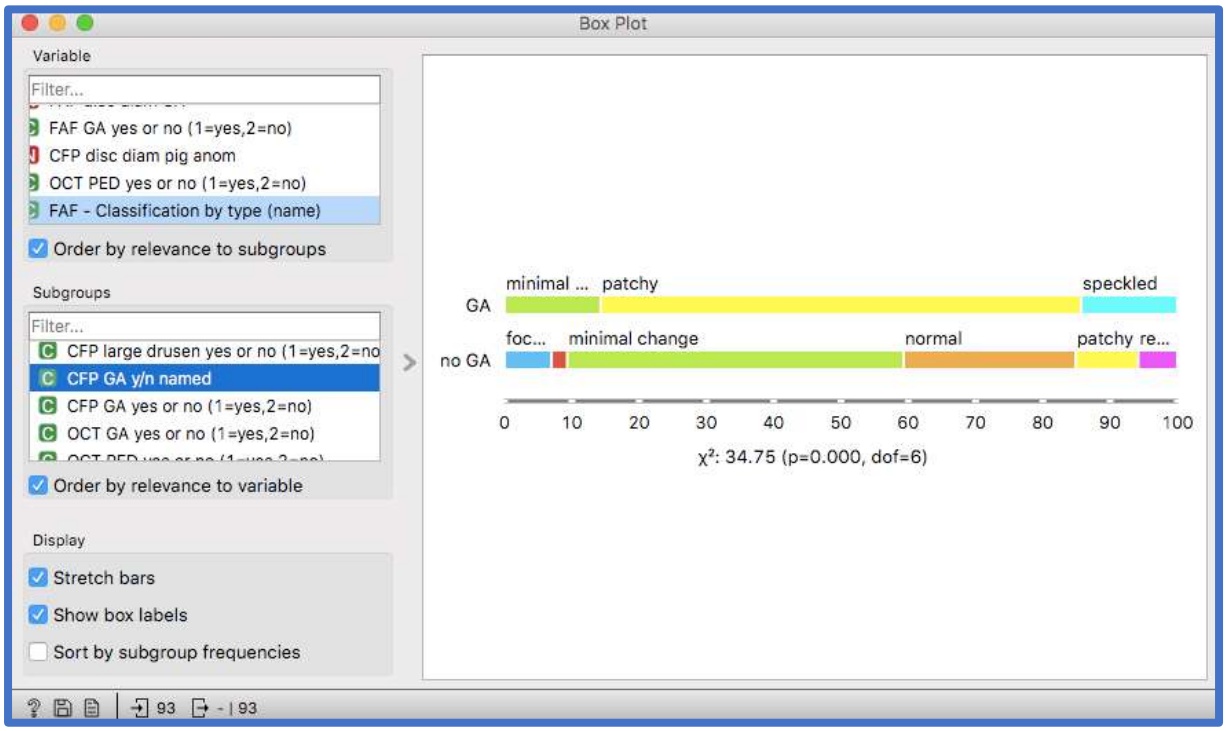

Figure 2.13 Illustration of stretch bars and box labels in the Orange Data Mining Boxplot

4. Data Sampler widget

The Data Sampler widget allows selection of a proportion of the total input data. In this study on FAF, this widget was utilised to examine the predictive abilities of model learners for increasingly larger percentages of the entire data set. This enabled a calculation of sample size adequacy, by measuring when graph of the degree of informedness (sensitivity + specificity -1) versus the percentage of the sample data utilised plateaued.

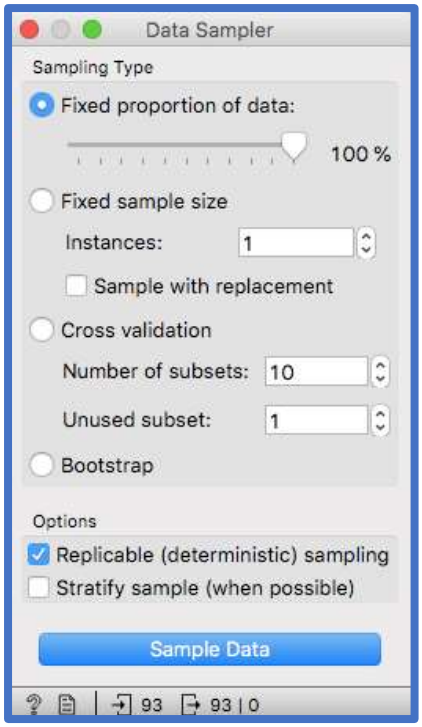

Figure 2.14 Screen shot taken from the Orange Data Mining sampler widget.

A fixed percentage of the data can be selected, as shown in the screen shot in Figure 2.14 above. However, it is also possible to select a fixed sample size, with or without replacing the selected data. If replaced, the next selection is made from the entire data set. Even if 100% of the data, or all instances are selected, shuffling still occurs. Cross validation splits the data into subsets, one subset is held back as “remaining data”, the others are outputted as the data sample. Bootstrapping is a process whereby the sample data is resampled to make an inference about the sample. In this way a model for inferring the actual population from the sample can be created. “Replicable sampling” ensures sampling that can be used by a variety of models, whereas “Stratify sample” mirrors the input dataset’s composition. For this study “Replicable sampling” was utilised.

5. Select Columns widget

This widget is used to select specific variables from the input data to be used in the analysis. Below, in Figure 2.15 can be seen an example of how the widget was utilised for this study on FAF. Ignored variables, in the left-hand box, include those related directly to the FAF result, which would contaminate the data being explored as a predictor of the FAF outcomes as normal or abnormal. Also ignored was the "years of smoking cessation", with a binary variable "Ceased smoking ≥ 20 years/never smoked" selected for analysis instead. This is because a non-smoker would either have a lifetime (which will vary naturally according to age), or zero years of smoking cessation, both of which could be confusing for the statistical results. The selected variables for analysis are listed in the right-hand box. The target variable, which the selected variables are attempting to predict, is placed in the "Target" box. Meta attributes, which may be placed in the bottom right-hand box, can be useful for labelling, but are not used in modelling processes. It can be seen that Orange labels each variable with a type, for example, "FAF classification (1-8)" is labelled "N" for numeric, and "FAF Classification by type (name)" is labelled "C" for categorical.

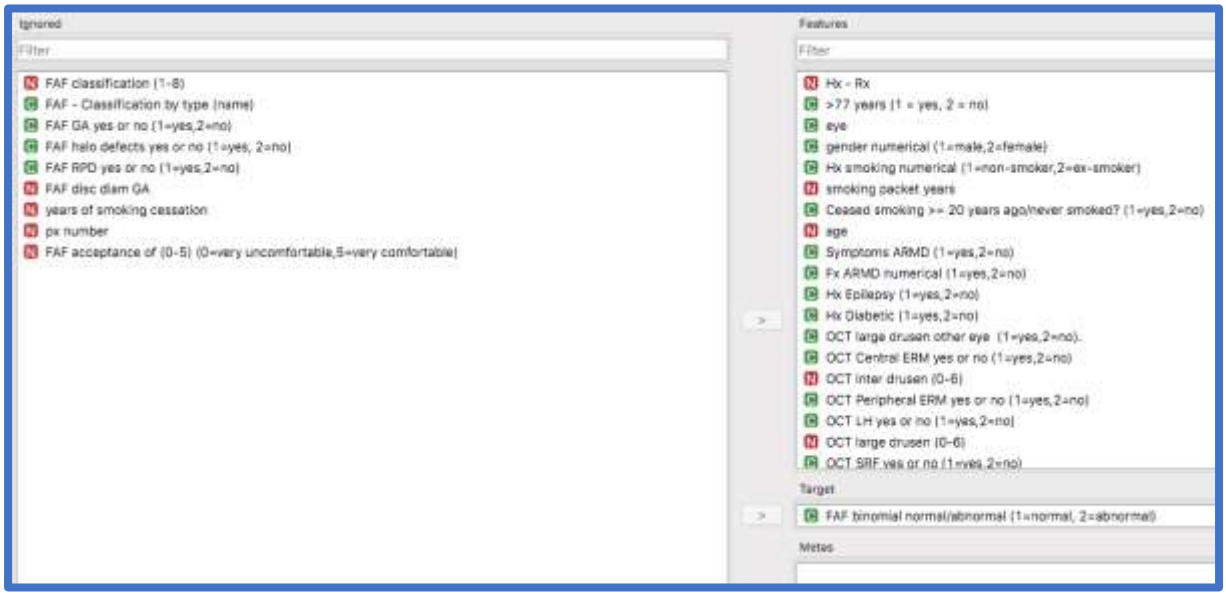

Figure 2.15 Illustration of the Orange Data Mining Select columns widget

6. Preprocess widget

Preprocessing of data is a crucial step that must be taken prior of analysis of data. ODM provides the Preprocess widget, shown below in Figure 2.16, which enables several manipulations to improve the data quality. It can make continuous variables discrete (or the opposite function), input missing values, select the most relevant variables for analysis by a variety of scoring methods (e.g. in this study the 18 most informative variables were selected by the Information Gain method), and finally normalise variables. Note that this "normalisation" function is not referring to the "normal" distribution of data, but in this context is transforming the values into relative terms e.g. to a scale from 0 to 1.

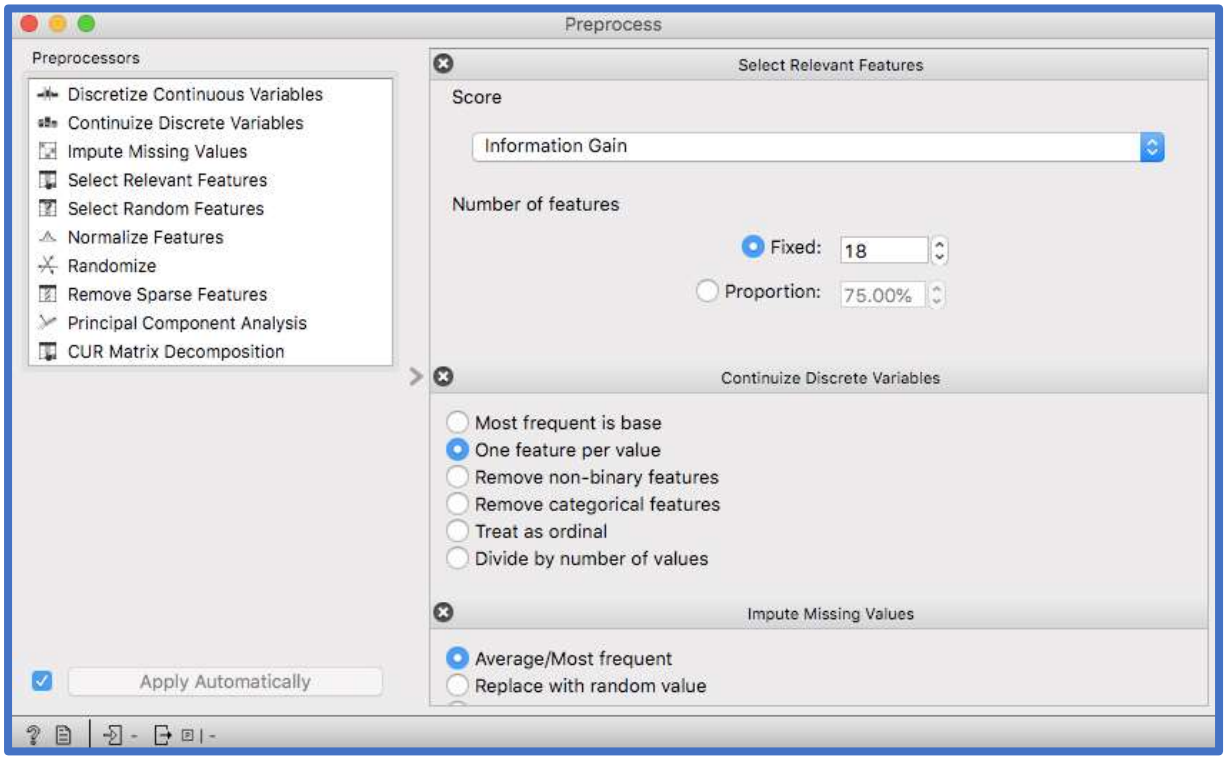

Figure 2.16 Illustration of the use of the Orange Data Mining Preprocess widget

7. Test and score widget

This widget tests the model learners (i.e. the algorithms). Firstly, it displays the different performance measures, including area under the curve (AUC), Recall (sensitivity), and Specificity. It also produces an output that enables other widgets to analyse the performance of classifiers, e.g. the Confusion Matrix and ROC (Receiver Operating Characteristic) Analysis. An example is shown below in Figure 2.17.

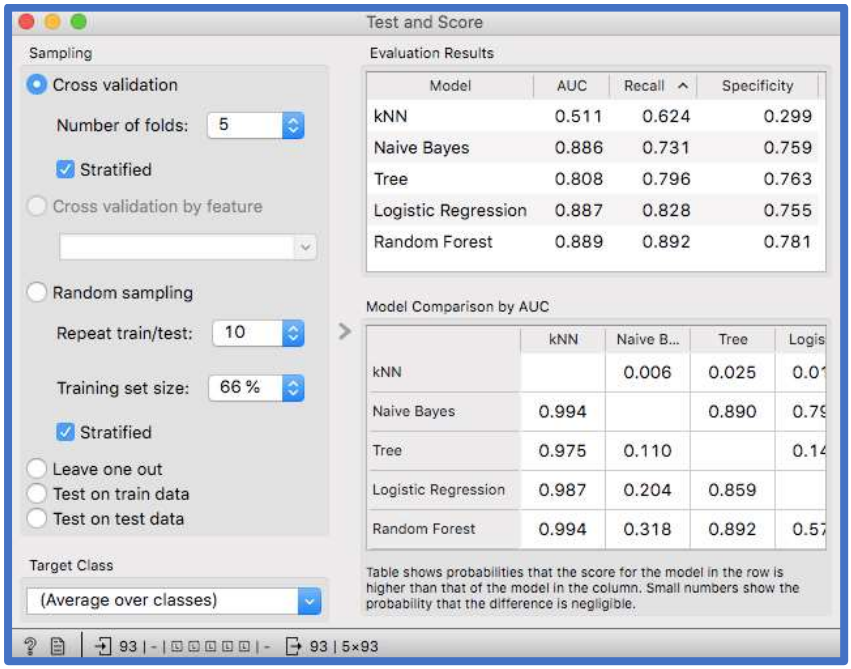

Figure 2.17 Illustration of the Orange Data Mining Test and Score widget

“Cross-validation” splits the data into a selected number of folds. The algorithm is then trained on all the folds bar one, and then tested on the held-out fold. This is then repeated for all the folds. “Leave one out” is similar, but instead of an entire fold being held-out, a single instance is kept back from the initial training before that single instance is classified by the algorithm and this is then repeated for all the data (this method is very reliable but with the disadvantage of being very slow). “Random sampling” is a third option, where the data is split into a training and testing set in a given proportion, and this is then repeated a set number of times. “Test on train” data is not recommended due to often generating incorrect results. “Test on test” data allows the researcher to input another data set with testing examples. Note that in this study the “Average over classes” option in the Target Class box was selected by default. The researcher may also select the scores for predicting either a normal (class 1) or an abnormal (class 2) FAF result.

8. Scatter Plot widget

The Scatter Plot widget allows 2-dimensional visualisation of the data as shown below. Below in Figure 2.18, the “show colour regions” illustrates the trend for a greater number of large drusen, as detected by OCT, with increasing patient age, accompanied by a greater likelihood of an abnormal FAF result.

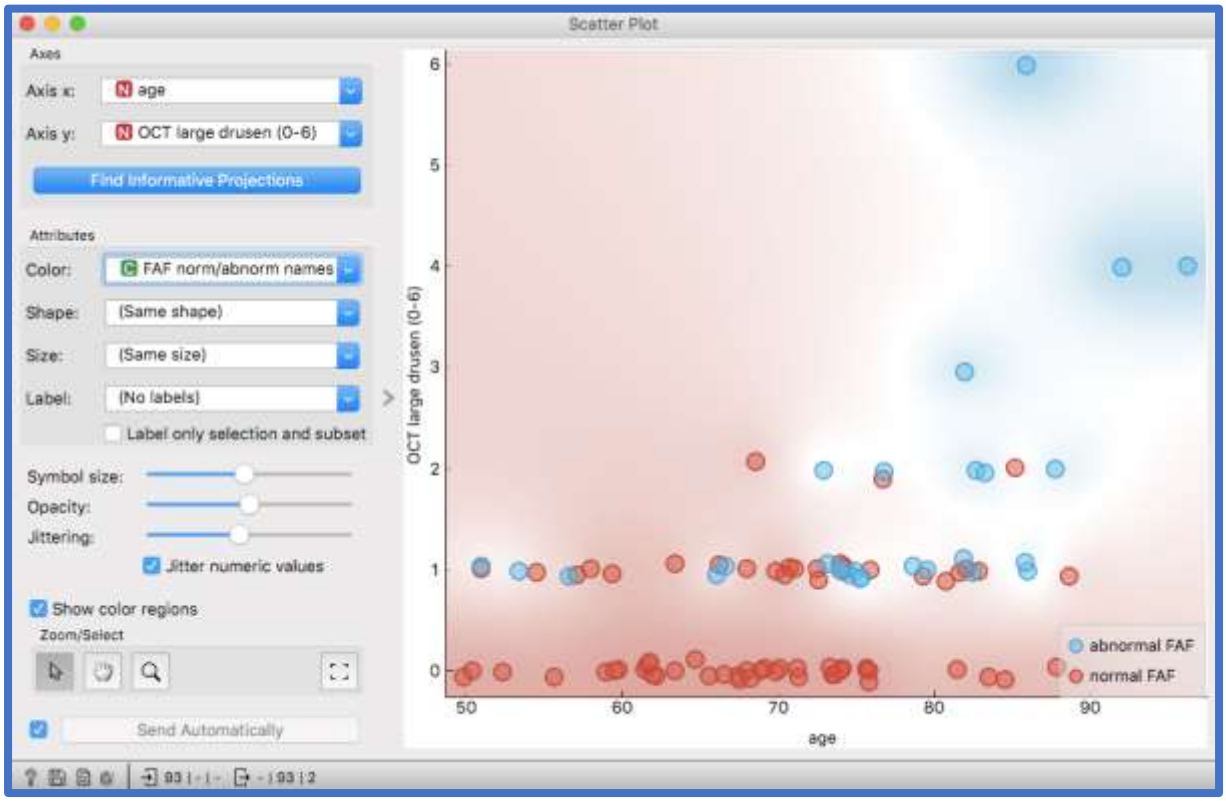

Figure 2.18 Illustration of the Orange Data Mining Scatterplot widget. Scatter plot taken from ODM showing a graph of age versus the number of large drusen as measured by OCT.

The “Find informative Projections” option, shown below in Figure 2.19, enables a rapid scan to explore the data for useful Scatter plots. This is performed by the function finding the 10 nearest neighbours for each data instance, in the graph of the two chosen features plotted as shown above in Figure 2.18. It then identifies how many of the 10 have the same colour. The score allocated to the projection is the average number of neighbours with the same colour as the data instance. A league table of scores is then produced as shown below. From the Score Plot it is revealed that Random Forest scores the highest for predicting FAF normality/abnormality and the “CFP DD GA” (colour fundus photography disc diameters of geographic atrophy) is the most informative variable for this particular model learner.

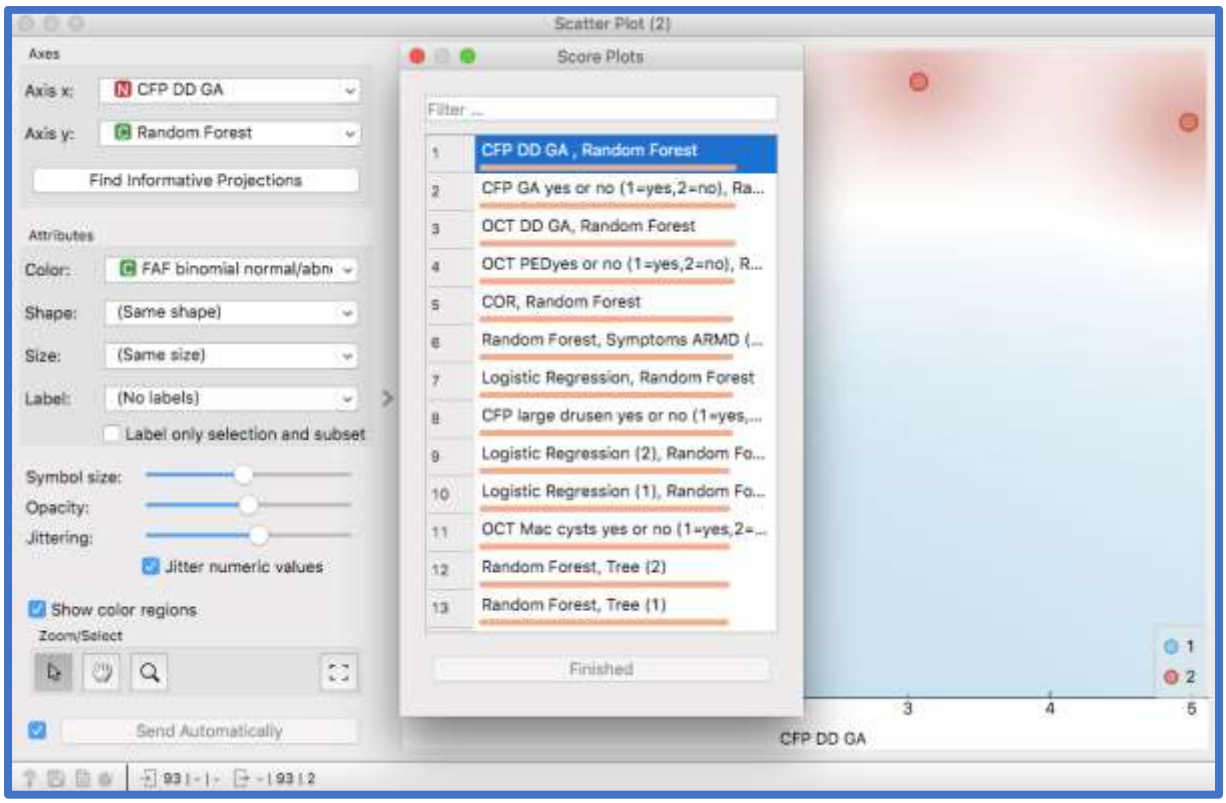

Figure 2.19 Illustration of the Orange Data Mining Scatterplot Informative Projections Score plot.

9. Confusion Matrix widget

The Confusion Matrix widget gives the number or proportion of instances, for the chosen learner, that were identified correctly (highlighted in the diagonal in blue) and those identified incorrectly (highlighted in the other squares in pink), shown in Figure 2.20 below.
Along the bottom are selection options where it is possible to select only the correctly or incorrectly identified instances in order to feed further widgets, e.g. a data table or scatter plot in order to clarify why specific cases were misclassified. It is also possible to select individual boxes. In the top right corner, in “Show”, it is possible to select “Number of instances” as shown. Alternatively, “Proportions of predicted” or “Proportions of actual” can be selected. The former shows cases classified as a percentage of the total predicted instances, whereas the latter shows the cases classified as a percentage of the total actual instances.

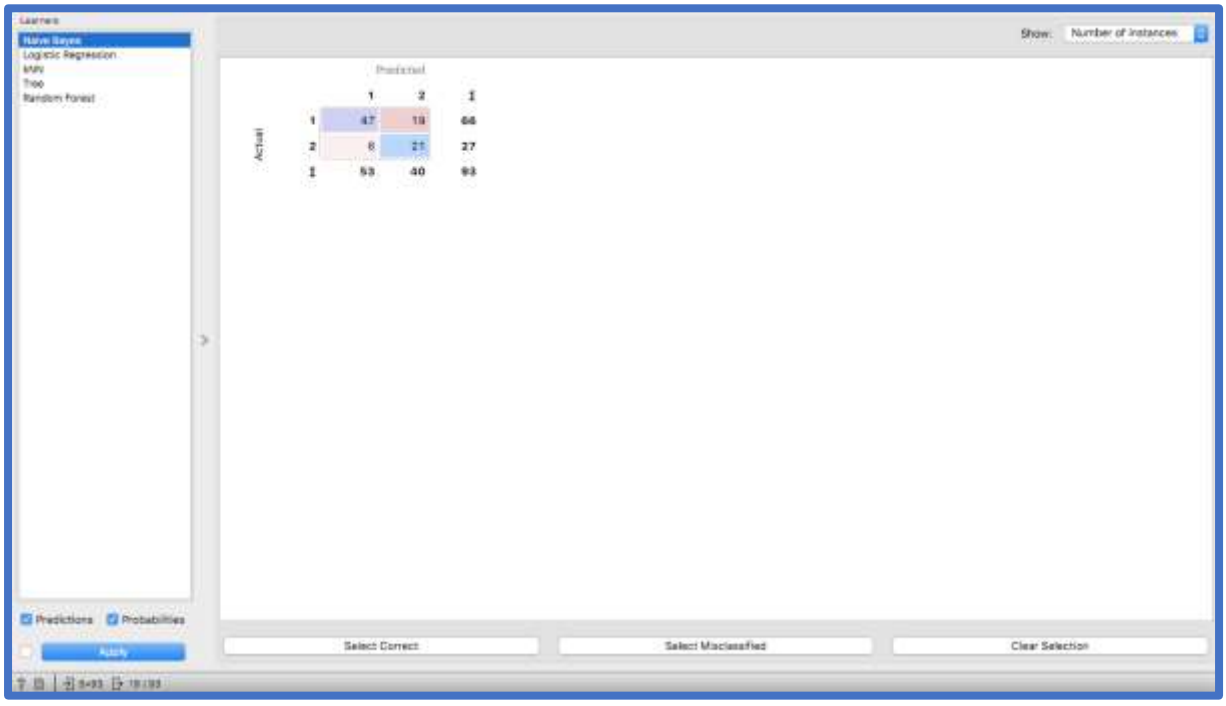

Figure 2.20 Illustration of the Orange Data Mining Confusion matrix contents.

10. Rank widget

The Rank widget scores variables for their association/correlation with the target variable, from a variety of internal scoring methods, i.e. Information Gain, Information Gain Ratio, Gini Decrease, ANOVA, Chi², ReliefF and FCBF, and any connected external models. These scoring methods are defined in Table 2.3 below:

Internal scoring method	**Definition of scoring method**
Information Gain	the expected amount of information
Information Gain Ratio	ratio of Information Gain to the feature's intrinsic information
Gini Decrease	the inequality of the values of a frequency distribution
ANOVA	the difference in the average of a feature's values within a class
Chi²:	association of feature and class as measured by Chi-squared
ReliefF	A variable's prediction of class for similar amounts of data
FCBF	fast correlation-based feature – detects lack of order (entropy)

Table 2.2 Internal scoring methods definitions.

The Rank widget can also utilise unsupervised data from external scorers, e.g. principal component analysis (PCA).

An example of the output from the Rank widget is shown below in Figure 2.21. In the top left corner, the “Scoring Methods” can be selected. In “Select Attributes” in the bottom left corner none, all, manual and “n” best ranked Attributes can be chosen.

Figure 2.21 Illustration of the Orange Data mining Rank widget output.Table from ODM showing the 18 best ranked variables on the “Information Gain”.

11. Principal component analysis (PCA) widget

This widget is covered in more detail in Chapter nine, dedicated to PCA. In summary, PCA summarises large data sets containing many variables, into a smaller number of “factors”, whereby the data can be more easily visualised and analysed.

12. Nomogram widget

The Nomogram widget allows the Logistic Regression and Naïve Bayes model learner outputs to be visually represented. By moving the sliders up and down the horizontal bars of each variable, the effect of the variable's value on the class probabilities can be directly observed. When there are too many variables, the "best ranked" can be selected as shown in the "Display features" section on the left of the diagram below in Figure 2.22. They can be ranked in a variety of ways, including by name and absolute importance (for both Naïve Bayes and Logistic Regression), and also by positive influence and negative influence (for Naïve Bayes only). The scale along the top right can be either represented as log odds ratios or for simplicity as a point scale as illustrated in the example, with the maximum absolute log odds ratio represented by 100 points.

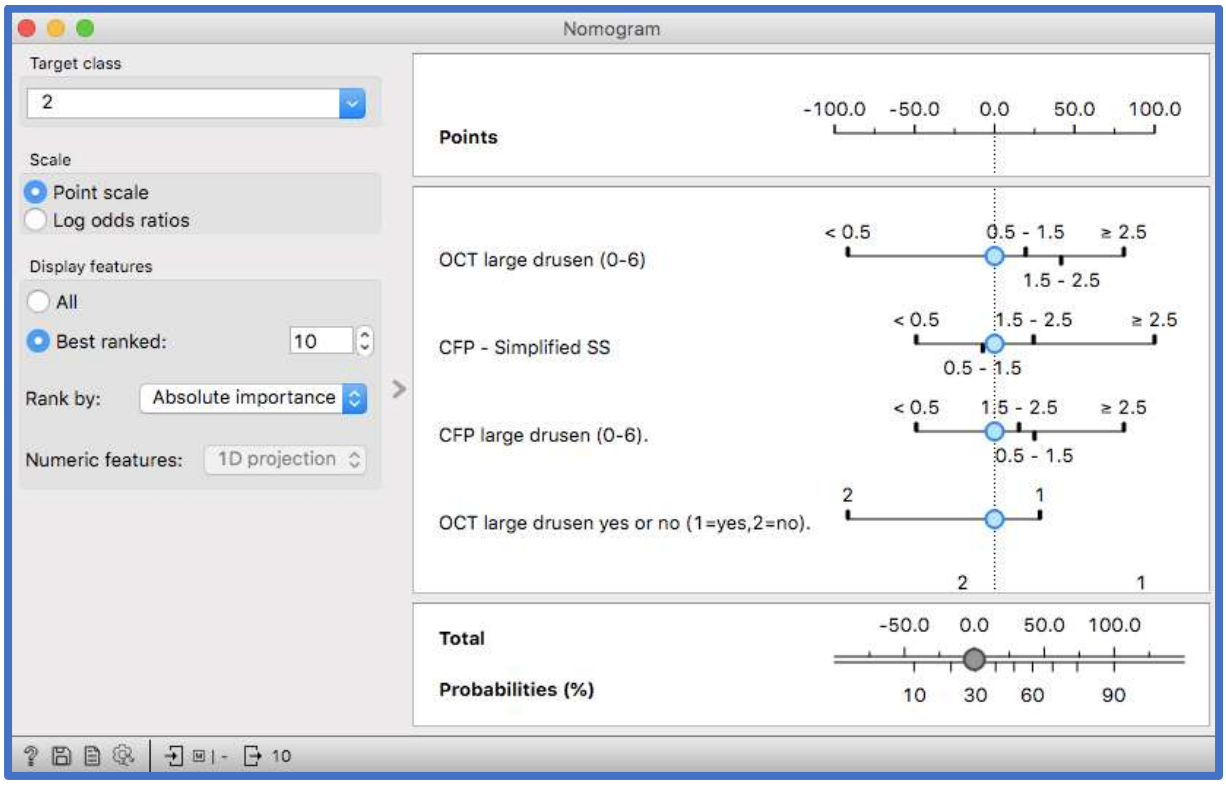

Figure 2.22 Illustration of the Orange Data Mining Nomogram widget output for the probability of an abnormal FAF (Target class 2).

13. Receiver Operating Characteristic (ROC) Analysis widget

Originally designed for operatives of military radars in the early 1950s,(168) the ROC curve represents the discriminative ability of a binary classifier for a variety of discriminative thresholds. Below in Figure 2.23 is an illustration of the ROC Analysis widget. The graph plots the false positive (FP) rate (1- specificity) on the x-axis against true positive (TP) rate (sensitivity) on the y-axis for a single model learner at a variety of thresholds, i.e. the model learner is determining the *probability* than an observation belongs to a specific class, and this threshold probability can be varied to produce the ROC curve. The closer the curve is to the left border the better the model learner. The prediction class must be selected, e.g. in this study the choice was the prediction of a normal or abnormal FAF result. Note that there is no third “Average over classes” result as offered by the Test and Score widget. The curve averaging options are Merge predictions from folds (selected by default for this study), Mean TP rate, Mean TP and FP at threshold, and show individual curves. Finally, the solid black performance line indicates the iso-performance line in the ROC space so that all the points on the line have the same false positive and false negative cost.

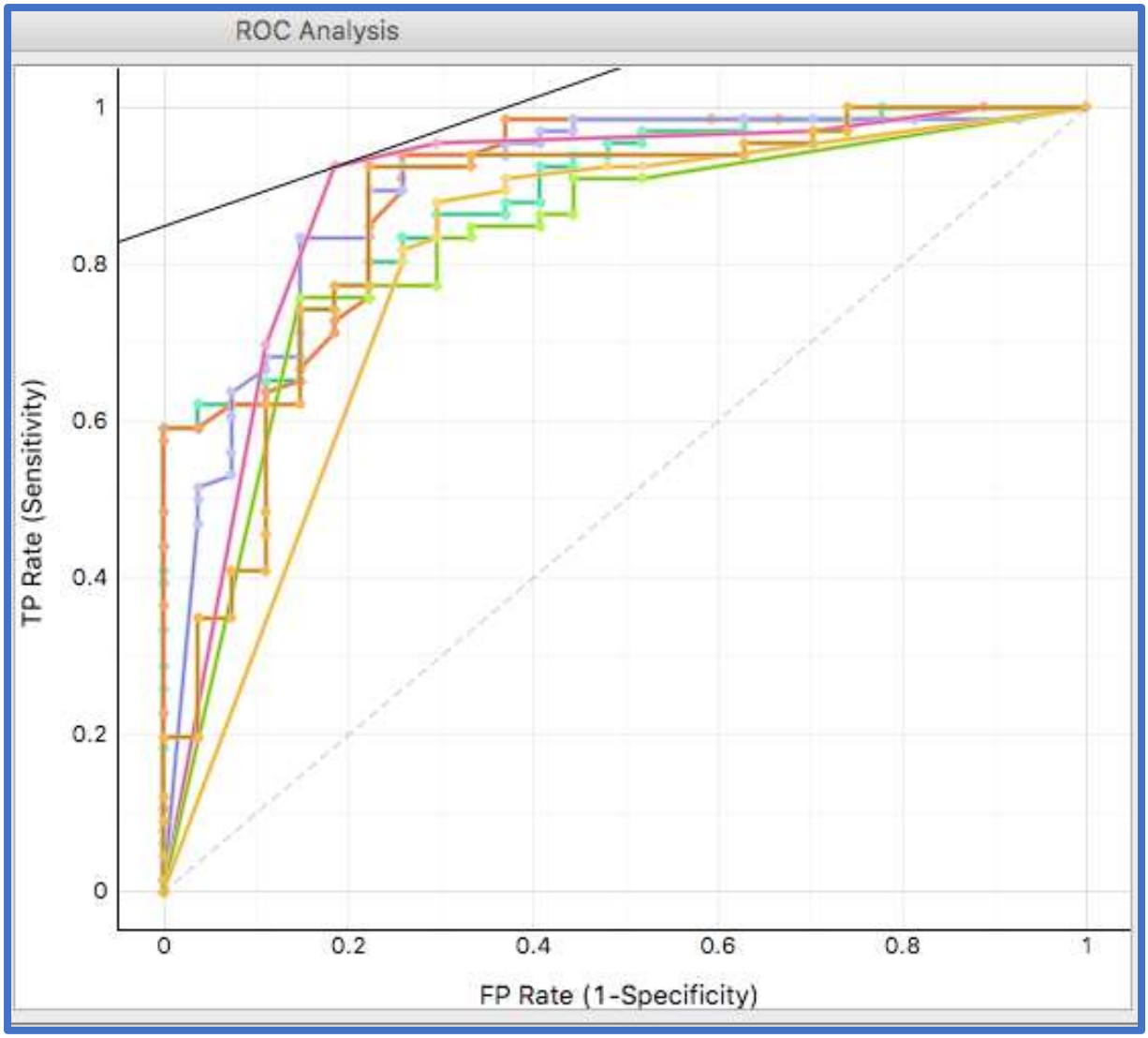

Figure 2.23 Illustration of the Orange Data Mining ROC analysis widget, for predicting a normal FAF result by a variety of model learners. The solid black line represents the performance line, the dashed black line represents a model with no predictive ability.

Additional Orange widgets were used for unsupervised and supervised machine learning workflows for image analysis in Chapter seven on image analysis as follows:

14. Import images and Image embedding widgets

Images are uploaded via the Import images widget to the Image embedding widget and deep learning models are used to create a vector of numbers for each image. These vectors are outputted in the form of a data table with additional columns. For this study, the images (CFP, FAF and OCT) were uploaded. Embedding passes the image through an existing deep network to calculate its representation as vectors. Orange contains a choice of embedders including Google's InceptionV3[6] - a convolutional 48-layered neural network

trained on 1.2 million images from ImageNet. ImageNet contains a diverse range of images, including e.g. everyday real-life objects.(169) Only one embedder, however, SqueezeNet, which is also trained on ImageNet and features an 18-layered neural network, can be used locally on the user's own PC, offering a more rapid vector computation compared to others, as well as maintaining privacy as the images are never uploaded to a remote server (unlike Google's Inception V3[6]).(169) SqueezeNet offers AlexNet-level precision with 50 times fewer parameters, and is the default embedder used by Orange. In recent research, no major differences were noticed between the performance accuracy of SqueezeNet and the other more complex embedders (including Google's Inception V3[6]) in four case studies.(169)

15. Distances widget

This widget creates a matrix by computing the distances between rows and columns in a dataset. Column-wise normalisation ensures equal treatment of individual features. Groups in the data can be identified by feeding the matrix data into the Hierarchical Clustering widget, Distance Map or Distance Matrix widget which can all be used to visualise the data, and finally the Multidimensional Scaling (MDS) widget can map the distances.
Once computation of the distances between either rows or columns has been selected, the user can choose the specific Distance Metric – for this study the Cosine Metric was selected. Cosine is the angle between two vectors of an inner product space – Orange computes the cosine distance, defined as 1- the similarity.
Distances between rows (i.e. instances), looks for groups of instances within the data, whilst the distances between columns describe similarities between features.

16. Hierarchical clustering widget

This widget creates a dendrogram (a tree-like structure) from the matrix of distances calculated by the Distances widget. The dendrogram can be pruned, and clusters can be selected via one of three methods: Manual, Height and Top N. The results can also be fed into the Boxplot widget for further visualisation and analysis.

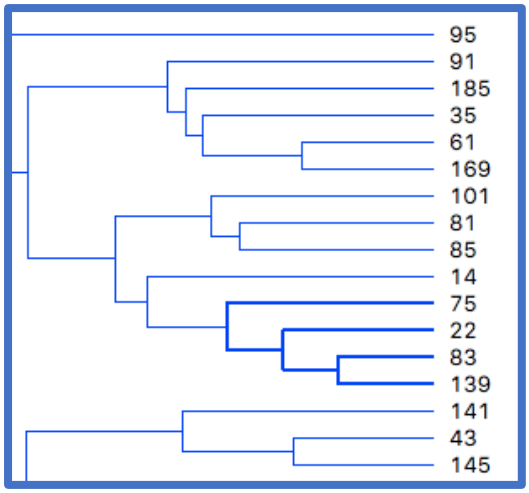

Figure 2.24 Illustration of the output from the Orange Data Mining Hierarchical clustering widget

17. Multidimensional scaling widget

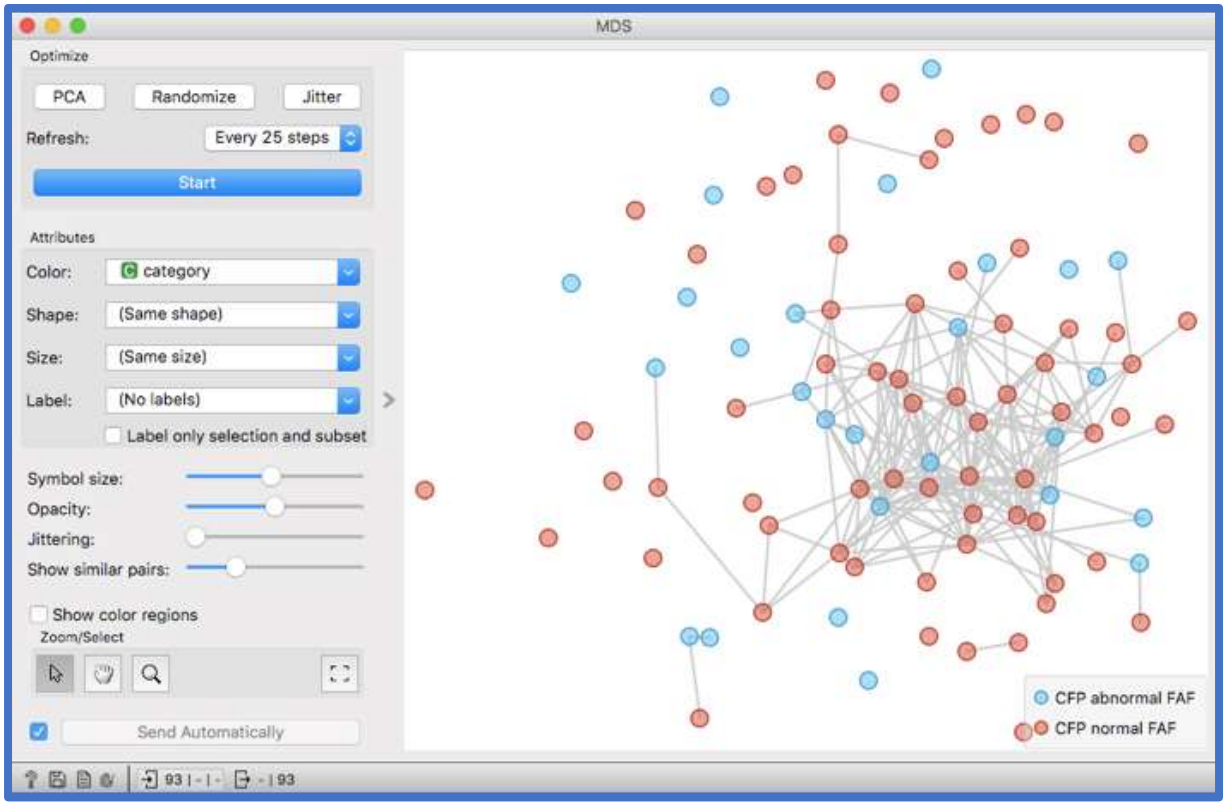

Figure 2.25 Illustration of the output from the Orange Data Mining MDS widget.

The MDS widget attempts to project points in a low-dimensional fashion, as effectively as possible, from a matrix of distances or a dataset, into a 2-dimensional space. A perfect fit is usually impossible as the data is high-dimensional. PCA sets the initial points along the principle coordinate axes. Randomise allows the initial points to be positioned at random before readjustment. Jitter adjusts the points to eliminate overlapping, whilst Refresh allows

selection of how often the visual mapping is refreshed. This can be at every iteration, never, or at every 5, 10, 25 and 50 steps. How the points are visualised can also be selected within the Attributes section, with the options of colour, shape, size and label.

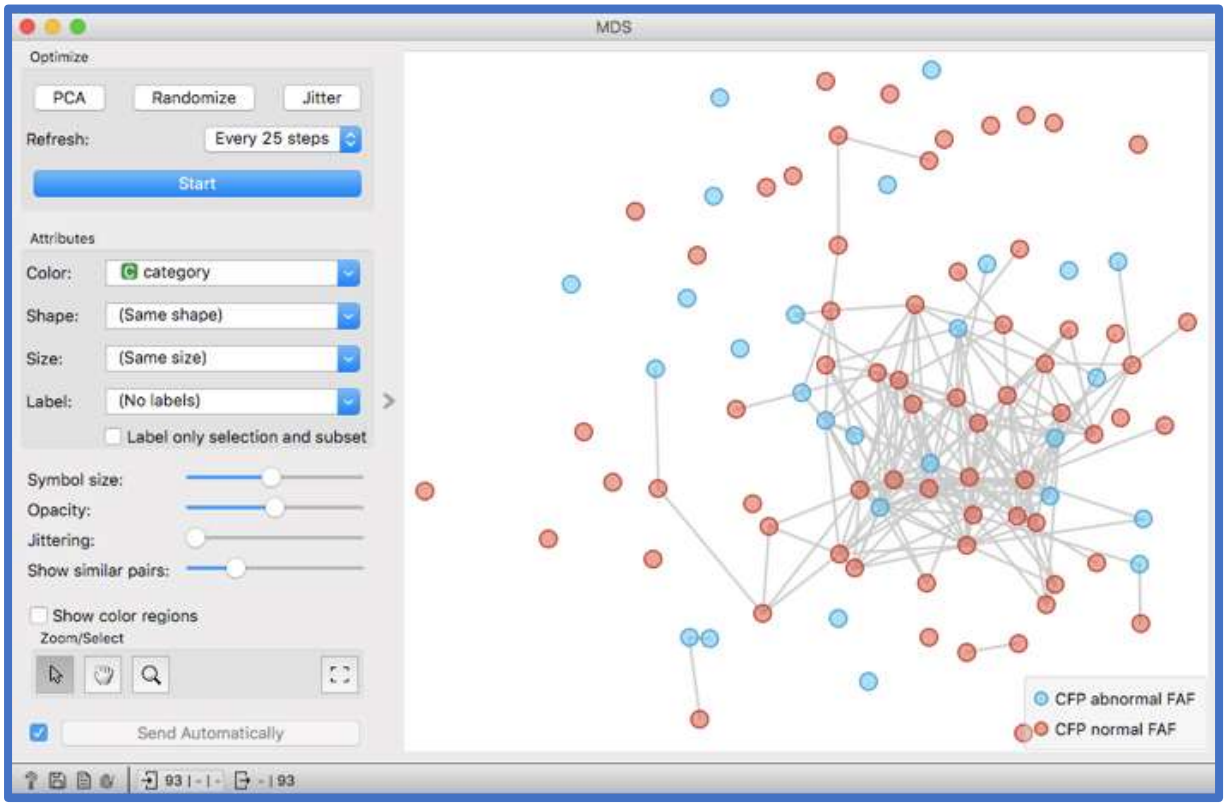

Figure 2.26 Illustration of the output from the Orange Data Mining MDS widget for a normal and abnormal FAF result using CFP.

2.11 Normal distribution tests for data collected in the study

Parametric statistical tests were employed throughout this study wherever possible, however, parametric tests assume that data is normally distributed. To ensure that this was the case, the normality of the data was tested from several different aspects using the following methods via SPSS software:

Standard errors of skew and kurtosis
Frequency distribution histograms with fitted normal curves
Normality tests (Kolmogorov-Smirnov test)
Q-Q plots
Boxplots

Where data was considered not to be normally distributed, an appropriate non-parametric test was performed in addition to the parametric test and the results from both tests reported.

2.12 Sample size calculations

Sample size for the Chapter three study on pupillary dilation was calculated using previously reported data regarding macular degeneration lesion size measured using FAF imaging.(170) Using these published mean and standard deviation data, an effect size of 2mm² was used to determine a sample size of 15 in order to allow a power of 0.80 and a 0.05 alpha error level when assessing clinical repeatability and reproducibility of the FAF imaging process in practice. The first 25 subjects were selected for the Chapter three study to ensure that an adequate sample had been chosen to achieve statistical significance. The reliability data produced during the Chapter three study informed the sample size for the subsequent study Chapters (covered in Chapters four to nine).

2.12.1 Sample size calculations for Chapters four to nine based on the results from Chapter three.

For the sample size required for Chapters four to nine of the study, the results from grader RS in Chapter three on FAF image quality, before and after dilation, were used to calculate the Odds ratio and Proportion of Discordant pairs. This data was then used within GPower to estimate the sample size, as shown below.

For grader RS		Before dilation	Before dilation
		acceptable	not-acceptable
After dilation	acceptable	18 (a)	6 (b)
After dilation	not-acceptable	0 (c)	1 (d)

Table 2.3 Agreement of FAF image grading before and after pharmacologically induced pupil dilation for grader RS.

Odds ratio calculated from Table 2.4

= a*d/b*c

= 18*1/6*0

= infinity, therefore replace c with 1 to enable a meaningful calculation

= 18*1/6*1

= 18/6

= 3.0

Proportion of discordant pairs calculated from Table 2.4

= (6 + 0)/(18+ 1)

= 6/19

= 0.3158

Using GPower version 6, with an “Exact” test family, a McNemar Test, a two tailed test, an odds ratio of 3.0, an alpha error value of 0.05, a power of 0.80 and the proportion of discordant pairs of 0.3158, gives a sample size of 95.

However, Orange also provides a facility to calculate whether the quantity of data collected is sufficient by tracking the informedness via the Test and Score widget. This was performed once the study had almost reached the required sample size as indicated by GPower, i.e. as the number of instances neared 95, by the following procedure: Firstly, the concordance statistic (C), which is a measure of goodness of fit for binary outcomes with the Logistic Regression learner, was varied to give an optimum trade-off between a low C with low complexity (tendency to underfit), and a high C with high complexity (tendency to overfit), i.e. a compromise between bias and variance respectively. The results for "Test on train" and "10-fold cross-validation" informedness were matched as closely as possible to reach the trade-off. Below are graphs of how the informedness varies with the use of data from CFP alone (Figure 2.27) and from combined CFP and OCT (Figure 2.28) for a variety of C values using Logistic Regression for both "Test on train" (orange line) and "10-fold cross-validation" (grey line). The optimum C value was then selected as the best trade-off between overfitting and underfitting as shown below. No optimisation of the C value was possible for the Naïve Bayes learner, as this model cannot be hyperparameter fine-tuned.

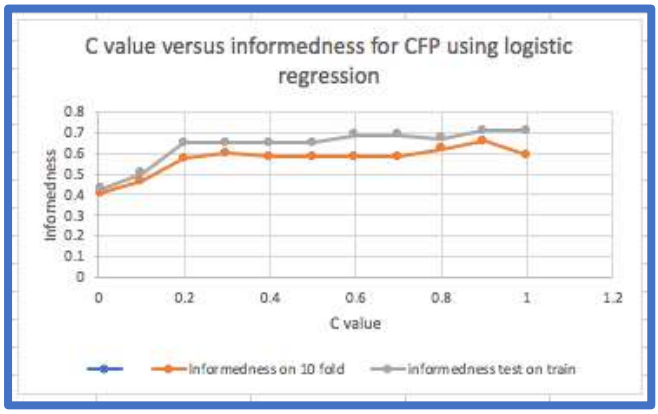

Figure 2.27 C versus informedness for CFP alone using Logistic Regression for both 10-fold cross validation (orange line) and Test on train (grey line). The best trade-off was for a C value of 0.30.

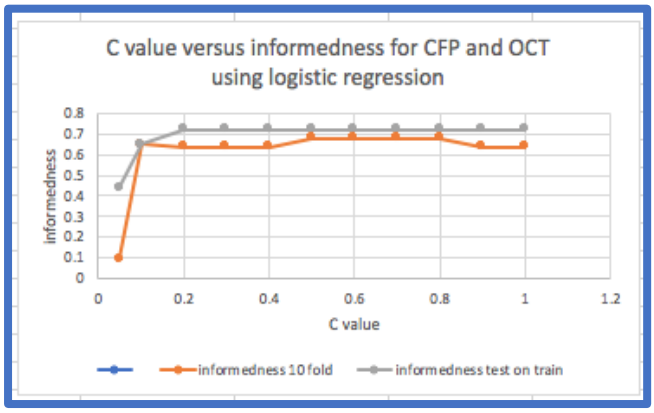

Figure 2.28 C versus informedness for CFP and OCT combined using Logistic Regression for both 10-fold cross validation (orange line) and test on train (grey line). The best trade-off was for a C value of 0.50.

Figures 2.29 below shows graphically how the level of informedness achieved varies with the quantity of data used for hyperparameter fine-tuned Logistic Regression and Naïve Bayes learners, for CFP alone, for predicting FAF outcomes. We can see that the level of informedness levels off at approximately 40% of the data collected for Naïve Bayes, i.e. at approximately 37 patients, and for Logistic Regression at approximately 80% of the data collected, i.e. at approximately 74 patients.

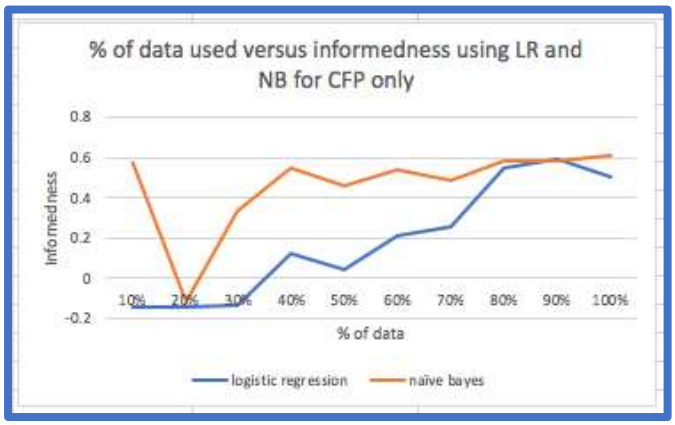

Figure 2.29 Percentage of data used versus informedness using Logistic Regression (blue line) and Naïve Bayes (orange line) for CFP data only.

In Figure 2.30 below for combined CFP and OCT, this levelling off occurs at approximately 20% of the data for both model learners, i.e. at 19 patients. After 93 cases it was concluded that enough data had been collected by virtue of the levelling off of the value of informedness achieved for both learners, for CFP alone and for CFP and OCT combined, and so data collection was concluded at this point.

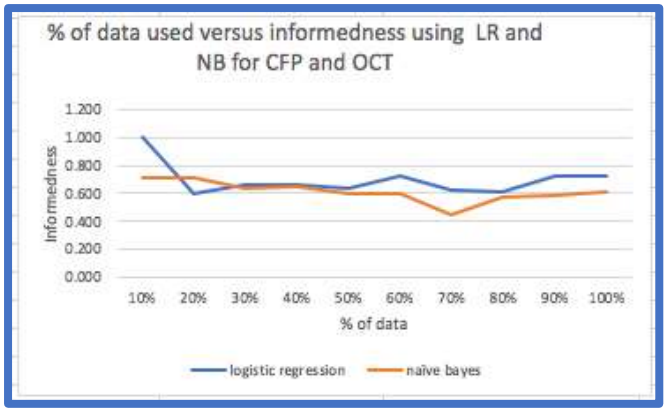

Figure 2.30 Percentage of data used versus informedness using Logistic Regression (blue line) and Naïve Bayes (orange line) for CFP and OCT data combined.

These results indicate that the addition of OCT to CFP allows maximum informedness to be achieved with less data. We can also see that the level of informedness achieved with CFP alone is superior with Naïve Bayes when to compared to Logistic Regression. Overall, the optimum informedness was achieved with hyperparameter fine-tuned Logistic Regression when information from both CFP and OCT was utilised.

2.12.2 Summary of sample size calculations

Hyperparameter fine-tuned Logistic Regression and Naïve Bayes AI machine learners were used to calculate informedness gained for various quantities of the study data. When increasing the amount of data no longer improved the degree of informedness, i.e. the graph of percentage of data used versus informedness levelled off, for both CFP alone and for CFP and OCT combined, then the quantity of data collected was considered to be sufficient and data collection was concluded. This levelling off had occurred by the collection of 93 cases, and hence data collection was stopped at this point.

2.13 Personnel

All patients were examined by the primary researcher Roger Smyth, optometrist at Andrew and Rogers Optometrists, Beverley. Roger Smyth is also a Principal Optometrist, employed by the NHS at the Hull University Teaching Hospitals NHS Trust. Part of his role involves working as a grader in the treat and extend injection service for the management of ARMD, retinal vein occlusions and diabetic retinopathy. RS has his work quality controlled on a regular basis, in addition to attending regular team meetings including feedback from consultant ophthalmologists on grading methods and protocols.

2.14 Scope and boundaries

Only patients who fulfilled the inclusion and exclusion criteria as detailed above, were included in this study, and all patients were gathered from one optometry practice, Andrew and Rogers Optometrists, Beverley, in The East Riding of Yorkshire.

2.15 Research data management

The information from clinical tests was entered into a computer spreadsheet at the earliest opportunity. The computer storing the information was backed up weekly to a remote hard drive and the backup disc was also stored off the practice premises in a locked cupboard.

2.16 Project timetable

The project began as soon as ethical approval was granted. The data collection for the study ran from February 2020 until February 2021.

2.17 Risk assessment

1. Data is lost due to computer failure – action: back up weekly and keep hard drive off the practice premises in a locked cupboard.

2. Project not completed in time – action: extra time built in to allow for this and the project supervisor regularly updated on the study progress (monthly).

3. OCT fails – Action: this is an integral part of the business's strategy and would be replaced by a new device/repaired as soon as possible if this occurred.

2.18 Resources and costs

Extra appointments were required for some of the subjects, but this small extra cost was absorbed by the business.

2.19 Place where the research work was undertaken

Andrew and Rogers Optometrists
15 North Bar Within
Beverley
East Riding of Yorkshire
HU17 8AP

Chapter 3 The impact of pharmacologically induced pupillary dilation on the quality and clinical usefulness of FAF imaging

3.1 Introduction

Useful FAF imaging relies on the quality of the image obtained, enabling the clinician to identify subtle abnormalities that may be present. The manufacturers of the "Nidek Retinascan Duo RS-330" recommend pupillary dilation (by verbal discussion between the distributors of the instrument, Birmingham Optical Limited, and the primary researcher of this project (RS)) for FAF image acquisition, however, this advice is not based on the result of formal research. A study from 2018 examined the need for pupillary dilation in fluorescence lifetime imaging ophthalmoscopy (FLIO), a technique that measures in vivo autofluorescence intensity decay of endogenous fluorophores in the retina over time. The paper concluded that FLIO has a longer image acquisition time without dilation, and therefore recommended pupillary dilation for all images taken with this technique.(171) Whilst FLIO is not identical to the technique of fundus camera-based FAF imaging conducted in this study, it may be appropriate to infer that pupillary dilation would also have a positive effect on the quality of the FAF images obtained.

The work in Chapter three was therefore conducted to identify whether pharmacologically induced pupillary dilation would lead to a significant improvement in FAF image quality. This was considered to be important, as pupillary dilation may affect patients' willingness to participate in the study due to the extra medical procedures required, as well as causing practical problems for patients e.g. longer time taken to complete the examination, sensitivity to light and the inability to drive for a period following the procedure. Avoiding pupillary dilation would also offer the benefit of avoiding the risk of precipitating an acute angle closure event, or transient intraocular pressure spikes in susceptible individuals. If pupillary dilation prior to FAF could be avoided, this may therefore offer significant advantages to both the clinician and the patient.

3.2 Methods

Chapter three was initially carried out to ascertain whether there was a statistically significant difference in the quality of FAF images collected before and after pharmacological pupillary dilation. 25 eyes from the first 25 study subjects were imaged before and after pupil dilation, giving a total of 50 images. The eye which was judged to have the most advanced changes related to ARMD was selected as the study eye.

The FAF images were taken before, and after pupillary dilation with one drop of guttae Tropicamide 1% instilled into the lower fornix of the study eye. Participants were screened prior to dilation to ensure that they were not at risk of acute angle closure as described previously.

The 50 images collected for Chapter three pre and post-dilation were then graded by two optometrists, the first, the project supervisor HB, a Reader in Optometry at Aston University, and the second, the student conducting the research RS, a retinal grader in the ARMD treat and extend service at The University of Hull Teaching Hospitals NHS Trust and the primary researcher for this study. The 50 images were allocated a randomly generated three-digit number and then ordered numerically, hence producing a random order of presentation to the graders. The images were graded according to a three-point scale as “good”, “acceptable” and “not acceptable”. This was considered to be appropriate after previous studies on the grading of retinal images were considered as discussed in section 3.2.1 below.

3.2.1 Creating a grading scale for FAF image quality

In a study conducted in 2003, dilated and un-dilated retinal images taken with digital photographic equipment were compared for the purposes of improving community diabetic screening initiatives. Three levels were used for the grading of image quality; “fully”, “partially” and “not accessible”. These three levels were based on a) whether small blood vessels were visible in the retinal temporal arcades with “reasonable clarity”, b) whether large blood vessels were visible in the retinal temporal arcades with “reasonable clarity”, and c) whether large blood vessels in the retinal temporal arcades were blurred or >1/3 of the total picture was blurred.(172)

A later study in 2009 looked at developing an automated method of evaluating the quality of digital retinal photographs. This study utilised a four-point scale, namely: “very good”, “good”, “acceptable” and “not acceptable”, viewed by six independent observers (three ophthalmologists and three experienced ophthalmic nurses). No significant difference was found between the rating of the “very good “ and “good” images among these observers, and therefore the grading was simplified into three categories of “good”, “borderline” and “unacceptable”.(173) Unfortunately, the study did not specify how the observers were instructed to draw boundaries between these grades.

Another more recent study published in 2020 considered the imaging quality delivered by a non-mydriatic camera for the purposes of a telemedical approach to retinal screening of disease graded by two experienced and masked retinal graders. Image quality was again evaluated on a three-point scale: “excellent”, “good” and “poor”. This study based the grading on areas of blur. Grading was “excellent” if blurred areas comprised less that 25% of the image, “good” if blurred areas were present in 25-50% of the image, and “poor” if the image was blurred in more than 50% of the of image.(174)

In summary, three studies, which considered digital retinal photography for the detection of diabetic retinopathy and general retinal disease concluded that a three-point scale, graded by at least two experienced observers, was adequate for the purposes of recording fundus image quality. The gradings were based on the visibility of small and large retinal blood vessels in the retinal temporal arcades and the percentage of the retinal image that is not in shadow/blurred. For the purposes of the current study, an amalgam of these grading boundaries was formulated with the two optometrists HB and RS advised to grade the quality of the FAF images on a three-point scale, “good”, “acceptable” and “not acceptable”. See Table 3.1 below.

	Over 75% of image not in shadow/blurred	**50-75% of image not in shadow/blurred**	**Less that 50% of image not in shadow/blurred**
Small blood vessels visible	good	acceptable	not acceptable
Large blood vessels visible	acceptable	acceptable	not acceptable
No blood vessels visible	not acceptable	not acceptable	not acceptable

Table 3.1 Preliminary study guidance for graders of FAF images

50 FAF images (25 taken prior to, and 25 after pupillary dilation), were taken from the first 25 patients enrolled in the study, and graded by the two optometrists, HB and RS independently, without reference to each other’s work or any collaboration apart from a virtual meeting to initially discuss the grading system prior to the images being shared. An Intraclass Kappa test was performed to calculate, firstly, if there was a significant difference between the grading performed by the two optometrists. Secondly, a McNemar test was performed to calculate whether pupillary dilation made a significant difference between the quality of FAF images taken prior to and post pupillary dilation. The results Chapter three guided the methods of all subsequent FAF image collection for the purposes of the rest of the study (for Chapters four to nine).

FAF images before and after pharmacological pupillary dilation from the first 25 participants were analysed by graders HB and RS as described previously in Chapter two.

Intraclass Kappa value and McNemar statistical tests were performed manually and via IBM’s SPSS statistics software (Version 28).

3.3 Results

Sex	Male	40%
	Female	60%
Age		71.60 ± 9.0 years
Gullstrand refractive error		-1.79 ± 2.90 dioptres
ARMD status (based on ARMD definition by Bird et al(11))	None	48%
	Early	44%
	Late	8%
Smoking status	Ex-smoker	12%
	Non-smoker	88%
Family history of ARMD	No	80%
	Yes	20%
Cataract type present	NUC	68%
	COR	12%
	PSC	8%
Lens status	Phakic	68%
	Pseudophakic	32%

Table 3.2 Demographic features of the first 25 participants within the study, used for the statistics on the effect of pupillary dilation on FAF image quality.

For grader HB	Before dilation	After dilation
good	9	18
acceptable	6	3
not acceptable	10	4
Total	25	25

Table 3.3 Summary of the first 25 participants FAF images quality for grader HB before and after pupil dilation

For grader RS	Before dilation	After dilation
good	11	21
acceptable	7	3
Not acceptable	7	1
Total	25	25

Table 3.4 Summary of first 25 participants FAF images quality for grader RS before and after pupil dilation

For grader HB	Before dilation	After dilation
acceptable	15	21
not acceptable	10	4
Total	25	25

Table 3.5 Summary of the first 25 participants FAF images quality for grader HB before and after dilation expressed dichotomously

For grader RS	Before dilation	After dilation
acceptable	18	24
not acceptable	7	1
Total	25	25

Table 3.6 Summary of the first 25 participants FAF images quality for grader RS before and after dilation expressed dichotomously

Summary of results for the comparison of FAF image quality grading by HB and RS:

For details of the statistical calculations see Appendix 1

The Intraclass Kappa value (95% Confidence Range 0.3752-0.7880) indicated minimal to moderate agreement between the two graders, however Kappa may be influenced by the proportion of positive and negative cases in the sample. To overcome this problem, positive and negative agreement (PA and NA respectively) were also calculated. Using the lowest values, (within the 95% confidence limits) from two methods for the calculation of PA and NA found that PA had at least a strong agreement (0.9642), and NA at least a moderate agreement (0.8411) for the comparison of image quality grading between HB and RS.

Summary of results for the comparison of FAF image quality before and after pupillary dilation calculated individually for both HB and RS:

For details of the statistical calculations see Appendix 2.

The McNemar test indicated, to the $p < 5\%$ level ($p = 0.03125$ exact p-value binomial test), that there is a significant difference in the quality of image when comparing FAF performed before and after pupillary dilation, with the images captured after dilation being of a significantly higher quality. The proportion of images that were acceptable (expressed dichotomously) prior to dilation and after dilation for grader HB were 60% and 84% respectively, whereas for grader RS these were 72% and 96% respectively. In total the proportions for both graders were 66% acceptable prior to dilation and 90% acceptable after dilation.

Below are some examples of FAF images taken during this study, before and after dilation.

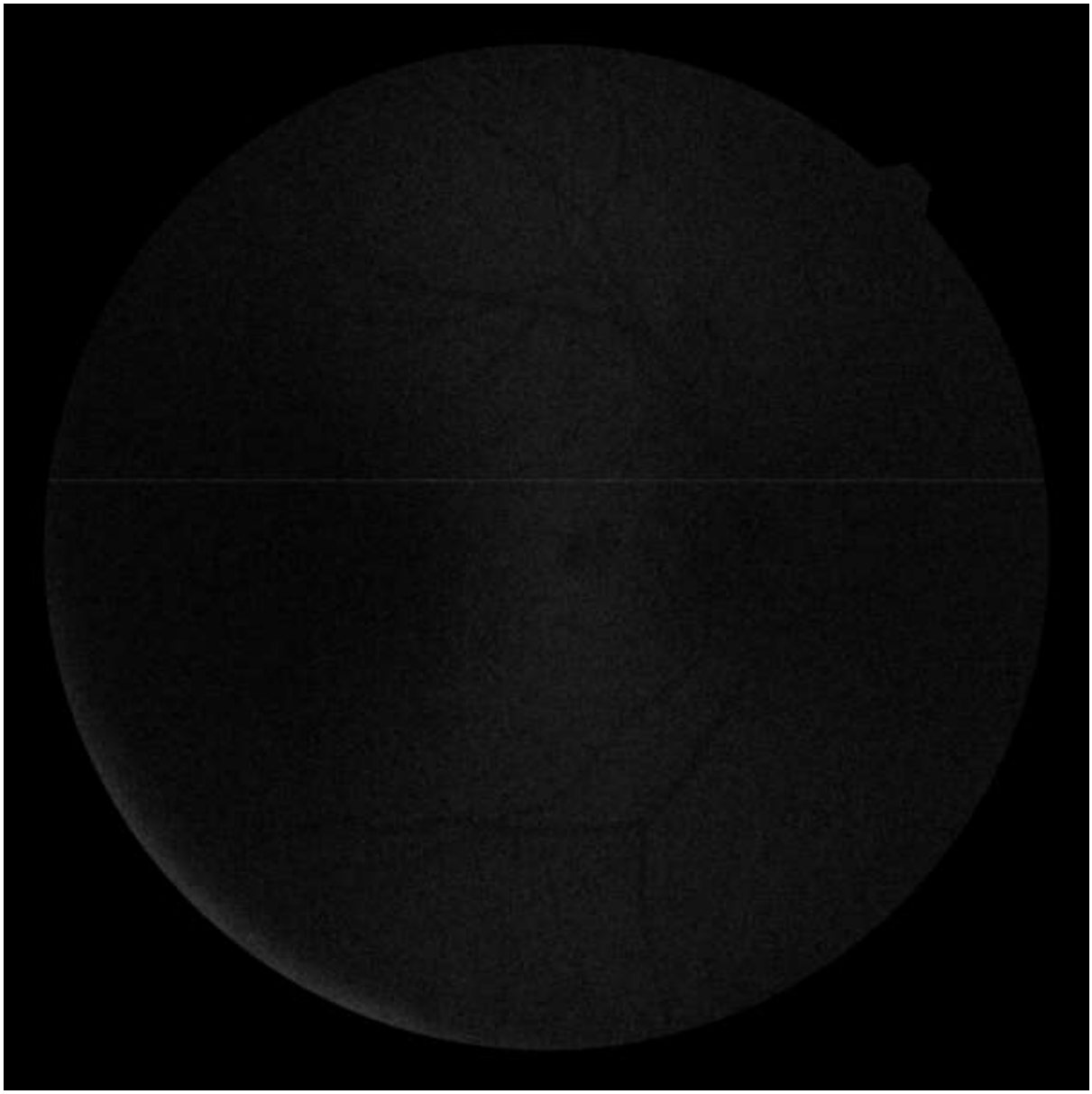

Figure 3.1 Example of a not acceptable FAF image taken prior to pupillary dilation.

Figure 3.2 Example of a “good” FAF image taken from the same patient as in Figure 3.1, after pupillary dilation

Figure 3.3 Example of an acceptable FAF image from a second participant, taken prior to pupillary dilation

Figure 3.4 Example of a “good” FAF image, taken from the same patient as Figure 3.3, after pupillary dilation

3.4 Discussion

From the results of Chapter three, based on positive and negative agreement, there was at least moderate agreement between the grading of FAF images from the first 25 participants in the study as “good”, “acceptable” and “not acceptable” between the graders HB and RS. RS was responsible for grading all further images recorded in this study.

There was also a significant difference in the quality of image when comparing FAF performed before and after pupillary dilation, with the images captured after dilation being of a significantly higher quality according to the results of the McNemar test performed. From Tables 3.3 and 3.4 there were almost twice as many “good” images collected after pupillary dilation compared to before dilation, and only five “not acceptable” images after dilation compared to 17 before dilation. From Tables 3.5 and 3.6, when the quality of images is expressed dichotomously (i.e. “good” and “acceptable” = “acceptable”, and only “not acceptable” = “not acceptable”) the number of “acceptable” images increases by over 36% post pupillary dilation, whilst the number of “not acceptable” images falls by over 70%. For both graders combined the percentage of “acceptable” FAF images prior to dilation was 66%, rising to 90% post dilation. This result indicated that all FAF images taken following Chapter 3 of the study should be captured after pharmacologically induced pupillary dilation, as dilation is likely to significantly enhance FAF image quality.

3.5 Conclusion

According to the McNemar statistical test comparing the quality of FAF images before and after pupil dilation for the first 25 participants of this study for both graders HB and RS, it can be concluded that pupil dilation makes a significant improvement to the quality of FAF images obtained.

Chapter 4 The impact of cataract on the quality of fundus autofluorescence imaging

4.1 Introduction

Cataract is a leading cause of visual impairment worldwide(175) and in 2017 accounted for 65.2 million cases of visual impairment globally.(176) Artificial intelligence (AI) has been utilised widely within the field of ophthalmology, however, development in the field of AI for cataract classification is still relatively unexplored.(177) Whilst AI classification of cataract is beyond the scope of this study, ODM software was utilised in Chapter four to explore the relationship between FAF image quality and the type of and severity of cataract present.

It has been previously reported that structures anterior to the retina naturally emit autofluorescence which can cause interference to FAF imaging, and thus adversely affect image quality. The crystalline lens has a natural excitation peak of 420-430nm, with an emission peak of 520nm,(19) with cataracts increasing light absorption and scatter by the lens, leading to poor contrast FAF images.(20) Below is a reminder of the grading boundaries for FAF image quality developed for this study.

4.2 Methods

4.2.1 Grading of the quality of FAF images

	Over 75% of image not in shadow/blurred	**50-75% of image not in shadow/blurred**	**Less that 50% of image not in shadow/blurred**
Small blood vessels visible	good	acceptable	not acceptable
Large blood vessels visible	acceptable	acceptable	not acceptable
No blood vessels visible	not acceptable	not acceptable	not acceptable

Table 4.1 A reminder of the method by which FAF image quality was graded for this project.

4.2.2 World Health Organisation Simplified Cataract Grading System

The Simplified Cataract Grading system used in this project was based on that published in 2002 by Thylefors et al.(178) A panel of experts developed the system with the aim of producing a scale for grading the most common forms of cataract that could easily be used by relatively inexperienced observers with minimal training, by means of slit-lamp examination. The scale was evaluated in hospitals at four different sites, with very good to fair interobserver agreement. It is recommended that a new observer should be trained via 20 initial cases containing various stages of cataract severity, in collaboration with a more senior and experienced mentor, acting as a “gold standard”. For this project the period of training was carried out prior to data collection in the primary researcher’s (RS’s) place of work, the Hull University Teaching Hospitals NHS Trust.
It is recommended that the slit-lamp is kept in good condition, with a bulb that is fully functional. The slit-lamp in primary researcher’s practice in Beverley is well maintained to this end, and is serviced annually.

The most common forms of cataract included in the grading system are:

Nuclear sclerotic cataract (NUC): a gradual opacification of the nucleus of the lens.
Cortical cataract (COR): opacities involving the cortex of the lens with typical wedge-shaped spokes.
Posterior subcapsular cataract (PSC): distinct opacities centrally or para-centrally on the posterior capsule.

It is recommended that the slit-lamp examination is made with 10X magnification, a beam width of 0.1mm, a height tall enough to cover the grading zone, and an incident illumination angle of 45°.

Mydriasis is recommended to be carried out at least 20 minutes prior to examination, and that a drop of anaesthetic is instilled initially to increase the absorption of subsequent drugs, followed by a drop of 0.5% Tropicamide and a drop of 2.5% Phenylephrine. A pupil size of at least 6.5mm should be aimed for to enable a complete examination. For this project one drop of 1.0% Tropicamide alone was used prior to waiting for 20 minutes for dilation to occur. This decision was made in order to limit the period of dilation, allowing participants to drive later in the day if required. Also, RS has observed, through over 25 years of optometry locum work in the local area, that the majority of optical practices only carry one type of dilating drop in stock, that being 1% Tropicamide. This would therefore make the findings of

this project practically meaningful, and the investigations more repeatable in a local high street optometric setting. The median pupil size obtained during the current project was 6.0mm.

4.2.3 Nuclear Sclerotic Cataract grading

In NUC cataract there are two principal processes. Clouding and colouration (or browning) of the lens. For this grading system, only the clouding aspect should be considered.

The “grading” zone is defined as having its anterior and posterior borders limited by the anterior and posterior nuclear shells. Between these shells lie the anterior embryonic nucleus (AEN) and the posterior embryonic nucleus (PEN). Between these four structures are three clear zones, the largest being the central clear zone.

The grading of NUC cataract should follow these recommendations:

0. NUC cataract considered less than grade 1. (Note, however, for this study, grading was started from 0.1, on a decimal scale from 0.1-3.0 and any increase in the visibility of the AEN or PEN/opaqueness of the central zone was recorded).
1. Clinically significant NUC cataract (the AEN and PEN are more visible than normal, however, the central zone is still easily/entirely visible).
2. Moderately advanced NUC cataract (the nuclear zone is uniformly opaque with the central clear zone not clearly visible and the red reflex reduced).
3. Sufficiently advanced to consider surgery (the nuclear zone is densely opaque and the red reflex is dull).

The observer compares the level of cataract by reference to standard photographs and grades the cataract appropriately. If there is severe corneal opacification, advanced cortical changes obscuring the nucleus, or if there is a Morgagnian cataract then the NUC cataract is considered to be “ungradable” and the cataract grading level of 9 should be recorded.

4.2.4 Cortical Cataract grading

For COR cataract a relatively short and broadly focussed beam should be used, positioned between three and nine o'clock of the pupil border. Sharply defined cortical opacities seen via retro-illumination are graded. The beam should be swept around so that all of the lens can be examined within the red reflex. The circumferential extent of the opacities is added to give a total figure for grading purposes.

The grading of COR cataract should follow these recommendations:

0. COR cataract covering <1/8th of the lens circumference. (Note, however, for this study, grading was started from 0.1, on a decimal scale from 0.1-3.0 with any element of COR cataract recorded, i.e. one tenth of an 1/8th was recorded as 0.1).
1. COR cataract covering ≥1/8th and <1/4 of the lens circumference.
2. COR cataract covering ≥1/4 and <1/2 of the lens circumference.
3. COR cataract covering ≥1/2 of the lens circumference.

Cortical opacities do not need to extend to the periphery of the lens to be included in the circumferential grading, but are extrapolated and included. Again, ungradable cases should be recorded as grade 9 as with NUC cataract.

4.2.5 Posterior Subcapsular Cataract grading

PSC cataract is seen in retro-illumination, with only the portion of cataract in focus at the level of the posterior capsule included in the grading. PSC cataract is graded according to its vertical diameter, read off the slit-lamp beam height graticule. PSC is typically centred near the posterior pole and may extend towards the lens equator. PSC cataracts generally have a "feathered" appearance compared to the sharp definition observed with a COR cataract.

The grading of PSC cataract should follow these recommendations:

0. Vertical dimension of <1mm. (Note, however, for this study, grading was started from 0.1, on a decimal scale from 0.1-3.0 and any element of PSC cataract recorded, i.e. a height of 0.1mm was graded as 0.1, 0.2mm as 0.2 etc).
1. Vertical dimension ≥1mm<2mm.

2. Vertical dimension ≥2mm<3mm.
3. Vertical dimension ≥3mm. (May require surgery).

If the red retro-illumination reflex cannot be visualised due to advanced NUC, COR or corneal opacities, then the cataract should be graded as 9 as with NUC and COR cataract.

All three types of cataract can be graded decimally, on a scale of 0.0-3.0 in 0.1 steps. This decimal system of grading was utilised throughout the current study.

4.3 Results

4.3.1 Patient demographics

Sex	Male	40%
	Female	60%
Age		71.59 ± 10.2 years
Gullstrand refractive error		-1.21 ± 2.95 dioptres
ARMD status (based on ARMD definition by Bird et al(11))	None	40%
	Early	52%
	Late	8%
Smoking status	Ex-smoker	40%
	Non-smoker	60%
Range of pack years		0-50 years
Mean pack years		4.94 ± 8.89 years
Family history of ARMD	No	73%
	Yes	27%
Symptoms of ARMD	No	5%
	Yes	95%
Cataract type present	NUC	71%
	COR	15%
	PSC	8%
NUC cataract	≤ grade 1.0	41 participants (43%)
	1.1 to 2.0	25 (26%)
	2.1 to 3.0	1 (1%)
COR cataract	≤ grade 1.0	3 (3%)
	1.1 to 2.0	9 (10%)
	2.1 to 3.0	2 (2%)
PSC cataract	≤ grade 1.0	1 (1%)
	1.1 to 2.0	3 (3%)
	2.1 to 3.0	3 (3%)
Lens status	Phakic	72%
	Pseudophakic	28%
Acceptable images	First definition	59%
	Second definition	41%
Non acceptable images	First definition	90%
	Second definition	10%
FAF images	Normal	71%
	Abnormal	29%

Table 4.2 Demographic features including all 93 participants within the study

4.3.2 Examples of each stage of cataract

Figure 4.1 Example of FAF image taken with cataract graded as NUC 0.5, COR 0.0 and PSC 0.0

Figure 4.2 Example of FAF image taken with a cataract graded as NUC 1.0, COR 0.0 and PSC 0.0

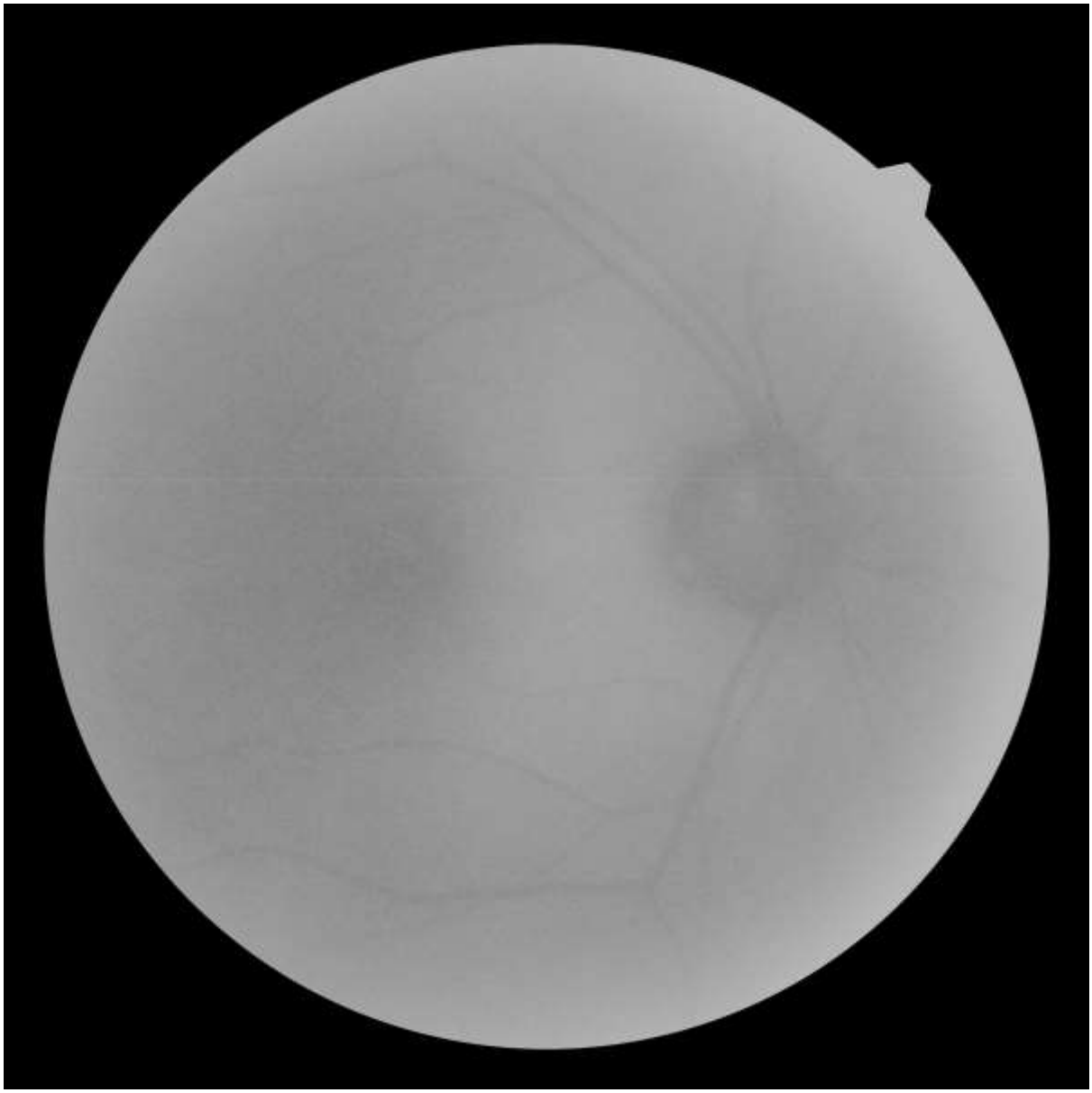

Figure 4.3 Example of a FAF image taken with a cataract grade of NUC 2.0, 0.0 COR and 0.0 PSC

Figure 4.4 Example of a FAF image taken with a cataract grade of NUC 0.2, COR 0.0 and PSC 1.2

Figure 4.5 Example of a FAF image taken with a cataract grade of NUC 0.3, COR 2.0 and PSC 0.0

A Chi squared test was initially used to test whether being pseudophakic gave rise to a significantly higher proportion of acceptable FAF images. The data used for this analysis is shown in table 4.3 below:

Image quality	Phakic	Pseudophakic	Totals
Good	36	19	55
Acceptable	22	7	29
Poor	9	0	9
Totals	67	26	93

Table 4.3 Proportions of good, acceptable and poor images for phakic and pseudophakic participants within the study

Chi squared = 4.89, degrees of freedom = 2, $p = 0.087$, indicating that p was not significant to the $p < 0.05$ level. However, as one value was < 5, the results may not be reliable. This indicates that there may not be a significant difference between the proportions of good, acceptable and poor images in phakic and pseudophakic participants within the study.

The following two sections explore, in more detail, the effect that cataract has on the quality of FAF images. For these sections dichotomous grading systems were created to facilitate statistical exploration of the results via ODM and SPSS.

4.3.3 First analysis

The first analysis is concerned with high FAF image quality, and considers what level of cataract grading will still allow good FAF images to be captured, i.e. when small blood vessels are visible and over 75% of the image is not in shadow/blurred. For this analysis the dichotomous grading system was as follows: only good images are graded as good, with acceptable and not acceptable images graded together as not acceptable.

4.3.4 Second analysis

The second analysis is concerned with low FAF image quality, and considers what level of cataract grading will cause such a poor image quality as to render FAF imaging clinically not acceptable. The assumption was made that if a FAF image was graded as not acceptable, i.e. < 50% of the image was not in shadow/blurred, and/or no blood vessels were visible, then the image was not clinically useful. For this analysis the dichotomous grading system

was as follows: good and acceptable images are graded together as good, and only not acceptable images are graded as not acceptable.

	First analysis concerned with <u>high</u> image quality	**Second analysis concerned with <u>low</u> image quality**
good image quality	good images	good and acceptable images
not acceptable image quality	acceptable and not acceptable images	not acceptable images

Table 4.4 The dichotomous grading system facilitating analysis via Orange Data Mining and SPSS.

4.3.5 Results of the first analysis

The results are illustrated in Figure 4.6 below. Out of the 93 images taken for this study, 55 were classified as good by the definition stipulated for the first analysis. An eye with a good FAF image quality had a mean nuclear sclerotic (NUC) cataract score of 0.487±0.50, whilst eyes with a not acceptable image quality had a mean score of 0.924±0.62. The median (indicated by the vertical yellow line in the Boxplot) for a good quality FAF image was 0.30 and for a not acceptable quality image the median was 1.0. The Student's t test found that there was a significant difference at the 1% ($p=0.001$) level between the mean NUC WHO scale scores of participants with a good and not acceptable FAF image quality. Overall, there were 66 participants in the study graded as having NUC cataract out of the total sample of 93 participants (71.0%).

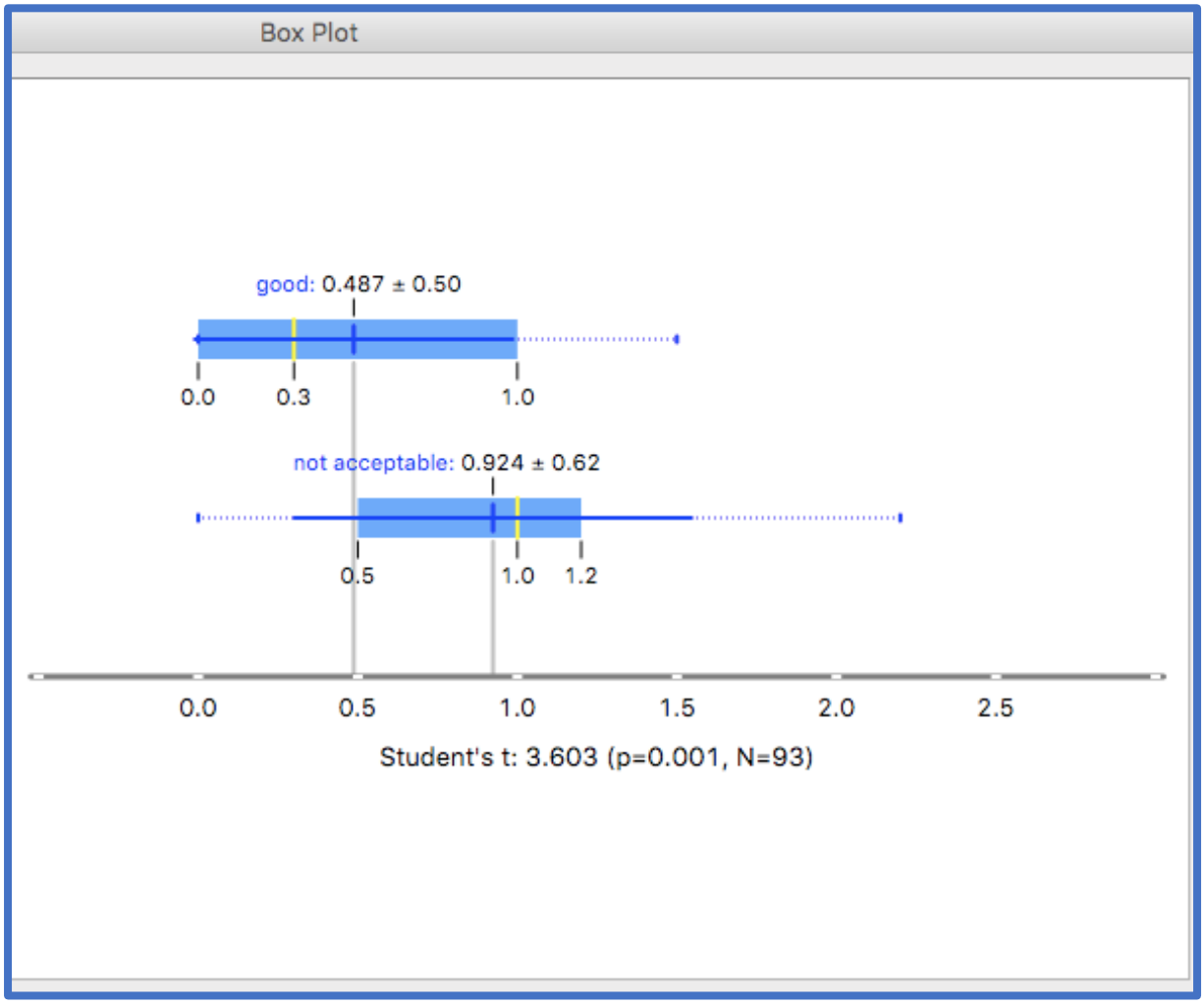

Figure 4.6 First analysis Boxplot of FAF Image quality and NUC cataract.

Boxplot above from ODM showing how FAF image quality is related to NUC cataract expressed on the WHO scale of cataract grading scale (0.0–3.0) along the bottom. Only good images were graded as good, with acceptable and not acceptable images graded as not acceptable to create a dichotomous scale.

Below in Figure 4.7 is from the Distributions widget output for NUC cataract grading for image quality graded as good and not acceptable for the first analysis. At approximately a WHO scale point of 1.0 for NUC cataract, there is a 50% chance of the FAF image quality being not acceptable. There were no instances of a good FAF image quality once the NUC WHO score was > 1.5. Note that 35 of the 66 participants (53.0%) of the participants graded as having NUC cataract were graded as having a good FAF image quality. If the WHO score is > 1.5 there appears to be an approximately 85% probability of the FAF image not being of high quality.

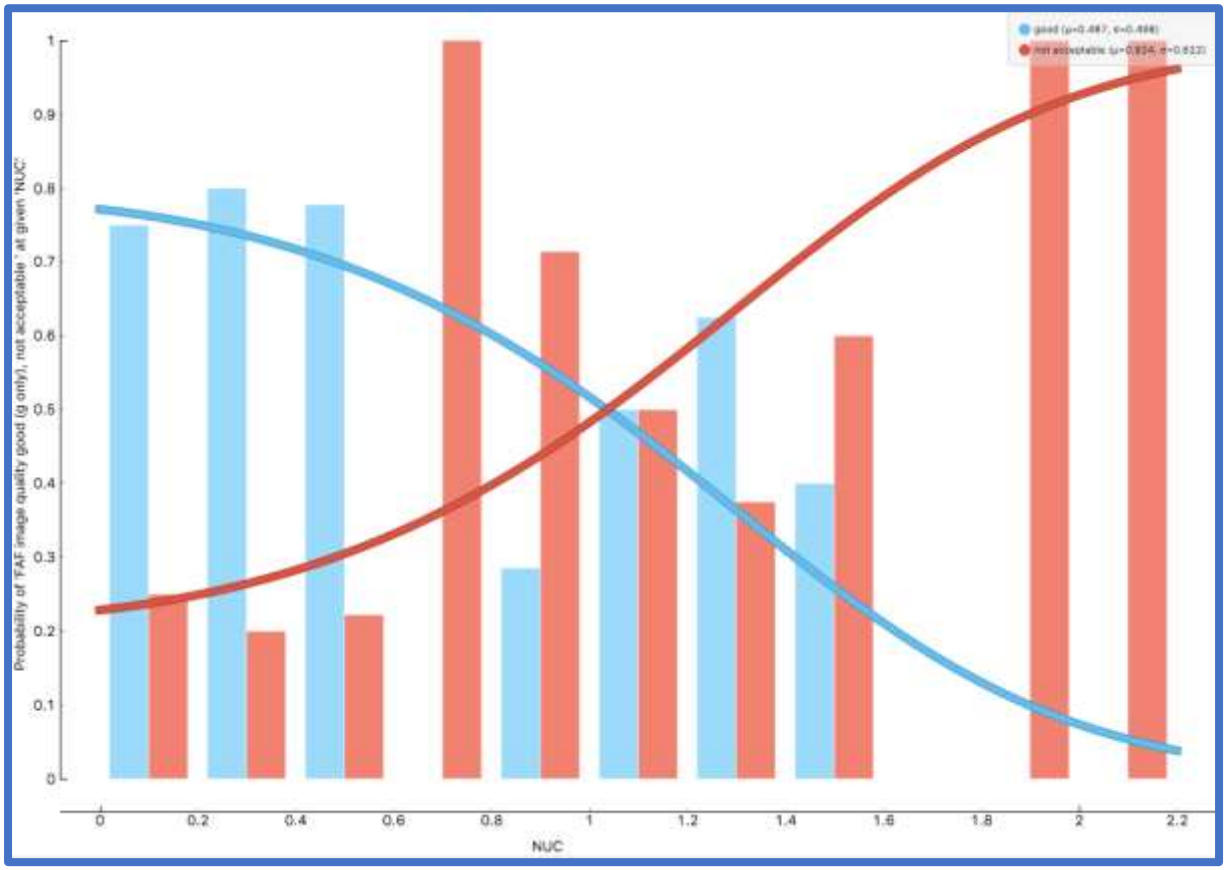

Figure 4.7 First analysis Distribution of FAF image quality and NUC cataract.

Distribution widget output for NUC cataract on the WHO scale versus the probability of a good or not acceptable FAF image (blue bars represent good image quality and red bars represent not acceptable image quality). Only good images were graded as good, with acceptable and not acceptable images graded as not acceptable to create a dichotomous scale.

Below in Figure 4.8 is a similar exploration for cortical crystalline lens opacities (COR). This time the Student's t test was significant at the 5% (p=0.022) level when comparing the mean WHO scores. (This finding was supported by non-parametric testing (Mann Whitney U test giving a significance value of 0.003 as the normality of the COR data was not perfect). Images with a good FAF image quality had a mean grading on the WHO scale of 0.1±0.43, and for not acceptable images the mean was 0.403±0.71. This time the median for both good and not acceptable images was 0.0. This is most likely due to there being a large proportion of the sample having no COR cataract - only 14 participants out of the total sample of 93 participants (15.1%).

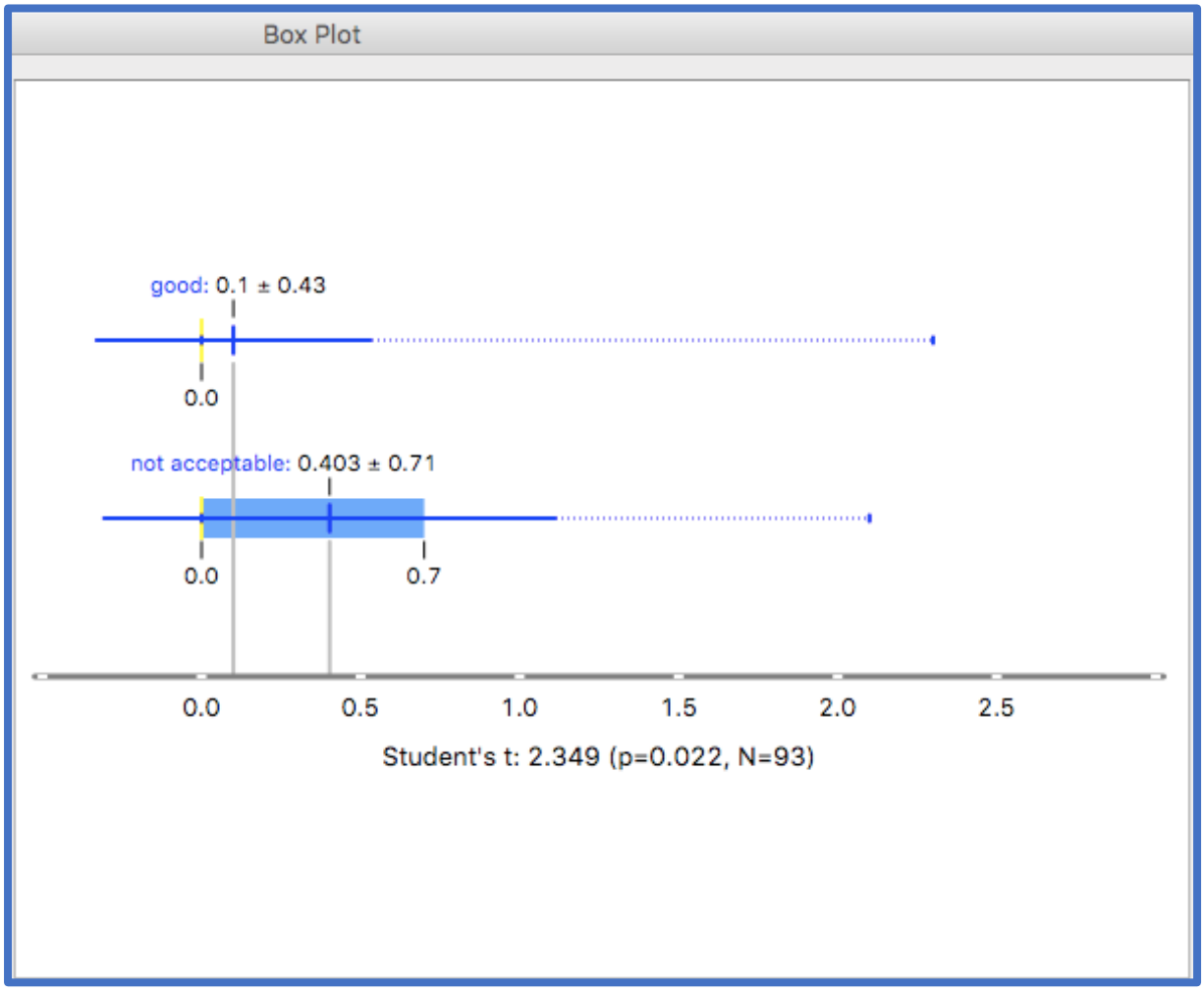

Figure 4.8 First analysis Boxplot of FAF image quality and COR cataract.

Boxplot from ODM showing how FAF image quality is related to COR cataract expressed on the WHO scale of cataract grading scale (0.0-3.0) along the bottom. Only good images were

graded as good, with acceptable and not acceptable images graded as not acceptable to create a dichotomous scale.

Below in Figure 4.9 is from the Distributions widget output for the same comparison of COR cataract grading for image quality graded as good and not acceptable for the first analysis. At a WHO scale point of approximately 0.70 for COR cataract, there was approximately a 50% chance of the FAF image being not acceptable. There was one case of a good FAF image being recorded with a COR cataract graded as 2.3 on the WHO scale. Note that 3 of the 14 participants (21.4%) with COR cataract were graded as having a good FAF image quality.

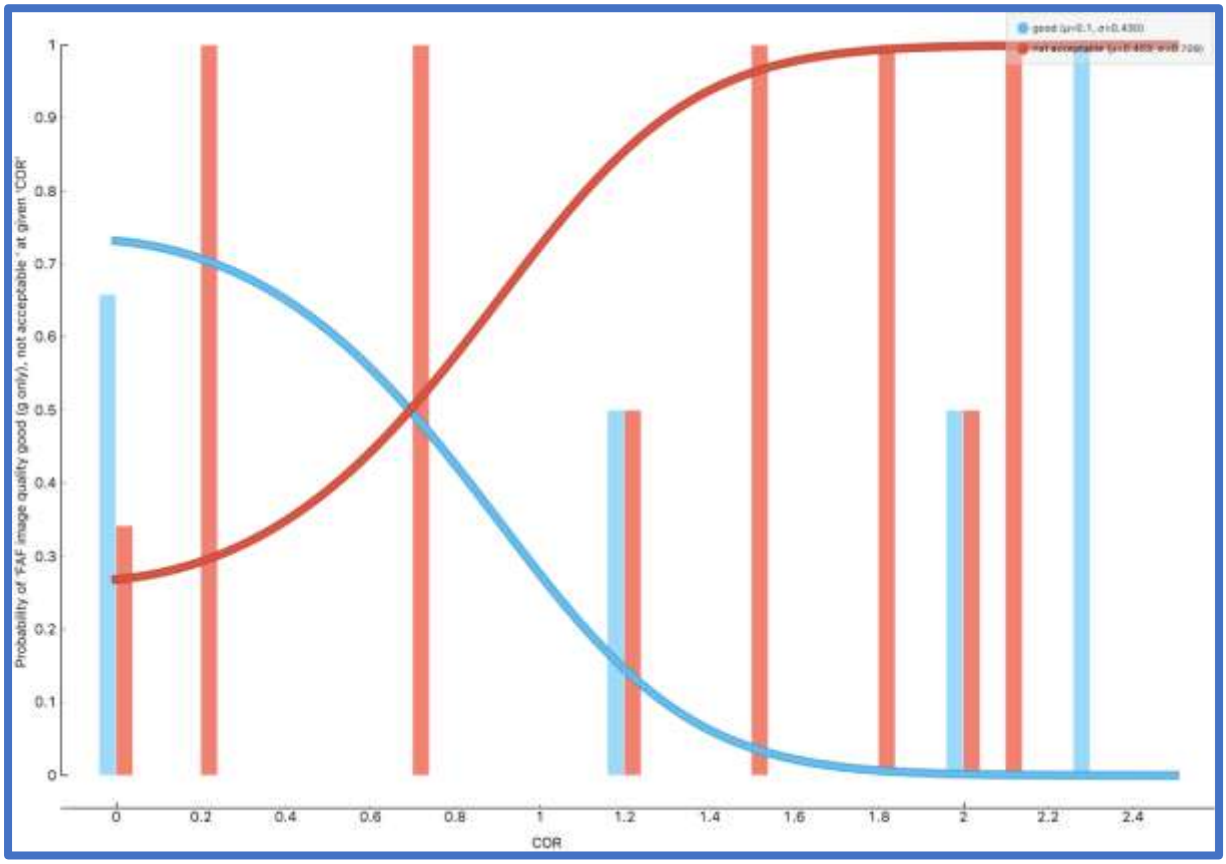

Figure 4.9 First analysis Distribution of FAF image quality and COR cataract.

Distribution widget output for COR cataract on the WHO scale versus the probability of a good or not acceptable FAF image (blue bars represent good image quality and red bars represent not acceptable image quality). Only good images were graded as good, with acceptable and not acceptable images graded as not acceptable to create a dichotomous scale.

Finally, in Figure 4.10 below is the Boxplot output for posterior subcapsular cataract (PSC). The Student's t test is significant to the 5% (p=0.026) level, with good images having a mean grading of 0.02±0.16, and not acceptable images having a grading of 0.255±0.64 (This finding was again supported by the Mann Whitney U test giving a significance value of 0.011 as the normality of the PSC data was not perfect). As with COR cataract, the median for both good and not acceptable FAF images is 0.0. Again, as with COR cataract, this is most likely due to there being relatively few cases of participants with PSC cataract (7 out of the total sample of 93 participants (7.5%)).

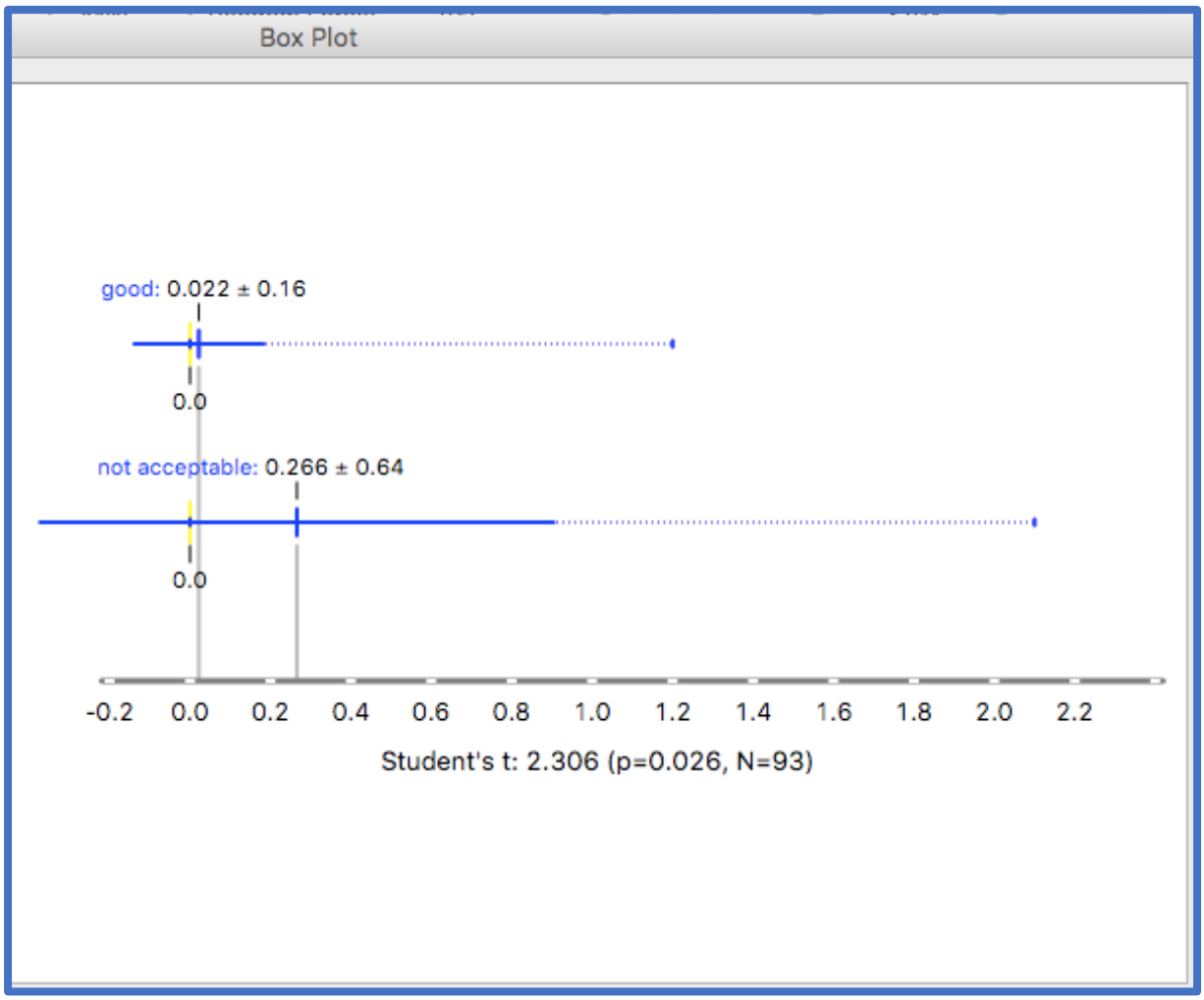

Figure 4.10 First analysis Boxplot of FAF image quality and PSC cataract.

Boxplot above from ODM showing how FAF image quality is related to PSC cataract expressed on the WHO scale of cataract grading scale (0.0-3.0) along the bottom. Only good images were graded as good, with acceptable and not acceptable images graded as not acceptable to create a dichotomous scale.

Below in Figure 4.11 is from the Distribution widget output for the same comparison of PSC cataract grading for FAF image quality graded as good and not acceptable for the first analysis. At approximately a WHO scale point of 0.30 for PSC cataract, there is approximately a 50% chance of the image being graded as not acceptable. There were no instances of a good quality FAF image once the PSC WHO score was over 1.2. Note that only one of the 7 participants (14.3%) who had PSC cataract were graded as having a good FAF image quality.

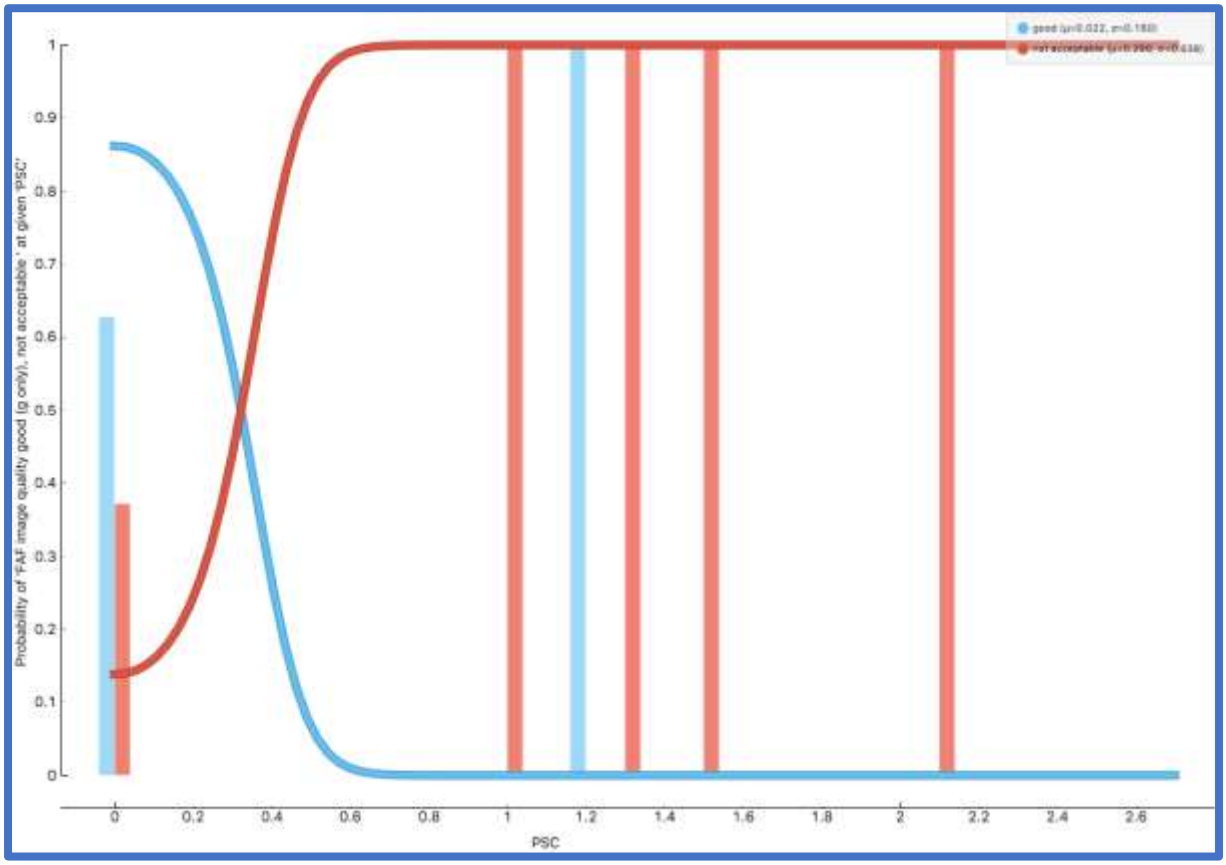

Figure 4.11 First analysis Distribution of FAF image quality and PSC cataract

Distribution widget output for PSC cataract on the WHO scale versus the probability of a good or not acceptable FAF image (blue bars represent good image quality and red bars represent not acceptable image quality). Only good images were graded as good with acceptable and not acceptable images graded as not acceptable to create a dichotomous scale.

Distribution widget outputs

Table 4.5 below shows how the mean cataract scores for NUC, COR and PSC cataract graded on the WHO scale vary with FAF images graded as good and not acceptable, by the definitions set out in the first analysis.

Cataract type	Mean cataract score for good FAF image	Mean cataract score for not acceptable FAF image	Upper limit within 1 SD of the mean for good image	Lower limit of 1 SD of the mean for good image	Upper limit within 1 SD of the mean for not acceptable image	Lower limit of 1 SD of the mean for not acceptable image
NUC	0.487	0.924	0.987	0.000	1.544	0.304
COR	0.100	0.403	0.530	0.000	1.113	0.000
PSC	0.022	0.255	0.182	0.000	0.895	0.000

Table 4.5 First analysis summary of results from the ODM Distribution and Boxplot widgets.

From the results of the Boxplot and Distribution widgets in the first analysis it is possible to conclude that cataract of all three types examined (NUC, COR and PSC) do have a negative effect on FAF image quality which is statistically significant.

For the first analysis, further statistical tests were carried out to analyse the differences between the means for the different types of cataract, i.e. NUC, COR and PSC by way of an ANOVA, the results are shown below:

IBM's SPSS Statistics (Version 28) was used to perform a within-subjects (as all participants were graded for all three types of cataract) repeated measures one-way ANOVA (parametric statistical test for three or more dependent groups) to compare the level of grading of the three different types of cataract among images defined to have a not acceptable FAF image, in participants with some form of cataract (31 eyes in total (note that the majority of the eyes had more than one type of cataract)).

The hypotheses:
Participants with not acceptable FAF images have significantly different levels of the three types of cataract studied, i.e. nuclear sclerotic (NUC), cortical (COR) and posterior subcapsular (PSC).

Descriptive statistics suggest that NUC cataract has a higher mean grading, i.e. NUC cataracts were worse than COR and PSC in the participants of this study who were graded as having a not acceptable FAF image.

Assuming sphericity
The Within subjects repeated measures one-way ANOVA assumes sphericity, i.e.
a) the variances of all the levels of the within subject factors are equal
b) the correlation between all within subject factors are equal
Mauchly's Test of Sphericity determines whether this assumption has been violated. The p-value (Sig.) is 0.304 and therefore indicates the assumption has not been violated.

Results: Tests of within-subjects effects with sphericity assumed: F(12.517), the degrees of freedom for cataracts and error(cataract) (2,60), the p-value of (Sig., <0.001) and partial eta squared (0.294).

Pairwise comparisons are relevant because the type of cataract had a statistically significant effect on the level of cataract grading.

After applying Bonferroni corrections for multiple comparisons, the following statistically significant differences arose:

Nuclear sclerotic cataract (1) had a higher mean value for not acceptable FAF images than Cortical cataract (2) (p=0.001).
Nuclear sclerotic cataract (1) had a higher mean value for not acceptable FAF images than Posterior subcapsular cataract (p=0.000) (3).

Determining the effect size: the effect in the form of partial eta squared is taken from the "Tests of Within-Subjects Effects" table, since sphericity has not been violated (incidentally, if sphericity had been violated, then the "Multivariate Tests" table would have been used).

Eta squared	Effect size
0.01	Small
0.06	Medium
0.14	Large

Table 4.6 Eta squared values with corresponding effect size, according to Cohen(179).
In this case partial eta squared is = 0.294, so the effect was large.

These findings support the hypothesis that participants with cataract and not acceptable FAF images (31 in total, 7 had not acceptable image quality with no cataract) had significantly different levels of the three types of cataract studied, i.e. NUC, COR and PSC cataract, with NUC cataract having a higher grading level than both COR and PSC cataract. There was no significant difference found between the grading means of COR and PSC cataract.

However, with the number of NUC cataracts in the sample being significantly greater than that of COR and PSC cataracts, a larger sample would be required to ascertain whether the three different types of cataract affect the FAF image quality to a greater or lesser degree. The ANOVA was not carried out in the second analysis due to there being only nine not acceptable FAF images by the definition in the second analysis.

How could the effect size be used to inform future studies?

GPower was used as shown in Figure 4.8 below to calculate the sample size required for future studies as shown below. The sample size indicated is 22, which is lower than the sample size used for this analysis (where 31 not acceptable FAF images were used for this analysis).

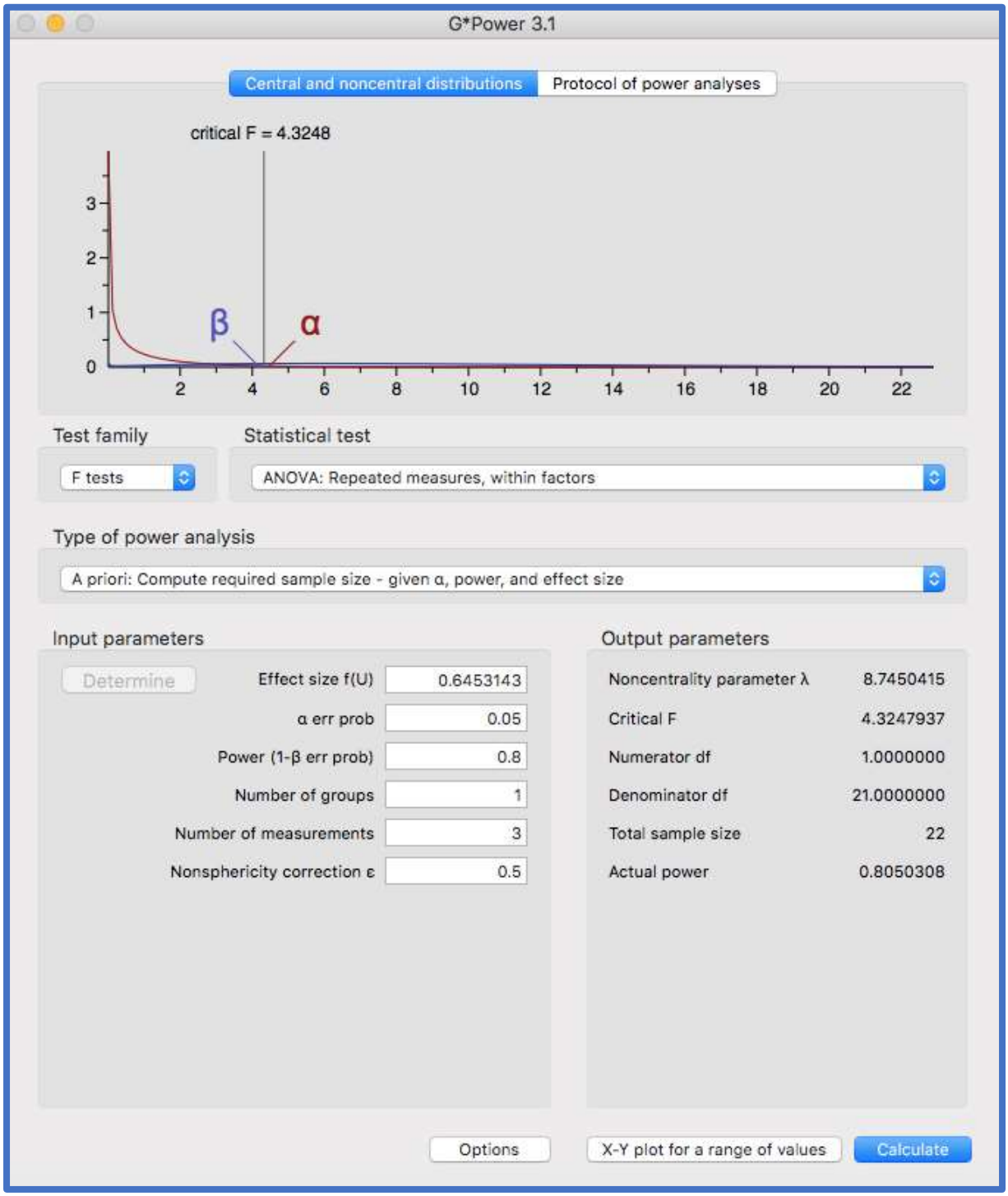

Figure 4.12 GPower sample size calculation

4.3.6 Results of the second analysis

In the second analysis, if the FAF image was graded as good or acceptable quality it was recorded as good, whilst only not acceptable images were recorded as not acceptable, the assumption being that if a FAF image was graded as not acceptable i.e. < 50% of the image was not in shadow, and/or no blood vessels were visible, then the image was not clinically useful. Out of the 93 images taken for this study, only nine were classified as not acceptable by the definition stipulated for the second analysis. The results are shown below. For not acceptable images defined as not clinically useful, the mean NUC score was 1.344±0.57 and a median of 1.2. For clinically useful images the mean NUC cataract score was 0.593±0.55, and a median of 0.5. The Student's t test found that there was a significant difference at the 1% ($p=0.004$) level between the mean NUC WHO scale scores of participants with a good and not acceptable FAF image quality. As a reminder, there were 66 participants in the study graded as having NUC cataract out of the total sample of 93 participants.

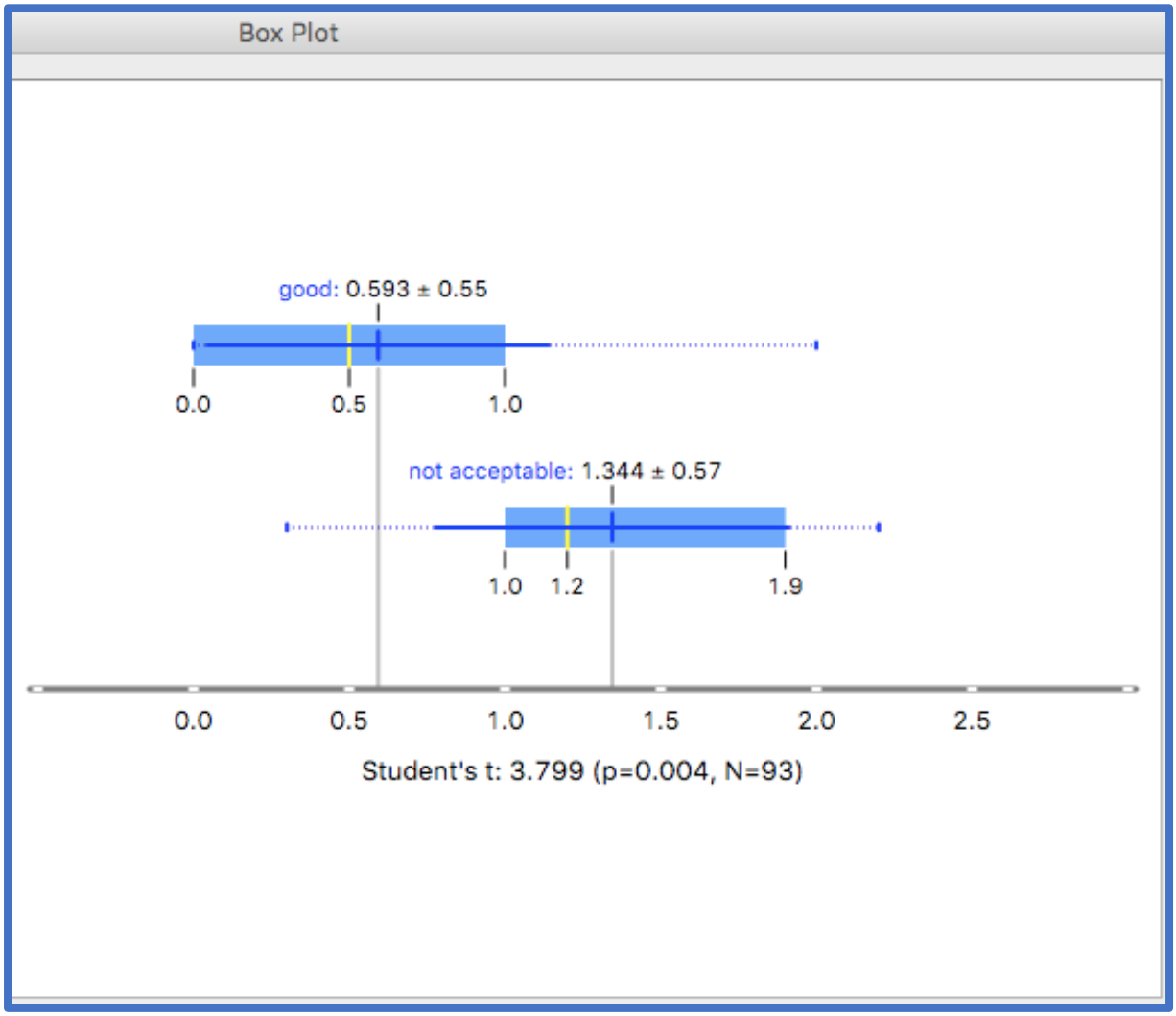

Figure 4.13 Second analysis Boxplot of FAF image quality and NUC cataract.

Boxplot from ODM showing how participant FAF image quality grading is related to NUC cataract score graded on the WHO scale of cataract grading (0.0-3.0) along the bottom. Good and acceptable images were graded as acceptable, and only not acceptable images graded as not acceptable to create a dichotomous scale.

Below is shown the Distribution widget output for the same comparison of NUC cataract grading for image quality graded as good and not acceptable for the second analysis. As the NUC score approaches the level of 1.9 on the WHO scale, there is approximately a 50% chance that the FAF image will be of such a poor, or not acceptable quality as to be not clinically useful. There were, however, two instances where the NUC score was 2.0 and the images were still graded as good. Note that all nine participants (100%) who were graded as having a not acceptable FAF image quality had some degree of NUC cataract.

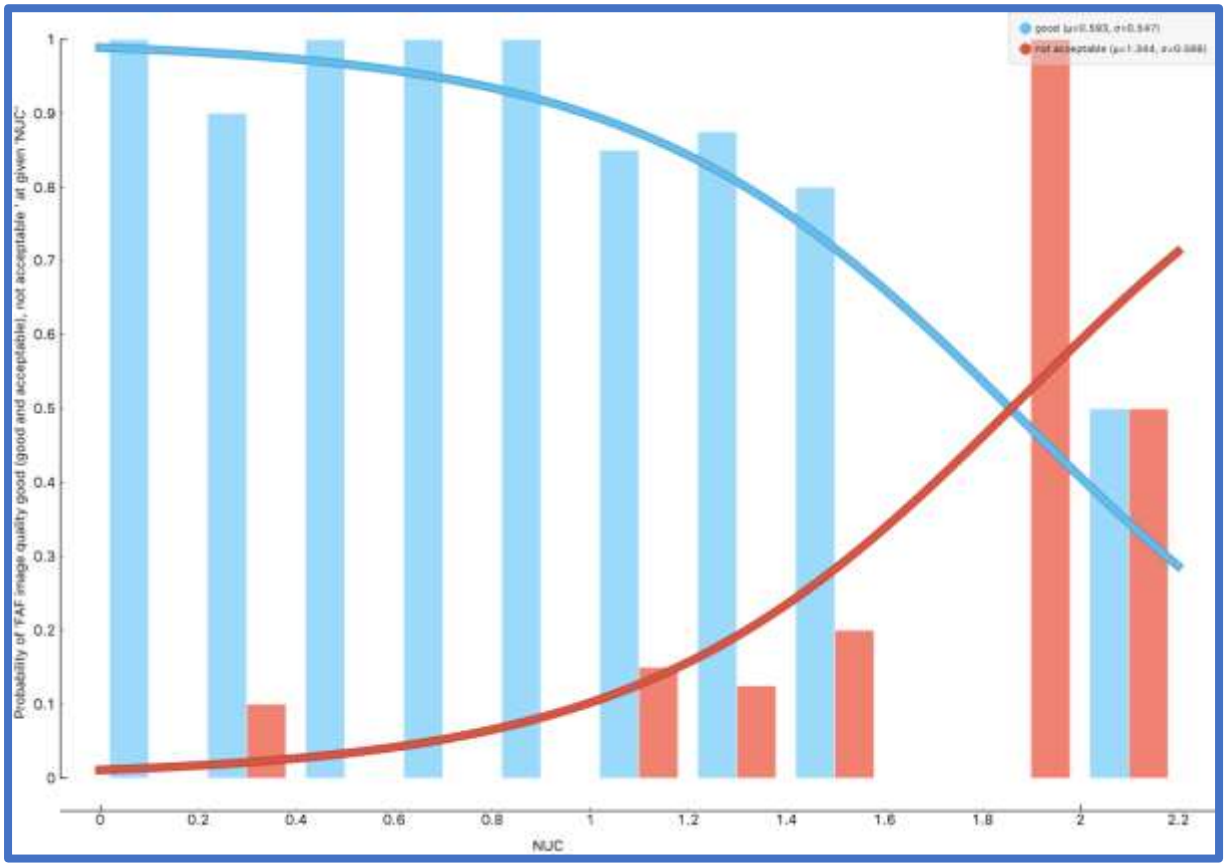

Figure 4.14 Second analysis Distribution of FAF image quality and NUC cataract.

Distribution widget output for NUC cataract on the WHO scale versus the probability of a good or not acceptable FAF images (blue bars represent good image quality and red bars represent not acceptable image quality). Good and acceptable images were graded as good with only not acceptable images graded as not acceptable to create a dichotomous scale.

The same analysis was carried out for COR cataract, the results are shown below. For not acceptable images considered not clinically useful, the mean COR score is 1.030±0.79 and a median of 1.2. For clinically useful images the mean COR cataract score was 0.137±0.48, and a median of 0.0. The Student's t test found that there was a significant difference at the 1% (p=0.009) level between the mean COR WHO scale scores of participants with a good and not acceptable FAF image quality. (This finding was supported by the Mann Whitney U test giving a significance value of <0.001 as the normality of the COR data was not perfect).

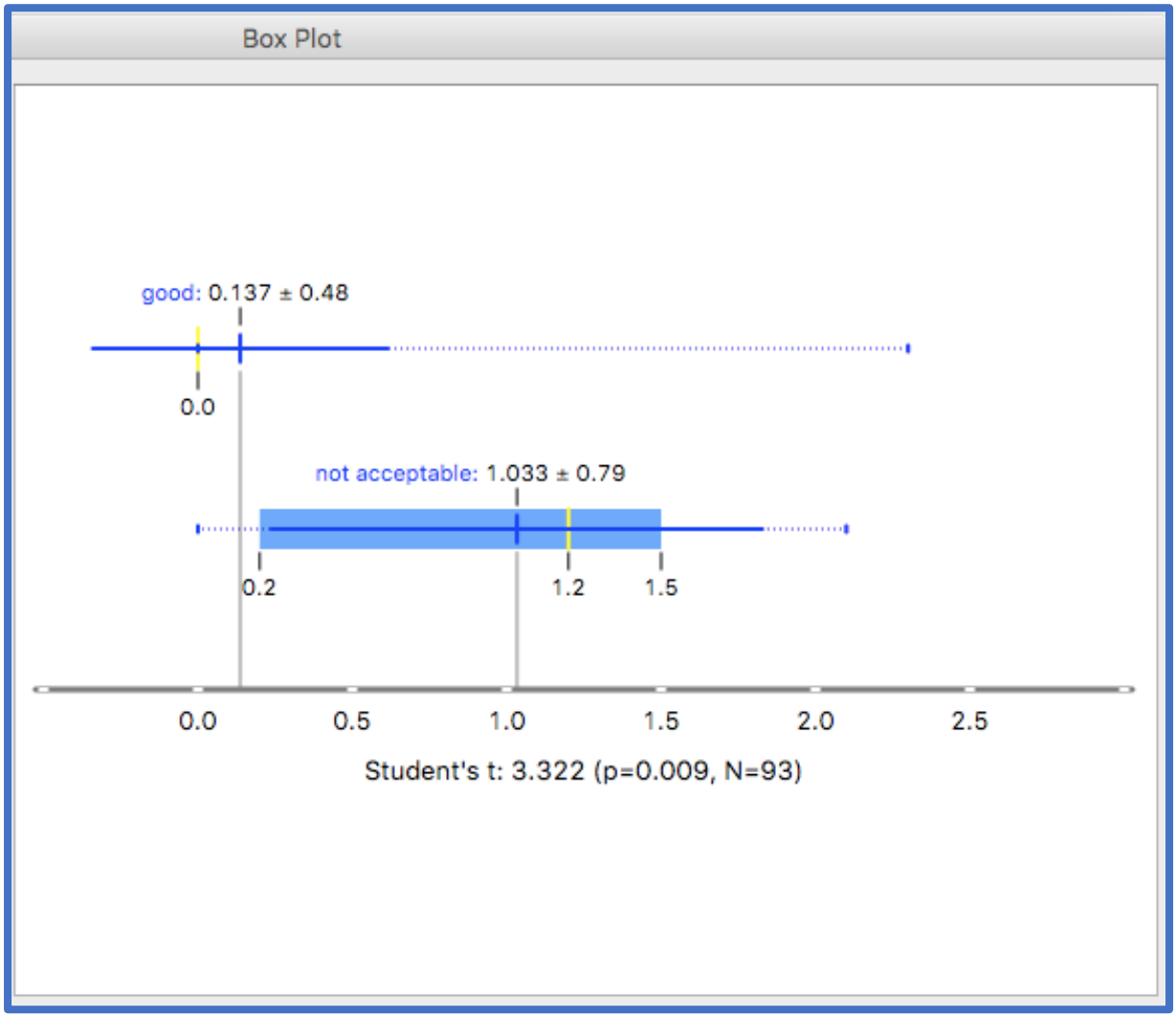

Figure 4.15 Second analysis Boxplot of FAF image quality and COR cataract.

Boxplot from ODM showing how participant FAF image quality grading varies with COR cataract score graded on the WHO scale. Good and acceptable images were graded as good with only not acceptable images graded as not acceptable to create a dichotomous scale.

Below is shown the Distribution widget output for the same comparison of COR cataract grading for image quality graded as good and not acceptable (1 = good, 2 = not acceptable) for the second analysis. As the COR score approaches the level of 1.2 on the WHO scale, there is approximately a 50% chance that the images will be not acceptable. Also, as the cataract score approaches 2.0 on the WHO scale, it becomes highly likely that the FAF image will be of such a not acceptable quality as to be not clinically useful. However, there were three instances of the COR cataract being graded as ≥ 2.0 with the FAF image still graded as good. Note that six of the nine participants (66.7%) who were graded as having a not acceptable FAF image quality had some degree of COR cataract.

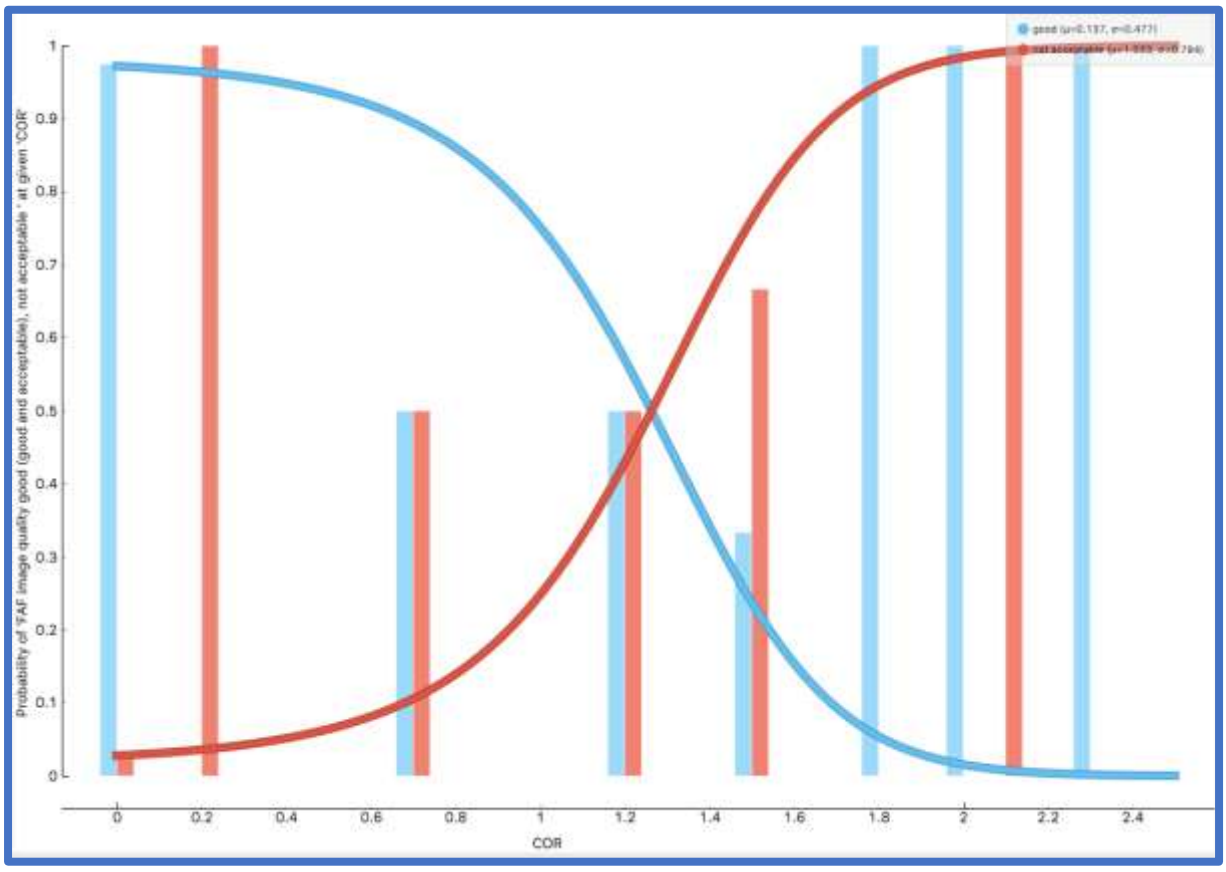

Figure 4.16 Second analysis Distribution of FAF image quality and COR cataract.

Distribution widget output for COR cataract on the WHO scale versus the probability of a good or not acceptable FAF images (blue bars represent good image quality and red bars represent not acceptable image quality). Good and acceptable images were graded as good with only not acceptable images graded as not acceptable to create a dichotomous scale.

The same analysis was carried out for PSC cataract, the results are shown below. For not acceptable images considered not clinically useful, the mean PSC score is 0.344±0.69 and a median of 0.0. For clinically useful images the mean PSC cataract score was 0.098±0.40, and a median of 0.0. The Student's t test found that there was no significant difference between the mean PSC WHO scale scores of participants with a good and not acceptable FAF image quality. (This finding was again supported by the Mann Whitney U test giving a significance value of 0.086 as the normality of the PSC data was not perfect).

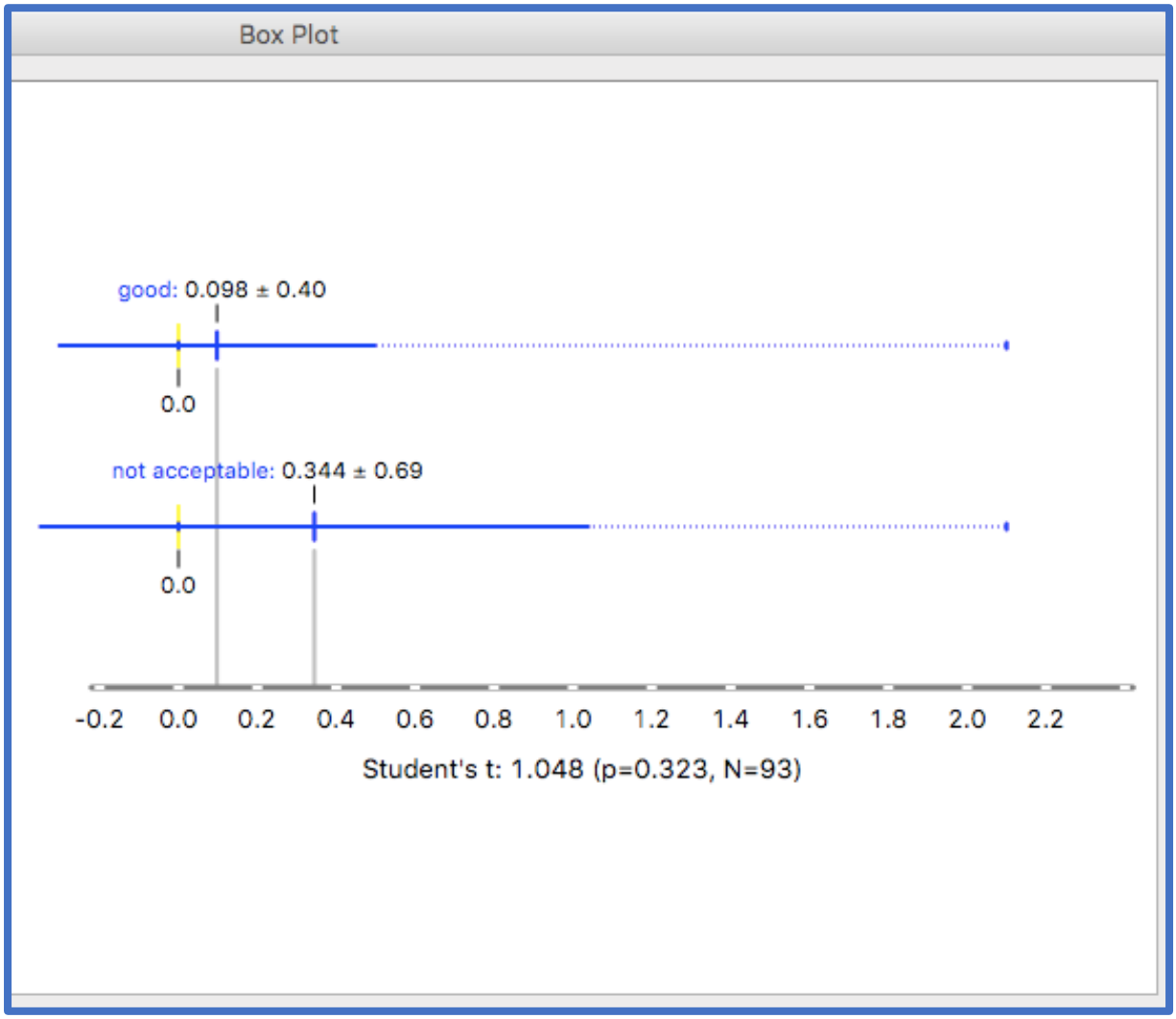

Figure 4.17 Second analysis Boxplot of FAF image quality and PSC cataract.

Boxplot from ODM showing how participant FAF image quality grading varies with PSC cataract score graded on the WHO scale. Good and acceptable images were graded as good with only not acceptable images graded as not acceptable to create a dichotomous scale.

Below is shown the Distribution widget output for the same comparison of PSC cataract grading for image quality graded as good and not acceptable (1 = good, 2 = not acceptable) for the second analysis. As the PSC score approaches the level of 1.1 on the WHO scale, the chance that the FAF image will be of such a poor, or not acceptable quality as to be not clinically useful is approximately 50%. There were, however, two instances of a good FAF image quality with a PSC WHO grading of = 2.1. Note that only two of the nine participants (22.2%) who were graded as having a not acceptable FAF image quality had some degree of PSC cataract.

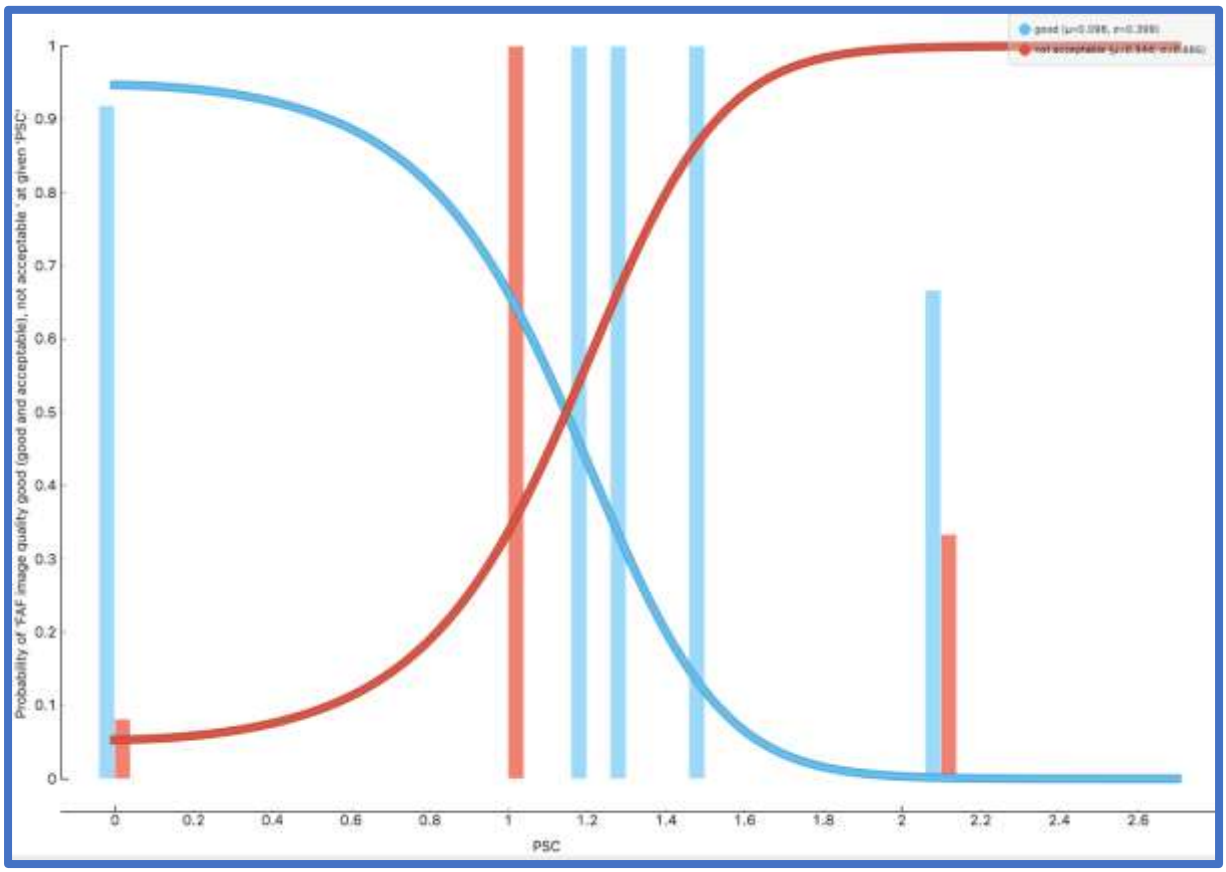

Figure 4.18 Second analysis Distribution of FAF image quality and PSC cataract.

Distribution widget output for PSC cataract on the WHO scale versus the probability of a good or not acceptable FAF images (blue bars represent good image quality and red bars represent not acceptable image quality). Good and acceptable images were graded as good with only not acceptable images graded as not acceptable to create a dichotomous scale.

Distribution widget outputs

Cataract type	Mean cataract score for good FAF image	Mean cataract score for not acceptable FAF image	Upper limit within 1 SD of the mean for good image	Lower limit of 1 SD of the mean for good image	Upper limit within 1 SD of the mean for not acceptable image	Lower limit of 1 SD of the mean for not acceptable image
NUC	0.593	1.344	1.143	0.043	1.914	0.774
COR	0.137	1.033	0.617	0.000	1.823	0.243
PSC	0.098	0.344	0.498	0.000	1.034	0.000

Table 4.7 Second analysis summary of results from ODM Distribution and Boxplot widgets.

The table above showing how the mean cataract scores for NUC, COR and PSC cataract graded on the WHO scale vary with FAF images graded as good and not acceptable, by the definitions set out in the second analysis.

4.3.7 Summary of results of the first and second analysis

The table below was created to illustrate how the quality of images varied with the level of cataract from these two analyses. The first analysis was concerned with **high** image quality, where only good images were graded as good, and both acceptable and not acceptable were graded as not acceptable. The second analysis was concerned with **low** quality images that were not clinically useful, with both good and acceptable images graded together as good, and only not acceptable images graded as not acceptable.

Cataract type	Level at which 50% of images are of **high** quality	Level at which 50% of images are of **low** quality (not clinically useful)	Level above which no image was graded as **high** quality	Level above which all images were **low** quality (not clinically useful)
NUC	1.0	1.9	1.5	No upper limit found
COR	0.7	1.2	No upper limit found	No upper limit found
PSC	0.3	1.1	1.2	No upper limit found

Table 4.8 The levels of cataract and the effect on FAF image quality.

Type of cataract	Number and % of participants with type of cataract in the study	Number and % of participants with this type of cataract having a **high** image quality	Number of participants and % with **high** image quality with type of cataract	Number and % of participants with this type of cataract and **not acceptable image quality**	Number of participants and % with **not acceptable** image quality with type of cataract
NUC	66/93 (71.0%)	35/66 (53.0%)	35/55 (63.6%)	9/66 (13.6%)	9/9 (100%)
COR	14/93 (15.1%)	3/14 (21.4%)	3/55 (5.5%)	6/14 (42.9%)	6/9 (66.7%)
PSC	7/93 (7.5%)	1/7 (14.3)	1/55 (1.8%)	2/7 (28.6%)	2/9 (22.2%)

Table 4.9 Image quality and cataract percentages.

4.4 Discussion

4.4.1 First analysis discussion

The first analysis was concerned with exploring the level of cataract that will still allow a high image quality. Out of the 93 images taken for this study, 55 were classified as good by the definition stipulated for the first analysis (59.1%). It is reasonable to conclude from the results that cataract has a significantly negative effect on FAF image quality, as the Boxplot widgets showed that for all three types of cataract studied, that there was a significantly higher level of cataract on the WHO grading score for FAF images graded as not acceptable quality compared to those images graded as good quality to at least the 5% level. NUC cataract was the most prevalent type of cataract in the sample (71.0% of the total sample of 93 were graded as having NUC cataract, 15.1% as having COR cataract, and only 7.5% as having PSC cataract). No FAF image was graded as good once the NUC score was over 1.5, for PSC this figure was over 1.2. For COR cataract there was no clear upper cut off point for good images, with one instance of a good FAF image in an eye with a COR cataract WHO score of 2.3. These findings indicate that practitioners should not expect high FAF image quality from patients who have a NUC WHO cataract score of over 1.5, and over 1.2 for PSC cataract. For COR cataract it may be possible to record high FAF image quality even with a cataract graded as high as 2.3. NUC cataract, being the most common type of cataract detected in the participants, appears to be responsible for the majority of FAF image degradation due to cataract in this study. An ANOVA was also carried out on the data from the first analysis, on the 31 images from participants who had cataract and whose FAF images were graded as not acceptable quality. The analysis showed that there was significantly higher NUC grading in these participants than either COR and/or PSC cataract, however, with only 14 and 7 participants in the study having COR and PSC cataracts respectively, a larger sample would most likely be required to draw firm conclusions as to the relative effect of the different types of cataract on FAF image quality by the first analysis definition.

4.4.2 Secondary analysis discussion

The second analysis was concerned with exploring the level of cataract that will cause the FAF image quality to be so poor as to be not clinically useful. This information could be beneficial as it would inform the practitioner of the level of cataract that would be likely to render FAF imaging of such poor quality as to be clinically uninformative, freeing up time for

more informative clinical tests. Out of 93 images taken in the study, only nine were graded as not clinically useful by the second analysis definition (9.7%).

The analysis shows that for NUC cataract, there is a significantly higher cataract WHO score for images that were graded as not acceptable compared to those that were graded as good, to the 1% level. At a level of NUC cataract of approximately 1.9, the probability of the FAF image being not acceptable was 50%, however, there was no clear cut off for when the FAF image is highly likely to be of not acceptable quality and therefore not clinically useful, with there being two instances when the NUC score was = 2.0 and the FAF image was still graded as good. Note that all nine images graded as not acceptable by the definition stipulated in the second analysis had some degree of NUC cataract.

For COR cataract, there was once again a significantly higher cataract WHO score for images that were graded as not acceptable compared to those that were graded as good, to the 1% level. At a level of COR cataract of approximately 1.2, the probability of the FAF image being not acceptable was 50%, however, as with NUC cataract, there was no clear cut off for when the FAF image is highly likely to be of not acceptable quality and therefore not clinically useful, there being three instances when the COR score was $\geq$ 2.0 and the FAF image quality was still graded as good. Note that six of the nine images graded as not acceptable had some degree of COR cataract.

For PSC cataract, there was not a significant difference in cataract WHO score between FAF images graded as not acceptable compared to those graded as good. At a level of PSC cataract of approximately 1.1, the probability of the FAF image being not acceptable was 50%. As with the NUC and COR cataract, there was no clear cut off point for when the FAF image is highly likely to be of not acceptable quality and therefore not clinically useful, there being two instances of when the PCS cataract was = 2.1 and the FAF image quality was still graded as good. Note that only two of the nine images graded as not acceptable had some degree of PCS cataract. It should be noted, however, as in the first analysis, that the number of COR and PSC cataracts in the sample were relatively small, so more data would most likely be required to provide stronger evidence regarding the relative effect of different cataract types on FAF image quality by the second analysis definition.

4.5 Conclusion

1. Cataract of all three types studied (NUC, COR and PSC) have a significantly negative effect on FAF image quality to at least the 5% level.
2. Nuclear sclerotic cataract was the most prevalent form of cataract in the sample. (71.0% of the sample were graded as having NUC cataract, 15.1% as having COR cataract, and 7.5% as having PSC cataract).
3. FAF image quality has a 50% probability of being of high quality at a WHO scale grading of 1.0 for NUC cataract, 0.7 for COR cataract and 0.3 for PSC cataract.
4. High FAF image quality should not be expected once the following levels of cataract scores on the WHO score are reached: for NUC cataract > 1.5 and for PCS cataract > 1.2. For COR cataract there was no clear cut off point.
5. FAF image quality has a 50% probability of being of such poor quality that it is rendered clinically not useful at a WHO grading score of 1.9 for NUC cataract, 1.2 for COR cataract and 1.1 for PCS cataract.
6. For all three types of cataract studied, there was no clear cut off point over which all FAF images are expected to be not clinically useful.

4.5.1 Relevance of the findings of Chapter four to primary care optometrists

1. If there is nuclear sclerotic cataract of > 1.5 on the WHO grading scale, then there is a high probability (approximately 85%) that FAF images will not be of high quality, but they may still be clinically useful.
2. If there is a nuclear sclerotic cataract of > 1.9 on the WHO grading scale, then there is approximately a 50% chance that the FAF images will not be clinically useful.
3. There is no level of nuclear sclerotic cataract, measured on the WHO scale, above which there is a high probability that all FAF images will not be clinically useful.

Chapter 5 The impact of FAF on patient comfort

5.1 Introduction

FAF imaging relies on exposing the patient's eye to a relatively bright flash of visible light. This is necessary due to the very low fluorescence signal obtained when attempting to capture in vivo FAF images in human subjects (the Quantum Yield of A2E, a major constituent of LF, is minimal at 0.003 ± 0.001).(17) For patient comfort and ocular safety reasons, using a very high-power excitation light source is not practically possible,(180) however, the light flash from FAF imaging is still considerably more intense than that generated during colour fundus photography (typically 300 Watt seconds for FAF compared to 60 Watt seconds for CFP).(181) Chapter five of the study was therefore designed to explore the experience, regarding visual/ocular comfort, of patients exposed to the FAF excitation stimulus produced by the RS-330 instrument. This part of the study could assist primary care optometrists by providing information on the likely patient acceptance of greater routine use of FAF imaging in primary care optometry.

5.2 Methods

A Likert six-point scale of 0-5 was utilised for patient comfort, with 0 being very uncomfortable, and 5 being very comfortable. This question was asked immediately after the imaging process in the following way, "on a scale of 0 to 5, with 0 being very uncomfortable and 5 very comfortable, how would you score the flash that you've just experienced?".

5.3 Results

Results were analysed via ODM software in order to explore any data patterns regarding patient comfort during FAF imaging, the results of which may help to inform primary care optometrists of the most appropriate way to incorporate FAF imaging into their routine clinical practice. Significance levels were checked for all the *categorical* variables via the Boxplot widget with two variables appearing to cause a significant effect on comfort during FAF imaging. These were the crystalline lens status (i.e. phakic or pseudophakic), and the normality or abnormality of the FAF image. The results are shown below.
Below is the Boxplot widget comparing the FAF image acquisition acceptance score from 0-5 (0 = very uncomfortable to 5 = very comfortable) for pseudophakic and phakic participants. There is a significant difference in comfort score to the 1% ($p = 0.004$) level with

pseudophakics having a lower mean of comfort score of 3.92±0.8 (median 4.0), and phakic individuals having a higher mean comfort score of 4.48±0.8 (median 5.0).

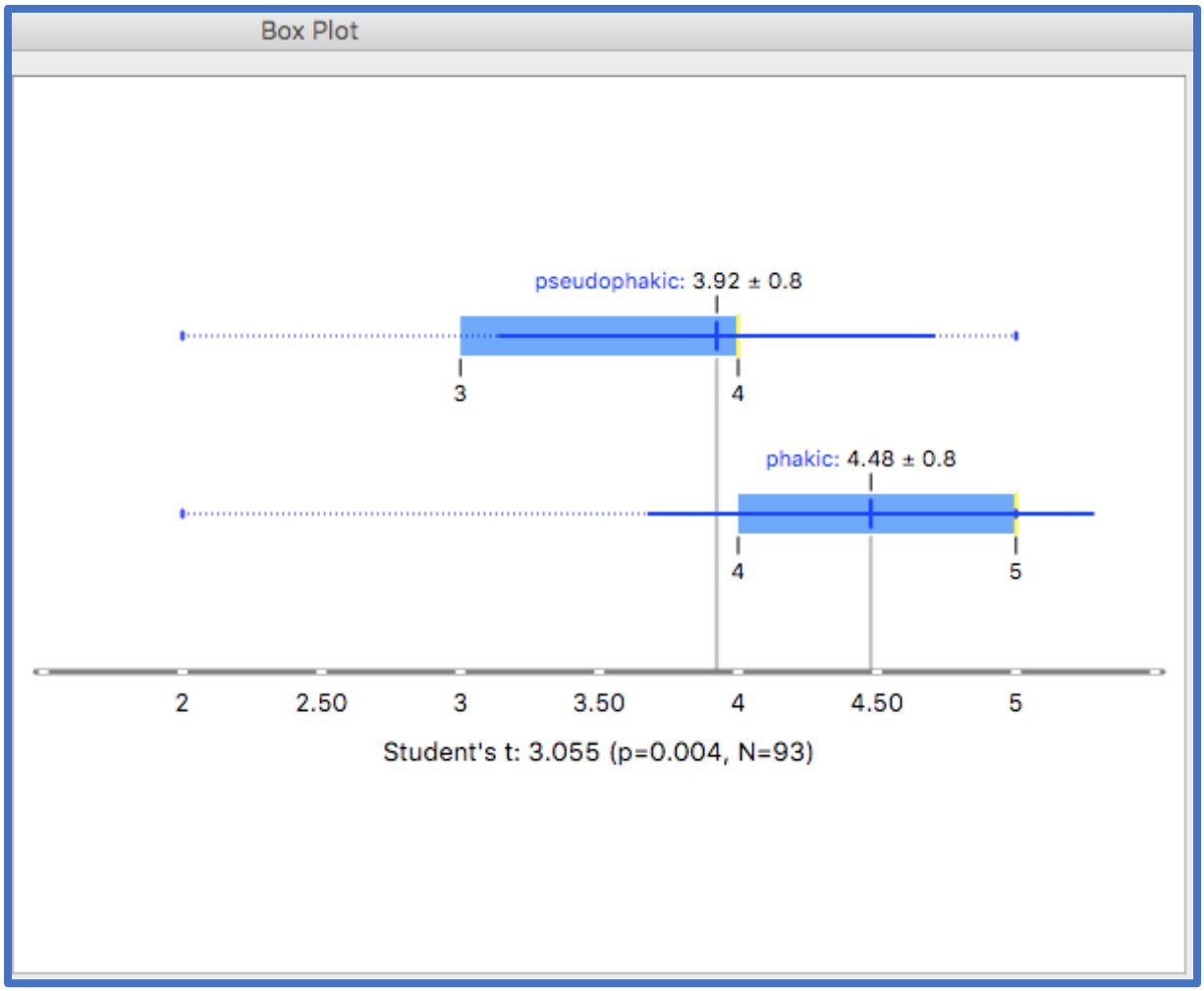

Figure 5.1 Boxplot of lens status and FAF comfort score.

Boxplot from ODM showing how participant lens status varies with FAF image acquisition comfort score along the bottom.

Below is the Boxplot widget comparing the FAF image acquisition comfort score for participants demonstrating a normal and abnormal FAF image. There is a significant difference in the comfort score to the 1% (p = 0.01) level, with participants demonstrating an abnormal FAF image having a lower mean comfort score of 3.96±0.8 (median 4.0), and those demonstrating a normal FAF having a higher mean comfort score of 4.47±0.8 (median of 5.0).

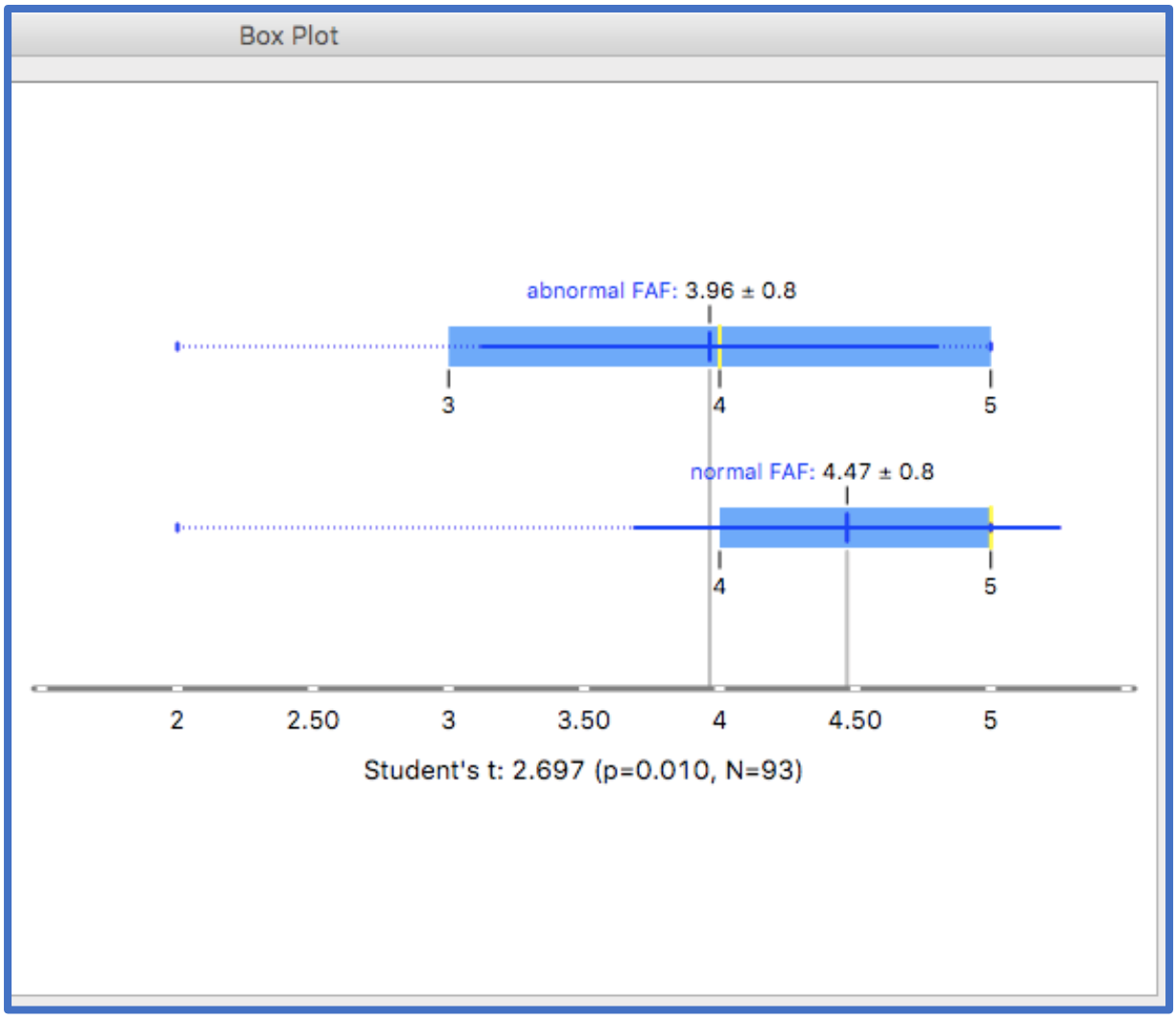

Figure 5.2 Boxplot FAF image normality/abnormality and FAF comfort score.

Boxplot from ODM showing how participant FAF dichotomous image grading varies with FAF image acquisition comfort score along the bottom.

Later, in Chapter six, it was found that as age increases, so too does the probability of pseudophakia *and* an abnormal FAF image. Therefore, the worsening comfort seen with pseudophakia and an abnormal FAF image could simply be down to a worsening comfort

score with increasing age. To test this hypothesis, it was necessary to create a categorical age column in the Microsoft Excel spreadsheet, as neither the Boxplot nor Distribution ODM widget enables comparison of two variables with numerical values - one must be categorical. An age cut off of 77 years was selected. This was based on the finding from the Boxplot widget that the mean age of a participant with an abnormal FAF image was 76.56 years, and the mean age of a pseudophakic participant was 77.54 years. Below is the Boxplot widget showing the results. The mean comfort score for > 77 years was 3.96±0.9 (median 4.0), and for < 77 years the mean comfort score was 4.46±0.8 (median 5.0) $p = 0.013$, i.e. the difference between the comfort score of FAF in under and over 77-year-olds was significant to the 5% level, but does not quite reach the 1% significance level. Therefore, it is possible to conclude that age could be a confounding factor in the previous findings regarding the lower comfort of FAF grading in patients who were pseudophakic or who demonstrated an abnormal FAF image, with older participants experiencing more discomfort during FAF image acquisition.

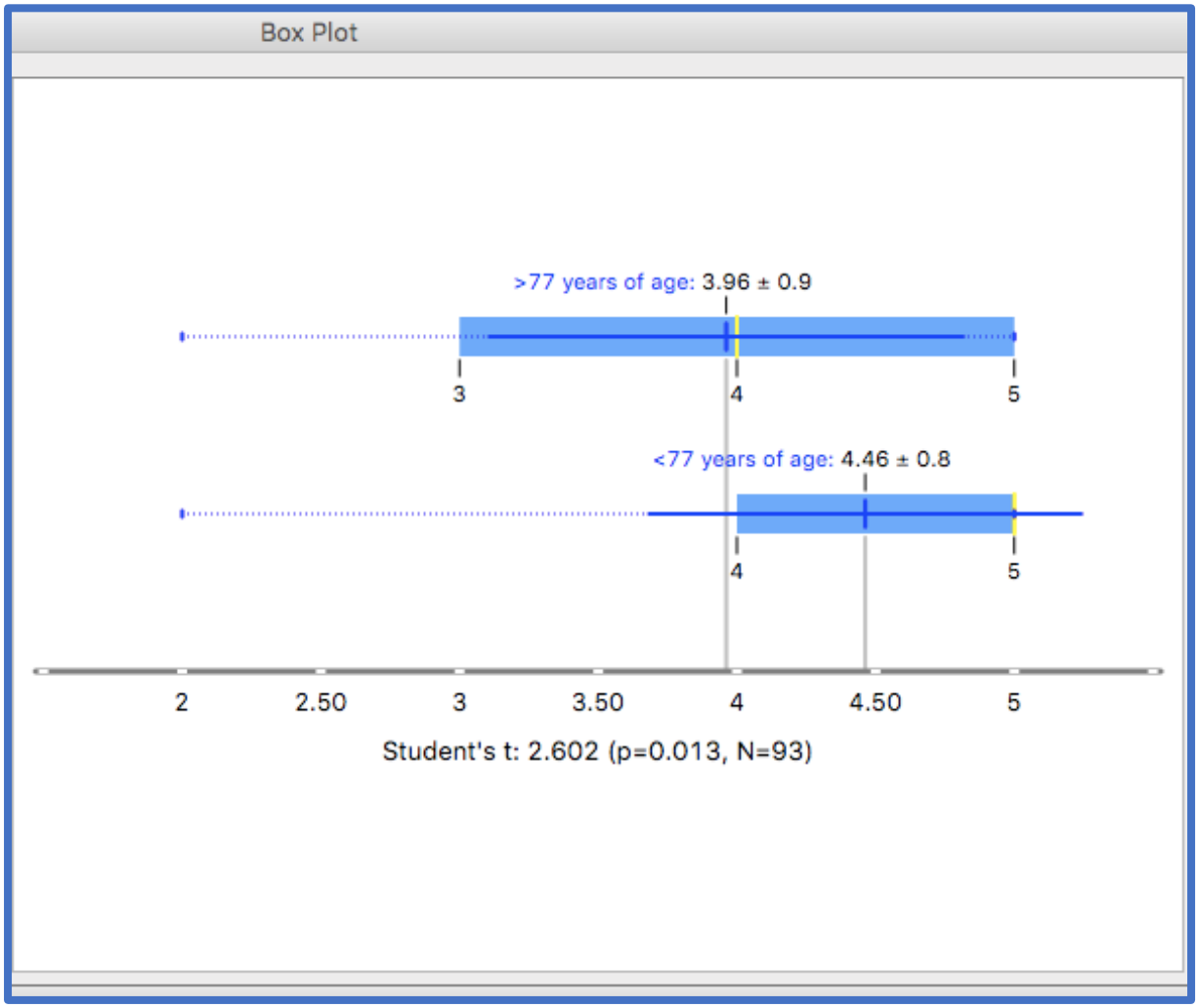

Figure 5.3 Boxplot of participant age and FAF comfort score.

Boxplot from ODM showing how participant age varies with FAF image acquisition comfort score along the bottom.

Below is the Distribution widget from ODM showing how there is a general trend for an increase in the probability of being under 77 years of age (red bars), and a decrease in the probability of being over 77 years of age (blue bars), as the FAF comfort score increases.

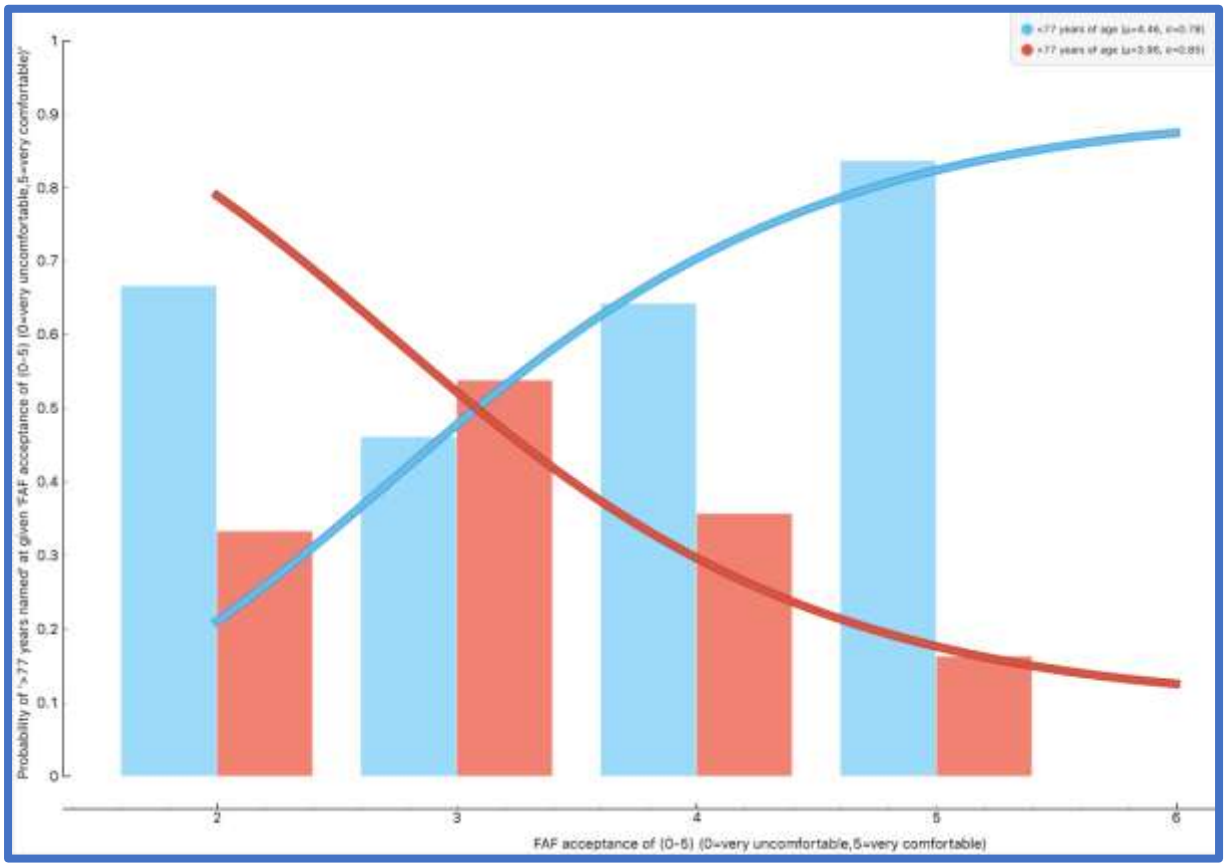

Figure 5.4 Distribution of under and over 77 years of age and FAF comfort score.

Distribution widget from ODM showing how the probability of being over 77 years of age varies with the comfort score (along the bottom) during FAF imaging. Blue bars represent the probability of participants being over 77 years of age, red bars represent the probability of participants being under 77 years of age.

An alternative strategy for testing whether age is a confounding factor for the finding that FAF comfort is lower in pseudophakics and in participants with an abnormal FAF result, was to create a sub-group composed of two age matched groups (N=48), one for phakics and another for pseudophakics. This technique does however, suffer from the disadvantage that the number of participants included in the analysis is lower, to achieve age-matching of the groups. The results are illustrated below in the Figure below. The Boxplot shows that whilst

pseudophakics still have a lower mean comfort score, the difference is now not statistically significant.

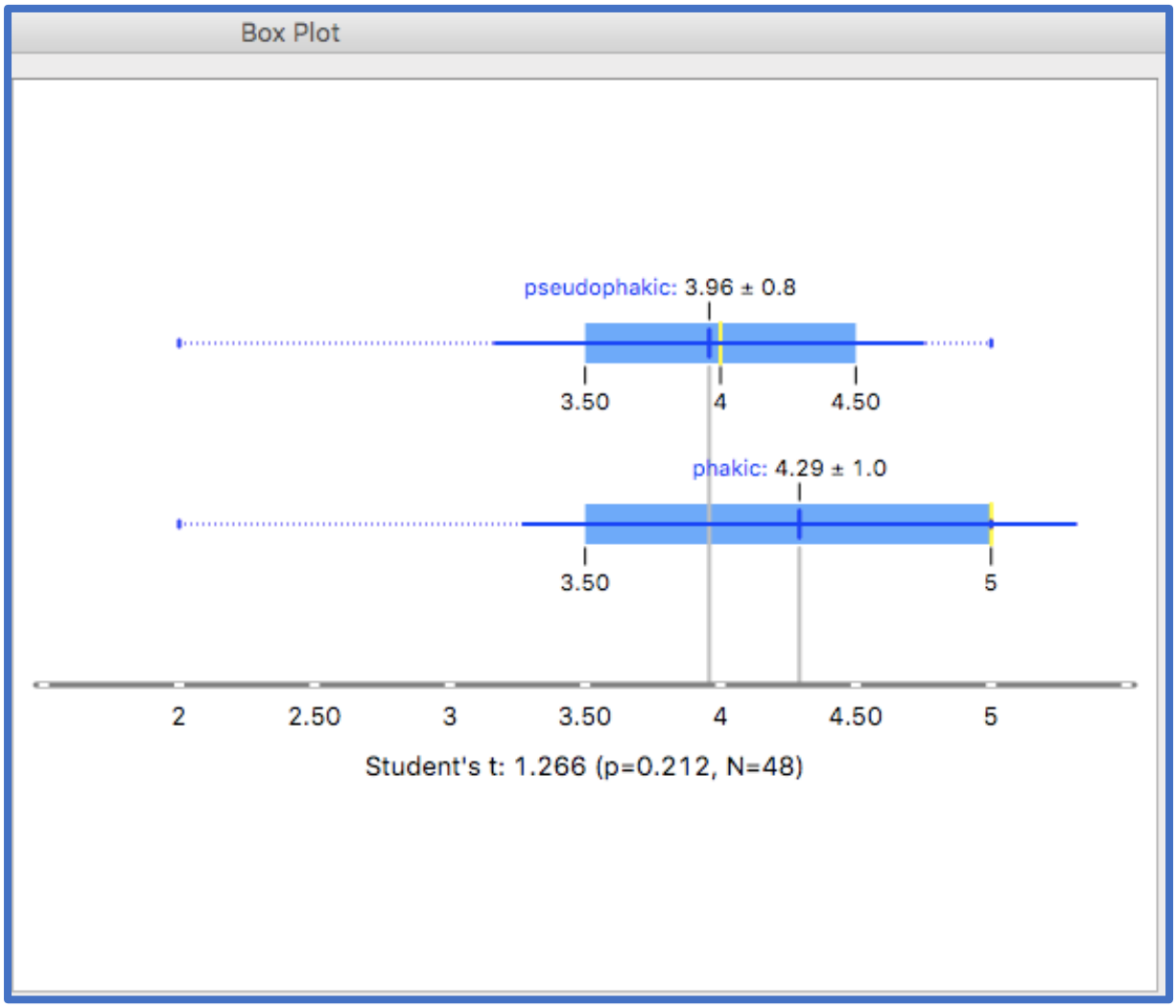

Figure 5.5 Boxplot of lens status and FAF comfort score for age-matched groups.

Boxplot from ODM showing how participant lens status (upper group 1 = pseudophakic, lower group 2 = phakic) varies with FAF image acquisition comfort score (along the bottom) for a sub-group composed of two age matched groups of participants (N=48).

The same lack of statistical difference is seen in the following Boxplot output in Figure 5.6 for an abnormal FAF result in the age matched groups; again, the participants with an abnormal FAF result still have a lower FAF comfort score, but this is no longer statistically significant.

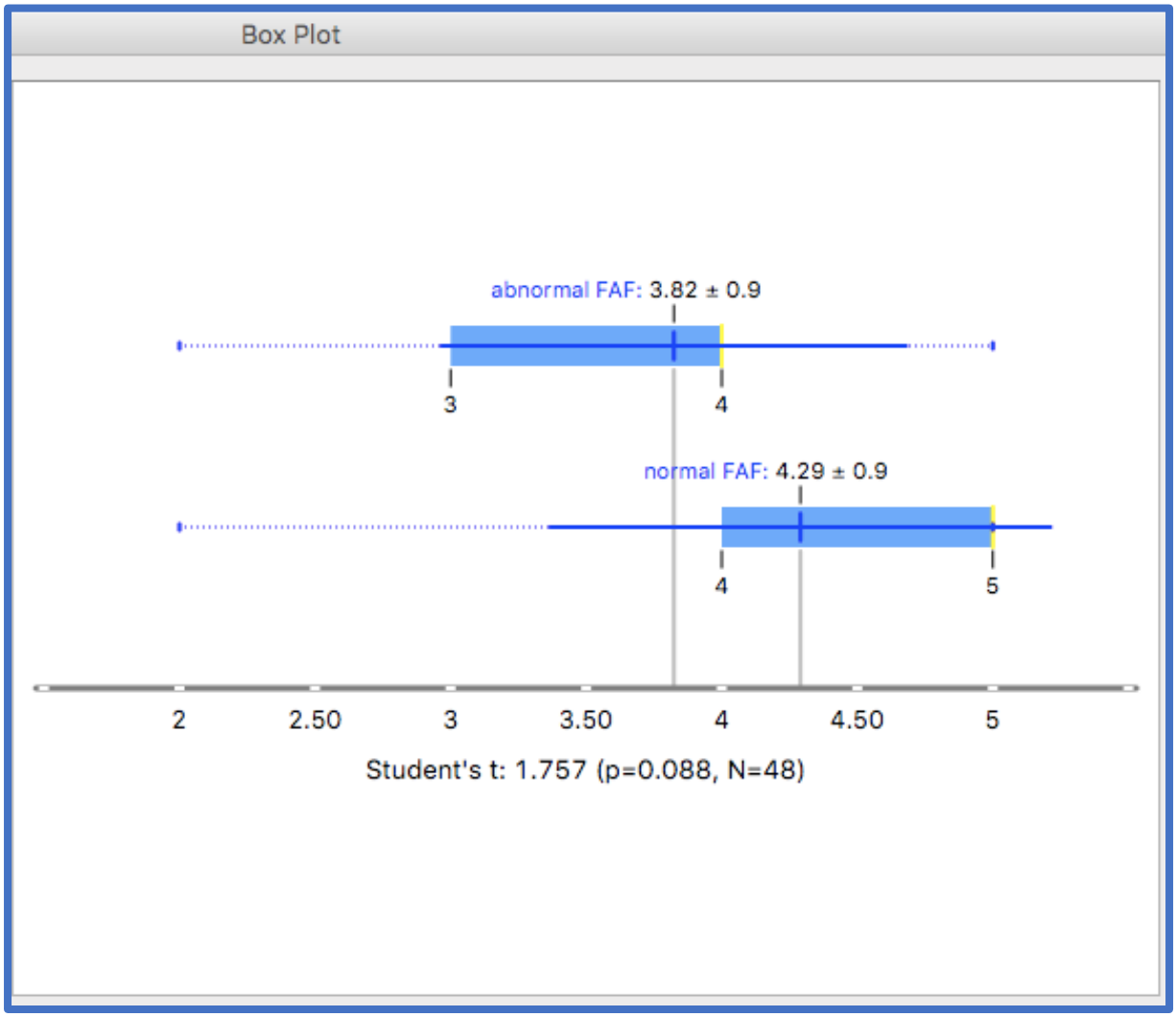

Figure 5.6 Boxplot FAF image normality/abnormality and FAF comfort score for age-matched groups

Boxplot from ODM showing how participant FAF image normality/abnormality varies with FAF image acquisition comfort score (along the bottom) for a subgroup composed of two age matched groups of participants (N=48).

Another numerical variable examined was pupil size. It would be reasonable to assume that pupil size post dilation may also be an important factor in patient comfort during the FAF imaging process, and that a larger pupil size would increase the light exposure of the retina and therefore have a negative effect on the comfort score.

To explore the variable of pupil size post dilation thoroughly, it was necessary to create a categorical pupil size post dilation grading system, so that the data could be handled by the ODM software. The mean pupil size post dilation for under 77 years was 5.597mm±0.61mm, and for over 77 years was 5.538mm±0.69mm.
Therefore, an arbitrary cut off point of 5.50mm was chosen so that two categories of < 5.50mm and ≥ 5.50mm were created. The Boxplot shown in Figure 5.7 below indicates that there is no significant difference in the mean FAF acquisition comfort scores for those with a pupil size post dilation of < 5.50mm and ≥ 5.50mm, with the group that had a larger pupil size post dilation actually having a higher mean and median comfort score. Also, from a contextual perspective, pupil size has previously been shown to decrease linearly as function of age for all levels of illuminance,(182) (so again age could be a confounding factor) and if a larger pupil size post dilation was an important factor in lowering FAF comfort scores, one would expect the comfort score to *increase* with age rather than to decrease.

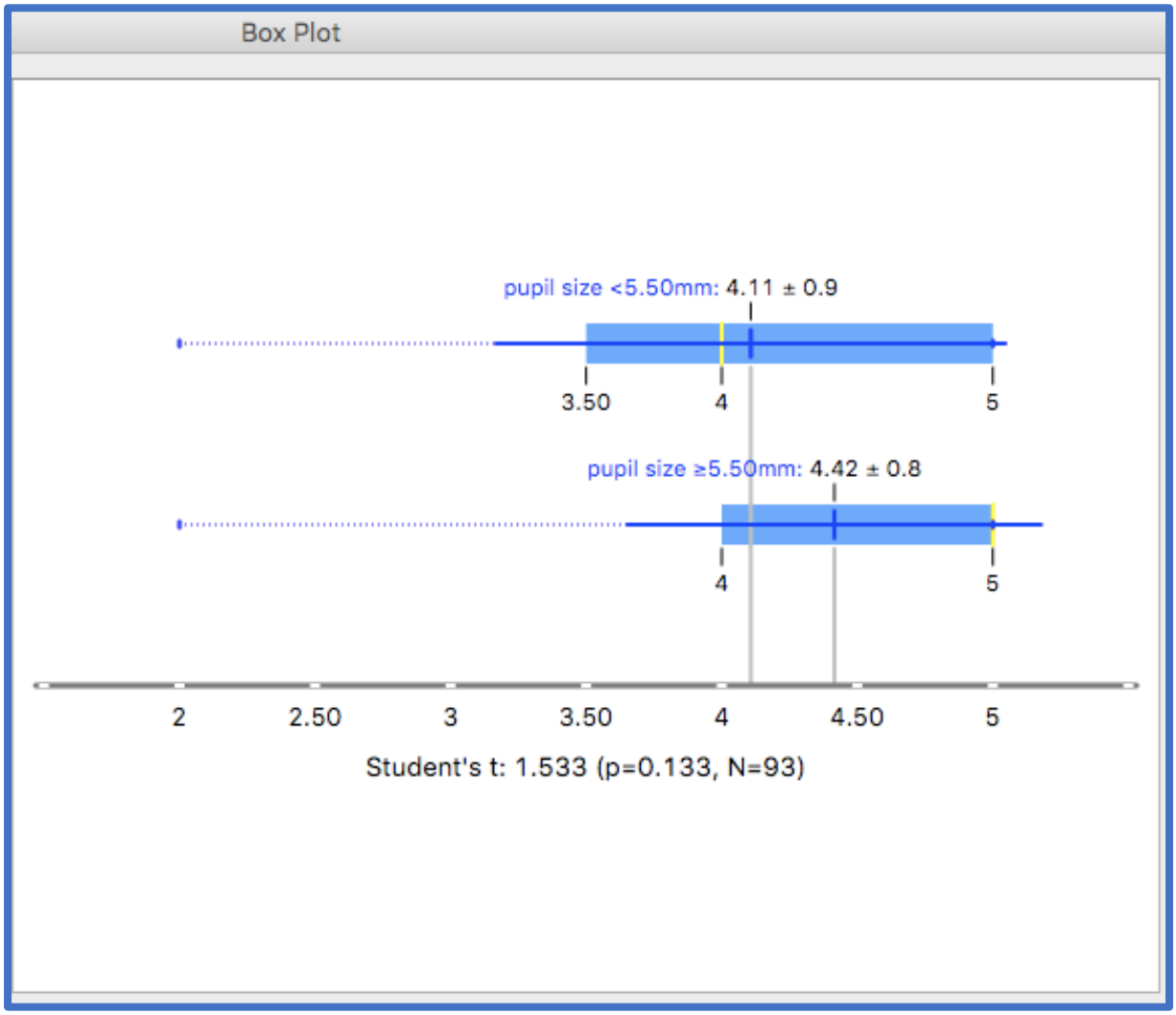

Figure 5.7 Boxplot of pupil size post dilation and FAF comfort score.

Boxplot from ODM showing how participant post dilation pupil size varies with FAF image acquisition comfort score (along the bottom).

Migraines have been known to have been triggered by bright flashes of light,(183) however, from the Boxplot output below shown in Figure 5.8 there was no significant difference in mean comfort score for participants who reported to be migraine sufferers and those who did not. (Note, however, that photosensitive migraine sufferers were excluded from the study).

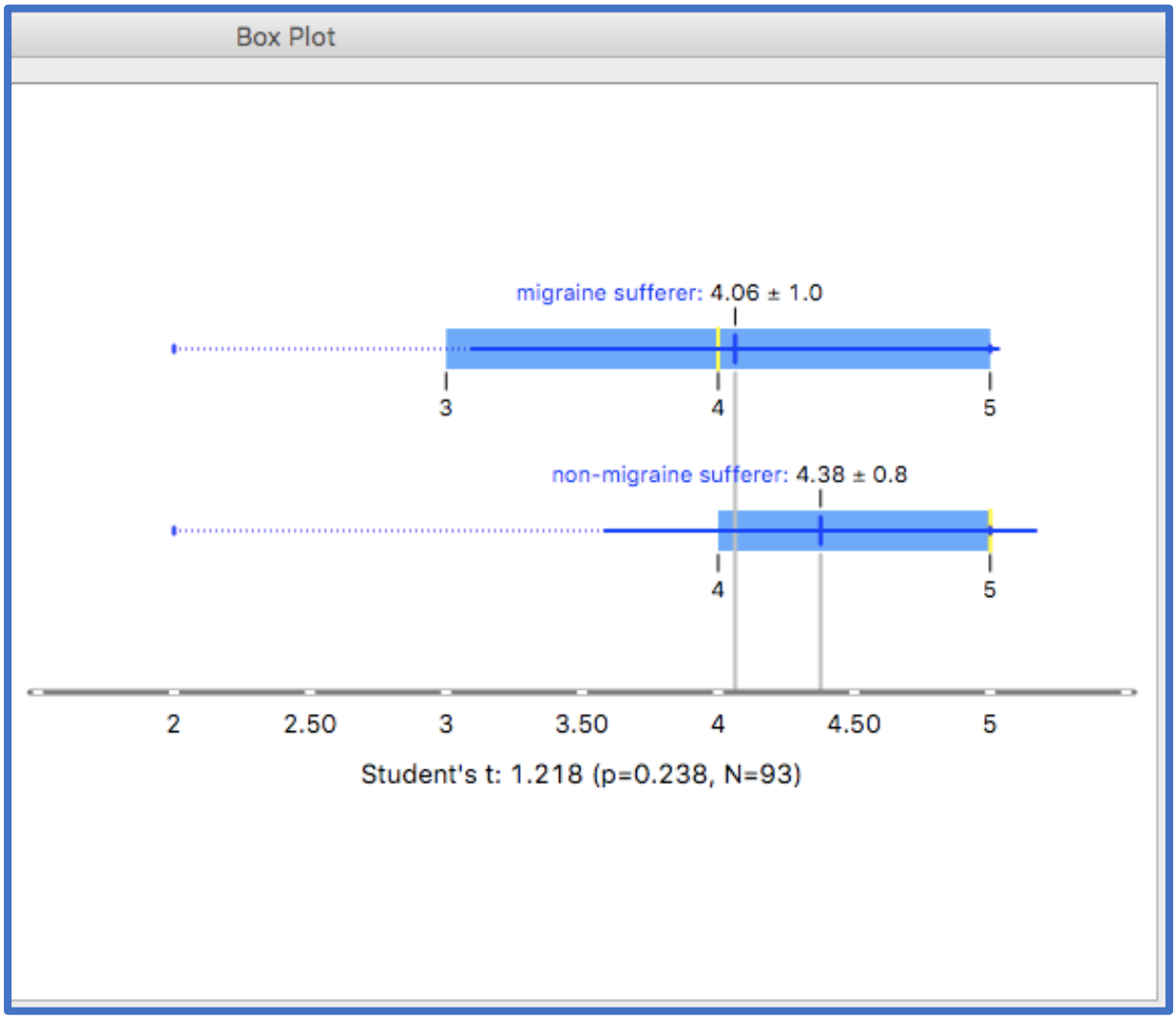

Figure 5.8 Boxplot of migraine status and FAF comfort score

Boxplot from ODM showing how participant migraine status) varies with FAF image acquisition comfort score (along the bottom).

There was only one case of non-photosensitive epilepsy in the cohort for the study (photosensitive epilepsy was an exclusion criterion), and therefore ODM was unable to compute significance for this variable. Therefore, more research would be needed to draw useful conclusions regarding comfort in non-photosensitive epileptics during the FAF image acquisition process.

Lastly, to enable a complete exploration of all remaining *numerical* variables via ODM, a categorical grading system for comfort was applied, with grading of comfort ≤ 3 and >3. This again highlighted the association between age and comfort as shown in Figure 5.9 below. There is a statistically significant difference in the mean age in the two groups to the 5% level (p=0.018), with those participants in group 1 who had a mean comfort score of ≤ 3 being significantly older than those in group 2 who had a mean comfort score of > 3. No other non-FAF related numerical variables, e.g. refractive status, vision or cataract status yielded a statistically significant difference in mean FAF comfort score.

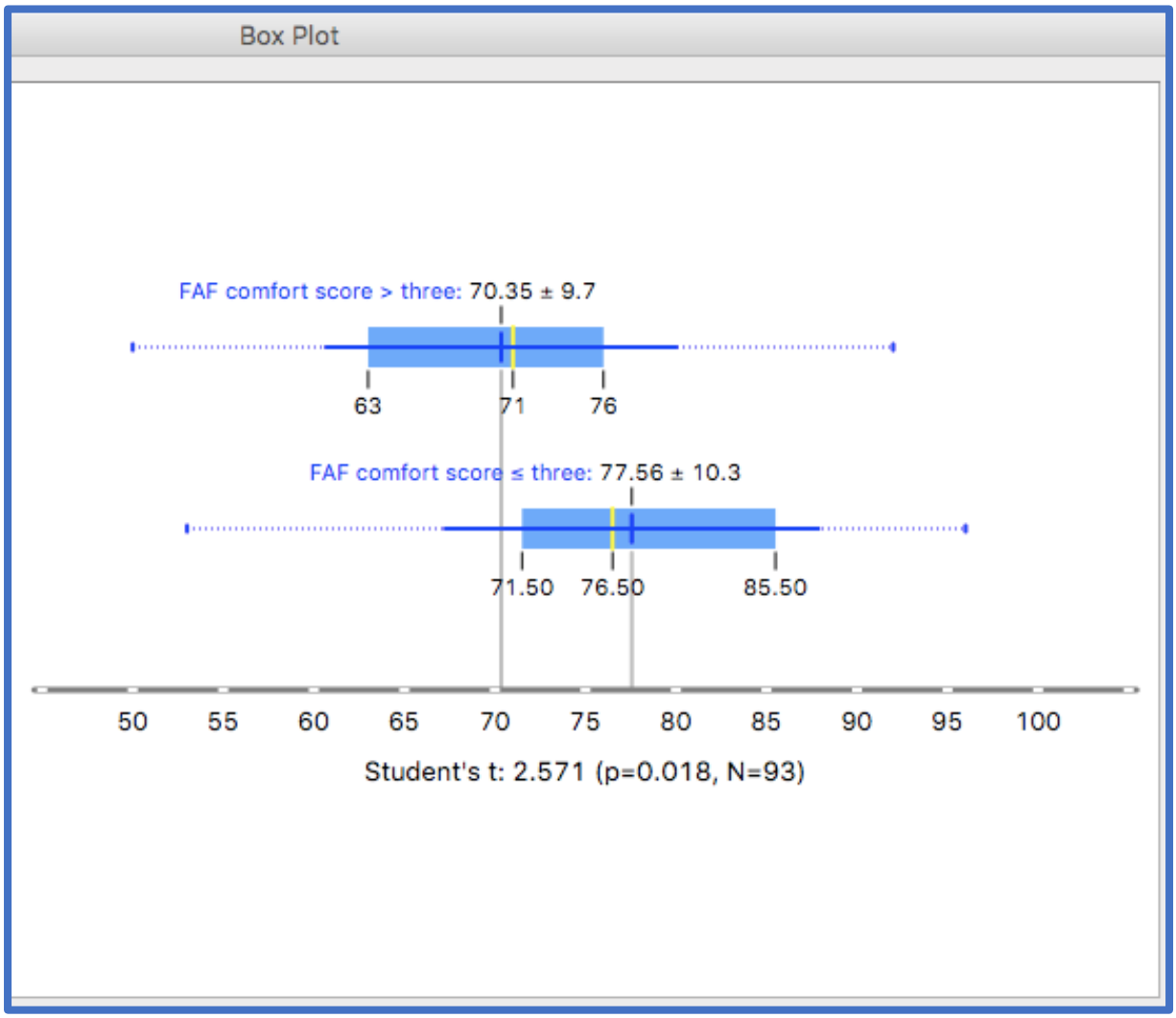

Figure 5.9 Boxplot of comfort score and age

Boxplot from ODM showing how participant FAF image acquisition comfort score (on a scale of 0 = very uncomfortable, and 5 = very comfortable) varies with age (along the bottom).

The Distribution widget output in Figure 5.10 below shows the increasing proportion of comfort scores ≤ 3 with increasing age. For age < 60 years the proportion of scores ≤ 3 is 7.7%, for ≥ 60 ≤ 80 years of age this rises to 12.7% and for > 80 years this rises again to 17.2%.

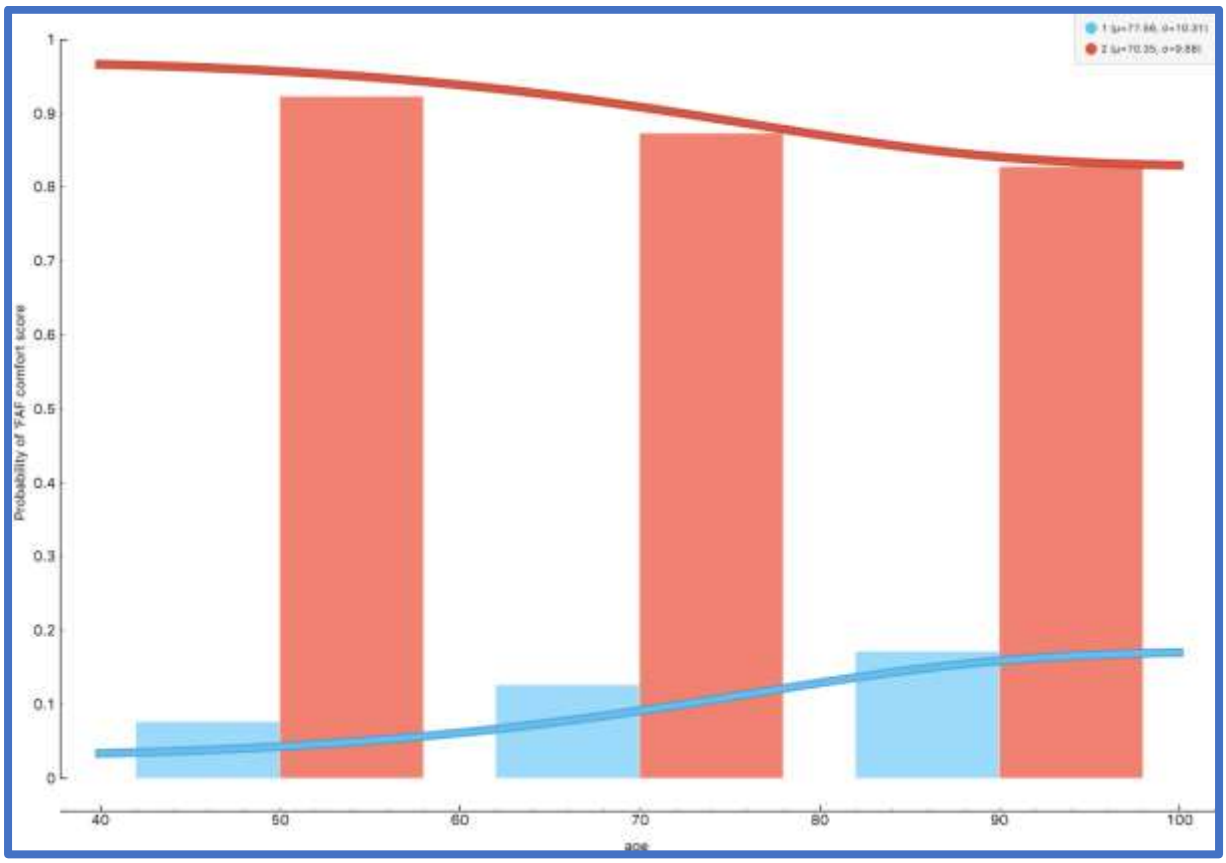

Figure 5.10 Distribution of age and probability of comfort score ≤ 3 and > 3.

Distribution widget from ODM showing how the probability of having a FAF comfort score of ≤ 3, or > 3 varies with participant age (along the bottom). Blue bars represent a comfort score of ≤ 3, red bars represent a comfort score of > 3.

5.4 Discussion

The results indicate that having an abnormal FAF image, being pseudophakic, and being > 77 years of age are all linked to a significantly lower mean comfort score. However, a subgroup composed of two age-matched groups was created (N=48), and the subsequent analysis demonstrated that an abnormal FAF image and/or pseudophakia made no significant difference to the mean comfort score. Therefore, it can be concluded that older age was a confounding factor for these two variables. This finding was highlighted again by another statistical strategy which had the advantage of using all 93 participants' data, by creating two categorical groupings for FAF comfort score, with the two options of ≤ 3 and > 3. This analysis highlighted that as age increases the proportion of FAF comfort scores ≤ 3 (lower comfort) in the group increases, whilst the proportion of comfort scores > 3 (better comfort) decreases. The mean age of a participant with a comfort score of ≤ 3 was 77.56 years±10.3 years (median 76.50 years), and for > 3 the mean was 70.35 years±9.7 years (median 71 years).

One might expect a larger pupil size post dilation to also have a negative effect on comfort during FAF imaging, however, there was no significant difference between the two groups, one which had a pupil size post dilation of < 5.50mm, and the other a pupil size post dilation of ≥ 5.50mm, with the mean comfort score actually being slightly higher in the group with the larger pupils. It is possible to conclude from these findings that a larger pupil size post dilation seen in clinical practice will not indicate that discomfort during FAF image acquisition will be any greater than those patients with a smaller pupil size post dilation.

Therefore, older age alone appears to increase the probability that the patient will experience relatively greater ocular/visual discomfort during the FAF image acquisition. Practically this could mean explaining the process to the patient, its importance, providing a quiet area to recover from the flash after the imaging and offering the patient assistance to this area of the premises. Patients who are migraine sufferers should be carefully questioned regarding potential triggers for their migraines, however, the findings of this study do not indicate a statistically significant link between the FAF image acquisition process and discomfort in non-visually evoked migraine sufferers. More research would be needed to explore any associations between non-photosensitive epilepsy and discomfort during FAF

imaging, as there were not enough participants who suffered from the condition in this study to draw useful conclusions.

For future studies involving comfort during image acquisition in optometry, it may be worth considering and recording qualitative data surrounding other possible sources of discomfort experienced by older patients. For example, headaches experienced following eye imaging or the awkward physical positioning of the legs, back, torso and neck required to obtain optimal images, rather than simply the discomfort glare from the flash as considered in this study.

In summary, the median FAF visual/ocular comfort score for participants under 77 years age was 5 (mean was 4.46±0.80), and for those over the age of 77 years of age the median was 4 (mean was 3.96±0.90), (0 = very uncomfortable, 5 = very comfortable). Only three of the total cohort of 93 participants graded the comfort score as 2, which was the lowest score recorded during the study. In conclusion, FAF image acquisition, with specific precautions and patient advice, is suitable in terms of visual/ocular comfort to patients examined in clinical practice.

5.5 Conclusions

1. Being > 77 years of age has a significantly negative effect on patient visual/ocular comfort during the FAF image acquisition process, with increasing age leading to a higher probability of a FAF comfort score of ≤ 3.
2. Having a larger pupil size post dilation does not negatively affect patient visual /ocular comfort during the FAF image acquisition process.
3. Being a non-visually evoked migraine sufferer does not negatively affect patient visual/ocular comfort during the FAF image acquisition process.
4. More research is required to draw conclusions regarding the comfort of non-photosensitive epilepsy sufferers during the FAF image acquisition process.
5. Refractive status, visual acuity and cataract status do not have a statistically significant effect on the FAF image acquisition comfort score.

Chapter 6 The impact of performing FAF in addition to CFP and OCT in clinical practice

6.1 Introduction

The aim of Chapter six was to build a database of patients on which a series of clinical tests have been carried out and a number of variables collected. ODM Artificial Intelligence open source software was used to investigate whether FAF imaging offers information over and above that already obtained with CFP and OCT, and therefore adds clinical value to eye examinations (regarding findings relevant to ARMD).

Questions that Chapter six will attempt to answer:

1. Does age have an impact on the likelihood of an abnormal FAF result? (Hypothesis – age has no impact on the likelihood of an abnormal FAF result).(114, 184)
2. Does refractive status have an impact on the likelihood of an abnormal FAF result? (Hypothesis – refractive status has no impact on the likelihood of an abnormal FAF result).(185)
3. Does smoking history have an impact on the likelihood of an abnormal FAF result? (Hypothesis - Smoking history has no impact on the likelihood of an abnormal FAF result).(184)
4. Does having a first degree relative with ARMD have an impact on the likelihood of an abnormal FAF result? (Hypothesis – having a first degree relative with ARMD has no impact on the likelihood of an abnormal FAF result).(186)
5. Does the number of drusen, as detected by CFP and/or OCT have an impact on the likelihood of an abnormal FAF result? (Hypothesis – the number of drusen has no impact on the likelihood of an abnormal FAF result).(114)
6. How does the simplified severity score impact on the likelihood of an abnormal FAF result? (The simplified severity score has no impact on the likelihood of an abnormal FAF result).(114)
7. How does the presence of pigmentary abnormalities impact on the likelihood of an abnormal FAF result? (The presence of pigmentary abnormalities have no impact on the likelihood of an abnormal FAF result).(114)
8. How does the presence of reticular pseudodrusen impact on the likelihood of an abnormal FAF result? (The presence of reticular pseudodrusen has no impact on the likelihood of an abnormal FAF result).(165)

6.2 Statistics employed

ODM software was used to explore the collected data, with the help of the property "Informedness" for the reasons discussed below.

There are many different statistical expressions to help to evaluate results from a dichotomous classification as required in this study.(187)

Below is an example of a four-cell contingency table summarising the predictions of a classifier in the context of a binary dichotomous problem. The table assumes that we are predicting a single condition, with either a positive or negative result, and that there is one model predicting the outcome (predictions), and a gold standard reference test (real). The green squares indicate correct counts and the yellow incorrect. The counts within the table are simply to act as an example and do not refer to the findings of this study.

	Real positives	Real negatives	
Predicted positive	true positives = 8 (tp)	false positives = 4 (fp)	Total predicted positives = 12 (pp)
Predicted negative	False negatives = 2 (fn)	true negatives = 86 (tn)	Total predicted negatives = 88 (pn)
	Total true positives = 10 (rp)	Total true negatives = 90 (rn)	Total number of cases 100 (N)N.

Table 6.1 Example of a four-cell contingency table

Abbreviations used:

tp = true positives
fp = false positives
fn = false negatives
tn = true negatives
rp = real positives = tp+fn
rn = real negatives = fp+tn
pp = predicted positives = tp+fp
pn = predicted negatives = fn+tn

6.2.1 Statistical expressions defined

a. Sensitivity (or Recall), is the proportion of *real* positive cases that are correctly predicted as positive. It is a measure of how good a test is at detecting the positives. Note that performance of handling negative cases is completely ignored. Mathematically it is expressed as tp/rp. Sensitivity is also one of the legs on which the receiver operating characteristic (ROC) analysis stands. It is also known as the true positive rate (tpr).
b. Precision is the proportion of *predicted* positive cases that were actually real positives. It is a measure of how many of the positively classified cases were positive. Mathematically it is expressed as tp/pp. It is also known as the true positive accuracy (tpa).
c. Specificity (or inverse recall) is defined as the proportion of *real* negative cases that are correctly predicted as negative. Mathematically expressed as tn/rn. It is also known as the true negative rate (tnr).
d. Inverse precision is the proportion of *predicted* negative cases that were actually real negatives. Mathematically it is expressed as tn/pn. It is also known as the true negative accuracy (tna).
e. Rand Index is a measure of the percentage of correct decisions made. Mathematically it is expressed as (tp + tn)/N.
f. The Jaccard similarity coefficient compares two sets by giving a percentage of how many members are present in both sets, and ranges from 0 to 100%.
g. Fallout is also known as the false alarm ratio or false positive rate (fpr). Mathematically it is expressed as fp/rn. False positive rate is the second of the legs on which ROC analysis is based.
h. Miss rate is also known as the false negative rate (fnr) Mathematically it is expressed as fn/rn.
i. Informedness represents the probability that a prediction is informed in relation to the condition versus chance. (The condition is the experimental outcome we are trying to predict by indirect means). Mathematically it is expressed as Sensitivity + Specificity - 1.
j. Markedness represents the probability that a condition is marked by the predictor versus chance. (The marker/predictor is the indicator we are using to determine the outcome). Mathematically it is expressed as Precision + Inverse Precision – 1.
k. ROC analysis plots the true positive rate (tpr) or sensitivity (on the vertical axis), against the false positive rate (fpr) or fallout (on the horizontal axis).

6.3 The uses of Informedness and Markedness

Informedness is appropriate for testing effectiveness relative to set of conditions, whilst markedness is appropriate for testing effectiveness relative to a set of predictions. In general terms, we can use informedness if we wish to know which solution best solves a problem, and also which problem is usually best solved by a particular solution, whilst markedness may be more appropriate for predicting, for example, which test will be effective across a wide range of complaints.

If we have a large number of documents containing the required information, it can be assumed that we do not need to find all the relevant documents, and a small set of documents can be expected to provide that information with confidence. In this scenario markedness is more relevant, however, if we have a small, specific set of documents for which we need to be confident that virtually all of them have been accessed, then informedness is a more appropriate measure. In this study, informedness was therefore selected as the most appropriate measure to test model performance.(187)

A strategy was devised to ascertain whether carrying out FAF imaging offers optometrists extra information over and above standard eye examination tests, i.e. CFP and OCT, for the detection of retinal changes associated with ARMD. This was achieved by measuring the degree of informedness (sensitivity + specificity -1) for predicting FAF imaging results, achieved by using the standard tests of CFP alone, and subsequently with both CFP and OCT combined. If FAF imaging results could be accurately predicted, i.e. with a high level of informedness, it may seem reasonable to assume that FAF adds little clinical information over and above standard tests.

6.4 Variables collected

CFP, OCT and FAF images were analysed and a number of variables were collected. (Note that all retinal variables were measured within the major temporal vascular arcades of the study eye unless otherwise stated). The variables collected are tabulated below:

Non-imaging variables collected	**CFP**	**OCT**	**FAF**
Age	Number of small, intermediate and large drusen counted within the temporal vascular arcades	Number of small, intermediate and large drusen counted within the temporal vascular arcades	Classification of the FAF pattern
Average spherical refraction (Gullstrand)	Presence of large drusen in the fellow eye	Presence of large drusen in the fellow eye	Presence of reticular pseudodrusen
Gender	Presence of geographic atrophy (GA)	Presence of GA	Presence of GA
Symptoms of distortion in the central vision i.e. a distortion of the central vision, a “kink” in the central vision or central scotoma	Disc diameters of GA	Disc diameters of GA	Disc diameters of GA
A history of or ongoing anti-VEGF injections	Presence of pigmentary anomalies (PA)	Minimum and average central foveal thickness	Presence of halo defects surrounding areas of GA
Smoking status	Disc diameters of PA	Presence of RPD	Acceptance of test in terms of patient comfort
Packet years of smoking	Simplified severity score (see section 6.5 below)	Presence of central or peripheral epiretinal membrane	
Years since smoking cessation	Presence of reticular pseudodrusen (RPD)	Presence of vitreomacular traction	
Epilepsy status		Presence of lamellar hole	
Migraine status		Presence of full thickness macular hole	
Diabetic status		Presence of macular cysts	

Family history of ARMD		Presence of subretinal fluid	
Visual acuity in the study and non-study eye		Presence of subretinal hyperreflective material (SHRM)	
Pupil size post dilation		Presence of pigment epithelial detachment	
Cataract type and grading			
Whether phakic or pseudophakic			
Degree of posterior capsular opacification if pseudophakic			

Table 6.2 Variables collected for the analysis in Chapter six.

6.5 Simplified Severity Score

The simplified severity score for ARMD from the Age-Related Eye disease Study (AREDS) report number 18 was utilised as a variable within this study, which is based on findings from CFP imaging. "The scoring system developed for patients assigns to each eye 1 risk factor for the presence of 1 or more large (≥ 125 μm, the approximate width of a normal large vein as it crosses the optic disc margin) drusen and 1 risk factor for the presence of any pigment abnormality. Risk factors are summed across both eyes, yielding a 5-step scale (0-4) on which the approximate 5-year risk of developing advanced ARMD in at least one eye increases in this easily remembered sequence: 0 factors, 0.5%; 1 factor, 3%; 2 factors, 12%; 3 factors, 25%; and 4 factors, 50%"(188) (In addition, for participants with no large drusen in either eye, the presence of bilateral intermediate drusen is counted as 1 risk factor. 2 risk factors are assigned for the presence of advanced ARMD in the fellow eye). Advice from the "Clinical Decision-Making Aid for Nutrition in Age Related Macular Degeneration" designed at Aston University was utilised by reference to the flow chart to offer nutritional advice to patients.(189)

Note that the AREDS Report number 17 recommends a 9-point severity scale for ARMD, and reports that drusen area is a stronger and more consistent risk factor than drusen size. However, with there being a strong association between the two variables of drusen area and size, and with the assumption that eye care professionals could more effectively assess maximum drusen size than drusen area, the simplified severity score from AREDS report number 18 utilises drusen size rather than area.(188) The AREDS report number 17 also

considered whether the type of drusen should be taken into account for the severity scale grading (i.e. hard drusen or soft indistinct drusen), however, in view of the complexity of adding a second drusen characteristic, and the low additional predicative power that this would provide, it was decided not to introduce this variable into the scale.

6.6 Methods

Whilst computer programs are available for the counting and measuring of drusen,(190, 191) for the purposes of this study it was decided that manual counting of drusen should be employed within anatomically defined regions of the retina, and for drusen size calculation reference was made to anatomical structures visible within retinal images. The reason for these decisions was two-fold. Firstly, since this study is based within a primary care optometry environment, it was considered appropriate to utilise methods that were easily accessible to the average community-based optometrist. This could facilitate auditing their own patient cohorts for comparisons of their results with this study's results. Secondly, any "rules of thumb" that are highlighted by the findings of this study could easily be translated into the everyday work of community optometrists without the need for specific software packages or equipment to which most community clinicians would not have easy access. Therefore, the number of drusen of specific sizes were manually counted by means of viewing the CFP/OCT images.

For CFP: Large drusen were taken to be those with a diameter $\geq 125\mu m$ (width of the average normal retinal vein at the disc margin), intermediate drusen were those $\geq 63\mu m$ in diameter (i.e. half that of a large drusen), and small drusen $< 63\mu m$ in diameter. The number of drusen of each size was recorded as none (grade 0), 1-9 (grade 1), 10-19 (grade 2), 20-29 (grade 3), 30-39 (grade 4), 40-49 (grade 5) and 50+ (grade 6), i.e. on a seven-point scale.

For OCT: Features were measured within the major temporal vascular arcades of the study eye. The "measure" function of the OCT was also utilised to estimate drusen size (in μm, by stretching an arrow between two points by means of the PC mouse) at its widest horizontal diameter, with the same size grading used as for CFP. The grader scrolled through the "edge to edge" macular map scan and manually counted the number of drusen of each size within the area bounded approximately by the major temporal vascular arcades. Again, the number of drusen of each size was recorded as none (grade 0), 1-9 (grade 1), 10-19 (grade

2), 20-29 (grade 3), 30-39 (grade 4), 40-49 (grade 5) and 50+ (grade 6), i.e. on a seven-point scale, in a similar manner to the method used to count drusen with CFP images as detailed above.

For FAF: Classifications of FAF patterns were made according to a classification system proposed in 2005 by Bindewald et al for early ARMD, which classifies FAF patterns into one of eight categories,(165) i.e. normal pattern, minimal change pattern, focal increased pattern, patchy pattern, linear pattern, lacelike pattern, reticular pattern and speckled pattern (Hereafter referred to as "Bindewald's classification system". Note that Bindewald's classification system was designed for "early ARMD", and that having more "advanced" ARMD was not an exclusion criterion for this study. However, only eight of the 93 participants were recorded as having "advanced" ARMD in the form of treated wet ARMD or GA, and therefore it was still considered appropriate for Bindewald's classification system to be employed throughout the study as over 90% of the participants were defined as having "early" or no ARMD. Later, to facilitate 10-fold cross validation via ODM statistical analysis, these patterns were also simplified and expressed dichotomously as "normal" if the pattern was normal or minimal change, and "abnormal" if the pattern was any of the other six remaining patterns according to Bindewald's classification system.

6.7 Orange Data Mining (ODM) Free Open Source Software

This section will explain the functioning of the commonly used ODM widgets as well as taking the reader through ways in which this software can be utilised and how to avoid common pitfalls which may have a detrimental effect on analyses.

ODM was first released in 1996, and is a free, open source AI data analysis and visualisation tool, developed at the Bioinformatics Laboratory of The University of Ljubljana in Slovenia, and is based on Python scripting. The program presents the researcher with a “canvas”, which is then used to create personalised workflows by connecting a series of “widgets”. The widgets are arranged in groups under the headings: Data, Visualise, Model, Evaluate, Unsupervised and Image Analytics, and through selecting these the user can read the data, present data tables, select features, train predictors, compare algorithms, and visualise the results in a number of formats.

ODM can remember previous choices and suggest to the user the most frequent combinations. The software can also visualise data via scatterplots, trees, bar charts, dendrograms, heatmaps etc.

Figure 6.1 below shows the ODM workflow designed for this study.

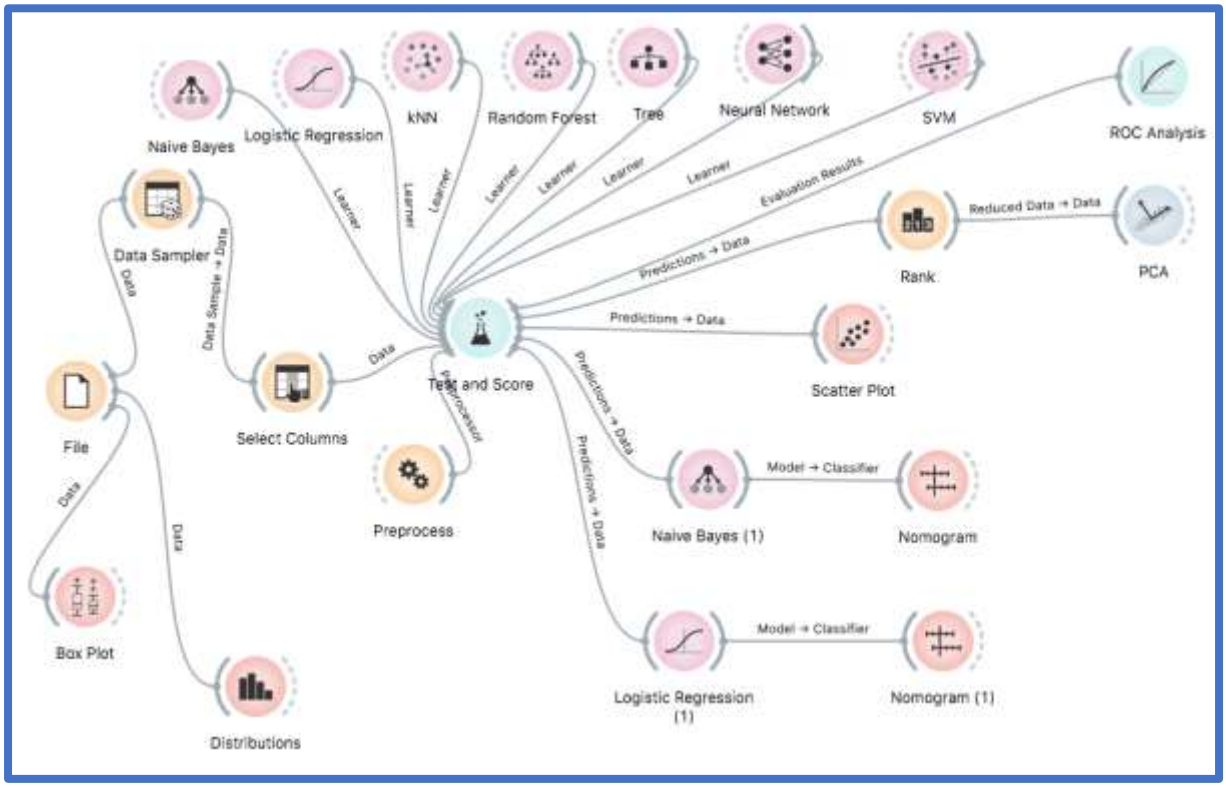

Figure 6.1 Orange data mining workflow used in Chapter 6

6.7.1 Avoiding pitfalls in Orange Data Mining

6.7.1.1 Introduction

The ODM toolbox can join compatible widgets via input and output threads on the left and right of the widgets respectively. These threads can thus create a flow of data through the pipeline created by the user. Orange performs calculations as the pipeline is constructed, indicated by a grey ring that gradually forms (indicating progress) around the widget performing the analysis. Double-clicking on the widget reveals its content. Any change further upstream in the pipeline is automatically filtered downstream.(192)

6.7.1.2 Pre-processing of data

Discussed below are four main areas where the researcher may choose to conduct pre-processing of data to avoid common pitfalls encountered in data analysis using ODM software.

Pre-processing 1. *Normalising, shuffling, filling missing values, removing outliers and continuising data*

Before analysing data, the researcher may decide to pre-process the data, e.g. normalise (scaling data), inputting any missing data, continuising discrete data, shuffling, removing outliers etc. prior to the actual machine learning process. This initial pre-processing step is crucial if the machine learner is to perform its tasks effectively. ODM assists in many of these initial tasks e.g. the Preprocess widget will handle the first three of these tasks, i.e. normalising, inputting missing values (by either replacing the gaps with the average or most frequent missing values) and continuising discrete variables. Shuffling is carried out by the Test and Score widget. In this study outliers were screened manually from the data.

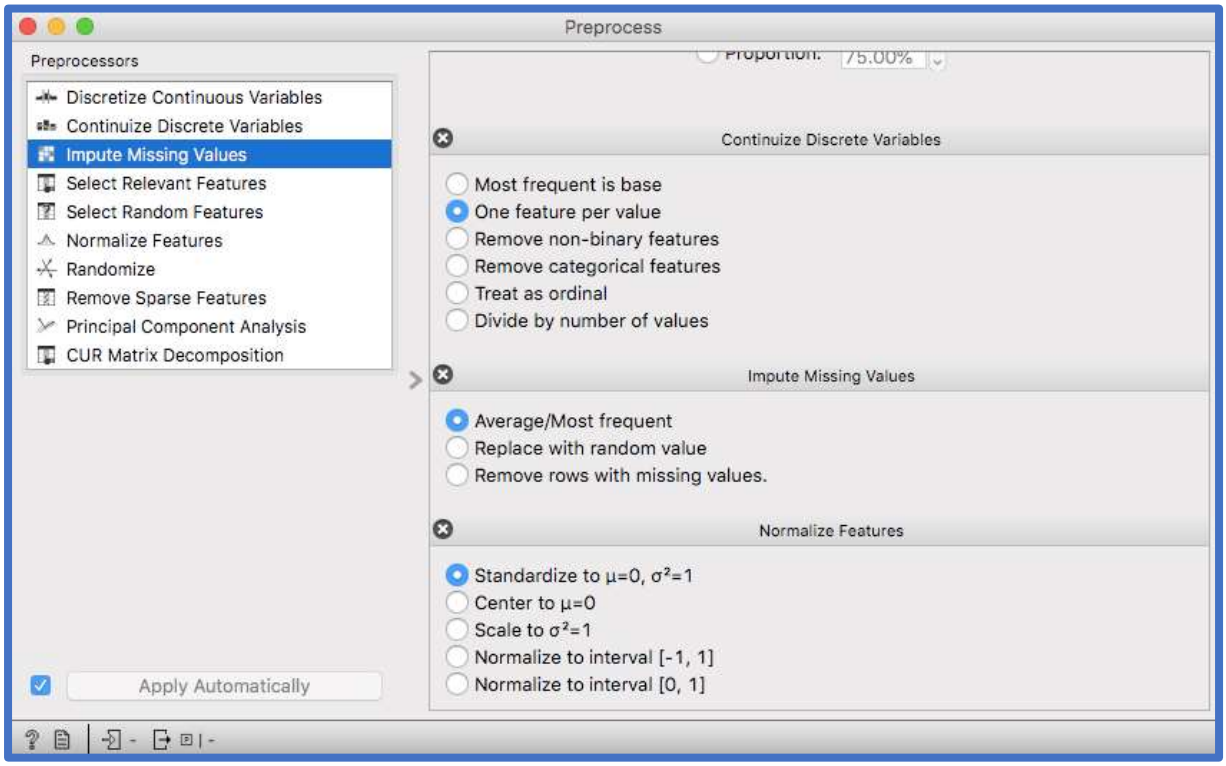

Figure 6.2 Normalising, continuising and inputting missing data via the Orange Data Mining Preprocess widget

Pre-processing 2. *The ratio of data instances to variables*

The researcher may be misled by the machine learning process when the number of variables greatly exceeds the number of data instances.(192) Some authors have suggested that there should be at least ten times as many data instances as there are variables,(193) but this may depend on the quality of the relationships within the data. Essentially, the number of variables may need to be stripped back to a workable number of the most informative variables in order to avoid the overfitting problem (finding patterns within the sample data that don't generalise well to novel data). However, the most informative variables must be chosen on a separate training slice of the data, otherwise we again may find variables that are randomly correlated with the class (or target) variable simply because there are so many variables to choose from. ODM facilitates this procedure via the "Preprocess" widget. This widget must be joined directly to the "Test and Score" widget so that the data is not pre-processed before it is split into test and training slices. The same rule applies to other pre-processing of data e.g. hyperparameter fine-tuning and model learner selection.

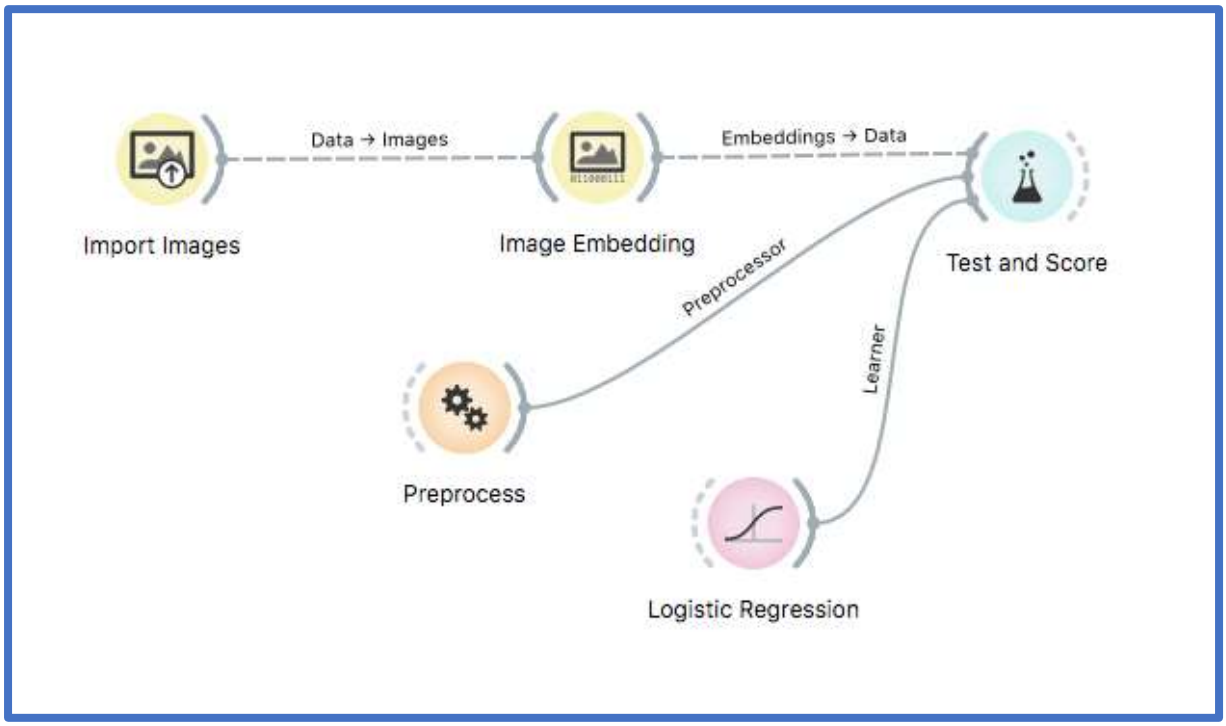

Figure 6.3 Illustration of the correct order for pre-processing within an Orange Data Mining workflow
Example of how to create a workflow that does not pre-process the data prior to splitting into test and training data. Note how the ODM Preprocess widget is joined directly to the Test and Score widget.

A similar problem arises when visualising data analysed by machine learners. Again, the researcher must split the data using one part to select the most informative variables and then visualise the remaining data using the selected variables. If we select variables based on the whole dataset, we may inadvertently select those that are highly correlated with the class (or target) variable simply by chance.(192)

Pre-processing 3. *The importance of not testing on training data*

A machine learning algorithm is a computational method based on statistics, implemented in software,(193) and has the ability to rapidly analyse large data sets in a way that a human researcher never could, formulating hypotheses capable of making reliable predictions based on the hidden patterns detected.(192-194) However, this strength is also a weakness, as computers can also effectively memorise datasets, giving rise to algorithms that do not generalise well to novel data.(194)

To avoid machine learners memorizing data the researcher must ensure that the data selected for training purposes is not the same data used for testing the hypotheses created (in a similar way to the selection of the most informative variables as described above). This ensures that the hypothesis is applicable to novel data. The “Test and Score” widget facilitates this process, as seen in Figure 6.3 below. This widget effectively carries out “k-Cross Validation”, whereby a randomly selected proportion or slice of the data is used to build a hypothesis (training), with the rest of the data tested on this model (testing). Subsequently the next slice of data is used for training, and the remainder again used for testing purposes. This process continues until all the data has been used for training the model. By this method, memorisation of the data by the machine learner is minimised.

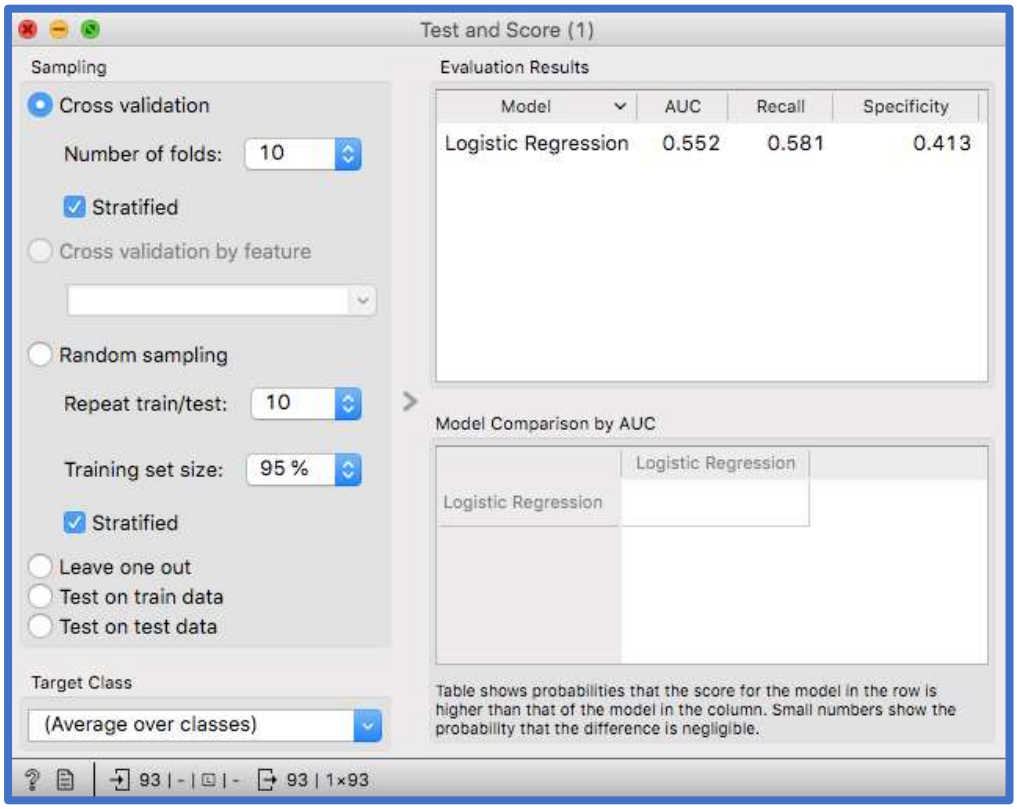

Figure 6.4 Illustration of 10-Cross fold validation within the Orange Data Mining Test and Score widget

Pre-processing 4. *The imbalanced data problem*

It can be problematic to train a model learner to be able to predict both positive and negative data instances if there is a large difference in the proportions of positive and negative instances within the data. In this study, the proportion split was approximately 29% abnormal FAF and 71% normal FAF, and therefore this data imbalance was considered to be mild. If there was a large difference in proportions, it is possible to fix the problem by manipulating the training set in order to over-represent the lower proportion. It is also possible to avoid misleading results due to the imbalanced data problem by using the Matthews correlation coefficient when measuring model learner prediction performances, which takes account of the ratios within the confusion matrix.(193)

6.8 Results

For patient demographics see Table 4.2 in Chapter 4.

6.8.1 Results from the Orange Data Mining Distribution and Boxplot widgets

The data was explored firstly via the ODM Distribution and Boxplot widgets, looking for general trends within the data.

The example below in shows that as age increases, the probability of a normal FAF decreases, whilst the probability of an abnormal FAF increases.

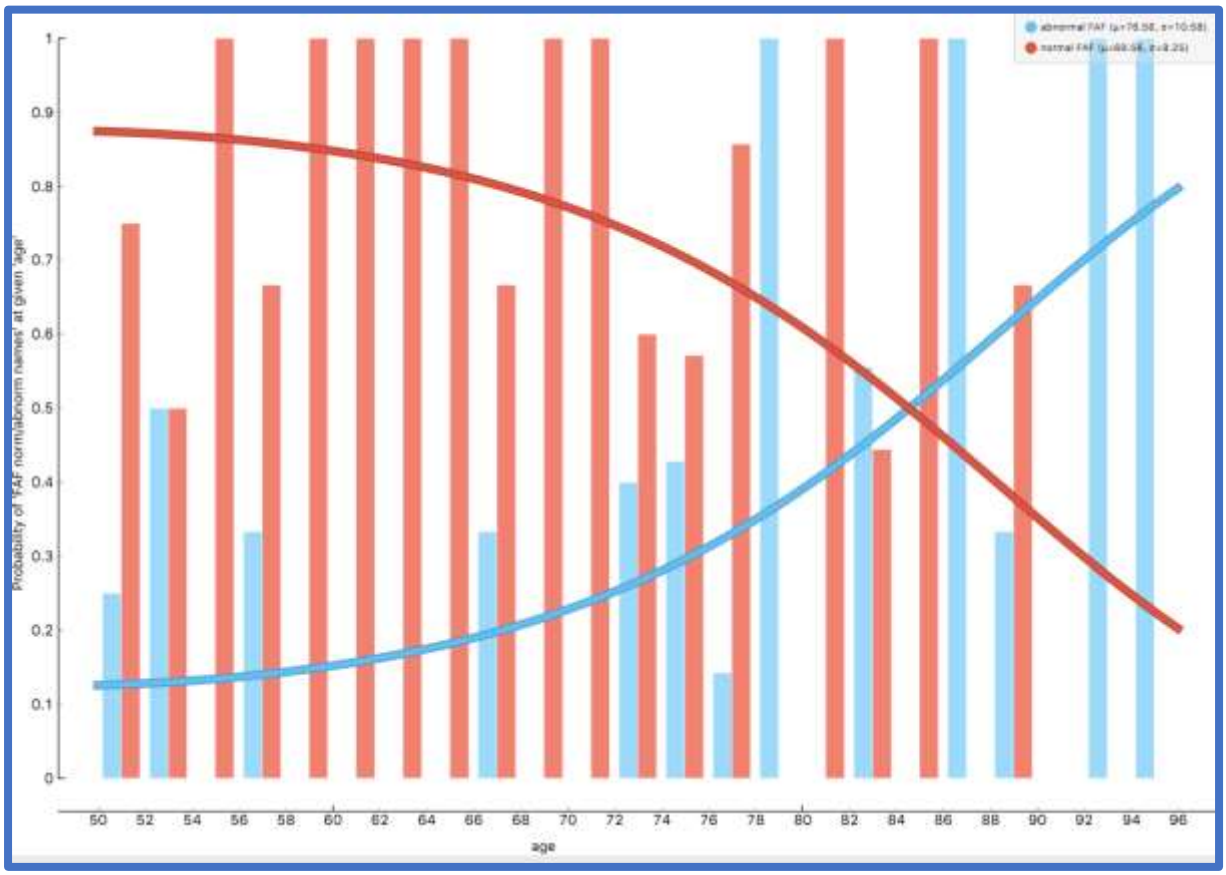

Figure 6.5 Distribution of Age versus probability of FAF normality/abnormality

Image taken from ODM showing age against the probability of a normal/abnormal FAF. The blue bars represent an abnormal FAF, and red bars a normal FAF.

The finding above is supported by the results of the Boxplot analysis shown below, which confirms that there is a statistically significant difference between the mean age of participants with normal and abnormal FAF result to the 1% level (p = 0.005).

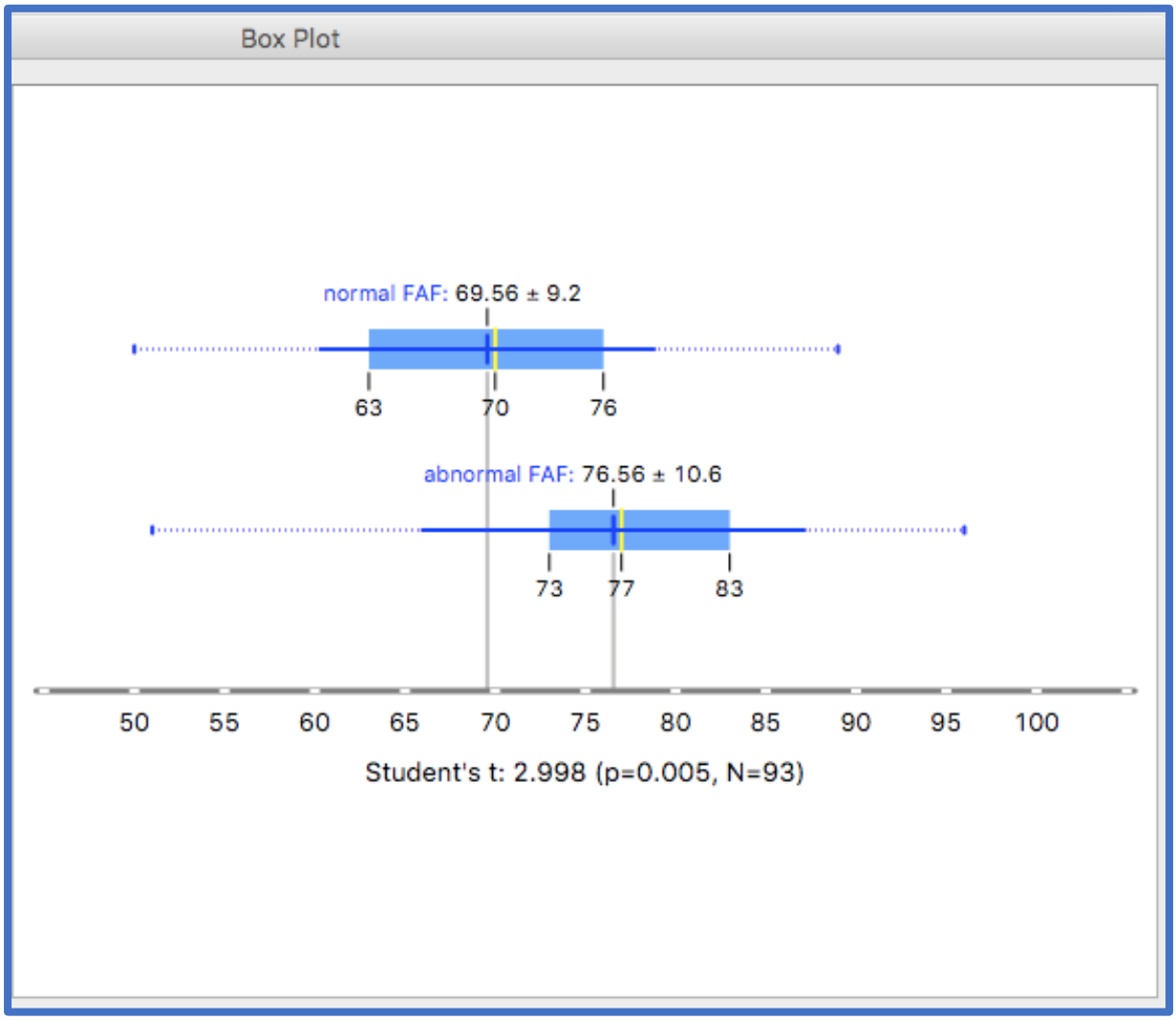

Figure 6.6 Orange Data Mining Boxplot of age versus FAF normality/abnormality.

ODM Boxplot widget showing the results of a Student's t test for a normal/abnormal FAF result against patient age (along the bottom).

The refractive state of the eye, as measured by the Gullstrand refraction function of the OCT, shows that this variable has very little influence over whether the FAF is normal or abnormal.

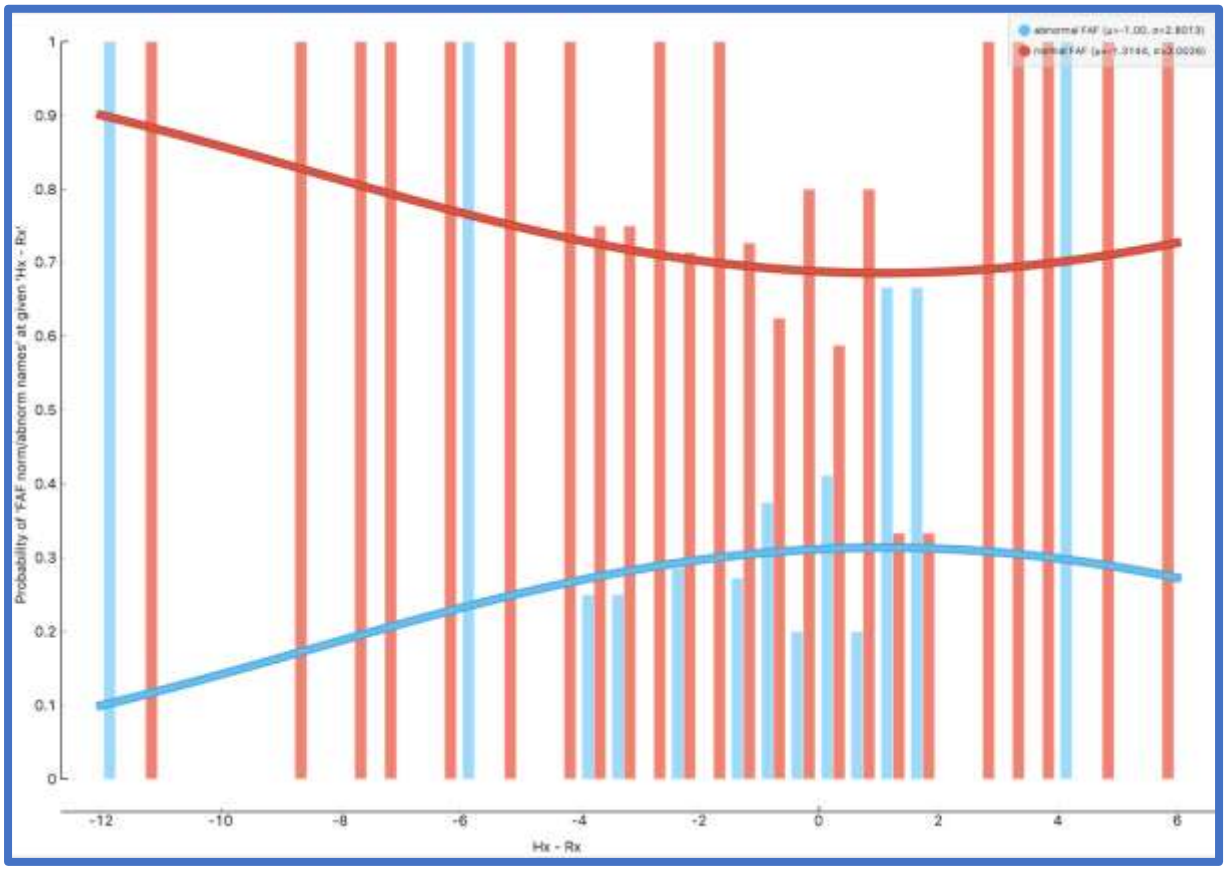

Figure 6.7 Distribution of refractive status and normality/abnormality of FAF

Image from ODM showing Gullstrand refraction (along the bottom) against the probability of a normal or abnormal FAF. The blue bars represent an abnormal FAF, and the red bars represent a normal FAF.

The finding above is supported by the results of the Boxplot analysis shown below, which confirms that there is no statistically significant difference between the mean refractive state as measured by the OCT Gullstrand refraction function of participants with a normal/abnormal FAF result.

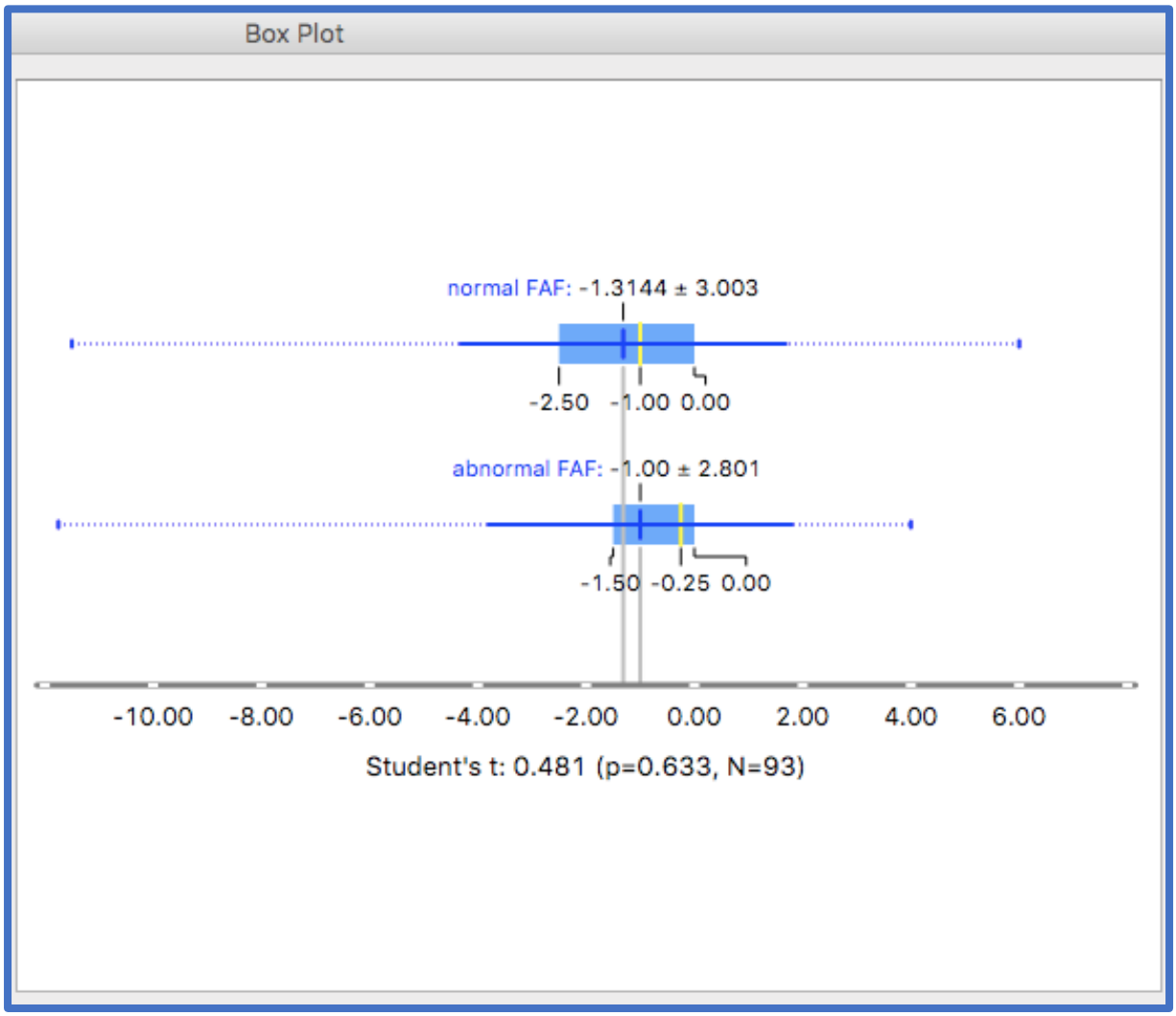

Figure 6.8 Boxplot of refractive status and FAF normality/abnormality

ODM Boxplot widget showing the results of a Student's t test for a normal for abnormal FAF result against the OCT Gullstrand refraction (along the bottom).

From the Figure below, it can be seen that as the number of packet years (number of packets per day times the number of years of smoking) increases, the probability of an abnormal FAF increases.

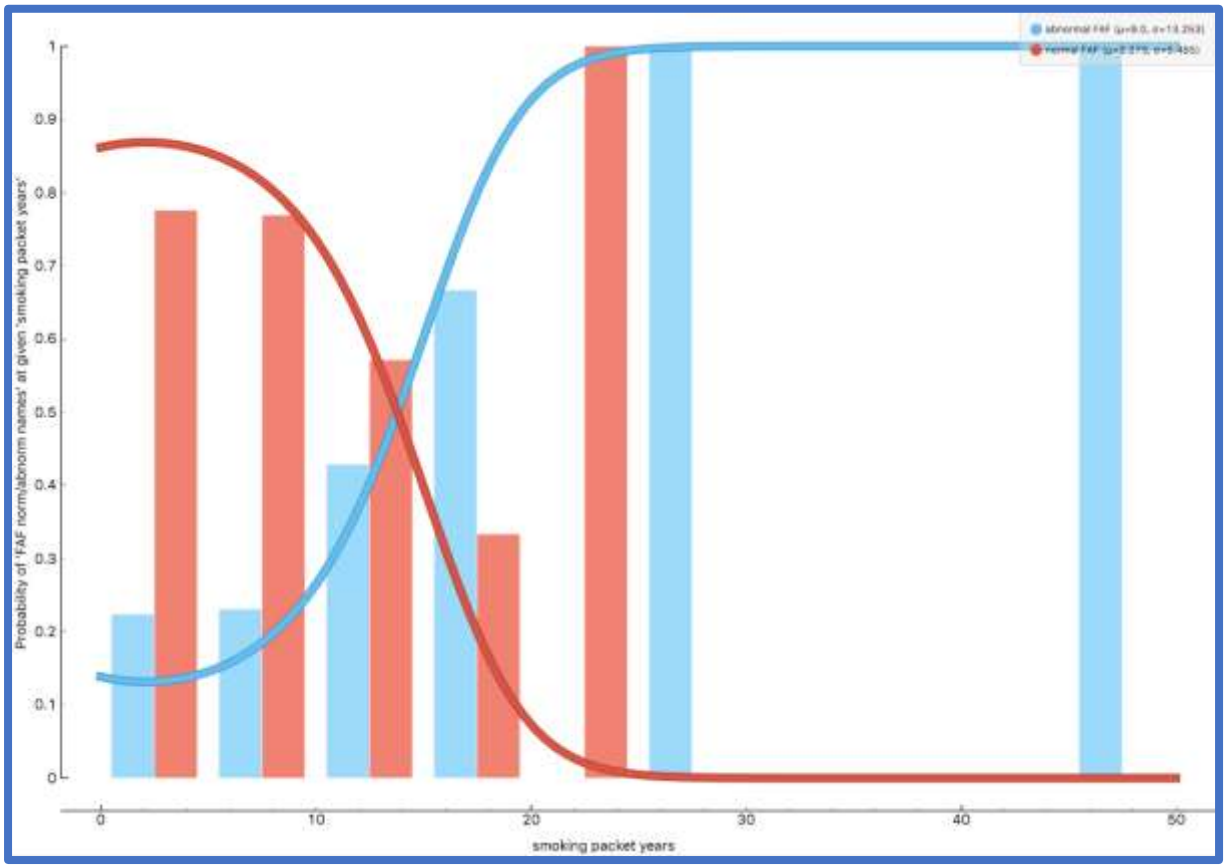

Figure 6.9 Distribution of smoking packet years and probability of normal/abnormal FAF

Image from ODM showing smoking packet years against the probability of a normal/abnormal FAF result. The blue bars represent an abnormal FAF result and the red bars a normal FAF.

The Figure below confirms that there is a statistically significant difference between the mean number of smoking packet years for participants with a normal/abnormal FAF result to the 5% level (p = 0.038), with a higher mean number of packet years seen with an abnormal FAF result.

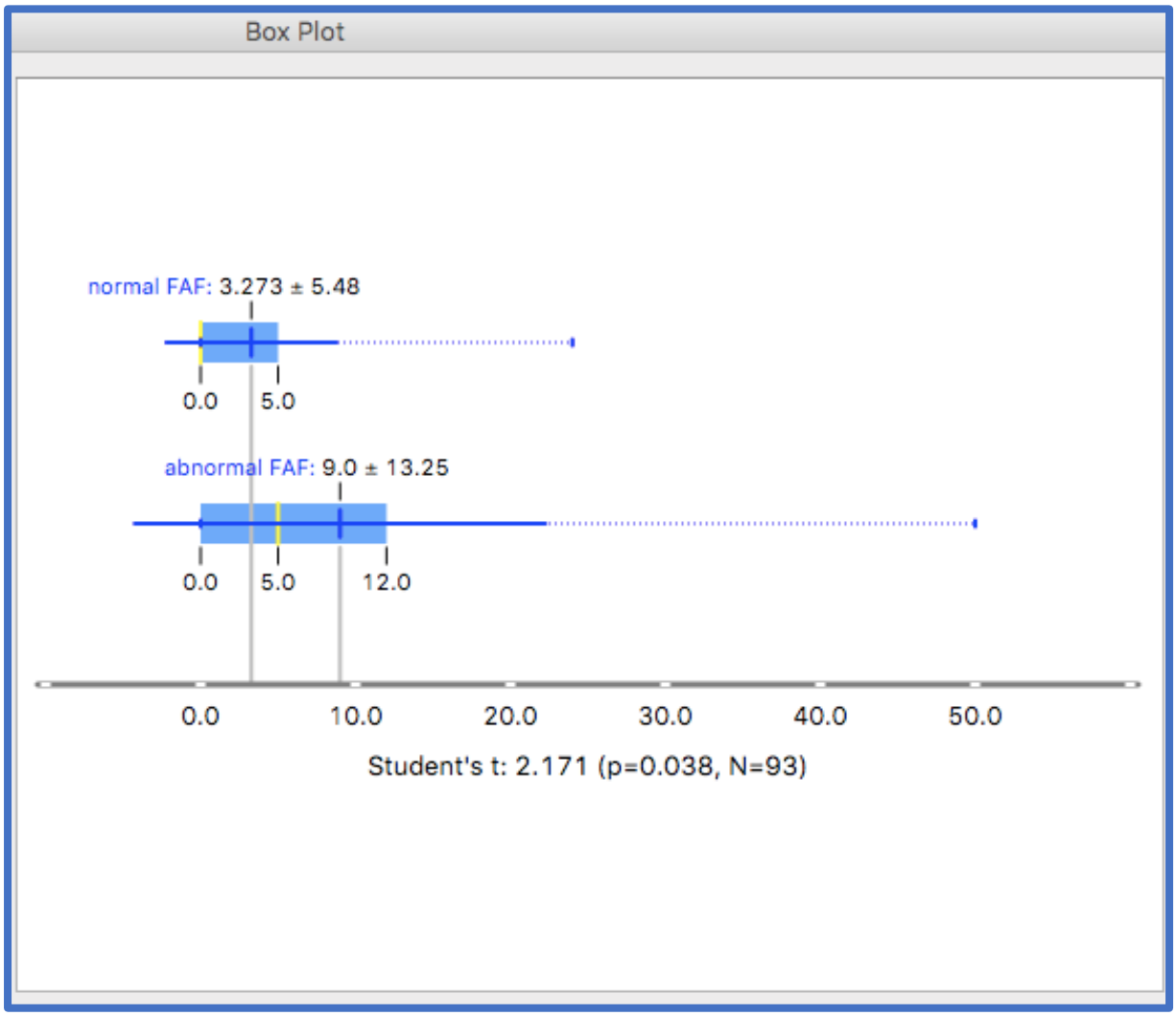

Figure 6.10 Boxplot of smoking packet years and normality/abnormality of FAF

ODM Boxplot widget showing the results of a Student's t test for normal and an abnormal FAF result against smoking packet years (along the bottom).

However, age could be a confounding factor, as older patients may have smoked more in their lives simply as a function of their age, and are more likely to have smoked in their youth due to the relative popularity of smoking in days gone by. Therefore, a sub-group of 24 participants, composed of two age matched groups of ex-smokers were created, with the group for those with a normal FAF result having a mean age of 76.33 years, and those with an abnormal FAF result a mean age of 76.53 years. This time, no significant difference was found (p=0.759) between the number of packet years smoked for the two groups, as shown in the Figure below.

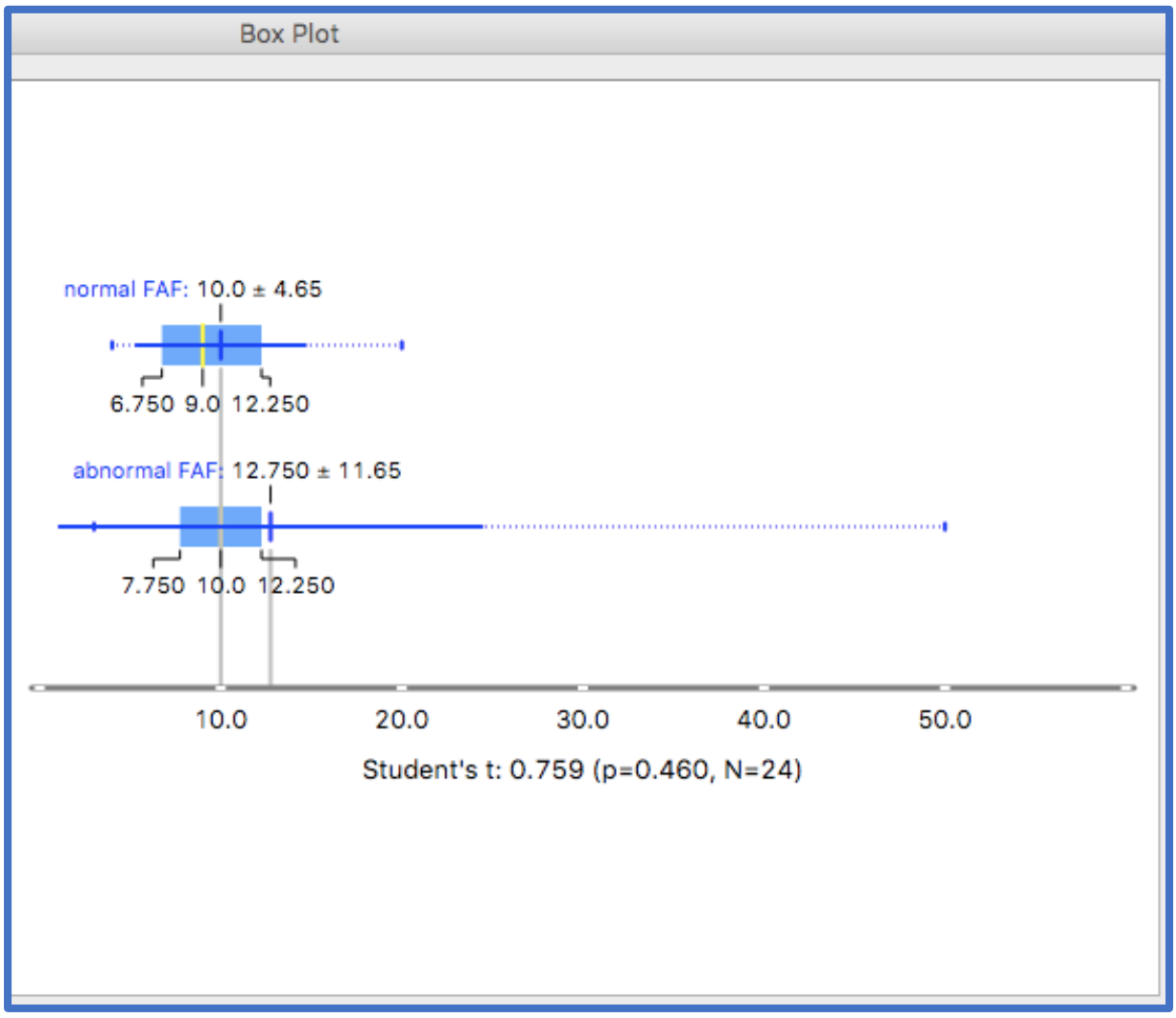

Figure 6.11 Boxplot of smoking packet years and normality and abnormality of FAF

ODM Boxplot widget showing the results of a Student's t test for normal and an abnormal FAF results against smoking packet years (along the bottom) for a sub-group of ex-smokers composed of two age matched groups (N = 24).

The chart in the Figure below indicates, by subjective comparison, that there is no significant difference in the probability of an abnormal FAF when there is/is not a first-degree family history of ARMD, using the entire sample.

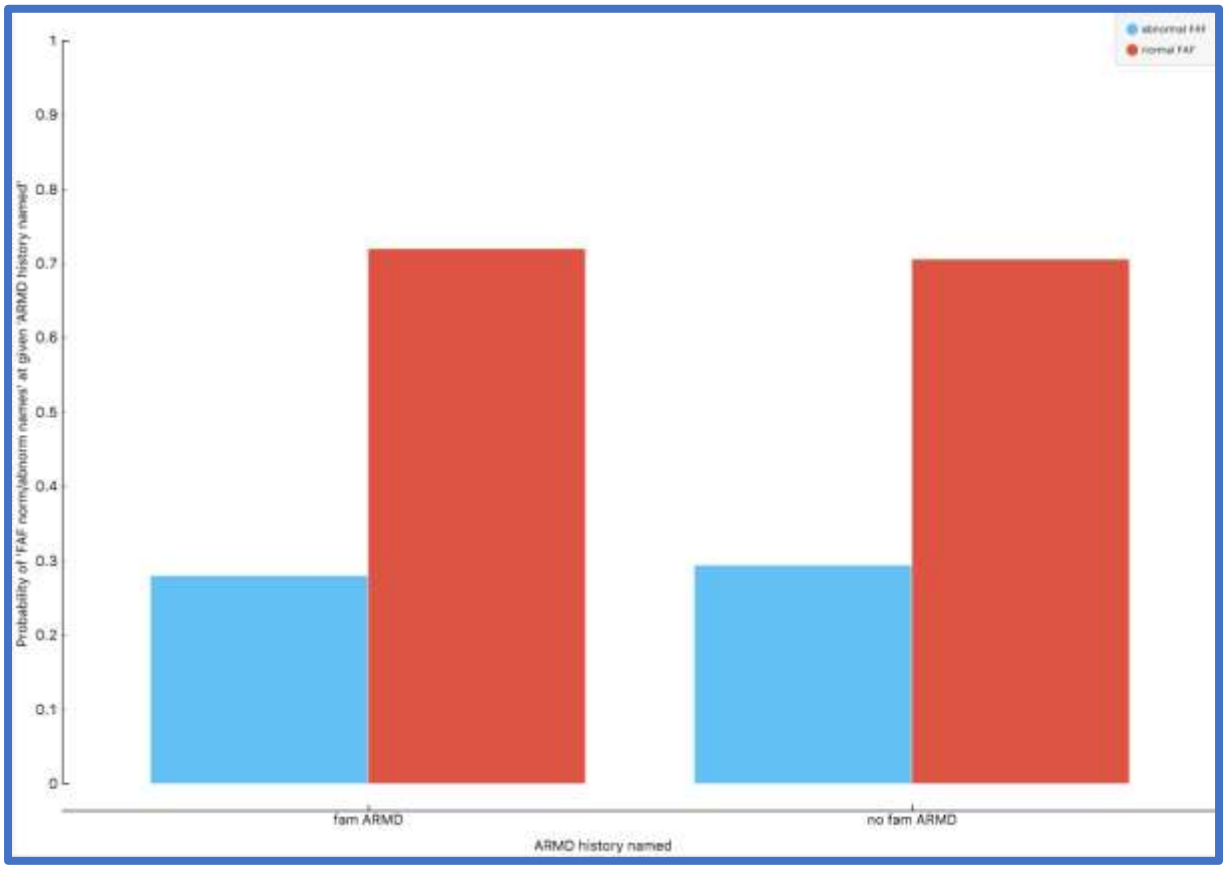

Figure 6.12 Distribution of family history of ARMD and probability of normal/abnormal FAF

Image from ODM showing the presence of a first-degree family history of ARMD against FAF probability. The blue bars represent an abnormal FAF and red bars a normal FAF. Group 1 on the left represents a positive family history of ARMD, and group 2 on the right represents a negative family history of ARMD.

The Figure shows that the proportions of normal and abnormal FAF results are not significantly different for participants who did and did not have a first degree relative with ARMD.

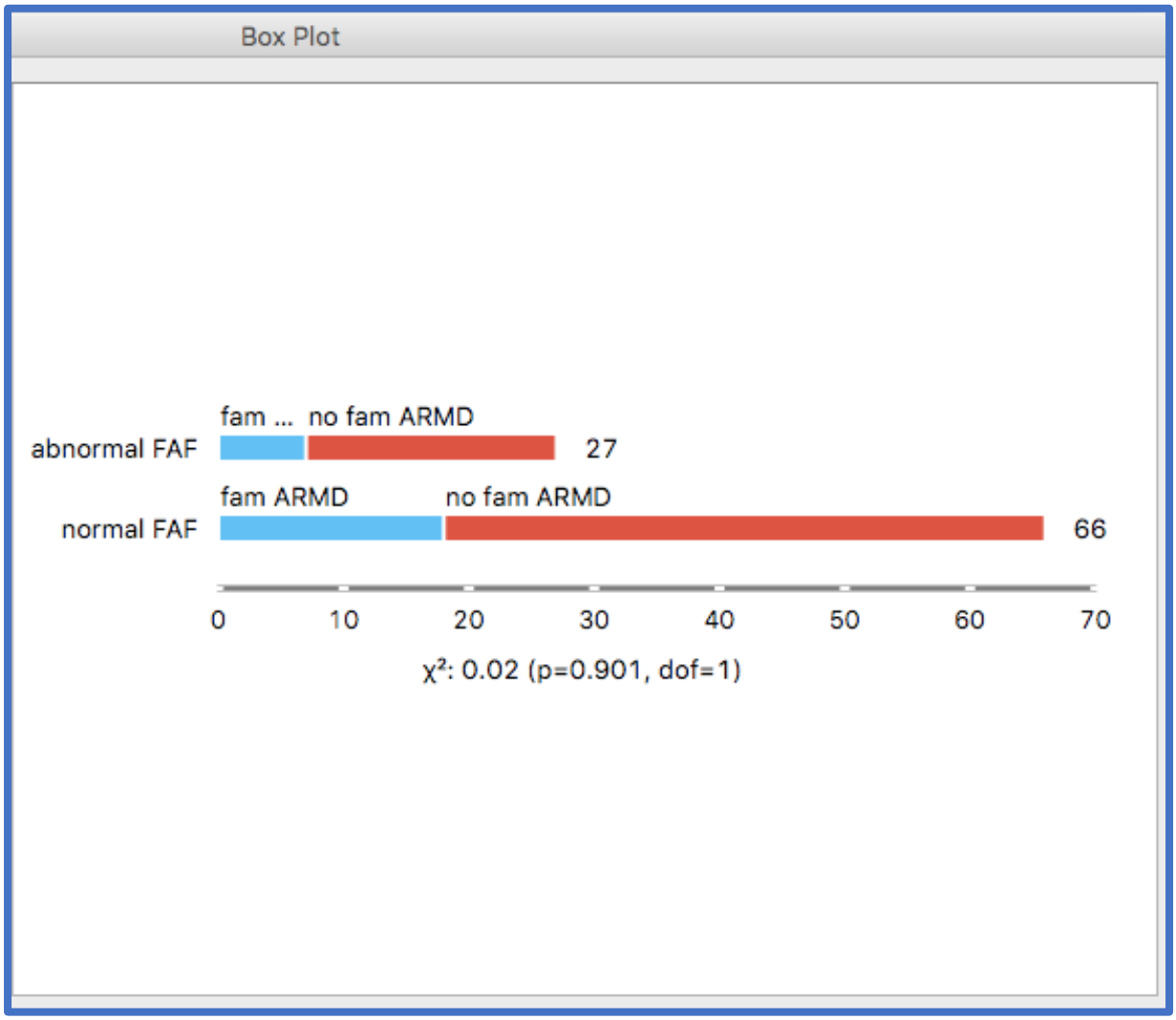

Figure 6.13 Boxplot of family history of ARMD and normality/abnormality of FAF

ODM Boxplot widget showing the results of a Chi squared test for those participants with a first degree relative with ARMD (blue bars) and those without (red bars), for a normal and abnormal FAF result.

The chart below suggests that as the number of large drusen seen by CFP increase, so does the probability of an abnormal FAF. There appears to be a cut-off point around large drusen grade “3”, i.e. 20-29 large drusen. This indicates that if there are 20 or more large drusen then the FAF will likely be abnormal, with a probability of over 90%. Conversely, if there are no large drusen, then the probability of an abnormal FAF is approximately 7%.

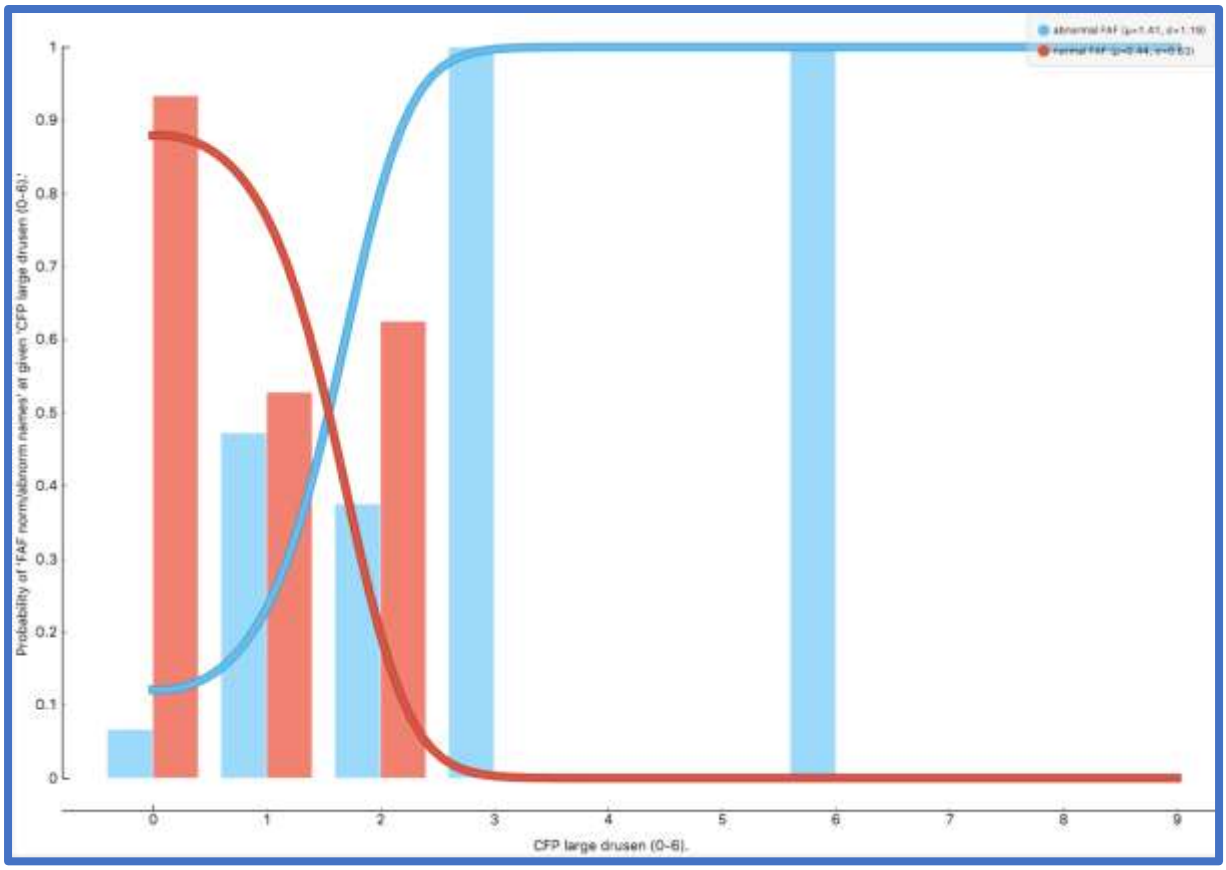

Figure 6.14 Distribution of CFP large drusen and probability of normal/abnormal FAF

Image from ODM showing the number of large drusen detected by CFP versus the probability of an abnormal (blue bars) and a normal (red bars) FAF result.

The Figure below shows that there is a statistically significant difference between the mean grading level (0-6) of large drusen as detected by CFP for participants with a normal (upper group 1) and abnormal (lower group 2) FAF result to the 1% level (p = 0.000).

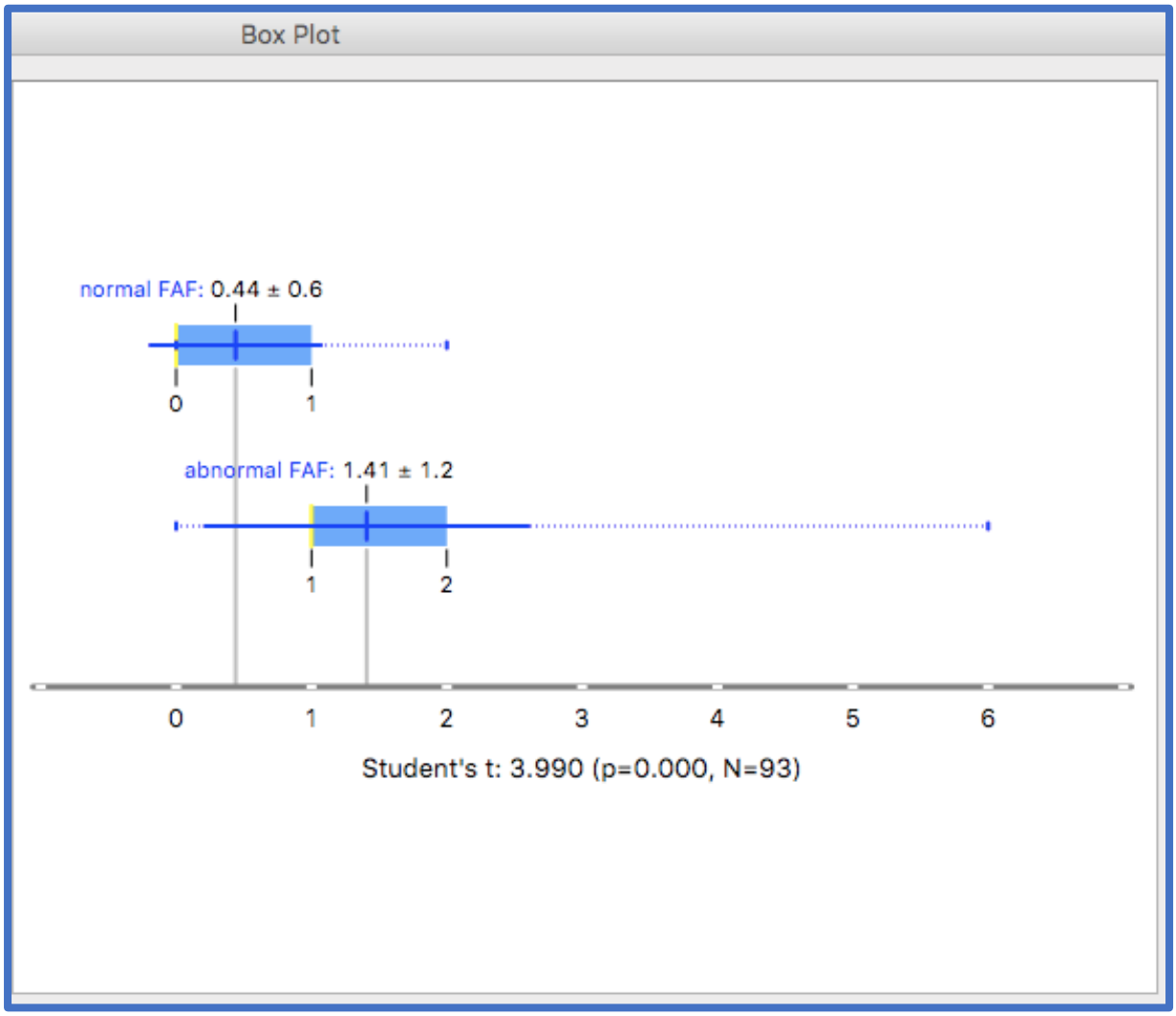

Figure 6.15 Boxplot of CFP large drusen (grade 0-6) FAF normality/abnormality

ODM Boxplot widget showing the results of a Student's t test for participants with a normal and abnormal FAF result against the grading level (0-6) of large drusen as detected with CFP (along the bottom).

Below is a Scatterplot output helping to illustrate the associations between age, CFP large drusen grading score (0-6) and FAF result. The blue dots indicating an abnormal FAF result are clustered in the top right corner of the graph.

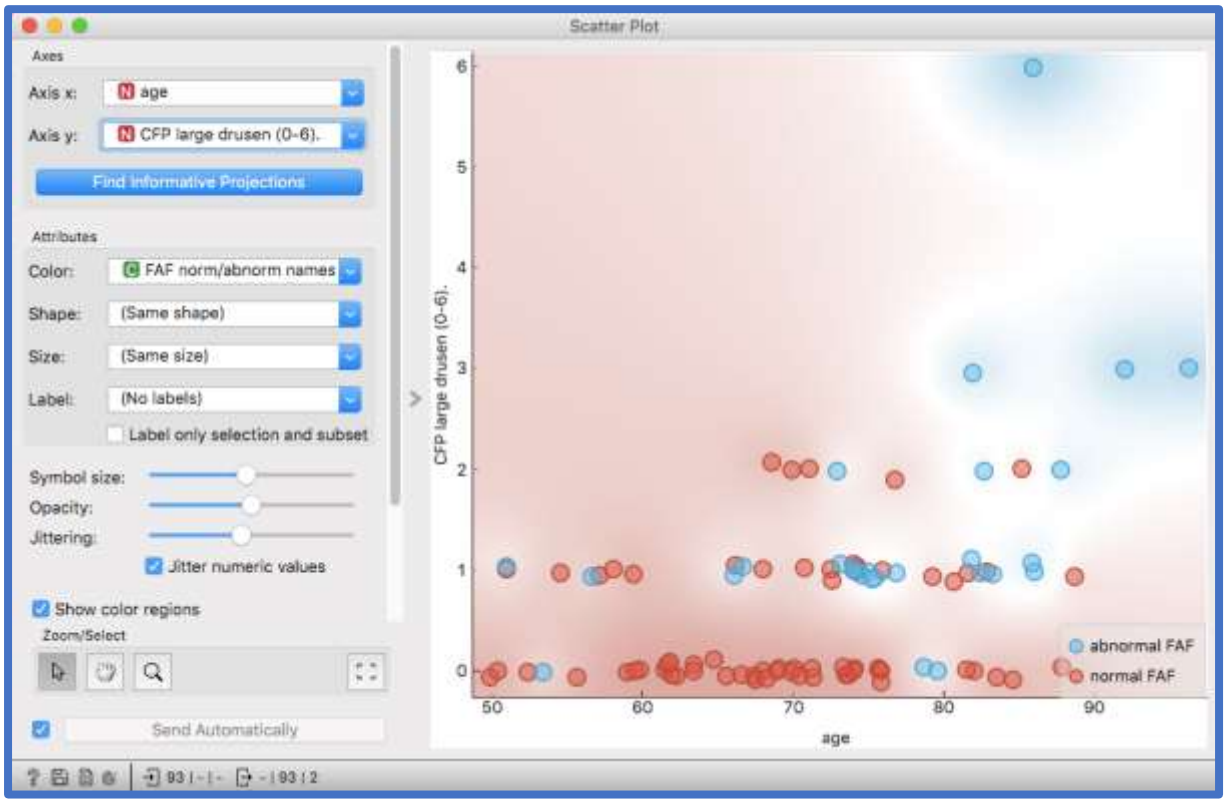

Figure 6.16 Scatterplot of age, CFP large drusen and normal/abnormal FAF

ODM Scatterplot widget output helping to illustrate the relationship between age (x axis) and CFP large drusen score (0-6) (y axis). Blue dots indicate an abnormal FAF and red dots a normal FAF.

For intermediate sized drusen detected by CFP the results are less clear cut, with two instances where the number of intermediate sized drusen are 50 or over in the presence of a normal FAF, however, for grading of drusen numbers from 0-3, there is again the same upward trend in the probability of an abnormal FAF with a greater number of drusen as shown in the Figure below.

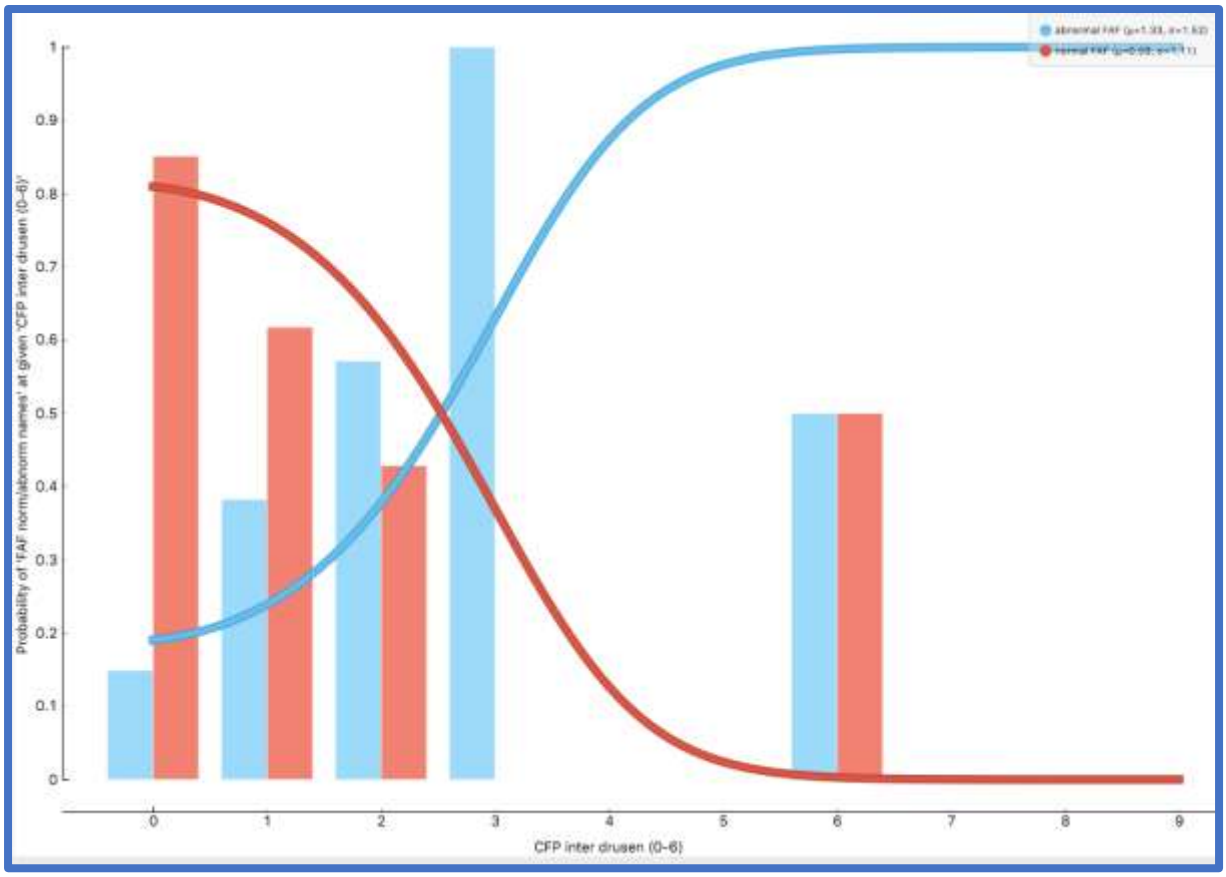

Figure 6.17 Distribution of intermediate drusen and probability of normal/abnormal FAF

Image from ODM showing the grading level of intermediate drusen detected by CFP against the FAF probability of an abnormal (blue bars) and normal FAF result (red bars).

Figure 6.17 shows that there is a statistically significant difference between the mean grading level (0-6) of intermediate drusen as detected by CFP for participants with a normal and abnormal FAF result to the 5% level (p = 0.027).

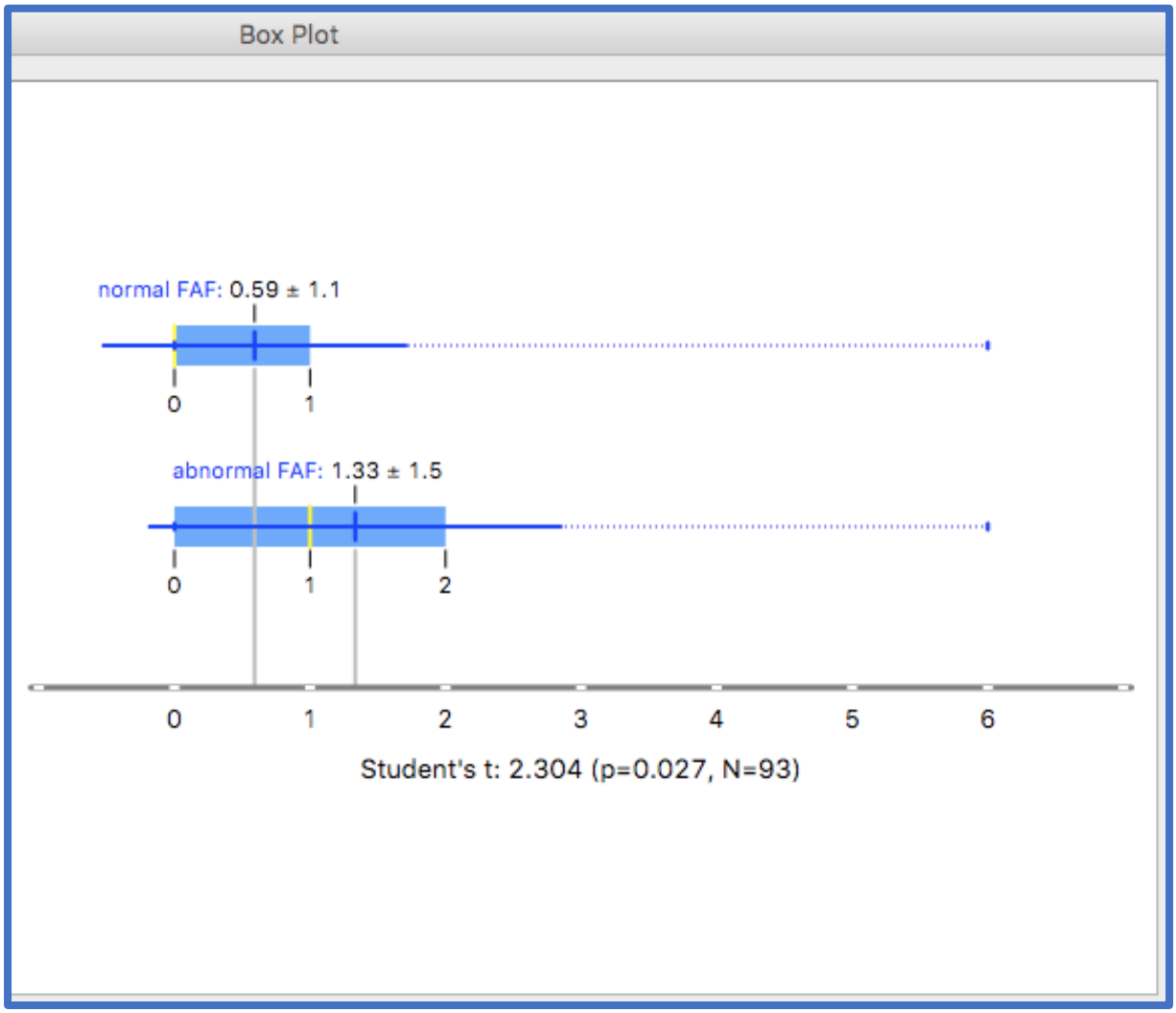

Figure 6.18 Boxplot of CFP intermediate drusen (0-6) and FAF normality/abnormality

ODM Boxplot widget showing the results of a Student's t test for participants with a normal and an abnormal lower FAF result against the grading level (0-6) of intermediate drusen as detected with CFP (at the bottom).

For small drusen detected with CFP there were two instances of the number of drusen as 50 or over graded with a normal FAF. For grading of drusen numbers from 0-3, there is the same upward trend in the probability of an abnormal FAF with a greater number of drusen as shown in the Figure below.

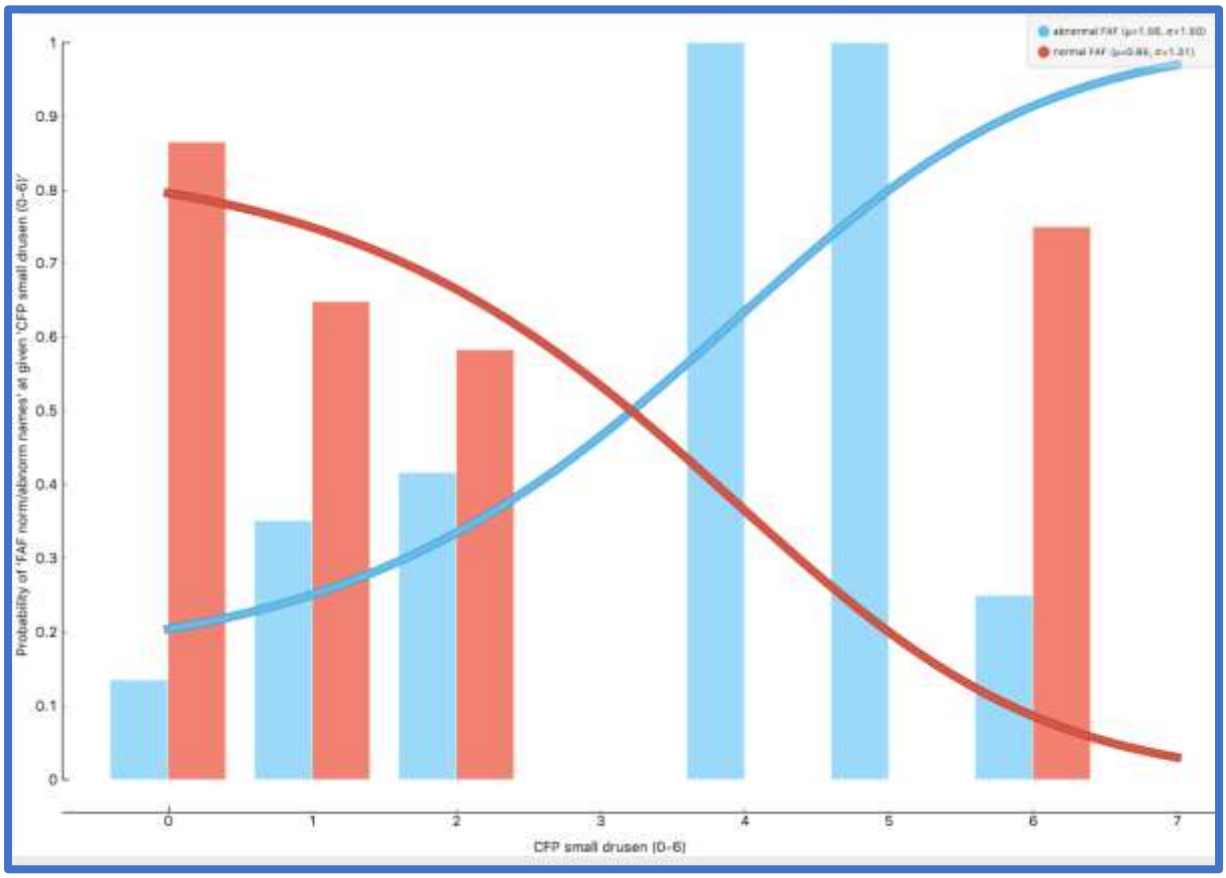

Figure 6.19 Distribution of CFP small drusen (0-6) and probability of normal/abnormal FAF

Image from ODM showing the grading level of small drusen (0-6) detected by CFP against the probability of a normal and abnormal FAF result. The blue bars represent an abnormal and the red bars a normal FAF result.

The Figure below shows that there is a statistically significant difference between the mean grading (0-6) level of small drusen as detected by CFP for participants with a normal and abnormal FAF result to the 5% level (p = 0.038).

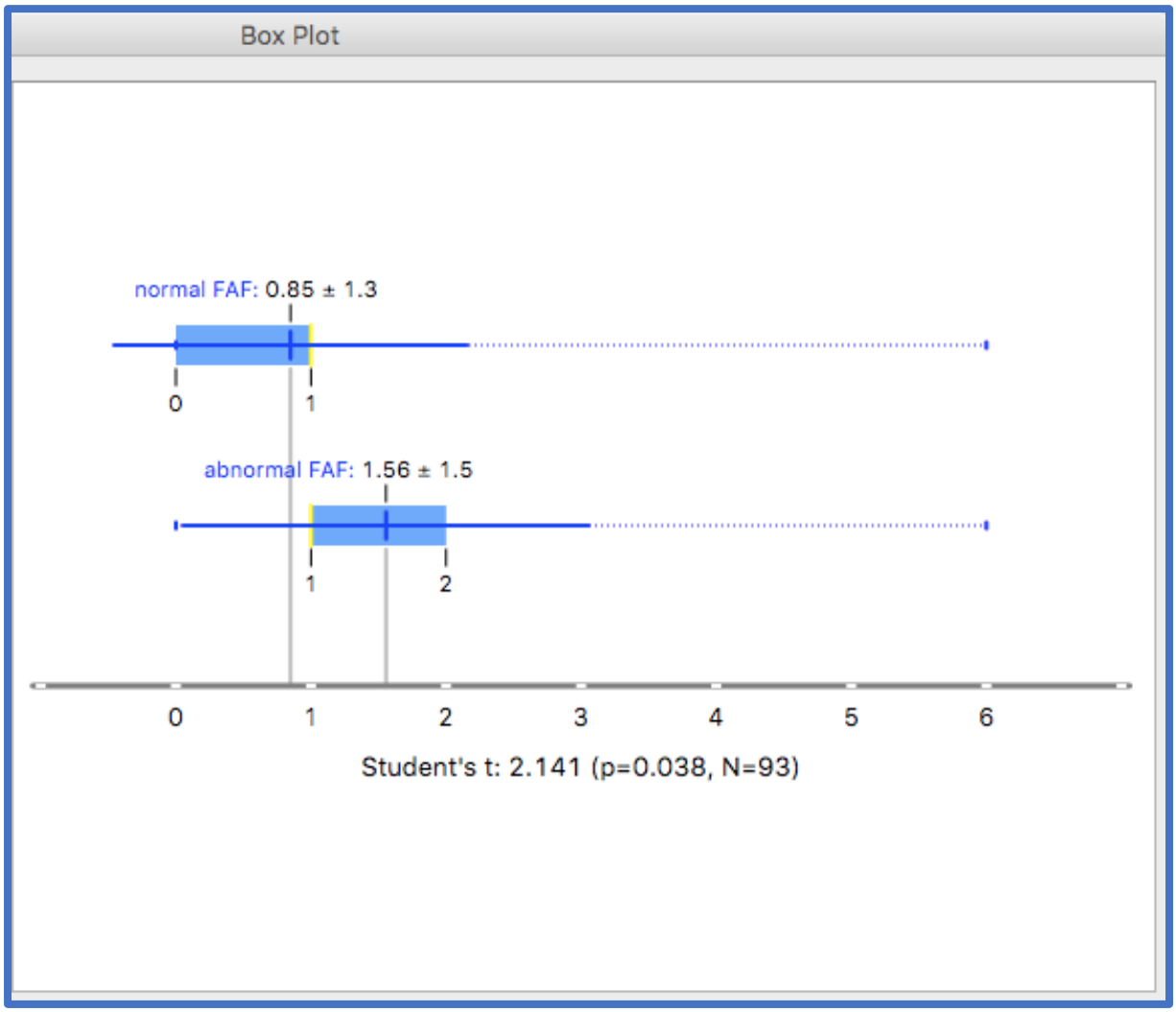

Figure 6.20 Boxplot of CFP small drusen (0-6) and normal/abnormal FAF

ODM Boxplot widget showing the results of a Student's t test for participants with a normal and abnormal FAF result against the grading level (0-6) of small drusen as detected with CFP (along the bottom).

Similar findings were found for large drusen as detected by OCT, i.e. if there were more than 20 large drusen, the probability of an abnormal FAF was high at over 90%, as illustrated in the graph below.

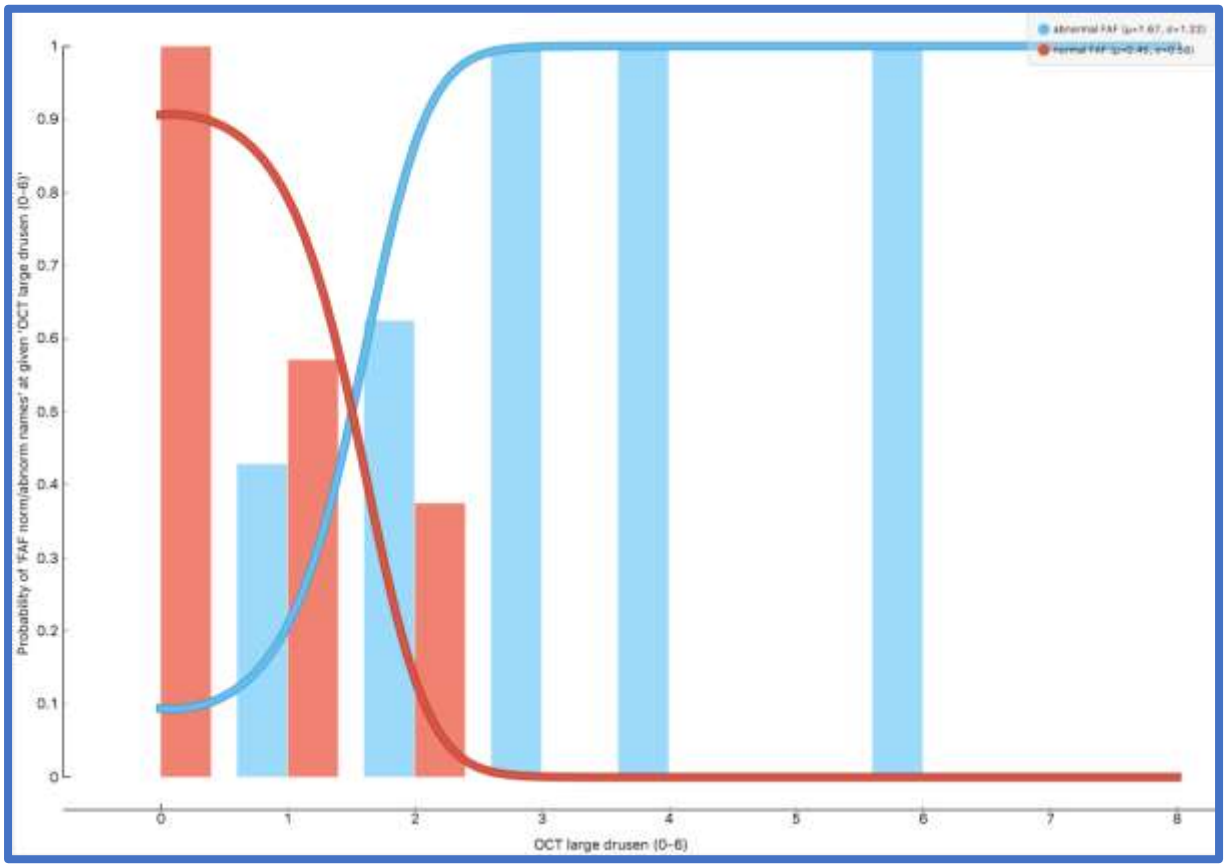

Figure 6.21 Distribution of OCT large drusen (0-6) and probability of normal/abnormal FAF

Image from ODM showing the grading level of large drusen (0-6) detected by CFP against the probability of a normal and abnormal FAF result. The blue bars represent an abnormal and the red bars a normal FAF result.

The Figure below shows that there is a statistically significant difference between the mean grading (0-6) level of large drusen as detected by OCT for participants with a normal and abnormal FAF result to the 1% level (p = 0.000).

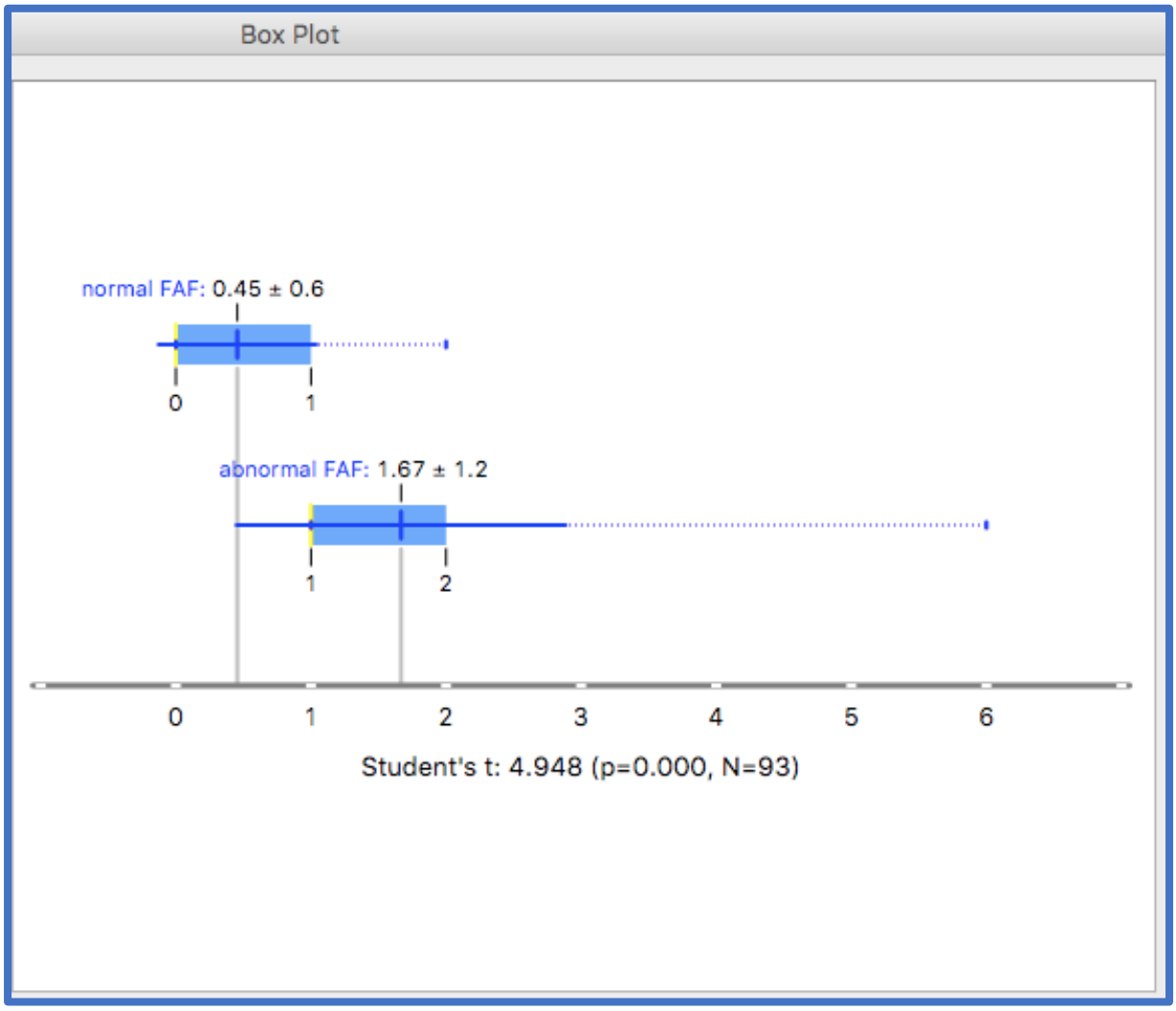

Figure 6.22 Boxplot of OCT large drusen (0-6) and normal/abnormal FAF

ODM Boxplot widget showing the results of a Student's t test for participants with a normal and abnormal FAF result against for the grading level (0-6) of large drusen as detected by OCT (along the bottom).

Below is a Scatterplot output helping to illustrate the associations between age, OCT large drusen score (0-6) and FAF result. The blue dots indicating an abnormal FAF result are clustered in the top right corner of the graph.

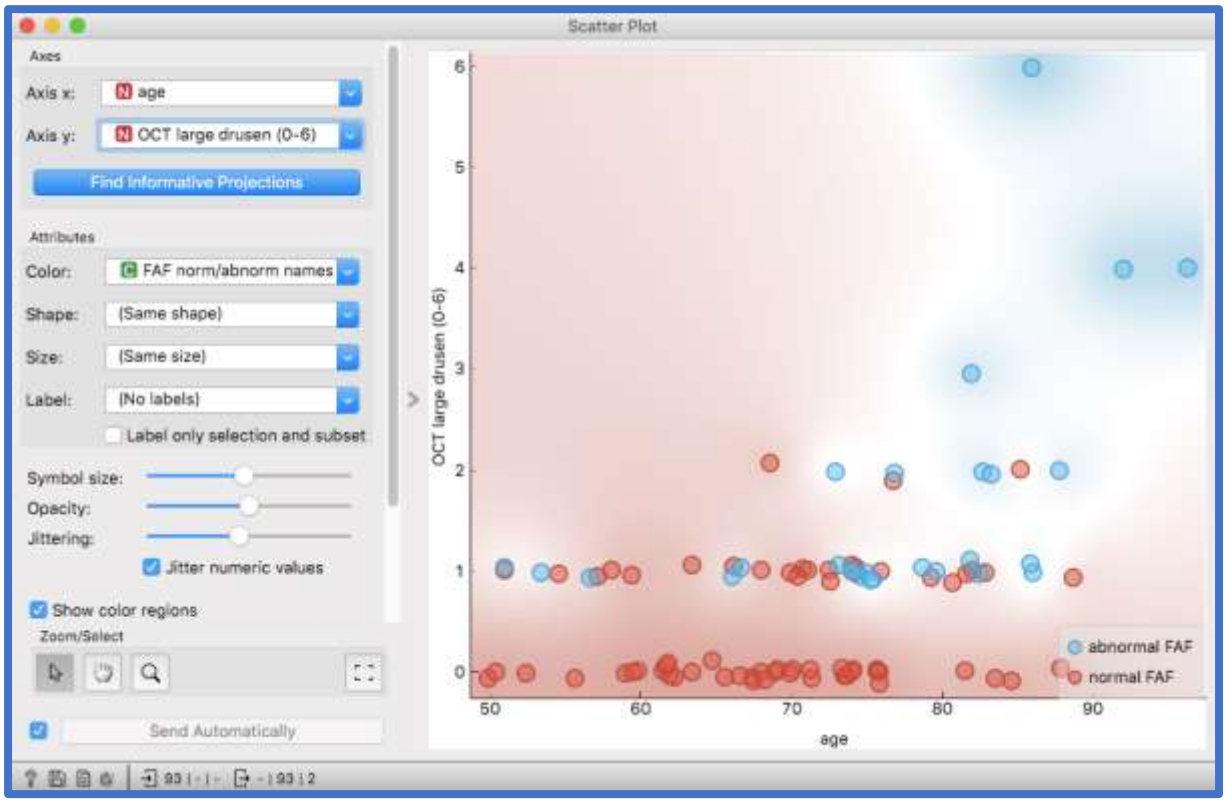

Figure 6.23 Scatterplot of age, OCT large drusen (0-6) and normal/abnormal FAF

ODM Scatterplot widget output helping to illustrate the relationship between age (x axis) and OCT drusen score (0-6) (y axis). Blue dots indicate an abnormal FAF and red dots a normal FAF result.

Similar results for OCT were found as with CFP for intermediate drusen. As with CFP, the general trend is for a higher probability of an abnormal FAF with more intermediate drusen, however, there were two instances of a normal FAF with 50 or more intermediate drusen as shown in the Figure below.

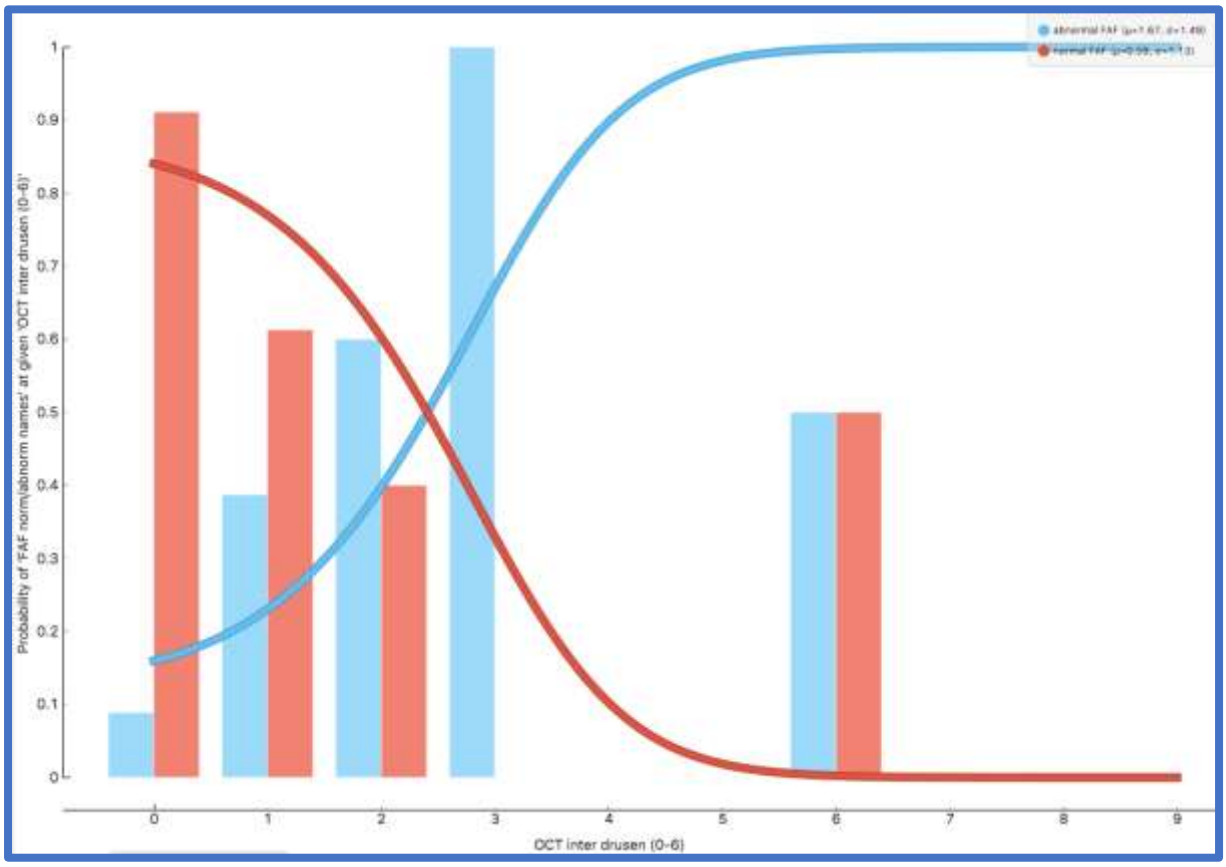

Figure 6.24 Distribution of OCT intermediate drusen (0-6) and probability of normal/abnormal FAF

Image from ODM showing the grading level of intermediate drusen detected by OCT against the probability of a normal and abnormal FAF result. The blue bars represent an abnormal and the red bars a normal FAF result.

The Figure below shows that there is a statistically significant difference between the mean grading level of intermediate drusen as detected by OCT for participants with a normal and abnormal FAF result to the 1% level (p = 0.002).

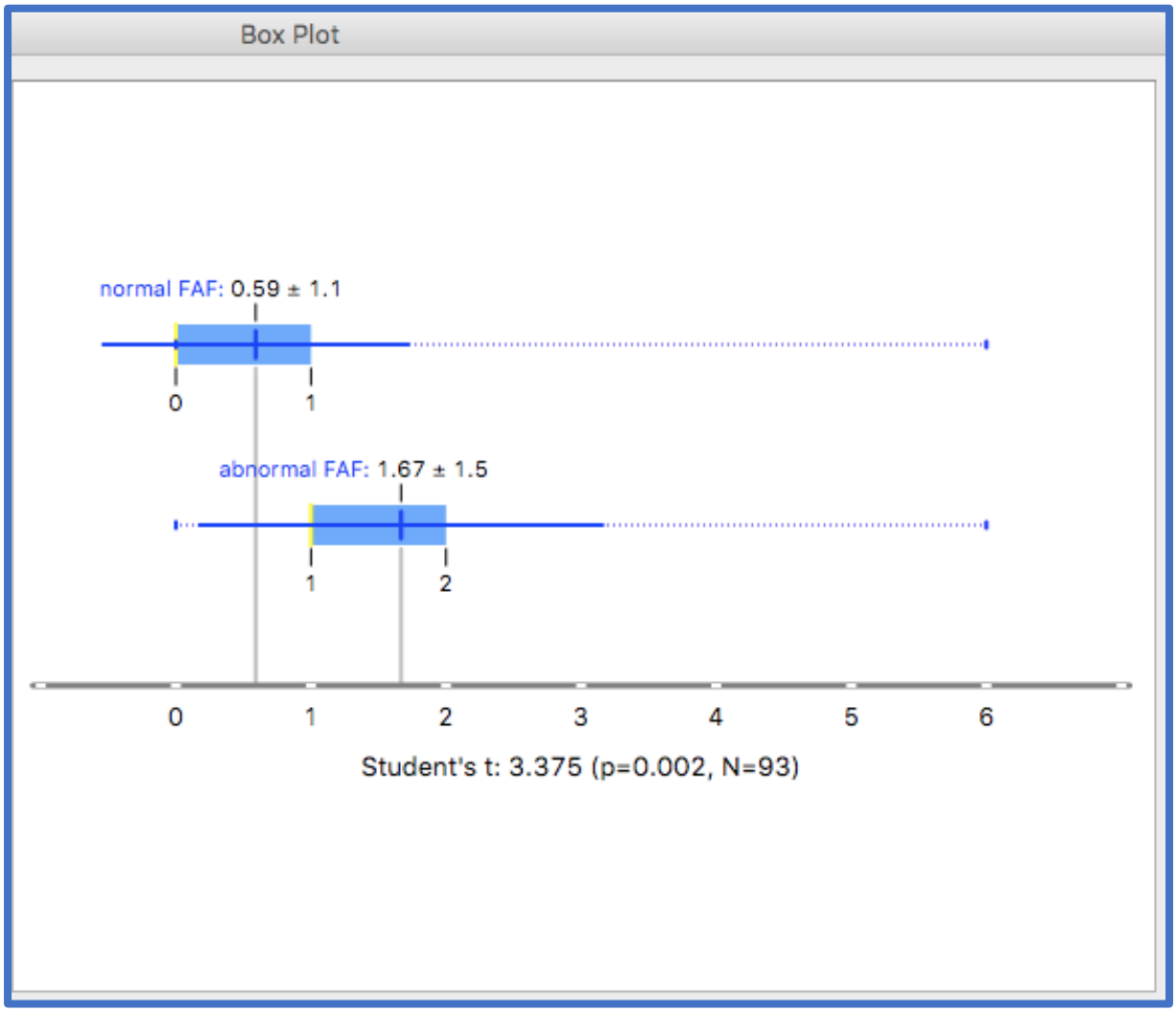

Figure 6.25 Boxplot of OCT intermediate drusen (0-6) and normal/abnormal FAF

ODM Boxplot widget showing the results of a Student's t test for participants with a normal and abnormal FAF result against the grading (0-6) level of intermediate drusen as detected with OCT (along the bottom).

Similar results for OCT were found as with CFP for small drusen. As with CFP, the general trend is for a higher probability of an abnormal FAF with a greater number of small drusen, however, there were two instances of a normal FAF with 50 or more small drusen, and one instance of a normal FAF with between 40 and 49 small drusen as shown below.

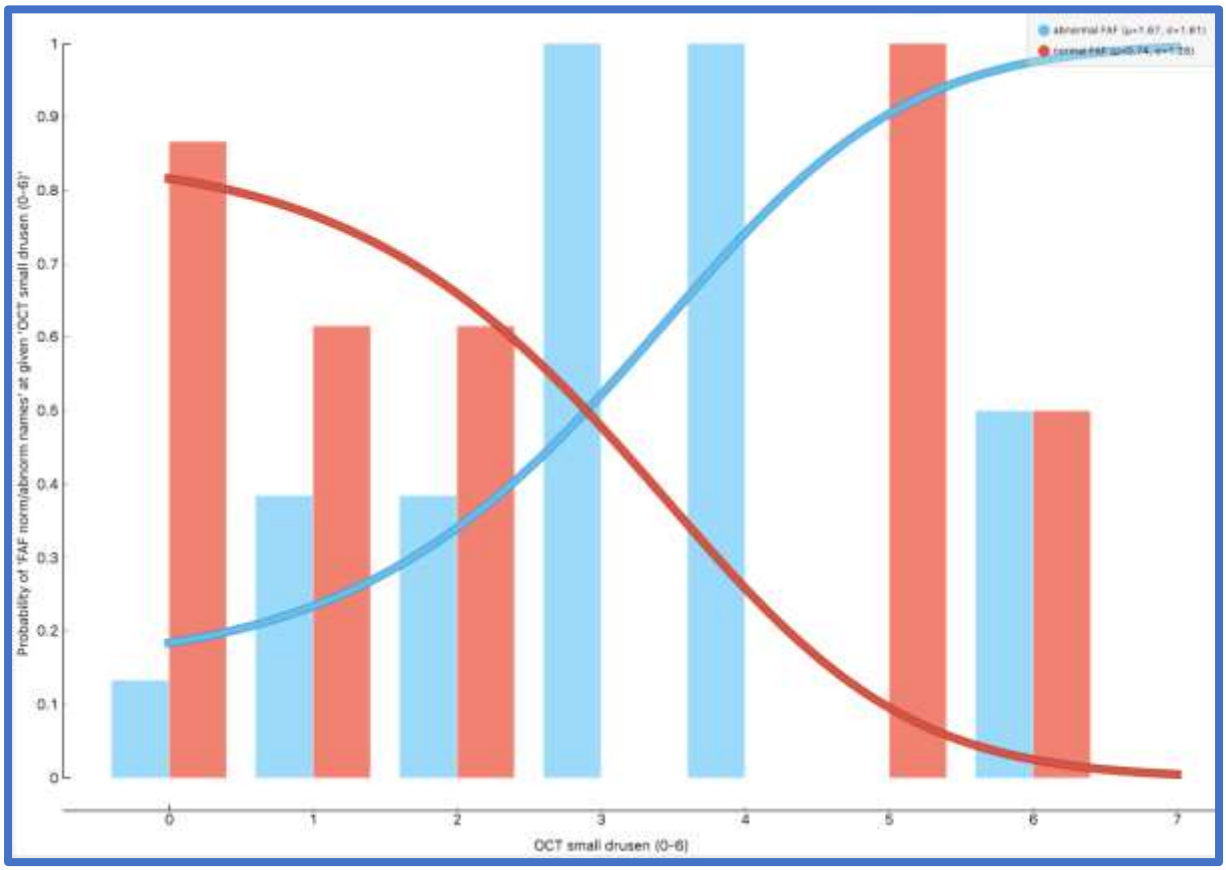

Figure 6.26 Distribution of OCT small drusen (0-6) and normality/abnormality of FAF

Image from ODM showing the grading level (0-6) of small drusen detected by OCT against the probability of an abnormal (blue bars) and a normal (red bars) FAF result.

The Figure below shows that there is a statistically significant difference between the mean grading (0-6) level of small drusen as detected by OCT for participants with a normal and abnormal FAF result to the 5% level (p = 0.011).

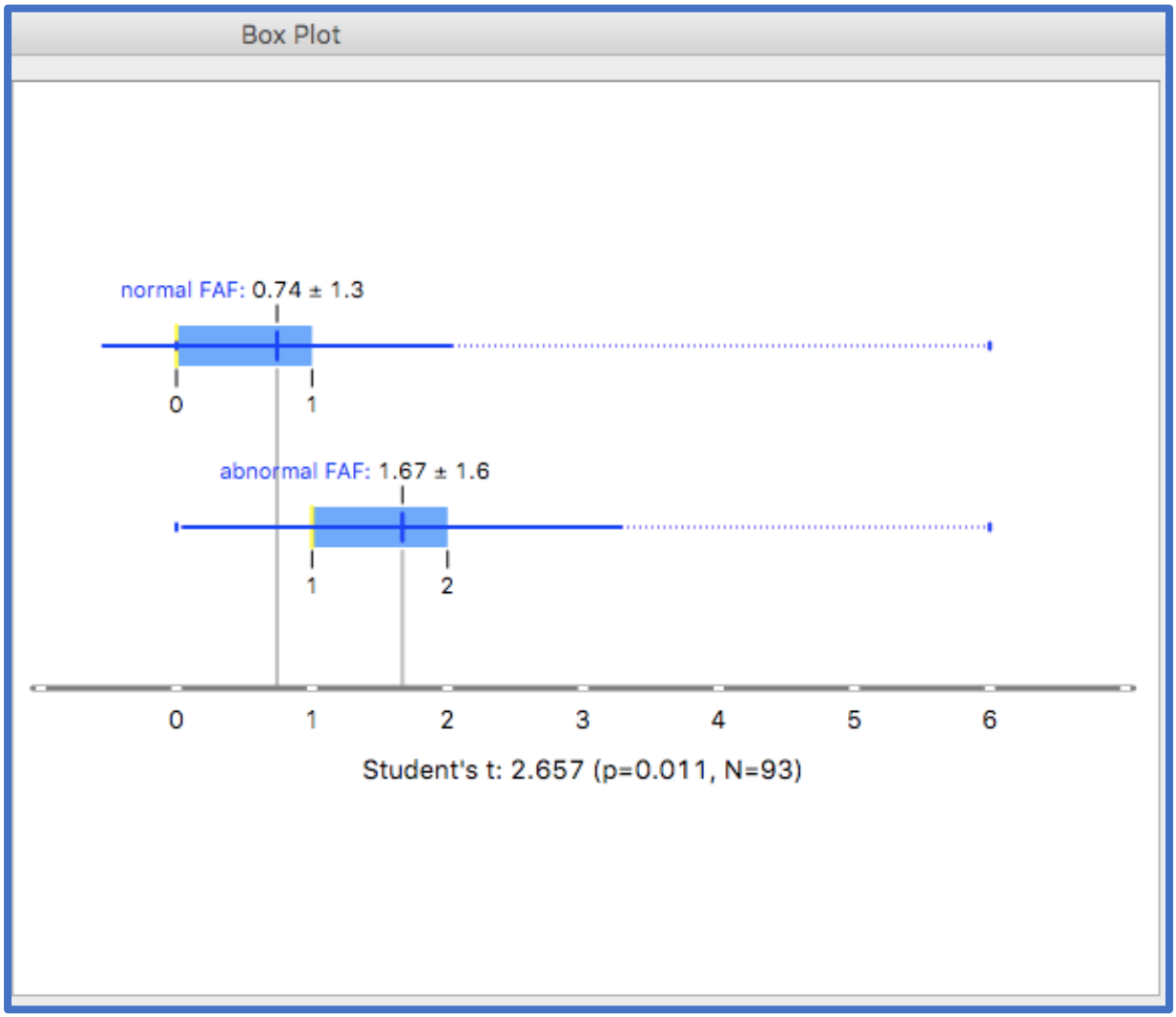

Figure 6.27 Boxplot of OCT small drusen and normal/abnormal FAF

ODM Boxplot widget showing the results of a Student's t test for participants with a normal and abnormal FAF result against for the grading level (0-6) of small drusen as detected by OCT (along the bottom).

The chart below helps to illustrate the correlation between the simplified severity score (SSS) and the probability of an abnormal FAF. The graph shows that there is a cross-over point at the SSS value of “2”, i.e. the data from this project supports the rule that if the SSS value is > 2, then the likelihood of an abnormal FAF result is high, with a probability of over 90%. At a level of SSS = 2, the probability of an abnormal FAF is approximately 50%.

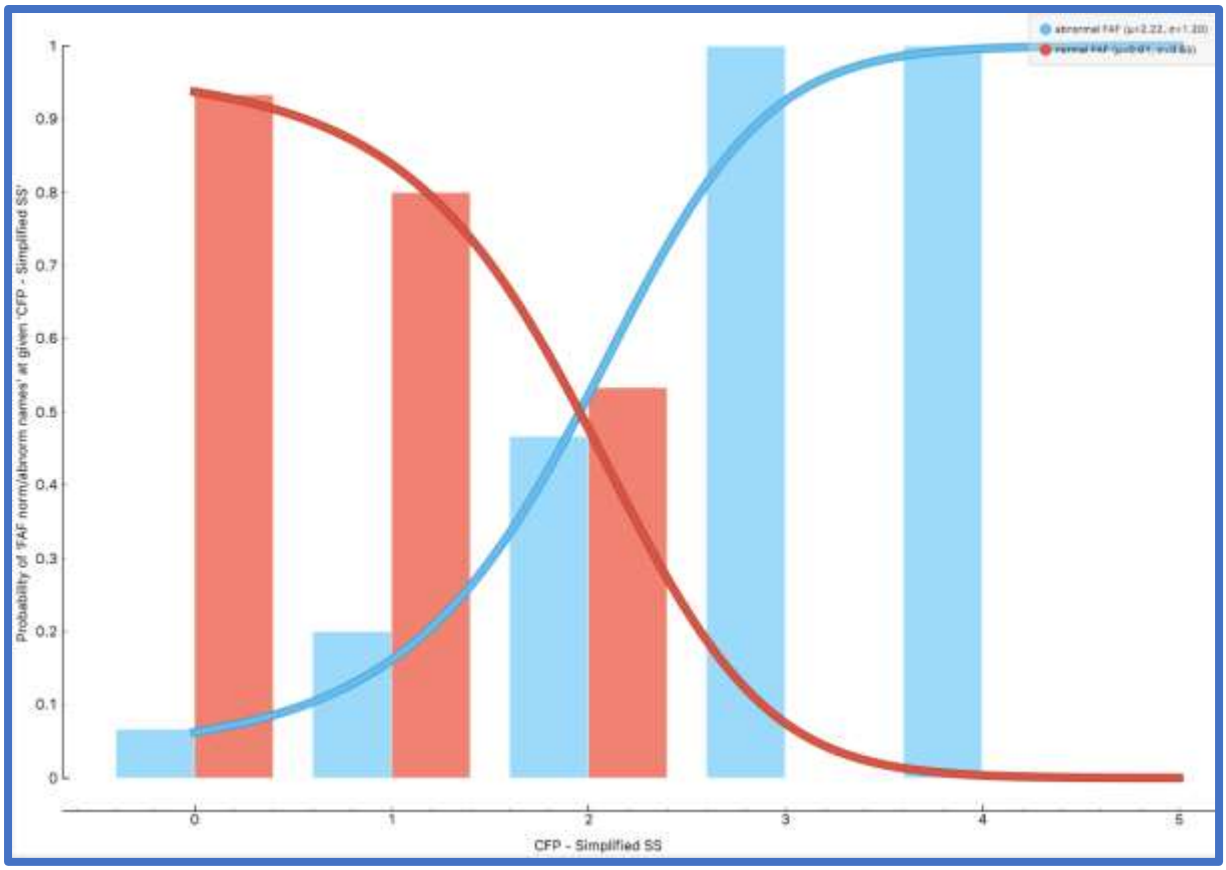

Figure 6.28 Distribution of CFP SSS and probability of normal/abnormal FAF

Image from ODM showing probability of an abnormal (blue bars) and a normal (red bars) FAF result against the simplified severity score as measured by CFP (along the bottom).

The Figure below shows that there is a statistically significant difference between the mean SSS for participants with a normal and abnormal FAF result to the 1% level (p = 0.000). For participants with a normal FAF the mean SSS is 0.61±0.9, and for an abnormal FAF the value is 2.22±1.2.

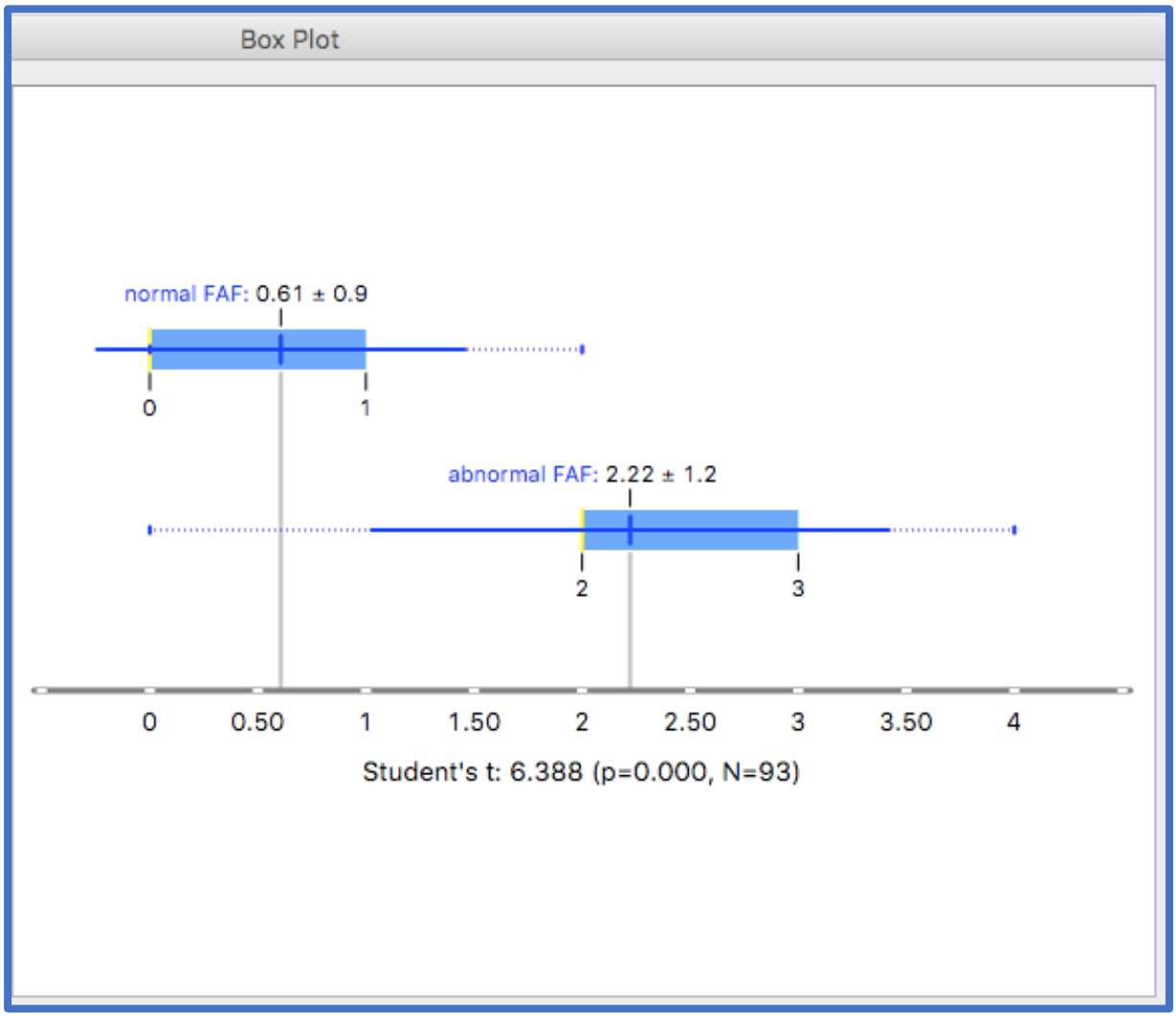

Figure 6.29 Boxplot of SSS and normal/abnormal FAF

ODM Boxplot widget showing the results of the Student's t test participants with a normal and abnormal FAF result against for the simplified severity score (along the bottom).

Below is a Scatterplot output helping to illustrate the associations between age, simplified severity score and FAF result. The blue dots indicating an abnormal FAF result are clustered in the top right corner of the graph.

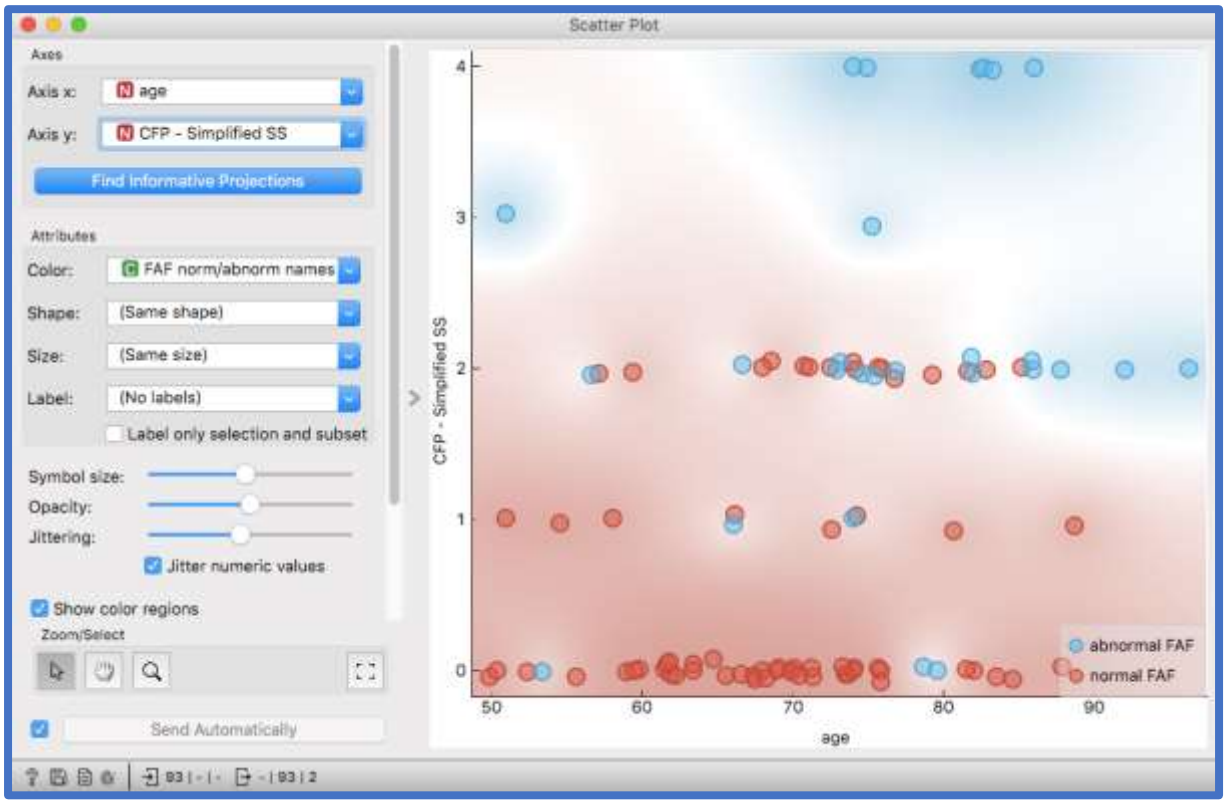

Figure 6.30 Scatterplot of age, CFP SSS and normal/abnormal FAF

ODM Scatterplot widget output helping to illustrate the relationship between age (x axis) and simplified severity score (y axis). Blue dots indicate an abnormal, and red dots a normal FAF result.

The chart below shows that pigmentary abnormalities detected by CFP are closely associated with an abnormal FAF. If pigmentary abnormalities are present, then the probability of an abnormal FAF is over 90%. Conversely if pigmentary abnormalities are absent, then the probability of an abnormal FAF falls to approximately 18%.

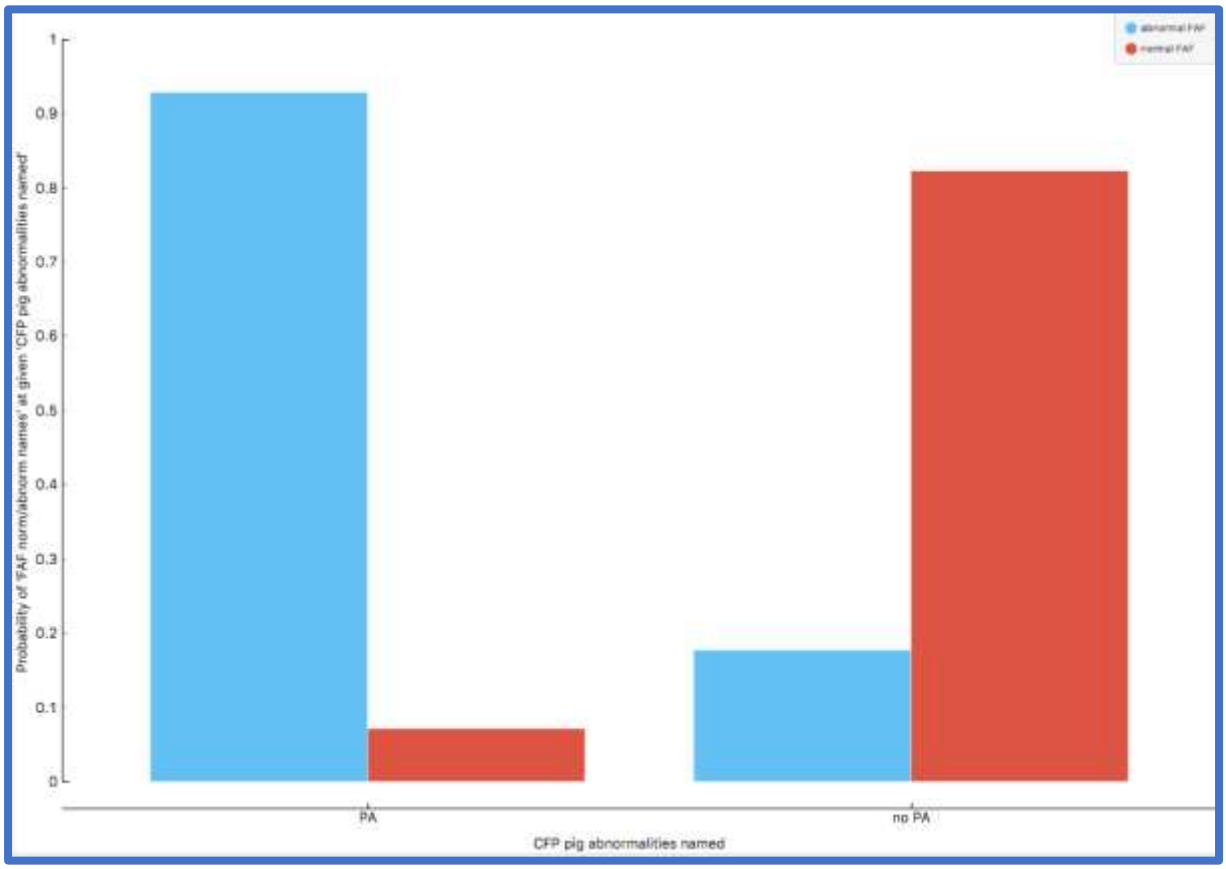

Figure 6.31 Distribution of CFP pigmentary abnormalities and probability of normal/abnormal FAF
Image from ODM showing the presence (group 1 on the left) or absence (group 2 on the right) of pigmentary abnormalities as detected by CFP against the probability of an abnormal (blue bars) and a normal (red bars) FAF result.

The Figure below shows that there is a statistically significant difference between the presence or absence of pigmentary abnormalities for participants with a normal and abnormal FAF result to the 1% level (p = 0.000).

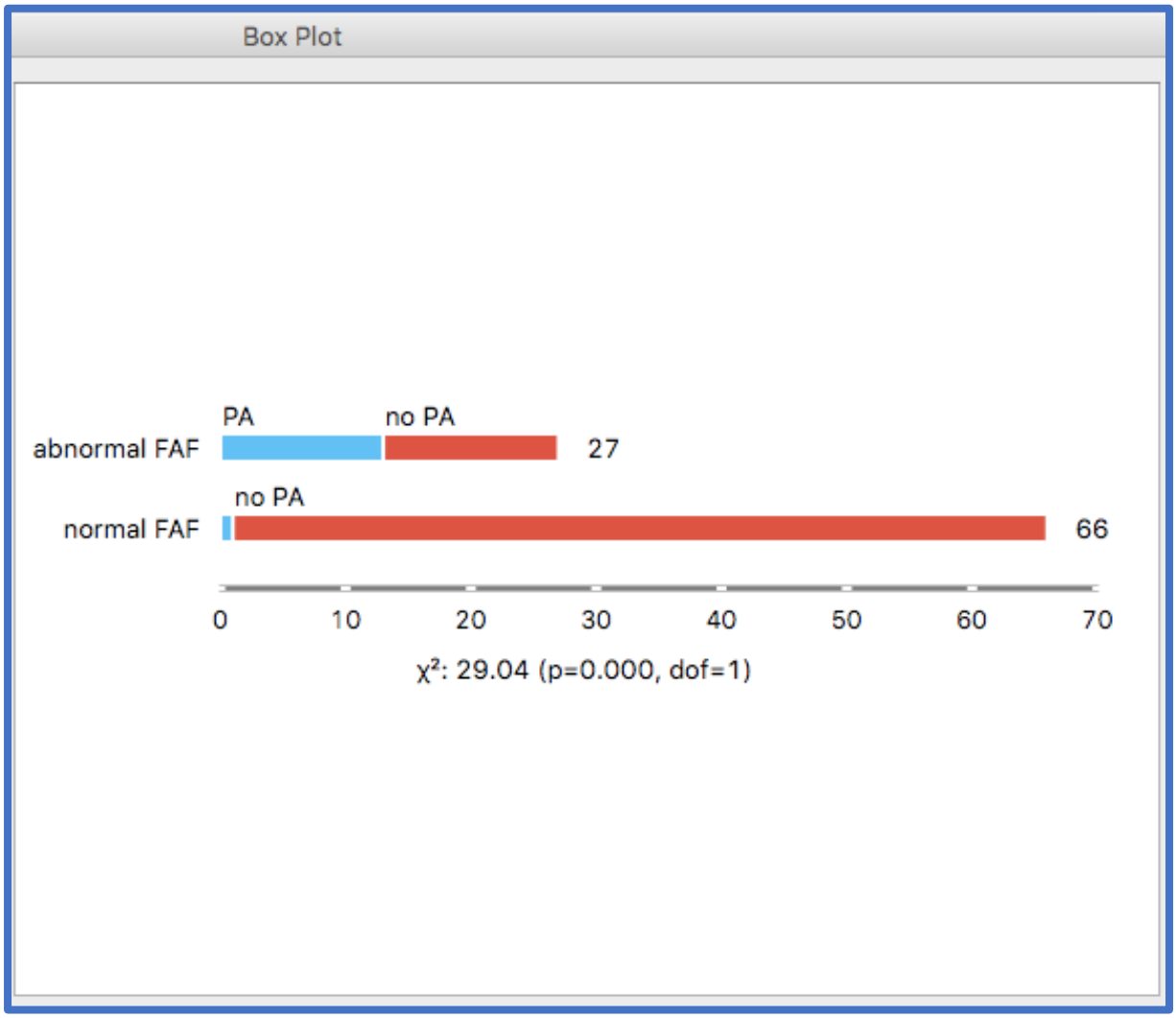

Figure 6.32 Boxplot of pigmentary abnormalities and FAF normality/abnormality

ODM Boxplot widget showing the results of a Chi squared test for the presence (blue bars) or absence (red bars) of pigmentary abnormalities for participants with a normal and abnormal FAF result.

The graph below indicates that if reticular pseudodrusen (RPD) are detected by CFP, then the probability of an abnormal FAF is approximately 75%, however, there were only four instances of RPD being detected by CFP, so this finding should be viewed with caution as the evidence is weak due to the low number of instances.

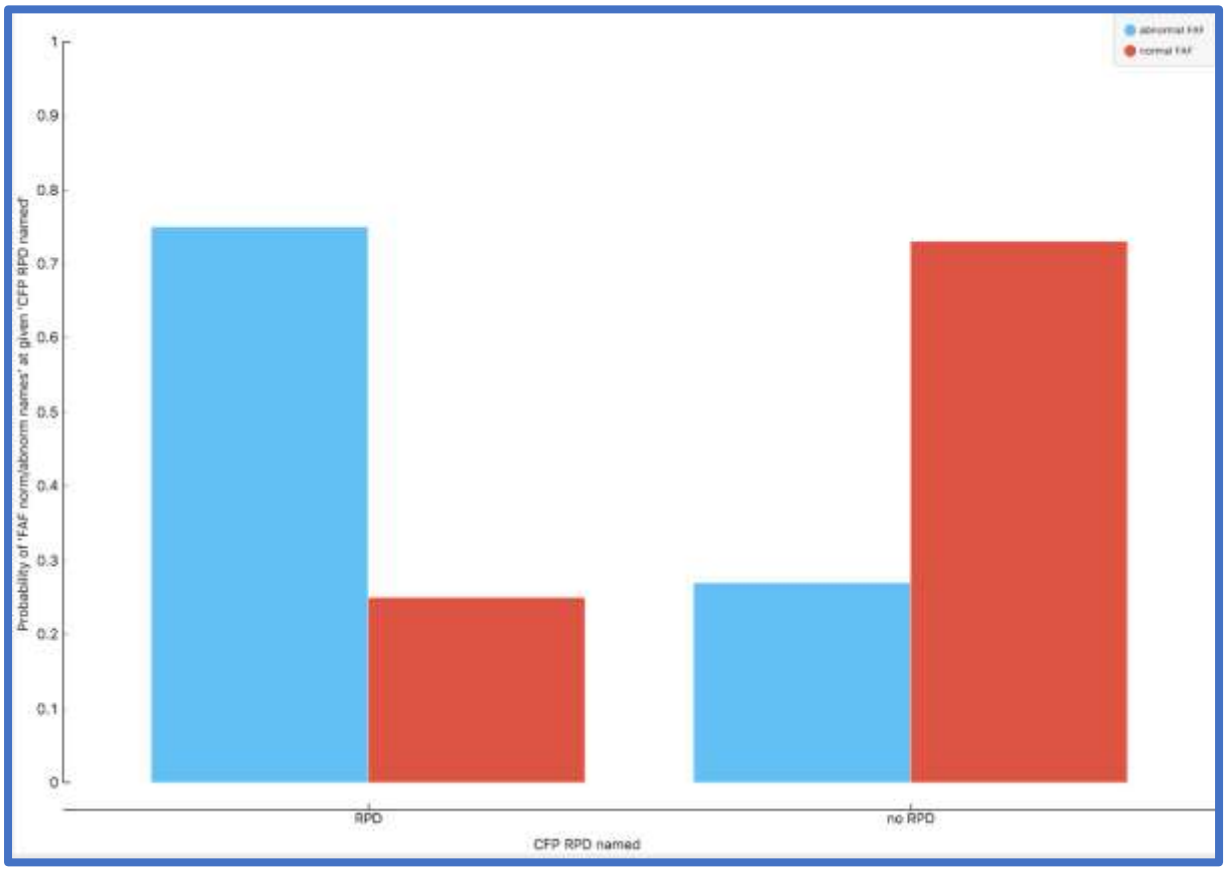

Figure 6.33 Distribution of CFP RPD and probability of FAF normality/abnormality

Image from ODM showing the presence (group 1 on the left) or absence (group 2 on the right) of RPD as detected by CFP against the probability of an abnormal (blue bars) and a normal (red bars) FAF result.

The Figure below shows that there is not a statistically significant difference between the presence or otherwise of RPD as detected by CFP for participants with a normal and abnormal FAF result (p = 0.132).

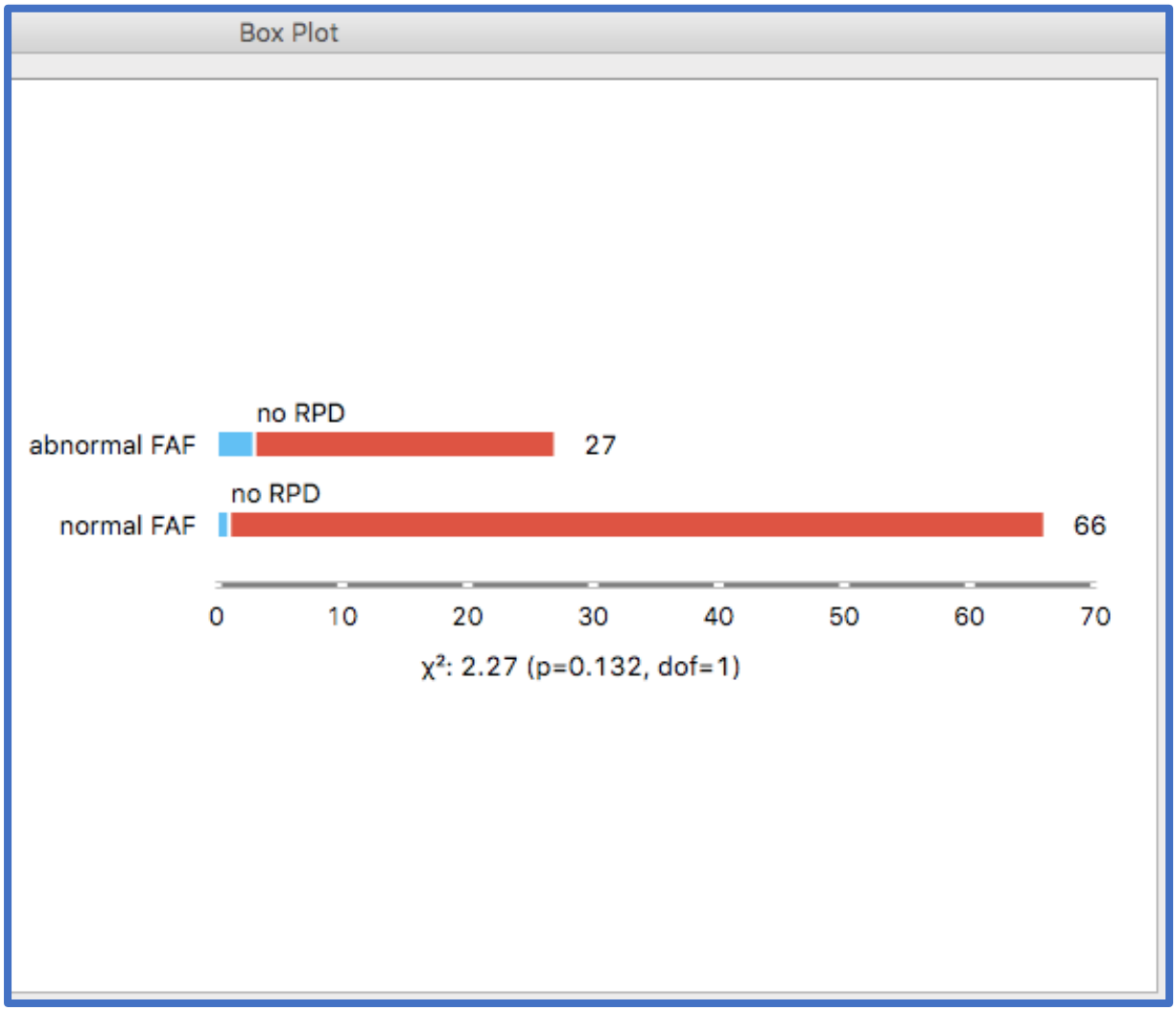

Figure 6.34 Boxplot of CFP RPD and FAF normality/abnormality

ODM Boxplot widget showing the results of a Student's t test for the presence (blue bars) or absence (red bars) of RPD as detected by CFP for participants with a normal and abnormal FAF result.

More instances of RPD detection occurred with OCT (10 in total), and the graph below indicates that if RPD are detected then the probability of an abnormal FAF is high at approximately 80%. If RPD are not detected the probability falls to approximately 23%.

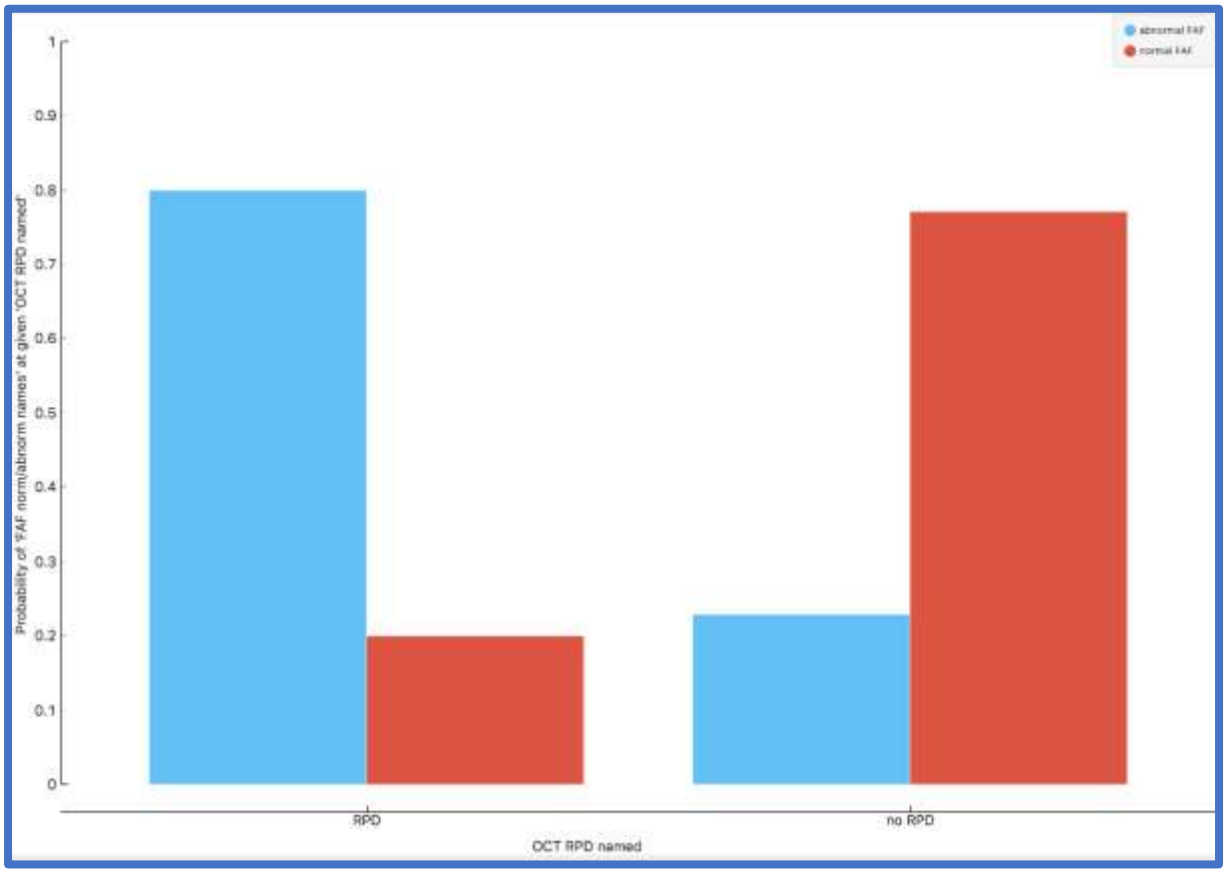

Figure 6.35 Distribution of OCT RPD and probability of normality/abnormality of FAF

Image from ODM showing the presence (group 1 on the left) or absence (group 2 on the right) of RPD as detected by OCT against the probability of an abnormal (blue bars) and a normal (red bars) FAF result.

The Figure below shows that there is a statistically significant difference between the presence or absence of RPD as detected by OCT for participants with a normal and abnormal FAF result to the 1% level (p = 0.001), indicating that there is a significantly higher number of cases with an abnormal FAF result in the group with RPD detected via OCT.

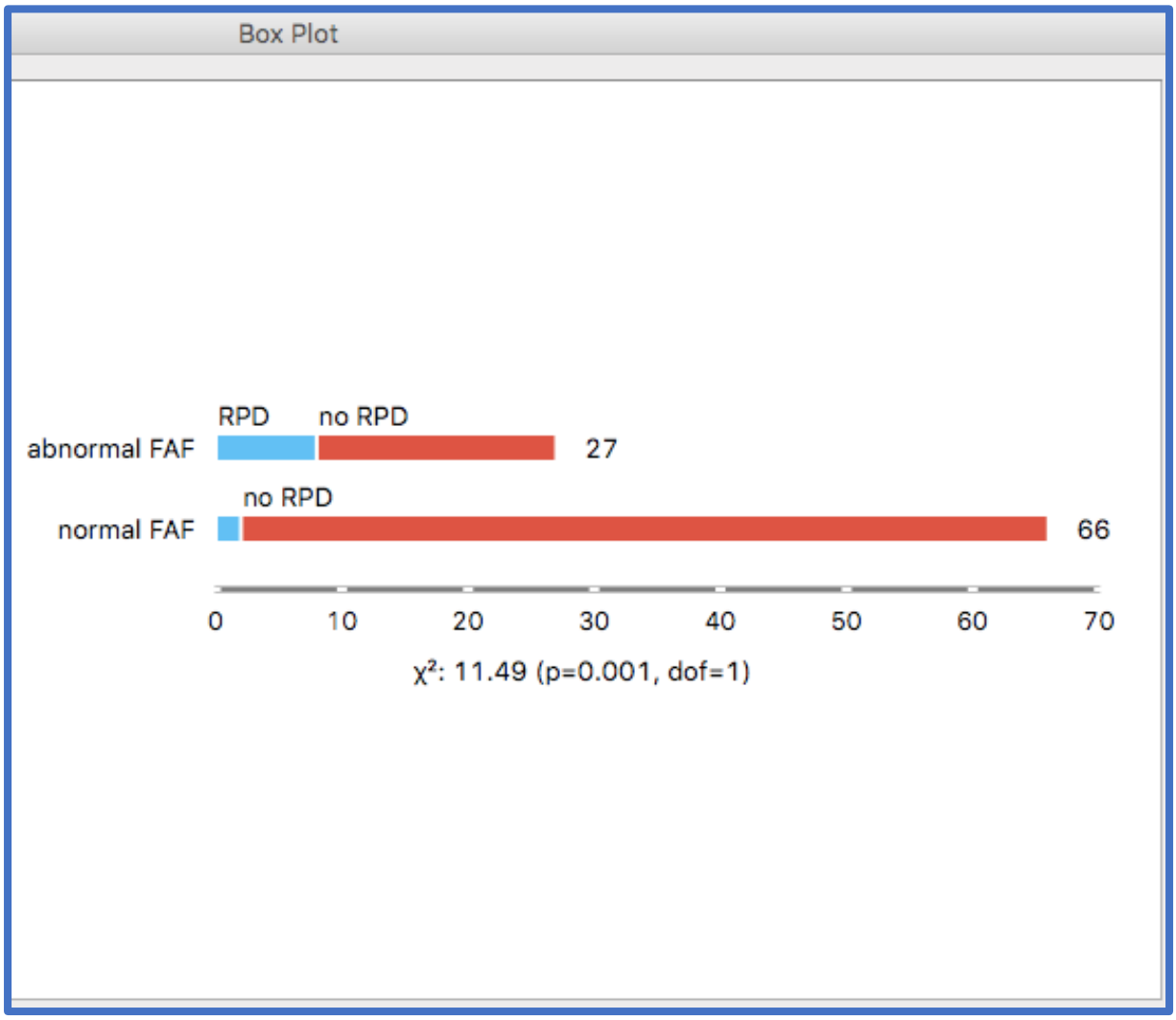

Figure 6.36 Boxplot of OCT RPD and FAF normality/abnormality

ODM Boxplot widget showing the results of a Student's t test for the presence (blue bars) or absence (red bars) of RPD as detected by OCT for participants with a normal and abnormal FAF result.

6.9 Summary of the results of the Distribution and Boxplot widgets

In summary, the ODM Distribution and Boxplot widgets have helped to identify some potential trends in the data (with reference to the normal fitted probability curves and the show probabilities function), helping to answer the questions posed in section 6.1 at the beginning of Chapter six as shown below:

1. Participants with a normal FAF result were significantly younger than those with an abnormal FAF result. The mean age for a participant with a normal FAF result was 69.56 years±9.2 years, and for those with an abnormal FAF the mean age was 76.56 years±10.6 years. The median age for a normal FAF result was 70 years and for an abnormal FAF the median was 77 years.
2. Refractive status of the eye made no significant difference to the likelihood of the FAF result being normal or abnormal.
3. Participants with a normal FAF result had a significantly lower number of packet years of smoking compared to those participants with an abnormal FAF result. (Mean number of packet years in participants with a normal FAF result was 3.3±5.5; mean number of packet years with an abnormal FAF was 9.0±13.3. Median packet years for a normal FAF was 0, and median for an abnormal FAF was 5 years). However, when the analysis was repeated with a subgroup composed of two age matched groups of ex-smokers (N = 24) the difference in packet years between participants with abnormal and normal FAF images is no longer significant.
4. Participants with a first degree relative with ARMD did not have a significantly higher probability of having an abnormal FAF.
5. Participants with a normal FAF result had a significantly lower drusen scale score as measured by either CFP and OCT compared to participants with an abnormal FAF result. This difference held for small, intermediate and large drusen.
6. Participants with a normal FAF result had a significantly lower simplified severity score compared to participants with an abnormal FAF result. If the SSS was > 2 then the probability of an abnormal FAF result was over 90%.
7. Participants in whom pigmentary abnormalities were detected had a significantly higher probability of having an abnormal FAF result. If pigmentary abnormalities were detected, the probability of an abnormal FAF result was approximately 90%, if pigmentary abnormalities were not detected the probability of an abnormal FAF result fell to approximately 18%.

8. Participants in whom reticular pseudodrusen (RPD) were detected by OCT had a significantly higher probability of having an abnormal FAF result. If RPD were detected by OCT, the probability of an abnormal FAF result was approximately 80%. If RPD were not detected by OCT, the probability of an abnormal FAF result fell to approximately 23%. Evidence for RPD as detected by CFP and the association with a normal/abnormal FAF result was weak due to a lack of data/poor RPD detection rate via CFP in this study.

6.10 Results from Machine learning using Orange Data Mining

The next step in analysing the data, after the initial testing with the Distribution and Boxplot widgets, is to perform an Artificial Intelligence/Machine Learning workflow as shown below:

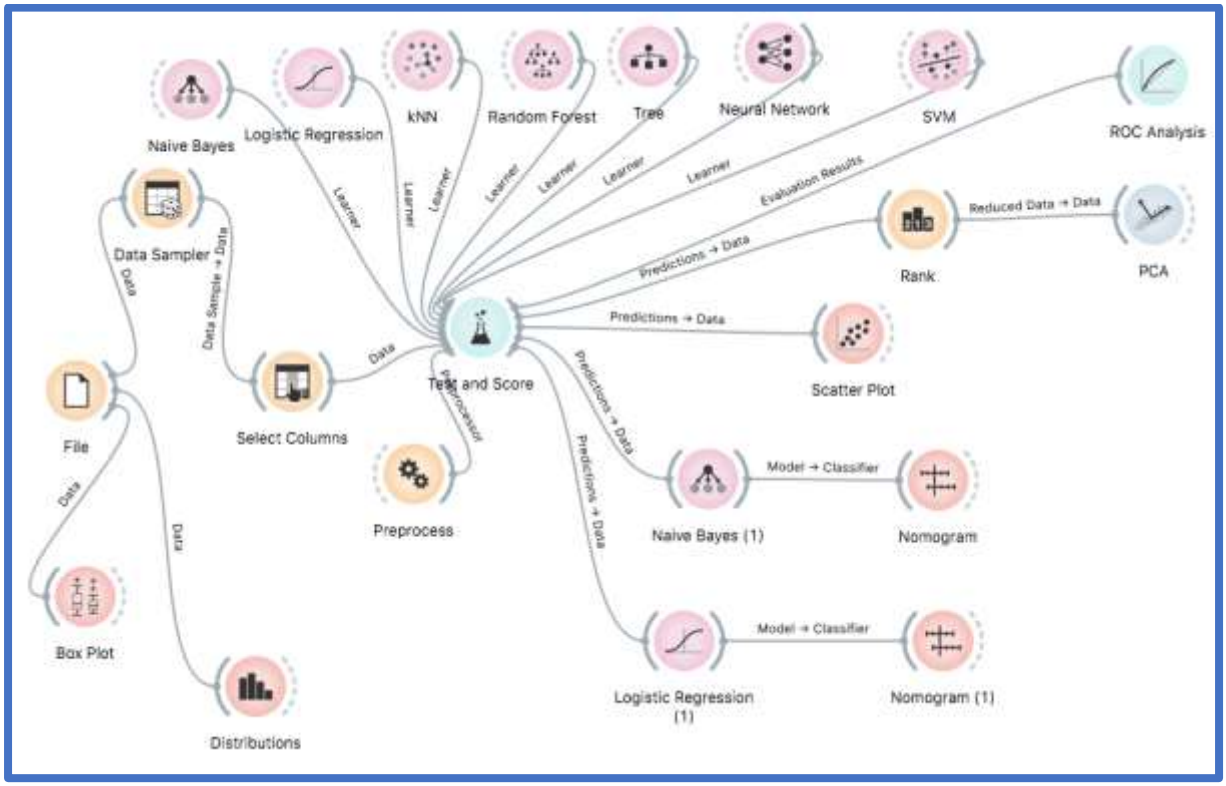

Figure 6.37 The Orange Data Mining workflow utilised in this study.

The first step was to upload a Microsoft Excel spreadsheet containing the data collected into the File widget. Secondly, the variables were selected to be used in the analysis via the Select Columns widget. This was done twice, once for CFP alone, and once for CFP and OCT combined, in order to observe how predictions changed when information from OCT was added to that from CFP alone. For CFP alone, all variables were ignored that related to FAF and OCT results. For CFP and OCT combined only variables relating to FAF were ignored. Also ignored in both cases was the "years of smoking cessation", with a binary attribute "ceased smoking ≥ 20 years/never smoked" selected for analysis instead. This is because a non-smoker would either have a lifetime (which will vary naturally according to age), or zero years of smoking cessation, both of which would be confusing for the statistical analysis. Also ignored for both cases was patient number (patient (px) number) (this was a randomly allocated number for patient identification).

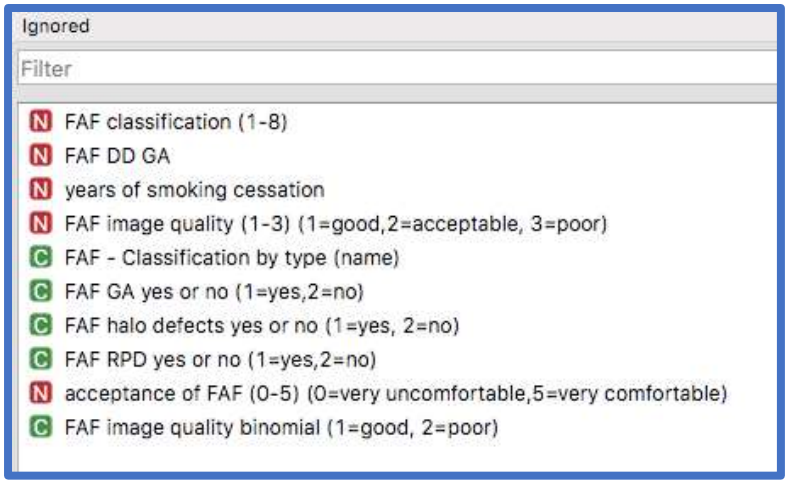

Figure 6.38 Section of the select columns widget of screened out variables

Part of the Select Columns widget from ODM showing which variables were screened out prior to Machine Learning being applied.

Within the Select Columns widget, a “Target Variable” was also selected, i.e. the FAF binomial normal/abnormal variable, as shown below in the Figure below.

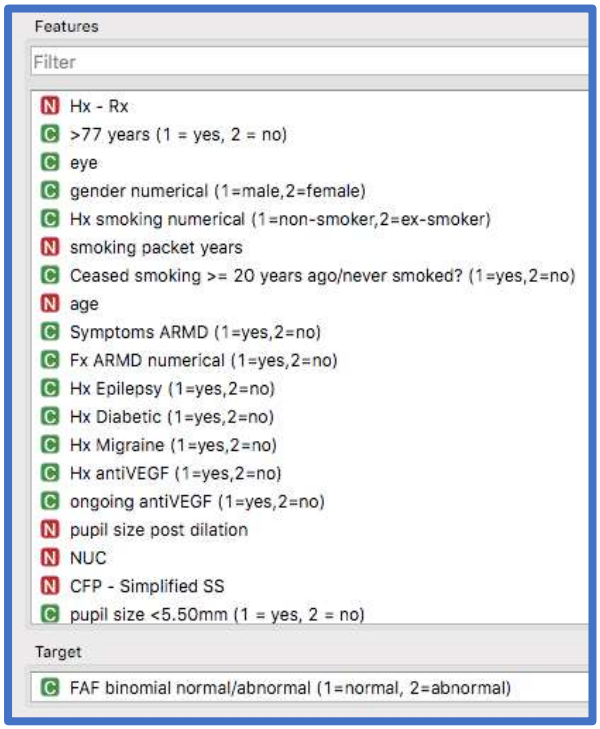

Figure 6.39 Selected and target variables selected in the Orange Select Columns widget.

Part of the Select Columns widget from ODM showing some of the variables selected prior to Machine Learning being applied. The “Target Variable” is shown in the box at the bottom of the screenshot.

The third step was to pre-process the data with the Preprocess and Rank widgets. The Preprocess widget performs the important task of reducing the number of variables down to an acceptable level, ensuring that the instances to variables ratio is no less than 5 to 1 as discussed in more depth in Chapter nine on PCA. The Rank widget allows the researcher to identify which variables have been retained by the Preprocess widget. The Preprocess widget has also normalised the variables, imputed any missing values with the average/most frequent values and continuised the data. The data was manually screened for outliers.

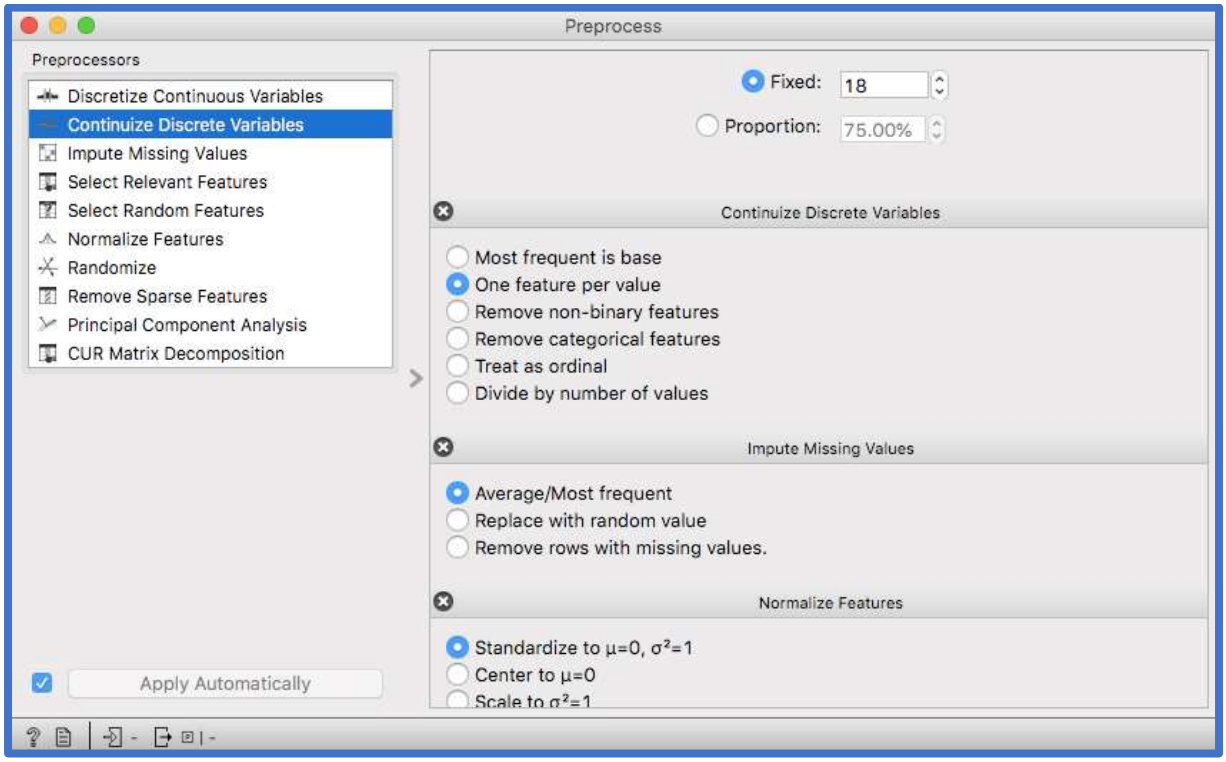

Figure 6.40 Orange Data Mining Preprocess widget selections

Preprocess widget output showing that the 18 most informative variables based on Information Gain were selected. Normalise features, impute missing values and continuise data was also performed.

The fourth step was to perform the hyperparameter fine-tuned Logistic Regression and Naïve Bayes model learning process. For CFP alone the hyperparameter fine-tuned logistic regression gives a best trade off C value of 0.30, and for CFP and OCT combined the best trade off C value is 0.10.

Using these values (all cases an "average over classes" of normal and abnormal FAF results) gives the following results shown below in the Table below:

For CFP alone

	Sensitivity	**Specificity**	**Informedness**	**AUC**
Naïve Bayes	0.796	0.807	0.603 (60.3%)	0.877
Logistic Regression (tuned to C = 0.30)	0.871	0.750	0.621 (62.1%)	0.875

Table 6.3 Results from CFP data alone

Table above showing the Sensitivity, Specificity, Informedness and AUC for the Naïve Bayes and Logistic Regression model learner for CFP alone.

For CFP and OCT combined

	Sensitivity	**Specificity**	**Informedness**	**AUC**
Naïve Bayes	0.774	0.820	0.594 (59.4%)	0.887
Logistic Regression (tuned to C = 0.10)	0.860	0.746	0.606 (60.6%)	0.903

Table 6.4 Results from CFP and OCT data combined

Table above showing the Sensitivity, Specificity, Informedness and AUC for the Naïve Bayes and Logistic Regression model learner for CFP and OCT combined.

The fifth and final step was to examine the Nomogram widget for CFP, and for CFP and OCT combined:

The 18 most informative variables for the two model learners for CFP alone are shown in the Table below, identified via the Rank widget.

Naive Bayes	**Logistic Regression**	**Most informative variables:** **Rank position 1 = most informative** **Rank position 18 = least informative**
CFP – Simplified Severity Score	CFP – Simplified Severity Score	1
CFP large drusen (0-6)	History – refraction	2
CFP pigmentary anomalies yes or no (1 = yes, 2 = no)	Smoking packet years	3
CFP disc diameters of pigmentary anomalies	Pupil size post dilation	4
Symptoms of ARMD (1 = yes, 2 = no)	CFP pigmentary anomalies yes or no (1 = yes, 2 =no)	5
CFP disc diameters of geographic atrophy	CFP disc diameters of geographic atrophy	6
CFP large drusen yes or no (1 = yes, 2 = no)	CFP large drusen (0-6)	7
CFP geographic atrophy yes or no (1 = yes, 2 = no)	Nuclear sclerotic cataract (WHO scale)	8
Age	Age	9
CFP small drusen (0-6)	>77 years of age (1 = yes, 2 = no)	10
CFP intermediate drusen (0-6)	Eye (Right or Left)	11
CFP drusen large other eye (1 = yes, 2 = no)	Gender numerical (1= male, 2 = female)	12
CFP reticular pseudodrusen yes or no (1 = yes, 2 = no)	History - smoking numerical (1 = non-smoker, 2 = ex-smoker)	13
History of anti-VEGF (1 = yes, 2 = no)	Ceased smoking >= 20 years ago/never smoked (1 = yes, 2 = no)	14
Ongoing anti-VEGF (1 = yes, 2 = no)	Symptoms of ARMD (1 = yes, 2 = no)	15
Posterior capsular opacification (1= yes, 2 = no)	Family history of ARMD numerical (1 = yes, no = 2)	16
Pupil size post dilation (mm)	History - epilepsy	17
Nuclear sclerotic cataract (WHO scale)	History – diabetic	18

Table 6.5 Most informative variables for CFP alone for NB and LR learners

Table above showing the 18 most informative variables for the two model learners for CFP alone, identified via the Rank widget.

The two learners agree on the following eight variables (for CFP alone):

Variable	**Naïve Bayes Rank position**	**Logistic Regression Rank position**	**Rank position total score (Naive + Logistic Regression Rank position)**
CFP – Simplified Severity Score	1	1	2
CFP pigmentary anomalies yes or no (1 = yes, 2 =no)	3	5	8
CFP large drusen (0-6)	2	7	9
CFP disc diameters of geographic atrophy	6	6	12
Age	9	9	18
Symptoms of ARMD (1 = yes, 2 = no)	5	15	20
Pupil size post dilation	17	4	21
Nuclear sclerotic cataract (WHO scale)	18	8	26

Table 6.6 The common most informative variables for CFP alone for NB and LR learners

The 8 most informative variables, agreed upon by the two model learners for CFP alone, identified via the Rank widget.

A clinical support tool could therefore be developed using these eight variables, e.g. for a patient A with a simplified severity score of 2, no pigmentary abnormalities, CFP large drusen score of > 10 < 20 (grade 2), no disc diameters of geographic atrophy, 80 years of age, no symptoms of ARMD, a pupil size post dilation of 6mm and a nuclear sclerotic cataract score on the WHO scale of 1.0. The results are shown below:

What is the probability of an abnormal FAF result for patient A?

For Naive Bayes using only data from CFP

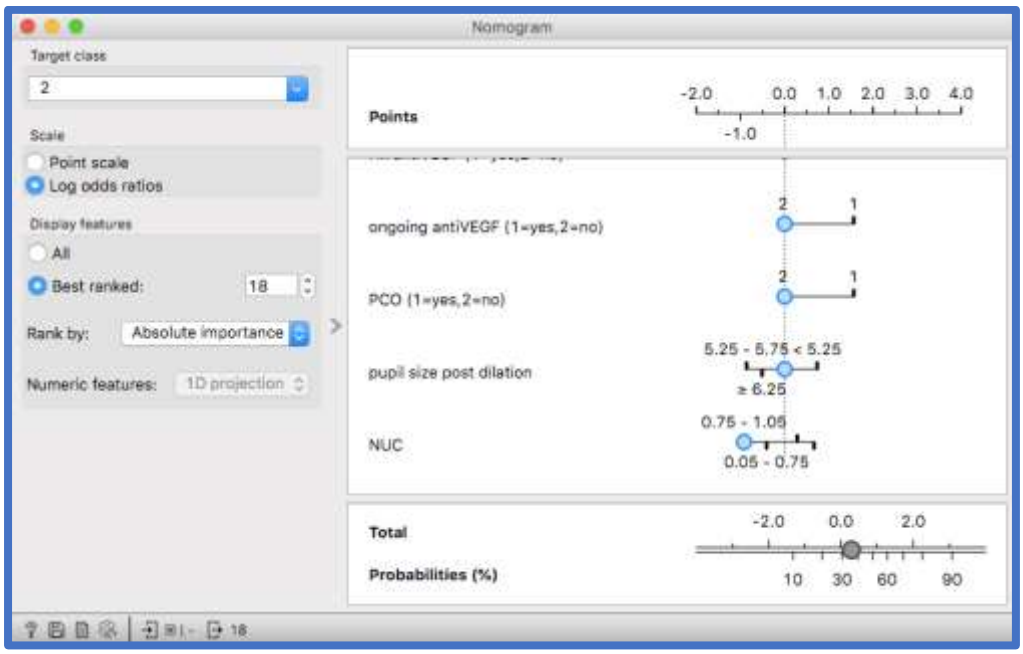

Table 6.7 Nomogram output for Naïve Bayes for CFP data alone for patient A.

Naïve Bayes suggests the probability of an abnormal FAF result for patient A of 36% (informedness of model 60.3%, AUC 0.877)

For Logistic Regression using only data from CFP

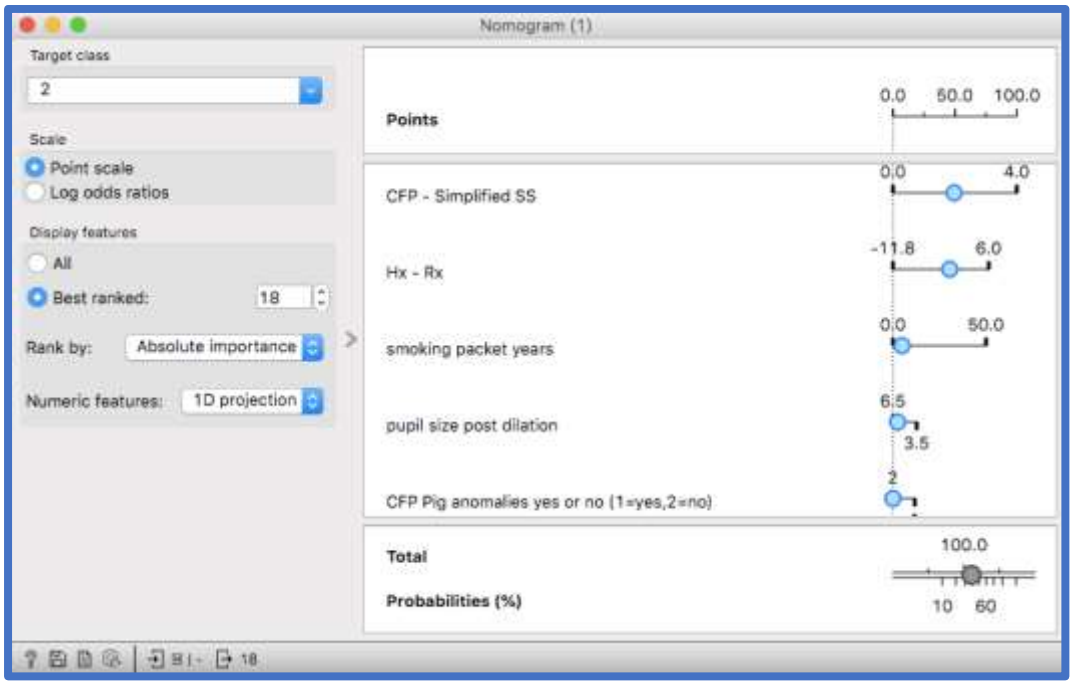

Table 6.8 Nomogram widget output for Logistic Regression for CFP alone for patient A

Logistic Regression suggests the probability of an abnormal FAF result for patient A of 39% (informedness of model 62.1%, AUC 0.875)

Therefore, the two model learners are in close agreement for CFP alone, with Naive Bayes indicating the probability of an abnormal FAF result for our patient A of 36% (informedness of 60.3%, AUC 0.877), whilst Logistic Regression suggests 39% (informedness of 39%, AUC 0.875).

Next, the 18 most informative variables for the two model learners for CFP and OCT combined are shown in the table below, identified via the Rank widget.

Naïve Bayes	**Logistic Regression**	**Most informative variables:** **Rank position 1 = most informative** **Rank position 18 = least informative**
OCT large drusen (0-6)	OCT large drusen (0-6)	1
CFP simplified severity score	Hx -Rx (Refractive state of the eye as measured with the OCT Gullstrand function)	2
CFP large drusen (0-6)	CFP Simplified Severity Score	3
OCT large drusen yes or no (1 = yes, 2 = no)	CFP pigmentary abnormalities yes or no (1 = yes, 2 = no)	4
CFP pigmentary abnormalities yes or no (1 = yes, 2 = no)	CFP disc diameters of geographic atrophy	5
CFP disc diameters of pigmentary abnormalities	Smoking packet years	6
Symptoms of ARMD (1 = yes, 2 = no)	CFP inter drusen (0-6)	7
OCT inter drusen (0-6)	OCT – min foveal thickness	8
CFP disc diameters of geographic atrophy	Pupil size post dilation	9
CFP large drusen yes or no (1 = yes, 2 = no)	NUC cataract	10
CFP geographic atrophy yes or no (1 = yes, 2 = no)	OCT average foveal thickness	11
OCT small drusen	>77 years (1 = yes, 2 = no)	12
OCT RPD yes or no (1 = yes, 2 = no)	Age	13
OCT disc diameters of geographic atrophy	Eye (Right or Left)	14
OCT PED yes or no (1 = yes, 2 = no)	gender numerical (1 = males, 2 = female)	15
OCT geographic atrophy yes or no (1 = yes, 2 = no)	Hx smoking numerical (1 = non-smoker, 2 = ex-smoker)	16
Age	Ceased smoking >=20 years ago/never smoked (1 = yes, 2 = no)	17
OCT SRF yes or no (1 = yes, 2 = no)	Symptoms of ARMD (1 = yes, 2 = no)	18

Table 6.9 Most informative variables for CFP and OCT combined for NB and LR learners

The 18 most informative variables for the two model learners for CFP and OCT combined, identified via the Rank widget.

The two learners agree on the following six variables (for CFP and OCT combined):

Variable	**Naïve Bayes Rank position**	**Logistic Regression Rank position**	**Rank position total score (Naive + Logistic Regression Rank position)**
OCT large drusen (0-6)	1	1	2
CFP Simplified Severity Score	2	3	5
CFP pigmentary abnormalities yes or no (1 = yes, 2 = no)	5	4	9
CFP disc diameters of geographic atrophy	9	5	14
Symptoms of ARMD	7	18	25
Age	17	13	30

Table 6.10 Most informative variables for CFP and OCT combined

The 6 most important variables, agreed upon by the two model learners for CFP and OCT combined, identified via the Rank widget.

A clinical support making tool could therefore be developed using these six variables, e.g. for a patient B, with > 10 < 20 large drusen on OCT (grade 2), a CFP simplified severity score of 2, no pigmentary abnormalities observed on CFP, no disc diameters of geographic atrophy detected with CFP, no symptoms of ARMD and of 70 years of age. The results are shown below:

What is the probability of an abnormal FAF result for patient B?

For Naïve Bayes using data from both CFP and OCT

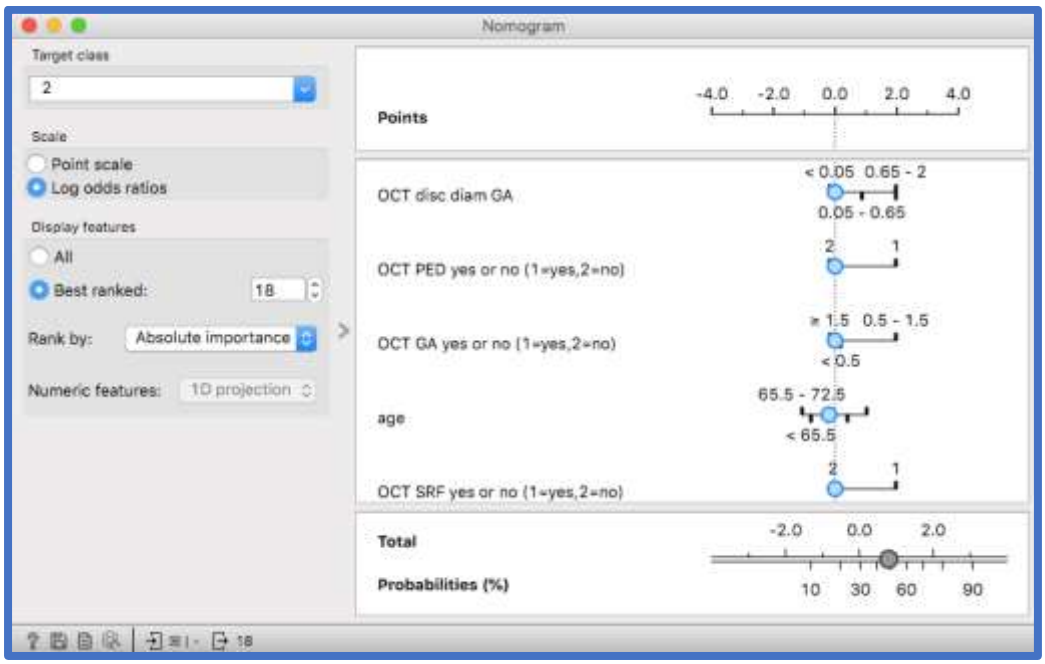

Figure 6.41 Nomogram for Naïve Bayes using data from both CFP and OCT for patient B.

Naïve Bayes suggests the probability of an abnormal FAF result for patient B of approximately 48% (informedness of model 59.4%, AUC 0.887)

For Logistic Regression using data from both CFP and OCT

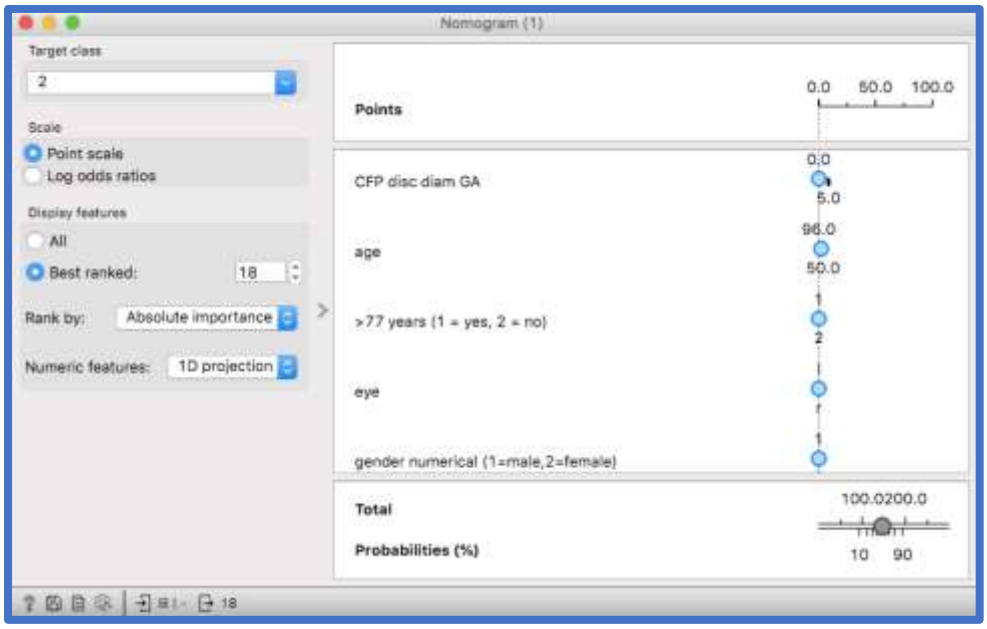

Figure 6.42 Nomogram for Logistic Regression for both CFP and OCT combined.

Logistic Regression suggests the probability of an abnormal FAF result for patient B of approximately 54% (informedness of model 60.6%, AUC 0.903)

Therefore, as with CFP alone, for CFP in combination with OCT, the two model learners are in fairly close agreement, with Naive Bayes indicating the likelihood of an abnormal FAF result for our patient B of 48% (informedness of model 59.5%, AUC 0.887), with Logistic Regression suggesting 54% (informedness of model 60.6%, AUC 0.903). These findings suggest that these two model learners could enable the development of a clinical support tool based on variables collected during eye examinations.

Other Artificial Intelligence Model Learners

Finally, the same data for stand-alone CFP and combined CFP and OCT were explored via the other AI Machine Learners available through the ODM Software, i.e. kNN, Tree, Random Forest, Support Vector Machine (SVM) and Neural Network. The table below summarises the results. (Again, in all cases an “average over classes” of normal and abnormal FAF results in selected).

6.10.1 Summary of results from all the artificial intelligence model learners utilised in this study

For CFP alone, the best model for Informedness is Logistic Regression hyperparameter fine-tuned to C = 0.3 (62.1%), however, the best AUC (0.877) is achieved via Naïve Bayes (see Table 6.11 below). For combined CFP and OCT combined the best informedness was from SVM (an impressive 70.2%), with Logistic Regression fine-tuned to C = 0.1 giving the best AUC (0.903) (see Table below). The results with the best performances highlighted in bold.

CFP alone

Model Learner	**Sensitivity**	**Specificity**	**Informedness**	**AUC**
kNN	0.839	0.671	0.510 (51.0%)	0.850
Tree	0.763	0.684	0.447 (44.7%)	0.742
Random Forest	0.860	0.702	0.562 (56.2%)	0.852
SVM	0.849	0.763	0.612 (61.2%)	0.847
Neural Network	0.828	0.755	0.583 (58.3%)	0.818
Naïve Bayes	0.796	0.807	0.603 (60.3%)	**0.877**
Logistic Regression (fine-tuned to C = 0.3)	0.871	0.750	**0.621 (62.1%)**	0.875

Table 6.11 Sensitivity, Specificity, Informedness and AUC for model learners for CFP alone.

CFP and OCT combined

Model Learner	**Sensitivity**	**Specificity**	**Informedness**	**AUC**
kNN	0.882	0.777	0.659 (65.9%)	0.889
Tree	0.817	0.706	0.523 (52.3%)	0.800
Random Forest	0.871	0.750	0.621 (62.1%)	0.881
SVM	0.882	0.820	**0.702 (70.2%)**	0.874
Neural Network	0.835	0.796	0.631 (63.1%)	0.741
Naïve Bayes	0.774	0.820	0.594 (59.4%)	0.887
Logistic Regression (fine-tuned to C = 0.1)	0.860	0.746	0.606 (60.6%)	**0.903**

Table 6.12 Sensitivity, Specificity, Informedness and AUC for model learners for CFP and OCT combined.

Below is shown the ROC analysis for a target class of 1, i.e. a normal FAF result, for CFP and OCT data combined. Unlike the Test and Score widget, the ROC widget does not offer an “Average over classes” option, as a weighted average of the performance for predicting both normal and abnormal FAF results. From the graph it can be seen that kNN is the best performing model learner by reference to the black solid performance line.

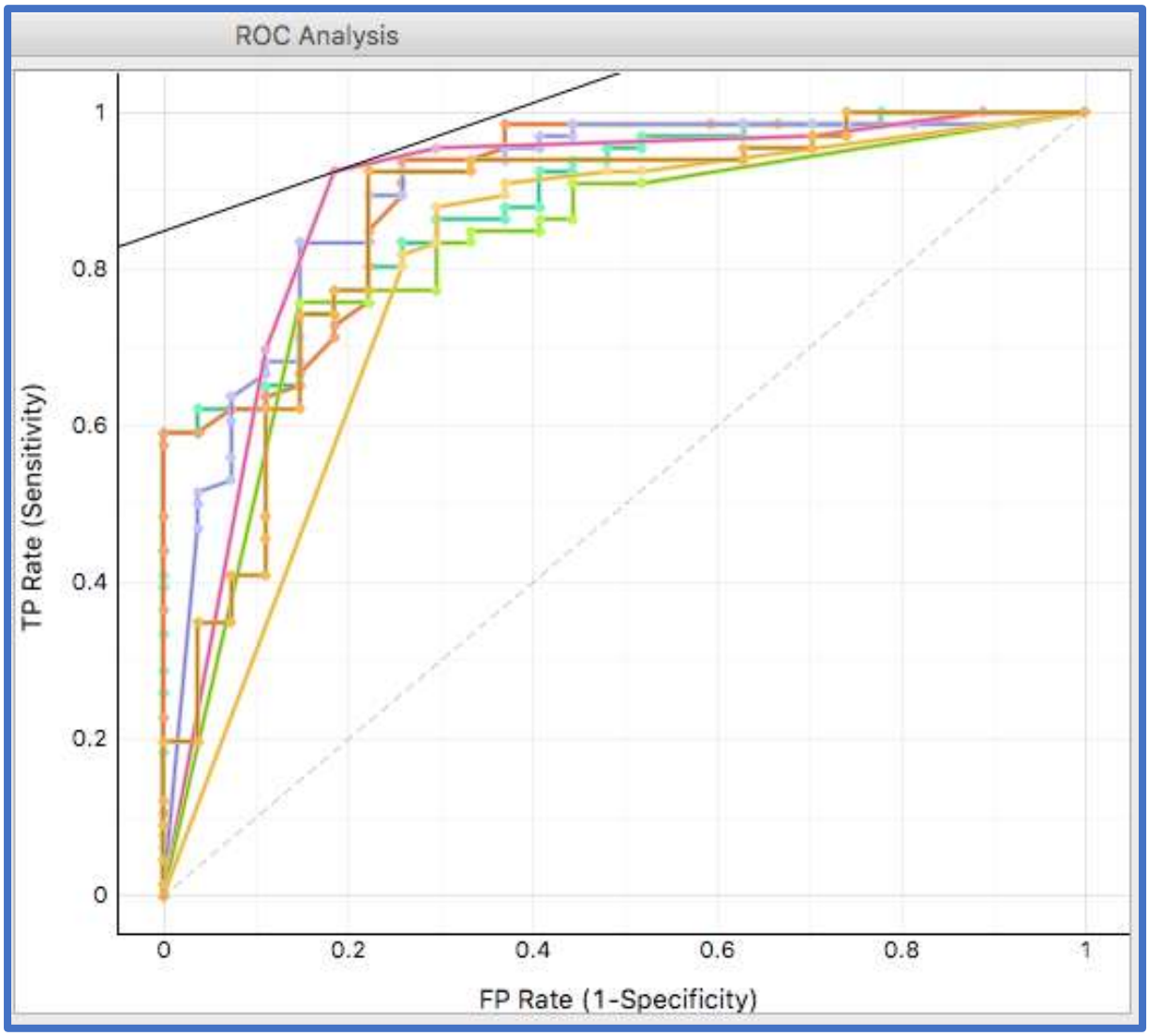

Figure 6.43 ROC analysis for target class 1, i.e. a normal FAF result prediction. The pink line represents the best model learner performance (kNN) with reference to the solid black performance line.

Below is shown the ROC analysis for a target class of 2, i.e. an abnormal FAF result, for CFP and OCT data combined. From the graph it can be seen that kNN is again the best performing model learner, by reference to the black solid performance line. It is interesting to note that according to the “Average over classes” results, that kNN was not the best performing model learner for Informedness nor AUC, however, it was second on both scores out of the seven models tested.

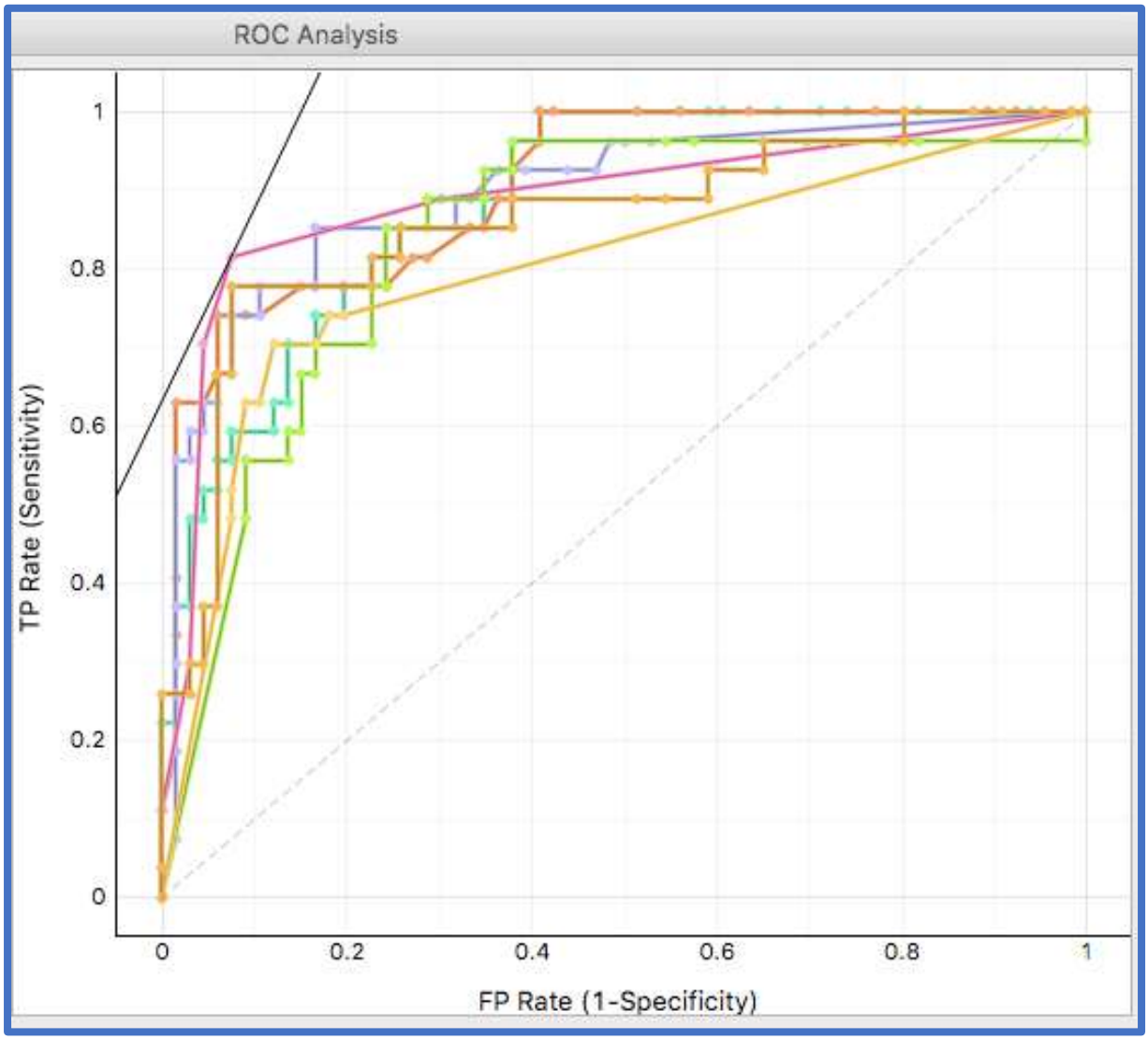

Figure 6.44 ROC analysis for target class 2, i.e. an abnormal FAF result prediction. The pink line represents the best model learner performance (kNN) with reference to the solid black performance line.

6.11 Dietary supplementation and the role of FAF imaging

An addition question was also addressed in the following section. Could performing FAF imaging have an impact on the clinicians’ advice to patients with regards to prescribing dietary supplements? Previously, the recommendation of dietary supplementation has been

based on the findings of the AREDS 2 study(149) – i.e. all of the following must hold for dietary supplementation to be recommended: an abnormal macular appearance, age of 50-85 years, large drusen at the macula, large drusen in the other eye or GA/neovascular ARMD in either eye.(189) The trigger for recommending dietary supplements in non-advanced ARMD (i.e. when there are no GA or neovascular changes present), is therefore the finding of large drusen in both eyes.

There are unfortunately no guidelines based on FAF imaging to assist practitioners regarding when to, and when not to offer dietary supplements. Therefore, certain assumptions have to be made in order to test whether performing FAF in addition to CFP and/or OCT might alter the recommendations given to patients (without advanced ARMD). The assumption that seems reasonable is to suggest that if the FAF imaging in the study eye was “abnormal” (i.e. not normal or minimal change from the classification system suggested by Bindewald in 2005, but one of six more advanced FAF patterns), then dietary supplements should be recommended. It was assumed that if the FAF image was “abnormal”, then it can be assumed that the retina is under stress, or in other words, chemically malfunctioning, and supplements would be of assistance. This will be referred to this as the “FAF assumption”. ODM was then used to test how often advice would change on this basis compared to the traditional method of recommending supplementation based on the finding of large drusen in both eyes via CFP or OCT. However, in this study FAF was only carried out on one eye, so a comparison of large drusen and FAF in both eyes is not possible. Therefore, the comparison performed was to test whether FAF may *potentially* change the supplementation advice, i.e. does FAF detect an abnormal signal in the absence of large drusen, simply based on one eye. Patients from this study who had active neovascular ARMD or GA were removed from the analysis, as the aim is to detect the subtle changes that are likely to occur in early/moderate ARMD rather than the more advanced retinal changes seen in the late disease (this decision resulted in eight patients being removed from the analysis in total). It appears from Figure 6.44 below, that when supplements are not recommended based on the absence of large drusen detected by OCT, the “FAF assumption” agreed in all cases, however, there appears to be a significant percentage (56%) of cases when recommendations based on large drusen findings will indicate that supplementation is required, but the “FAF assumption” will not. This would therefore indicate that FAF would not change the supplementation advice, so long as the presence of large drusen were allowed to “trump” the FAF findings.

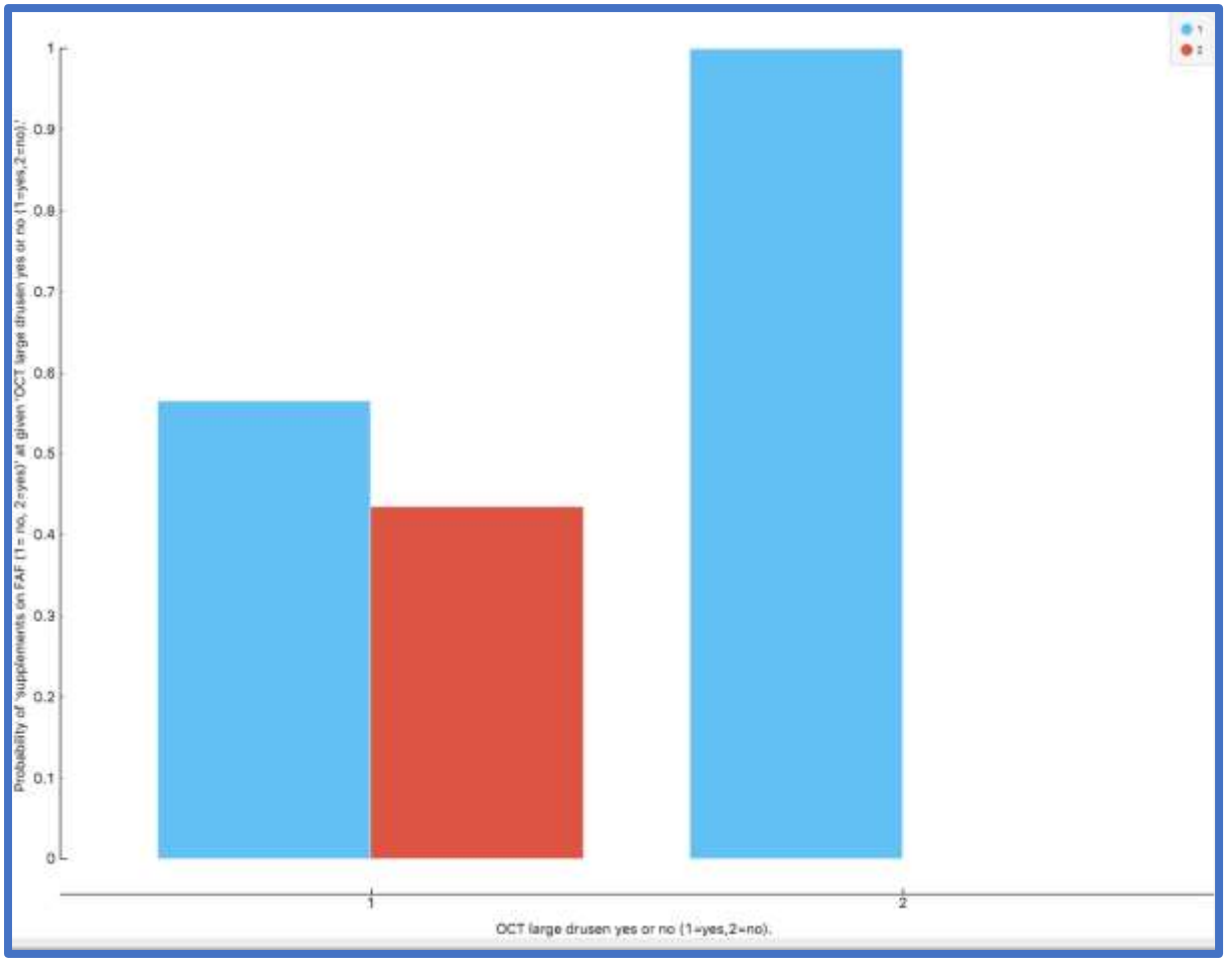

Figure 6.45 Distribution showing how large drusen detected by OCT on the x axis (1 = yes, 2 = no) are related to the probability of recommending dietary supplements based on "The FAF assumption" on the y axis (blue bars = probability of not giving supplements based on the "FAF assumption", red bars = probability of giving supplements based on the "FAF assumption").

However, when the same analysis was run for large drusen as detected by CFP alone, shown in the Figure below, there were three cases when CFP did not detect large drusen, but the "FAF assumption" did recommend supplementation. This was due to OCT detecting large drusen more often than CFP. This phenomenon has been reported previously in 2021, with the size of drusen being smaller as measured by CFP compared to OCT. The paper reported that large drusen of > 125μm on CFP had a diameter of ≥ 145μm on OCT, medium drusen between 63 and 124μm on CFP measured as 100 to 144μm on OCT, and small drusen of < 63μm on CFP had a diameter of < 100μm on OCT. The paper also created an algorithm: drusen diameter on SD-OCT = 0.77 * (drusen diameter on CFP) + 50μm.(195) This underestimation of drusen size on CFP could be due to the tendency of drusen to be wider at their outer retinal base than at their inner retinal apex, with CFP not allowing the clinician to appreciate the full size of drusen due to the superficial nature of CFP imaging. OCT benefits from a cross-sectional profiling, and therefore appears to be superior for

appreciating the true lateral extent of drusen dimensions. Practitioners should therefore be cautious when utilising CFP alone to estimate the size of drusen for the purposes of patient risk profiling and supplementation advice, with OCT being a superior modality for this purpose.

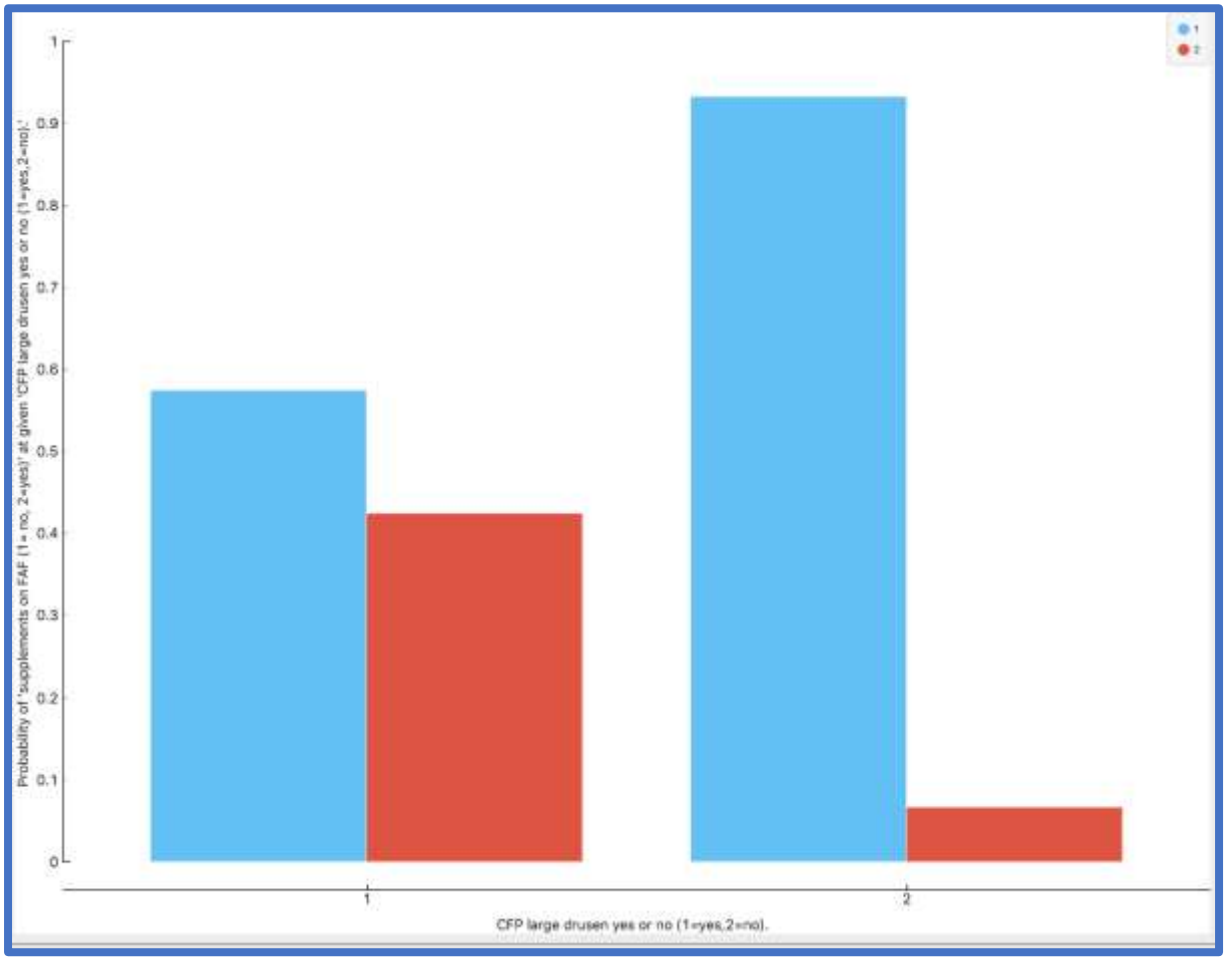

Figure 6.46 Distribution showing how large drusen detected only by CFP on the x axis (1 = yes, 2 = no) are related to the probability of recommending dietary supplements based on "The FAF assumption" on the y axis (blue bars = probability of not giving supplements based on the "FAF assumption", red bars = probability of giving supplements based on the "FAF assumption").

6.12 Discussion

Chapter six has examined the data from this study with the ODM Boxplot and Distribution widgets, before finally carrying out an AI/Machine Learning model analyses with Test and Score and Nomogram widgets. From the Boxplot and Distribution widgets the following trends in the data were identified (a facsimile from the previous section on Boxplot and Distribution widgets above in section 6.9):

6.12.1 Summary of results from data analysed via Orange Data Mining software

In summary, the Distributions and Boxplot widgets have helped to identify some potential trends in the data that can be further investigated within the ODM software. These are:

1. Participants with a normal FAF result were significantly younger than those with an abnormal FAF result. The mean age for a participant with a normal FAF result was 69.56 years±9.2 years, and for those with an abnormal FAF the mean age was 76.56 years±10.6 years. The median age for a normal FAF result was 70 years and for an abnormal FAF the median was 77 years.
2. Refractive status of the eye made no significant difference to the likelihood of the FAF result being normal or abnormal.
3. Participants with a normal FAF result had a significantly lower number of packet years of smoking compared to those participants with an abnormal FAF result. (Mean number of packet years in participants with a normal FAF result was 3.3±5.5; mean number of packet years with an abnormal FAF was 9.0±13.3. Median packet years for a normal FAF was 0, and median for an abnormal FAF was 5 years). However, when the analysis was repeated with a subgroup composed of two age matched groups of ex-smokers (N = 24) the difference in packet years between participants with abnormal and normal FAF images is no longer significant.
4. Participants with a first degree relative with ARMD did not have a significantly higher probability of having an abnormal FAF.
5. Participants with a normal FAF result had a significantly lower drusen scale score as measured by either CFP and OCT compared to participants with an abnormal FAF result. This difference held for small, intermediate and large drusen.
6. Participants with a normal FAF result had a significantly lower simplified severity score compared to participants with an abnormal FAF result. If the SSS was > 2 then the probability of an abnormal FAF result was over 90%.

7. Participants in whom pigmentary abnormalities were detected had a significantly higher probability of having an abnormal FAF result. If pigmentary abnormalities were detected, the probability of an abnormal FAF result was approximately 90%, if pigmentary abnormalities were not detected the probability of an abnormal FAF result fell to approximately 18%.
8. Participants in whom reticular pseudodrusen (RPD) were detected by OCT had a significantly higher probability of having an abnormal FAF result. If RPD were detected by OCT, the probability of an abnormal FAF result was approximately 80%. If RPD were not detected by OCT, the probability of an abnormal FAF result fell to approximately 23%. Evidence for RPD as detected by CFP and the association with a normal/abnormal FAF result was weak due to a lack of data/poor RPD detection rate via CFP in this study.

In the AI/Machine Learning model Nomogram widget section, the findings from CFP alone, and subsequently from CFP in combination with OCT are examined. For CFP alone, and for CFP plus OCT, the informedness levels achieved for both the Naïve Bayes and Logistic Regression model learners via the Test and Score widget all ranged from 59.4 to 62.1%, (with an AUC from 0.875 to 0.903), with the best informedness delivered by the hyperparameter fine-tuned Logistic Regression learner using data from CFP alone achieving a result of 62.1%. This result may be surprising, given that so much more information is gathered to assist in predicting an abnormal FAF result with OCT. However, these findings may be explained when the data is examined in more detail as discussed in the following section 6.12.2 based on CFP data alone.

6.12.2 Discussion of machine learning results based on CFP alone

Both the Naïve Bayes and Logistic Regression learners for CFP alone place the Simplified Severity Score in first place for delivering informedness for predicting a normal/abnormal FAF result. Also, the presence or absence of pigmentary abnormalities was the second-best performing variable based on CFP alone for predicting an abnormal FAF result when the results from both the model learners are taken into consideration together as a weighted average. Both of these variables are heavily influenced by pigmentary abnormalities, which is a variable determined by CFP and not specifically identified within OCT images. The other variables on which the two model learners agree are key to predicting an abnormal FAF result with CFP alone are:

Large drusen (grade 0-6) – this is expected, as from the Boxplot and Distribution widgets section we had established that there is a tendency for a higher probability of an abnormal FAF result with a greater number of large drusen.

Disc diameters of geographic atrophy – this is expected, as geographic atrophy is known to cause hypo-autofluorescence of the fundus.

Age – this is expected – from the Boxplot and Distribution section there was an increase in the probability of an abnormal FAF result with increasing age.

Symptoms of ARMD – from examining the data, five of the 27 participants who had an abnormal FAF result, and none of the remaining 66 who had a normal FAF result, had symptoms of ARMD (i.e. a distortion of the central vision, a “kink” in the central vision or central scotoma). Having symptoms related to dry ARMD would indicate that there are patch(es) of central geographic atrophy, and therefore an abnormal FAF result would be expected in these cases. Likewise, for the sudden onset of symptoms attributable to wet or neovascular ARMD, one would expect disruption of the central macula with a subsequent alteration of the FAF signal due to haemorrhage or intra/subretinal fluid.

Pupil size post dilation – that the two model learners place this variable in the top eighteen for delivering informedness is surprising, albeit that Naïve Bayes places it in 17th position. A Boxplot widget examination reveals that the mean pupil size post dilation is on average smaller for those participants with an abnormal FAF, however, the result does not reach the

level of statistical significance. One explanation for this finding is that the smaller pupil size post dilation may simply be indicating an older participant.

Nuclear sclerotic cataract (WHO scale) – that nuclear sclerotic cataract could be associated with an abnormal FAF result is another result that may be expected, given the association of worsening cataract with increasing age, however, in this study the reverse is found, i.e. that participants with a normal FAF result had a higher mean nuclear cataract score, although the difference was not statistically significant when explored via a students t test. One explanation is that the cataract is masking subtle FAF abnormalities and that FAF abnormalities are therefore more easily detected in eyes with clearer ocular media.

6.12.3 Discussion of machine learning results based on CFP and OCT combined

For CFP and OCT combined it was found that the best performing variables for predicting an abnormal FAF result are as follows: in first place is OCT large drusen (0-6), followed by CFP Simplified Severity Score and pigmentary abnormalities detected by CFP (yes or no), so again those variables associated with pigmentary abnormalities are ranked highly as with the variables from CFP alone. CFP disc diameters of geographic atrophy, Symptoms of ARMD and Age make up the remainder of the table – again, these variables would be expected to be related to an abnormal FAF result as already discussed.

The results indicate, therefore, that pigmentary abnormalities, large drusen (detected by both CFP and/or OCT) and the Simplified Severity Score (essentially an amalgam of pigmentary abnormalities and large drusen as detected by CFP), are highly ranked and are an important predictor (according to the Naïve Bayes and Logistic Regression model learners) of an abnormal FAF result, when both CFP and OCT data is utilised. This explains why CFP at least matches the informedness of CFP and OCT combined in this study, for the artificial intelligence prediction of an abnormal FAF result via the two model learners selected, as findings based on CFP are such important predictors of an abnormal FAF result.

Recent research into pigmentary abnormalities, from a study based in South Korea in patients without significant drusen, appear to concur with these findings.(196) The authors suggest that pigmentary abnormalities have as much clinical significance as drusen, and therefore deserve more attention. They also report that hypopigmented pigmentary

abnormalities are related to a specific OCT sign called shallow irregular RPE elevation (SIRE) and that previous studies have shown that hypopigmentation is a high risk for neovascular ARMD(197, 198). (SIRE, also known as the double-layer sign, is defined as RPE elevations with a greatest transverse linear dimension of $\geq 1000\mu m$, an irregular RPE layer with a height of predominantly less than $100\mu m$, and a non-homogenous internal reflectivity). The paper goes on to explain that hyperpigmented pigmentary abnormalities are related to intra-retinal hyperreflective foci (IHRF) on OCT. The paper concludes that by detecting these pigmentary variations, CFP has a favourable diagnostic performance for detecting OCT abnormalities. These findings appear to support the conclusion that pigmentary abnormalities are an important predictor of retinal health regarding ARMD, which agrees with the finding from the machine learning process in the current study that pigmentary abnormalities are an important predictor of FAF results.

When five more model learners available via the ODM software are considered however, (kNN, Tree, Random Forest, SVM and neural network), SVM and neural network, utilising both CFP and OCT, do provide an informedness which outperforms hyperparameter fine-tuned Logistic Regression using CFP alone, with SVM delivering the highest informedness of 70.2%, whilst hyperparameter fine-tuned Logistic Regression utilising data from both CFP and OCT continues to deliver the best AUC of 0.903. However, on analysing the best preforming model learner with reference to the performance line on the ROC analysis graph, kNN attains the best results for CFP and OCT data combined, for both detection of a normal (at a probability threshold of 0.800) or an abnormal FAF result (at a probability threshold of 0.400). kNN was the second best model for informedness (65.9%) and also second best for AUC (0.889). Therefore, it is possible to conclude that SVM, hyperparameter fine-tuned Logistic Regression and kNN are the best performing model learners for predicting the outcome of the FAF results in this study.

6.13 Conclusions

Summary of findings from Chapter six which could be used as a series of rules of thumb for primary care practitioners:

1. At 85 years of age, patients have approximately a 50% probability of an abnormal FAF result, and at 92 years this approaches 80%.
2. The number of packet years of smoking can be used for risk analysis, as between 10 to 15 packet years indicates approximately a 50% probability of an abnormal FAF, and at 30 years this is over 90%, although age could be a confounding factor in this finding.
3. For small drusen, as detected either by CFP and OCT, approximately a 50% probability of an abnormal FAF is reached at a level of > 20 < 30, with a high probability of over 90% reached at > 50 drusen.
4. For intermediate drusen, as detected either by CFP and OCT, approximately a 50% probability of an abnormal FAF is reached at >10 < 20 drusen, with a high probability of over 90% reached at > 40 drusen.
5. For large drusen, as detected either by CFP or OCT, approximately a 50% probability or an abnormal FAF is reached at > 10 drusen, with a high probability of over 90% reached at > 20 drusen.
6. For the simplified severity score, a score of 2 indicates approximately a 50% probability of an abnormal FAF result, and this reaches a high probability of over 90% at a score of > 2.
7. If pigmentary abnormalities are detected with CFP then there is over a 90% probability of an abnormal FAF result.
8. If reticular pseudodrusen are detected by OCT, then there is approximately an 80% probability of an abnormal FAF result.
9. Practitioners should be cautious when utilising CFP alone for estimating the lateral dimensions of drusen, with OCT B-scan proving a more accurate modality for these measurements.

Furthermore, the results from the Naïve Bayes and Logistic Regression ODM Nomogram widgets indicate that AI algorithms could be developed for predicting the probability of a patient demonstrating an abnormal FAF result due to ARMD, with an informedness of 60.6%

- 62.1% and an AUC of 0.875 - 0.903, without having to perform the FAF imaging (based on patient history and variables gathered from CFP and OCT). Note that the SVM model learner achieved a higher informedness than Naïve Bayes and Logistic Regression (70.2%), although not in the form of a Nomogram.

Chapter 7 Orange Data mining image analytics

7.1 Introduction

Artificial intelligence has the potential to revolutionise medicine including the rapidly advancing specialism of ophthalmology. IT giants such as Google, Apple and Microsoft have begun to invest billions of dollars into this fledgling industry, hoping to harness the power of computers to analyse data rapidly and effectively. Digital images are becoming utilised in many ophthalmological fields e.g. medical retina and glaucoma, and enormous databases of information are being amassed from screenings performed at many face to face and virtual patient appointments.

Machine learning is a subset of artificial intelligence which focuses on the learning aspect of intelligence and can be harnessed to explore and exploit these digital resources. In supervised machine learning, the computer instructed to refer to human image analysis, with each image in the training data given a descriptive label, whilst unsupervised machine learning searches for patterns within the data without reference to labelling.

The human brain consists of neural structures which computers mimic via artificial neural networks (ANNs) arranged in layers. Deep learning is a computer-based ANN with many layers enabling enhanced performance. As training data passes through the layers it is transformed, and patterns within the data are detected and stored.

Transfer learning involves training machine learners on an image data set applicable to the specific subject after initial pre-training with unrelated images. The object of this chapter was to utilise Transfer Learning for the prediction of the normality or otherwise of FAF images, without the need for the large training datasets that have been used in ophthalmology to date, e.g. by the DeepMind Health and Moorfields Eye Hospital NHS Foundation Trust collaboration. ODM features an “add on” capability which allows for “Transfer Learning” of machine learning models. Transfer learning is where a machine learner is initially trained on an image database which is unrelated to the field of study, with a final, lean image training step carried out to fine-tune the machine learner performance. This could mean that ophthalmologists and optometrists could obtain an “off the peg” pre-trained machine learning algorithm whose training could be “finished off” with a relatively small database of ophthalmological images, with the benefits of the saving of time and financial resources. For this chapter the SqueezeNet image embedding program available within ODM was used, which is trained on the ImageNet library of 14 million non-ophthalmological images of common objects.

7.1.1 Unsupervised machine learning

ODM is capable of importing images, embedding them in a vector space, calculating distances between the vectors, and performing machine learning via hierarchical clustering and data visualisation techniques. Clustering and data projection in this way are an example of “unsupervised” machine learning, and the workflow shown below is an illustration of this. Note that in unsupervised machine learning there is no target variable, with the software exploring patterns within the data without reference to any tagging of images by the researcher.

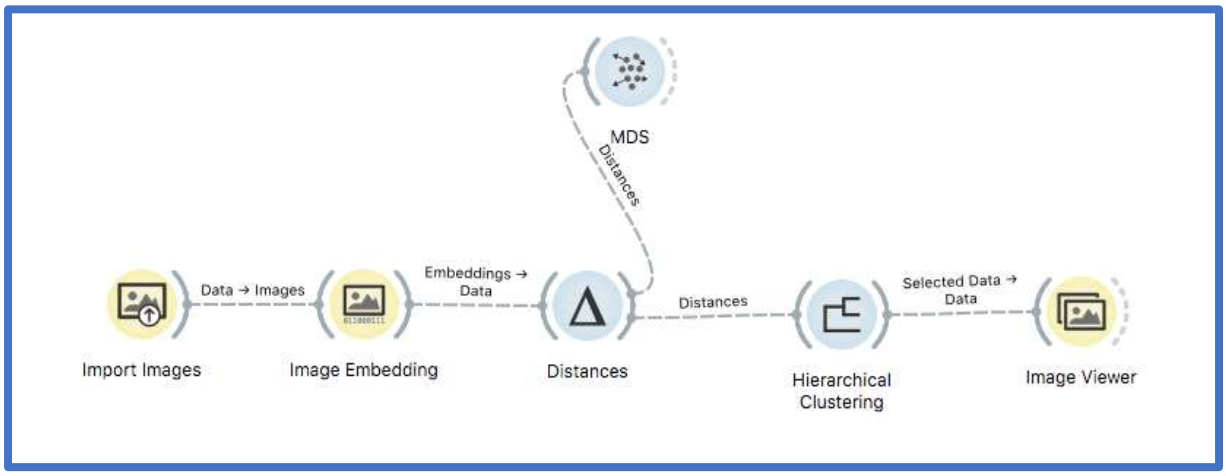

Figure 7.1 Illustration of an Unsupervised Orange Data Mining Image analytics workflow

7.1.2 Supervised machine learning

It is also possible to create a “supervised” machine learning workflow, using it to predict the image class. The workflow is similar to the “unsupervised” machine learning workflow, but with the Hierarchical Clustering widget replaced by the Test and Score widget (connected to model learner widgets), followed by a Confusion Matrix widget and an Image Viewer widget. Note that a Preprocess widget is also providing an input to the Test and Score widget. This allows reduction of the 2048 variables to 18 (i.e. 93/5 = 18.6 ensuring at least a 5 to 1 ratio of data instances to variables) to avoid overfitting which is a risk given the large number of variables and the relatively small number of data instances.

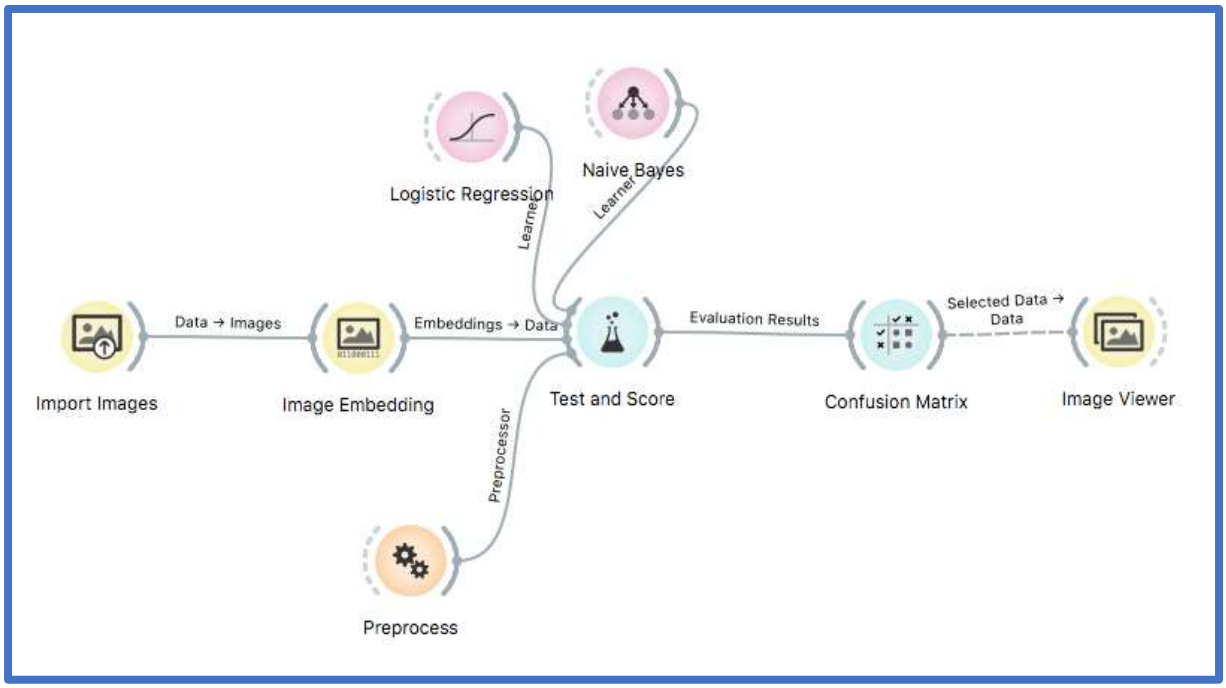

Figure 7.2 Illustration of a Supervised Orange Data Mining Image analytics workflow

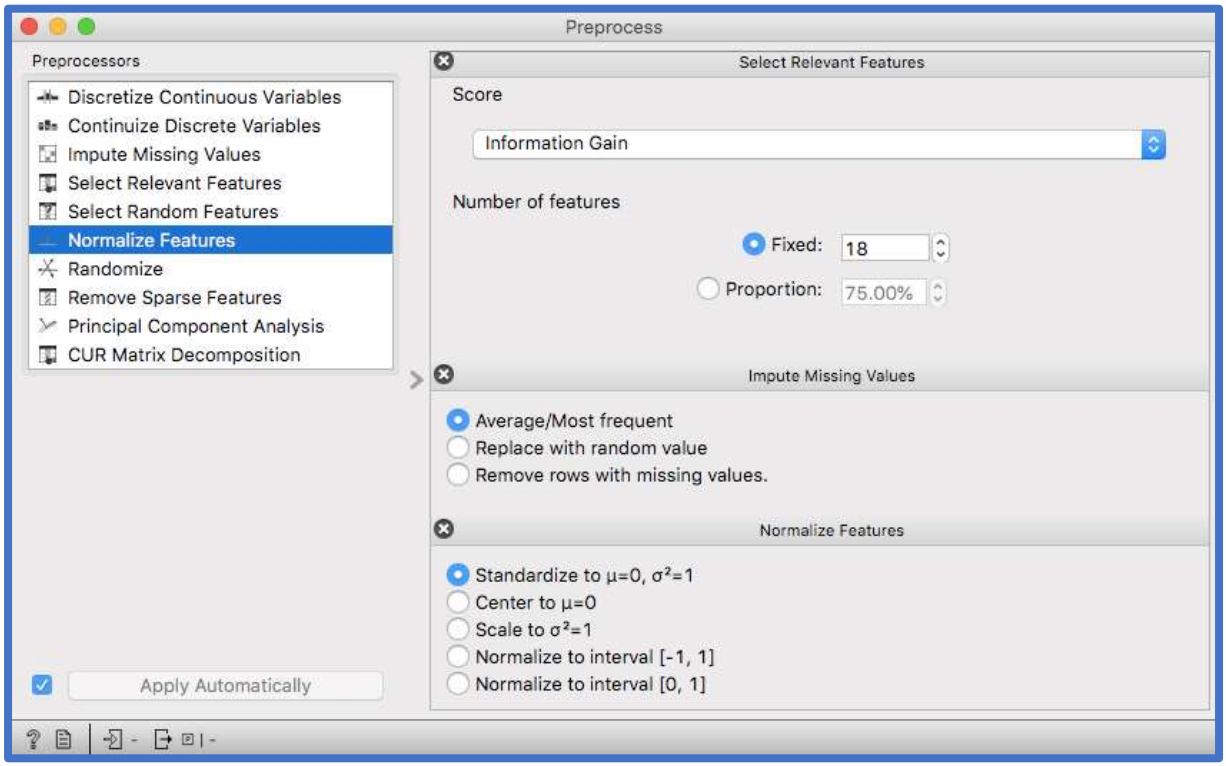

Figure 7.3 Orange Data Mining Preprocess widget selections

Illustration of the contents of the Preprocess widget, showing that the data has been normalised, missing values inputted and only the 18 most informative relevant features have been selected based on Information Gain.

7.2 Methods

7.2.1 Unsupervised machine learning methods

Images from the three imaging methods, colour fundus photography (CFP), optical coherence tomography (OCT) and fundus autofluorescence (FAF) were placed in three separate folders within the PC's hard drive. Each folder was subdivided into two subfolders, one containing the images taken from participants who demonstrated a normal FAF image, and the other containing images from participants with an abnormal FAF image. Unsupervised machine learning, without reference to human researcher tagging, was performed using the workflow above in Figure 7.1, for each of the three imaging methods.

7.3 Results

The data can be visualised in the Hierarchical Clustering widget, however, no clear clustering pattern for normal and abnormal FAF is evident, as shown below.

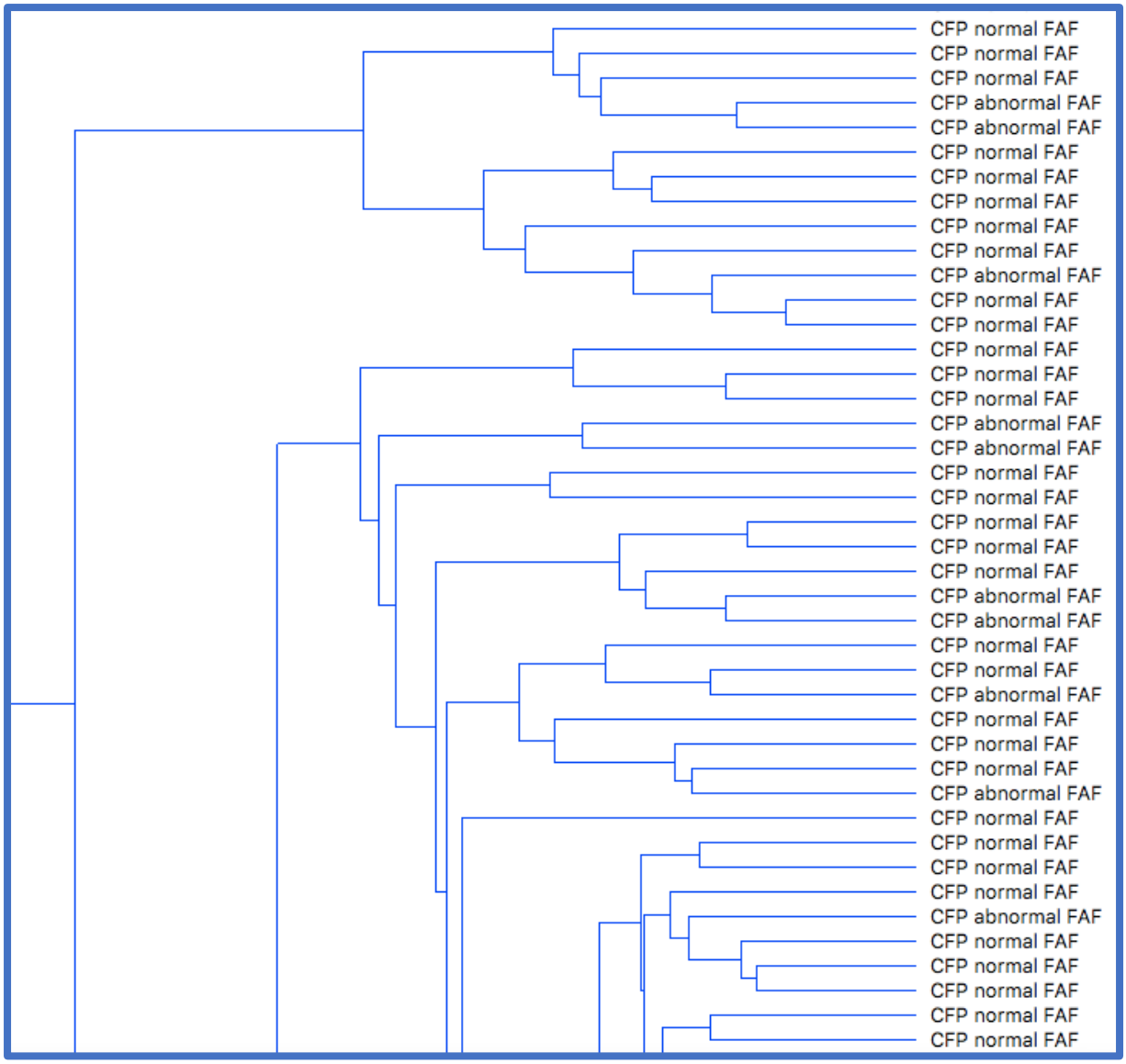

Figure 7.4 Orange Data Mining Hierarchical Clustering widget output for CFP images from participants with normal and abnormal FAF results.

Alternatively, as shown below in the Figure below, the output for CFP is illustrated by the Multidimensional Scaling (MDS) widget. Blue dots represent images with an abnormal FAF image, and red dots a normal FAF image.

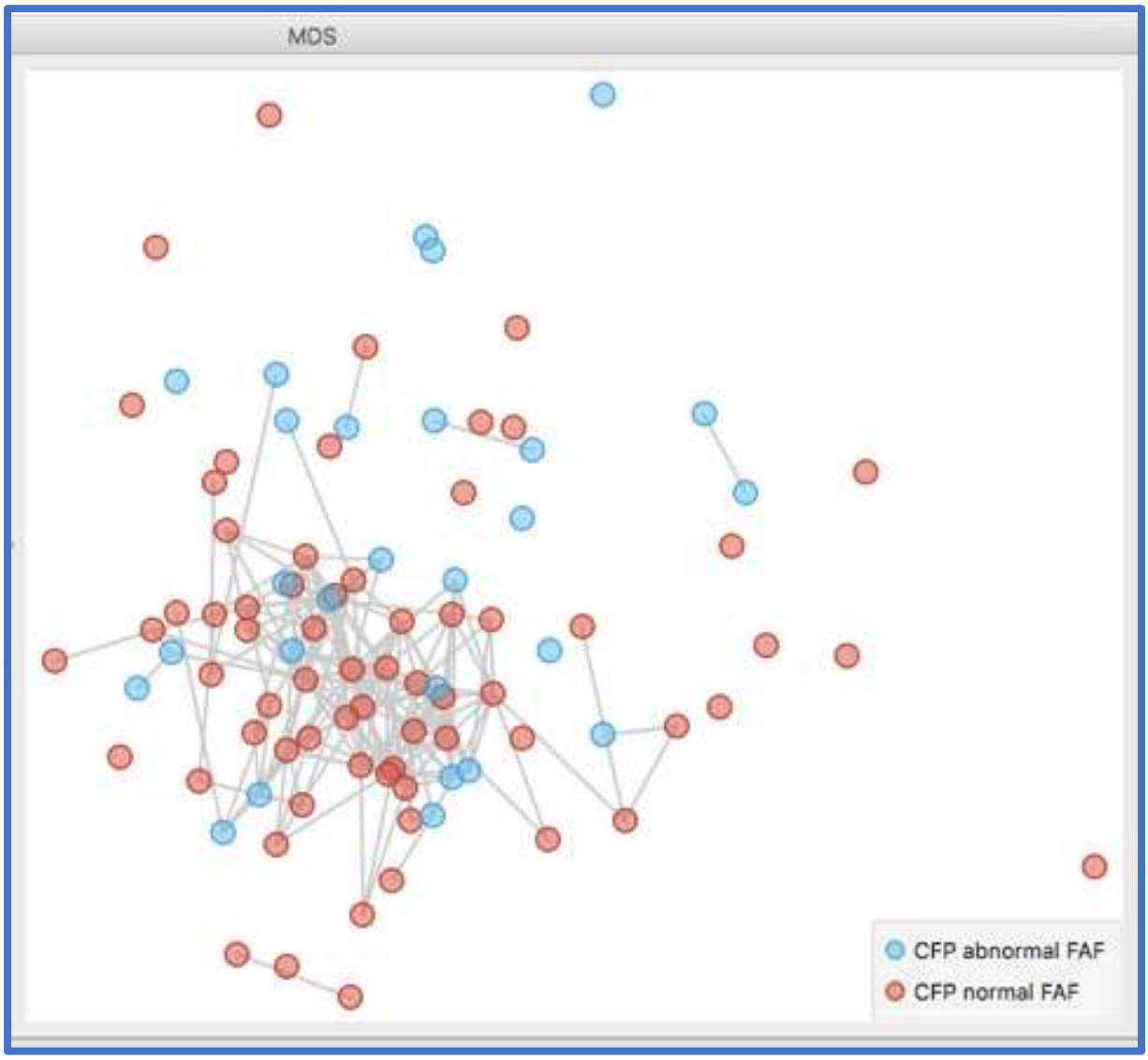

Figure 7.5 MDS output for unsupervised machine learning for CFP images, blue dots represent an abnormal FAF and red dots a normal FAF

To investigate clustering of the points in the Figure below, two blue dots from a similar location in the diagram have been selected and appear highlighted by a yellow “halo”.

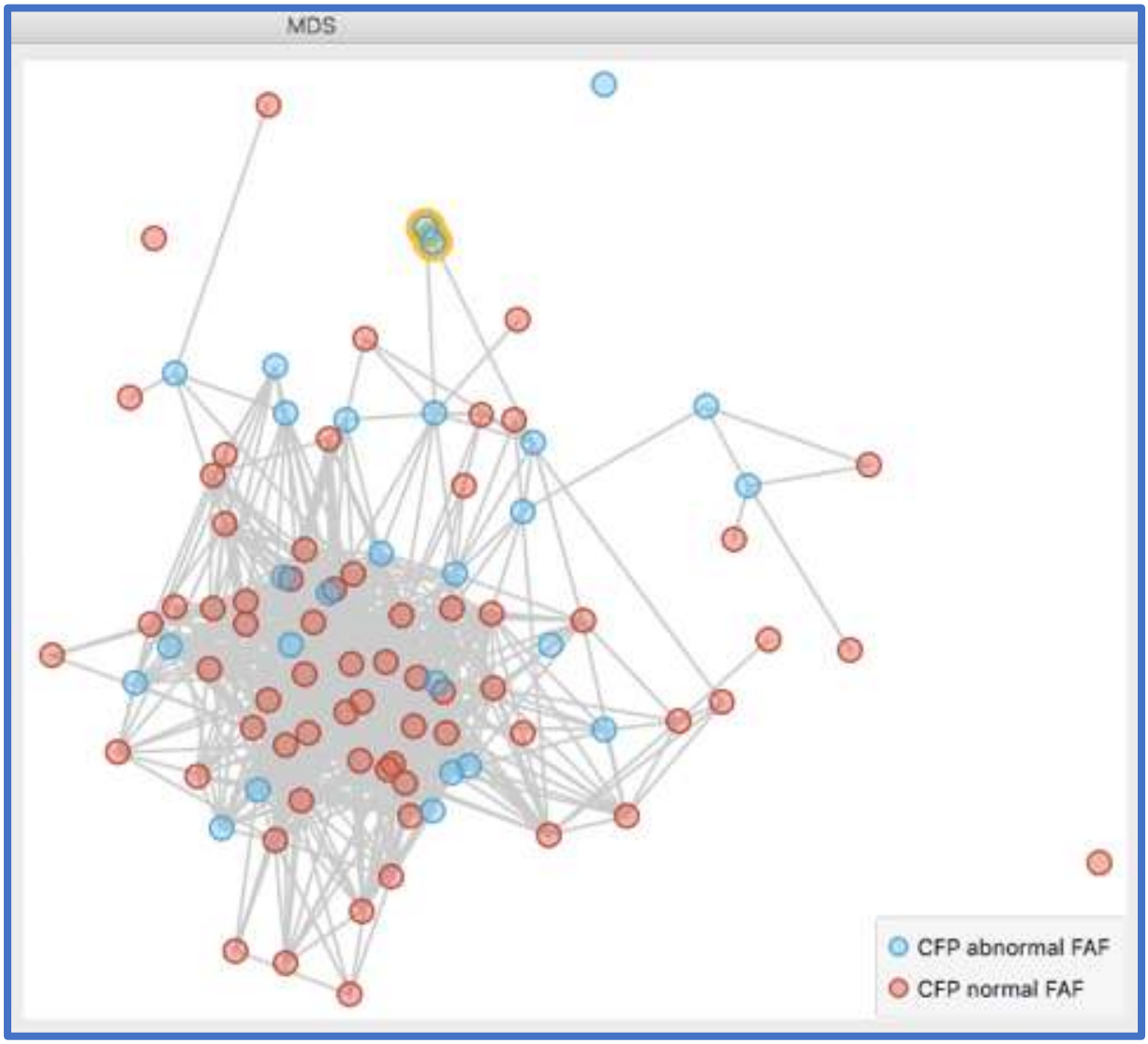

Figure 7.6 MDS output for unsupervised machine learning for CFP images with selections, blue dots represent an abnormal FAF and red dots a normal FAF

Below are shown the clinical features of the two closely clustered blue dots as selected. From this simple analysis it is difficult to ascertain the reason for the very close clustering of these two images. Both are pseudophakic, however, they differ greatly in the grading of the number of small, intermediate and large sized drusen. One demonstrates pigmentary abnormalities and geographic atrophy, the other does not. Neither case was recorded has having RPD or neovascular ARMD.

Case number	**58**	**38**
Pseudophakic (Y/N)	Y	Y
CFP Small drusen grade	6	1
CFP Intermediate drusen grade	6	1
CFP Large drusen grade	6	1
Pigmentary abnormalities (Y/N)	N	Y
Geographic atrophy (Y/N)	N	Y
RPD (Y/N)	N	N
Wet ARMD (Y/N)	N	N

Table 7.1 Showing the clinical features for the two images selected shown in Figure 7.6

Below, are the results from the same workflow carried out for OCT images, with the output from the MDS widget again illustrating no clear pattern in the spread of the data points.

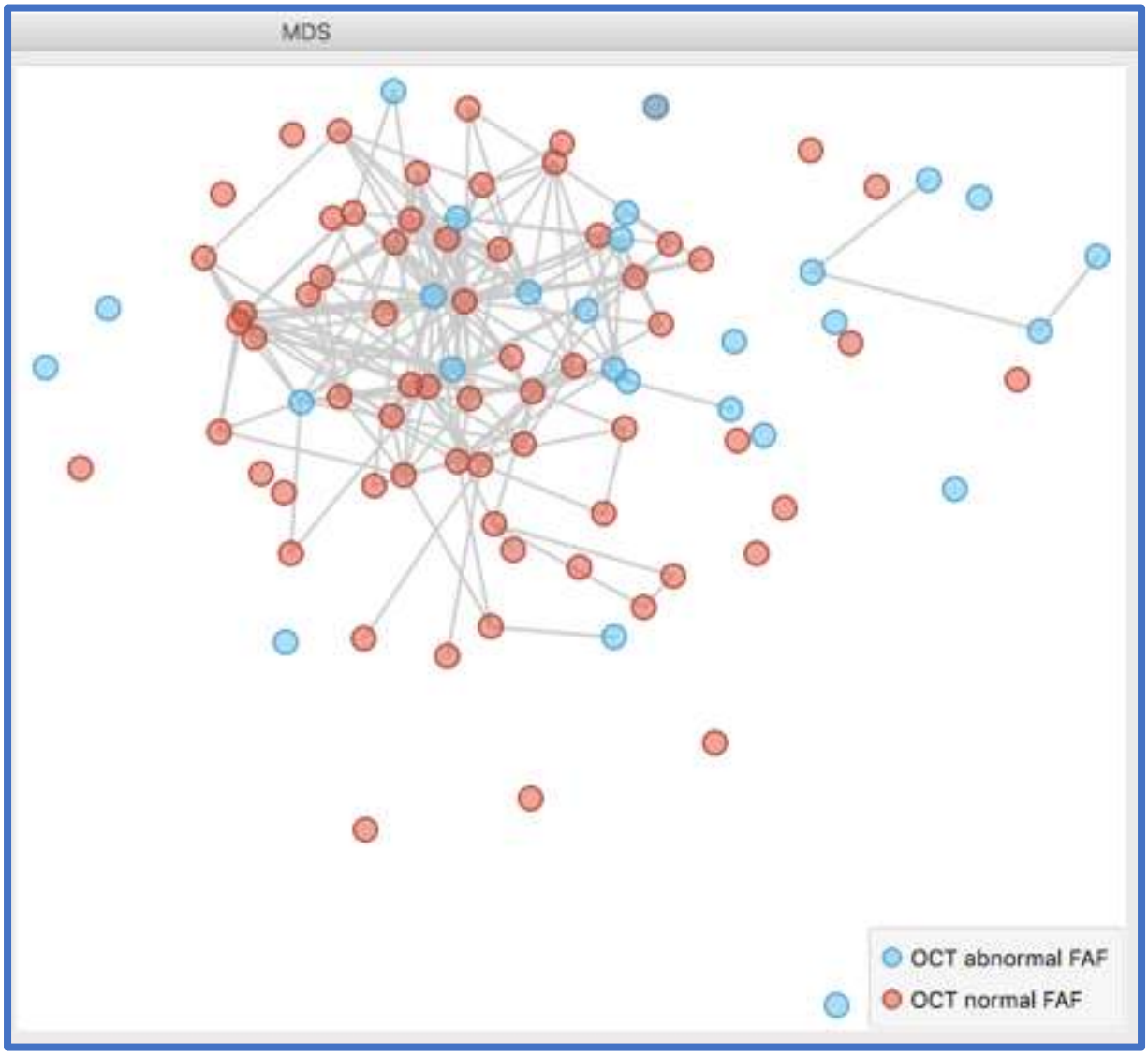

Figure 7.7 MDS output for unsupervised machine learner for OCT images, blue dots represent an abnormal FAF and red dots a normal FAF

Below is from the same workflow again, this time illustrating the spread of data points within the MDS widget from FAF images. From this data, there appears to be a heavier concentration of red dots in the upper left of the plot, with more blue dots towards the bottom.

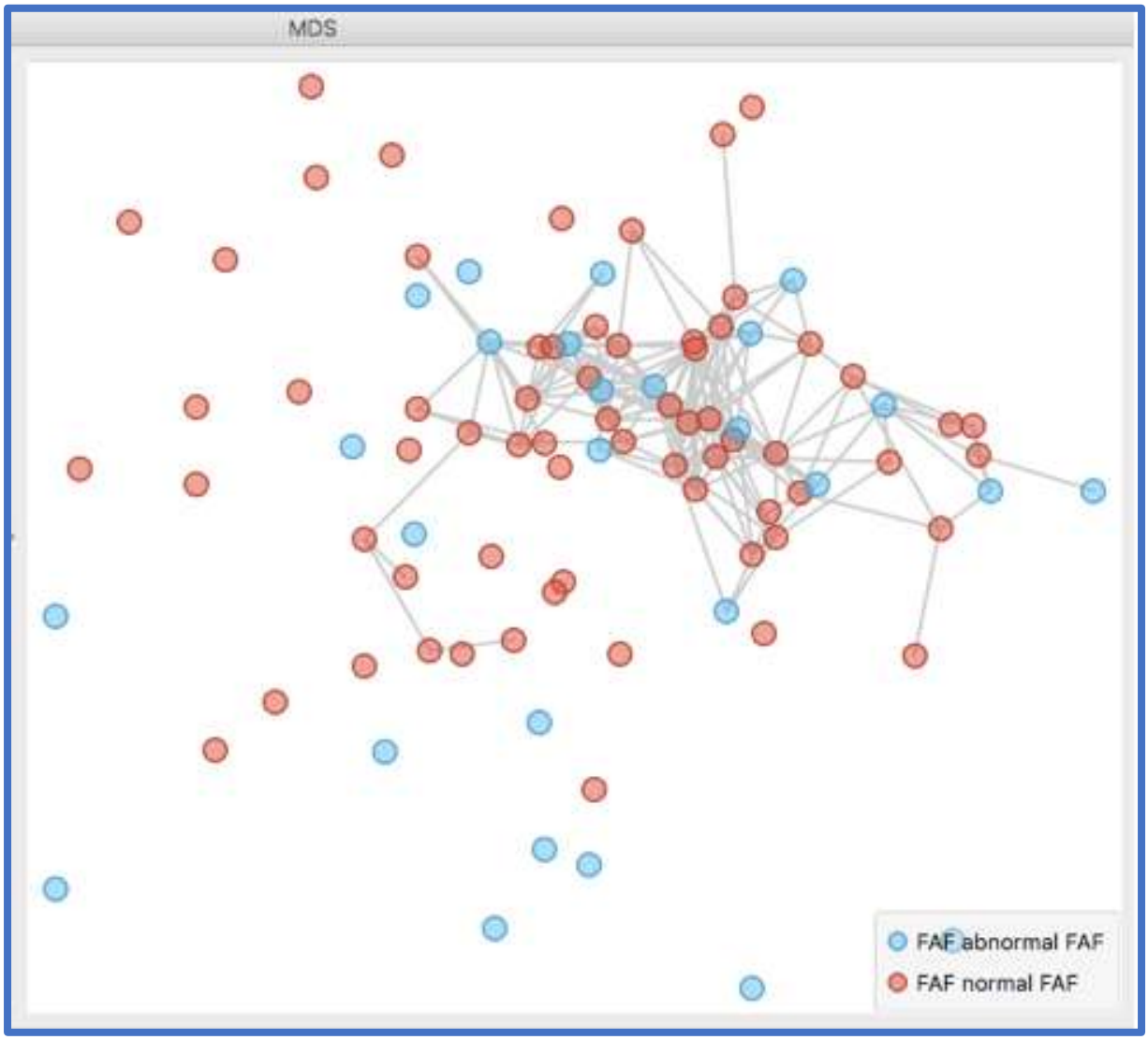

Figure 7.8 MDS output for unsupervised machine learner for FAF images, blue dots represent an abnormal FAF and red dots a normal FAF

To analyse the results in detail, we can select these images and view them via an Image Viewer widget joined to the MDS widget, as shown in Figures below.

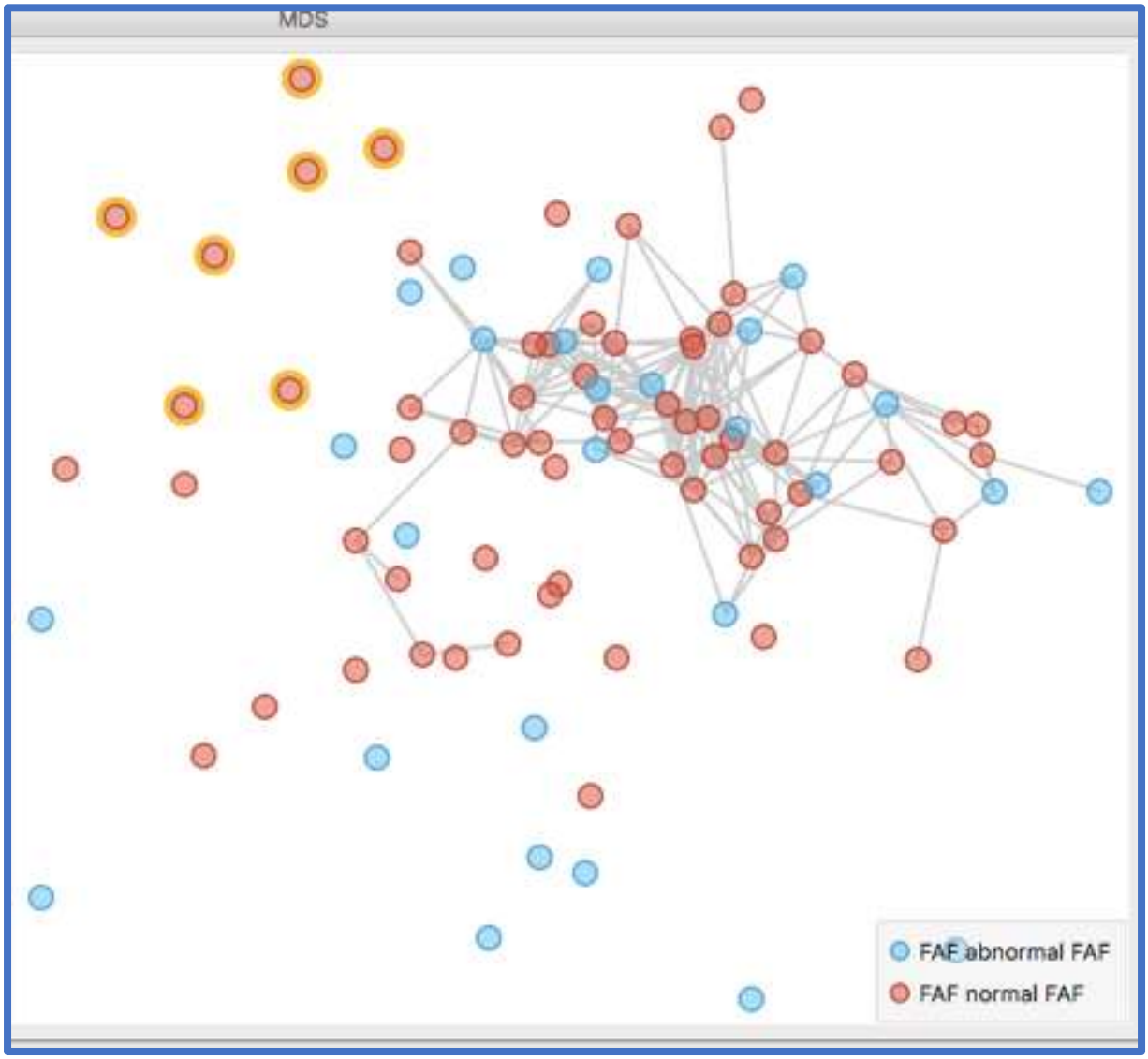

Figure 7.9 MDS for unsupervised machine learner for FAF images with selections. Blue dots represent abnormal FAF and red dots normal FAF.

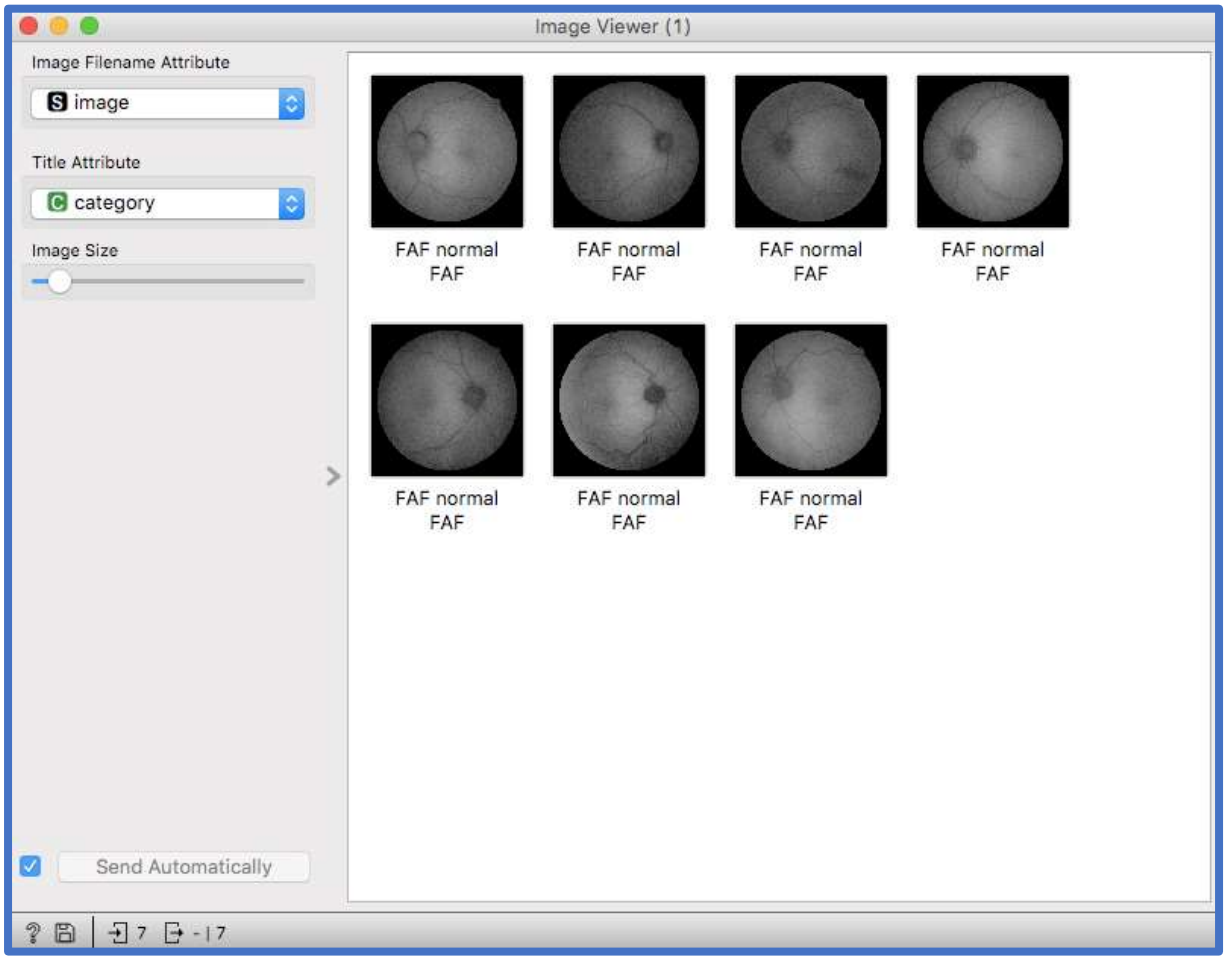

Figure 7.10 Image Viewer for unsupervised machine leaner for FAF images with normal selections.

The seven images selected in Figure 7.10 demonstrate a mixture of normal and minimal change FAF patterns, as shown in Table 7.2 below. Just over 47% of all 93 images taken in the study are classified as normal or minimal change. ODM may be separating images based on their FAF classification, however, more analysis, beyond the score of this study, would be required to corroborate how the images are being dealt with by the image analytics software. In Figures 7.8, and 7.9 the upper right area of the plots still contains many of the red and blue dots clustered together, so it is reasonable to assume that any clustering of normal and abnormal FAF images is weak at best.

Case number	**FAF classification**
5	Minimal change
4	Normal
93	Minimal change
81	Normal
28	Normal
27	Minimal change
89	Minimal change

Table 7.2 Showing FAF classification for each of the images selected shown in Figures 7.9 and 7.10

Figure 7.11 An example a normal FAF image from figure 7.10

Below in Figure 7.12 is the MDS output for unsupervised machine learning for FAF images, but this time with the cluster of blue dots representing abnormal FAF images highlighted with a yellow halo.

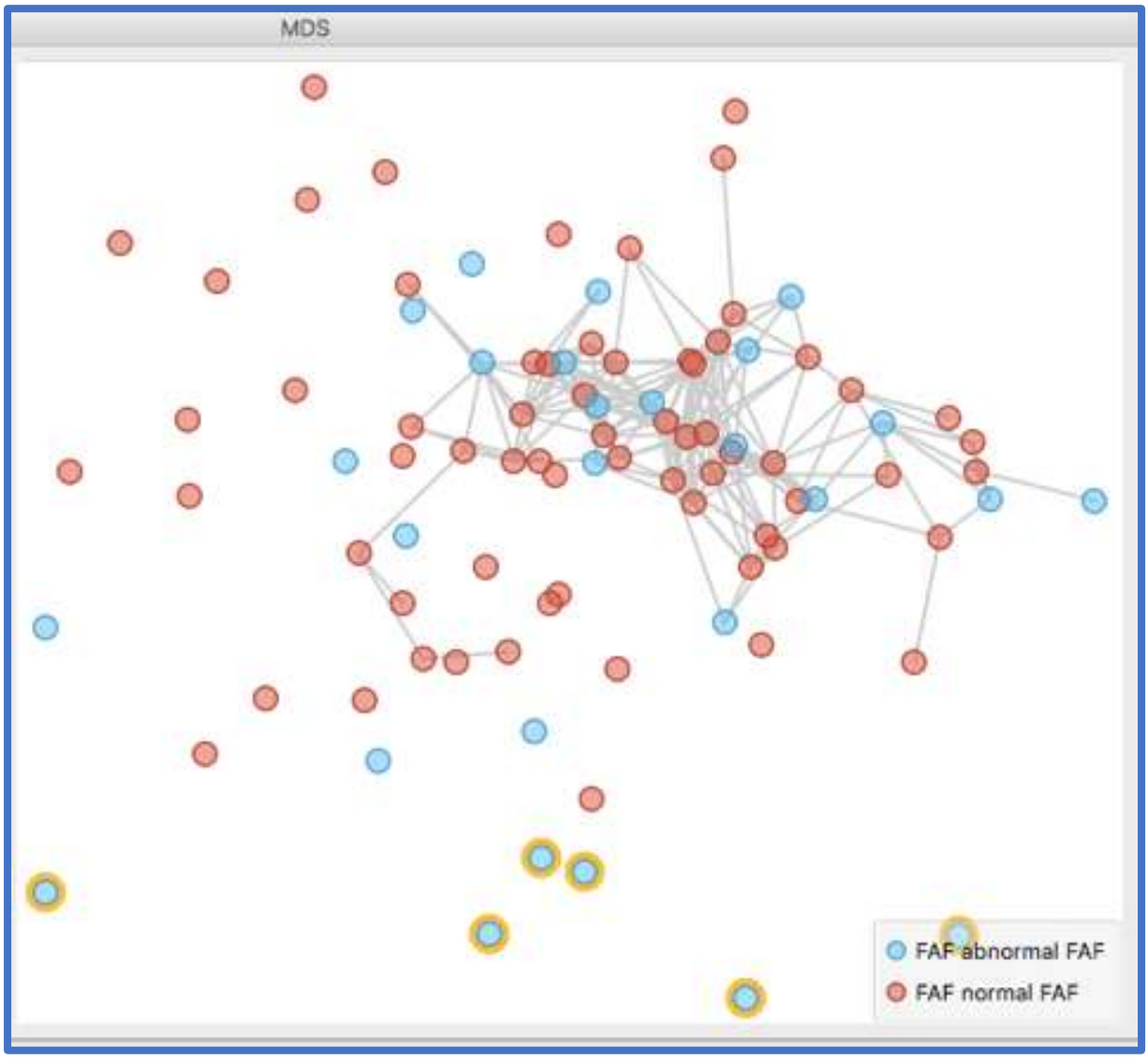

Figure 7.12 MDS for unsupervised machine learner for FAF images with abnormal selections blue dots represent an abnormal FAF and red dots a normal FAF.

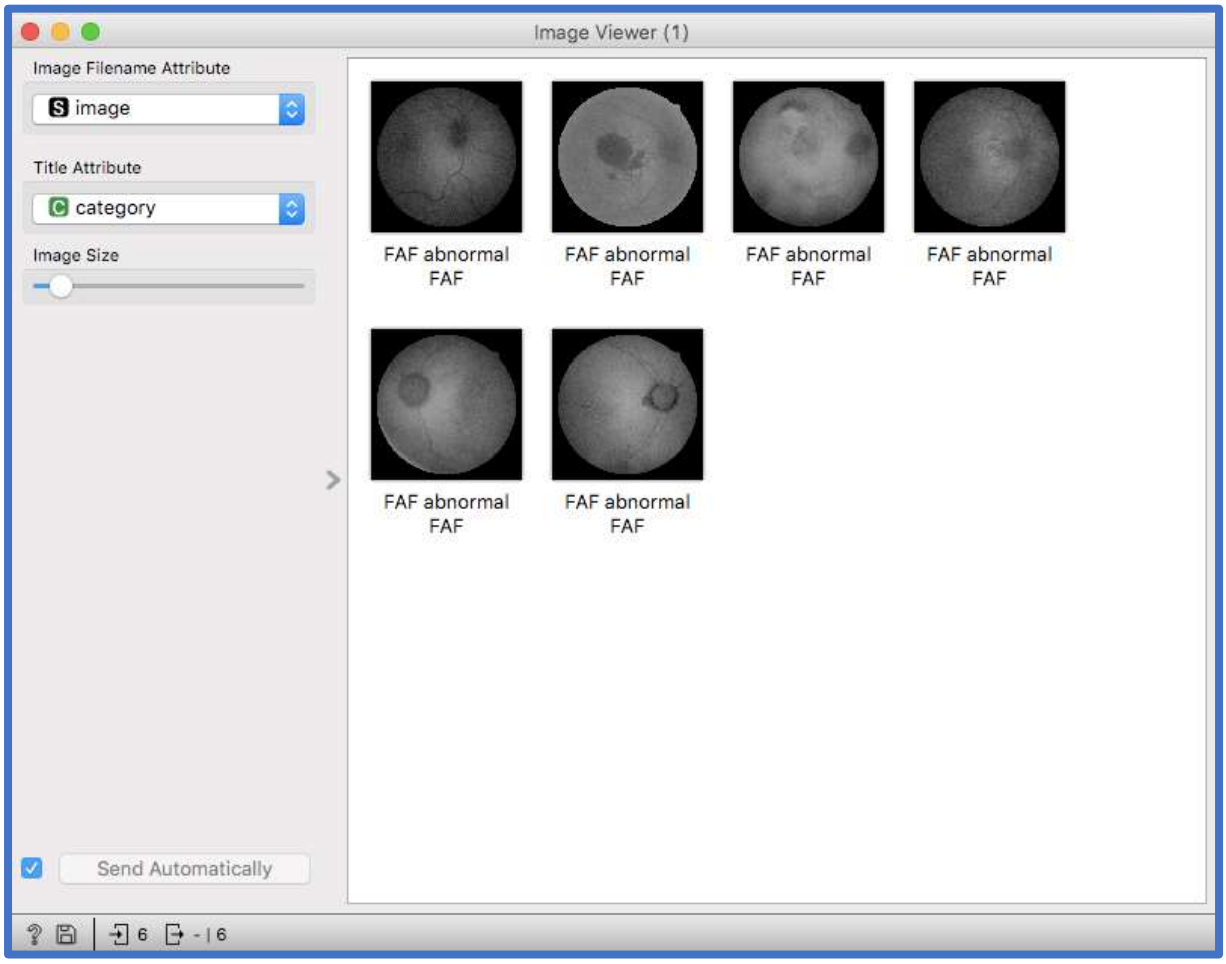

Figure 7.13 Image Viewer for unsupervised machine learner for FAF images with abnormal selections from Figure 7.12.

All six images selected in Figure 7.12 demonstrate a patchy or reticular pattern of FAF, as shown in Table 7.3 below. Just over 19% of all 93 images are classified as patchy or reticular. Again, ODM may appear to be having limited success in clustering images by virtue of their specific FAF classification, however, two thirds of all abnormal FAF images were graded in this study as either patchy or reticular, and only a third of all patchy or reticular images were included in this cluster. Therefore, we can again reasonably assume that any clustering of FAF images as normal or abnormal, or of a specific pattern within the abnormal class, is weak at best.

Case number	FAF classification
10	Reticular
58	Patchy
23	Patchy
38	Patchy
59	Patchy
60	Reticular

Table 7.3 Showing FAF classification for each of the images selected in Figures 7.12 and 7.13

Table 7.4 A typical FAF image from an abnormal result from Figure 7.13

7.4 Supervised machine learning methods

Supervised machine learning was also carried out using the workflow shown in Figure 7.2, reproduced again below for reference in Figure 7.14.

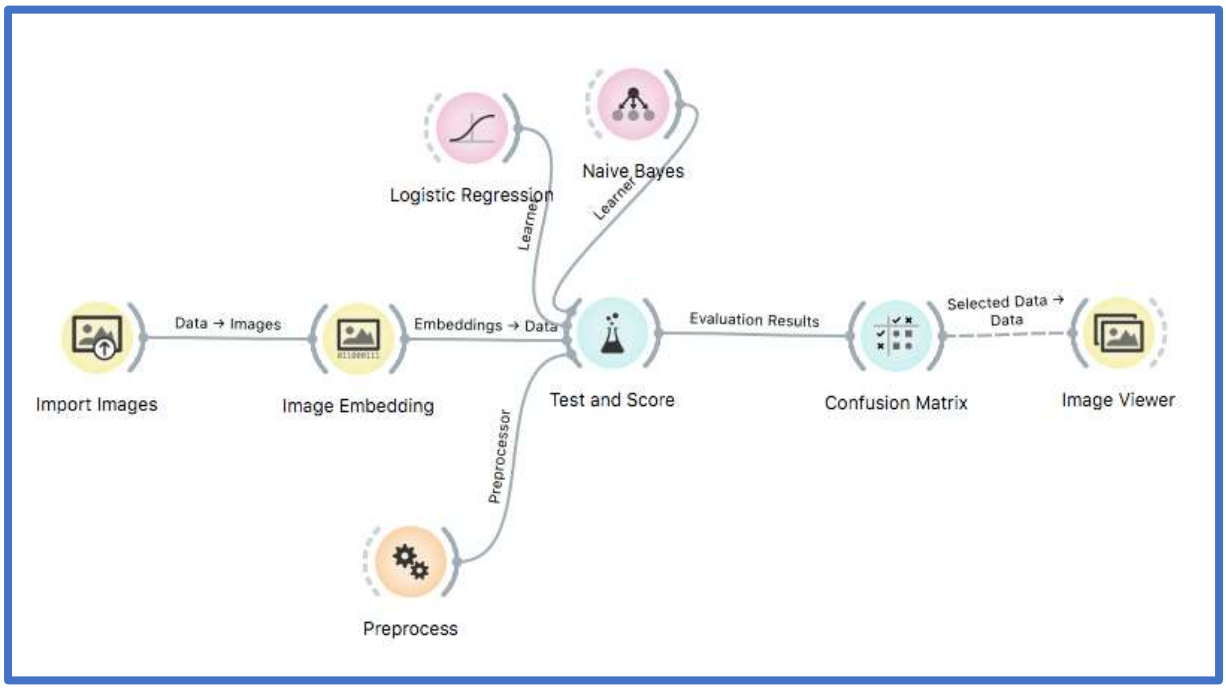

Figure 7.14 Supervised machine learning workflow taken from Orange Data Mining's Image Analytics "add-on".

Here the images were again uploaded in separate folders for CFP, OCT and FAF, with each folder containing two subfolders, one for participants with a normal FAF image, and another for those participants with an abnormal FAF image. Note that, by using the Test and Score widget, the workflow is now utilising two model learners (Naïve Bayes and Logistic Regression) to predict image class based on human researcher image tagging, with reference to a target variable (normality/abnormality of the FAF image), unlike with the unsupervised machine learning Distance widget which purely facilitates the graphical illustration of distances between vectors.

7.5 Results

The results are shown below in Table 7.5.

Imaging utilised by ML	Best performing model learner	AUC	Informedness	Best performing model learner with R flip	AUC with R flip	Informedness achieved with R flip
CFP alone	SVM	0.484	14.0%	kNN	0.627	13.8%
OCT alone	kNN	0.613	17.5%	Tree	0.584	21.7%
FAF alone	Logistic Regression	0.551	20.4%	Tree	0.727	40.4%
CFP and OCT combined	Naïve Bayes	0.550	19.4%	Logistic Regression	0.612	23.4%
CFP, OCT and FAF combined	kNN	0.551	13.8%	kNN	0.579	20.3%

Table 7.5 Showing Orange Data Mining supervised machine learning results.

For analysis of CFP alone, the Preprocessing widget was set to select only the 18 most informative variables based on Information Gain. This was to maintain an instance to variable ratio of at least five (93 instances/18 variables = 5.17). Informedness scores were poor for all of the model learners used. SVM gave the best result, however, this learner still only managed an informedness of 14.0% (0.731+0.409-1)*100 with an AUC of 0.484 (essentially the random guess AUC of 0.500)(199). The MDS widget below also shows disappointing results for CFP with the blue and red dots appearing on manual inspection to be randomly distributed as shown in Figure 7.15 below.

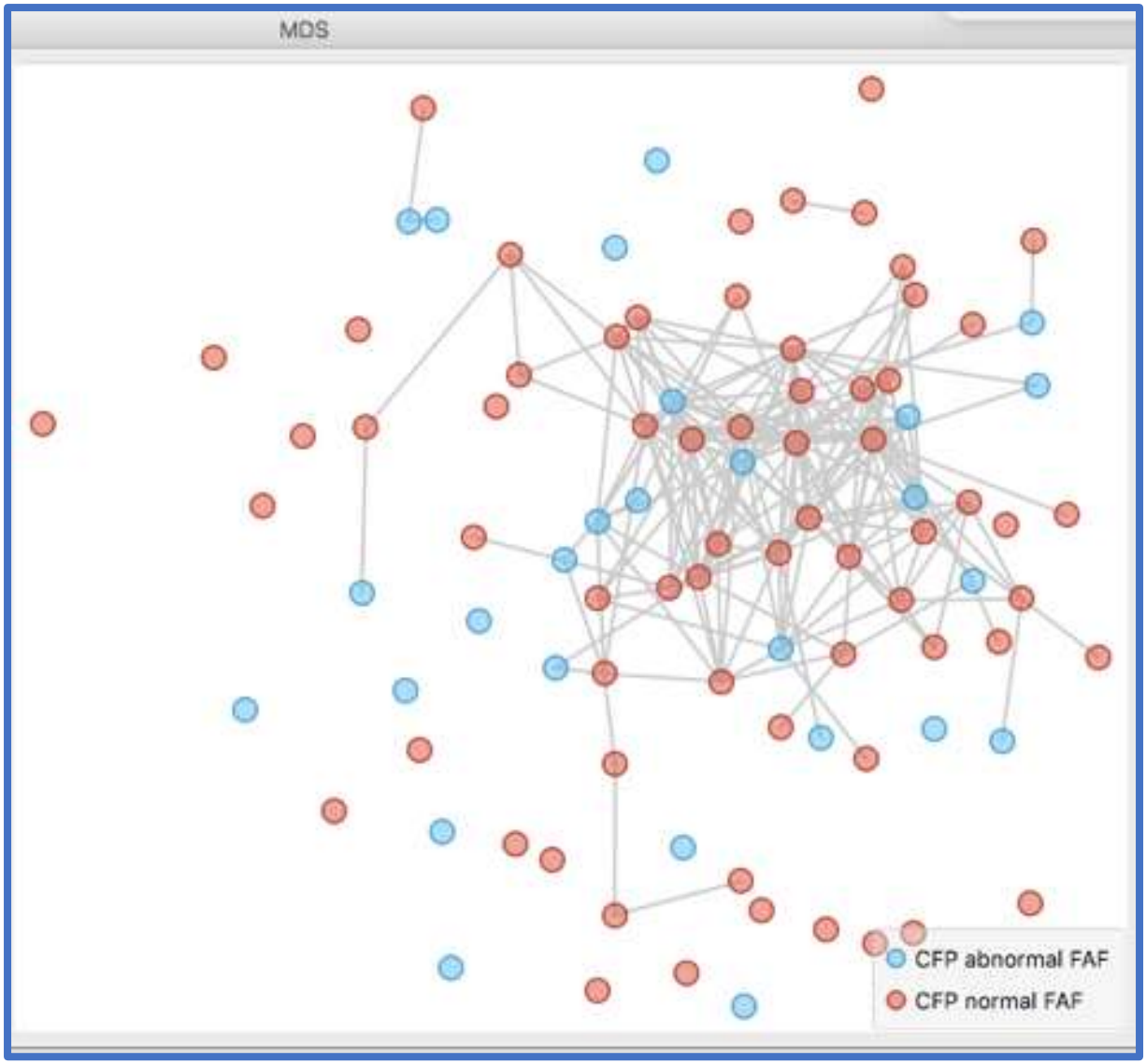

Figure 7.15 MDS for supervised machine learning for CFP images. Blue dots represent an abnormal FAF and red dots a normal FAF.

For analysis of OCT alone, the Preprocessing widget was again set to select only the 18 most informative variables based on Information Gain. The informedness scores are poor for all the model learners used. kNN had the best result, just outperforming SVM using CFP alone. However, the kNN learner still only managed an informedness of 17.5% (0.710+0.465-1)*100 with an AUC of 0.613 (an AUC of 0.6-0.7 is considered poor).(199) The MDS widget below also shows poor results for OCT, with the blue and red dots appearing on manual inspection to be randomly distributed as shown if Figure 7.16 below.

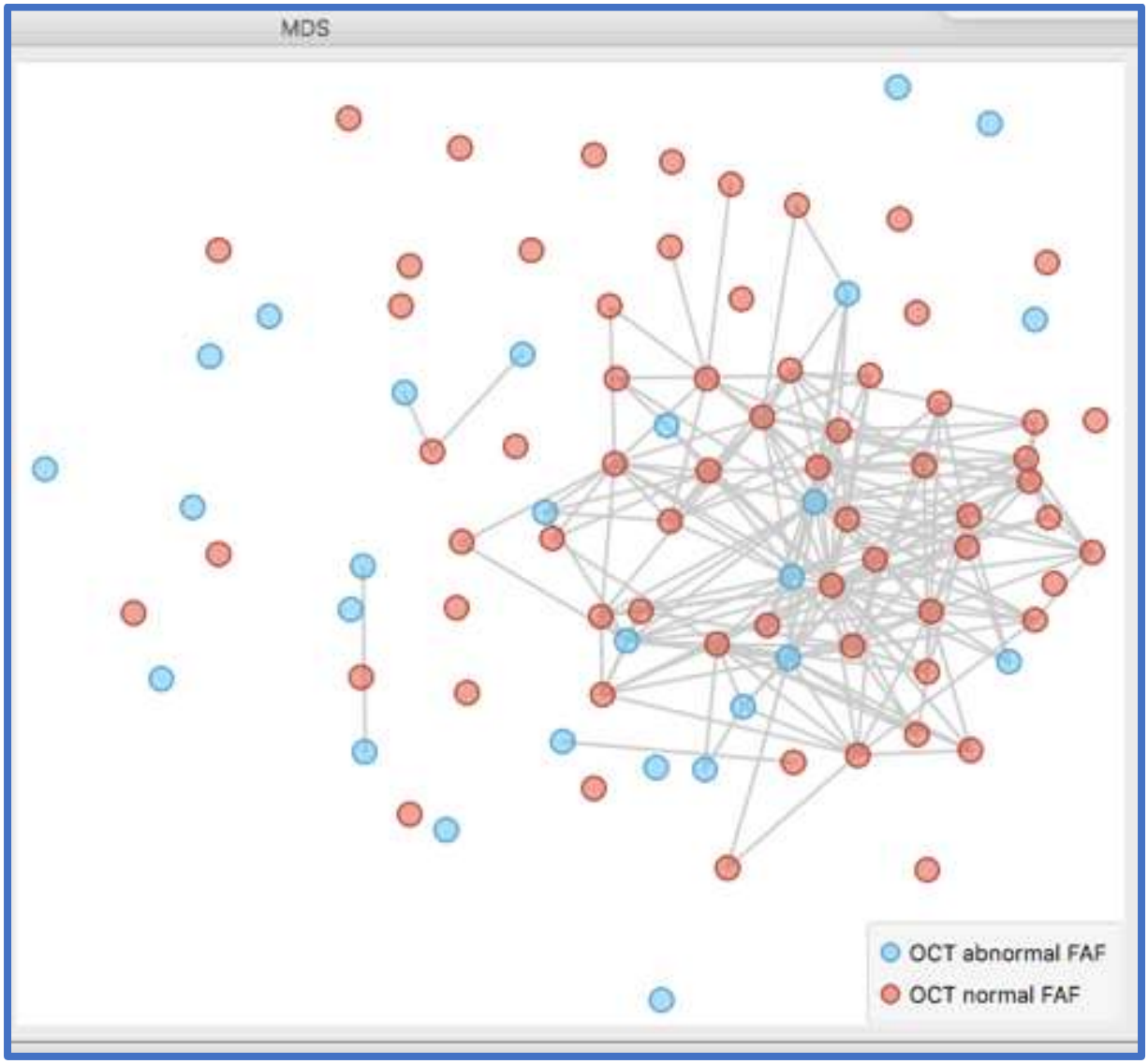

Figure 7.16 MDS for supervised machine learning for OCT images. Blue dots represent an abnormal FAF and red dots a normal FAF.

For analysis of FAF alone, the Preprocessing widget was again set to select only the 18 most informative variables based on Information Gain. Supervised machine learning results are informative, albeit with slightly better results than for CFP and OCT, with an informedness generated by hyperparameter fine-tuned Logistic Regression of 20.4% (0.699+0.505-1)*100 with an AUC of 0.551 (AUC of 0.5-0.6 is considered to have failed).(199) The MDS output is also poor, with no clear pattern of point separation emerging, shown in Figure 7.17 below.

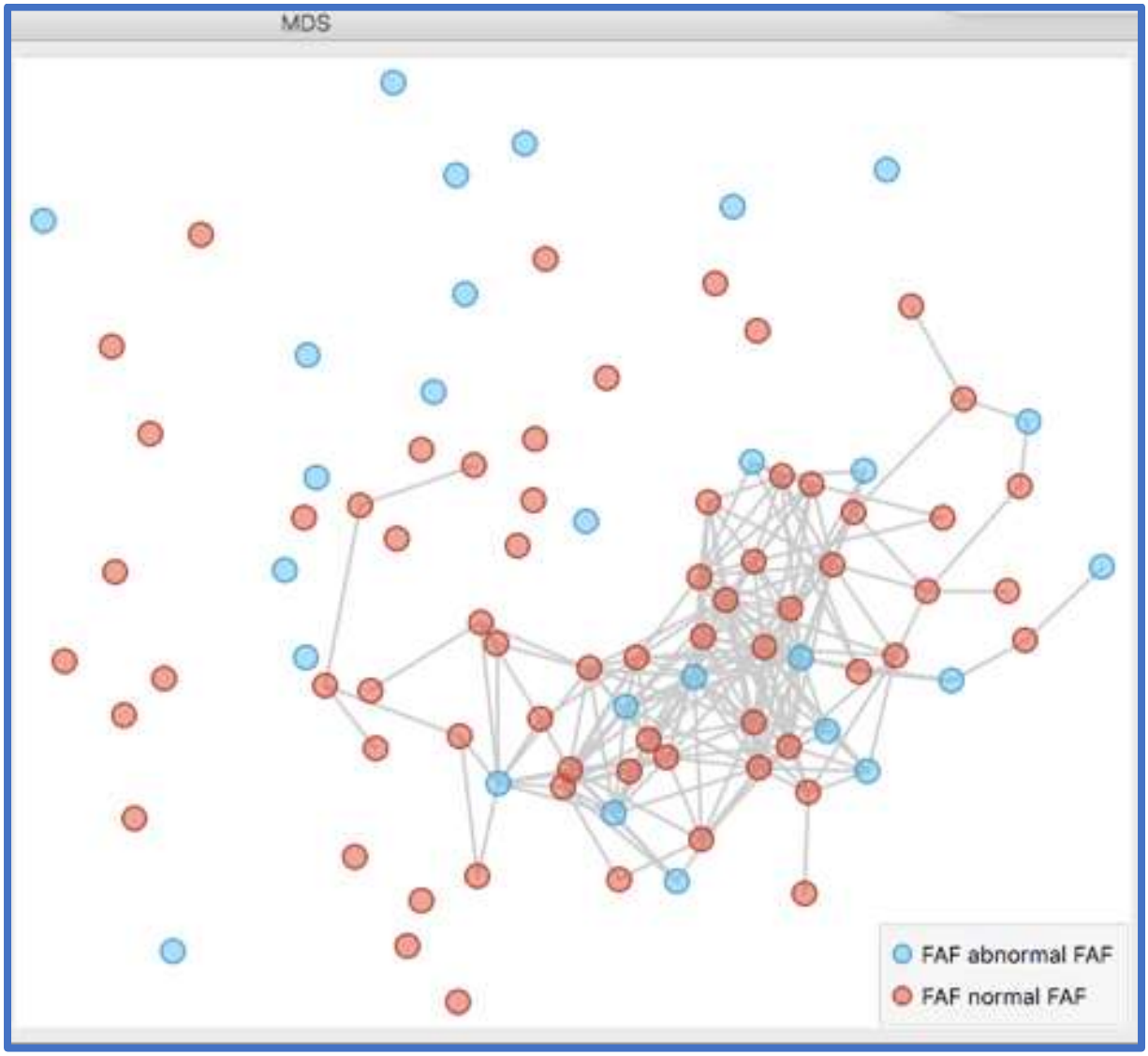

Figure 7.17 MDS for supervised machine learning for FAF images. Blue dots represent an abnormal FAF and red dots a normal FAF

A combined approach for predicting a normal or abnormal FAF result was also tested, with CFP and OCT combined, followed by CFP, OCT and FAF combined. Combining CFP and OCT analysis provides a better informedness via Naïve Bayes than utilising both imaging techniques separately at 19.4%, however, when combining the results of all three modalities the best informedness via kNN falls again to 13.8%. It is important to note that in most previous studies utilising transfer learning in ophthalmology, images had been extensively pre-processed prior to using machine learning on the data,(200),(201),(202),(203), and this may explain the poor results in this study, i.e. the image data used in this study was simply "too raw" for analysis.

The last three columns on the right in Table 7.5 show the results from the analysis after an additional step has been carried out, by performing a "horizontal flip", therefore making left images appear as a right eye for all three modalities, in an attempt to investigate whether this simple step would make a material difference to the informedness and AUC results. However, there was not a significant improvement in the informedness levels achieved, with the exception of the Tree model learner. Tree learners are, however, prone to overfitting, with Random Forest learners generally considered to provide superior accuracy due to their use of multiple trees which "vote" to give a consensus.(204) For the FAF analysis Random Forest achieved an informedness of 7.1% with an AUC of 0.577 (not shown in the table as this was not the best performing model), rising to only 9.1% with an AUC of 0.576 with a R flip, so it is reasonable to assume that the Tree learner was overfitting in this case.

7.5.1 An Alternative Approach Using Features Identified by Principal Component Analysis

Instead of attempting to use data from CFP, OCT and FAF image data to predict FAF outcomes, an alternative approach was to consider the clinical features selected as contributing strongly to the principal components identified in Chapter nine covered later in this paper. These clinical features were drusen (and in particular large drusen), geographic atrophy, pigmentary abnormalities, and age. All drusen types were included in principal component one, however, large drusen as detected by OCT was the most informative variable as determined via the Rank widget. Therefore, large drusen were selected for the subsequent closer analysis as follows.

7.5.1.1 Large Drusen

For large drusen, OCT was selected as the imaging method for exploration via the Test and Score widget. This rationale was based on the finding that there was close agreement between when large drusen were detected as being present by CFP or by OCT. In the 39 participants in whom large drusen were not detected by OCT, CFP agreed in all cases. In the remaining 54 participants in whom large drusen were detected by OCT, CPF disagreed in only 6 cases. For analysis of large drusen as detected by OCT, the preprocessing widget was again set to select only the 18 most informative variables based on Information Gain. Hyperparameter fine-tuned Logistic Regression was the best performing model learner delivering only 16.4% informedness (0.581+0.583-1) and an AUC of 0.573. The MDS widget output also gives a disappointing random spread indicating no clear separation of points.

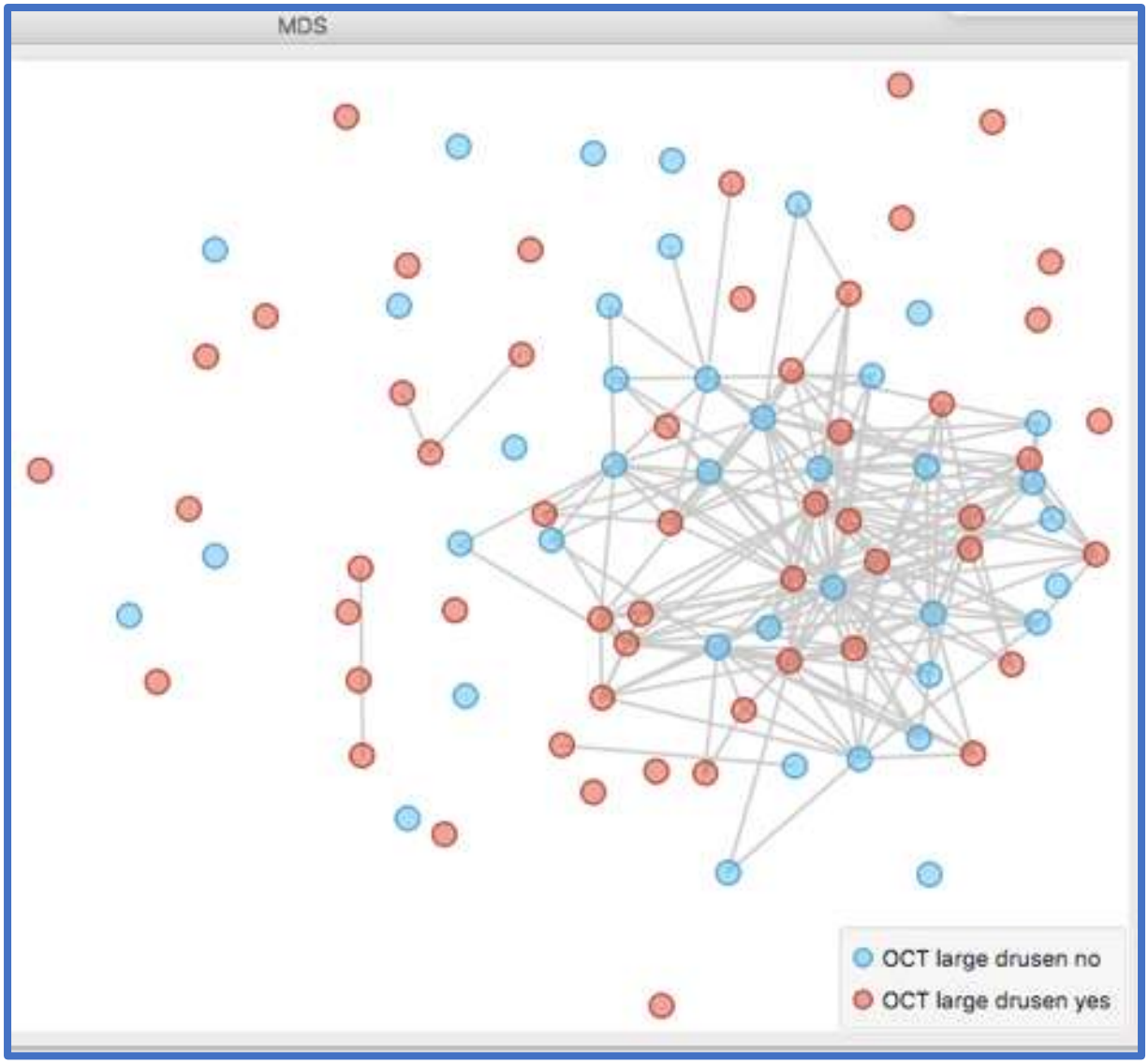

Figure 7.18 MDS for supervised machine learning for large drusen detected vis OCT imaging. Blue dots indicate no large drusen, red dots indicate large drusen present.

7.5.1.2 Geographic atrophy

For this analysis of geographic atrophy FAF imaging results were used. FAF was chosen rather than CFP or OCT due to FAF having an ability to demarcate areas of GA with high accuracy, according to the literature review in Chapter one.(48) However, there were only four instances of GA indicated by the FAF images in this study, and 89 cases without GA. Therefore, we have an imbalanced data problem, with the likelihood that the algorithm will learn how to spot the overrepresented negative cases rapidly, but will struggle to learn how to spot the rare positive instances. Chicco(193) suggests this can be mitigated by altering the training set with a 50% pick-up of the average value between 50% and the real proportion percentage. This calculation is shown below:

89 cases with no GA on FAF
4 cases with GA on FAF
= 93 cases in total

89/93 = 95.7% negative cases
4/93 = 4.3% positive cases

(95.7% + 50%)/2 = 72.85% rounded to 73%
(4.3% + 50%)/2 = 27.15% rounded to 27%

The 4 positive cases need to represent 27% of the total cases used in the analysis. Therefore, the total sample needs to be 4/0.27 = 14.8, rounded up to 15. Therefore only (15 – 4) 11 negative cases should be included in the sample (this is known as “under-sampling” of the majority class within the data). Note that for this calculation, “leave one out” was selected in the Test and Score widget, where the number of folds equals the number of instances, 15 in this case. This method is more appropriate for the smaller dataset that was left after allowance for the imbalanced data problem. The Preprocessing widget was also adjusted for this new number of instances. As previously, a minimum instance to variable ratio of five to one was selected, and therefore only the three (15/5) most informative variables were selected based on Information Gain. The results appear encouraging, however, with only four instances for GA present with FAF, the results must be viewed with caution. The best performing learner was Random Forest, giving an informedness of 90.9% (0.933+0.976-1) and an AUC of 0.955 (considered excellent).(199) The MDS widget, by the

projection of distances between points onto a 2-dimensional plane, also appears to having success in separating the cases where GA is present and not present as detected by FAF shown in the Figure below.

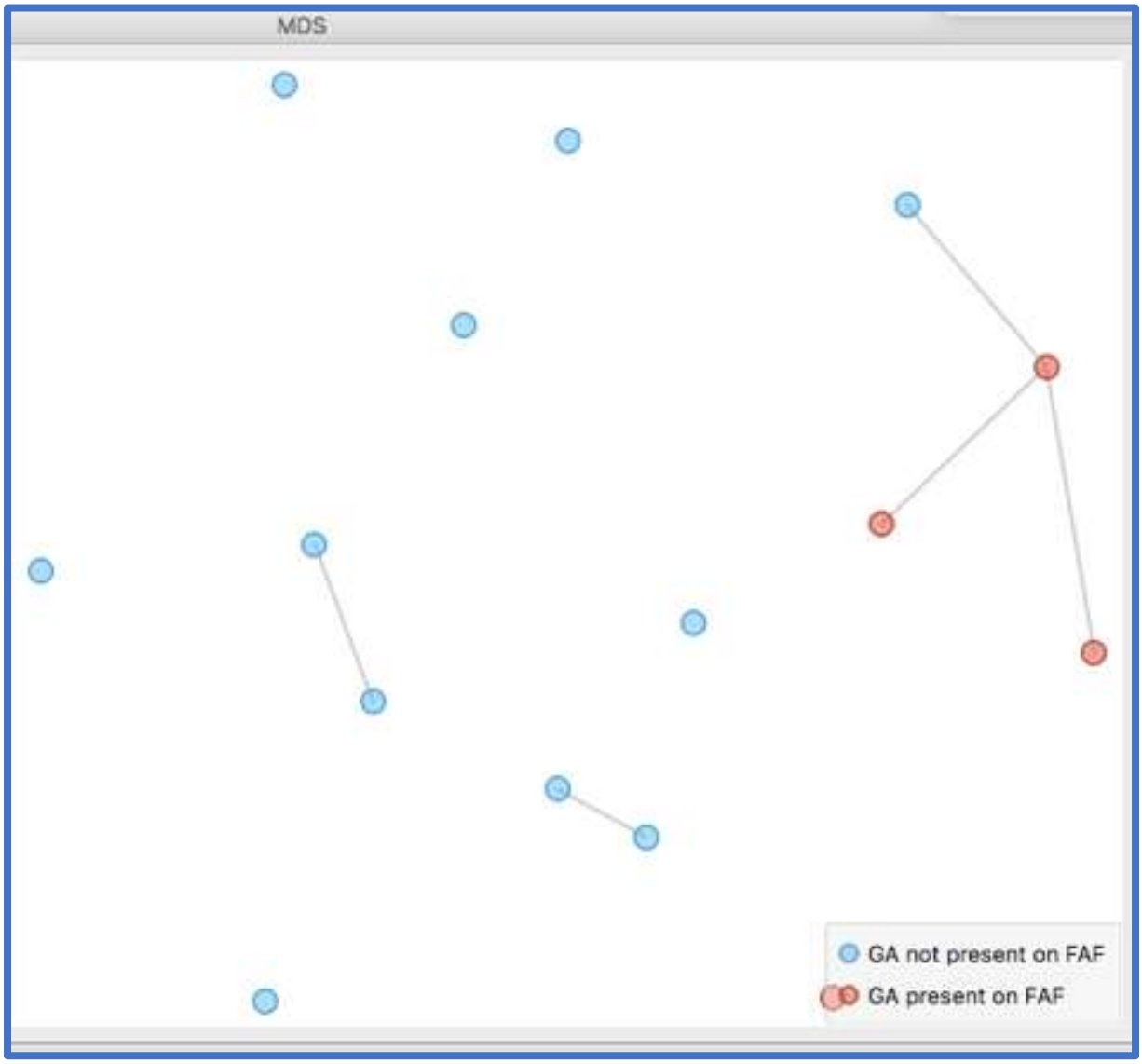

Figure 7.19 MDS for supervised machine learning for geographic atrophy in FAF images. Blue dots indicate no GA present and red dots GA present.

7.5.1.3 Pigmentary abnormalities

For analysis of pigmentary abnormalities as detected by CFP, the Preprocessing widget was again set to select only the 18 most informative variables based on Information Gain. However, again, the level of informedness is disappointing, with hyperparameter fine-tuned Logistic Regression having the best result of 9.7%. (0.785+0.312-1)*100, with an AUC of 0.534. Also, no clear pattern is seen from the spread of data points via the MDS widget shown in the Figure below.

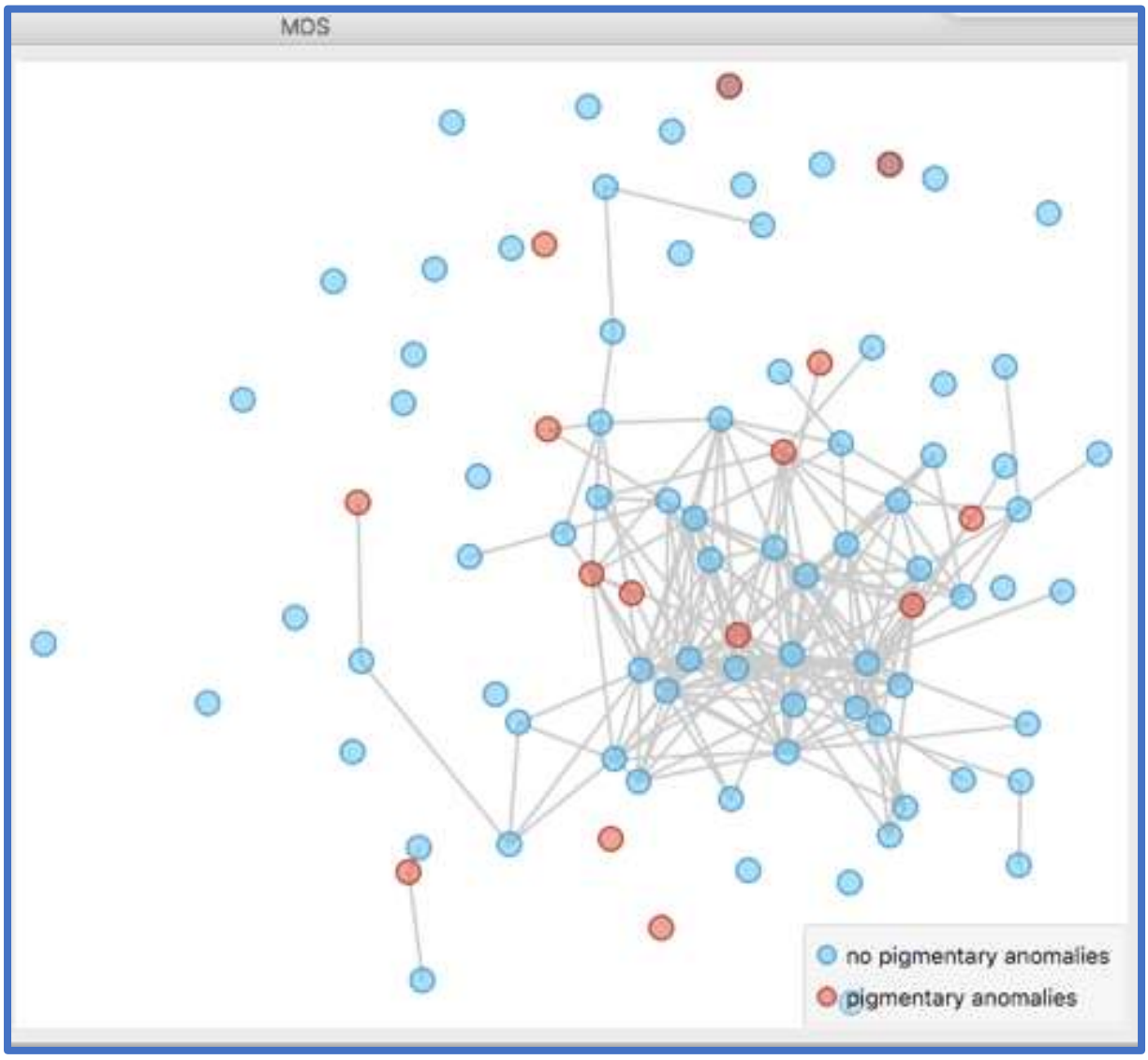

Figure 7.20 MDS for supervised machine learning for CFP images for pigmentary abnormalities. Blue dots indicate no pigmentary abnormalities present and red dots pigmentary abnormalities present.

7.5.1.4 Participant age

Finally, patient age (under and over the age of 80 years) and FAF results were analysed with supervised machine learning. The Preprocessing widget was again set to select only the 18 most informative variables based on Information Gain. From the Test and Score results, kNN was the best model learner, achieving an impressive level of 58.6% informedness. (0.828+0.758-1) with an AUC of 0.827 (considered good).(199) Also, from the MDS results in the Figure shown below, the algorithm does appear to having some success, with many more blue dots indicating participants under the age of 80 being situated at the lower part of the plot.

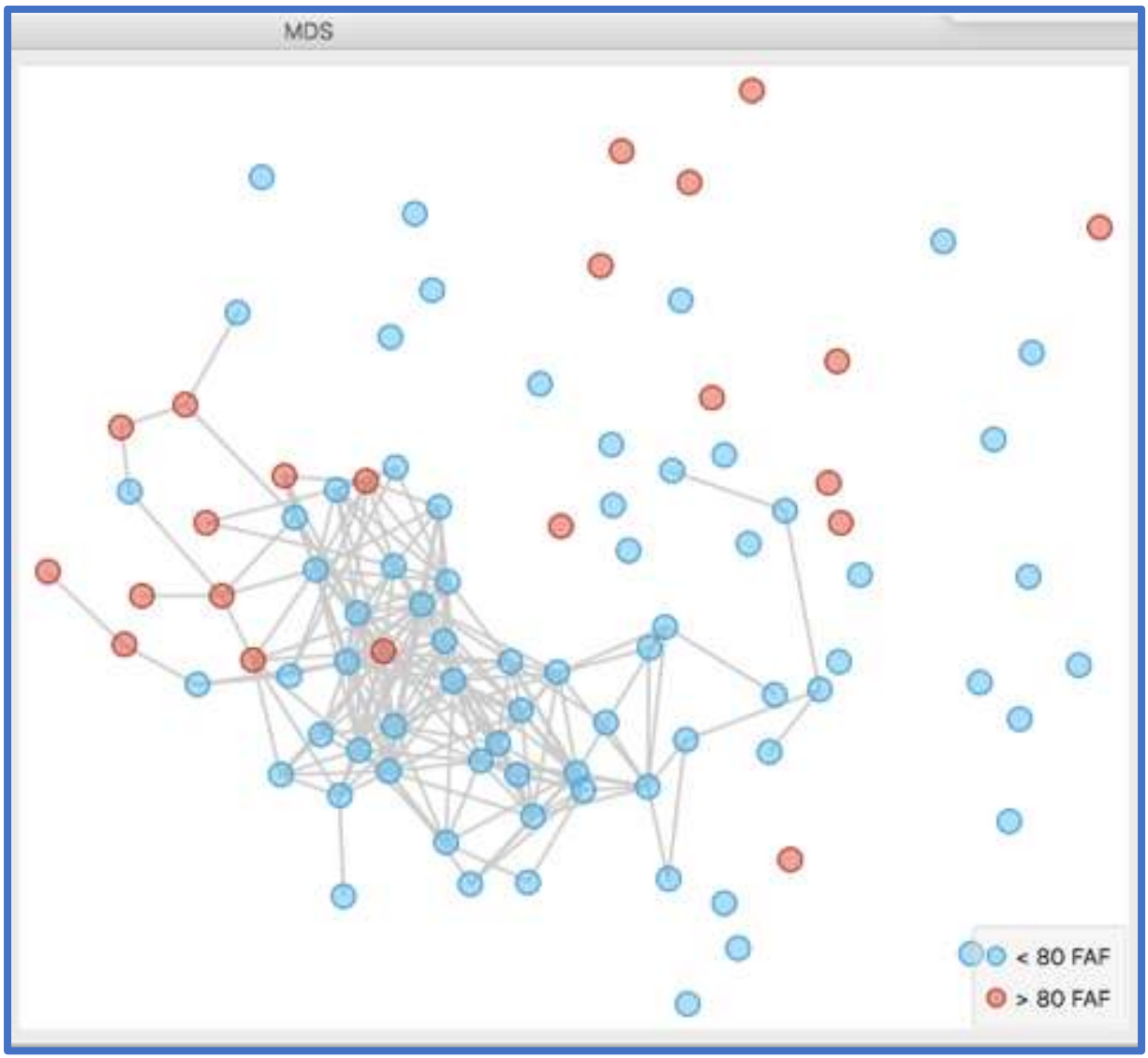

Figure 7.21 MDS for unsupervised machine learner for participant age for FAF images. Blue dots indicate < 80 years of age, and red dots indicate ≥ 80 years of age.

However, was the success of the machine learning for predicting patient age simply down to the image analysis detecting worsening cataract formation? To test this hypothesis, only pseudophakic patients (a subgroup of 26 in total) were included in the analysis. Again, the Preprocessing widget was adjusted to an instance to variable ratio of five to one, i.e. 26/5 = 5.2, so the five most informative variables based on Information Gain were selected. This sample includes 15 participants under, and 11 participants over the age of 80 years. Test and Score levels of informedness achieved are disappointing compared to those obtained from all the participants with both hyperparameter fine-tuned Logistic Regression giving 35.2% (0.654+0.698-1) with an AUC of 0.642 (considered poor),(199) indicating that cataract formation could be contributing to how the software is predicting age, and not entirely by analysis of the retinal changes detected by FAF. The MDS widget output concurs with a more random spread for pseudophakic participants compared to the entire cohort with regards to the spread of points for participants' age groupings.

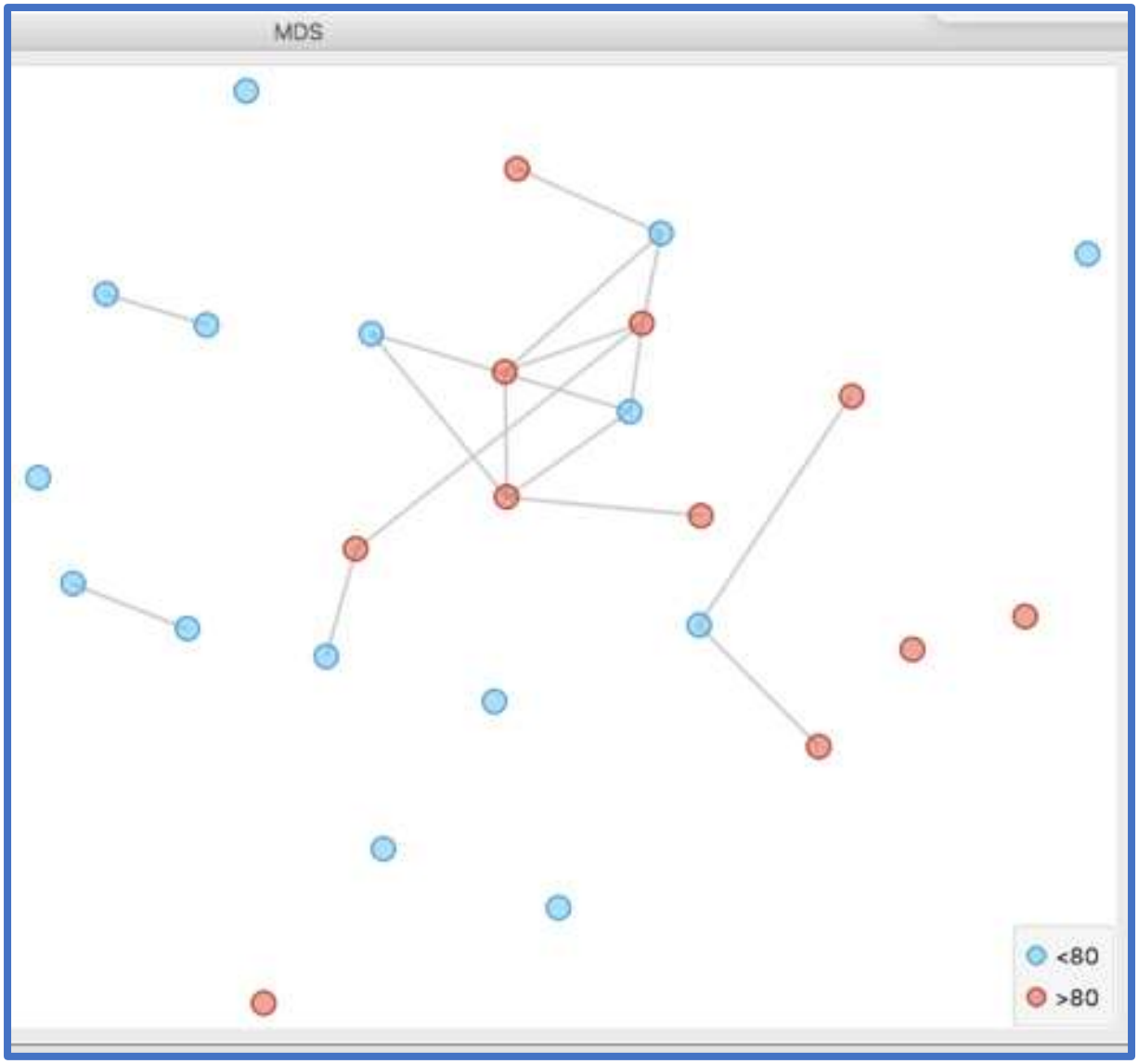

Figure 7.22 MDS for supervised machine learner for pseudophakic participants < and ≥ 80 years of age. Blue dots indicate < 80 years of age, red dots indicate > 80 years of age.

Clinical feature	Best performing model learner	AUC	Informedness
Large drusen	Logistic regression	0.581	16.4%
Geographic atrophy	Random Forest	0.955	90.9%
Pigmentary abnormalities	Logistic regression	0.534	9.7%
Age	kNN	0.827	58.6%
Age of pseudophakic participants	Logistic regression	0.642	35.2%

Table 7.6 Summary for clinical feature results for the best performing machine learners for the clinical features identified as contributing strongly to the principle components in Chapter nine of this paper.

7.6 Discussion

Utilising unsupervised machine learning, patterns within the data were explored without a specific target variable being selected. No clear patterns from the exploration of images captured via CFP and OCT emerged, with the images appearing to be randomly clustered within the vector space without clear separation of images with a normal and abnormal FAF result. The properties of two images, for both imaging modalities, closely situated within the MDS widget spaces were investigated, however, their clinical features appeared diverse with few common properties. For the FAF image exploration there did appear to be some evidence of clustering of normal and abnormal FAF images, however, with closer inspection the separation of the two groups was weak at best.

Utilising supervised machine learning, 18 most informative variables were selected based on Information Gain, to avoid overfitting, from the decision to maintain an instance to variable ratio of five to one. For CFP the best performing learner was SVM, but this model learner still only achieved an informedness of 14.0% (AUC 0.484). For OCT the kNN model learner managed an improved result of 17.5% informedness (AUC 0.613), and with the use of FAF images this rose again with hyperparameter fine-tuned Logistic Regression to 20.4% (AUC 0.551). Combinations of modalities were also tried, firstly CFP and OCT. For this analysis Naïve Bayes performed best, resulting in 19.4% informedness (AUC 0.550), however, when all three modalities, CFP, OCT and FAF were included, the best informedness level (via kNN) fell again to 13.8% (AUC 0.551). Horizontal “flipping” of the images, used as a simple initial image manipulation step to present all images as a right eye had very little effect on the results, with the exception of a dramatic improvement of the Tree model learner results. However, this is most likely due to the tendency of Tree to overfit, with Random Forest considered to be a more reliable model due to its use of multiple trees, which only achieved an informedness of 7.1% and an AUC of 0.577.

An alternative approach was tried using the specific clinical features which appear to be contribute strongly to the principal components as highlighted via the principle component analysis in Chapter nine. The ability of the model learners to find clustering of image properties for each feature was then explored. For large drusen as detected by OCT, Logistic Regression achieved an informedness of 16.4% (AUC 0.573), for geographic atrophy as detected by FAF, Random Forest managed an impressive 90.9% (AUC 0.955)

(however with only four data instances this result should be viewed with caution), for pigmentary abnormalities detected by CFP, hyperparameter fine-tuned logistic regression resulted in 9.7% informedness (AUC 0.534), and for participant age kNN recorded 58.5% (AUC 0.827). However, there is a possibility that the image degrading effect associated with the progression of cataract (which is particularly pronounced in FAF images) due to advancing age is assisting the model. To this end the participants who were pseudophakic (N = 26) were taken as a subgroup and the analysis repeated. In this new analysis, the best informedness result achieved by the hyperparameter fine-tuned Logistic Regression model learner fell to a lower figure of 35.2% (AUC 0.642), indicating that cataract formation may indeed by a confounding factor in the initial analysis.

Therefore collectively, the results from the unsupervised and supervised machine learning image analysis strand of this study indicate that all three modalities; CFP, OCT and FAF solely or collectively fail to build a reliable model to predict images that would be graded by a human observer as demonstrating a normal or abnormal FAF result. This appears to disagree with previous studies involving transfer learning in ophthalmology which boast impressive results for accuracy. For example, a study based in Chicago, USA, found that transfer learning utilising a convoluted neural network (CNN) called VGG16 and OCTA images could be applied to the identification of eyes with no diabetic retinopathy/with diabetic retinopathy with a sensitivity of 83.76% and a specificity of 90.82%.(200) However, it should be noted that all of the images were examined for shadow/motion artefacts, and any images considered to be qualitatively unacceptable were excluded from the study. Also, in a 2022 study on myopic eye disease from China which generated results at least equal to human observers, methods included cropping out the area of interest and normalising the image for the same size resolution and background colour.(201) An earlier study, this time based in India and utilising transfer learning and the CNN GoogleNet to study OCT images, managed an accuracy of 96% with the best performing model. Again, the images were heavily pre-processed which included smoothing the RPE, retinal flattening, the image repositioned within the frame (so that the RPE lower contour was in a fixed position), the image resized and finally filtered through BM3D three times (Block Matching and 3D filtering; this acts essentially as an image noise reducing tool).(202) Furthermore, a recent systematic review from China examined the evidence regarding artificial intelligence for the detection of ARMD from CFP, and looked at 19 studies (totalling 1.2 million images) 12 of which employed CNNs.(205) The results are encouraging, stating a collective sensitivity of 88.0% and a specificity of 90%. However, the studies included in the analysis varied in their definition of ARMD, and so it is difficult to draw conclusions as to how these results could be

compared to the analysis of CFP images for the prediction of FAF abnormalities conducted in the current study. In other words, advanced ARMD featuring geographic atrophy could flatter the performance of CNNs included within the review. Another recent study from France based on deep learning for the classification of retinal atrophy using FAF also reports encouraging results, however, the study was comparing images from patients with advanced ARMD (with geographic atrophy), and images demonstrating late stage retinal atrophic diseases (Stargardt and Pseudo-Stargardt Pattern Dystrophy).(206) The authors, who also used the ImageNet database for pretraining, reported an accuracy of 92.0% with an AUC of 0.981. However, as well as pre-processing the data, the training set also included new, randomly augmented data, generated to improve the robustness of the artificial intelligence models created. This consisted of the following steps; horizontal flip, a fill mode, width shift, height shift, rotation and zoom using Keras/ImageDataGenerator API.
Therefore, it can be concluded that the results from the current study offer lower predictive values compared to the results from other ophthalmological studies within the field of artificial intelligence based on transfer learning, with poor levels informedness achieved. However, in the current study, no selection by virtue of image quality or pre-processing of images was conducted, but rather every image that was considered to be suitable at the image capture stage was included. Also, no new data was created using image augmentation software to enhance the performance of the models. Whilst utilising advanced image pre-processing and image augmentation software was beyond the scope of the current study, a simple pre-processing step involving horizontal image flip, similar to that used in the image augmentation programs was simple to perform, and was therefore attempted. The results suggested that presenting only raw data involving right eyes to the machine learners makes little difference to the final informedness levels achieved.

7.7 Conclusions

In summary, image analysis results from this study overall offered low levels of informedness and AUC, with only the strand involving the identification of geographic atrophy via FAF imaging delivering an informedness level of over 60% (achieving 90.9%; AUC 0.955), however, the number of data instances for this calculation was low, (only four participants had geographic atrophy detected via FAF) and so the data uplift calculation (suggested by Chicco(193)) was required to correct the data imbalance problem. Therefore, the results of this strand could be considered to be less robust than other calculations involving a greater number of data instances. However, geographic atrophy usually

generates deep retinal defects differing in colour from the background retina, and domain expertise would lead the researcher to suspect that machine learners should perform well when identifying this condition in comparison to more subtle retinal lesions that could easily be masked by artefacts. With the exciting news, at the time of writing, of intravitreal Pegcetacoplan (Apellis' Syfovre®) being approved by the United States Food and Drug Administration for the treatment of retinal geographic atrophy (dry ARMD),(207) AI could therefore play a future role in identifying patients with this condition from screenings, enabling closer inspection on the selected cohorts to identify those patients with perilesional hyperautofluorecence and therefore potentially requiring referral to secondary care to be offered vision preserving medical treatment.

In conclusion, image pre-processing and the use of novel randomly augmented additional data appear to be a vital step when analysing ophthalmological images by machine learning processes when aiming for maximal machine learning performance. This indicates that any artificial intelligence utilised in primary care optometry would require instruments to feature built in software to enable image enhancement including, for example, cropping the area of interest and image normalisation for the same size, resolution and background colour, as well as potentially using image augmentation software to create new training images to improve the performance of machine learners, especially for the detection of relatively rare conditions. However, there may be a role for artificial intelligence transfer learning utilising raw data from primary care optometry, for conditions causing frank retinal changes such as the geographic atrophy. These may enable screening algorithms to be developed to identify patients who could benefit from more in-depth optometric/ophthalmological examinations and treatment.

Chapter 8 Clinical features and their relation to the specific FAF patterns

8.1 Introduction

A classification system enabling the identification of specific FAF phenotypes in ARMD is important as it enables clinicians to potentially highlight high-risk features and facilitates research involving the monitoring and treatment of macular degeneration. Such systems may also assist primary care clinicians in decision making processes regarding recall timings, patient advice and referrals to secondary care. In the previous chapters in this study, FAF classifications have been dichotomous to assist k-cross fold validation via ODM software, however, the aim of this chapter is to once again analyse the individual FAF classification categories to look for associations between specific patterns and the clinical features observed.

8.2 Methods

8.2.1 Classification systems for the grading of FAF images

Exploration of the literature has highlighted three proposed grading systems, the first aimed to grade/label FAF images from eyes with early ARMD, excluding more advanced features such as choroidal neovascularisation, geographic atrophy and pigment epithelial detachments. The second system involved the grading of the areas immediately surrounding patches of geographic atrophy in an attempt to risk stratify the likelihood of expansion of the atrophic areas. The third system was proposed from research into peripheral FAF abnormalities detected via ultra-widefield FAF in patients the majority of whom had ARMD.

8.2.2 The first FAF classification system considered

In 2005, a classification system was proposed by Bindewald et al to place FAF results from patients with early ARMD into one of eight distinct categories. This will from now on be referred to as the "Bindewald study". These were; normal, minimal change, focal increased, patchy, linear, lacelike, reticular and speckled.(165) To recap, a precise description of each phenotypic pattern in given below:

Normal: A homogenous background FAF with a gradual decrease in the inner macula toward the foveola due to the masking effect of the yellow macular pigment.

Minimal change: Very limited irregular increase and decrease of background FAF without a clear pattern.

Focal increased: Defined as having at least one area (< 200μm in diameter) of significantly increased FAF which is much brighter than the surrounding background's FAF signal. The borders are well defined, and the difference in FAF between the brighter area and its surrounding is not gradual. The brighter area may or may not demonstrate a darker halo surround.

Patchy: Characterised by presence of one large area (> 200μm in diameter) of markedly increased FAF. The borders tend to be less well defined than the Focal increased pattern.

Linear: This pattern features at least one linear area of markedly increased FAF. The borders of these areas are usually well defined and the difference in FAF between the brighter area and its surrounding is not gradual. These linear patterns usually map to hyperpigmented lines on CFP.

Lacelike: Multiple branching linear structures of increased FAF in a lace-like pattern which may map to hyperpigmentation on CFP. The borders may be difficult to define.

Reticular: Multiple small areas (< 200μm) of decreased FAF whose borders can be indistinct. This pattern tends to occur in the macular region, but also superotemporally in the retina. The pattern may or may not map to numerous small soft or hard drusen, or pigmentary abnormalities detected on CFP.

Speckled: A mixture of hypo and hyper-autofluorescent FAF abnormalities covering a large area which may extend beyond the macular area to cover the entire posterior fundus. These small areas may be punctate or linear. They may map to hypo and hyperpigmentation and/or multiple sub-confluent and confluent drusen.

The intraobserver variability for the two graders in the study that proposed this first classification system was 0.80 (95% confidence interval (CI) of 0.71-0.89), and 0.74 (95% CI 0.64-0.84), and for interobserver variability was 0.77 (95% CI 0.67-0.87), which were taken to indicate a relatively high-level agreement between and within the observers' results.(165)

8.2.3 The second FAF classification system considered

Later in 2005, another FAF grading system was proposed by Bindewald et al to classify phenotypes that may help to identify patients in whom there is likely to be an increased risk of the spread of retinal atrophy and vision loss. This will be referred to as the “Bindewald GA study”. The authors graded 149 eyes and identified eight different and distinct patterns surrounding areas of GA: normal, focal, banded, patchy and diffuse, with diffuse being further sub-divided into reticular, branching, fine granular and fine granular with peripheral punctate spots.(208) Further research identified that progression rates were highest in the banded and diffuse phenotypes(45) and in particular in a new, ninth pattern named as diffuse trickling. The diffuse trickling FAF phenotype has greyish areas of GA rather than a markedly decreased, darker FAF signal that tends to be present in other GA subtypes.(45)

8.2.4 The third FAF classification system considered

In 2013, Optos® ultra-wide FAF fundus imaging was used by Tan et al to classify abnormal peripheral (defined as outside the central 30 degrees centred on the fovea) FAF patterns with regard to location, extent and type. This will be referred to as the “Tan study”. 164 eyes were examined and three distinct phenotypes were identified: granular, mottled and nummular. The granular pattern was found to be highly correlated with peripheral drusen and was defined as small discrete areas of hyper-autofluorescence. Mottled was closely associated with RPE depigmentation and featured a decreased FAF signal distributed in an irregular pattern. Finally, nummular was defined as small to medium areas of discrete and uniform hypo-autofluorescence.(114)

8.2.5 Decision to adopt the “Bindewald study” method of classification

The first “Bindewald study” classification system discussed above was utilised within the current study due to the majority of the ARMD encountered in the cohort of 93 being of the early type. The second classification system concerned eyes with advanced changes involving GA and was therefore not considered to be as useful for comparisons and grading in the current study. The third grading system was also unsuitable for the current study due to its being based on peripheral FAF imaging results.

Prevalence of ARMD clinical features for the study participants

Clinical feature	Presence or grade of clinical feature	Percentage of participants with this clinical feature
Pigmentary abnormalities (detected by CFP)	Yes No	15% 85%
RPD (detected by OCT)	Yes No	11% 89%
Simplified severity score (detected by CFP)	Grade 0 Grade 1 Grade 2 Grade 3 Grade 4	48% 11% 32% 2% 7%
Large drusen (detected by OCT)	Yes No	58% 42%
Geographic atrophy (detected by FAF)	Yes No	4% 96%
Pigment epithelial detachment (detected by OCT)	Yes No	6% 94%
Symptoms of ARMD i.e. a distortion of the central vision, a “kink” in the central vision or central scotoma	Yes No	5% 95%

Table 8.1 Prevalence of clinical features for the study participants

8.3 Results

The similarities and differences from the research which led to the development of the “Bindewald study” classification system and the current study are discussed below. These comparisons were considered a useful exercise to enable a greater understanding of the prevalence of the various FAF pattern phenotypes, as well as an exploration of how clinical features seen with other modalities, such as CFP and SD-OCT, are related to specific FAF patterns.

8.3.1 Similarities to the current study

1. The number of participants involved was similar (100 in the “Bindewald study” versus 93 in the current study).
2. All participants were > 50 years of age in both studies.
3. Patients demonstrating pigmentary abnormalities and/or drusen were both included in the “Bindewald study” and the current study.

8.3.2 Differences from the current study

1. A cSLO was used in the “Bindewald study” for FAF imaging rather than a modified fundus camera-based system as in the current study.
2. All patients recruited for the “Bindewald study” had early ARMD (defined has having either pigmentary abnormalities, drusen, or both detected via CFP), whereas the current study recruited all eligible patients, regardless of ARMD status.
3. The age inclusion criterion was older in the “Bindewald study”, > 55 years of age, rather than > 50 years of age as in the current study.
4. Patients with choroidal neovascular membranes (CNV), geographic atrophy (GA) and pigment epithelial detachments (PEDs) were excluded from the “Bindewald study” unlike the current study. This is due to these features being considered as “advanced” changes.
5. The macula was defined as a 6000μm (approximately 20°) circle centred on the foveola in the “Bindewald study”; in the current study the macula was taken as a larger area bounded by the major temporal vascular arcades, in a circle whose diameter is defined as the length of the vertical line between the temporal vascular arcades. In a previous study this line has been measured, in the healthy control eyes of patients with a idiopathic macular hole in the fellow eye, as 9473±1974μm(166). This equates to 33°±7°.(167)

8.3.3 Prevalence of patterns

In the “Bindewald study” and the current studies the percentages of the patterns identified were as follows:

Numerical code of pattern in current study	Pattern name	Bindewald study	Current study	Participants with early ARMD in the current study (subgroup of 48) as defined by Bird et al(11)
1	Normal	19%	24%	6%
2	Minimal change	25%	47%	56%
3	Focal increased	23%	6.5%	10.5%
4	Patchy	15%	14%	13%
5	Linear	9%	0%	0%
6	Lacelike	3%	2%	4%
7	Reticular	2%	5.5%	10.5%
8	Speckled	26%	1%	0%

Table 8.2 Showing the prevalence of FAF patterns in the “Bindewald study” and the current studies compared

Comparing the percentages, in the current study the most common FAF pattern is “minimal change”, making up almost half of all graded images. The figure for “minimal change” was much lower in the “Bindewald study” involving patients with early ARMD at 25%. 24% had a “normal” FAF pattern in the current study compared to 19% in the “Bindewald study”. “Focal increase” made up almost a quarter of all patients in the “Bindewald study”, representing only 6.5% in the current study. No “linear patterns” were detected in any patients in the current study, making up only 9% in the “Bindewald study”. “Patchy”, “lacelike” and “reticular patterns” showed fair agreement between the two studies, however, “speckled” was much more prevalent in the “Bindewald study”. To summarise, there is fair agreement (within at least 9%) between the studies on “normal”, “patchy”, “linear”, “lacelike” and “reticular patterns”, with many fewer “minimal changes” and many more “focal increased” and “speckled” in the “Bindewald” study compared to the current study.

However, as noted previously, the “Bindewald study” excluded those patients with CNV, GA and PEDs. The tables below show how the percentages of the various FAF patterns change when patients with CNV, GA and PEDs are stripped out of the current study.

Table 8.2 below shows the types of FAF patterns recorded from the excluded patients, whilst table 8.3 shows how the FAF patterns percentages compare between the two studies after the stripping out process is performed. It appears that having GA and PEDs may lead to mainly "patchy" patterns, with 6 of the 9 patients who had these clinical features demonstrating this pattern. There were only two cases with CNV in the current study, so there is not sufficient data to draw conclusions regarding neovascular (wet) ARMD and the type of FAF pattern likely to be attributed to this condition.

Case number excluded from current study	**GA**	**CNV**	**PED**	**pattern**
3	y	n	y	speckled
24	y	n	n	patchy
28	n	n	y	minimal change
39	y	n	y	patchy
40	y	n	y	patchy
59	n	n	y	patchy
60	n	y	n	minimal change
69	y	n	n	minimal change
86	n	y	y	patchy
92	y	n	n	patchy

Table 8.3 Cases excluded from the "Bindewald study" with clinical feature and FAF pattern detected

Numerical code of pattern in current study	**Pattern name**	**"Bindewald study"**	**Current study with exclusions**	**Current study without exclusions**
1	Normal	19%	26.5%	24%
2	Minimal change	25%	49.0%	47%
3	Focal increased	23%	7.0%	6.5%
4	Patchy	15%	8.0%	14%
5	Linear	9%	0.0%	0.0%
6	Lacelike	3%	2.5%	2%
7	Reticular	2%	6.0%	5.5%
8	Speckled	26%	0.0%	1%

Table 8.4 Percentages of FAF patterns from the "Bindewald study" and the current study with and without exclusions

8.3.4 Prevalence of patterns: Summary

Therefore, it appears that having early ARMD (defined as having pigmentary abnormalities and/or drusen detected via CFP), moves the FAF pattern from “minimal change” into the categories of “focal increased” and “speckled”, as these were the Inclusion criteria for the “Bindewald study”, but were not a required feature to be part of the current study, with these two patterns being more prevalent in the “Bindewald study”.

However, in the current study, many of the participants (as defined by having drusen or pigmentary abnormalities without more advanced changes) had early ARMD (61.3%) and only 1% demonstrated a “speckled” pattern, so it may be concluded from this finding that the graders in the “Bindewald study” had a greater tendency to categorise a pattern as “speckled” compared to the grader of the current study. Also, the “patchy” and “speckled” patterns are arguably the two most similar patterns in the classification system and most likely to be confused with one another. The assumption has also been made that the different instrumentation employed by the two studies (cSLO versus fundus camera-based FAF) had little or no effect on the categorisation of the FAF patterns.

Studying the clinical features of participants from the current study that were stripped out due to having CNV, GA and PEDs indicates that having GA and/or PEDs may lead to a “patchy” pattern. The lack of CNV cases in the current study mean that conclusions cannot be drawn regarding the FAF patterns likely to be attributable to this condition. The results are summarised in Table 8.4 below.

Clinical feature	**Associated FAF pattern(s)**
Pigmentary abnormalities	Focal increased/Speckled
Drusen	Focal increased/Speckled
Geographic atrophy	Patchy
PED	Patchy

Table 8.5 Showing likely associations between clinical features and FAF patterns from examination of the results from the “Bindewald study” and current study results.

The last column in Table 8.2 also compares as subgroup (N=48) of participants within the current study who had “Early” ARMD as defined by Bird et al, the classification system adopted by the “Bindewald” study, i.e. participants identified as having either RPE

pigmentary abnormalities and/or large macular drusen. Note that only one of this subgroup had a PED detected by OCT. When comparing this group to the entire study cohort of 93, this subgroup has a lower percentage of “normal” patterns, down from 24% to 8%. “Minimal change” increases from 47% to 56%, “focal increased” increases from 6.5% to 10.5%, “patchy” remains similar at 13% down from 14%, with an increase of “lacelike” from 2% to 4%, “reticular” increasing from 5.5% to 10.5%. Finally, speckled falls from 1% to zero%. In summary, these findings indicate that having early ARMD increases the percentage of “minimal change”, “focal increased”, “lacelike” and “reticular”. These findings again back-up the findings from Table 8.5 that RPE pigmentary abnormalities and drusen are associated with “focal increase” FAF patterns, with an additional association of an increase in the percentages of “lacelike” and “reticular” patterns.

8.3.5 Using Orange Data Mining to explore the patterns

ODM was also used to investigate which clinical findings in the current study are related to which FAF patterns to test the above hypotheses. The Boxplot widget indicates that the following 18 clinical findings are the most influential in predicting the outcome of the FAF pattern grading, by using the "order by relevance of variable" function.

Clinical finding number	**Clinical finding ranked via ODM Boxplot widget for importance in determining FAF pattern**	**Rationale for further exploration, i.e. if considered relevant then analysed below via the ODM Distribution widget**
1	CFP disc diameters of pigmentary abnormalities	relevant
2	FAF RPD yes or no	reticular by definition
3	CFP – SSS (simplified severity scale)	relevant
4	CFP pig abnormalities yes or no	relevant, similar to 1
5	OCT large drusen 0-6	relevant
6	OCT RPD yes or no	relevant
7	OCT large drusen yes or no	relevant, similar to 5
8	CFP RPD yes or no	relevant, similar to 6
9	FAF GA yes or no	relevant
10	CFP large drusen yes or no	relevant, similar to 5
11	CFP GA yes or no	relevant, similar to 9
12	CFP large drusen 0-6	relevant, similar to 5
13	OCT PED yes or no	relevant
14	CFP large drusen other eye yes or no	N/A
15	OCT large drusen other eye yes or no	N/A
16	OCT intermediate drusen 0-6	relevant
17	Symptoms of ARMD yes or no	relevant
18	CFP disc diameters of GA	relevant

Table 8.6 Boxplot widget indicating the 18 most influential clinical findings for predicting the outcome of the FAF pattern grading and their relevance for further exploration via Orange Data Mining.

Colour coding	FAF pattern
Dark blue	Focal increase
Red	Lacelike
Green	Minimal change
Orange	Normal
Yellow	Patchy
Mauve	Reticular
Light blue	Speckled

Table 8.7 Colour coding for FAF patterns in Figures 8.1 to 8.9 below

A selection of the most relevant clinical findings investigated further via the ODM Distribution widget are shown in Figures 8.1 to 8.9 below:

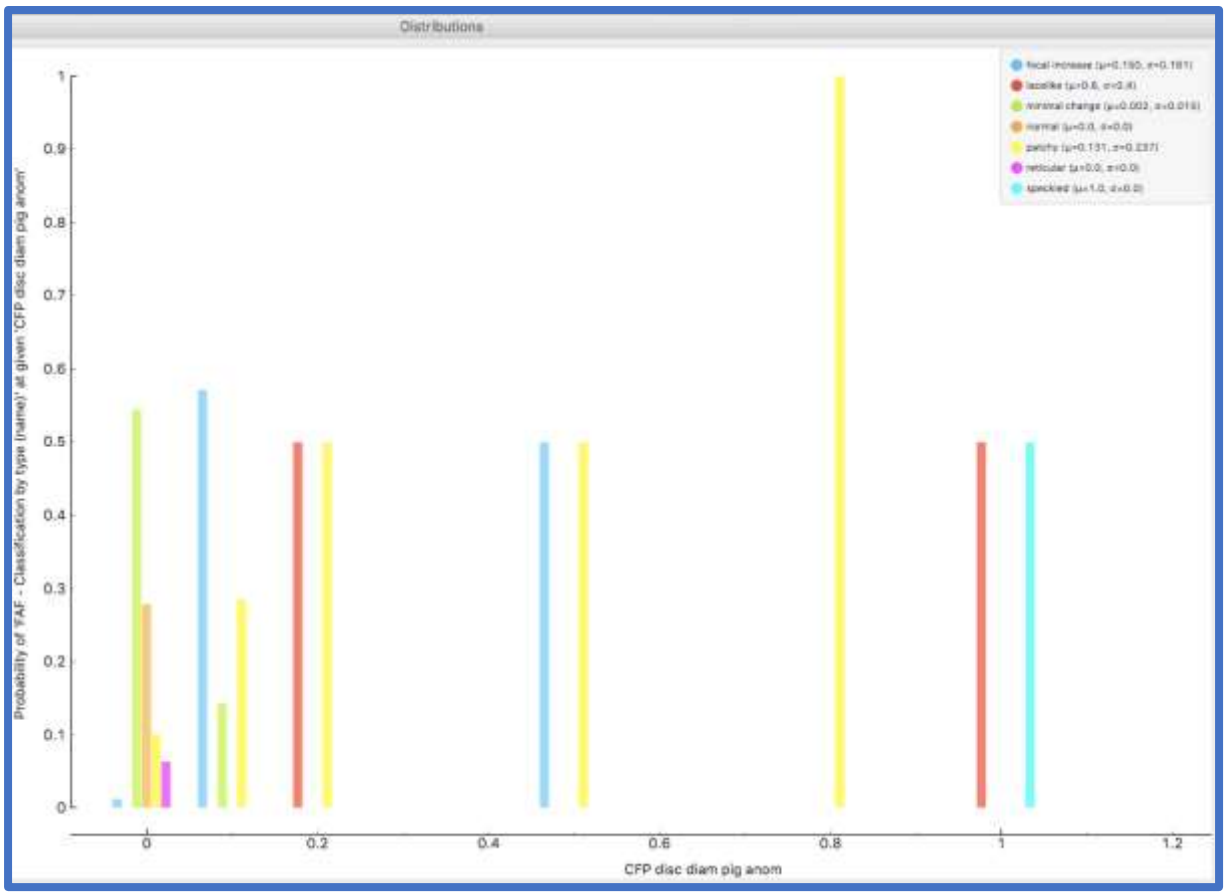

Figure 8.1 Distribution of CFP disc diameters of pigmentary abnormalities and probability of FAF patterns.

Clinical finding #1. ODM Distribution widget output indicating the general trend for an increase in the probability of a patchy FAF pattern (shown in the yellow bars) being detected with increasing disc diameters of pigmentary abnormalities as detected by CFP. 5 of the 13 participants with pigmentary abnormalities (38%) had a patchy FAF pattern.

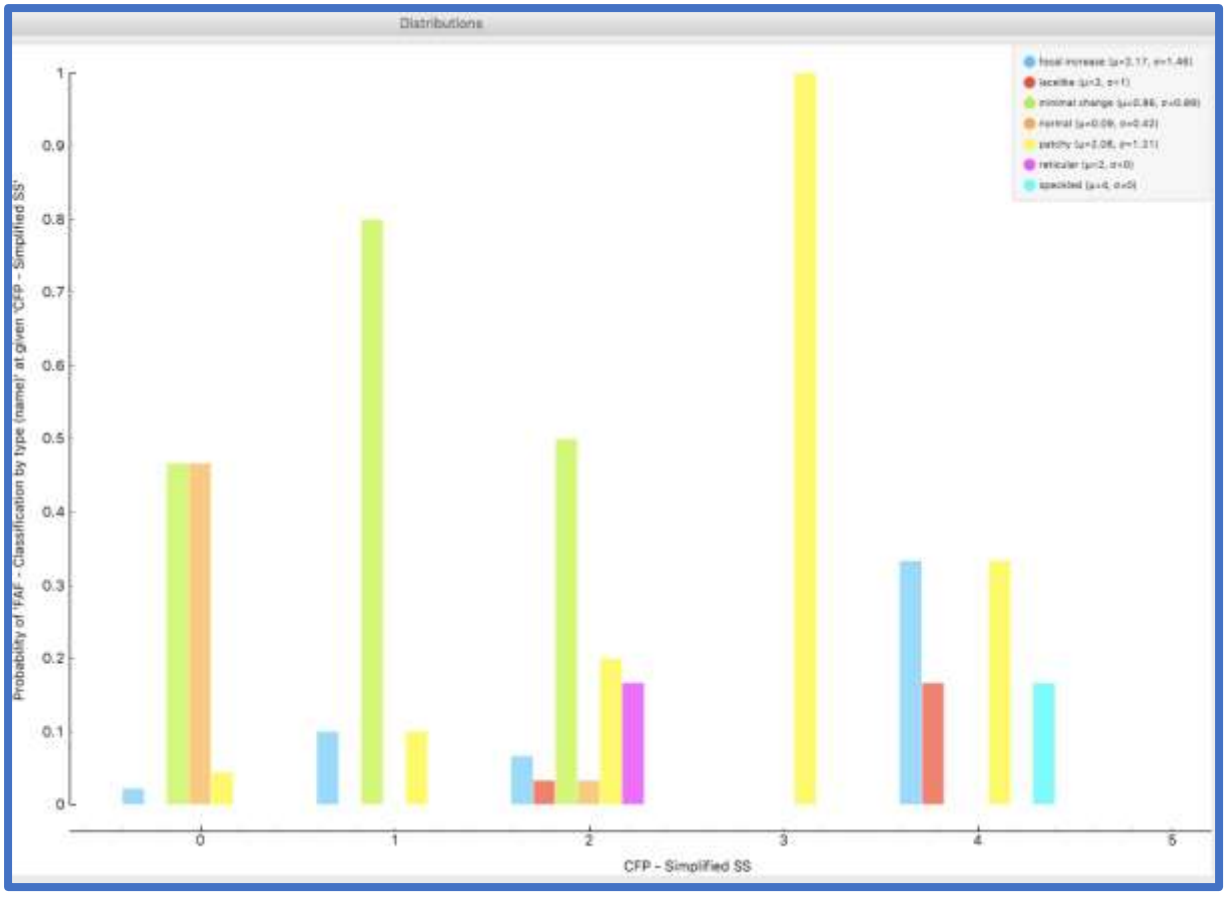

Figure 8.2 Distribution of CFP simplified severity score and FAF pattern

Clinical finding #3. ODM Distribution widget output indicating the general trend for an increase in the probability of a patchy (yellow bars) and/or focal increased (darker blue bars) FAF pattern with increasing simplified severity score. 11 of the 13 participants (85%) with a patchy FAF pattern had an SSS > 0, whilst five out of six participants (83%) with a focal increased pattern had a SSS > 0.

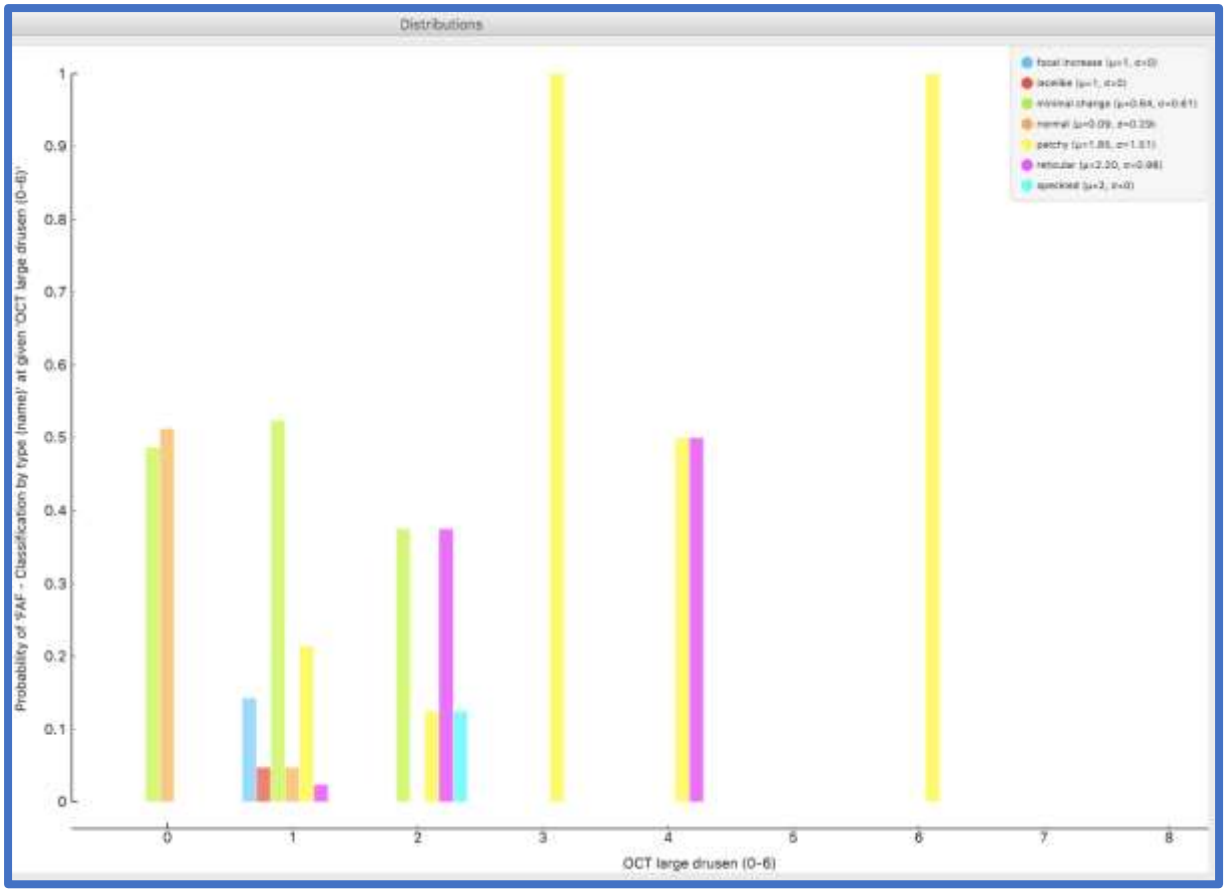

Figure 8.3 Distribution of OCT large drusen and probability of FAF pattern

Clinical finding #5. ODM Distribution widget output indicating the general trend for an increase in the probability of patchy (yellow bars) FAF pattern with increasing OCT Large Drusen score. All 13 participants (100%) with a OCT large drusen score of > 0 had a patchy FAF pattern.

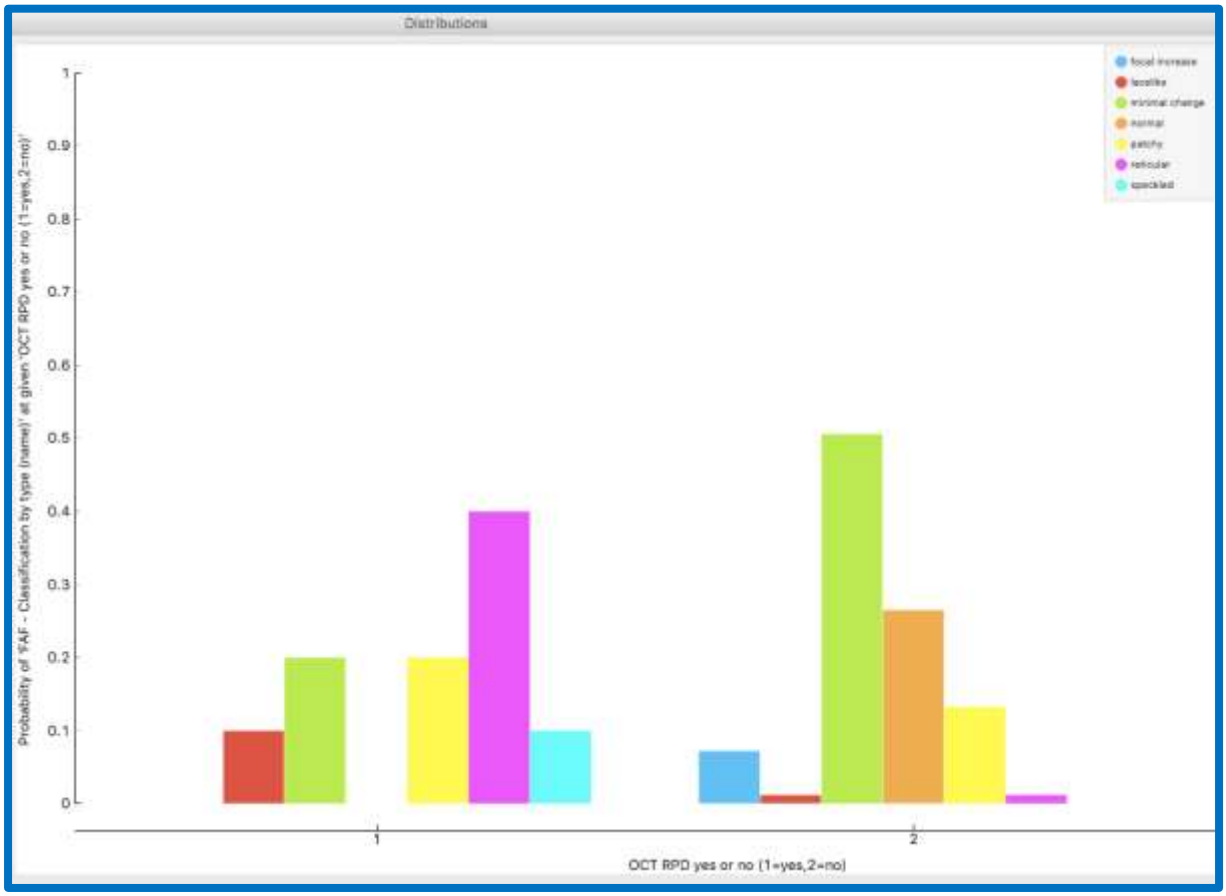

Figure 8.4 Distribution presence/absence of OCT RPD and probability of FAF patterns

Clinical finding #6.

ODM Distribution widget output indicating the general trend for an increase in the probability of a reticular FAF pattern (shown in the purple bars) in the presence of RPD as detected by OCT. 4 out of the 10 participants (40%) with RPD detected by OCT had a reticular" pattern.

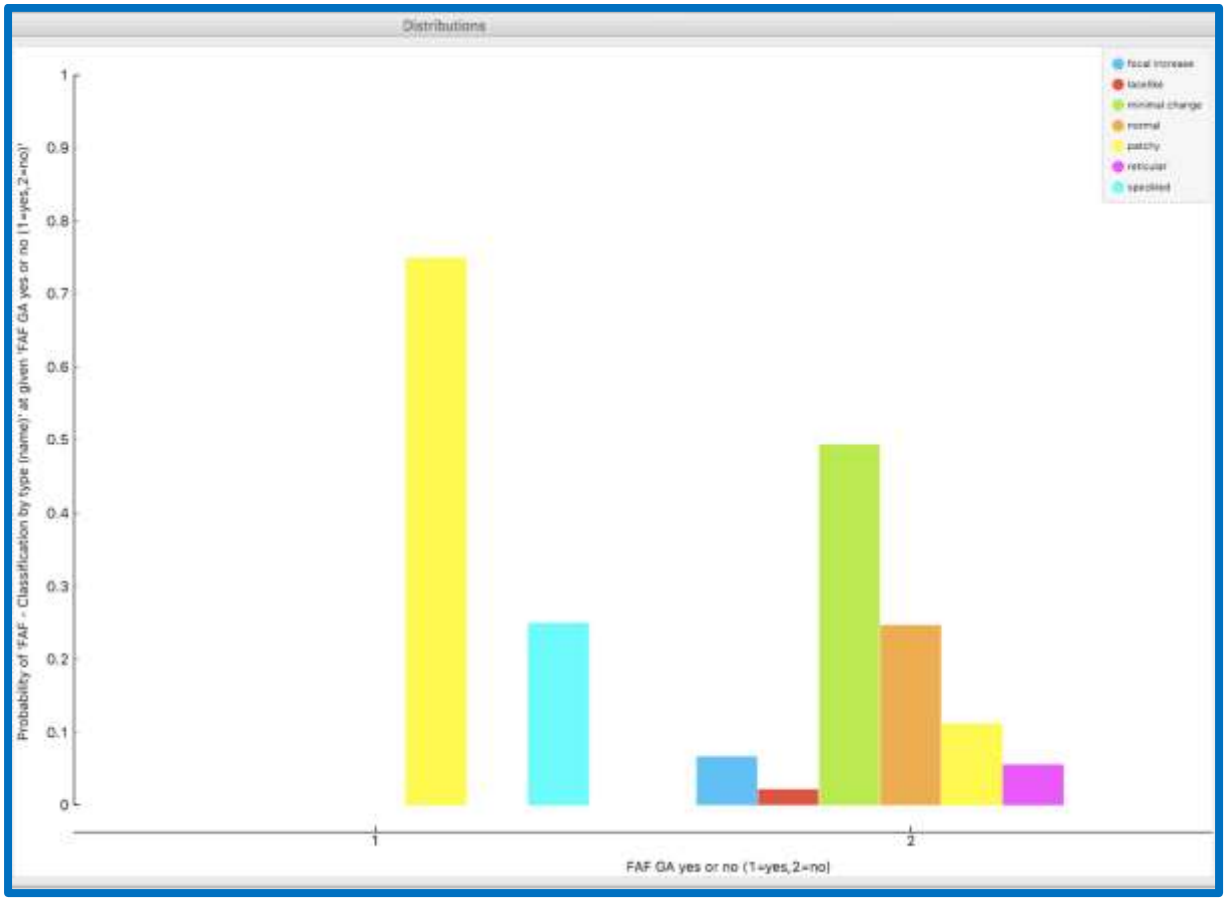

Figure 8.5 Distribution of FAF GA presence yes or no and FAF pattern.

Clinical finding #9. ODM Distribution widget output indicating the association between FAF GA presence yes or no and FAF pattern. The probability of a patchy (yellow bars) FAF pattern was approximately 77% when GA was detected by FAF and fell to approximately 12% when GA was not detected by FAF. 3 of the 4 participants with GA detected by FAF (75%) had a patchy FAF pattern.

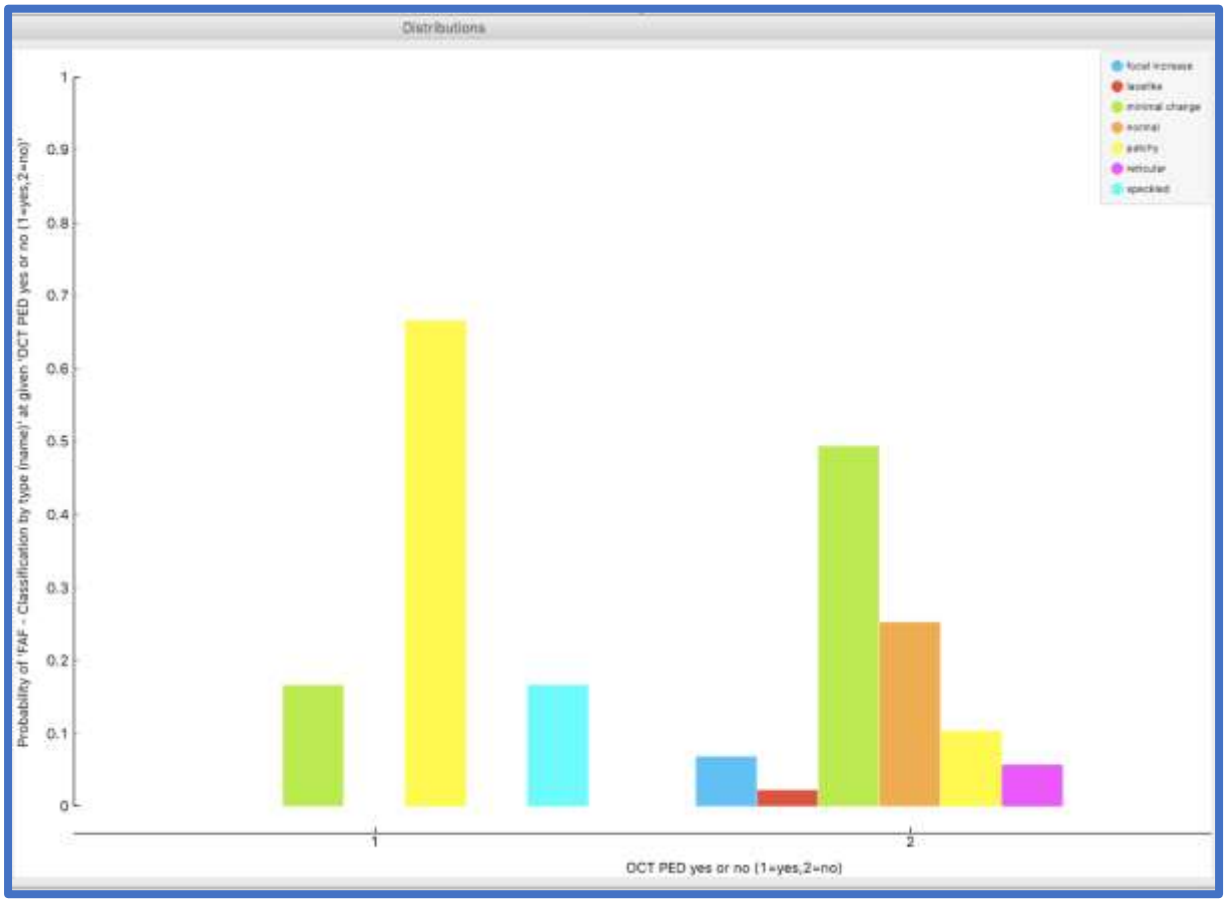

Figure 8.6 Distribution of OCT PED presence/absence and FAF pattern

Clinical finding #13. ODM Distribution widget output indicating the association between OCT PED detection and the FAF pattern. The probability of a patchy (yellow bars) FAF pattern was 66.7% when a PED was detected. This fell to just over 10% with a PED was not detected. 4 out of 6 participants with a PED detected by OCT (67%) had a patchy FAF pattern.

.

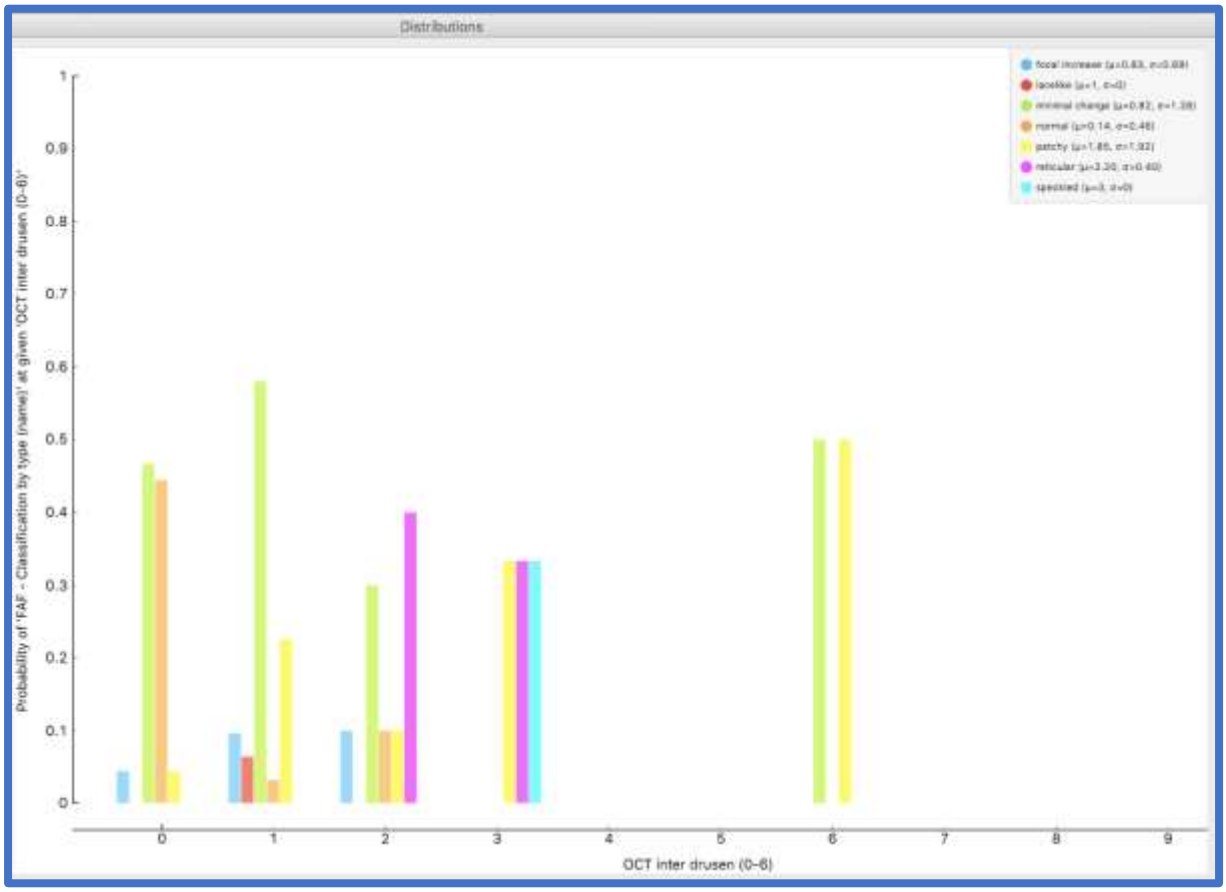

Figure 8.7 Distribution of OCT intermediate drusen and probability of FAF pattern

Clinical finding #16. ODM Distribution widget output indicating the general trend for an increase in the probability of patchy (yellow bars) FAF pattern with increasing OCT Intermediate Drusen score. 11 out if 13 participants with intermediate drusen detected by OCT (85%) had a patchy FAF pattern.

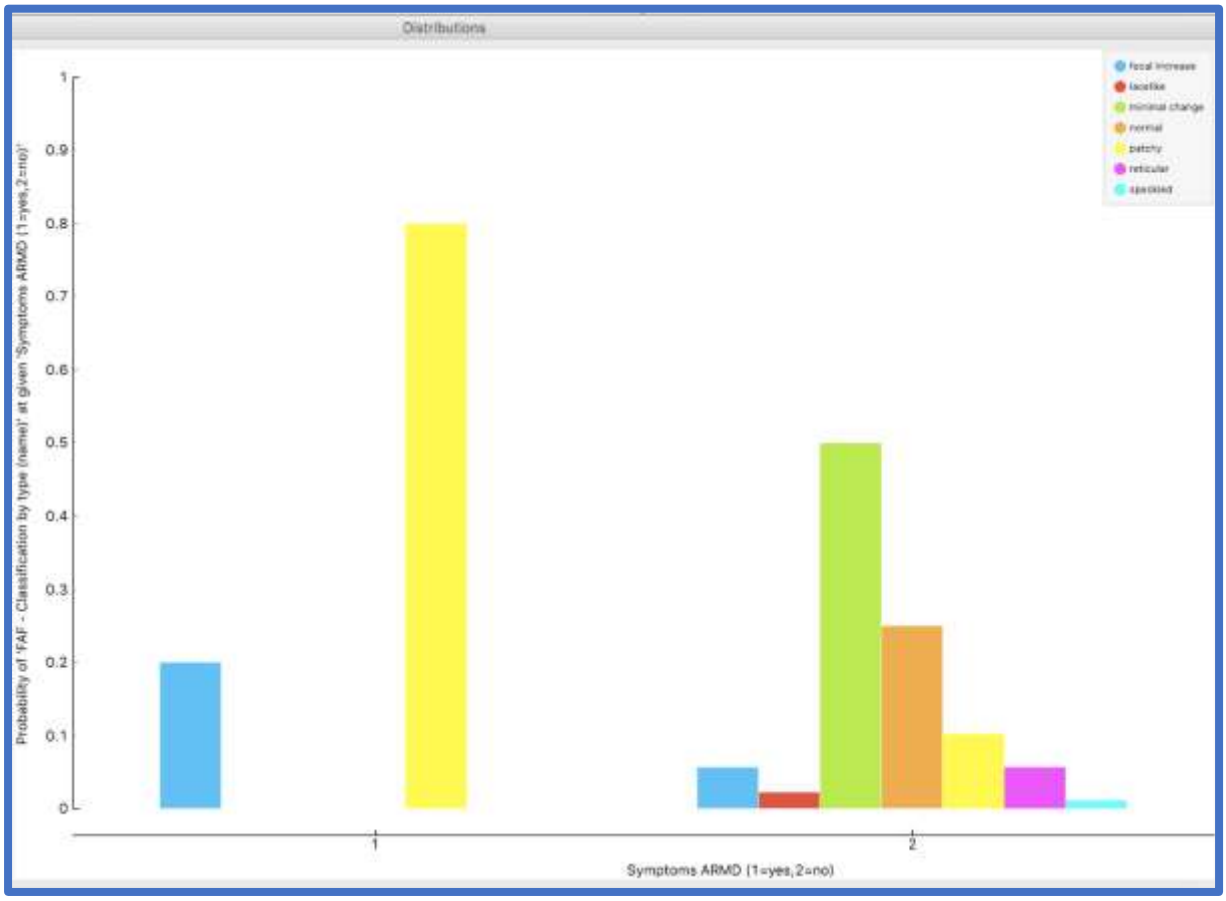

Figure 8.8 Distribution for Symptoms related to ARMD and the probability of FAF pattern

Clinical finding #17. ODM Distribution widget output indicating the association between symptoms of ARMD and the FAF pattern. Four out of the five participants with symptoms of ARMD (80%) had a patchy (yellow bars) pattern.

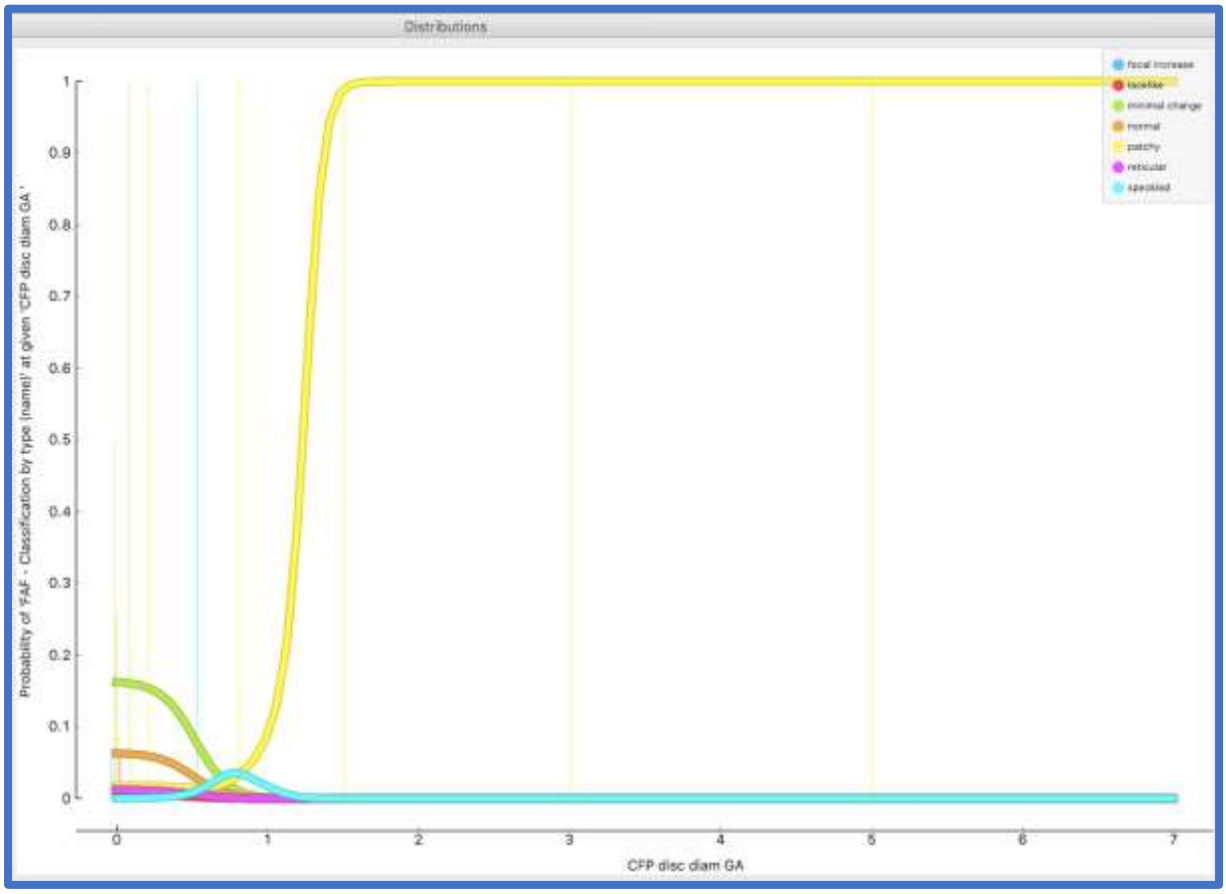

Figure 8.9 Distribution of CFP disc diameters of GA and probability of FAF pattern

Clinical finding #18. ODM Distribution widget output indicating the trend for an increase in the probability of patchy (yellow line) FAF pattern with increasing CFP disc diameters of GA. The probability of a patchy pattern appears to increase dramatically once the score reaches approximately 1.25 CFP disc diameters of GA and plateaus maximally at approximately 1.5 CFP disc diameters. 5 out of 7 participants (71%) with GA detected by CFP had a patchy FAF pattern.

8.3.6 Orange Data Mining exploration of the patterns: Summary

Clinical feature number	Clinical feature	Associated FAF patterns(s) within the current study	Predictions of associated FAF patterns(s) from examination of the "Bindewald study" and current study
1	CFP disc diameters of pigmentary abnormalities	Patchy	Focal increased/speckled
2	FAF RPD yes or no	Reticular (by definition)	N/A
3	CFP SSS (simplified severity score)	Patchy/Focal Increased	Focal increased/speckled
4	CFP pigmentary abnormalities yes or no	Patchy	Focal increase/speckled
5	OCT large drusen 0-6	Patchy	Focal increased/speckled
6	OCT RPD yes or no	Reticular	N/A
7	OCT large drusen yes or no	Patchy	Focal increased/speckled
8	CFP RPD yes or no	Reticular	N/A
9	FAF GA yes or no	Patchy	Patchy
10	CFP large drusen yes or no	Patchy	Focal increased/speckled
11	CFP GA yes or no	Patchy	Patchy
12	CFP large drusen 0-6	Patchy	Focal increased/speckled
13	OCT PED yes or no	Patchy	Patchy
14	CFP large drusen other eye	N/A	N/A
15	OCT large drusen other eye	N/A	N/A
17	OCT Intermediate drusen 0-6	Patchy	Focal increased/speckled
17	Symptoms of ARMD	Patchy	N/A
18	CFP disc diameters of GA	Patchy	Patchy

Table 8.8 Showing how clinical features are associated with FAF patterns based on an Orange Data Mining exploration of the results in the current study.

The columns in Table 8.8 above show agreement that GA and PEDs may be responsible for a patchy FAF pattern. There is also partial agreement on the simplified severity score, where both tables indicate that a focal increased FAF pattern may associated with a higher score. There appears to be disagreement on large/intermediate drusen and pigmentary abnormalities, with the current study column indicating a higher likelihood of a patchy pattern, with the prediction column a focal increased/speckled pattern. Reticular pseudodrusen are associated with a reticular FAF pattern by definition.

Summary of results

Clinical feature	**Associated FAF pattern from both analyses**	**Agreement/disagreement between current study and prediction columns**
Geographic atrophy	Patchy	agreement
Pigment epithelial detachment	Patchy	agreement
Simplified severity score	focal increased	partial agreement
Large/Intermediate drusen	Patchy/focal increased/speckled	disagreement
Pigmentary abnormalities	Patchy/focal increased/speckled	disagreement
Reticular pseudodrusen	Reticular	N/A

Table 8.9 Summary of clinical findings and associated FAF patterns from the current study.

8.4 Discussion

By studying the inclusion and exclusion criteria from a 2005 study(165) (referred to as the “Bindewald study” in this chapter) and the current study, it was possible to identify which clinical features may be responsible for certain FAF patterns, and the results of this analysis was compared to an alternative method using the ODM Distribution widget with data solely from the current study.

The results indicate that geographic atrophy and pigment epithelial detachments may be responsible for a patchy FAF pattern, with a higher simplified severity score (an amalgam of large or intermediate drusen and pigmentary abnormalities) being partially associated with a focal increased pattern. There appeared to be a discrepancy between the two methods with regards to pigmentary abnormalities, with one method indicating an association with a patchy pattern and the other a focal increased/speckled pattern. The same discrepancy occurred with large and intermediate sized drusen as detected by OCT. An association between RPD as detected via CFP and OCT and a reticular FAF pattern was also found, based on the ODM exploration of results from the current study.

8.5 Conclusions

In summary, the results of this chapter indicate that geographic atrophy and pigment epithelial detachments are associated with patchy FAF patterns. Furthermore, there is a weaker association between pigmentary abnormalities and large or intermediate drusen and a focal increased FAF pattern by virtue of this pattern's association with an increased simplified severity score, which is calculated as an amalgam of these three clinical features. A reticular FAF pattern is associated with the detection of reticular pseudodrusen by definition, and was also found to be associated with RPD as detected via CFP and OCT in the current study. The findings from this chapter indicate that there may be a role in clinical optometry and ophthalmology for the greater use of FAF imaging for the identification of clinical features.

Chapter 9 Principle Component Analysis to analyse CFP, OCT and FAF data collected within this study

9.1 Introduction

Principle component analysis (PCA) is a statistical process that enables the reporting of a large number of variables using fewer “components” (from now on referred to as “factors”) and may be described as a data reduction technique.(209),(210) Essentially, PCA allows the researcher to summarise the relationships between variables, by means of these factors and factor loadings. Alternatively, the factors may be thought of as being created as a result of the collected variables. PCA analysis was considered a useful statistical technique to employ as part of this research project, as the relationships between variables could assist in elucidating the physiological processes that may be occurring within the retina, that ultimately result in an abnormal FAF pattern.

9.2 Methods

For Chapter nine of this study, PCA was calculated by two alternative methods: by ODM software and by IBM's SPSS Statistics (Version 6) program. Each technique has its own strengths and weaknesses, with ODM allowing easier manipulation and visual representation of the data, and SPSS facilitating a more complete statistical exploration of the results. Using two methods also allows for a comparison of the results, enabling assessment of the agreement between the two statistical procedures. During this project 66 variables were collected, 55 of which were used for the PCA analysis after screening out 9 variables which were associated with FAF imaging itself, as we are attempting to search for relationships between variables that are not directly informative of the FAF imaging outcomes. Also screened out were the variables "patient number" and "years of smoking cessation", the former being a randomly generated number (and therefore not relevant to the analysis) and the latter replaced by "smoking packet years" to allow a less equivocal statistical analysis as discussed earlier.

9.3 Results

9.3.1 Subject to variables ratio

Subject-to-variables (or instances-to-variables) ratio is a crucial consideration when considering PCA, and several authors have suggested guidelines that should be followed when planning an analysis. Suhr et al recommend at least 100 cases,(211) with a subjects-to-variables ratio of no lower than five to one, however, others have suggested that much larger sample sizes are required, e.g. Hutcheson and Sofroniou (they suggest between150-300 cases)(212), and Norusis (at least 300 cases)(213). However, there appears to be a consensus in the literature that ratio criteria alone do not offer an accurate guide for the researcher, with Guadagnoli and Velicer(214) suggesting that the size of the sample depends on the strength of the factors identified, i.e. for weaker relationships larger samples are required. Therefore, it is not possible to predict whether the sample size is adequate until the PCA has been conducted and the relationship strength established. For the purposes of this study the recommendation of Suhr was applied (a subjects-to-variables ratio of at least five to one), i.e. 93 cases collected in this project indicated that a maximum of 18 variables (93/5 = 18.6) from the total of 55 were selected. Conveniently, ODM contains a “Rank” widget, allowing the most informative 18 variables to be selected. However, even within the Rank widget, there are several scoring methods. Therefore, a strategy was devised to select the scoring method that would provide at least 80% of the variance in the data with the lowest number of factors. The value of 80% was selected after reviewing recommendations from literature regarding the percentage of variance that needs to be explained by the selected factors.(214, 215) This was found to be the “Information Gain” scorer, which achieves 83.2% of variance with five factors. The resulting 18 variables were then fed into the “PCA” ODM widget and SPSS software to perform the process of PCA.

Rank

	#	Info. gain	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
OCT large drusen (0-6)		0.342	0.222	0.151	NA	26.847	0.056	0.000
CFP - Simplified SS		0.299	0.178	0.157	NA	36.836	0.173	0.000
OCT large drusen yes or no (1=yes,2=no).	2	0.288	0.294	0.122	NA	15.955	0.308	0.453
CFP disc diam pig anom		0.252	0.302	0.146	NA	51.980	0.046	0.000
CFP Pig anomalies yes or no (1=yes,2=no)	2	0.241	0.394	0.144	NA	4.905	0.354	0.482
CFP large drusen (0-6).		0.230	0.150	0.119	NA	20.445	0.050	0.000
CFP large drusen yes or no (1=yes,2=no).	2	0.182	0.182	0.094	NA	10.925	0.260	0.000
OCT inter drusen (0-6)		0.169	0.102	0.093	NA	22.076	0.033	0.000
OCT large drusen other eye (1=yes,2=no).	2	0.124	0.129	0.071	NA	6.222	0.080	0.000
age		0.124	0.062	0.068	NA	10.798	-0.001	0.000
CFP large drusen other eye (1=yes,2=no).	2	0.123	0.131	0.072	NA	5.734	0.108	0.000
CFP disc diam GA		0.106	0.211	0.060	NA	30.101	-0.000	0.000
OCT small drusen (0-6)		0.103	0.059	0.059	NA	13.943	0.012	0.000
Symptoms ARMD (1=yes,2=no)	2	0.101	0.336	0.057	NA	0.694	0.000	0.210
OCT RPD yes or no (1=yes,2=no)	2	0.099	0.201	0.063	NA	1.519	0.056	0.170
CFP inter drusen (0-6)		0.085	0.055	0.049	NA	10.563	0.023	0.000
OCT disc diam GA		0.084	0.189	0.048	NA	22.849	-0.001	0.000
CFP GA yes or no (1=yes,2=no)	2	0.083	0.216	0.052	NA	0.888	0.014	0.000

Figure 9.1 Portion of Orange Data Mining Rank widget showing the 18 most important variables ranked according to Information Gain.

9.3.2 Initial assumptions regarding the data prior to PCA

Certain assumptions regarding the data need to be made prior to performing PCA. These include the absence of outliers, low level of missing data, continuous data and lack of excessive multicollinearity. The data was initially manually checked for outliers and missing data. ODM continuises categorical variables, with the SPSS Component Correlation Matrix checking for excessive multicollinearity.

9.3.3 Factorability of the data

To initially check whether the data was suitable for PCA, the factorability of the data was checked by two methods via SPSS: Firstly, from observation of the correlation coefficients, and secondly by carrying out the Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) Test. The results of these observations are shown below:

1. From the Correlation coefficients Table, of 306 possible correlations (17 times 18), 158 are > ±0.30. This indicates that the majority of factors (although perhaps not the "vast" majority as suggested by Tabachnick and Fidell(216)) are > ±0.30, suggesting that the data is indeed factorable.

2. Bartlett's test of sphericity and the Kaiser-Meyer-Olkin Test of Sampling Adequacy (KMO) are used to provide complex measures for assessing relationship strengths and indicating variables factorability. Bartlett's test indicates whether or not linear correlations exist, whilst the KMO test is a measure of shared variance. In this project Bartlett's test of sphericity shows that factor analysis is suitable as the p value is statistically significant (i.e. N = 93, Chi squared = 2223.592, $p<0.001$). The KMO tests gives a value of 0.724, i.e. > 0.6, which also indicates that the data is factorable. A KMO value of 0.90 to 1.00 is considered marvellous, 0.80 to 0.89 meritorious, 0.70 to 0.79 middling, 0.60 to 0.69 mediocre, 0.50 to 0.59 miserable and 0.00 to 0.49 factoring not recommended. Therefore, the KMO value from this study falls into the middling category.

In summary, both of the above criteria (the Correlation coefficients and the Bartlett's test of sphericity/KMO tests) indicate that the data is factorable via PCA.

9.3.4 Considering how many factors to retain

The first factor determines the greatest amount of variance in the data, with subsequent factors explaining continually decreasing amounts of variance.(216) The aim of a PCA analysis is to select enough factors to adequately describe the data while ignoring factors that are not relevant.(217) Caution must be exercised, however, as research has demonstrated that it is less detrimental to retain more factors than required than to reject factors that are actually useful, but the researcher must also be careful not to retain too many factors which can result in weak factor loadings.(215) A solution which only contains one or two factors should arouse suspicion, as these may not accurately represent the data's structure.(218),(215) The following three selection methods may be employed to decide how many factors to retain: Kaiser's Criterion, the Scree plot analysis and finally the percentage of variance. The three methods are dealt with in greater depth below:

9.3.4.1 Kaiser's criterion:

Every factor will have an eigenvalue – this is a number which describes the quantity of variance in the data that can be explained by that factor.(215) The critical eigenvalue that will indicate whether a factor should be retained is most commonly calculated using Kaiser's Criterion. This states that if the eigenvalue is greater than or equal to one, then the factor should be retained. (219) Several authors, however, have criticised the Kaiser Criterion method, reporting that it can lead to over and under-extraction of factors.(217, 220, 221) In this study only four factors were found to have an eigenvalue of >1.0.

9.3.4.2 Scree plot:

A Scree plot is a graph where the x axis represents the factors with the eigenvalues forming the y axis. The bend or elbow in the graph is used to select only those factors that occur to the left of the elbow. This technique is, however, somewhat subjective, and can lead to an over-extraction of factors.(220) From the Scree plot, it was found that 2, 3, or 4 or factors could be retained, depending on the interpretation of where the elbow occurs.

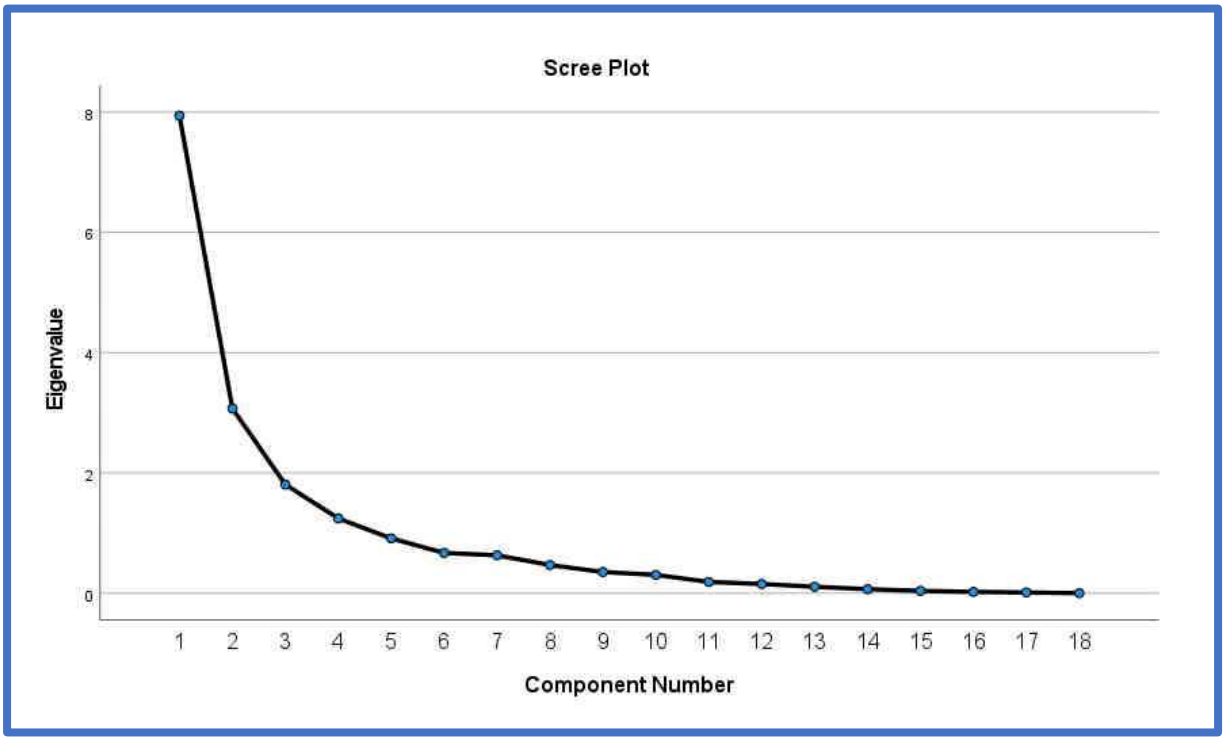

Figure 9.2 Scree plot taken from SPSS showing Factor (or Component number) on the x axis and the Eigenvalue on the y axis.

9.3.4.3 Variance explained by the data:

The advice regarding this property varies according to the different literature, however, a consensus suggests that between 75 and 90% of variance should be explained by the retained factors.(215, 222) For this project a figure of 80% was selected. From the proportion of variance method, both ODM and SPSS agree that five factors are required to explain at least 80% of the variance, with both ODM and PCA via SPSS finding that five factors explain 83.2% of the data. The table below lists the percentages of variance and eigenvalues as calculated by SPSS and ODM for one to five factors.

	Cumulative percentage of variance explained	**Cumulative percentage of variance explained**	**Eigenvalues from SPSS**	**Eigenvalues from ODM**
Number of factors	SPSS	ODM	Eigenvalue/18 = percentage of variance/100	Eigenvalue/18 = percentage of variance/100
1	44.1	44.0	7.9	7.9
2	61.2	60.0	3.1	2.9
3	71.2	70.2	1.8	1.8
4	78.1	77.9	1.2	1.4
5	83.2	83.2	0.9	1.0

Table 9.1 Showing cumulative percentage of variance explained by SPSS and Orange Data Mining per factor retained (rounded to one decimal place) and the corresponding eigenvalues for both methods.

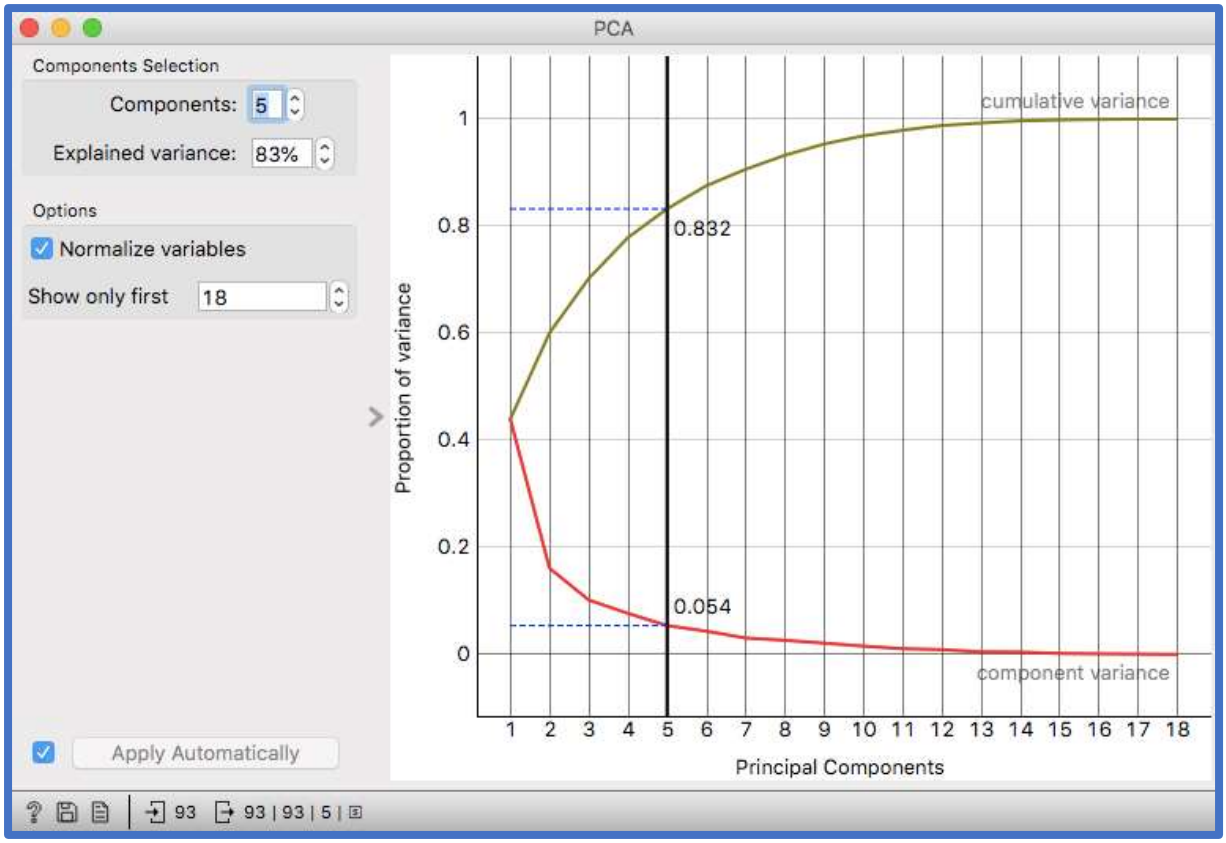

Figure 9.3 PCA widget output from Orange Data Mining

Graph from ODM's PCA widget output where the upper line in green shows the cumulative variance explained by the selected number of factors, whilst the lower red line is the variance explained per factor retained.

9.3.5 Factor retention methods summary

It has been recommended that several methods of factor selection should be compared, before deciding on the number of factors to retain.(219, 221) Following this recommendation, Kaiser's Criterion based on eigenvalues from SPSS suggests four factors are retained, the Scree plot, again based on eigenvalues from SPSS, suggests two, three, or four factors, and the percentage of variance method suggests five factors from calculations by both SPSS and ODM. All the methods employed are therefore in agreement that five or fewer factors should be retained. Therefore, for the purposes of this project's PCA calculations, five factors were retained, following the recommendation that it is more advantageous to retain an excess of factors than to eliminate too many.(215) However, it

has been recommended that the final decision on the factor retention should also be made in view of the overall context of the research.(211)

9.3.6 Factor rotation

As mentioned in the introduction to this section, PCA looks at the relationships between variables, expressed as factors and factor loadings. These are linear relationships, and do not have a single solution,(217) but rather an infinite number of alternative solutions or “rotations”.(216, 223) Rotation helps to improve interpretation of the results by reducing ambiguities that may exist.(224) There are two main types of factor rotation method, orthogonal and oblique. If the factors are unrelated, then orthogonal rotation is appropriate, however, if relationships exist between the factors, oblique rotation is the option of choice. It has been suggested that if the relationships between factors are unknown, then oblique rotation should be employed initially. If correlations between factors transpire to be low, then a switch to the use of an orthogonal rotation should be made.(217) However, when deciding on factor retention, the researcher should again bear in mind the overall context of the research when considering the type of rotation technique to use. In this study the SPSS Component Correlation Matrix indicated that there is a significant relationship (> ±0.32) between factors one and five (-0.477), and therefore it is acceptable to initially use an oblique rotation method. To achieve this, the Direct Oblimin method was selected within the SPSS program.

9.3.7 Interpreting factor rotation results

9.3.7.1 A simple solution

PCA is really a process whereby information is gathered by each step performed, and the results must be interpreted by way of both statistical analysis and domain expertise, with the latter being more relevant.(225) A “simple” solution is achieved when each factor is summarised by a number of variables (ideally three to five(214)) that load strongly onto that factor only.(215, 216)

9.3.7.2 Factor rotations

From Table 10.6 in Appendix 3, the first factor in this study is summarised by six variables, of which all six strongly load onto factor one (i.e. > 0.70), without any of these six variables loading significantly onto any other factor by > 0.40. This satisfies the loading cut-offs as suggested by Guadagnoli and Velicer,(214) Tabachnick & Fidell (216) and Garson (222) for a “simple” solution. However, if we apply the alternative, lower cut-offs recommended by and Tabachnick and Fidell (> 0.32 for cross loading), one variable cross-loads by a significant level onto factor four. As the aim of factor rotation is to explain the data in its simplest form, the former cut-off levels suggested (i.e. > 0.40) were applied to the other factors, and therefore three variables load strongly onto factor two, two variables load strongly onto factors three and four and only one variable onto factor five, in all cases with no significant cross-loadings. However, three variables, whilst not strongly loaded to any factor, did cross load relatively strongly (> 0.40) to more than one factor, these being “CFP simplified severity score”, “OCT large drusen other eye” and “CFP drusen large other eye”. It has been recommended that if a variable does not have a strong loading to any factor (< 0.30), and is not considered an important/essential item conceptually, then it should be removed. Furthermore, if the variable appears to be relatively strongly cross-loaded to more than one factor by > 0.40 it should also be removed.(221) Once these eliminations have been performed, the abridged data should be factored again to fine-tune the solution.(215) It has been also been recommended that each factor is adequately loaded to a sufficient number of variables (ideally three to five) to be considered a reliable and robust factor. However, the researcher must ensure that the variables and the subsequent factors identified can be explained in context of the research and prior knowledge of the subject.(211, 225) The

following steps were followed by an iterative process, repeating steps two and three until no more variable eliminations were required:

1. Eliminate variables with < 0.30 loadings.
2. Eliminate variables with cross-loadings > 0.40.
3. Refactor and return to step one until all variable eliminations are exhausted before moving on to step four.
4. Eliminate factors not considered robust (a factor is robust when a sufficient number of variables are loaded to it – ideally three to five).
5. Check results within the context of the research.

By this method, “CFP simplified severity score”, “OCT large drusen other eye” and “CFP large drusen in the other eye” were eliminated in the first factor rotation, followed on the second factor rotation by “CFP large drusen (0-6)” and “OCT large drusen (0-6)”. These variables were all eliminated by virtue of having significant loadings to more than one factor (> 0.40). This left, on the third factor rotation, 13 variables, with 12 loading strongly (> 0.70) onto a single factor, with four variables loading strongly to factor one, three variables loading strongly to factor two, two variables loading strongly onto factors three and four, and only one variable loading strongly to factor five. In each factoring, the Component Correlation Matrix’s shown in Appendix three indicate that Oblimin rotation is suitable as significant relationships (> 0.32) exist between factors one and five in each factoring.

9.4 PCA results

PCA results from SPSS revealed the presence of four factors with eigenvalues exceeding one (Kaiser's Criterion). These explained 44.1%, 17.1%, 10.0% and 6.9% of the variance respectively, with a cumulative total of 78.1%. For ODM the equivalent figures were similar at 44.0%, 16.0%, 10.2% and 7.7% with a cumulative total of 77.9%. However, inspection of the scree plot from SPSS revealed a clear break after the first three, four or five factors, suggesting the retention of two, three or four factors (with the identification of the elbow in the graph being subjective). The percentage of variance method from both SPSS and ODM suggested five factors should be retained when 80% was selected as the target variance explained by the chosen factors (five factors offer 83.2% of variance according to both SPSS and Orange). Therefore, five factors were retained as it is considered less detrimental to retain poor factors than to reject useful ones.

Summary below of strong loadings (> 0.70) to each of the first five factors after three factor rotations, offering a "simple solution".

Factor 1 CFP intermediate drusen (0-6), OCT small drusen (0-6), OCT intermediate drusen (0-6), OCT RPD yes or no (with the final addition of OCT large drusen (0-6) and OCT large drusen (0-6) from a contextual perspective as discussed below).
Factor 2 CFP disc diameters of GA, OCT disc diameters of GA and CFP GA yes or no.
Factor 3 CFP disc diameters of pigmentary abnormalities and CFP pigmentary abnormalities yes or no.
Factor 4 OCT large drusen yes or no and CFP large drusen yes or no.
Factor 5 Patient age.

Of the factors, only one and two are considered "robust" as these have between three and five variables strongly loaded to them.(219)

PCA was once again repeated after the three factor rotations with the final 12 variables. The percentages of variance for each factor are shown in the table below. Note that only four factors are now required to explain > 80% of the variance in the data.

	Cumulative % of variance explained	**Cumulative % of variance explained**	**Eigenvalues from SPSS**	**Eigenvalues from ODM**
Number of factors	SPSS	ODM	Eigenvalue/12 = percentage of variance/100	Eigenvalue/12 = percentage of variance/100
1	38.1	39.0	4.6	4.7
2	61.0	60.5	2.8	2.6
3	74.1	73.6	1.6	1.6
4	82.5	83.4	1.0	1.2
5	88.9	88.1	0.8	0.6

Table 9.2 Cumulative percentage of variance explained by SPSS and ODM per factor retained (rounded to one decimal place) and the corresponding eigenvalues after three factor rotations and eliminations.

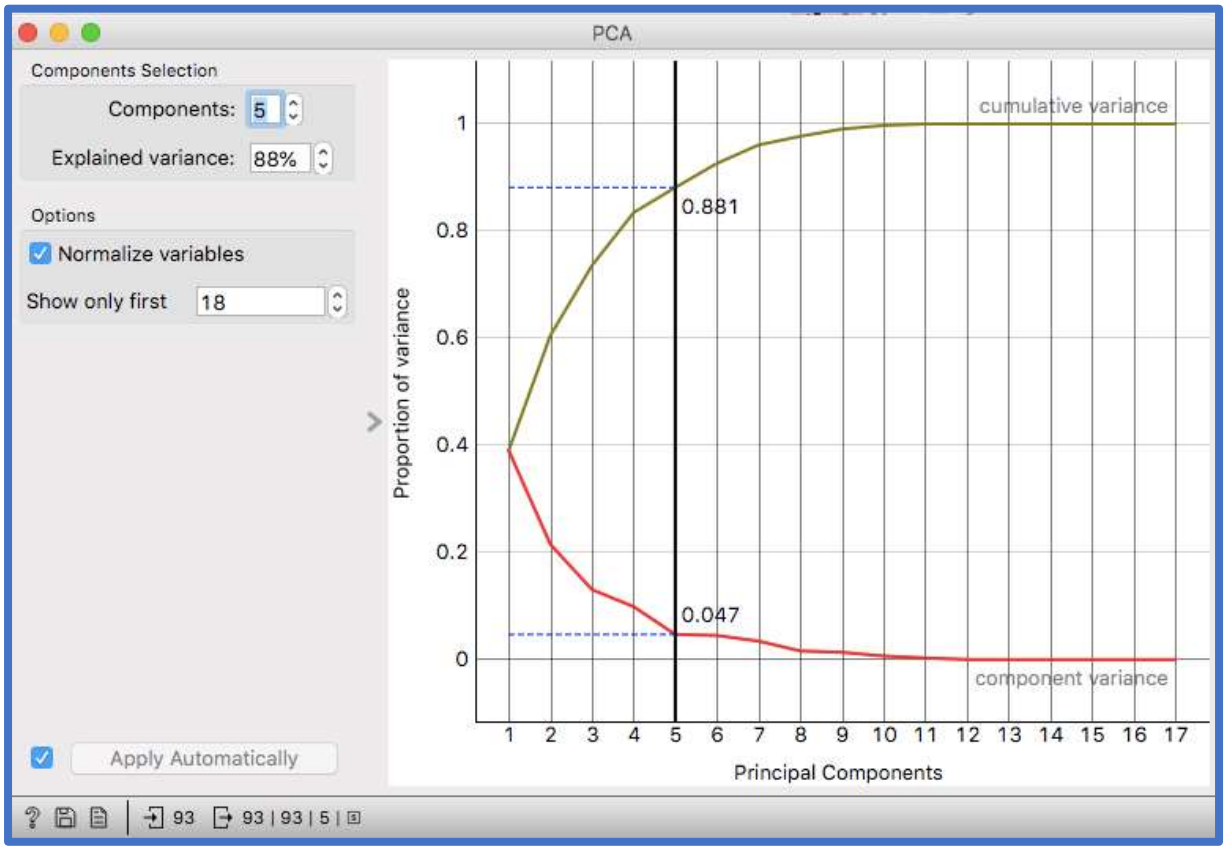

Figure 9.4 PCA widget output from Orange Data Mining after three factor rotations and eliminations

Graph from ODM where the upper line in green shows the cumulative variance explained by the final 12 selected of factors, after three factor rotations, whilst the lower red line is the variance explained per factor retained.

Therefore, the conclusions that can be drawn from PCA, carried out by both SPSS and ODM, are that the following variables are responsible for the majority of the variance in the data from the current study in descending order:

1. The number of small and intermediate drusen as detected by OCT, the number of intermediate drusen as detected by CFP and RPD as detected by OCT. (With the final addition of OCT large drusen (0-6) and OCT large drusen (0-6) from a contextual perspective as discussed below). This can be considered a robust factor and accounts for between 38.1 and 39.0% of variance in the data.
2. Disc diameters of geographic atrophy as detected by either CFP or OCT or both, and the binary finding of the presence or otherwise of geographic atrophy as found by CFP. This can be considered a robust factor and accounts for between 21.5 and 22.9% of variance in the data.
3. CFP detection of pigmentary abnormalities and CFP disc diameters of pigmentary abnormalities. This can be considered a non-robust factor and accounts for between 13.0 and 13.1% of variance in the data.
4. The binary finding of the presence or otherwise of large drusen detected by either CFP or OCT. This can be considered a non-robust factor and accounts for between 8.4 and 9.9% of the variance in the data.
5. Patient age. This can be considered a non-robust factor and accounts for between 4.7 and 6.4% of variance in the data.

Factor and clinical feature represented	SPSS % of variance explained	ODM % of variance explained	Average % of variance explained
1. Drusenoid changes	38.1	39.0	38.6
2. Geographic atrophy	22.9	21.5	22.2
3. Pigmentary abnormalities	13.1	13.0	13.1
4. Presence or absence of large drusen	8.4	9.9	9.2
5. Participant age	6.4	4.7	5.6

Table 9.3 Final percentages of variance for the final five retained principal factors with the clinical features they represent.

However, in the context of the research we should potentially retain CFP large drusen (0-6) and OCT large drusen (0-6) within the “drusenoid” factor one, as these two important variables were eliminated in factor rotations by virtue of strong cross-loadings to factor four, which, by the use of domain expertise, is to be expected as factor four is associated with the binary presence of large drusen detected via either CFP or OCT. These two variables are also considered important by virtue of their league table position in the ODM Rank widget output which placed them at numbers six and one respectively. This gives a final list of the variables responsible for the majority of the variance in the data as:

1. The number of small and intermediate drusen as detected by OCT, the number of intermediate drusen as detected by CFP and RPD as detected by OCT, with the final addition of CFP large drusen (0-6), and OCT large drusen (0-6) retained contextually. Considered a robust factor and accounts for between 38.1 and 39.0% of variance in the data.
2. Disc diameters of geographic atrophy as detected by either CFP or OCT or both, and the binary finding of the presence or otherwise of geographic atrophy as found by CFP. Considered a robust factor and accounts for between 21.5 and 22.9% of variance in the data.
3. CFP detection of pigmentary abnormalities and CFP disc diameters of pigmentary abnormalities. Considered a non-robust factor and accounts for between 13.0 and 13.1% of variance in the data.
4. The binary finding of the presence or otherwise of large drusen detected by either CFP or OCT. Considered a non-robust factor and accounts for between 8.4 and 9.9% of the variance in the data.
5. Patient age. Considered a non-robust factor and accounts for between 4.7 and 6.4% of variance in the data.

9.5 Discussion

Therefore, there are four factors cumulatively explaining > 80% of the variability in the data (the fifth being age), with only factors one and two considered robust. However, as discussed earlier, we should view the results with caution, as a solution with only one or two factors may not accurately represent the data’s structure.(210, 215, 218) With only factors

one and two passing the threshold for a robust factor (i.e. three to five variables strongly loaded onto these factors)(209, 219), one could conclude that factors three, four and five should be ignored, however this depends on the context of the analysis. In this case these less robust factors were retained as they may assist, at least in observational terms, in our understanding of how the variables interact with one another.

Returning to the initial question of whether the sample size of 93 instances, and the strength of the relationships between the variables was adequate for PCA to be carried out, Guadagnoli and Velicer(214) state that if four or more variables load strongly (> 0.60) onto the factors, as in factor one and factor two from this study, then the results can be interpreted regardless of the sample size. Therefore, the results indicate that only those variables relating to drusen, and geographic atrophy can, with certainty, be taken as important in explaining the variance within the data for this study. Note that using the > 0.60 (Guadagnoli and Velicer)(210, 214) stipulation, rather than > 0.70,(225) brings the "symptoms of ARMD" into a strong loading with factor two, indicating that symptoms are associated with geographic atrophy, which would appear plausible with the application of domain expertise, as patients with atrophic macular lesions would be very likely to be aware of defects within their central visual field.

9.6 Conclusions

The following results are an average of the percentages obtained from analyses via SPSS and ODM. It can be concluded that drusenoid changes of all sizes (including reticular pseudodrusen) are responsible for the greatest percentage of variance within the data (38.6%), followed by geographic atrophy (22.2%), then pigmentary abnormalities (13.1%), the presence or absence of large drusen (9.2%) and finally participant age (5.6%).

Chapter 10

Chapter three of this study has shown that pharmacologically induced pupillary dilation appears to make a significant improvement to the quality of FAF images captured via a modified fundus camera, the RS-330 Nidek Retinascan Duo. This finding indicates that in most cases, for FAF imaging to be as clinically useful as possible, it should be carried out post pupillary dilation. This has implications for both the practitioner and the patient, as extra time will be required for the eye examination, and additional tests will be required e.g. Van Herrick angle assessment and post dilation IOP measurements, or in some cases another visit to the practice. The patient may also be unable to drive for a period of time following the eye examination. For busy practitioners, the requirement for pupillary dilation may be barrier to effective clinic management, increasing costs and threatening the commercial viability of the practice.

Chapter four explored the quality of FAF images captured during the study, relating the findings to the measured level of cataract observed by slit-lamp examination. During the data analysis process, the images were graded as “good”, acceptable” and “not acceptable”. To facilitate the statistical analysis via ODM and SPSS, the three grading boundaries were arranged into only two categories: “good” and “not acceptable”. For the first analysis, only “good” images were labelled as “good” (the assumption being that these were of high quality), with both “acceptable” and “not acceptable” images labelled as “not acceptable”. For the second analysis, only “not acceptable” images were labelled as “not acceptable” (the assumption being that these were not clinically useful), with both “good” and “acceptable” images labelled as “good”.

Therefore, the first analysis was concerned with the level of cataract that would still allow a high FAF image quality whilst the second analysis was concerned with what level of cataract would degrade the image to such a degree as to be not clinically useful.

From the first analysis it was found that no FAF image was graded as “good” once the nuclear sclerotic (NUC) cataract reached a WHO grading of > 1.5 (on a scale of 0.0-3.0 with 3.0 indicating a dense cataract). For posterior subcapsular (PSC) cataract this figure was > 1.2 and for cortical (COR) cataract there was no clear cut-off point.

For the second analysis it was found that at a NUC cataract level of approximately 1.9 there was a 50% probability of the image being “not acceptable” (i.e. not be clinically useful as defined by a novel grading system designed specifically for this study), however, there was

no level over which all images collected were “not acceptable”. For COR and PSC cataract the levels found were approximately 1.2 and 1.1 respectively for a 50% probability of the image being “not acceptable”, with again no clear upper level of cataract over which all images collected would be graded as “not acceptable”.

The most common cataract detected in the study was NUC cataract, being present in 71.0% of the participants’ study eye. For COR and PSC cataracts the figures were 15.1% and 7.5% respectively. The number of participants with COR and PSC cataract in this study was relatively small, and so any conclusions drawn with regard to these types of cataract should be viewed with caution. The analysis in this Chapter four therefore indicates, with reasonable certainty, that:

1. If a patient has a NUC cataract of > 1.5 on the WHO grading scale, then some degradation of the FAF image is to be expected (i.e. there is approximately an 85% probability that the FAF image will not be of high quality).
2. There is a 50% probability of the FAF image not being clinically useful once the NUC cataract reaches a grading level of 1.9.

The rules of thumb above could help to guide primary care practitioners to decide when FAF imaging is likely to be a worthwhile exercise in patients with NUC cataract, and therefore likely to add clinically useful information to the examination results.

Chapter five was concerned with the visual comfort levels experienced by participants within the study who were exposed to the relatively bright flash of light associated with FAF imaging. Of the variables explored, only age was found to be negatively associated with the comfort score, with participants registering a comfort score of ≤ 3 out of 5 (0 = very uncomfortable, 5 = very comfortable) having a mean age of 77.56 years±10.3 years) and those with a comfort score of > 3 having a mean age of 70.35 years±9.7 years). This finding could lead practitioners to modify the way in which they would manage older patients before and after FAF imaging is performed, e.g. explaining the importance of the test and allowing more time for patient recovery following the procedure in a safe and comfortable environment. Other findings from Chapter five were that a larger pupil size post dilation does not have a negative effect on comfort during the FAF acquisition process, and also that being a migraine sufferer does not indicate that FAF imaging will be any more uncomfortable for the participant than for a non-migraine sufferer. However, it is important to note that patients who reported that their migraines were triggered specifically by bright lights (i.e.

photosensitive migraines) were excluded from the study. Overall, the FAF acquisition process appears to have been well tolerated during the study with no adverse events observed, with the mean comfort score for participants < 77 years of age being 4.46±0.80), and for those > 77 years of age 3.96±0.90).

Chapter six examined, with the help of artificial intelligence, the variables collected from patients during this study. The data was uploaded via a Microsoft Excel spreadsheet into an ODM software workflow, with an initial statistical analysis carried out via the Boxplot and Distribution widgets. For these analyses, the eight FAF patterns were converted into a dichotomous grading system to facilitate statistical calculations. Patterns labelled as normal and minimal change were redefined as “normal”, with the remaining patterns of focal increased, patchy, linear, lacelike, reticular, and speckled redefined as “abnormal”. The widgets indicated a number of trends within the data, showing that an abnormal FAF result was positively associated with increased age, the number of packet years of smoking, the number of drusen of all sizes but especially large drusen, the simplified severity score, pigmentary abnormalities, and the presence of reticular pseudodrusen. An abnormal FAF result was not significantly associated with refractive status of the eye (measured via the RS-330’s OCT Gullstrand function), nor with having a first degree relative with ARMD. There was also no significant association between the pupil size post dilation and the abnormality or otherwise of the FAF image result. Many of these findings would be expected, given the known associations between age, smoking, drusen number/size and pigmentary abnormalities with ARMD, with the only unexpected finding being the lack of evidence for an association of an abnormal FAF result for participants having a first degree relative with ARMD. The explanation for this could be that patients were poor historians regarding family history, or that the sample size was not large enough for the effect size involved.

The following rules of thumb for primary care practitioners, based on these findings, could be suggested:

Any of the following six factors indicate a probability of an abnormal FAF result of over 90%:

1. > 50 small drusen in the study eye
2. > 40 intermediate drusen in the study eye
3. > 20 large drusen in the study eye
4. > 92 years of age
5. > 30 packet years of smoking

6. the presence of any pigmentary abnormalities in the study eye

For the presence of RPD the probability figure for an abnormal FAF result is approximately 80%.

For the next section of Chapter six, machine learners were employed to rank the collected variables in order of their importance in predicting an abnormal FAF result. Note that clinical variables based on patient history were also included in all the following analyses. Initially the Naïve Bayes and Logistic Regression models were selected for this process (as both learners are the only ones in the ODM software that link to the useful Nomogram widget), with the former being a relatively simple machine learner and the latter more complex. The process was carried out using data from CFP alone, and then for data from CFP and OCT combined, with the results compared. It was found that the best performance for informedness came from hyperparameter fine-tuned Logistic Regression for CFP alone (62.1%), with the same machine learner using data from both CFP and OCT combined providing the best AUC (0.903). The former result may appear surprising, however, when the ranking of variables is studied, this becomes explainable by virtue of the high position in the rankings of variables associated with CFP alone, i.e. the Simplified Severity Score (calculated only from CFP findings), pigmentary abnormalities and disc diameters of geographic atrophy as detected via CFP. When other machine learners were also introduced, the extra information extracted via CFP and OCT combined appeared to be exploited by the SVM learner, which achieved a sensitivity of 0.882 and a specificity of 0.820 and the highest overall informedness level of 70.2%. *On average*, using the seven model learners, CFP alone achieved an informedness of 56.3% and an AUC of 0.837. For CFP and OCT combined this rose to an informedness of 61.9% and an AUC of 0.854. Therefore, in summary, CFP alone achieves, on average across seven model learners, an informedness 5.6% lower than CFP and OCT combined and an AUC 0.017 lower. This indicates that CFP alone may be the only test required to satisfy the practitioner who wishes to gather information from patients regarding their likelihood of having an abnormal FAF result, by virtue of the similar AUC results compared to CFP and OCT combined, although the informedness is somewhat lower.
Using the Nomogram widget available within ODM software, it was possible to show that clinical support tools could be developed that would practitioners to predict the probability of an abnormal FAF image result, by using data from CFP and OCT, without having to carry out FAF imaging itself using Naïve Bayes and hyperparameter fine-tuned Logistic Regression model learners.

Another final question was addressed in Chapter six. Could FAF imaging change the recommendation of dietary supplementation based on CFP and OCT results? To test this theory, an assumption was made that dietary supplementation would be advised if the FAF results did not fall into the normal or minimal change categories as outlined by Bindewald in 2005, but rather into one of the six more extreme patterns. These decisions were then compared via ODM to test whether advice based on the presence of large drusen might be potentially be changed by performing FAF imaging in this study. It was found that, so long as OCT was used to detect large drusen, then no decisions were overturned by FAF imaging, however, when CFP was relied upon to detect large drusen, in three cases FAF findings would lead a practitioner to recommend supplements when CFP findings would not. This may be due to the underestimation of drusen size via CFP compared to OCT; a finding that has been found in previous research, most likely due to the superficial enface nature of CFP imaging compared to cross-sectional OCT.

Chapter seven analysed the use of unsupervised and supervised transfer machine learning to examine image data collected within this study. In this context transfer learning uses machine learners already pretrained on images taken from objects not related to the field of ophthalmology and then retrains them with data from the study. For unsupervised machine learning no human image tagging is used with the machine learners simply detecting patterns within the data. For supervised machine learning, the algorithms utilise human tagging to learn how to place images into specific categories. Overall the results indicate poor levels obtained for informedness and AUC. This may be due to the data within this study being “too raw”. Previous ophthalmological studies that achieved greater success with transfer learning used heavily pre-processed images, leading to the conclusion that for transfer learning to be of practical use within primary care, image pre-processing is a vital step and should be integrated into equipment utilised by optometrists.

Chapter eight explored how clinical findings from CFP and OCT are associated with the eight different FAF image patterns identified in early ARMD as utilised in this study. Two methods were used to elucidate these associations, one based on the percentages of the various patterns found in the current and a previous study involving FAF patterns and the inclusion and exclusion criteria for each paper. The results from this exploration were compared with results from an ODM analysis based solely on data from the current study. These analyses indicate that geographic atrophy and pigment epithelial detachments may be linked to a patchy FAF pattern, whilst an increased simplified severity score (formed from an amalgam of large/intermediate drusen and pigmentary abnormalities) is associated with a

focal increased pattern. However, the latter finding is somewhat equivocal as the two methods appear to disagree when pigmentary abnormalities and large/intermediate drusen are examined separately, with one method (from the current study data alone) linking these clinical findings to a patchy pattern and the other (from comparing data from both studies) to a focal increased/speckled pattern.
Conclusions can therefore be drawn that geographic atrophy and pigment epithelial detachments appear to be linked to a patchy FAF pattern, with the association of pigmentary abnormalities and large/intermediate drusen and specific FAF patterns being less clear cut. A higher simplified severity score appears to be linked, partially at least, to a focal increased FAF pattern.

Chapter nine on principle component analysis (PCA) subjected 18 variables to analysis via two methods, ODM software and IBM's SPSS statistical program, in order the elucidate the principle components (factors) affecting the variability of the data collected in this study. Using two methods meant that the benefits of two very different statistical programs could be accessed as well as enabling corroboration of the results. SPSS facilitates initial exploration of the data to check for factorability via three different methods: inspection of the correlation matrix, Bartlett's test of sphericity and finally the Kaiser-Meyer-Olkin (KMO) test. All three methods indicated that the data was factorable, and so PCA was carried out. Five factors were retained to ensure that at least 80% of the variance in the data was explained by the chosen factors.

Three factors rotations of the initial 18 variables were carried out to leave 12 variables once exclusions had been made based on cross loadings between factors and weak loadings to factors. The five factors were:

1. Drusenoid changes explaining (on average between ODM and SPSS) 38.6% of the variance in the data
2. Geographic atrophy/Symptoms of ARMD explaining 22.2% of variance
3. Pigmentary abnormalities explaining 13.1% of variance
4. Presence or absence of large drusen explaining 9.2% of variance
5. Participant age explaining 5.6% of variance.

It should be noted that only factors one and two are considered "robust" given the stipulation advised by Costello and Osbourne in 2005,(219) i.e. that between three and five variables should load strongly onto each factor. Also, it should be borne in mind that solutions with

only one or two factors may not adequately explain the data's structure.(210, 215, 218) Domain expertise could, however, lead to an interpretation of the results which includes all five of the final factors, given the agreement between the two model learners and the high importance ranking of the selected variables for predicting the outcome of the FAF result shown in Chapter six. Therefore, in conclusion, drusen of all sizes (including reticular pseudodrusen), can be taken to be responsible for the greatest percentage of variability in the data (approximately 38.6%), followed by geographic atrophy (22.2%), pigmentary abnormalities (13.1%), the presence or absence of large drusen (9.2%) and finally participant age (5.6%).

10.2 Final conclusions, clinical implications and impact of this study

As a result of the work carried out in this study, the following is a list of the clinical implications and the impact on the work of clinicians:

1. It is advisable to carry out pharmacologically induced pupillary dilation prior to fundus autofluorescence (FAF) imaging via a modified fundus camera-based system.
2. This study indicates that if the nuclear sclerotic cataract score on the WHO grading scale is > 1.5, then there is a high probability (approximately 85%) that the FAF image will not be of high quality. If the nuclear sclerotic score is > 1.9, then there is approximately a 50% probability that the FAF image will not be clinically useful.
3. This study indicates that FAF imaging is acceptable in terms of visual/ocular comfort to patients, however, comfort levels progressively fall with increasing age.
4. This study indicates that FAF imaging results can be predicted with an informedness of up to 70.2% and an area under the curve of up to 0.903 by using data from patient history and colour fundus photography (CFP) combined with data from optical coherence tomography (OCT). Also, carrying out FAF imaging is unlikely to alter dietary supplementation recommendations from the current method based on the detection of large drusen, provided the large drusen size is measured via OCT B-scan rather than by CFP imaging. There is over a 90% probability of an abnormal FAF result in an eye with any of the following: > 50 small, > 40 intermediate, > 20 large drusen, > 92 years of age, > 30 packet years of smoking, any pigmentary abnormalities. If reticular pseudodrusen (RPD) are present then there is approximately an 80% probability of an abnormal FAF result.
5. This study indicates that for transfer learning to be useful in clinical optometric practice, pre-processing of images is likely to be required.

6. This study indicates that geographic atrophy (GA) and pigment epithelial detachments are related to a patchy FAF pattern, whilst the simplified severity score is partially associated with a focal increased FAF pattern. RPD are associated with a reticular FAF pattern.
7. This study has indicated that small, intermediate and large drusenoid changes (including RPD) appear to be related, and cumulatively explain 38.6% of the variance in the clinical data collected in this study. GA and symptoms of age-related macular degeneration also appear to be related, explaining 22.2% of the data variance. Pigmentary abnormalities explain 13.1% of the variance, presence or absence of large drusen explain 9.2% and age explains 5.6%.

10.3 Limitations

All patients were recruited from a single high-street optometric practice in a relatively wealthy area of the UK with poor ethnical diversity. The data from this study was collected by RS, an optometrist with primary and secondary care experience, however, the clinical grading decisions were not validated by another optometrist or ophthalmologist, except for the first 50 images in Chapter three. In particular, the grading of FAF images was a self-taught skill, using previously published research as a guide. To make the project more relevant to primary care clinicians, anatomical markers were utilized rather than computerised image analysis tools, however, this introduces the problem of subjective interpretation and individual anatomical variation between patients. Where possible, the order of images was randomized, and images from the same patient imaged via different modalities were graded on different days to avoid decision making bias, although this potential for bias cannot be entirely eliminated.

The comfort analysis was based on a single question involving the visual comfort of the flash associated with FAF imaging, however, other reasons for poor comfort, e.g. physical positioning could also have been investigated to help to investigate confounding factors.

In the cataract analyses, the data on cortical and posterior subcapsular cataract suffered from a lack of instances and larger sample sizes would have been beneficial.

10.3 Future work

Future work in the field of FAF could benefit from the collection of additional variables not recorded within this study including complete and incomplete retinal pigment epithelial and outer retinal atrophy (cRORA and iRORA respectively) and hyperreflective foci (HRF). HRF are related to hyperpigmentation and the double layer sign, and pigmentary abnormalities are strongly related to an abnormal FAF pattern, so recording these subtle features could have led to further interesting analyses.

This study has also indicated that further research may be advantageous in the following areas: the toxicity and risks of exposing the human eye to short wavelength blue light during retinal imaging, the uses of quantitative FAF for the phenotyping of retinal disease, peripheral FAF and its use in the risk profiling of patients and for the detection and monitoring of ARMD, FAF image analysis training for clinicians, the role of FAF in the monitoring of retinal health with regards to dietary supplementation, the role of FAF in the monitoring of retinal sensitivity, the impact of local lipid recycling in the formation of RPD, the role of cholesterol lowering medications in ARMD, computer analysis of FAF images and retinal image preprocessing.

10.4 A final note on pegcetacoplan

There has been great interest over recent years in medical treatments for slowing the growth of geographic atrophy. In the USA, intravitreal (IV) pegcetacoplan (Apellis Pharmaceutical's Syfovre ®) has only recently (17th of February 2023) been approved by the US Food and Drug Administration (FDA) for this purpose, and, at the time of writing, applications are pending for its deployment within the UK. With this intervention only just having been approved, and therefore a new development within the field of FAF, this thesis was not planned with the knowledge of pegcetacoplan in mind, and may well have focussed on different aspects of FAF if the study was being designed today. For example, the project may have centred more on perilesional FAF. The OAKS and DERBY studies,(226) which are concerned with the efficacy and safety of IV pegcetacoplan in the treatment of GA, listed their inclusion criteria as: $\geq$ 60 years of age, BCVA $\geq$ 24 letters (ETDRS), GA lesion size of $\geq$ 2.5 and $\leq$ 17.5mm² (may be either foveal or extrafoveal), if the GA is multifocal then at least one lesion must be $\geq$ 1.25mm², and finally the presence of perilesional hyper-autofluorescence. Those subjects who did not demonstrate perilesional hyper-autofluorescence were excluded from the studies. These criteria highlight the clinical value of FAF, and indicate that this imaging modality may become significantly important in

primary care optometry over the coming years with the potential availability of IV drugs to slow the progression of GA, and FAF findings could therefore become a crucial finding to trigger referrals to the hospital eye service. Indeed, one can envisage a time, not in the distant future, where medicolegal action could potentially be taken against optometrists who do not carry out FAF on a patient who subsequently loses vision as a consequence GA expansion.

Appendix 1

Statistical analysis of the comparison of the two graders HB and RS for assessing FAF image quality using intraclass Kappa and positive and negative agreement. (This is superior to a simple percentage of agreement calculation, as Kappa takes into consideration the possibility of an agreement occurring by chance).
Intraclass Kappa, and positive and negative agreement was used to test the agreement between to the two graders. (N=50).

	RS			**Total**
HB	Good	Acceptable	Not-acceptable	
Good	27 (a)	0 (b)	0 (c)	27 (a+b+c)
Acceptable	4 (d)	4 (e)	1 (f)	9 (d+e+f)
Not-acceptable	1 (g)	6 (h)	7 (i)	14 (g+h+i)
Total	32 (a+d+g)	10 (b+e+h)	8 (c+f+i)	50 (N)

Table 10.1 The grading of image quality selected by the two graders HB and RS

N = 50

The observed agreement (Po)
= (a+e+i)/N
= (27+4+7)/50 = 38/50 = 0.76

Standard error of Po (SE(Po))
= √ (Po * (1-Po)/N)
= √(0.76*(1-0.76)/50)
= √(0.003648)
= 0.0640

95% confidence limits for Po:

Lower limit
= Po – 1.96*SE(Po)
= 0.76 – 1.96*(0.0640)
= 0.76 – 0.1254
= 0.6346

Upper limit
= Po + 1.96*SE(Po)
= 0.76 + 1.96*(0.0640)
= 0.76 + 0.1254
= 0.8854

However, how does the agreement improve upon what might be expected by chance alone? To answer this, Pe was calculated (the calculation of agreement by chance).

Pe

= (((a+d+g)*(a+b+c)/N + (b+e+h)*(d+e+f)/N + (c+f+i)*(g+h+i))/N))/N

= ((32*27)/50 + (10*9)/50 + (8*14)/50)/50

= 17.28 + 1.80 + 2.24

= 21.32/50

= 0.4264

Validity is best assessed using intraclass Kappa for dichotomous data.

Calculation of the intraclass Kappa (K)

K

= (Po-Pe)/(1-Pe)

= (0.76-0.4264)/(1- 0.4264)

= 0.5816

The same analysis was also carried out via SPSS as shown below.

HB_first_grader * RS_second_grader Crosstabulation

			RS_second_grader			
			1.00	2.00	3.00	Total
HB_first_grader	1.00	Count	27	0	0	27
		% within HB_first_grader	100.0%	0.0%	0.0%	100.0%
		% within RS_second_grader	84.4%	0.0%	0.0%	54.0%
	2.00	Count	4	4	1	9
		% within HB_first_grader	44.4%	44.4%	11.1%	100.0%
		% within RS_second_grader	12.5%	40.0%	12.5%	18.0%
	3.00	Count	1	6	7	14
		% within HB_first_grader	7.1%	42.9%	50.0%	100.0%
		% within RS_second_grader	3.1%	60.0%	87.5%	28.0%
Total		Count	32	10	8	50
		% within HB_first_grader	64.0%	20.0%	16.0%	100.0%
		% within RS_second_grader	100.0%	100.0%	100.0%	100.0%

Figure 10.1 SPSS Crosstabulation output

The manual calculations from above agree with the Kappa value obtained with IBM's SPSS = 0.582 (i.e. moderate agreement) as shown below in the Symmetric Measures Table below.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
McNemar-Bowker Test	8.571	3	.036
N of Valid Cases	50		

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.582	.092	5.772	<.001
N of Valid Cases		50			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 10.2 SPSS Chi-squared Test results and Symmetric Measures output
Confidence limits for intraclass Kappa:

Standard error (SE) for intraclass Kappa

= SE(K)

= $\sqrt{((Po*(1-Po))/(N*(1-Pe)^2))}$

= $\sqrt{((0.76*0.24))/(50*(0.5736)^2))}$

= $\sqrt{(0.1824/16.4508)}$

= $\sqrt{(0.01109)}$

= 0.1053

95% confidence limits for Kappa:

Lower limit

= K – 1.96* SE(K)

= 0.5816 – 1.96*0.1053

= 0.5816 – 0.2064

= 0.3752

(minimal agreement)

Upper limit

= K + 1.96*0.1053

= 0.5816 + 1.96*0.1053

= 0.5816 + 0.2064

= 0.7880

(moderate agreement)

Interpreting Kappa values:

Kappa Value	Level of agreement
0.0 – 0.20	None
0.21 – 0.39	Minimal
0.40 – 0.59	Weak
0.60 – 0.79	Moderate
0.80 – 0.90	Strong
Over 0.90	Almost perfect

Table 10.2 Table of Cohen's kappa and level of agreement (227)

Summary of Kappa test findings:

For comparison of the two graders, intraclass Kappa has a 95% confidence range from 0.3752 to 0.7880. These values have a range from minimal to moderate agreement between the two graders for FAF image quality.

A known problem with intraclass Kappa is that the same test might yield different values depending on the proportions of positive and negative cases in the sample. It has been suggested that this problem can be overcome if positive agreement (PA) (analogous to sensitivity) and negative agreement (NA) (analogous to specificity) are calculated instead of Kappa. However, for calculation of PA and NA, a 2 by 2 grid is required.

To enable calculation of PA and NA, figures from Table 10.1 were amalgamated to produce a 2 by 2 grid. If agreement was not absolute, e.g. one grader selected "good", whilst the other selected "acceptable", this was considered to be a partial positive agreement of 50%, whilst if one grader selected "acceptable" and the other "not acceptable", this was considered to be 50% negative agreement. If both graders chose "acceptable", this was considered to be both 50% positive and negative agreement. Table 10.3 shows how these amalgamated figures were calculated.

	RS acceptable	RS not-acceptable
HB acceptable	31 (A) = (a + b/2 + d/2 + e/2) = 31	0.5 (B) = (c + b/2 + f/2) = 0.5
HB not-acceptable	6 (C) = (g + d/2 + h/2) = 6	12.5 (D) = (i + f/2 + h/2 + e/2) = 12.5

Table 10.3 Amalgam of decision for graders HB and RS creating a 2 by 2 grid to enable calculation of PA and NA.

N = 50

The positive agreement (PA) and its 95% confidence limits are calculated:

PA
= (2*A)/((2*A) + B + C)
= (2*31)/((2*31) + 0.5 + 6)
= 62/68.5
= 0.9051

SE(PA)
= (√((4*A)*(C + B)*(A + B + C)))/((2*A) + B + C)²
= (√((4*31)*(6 + 0.5)*(31 + 0.5 + 6)))/((2*31) + 0.5 + 6)²
= (√((124*6.5)*(36.5)))/(68.5)²
= (√((806)*36.5)/4692.25
= √(29419)/4692.25
= 171.5197/4692.25
= 0.03655

Lower limit of PA
= PA – 1.96*SE(PA)
= 0.9051 – 1.96*0.03655
= 0.9051 – 0.071638
= 0.8335

Upper limit of PA
= PA + 1.96*SE(PA)

= 0.9051 + 1.96*0.03655

= 0.9051 + 0.071638

= 0.9767

The negative agreement (NA) and its 95% confidence limits are calculated:

NA
= (2*D)/((2*D) + B + C)
= (2*12.5)/((2*12.5) + 0.5 + 6)
= 25/31.5
= 0.7937

SE(NA)
= (√((4*D)*(C + B)*(D + C + B)))/((2*D) + B + C)²
= (√((4*12.5)*(6 + 0.5)*(12.5 + 6 + 0.5))))/(2*12.5) + 0.5 + 6)²
= (√((50)(6.5)*(19)))/(31.5)²
= √((6175)/992.25
= (78.5812)/992.25
= 0.07919

Lower limit of NA
= NA – 1.96*SE(NA)
= 0.7937 – 1.96*0.07919
= 0.7937 – 0.1552
= 0.6385

Upper limit of NA
= NA + 1.96*SE(NA)
= 0.7937 + 1.96*0.0719
= 0.7937 + 0.1409
= 0.9346

Summary of PA and NA findings by the method described above from Table 10.3:

For the comparison of the two graders, PA has a 95% confidence range from 0.8335 to 0.9767, whilst NA has a 95% confidence range from 0.6385 to 0.9346. Therefore, PA ranges from a strong to almost perfect agreement whilst NA ranges from a moderate to strong agreement.

An alternative approach is to only allow for absolute agreement between graders, i.e. to define positive agreement as both graders grading the image as “good”, and negative agreement as both graders grading the image as “not acceptable”. This would also lead to the 2 by 2 grid shown below in Table 10.4 below which shows only absolute agreement, from Table 10.1, between graders HB and RS, creating a 2 by 2 grid, which enables calculation of PA and NA by an alternative method from that set out above in Table 10.3.

	RS acceptable	RS not-acceptable
HB acceptable	27 (a)	0 (c)
HB not-acceptable	1 (g)	7 (i)

Table 10.4 Absolute agreement between HB and RS.

N = 35

The positive agreement (PA) and its 95% confidence limits are calculated:

PA
= (2*a)/((2*a) + c + g)
= (2*27)/((2*27) + 0 + 1)
= 54/55
= 0.9818

SE(PA)
= √((4*a)*(g + c)(a + c + g))/((2*a) + c + g)²
= √((4*27)*(1)*(28))/((2*27) + 0 + 1)²
= √(108)*(28)/(54 + 1)²
= √(3024)/(3025)
= 54.9909/3025
= 0.01818

Lower limit of PA
= PA – 1.96*SE(PA)
= 0.9818 – 1.96*0.01818
= 0.9818 – 0.03563
= 0.9462

Upper limit of PA

= PA + 1.96*SE(PA)

= 0.9818 + 1.96*0.01818

= 0.9818 + 0.03563

= 1.0174

= recorded as 1.

The negative agreement (NA) and its 95% confidence limits are calculated thus:

NA

= (2*i)/((2*i) + c + g)

= (2*7)/((2*7) + 0 + 1)

= (14)/14 + 0 + 1)

= 14/15

= 0.9333

SE(NA)

= (√((4*i)*(c + g)*(c + g + i)))/((2*i) + c + g)²

= (√((4*7)*(0 + 1)*(0 + 1 + 7)))/((2*7) + 0 + 1)²

= (√((14)*(1)*(8)))/((14 +0 + 1)²

= (√112)/(15)²

= 10.5830/225

= 0.04704

Lower limit of NA

= NA – 1.96*SE(NA)

= 0.9333 – 1.96*(0.04704)

= 0.9333 – 0.09220

= 0.8411

Upper limit of NA

= NA + 1.96*SE(NA)

= 0.9333 + 1.96*(0.04704)

= 0.9333 + 0.09220

= 1.0255

= recorded as 1.

Summary of PA and NA findings by the alternative method shown above in Table 10.4:

For the comparison of the two graders, PA has a 95% confidence range from 0.9642 to 1, whilst NA has a 95% confidence range from 0.8411 to 1. Therefore, both PA and NA range from a strong to an almost perfect agreement.

Therefore, in final summary, for comparison of the two graders HB and RS, the Kappa value indicates minimal to moderate agreement, however, to take account of the number of positive and negative cases within the sample, PA and NA were also calculated by two alternative methods. The first method indicated a PA of strong to almost perfect agreement, whilst NA gave moderate to strong agreement. The second method indicated strong to almost perfect agreement for both PA and NA.

Appendix 2

Statistical analysis for the comparison of the quality of FAF images taken from 25 eyes pre and post dilation according to each grader using the McNemar test for repeat measurements.

Grader HB	**before dilation**	**after dilation**
good	9	18
acceptable	6	3
not-acceptable	10	4
Total	25	25

Table 10.5 Quality of FAF images from the first 25 participants graded by HB

Grader RS	**before dilation**	**after dilation**
good	11	21
acceptable	7	3
not-acceptable	7	1
Total	25	25

Table 10.6 Quality of FAF images from the first 25 participants before and after dilation for grader RS

Manual calculation of the McNemar test:

Again, as with positive and negative agreement, the McNemar test relies on a 2 by 2 grid, so a strategy was devised to reduce the 3 by 3 grid to a 2 by 2 grid as shown in Table 10.7 below. This was done by recording both “good” and “acceptable” images as “acceptable”, and “not acceptable” images as “not acceptable”, thereby creating a dichotomous system of grading.

Grader HB		**Before dilation**	**Before dilation**
		acceptable	not-acceptable
After dilation	acceptable	15 (a)	6 (b)
After dilation	Not acceptable	0 (c)	4 (d)

Table 10.7 Showing dichotomous grading of the FAF images from the first 25 participants before and after dilation for grader HB

X (McNemar test statistic)

$= (b-c)^2/(b + c)$

$= (6-0)^2/(6+0)$

$= 36/6$

$= 6$

$p = 0.03125$ (exact p-value binomial test), therefore the result is statistically significant at the 5% level and so it appears likely that FAF images are significantly improved after dilation. Therefore, for grader HB, it is possible to conclude that pupil dilation does make a significant difference to the quality of FAF images obtained.

HB_after_dilation * HB__before_dilation Crosstabulation

			HB__before_dilation		
			acceptable	not-acceptable	Total
HB_after_dilation	acceptable	Count	15	6	21
		% within HB_after_dilation	71.4%	28.6%	100.0%
		% within HB__before_dilation	100.0%	60.0%	84.0%
	not-acceptable	Count	0	4	4
		% within HB_after_dilation	0.0%	100.0%	100.0%
		% within HB__before_dilation	0.0%	40.0%	16.0%
Total		Count	15	10	25
		% within HB_after_dilation	60.0%	40.0%	100.0%
		% within HB__before_dilation	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	Exact Sig. (2-sided)
McNemar Test		.031[a]
N of Valid Cases	25	

a. Binomial distribution used.

Figure 10.3 SPSS screenshot of calculation of McNemar test statistic for comparison of pre and post dilation FAF images for grader HB which corroborates the manual calculation.

Grader RS		**Before dilation**	**Before dilation**
		acceptable	not-acceptable
After dilation	acceptable	18 (a)	6 (b)
After dilation	not-acceptable	0 (c)	1 (d)

Table 10.8 Showing dichotomous grading of FAF images from the first 25 participants before and after dilation for grader RS.

X (NcNemar test statistic)

$= (b-c)^2/(b + c)$

$= (6-0)^2/(6 + 0)$

$= 36/6$

$= 6$

p = 0.03125 (exact p-value binomial test), therefore the result is statistically significant at the 5% level and it appears likely that FAF images are likely significantly improved after dilation. Therefore, for grader RS, it is possible to conclude that pupil dilation does make a significant difference to the quality of FAF images obtained.

RS_after_dilation * RS_before_dilation Crosstabulation

			RS_before_dilation		
			1.00	2.00	Total
RS_after_dilation	acceptable	Count	18	6	24
		% within RS_after_dilation	75.0%	25.0%	100.0%
		% within RS_before_dilation	100.0%	85.7%	96.0%
	not-acceptable	Count	0	1	1
		% within RS_after_dilation	0.0%	100.0%	100.0%
		% within RS_before_dilation	0.0%	14.3%	4.0%
Total		Count	18	7	25
		% within RS_after_dilation	72.0%	28.0%	100.0%
		% within RS_before_dilation	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	Exact Sig. (2-sided)
McNemar Test		.031[a]
N of Valid Cases	25	

a. Binomial distribution used.

Figure 10.4 SPSS screenshot of calculation of McNemar test statistic for comparison of pre and post dilation FAF images for grader RS which corroborates the manual calculation.

Appendix 3

Details of the principle component analysis factor rotation process from SPSS.

First factor rotation:

Component Correlation Matrix

Component	1	2	3	4	5
1	1.000	-.120	-.138	-.322	-.477
2	-.120	1.000	.220	.193	.179
3	-.138	.220	1.000	.186	.162
4	-.322	.193	.186	1.000	.253
5	-.477	.179	.162	.253	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Figure 10.5 Component correlation Matrix from SPSS for the first factor rotation

Pattern Matrix[a]

	Component				
	1	2	3	4	5
CFP inter drusen (0-6)	.972				
OCT inter drusen (0-6)	.947				
OCT small drusen (0-6)	.856				
OCT RPD yes or no (1=yes,2=no)	-.720				
CFP large drusen (0-6).	.718			-.374	
OCT large drusen (0-6)	.711			-.319	
CFP disc diam GA		-.987			
OCT disc diam GA		-.973			
CFP GA yes or no (1=yes, 2=no)		.738	.326		
Symptoms ARMD (1=yes, 2=no)		.654			
CFP disc diam pig anom			-.911		
CFP Pig anomalies yes or no (1=yes,2=no)			.866		
CFP large drusen yes or no (1=yes,2=no).				.754	
OCT large drusen yes or no (1=yes,2=no).				.744	
CFP - Simplified SS			-.416	-.601	-.347
age					-.803
OCT large drusen other eye (1=yes,2=no).				.490	.689
CFP large drusen other eye (1=yes,2=no).				.527	.597

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 14 iterations.

Figure 10.6 Pattern Matrix from SPSS for the first factor rotation

Second factor rotation:

Component Correlation Matrix

Component	1	2	3	4	5
1	1.000	-.095	-.057	-.421	.435
2	-.095	1.000	.209	.191	-.126
3	-.057	.209	1.000	.193	-.146
4	-.421	.191	.193	1.000	-.248
5	.435	-.126	-.146	-.248	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Figure 10.7 Component correlation Matrix from SPSS for the second factor rotation

Pattern Matrix[a]

	Component				
	1	2	3	4	5
CFP inter drusen (0-6)	.943				
OCT inter drusen (0-6)	.885				
OCT small drusen (0-6)	.865				
OCT RPD yes or no (1=yes,2=no)	-.762				
CFP disc diam GA		-.982			
OCT disc diam GA		-.970			
CFP GA yes or no (1=yes, 2=no)		.719			
Symptoms ARMD (1=yes, 2=no)		.664			
CFP disc diam pig anom			-.923		
CFP Pig anomalies yes or no (1=yes,2=no)			.839		
OCT large drusen yes or no (1=yes,2=no).				.906	
CFP large drusen yes or no (1=yes,2=no).				.875	
CFP large drusen (0-6).	.499			-.541	
OCT large drusen (0-6)	.466			-.510	
age					1.006

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 8 iterations.

Figure 10.8 Pattern Matrix from SPSS for the second factor rotation.

Third (and final) factor rotation:

Component Correlation Matrix

Component	1	2	3	4	5
1	1.000	-.083	-.083	-.330	.406
2	-.083	1.000	.211	.185	-.122
3	-.083	.211	1.000	.213	-.164
4	-.330	.185	.213	1.000	-.175
5	.406	-.122	-.164	-.175	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Figure 10.9 Component correlation Matrix from SPSS for the third (and final) factor rotation.

Pattern Matrix[a]

	Component				
	1	2	3	4	5
CFP inter drusen (0-6)	.939				
OCT inter drusen (0-6)	.888				
OCT small drusen (0-6)	.866				
OCT RPD yes or no (1=yes,2=no)	-.763				
CFP disc diam GA		-.981			
OCT disc diam GA		-.968			
CFP GA yes or no (1=yes, 2=no)		.719			
Symptoms ARMD (1=yes, 2=no)		.665			
CFP disc diam pig anom			-.924		
CFP Pig anomalies yes or no (1=yes,2=no)			.829	.301	
OCT large drusen yes or no (1=yes,2=no).				.896	
CFP large drusen yes or no (1=yes,2=no).				.859	
age					1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Figure 10.10 Pattern Matrix from SPSS for the third (and final) factor rotation

Appendix 4

18 top ranking variables for explaining the variance within the data for FAF imaging normality/abnormality.

Grading of drusen number throughout this study (0-6) indicates: 0 = no drusen, 1 ≤ 10, 2 ≤ 20, 3 ≤ 30, 4 ≤ 40, 5 ≤ 50, 6 > 50.

OCT large drusen (0-6)
CFP simplified severity score
OCT large drusen present yes or no
CFP disc diameters of pigmentary abnormalities
CFP pigmentary abnormalities yes or no
CFP large drusen (0-6)
CFP large drusen yes or no
OCT intermediate drusen (0-6)
OCT large drusen presence other (fellow) eye yes or no
Patient age
CFP large drusen presence other (fellow) eye yes or no
CFP disc diameters of GA
OCT small drusen (0-6)
Symptoms of ARMD
OCT reticular pseudodrusen presence yes or no
CFP intermediate drusen (0-6)
OCT disc diameters of GA
CFP GA yes or no

References

1. Suetsugu T, Kato A, Yoshida M, Yasukawa T, Nishiwaki A, Hasegawa N, et al. Evaluation of peripheral fundus autofluorescence in eyes with wet age-related macular degeneration. Clinical Ophthalmology. 2016;10:2497-503.
2. Yehoshua Z, de Amorim Garcia Filho CA, Nunes RP, Gregori G, Penha FM, Moshfeghi AA, et al. Comparison of Geographic Atrophy Growth Rates Using Different Imaging Modalities in the COMPLETE Study. Ophthalmic surgery, lasers & imaging retina. 2015;46(4):413-22.

3. Meleth AD, Sen HN. Use of fundus autofluorescence in the diagnosis and management of uveitis. Int Ophthalmol Clin. 2012;52(4):45-54.
4. Ben Moussa N, Georges A, Capuano V, Merle B, Souied EH, Querques G. MultiColor imaging in the evaluation of geographic atrophy due to age-related macular degeneration. British Journal of Ophthalmology. 2015;99(6):842-7.
5. Bertolotto M, Borgia L, Iester M. Hyperautofluorescence in Outer Retinal Layers Thinning. Biomed Research International. 2014.
6. Bhutto I, Lutty G. Understanding age-related macular degeneration (AMD): Relationships between the photoreceptor/retinal pigment epithelium/Bruch's membrane/choriocapillaris complex. Molecular Aspects of Medicine. 2012;33(4):295-317.
7. Anderson DH, Mullins RF, Hageman GS, Johnson LV. Perspective - A role for local inflammation in the formation of drusen in the aging eye. American Journal of Ophthalmology. 2002;134(3):411-31.
8. Bora PS, Sohn JH, Cruz JMC, Jha P, Nishihori H, Wang Y, et al. Role of complement and complement membrane attack complex in laser-induced choroidal neovascularization. Journal of Immunology. 2005;174(1):491-7.
9. Cachulo L, Silva R, Fonseca P, Pires I, Carvajal-Gonzalez S, Bernardes R. Early markers of choroidal neovascularization in the fellow eye of patients with unilateral exudative age-related macular degeneration. 2011;225:144-9.
10. Danis RP, Lavine JA, Domalpally A. Geographic atrophy in patients with advanced dry age-related macular degeneration: current challenges and future prospects. Clinical Ophthalmology. 2015;9:2159-74.
11. Bird AEC, Bressler NM, Bressler SB, Chisholm IH, Coscas G, Davis MD, et al. AN INTERNATIONAL CLASSIFICATION AND GRADING SYSTEM FOR AGE-RELATED MACULOPATHY AND AGE-RELATED MACULAR DEGENERATION. Survey of Ophthalmology. 1995;39(5):367-74.
12. Yung M, Klufas MA, Sarraf D. Clinical applications of fundus autofluorescence in retinal disease. Int J Retina Vitreous. 2016;2:12.
13. Delori FC, Dorey CK, Staurenghi G, Arend O, Goger DG, Weiter JJ. IN-VIVO FLUORESCENCE OF THE OCULAR FUNDUS EXHIBITS RETINAL-PIGMENT EPITHELIUM LIPOFUSCIN CHARACTERISTICS. Investigative Ophthalmology & Visual Science. 1995;36(3):718-29.
14. Jablonski A. Efficiency of Anti-Stokes Fluorescence in Dyes. Nature. 1933;131(3319):839-40.
15. Lacowicz J. The Principles of Fluorescence Spectroscopy. New York: Springer; 2006.
16. Jacobkhed diagram of absorbance, non-radiative decay, and fluorescence. [Internet]. 2002 [cited 22/01/2023]. Available from: https://commons.wikimedia.org/wiki/File:Jablonski_Diagram_of_Floouresence_Only.png (22/01/2023).
17. Nafar Z, Wen R, Guan ZQ, Li YW, Jiao SL. Quantifying lipofuscin in retinal pigment epithelium in vivo by visible-light optical coherence tomography-based multimodal imaging. Scientific Reports. 2020;10(1).
18. Van Schaik HJ, Alkemade C, Swart W, Van Best JA. Autofluorescence of the diabetic and healthy human cornea in vivo at different excitation wavelengths. Experimental Eye Research. 1999;68(1):1-8.
19. Sparrow JM, Bron AJ, Brown NAP, Neil HAW. AUTOFLUORESCENCE OF THE CRYSTALLINE LENS IN EARLY AND LATE ONSET DIABETES. British Journal of Ophthalmology. 1992;76(1):25-31.
20. Sasamoto Y, Gomi F, Sawa M, Sakaguchi H, Tsujikawa M, Nishida K. Effect of Cataract in Evaluation of Macular Pigment Optical Density by Autofluorescence Spectrometry. Investigative Ophthalmology & Visual Science. 2011;52(2):927-32.

21. Abiko T, Abiko A, Ishiko S, Takeda M, Horiuchi S, Yoshida A. Relationship between autofluorescence and advanced glycation end products in diabetic lenses. Experimental Eye Research. 1999;68(3):361-6.
22. Yung M, Klufas MA, Sarraf D. Clinical applications of fundus autofluorescence in retinal disease. International Journal of Retina and Vitreous. 2016;2(1):12.
23. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunovic H. Artificial intelligence in retina. Progress in Retinal and Eye Research. 2018;67:1-29.
24. Lu W, Tong Y, Yu Y, Xing YQ, Chen CZ, Shen Y. Applications of Artificial Intelligence in Ophthalmology: General Overview. Journal of Ophthalmology. 2018;2018.
25. O'Leary DE. Artificial Intelligence and Big Data. Ieee Intelligent Systems. 2013;28(2):96-9.
26. Baro E, Degoul S, Beuscart R, Chazard E. Toward a Literature-Driven Definition of Big Data in Healthcare. Biomed Research International. 2015;2015.
27. Boyd D, Crawford K. Critical questions for big data - Provocations for a cultural, technological, and scholarly phenomenon. Informacios Tarsadalom. 2012;12(2):7-+.
28. Chen HC, Chiang RHL, Storey VC. BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. Mis Quarterly. 2012;36(4):1165-88.
29. Dumbill E. MAKING SENSE OF BIG DATA. Big Data. 2013;1(1):1-+.
30. Jacobs A. The Pathologies of Big Data. Communications of the Acm. 2009;52(8):36-44.
31. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine. 2018;24(9):1342-+.
32. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. Plos One. 2017;12(5).
33. Ahn JM, Kim S, Ahn KS, Cho SH, Lee KB, Kim US. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. Plos One. 2018;13(11).
34. Pan SJ, Yang QA. A Survey on Transfer Learning. Ieee Transactions on Knowledge and Data Engineering. 2010;22(10):1345-59.
35. Kermany DS, Goldbaum M, Cai WJ, Valentim CCS, Liang HY, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018;172(5):1122-+.
36. Treder M, Lauermann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. Graefes Archive for Clinical and Experimental Ophthalmology. 2018;256(2):259-65.
37. Devalla SK, Subramanian G, Pham TH, Wang XF, Perera S, Tun TA, et al. A Deep Learning Approach to Denoise Optical Coherence Tomography Images of the Optic Nerve Head. Scientific Reports. 2019;9.
38. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the Acm. 2017;60(6):84-90.
39. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision. 2015;115(3):211-52.
40. He KM, Zhang XY, Ren SQ, Sun J, Ieee, editors. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. IEEE International Conference on Computer Vision; 2015 Dec 11-18; Santiago, CHILE2015.
41. Qiu JT, Wang J, Yao S, Guo KY, Li BX, Zhou EJ, et al. Going Deeper with Embedded FPGA Platform for Convolutional Neural Network. Proceedings of the 2016 Acm/Sigda International Symposium on Field-Programmable Gate Arrays (Fpga'16). 2016:26-35.
42. Kedar KB, Nannaware SK, Sathavane GV, Saoji C, Patil P. Significance of Artificial Intelligence in ophthalmology: Review Article. Annals of Medical and Health Sciences Research. 2021;11:169-73.

43. Ferro Desideri L, Rutigliani C, Corazza P, Nastasi A, Roda M, Nicolo M, et al. The upcoming role of Artificial Intelligence (AI) for retinal and glaucomatous diseases. Journal of Optometry. 2022;15:S50-S7.
44. Frampton GK, Kalita N, Payne L, Colquitt J, Loveman E. Accuracy of fundus autofluorescence imaging for the diagnosis and monitoring of retinal conditions: a systematic review. Health Technology Assessment. 2016;20(31):1-+.
45. Holz FG, Bindewald-Wittich A, Fleckenstein M, Dreyhaupt J, Scholl HP, Schmitz-Valckenberg S. Progression of geographic atrophy and impact of fundus autofluorescence patterns in age-related macular degeneration. Am J Ophthalmol. 2007;143:463-72.
46. Chan H, Cougnard-Gregoire A, Delyfer MN, Combiflet F, Rougier MB, Schweitzer C, et al. Multimodal Imaging of Reticular Pseudodrusen in a Population-Based Setting: The Alienor Study. Investigative Ophthalmology & Visual Science. 2016;57(7):3058-65.
47. Schmitz-Valckenberg S, Braun MD, Thiele S, Ferrara D, Honigberg L, Gao SS, et al. Conversion from Intermediate Age-Related Macular Degeneration to Geographic Atrophy in a Proxima B Subcohort Using a Multimodal Approach. Ophthalmologica. 2021;244(6):523-34.
48. Kolomeyer AM, Nayak NV, Szirth BC, Khouri AS. Fundus autofluorescence imaging in an ocular screening program. Int J Telemed Appl. 2012;2012:806464.
49. Horani M, Mahmood S, Aslam TM. Macular Atrophy of the Retinal Pigment Epithelium in Patients with Neovascular Age-Related Macular Degeneration: What is the Link? Part I: A Review of Disease Characterization and Morphological Associations. Ophthalmology and Therapy. 2019;8(2):235-49.
50. Wu ZC, Ayton LN, Luu CD, Baird PN, Guymer RH. Reticular Pseudodrusen in Intermediate Age-Related Macular Degeneration: Prevalence, Detection, Clinical, Environmental, and Genetic Associations. Investigative Ophthalmology & Visual Science. 2016;57(3):1310-6.
51. Kim M, Kim ES, Seo KH, Yu SY, Kwak HW. Change of retinal pigment epithelial atrophy ater anti-vascular endothelial growth factor treatment in exudative age-related macular degeneration. Indian Journal of Ophthalmology. 2016;64(6):427-33.
52. Delori FC. SPECTROPHOTOMETER FOR NONINVASIVE MEASUREMENT OF INTRINSIC FLUORESCENCE AND REFLECTANCE OF THE OCULAR FUNDUS. Applied Optics. 1994;33(31):7439-52.
53. Holz FG, Bellman C, Staudt S, Schutt F, Volcker HE. Fundus autofluorescence and development of geographic atrophy in age-related macular degeneration. Investigative Ophthalmology & Visual Science. 2001;42(5):1051-6.
54. von Ruckmann A, Fitzke FW, Bird AC. Distribution of pigment epithelium autofluorescence in retinal disease state recorded in vivo and its change over time. Graefes Archive for Clinical and Experimental Ophthalmology. 1999;237(1):1-9.
55. Krebs I. Noemi Lois and John V. Forrester: Fundus Autofluorescence. Graefe's Archive for Clinical and Experimental Ophthalmology. 2011;249(2):309-.
56. Curcio CA, Sloan KR, Kalina RE, Hendrickson AE. HUMAN PHOTORECEPTOR TOPOGRAPHY. Journal of Comparative Neurology. 1990;292(4):497-523.
57. Holz FG, Bellmann C, Margaritidis M, Schutt F, Otto TP, Volcker HE. Patterns of increased in vivo fundus autofluorescence in the junctional zone of geographic atrophy of the retinal pigment epithelium associated with age-related macular degeneration. Graefes Archive for Clinical and Experimental Ophthalmology. 1999;237(2):145-52.
58. Eldred GE, Lasky MR. RETINAL AGE PIGMENTS GENERATED BY SELF-ASSEMBLING LYSOSOMOTROPIC DETERGENTS. Nature. 1993;361(6414):724-6.
59. Schutt F, Davies S, Kopitz J, Holz FG, Boulton ME. Photodamage to human RPE cells by A2-E, a retinoid component of lipofuscin. Investigative Ophthalmology & Visual Science. 2000;41(8):2303-8.
60. Sparrow JR, Nakanishi K, Parish CA. The lipofuscin fluorophore A2E mediates blue light-induced damage to retinal pigmented epithelial cells. Investigative Ophthalmology & Visual Science. 2000;41(7):1981-9.

61. Delori FC, Goger DG, Dorey CK. Age-related accumulation and spatial distribution of lipofuscin in RPE of normal subjects. Investigative Ophthalmology & Visual Science. 2001;42(8):1855-66.
62. Sparrow JR, Vollmer-Snarr HR, Zhou JL, Jang YP, Jockusch S, Itagaki Y, et al. A2E-epoxides damage DNA in retinal pigment epithelial cells - Vitamin E and other antioxidants inhibit A2E-epoxide formation. Journal of Biological Chemistry. 2003;278(20):18207-13.
63. Wu YL, Yanase E, Feng XD, Siegel MM, Sparrow JR. Structural characterization of bisretinoid A2E photocleavage products and implications for age-related macular degeneration. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(16):7275-80.
64. Sparrow JR, Cai BL, Jang YP, Zhou JL, Nakanishi K. A2E, a fluorophore of RPE lipofuscin, can destabilize membrane. Retinal Degenerative Diseases. 2006;572:63-8.
65. Wing GI Fau - Blanchard GC, Blanchard Gc Fau - Weiter JJ, Weiter JJ. The topography and age relationship of lipofuscin concentration in the retinal pigment epithelium. Investigative Ophthalmological Science. 1978;17(7)(0146-0404 (Print)):601-7.
66. Feeneyburns L, Hilderbrand ES, Eldridge S. AGING HUMAN RPE - MORPHOMETRIC ANALYSIS OF MACULAR, EQUATORIAL, AND PERIPHERAL CELLS. Investigative Ophthalmology & Visual Science. 1984;25(2):195-200.
67. Slotnick S, Sherman J. Panoramic Autofluorescence: Highlighting Retinal Pathology. Optometry and Vision Science. 2012;89(5):E575-E84.
68. Lois N, Owens SL, Coco R, Hopkins J, Fitzke FW, Bird AC. Fundus autofluorescence in patients with age-related macular degeneration and high risk of visual loss. American journal of ophthalmology. 2002;133(3):341-9.
69. Delori FC. Spectrophotometer for non-invasive measurement of intrinsic fluorescence and reflectance of the ocular fundus. 1994;33:7429-52.
70. Keilhauer CN, Delori FC. Near-infrared autofluorescence imaging of the fundus: Visualization of ocular melanin. Investigative Ophthalmology & Visual Science. 2006;47(8):3556-64.
71. Burke JM, Kaczara P, Skumatz CMB, Zareba M, Raciti MW, Sarna T. Dynamic analyses reveal cytoprotection by RPE melanosomes against non-photic stress. Molecular Vision. 2011;17(309-12):2864-77.
72. Wang Z, Dillon J, Gaillard ER. Antioxidant properties of melanin in retinal pigment epithelial cells. Photochemistry and Photobiology. 2006;82(2):474-9.
73. Warrant EJ, Nilsson DE. Absorption of white light in photoreceptors. Vision Research. 1998;38(2):195-207.
74. Morgan JIW, Pugh EN. Scanning Laser Ophthalmoscope Measurement of Local Fundus Reflectance and Autofluorescence Changes Arising from Rhodopsin Bleaching and Regeneration. Investigative Ophthalmology & Visual Science. 2013;54(3):2048-59.
75. Silva R, Cachulo ML, Fonseca P, Bernardes R, Nunes S, Vilhena N, et al. Age-Related Macular Degeneration and Risk Factors for the Development of Choroidal Neovascularisation in the Fellow Eye: A 3-Year Follow-Up Study. Ophthalmologica. 2011;226(3).
76. McBain VA, Townend J, Lois N. Fundus autofluorescence in exudative age-related macular degeneration. British Journal of Ophthalmology. 2007;91(4):491-6.
77. Einbock W, Moessner A, Schnurrbusch UEK, Holz FG, Wolf S, Grp FAMS. Changes in fundus autofluorescence in patients with age-related maculopathy. Correlation to visual function: a prospective study. Graefes Archive for Clinical and Experimental Ophthalmology. 2005;243(4):300-5.
78. Schmitz-Valckenberg S, Bindewald-Wittich A Fau - Dolar-Szczasny J, Dolar-Szczasny J Fau - Dreyhaupt J, Dreyhaupt J Fau - Wolf S, Wolf S Fau - Scholl HPN, Scholl Hp Fau - Holz FG, et al. Correlation between the area of increased autofluorescence surrounding geographic atrophy and disease progression in patients with AMD. (0146-0404 (Print)).

79. Smith RT, Chan JK, Busuoic M, Sivagnanavel V, Bird AC, Chong NV. Autofluorescence characteristics of early, atrophic, and high-risk fellow eyes in age-related macular degeneration. 2006;47:5495-504.
80. Bandello F, Sacconi R, Querques L, Corbelli E, Cicinelli MV, Querques G. Recent advances in the management of dry age-related macular degeneration: A review. F1000Res. 2017;6:245-.
81. Fang V, Gomez-Caraballo M, Lad EM. Biomarkers for Nonexudative Age-Related Macular Degeneration and Relevance for Clinical Trials: A Systematic Review. Molecular Diagnosis & Therapy. 2021;25(6):691-713.
82. Göbel AP, Fleckenstein M, Heeren TF, Holz FG, Schmitz-Valckenberg S. In-vivo mapping of drusen by fundus autofluorescence and spectral-domain optical coherence tomography imaging. Graefes Arch Clin Exp Ophthalmol. 2016;254(1):59-67.
83. Olcay K, Cakir A, Sonmez M, Duzgun E, Yildirim Y. Analysing the Progression Rates of Macular Lesions with Autofluorescence Imaging Modes in Dry Age-Related Macular Degeneration. Turk Oftalmoloji Dergisi-Turkish Journal of Ophthalmology. 2015;45(6):235-8.
84. Fleckenstein M, Adrion C, Schmitz-Valckenberg S, Göbel AP, Bindewald-Wittich A, Scholl HP, et al. Concordance of disease progression in bilateral geographic atrophy due to AMD. Investigative ophthalmology & visual science. 2010;51(2):637-42.
85. Sunness JS, Gonzalez-Baron J, Applegate CA, Bressler NM, Tian Y, Hawkins B, et al. Enlargement of atrophy and visual acuity loss in the geographic atrophy form of age-related macular degeneration. Ophthalmology. 1999;106(9):1768-79.
86. Batioglu F, Oguz YG, Demirel S, Ozmert E. Geographic Atrophy Progression in Eyes with Age-Related Macular Degeneration: Role of Fundus Autofluorescence Patterns, Fellow Eye and Baseline Atrophy Area. Ophthalmic Research. 2014;52(2):53-9.
87. Bearelly S, Khanifar AA, Lederer DE, Lee JJ, Ghodasra JH, Stinnett SS, et al. Use of fundus autofluorescence images to predict geographic atrophy progression. Retina. 2011;31(1):81-6.
88. Hwang JC, Chan JW, Chang S, Smith RT. Predictive value of fundus autofluorescence for development of geographic atrophy in age-related macular degeneration. Invest Ophthalmol Vis Sci. 2006;47(6):2655-61.
89. Schmitz-Valckenberg S, Bindewald-Wittich A, Dolar-Szczasny J, Dreyhaupt J, Wolf S, Scholl HP. Correlation between the area of increased autofluorescence surrounding geographic atrophy and disease progression in patients with AMD. Invest Ophthalmol Vis Sci. 2006;47:2648-54.
90. Yonekawa Y, Miller JW, Kim IK. Age-Related Macular Degeneration: Advances in Management and Diagnosis. Journal of Clinical Medicine. 2015;4(2):343-59.
91. Nunes RP, Gregori G, Yehoshua Z, Stetson PF, Feuer W, Moshfeghi AA, et al. Predicting the Progression of Geographic Atrophy in Age-Related Macular Degeneration With SD-OCT En Face Imaging of the Outer Retina. Ophthalmic Surgery Lasers & Imaging Retina. 2013;44(4):344-59.
92. Laiginhas R, Liu J, Shen MX, Shi YY, Trivizki O, Waheed NK, et al. Multimodal Imaging, OCT B-Scan Localization, and En face OCT Detection of Macular Hyperpigmentation in Eyes with Intermediate AMD. Investigative Ophthalmology & Visual Science. 2022;63(7).
93. Domalpally A, Danis R, Agrón E, Blodi B, Clemons T, Chew E. Evaluation of Geographic Atrophy from Color Photographs and Fundus Autofluorescence Images: age-Related Eye Disease Study 2 Report Number 11. Ophthalmology. 2016;123(11):2401-7.
94. Zarubina AV, Neely DC, Clark ME, Huisingh CE, Samuels BC, Zhang YH, et al. Prevalence of Subretinal Drusenoid Deposits in Older Persons with and without Age-Related Macular Degeneration, by Multimodal Imaging. Ophthalmology. 2016;123(5):1090-100.
95. Elfandi S, Ooto S, Ueda-Arakawa N, Takahashi A, Yoshikawa M, Nakanishi H, et al. Clinical and Genetic Characteristics of Japanese Patients with Age-Related Macular Degeneration and Pseudodrusen. Ophthalmology. 2016;123(10):2205-12.

96. Monge M, Araya A, Wu LH. Subretinal drusenoid deposits: An update. Taiwan Journal of Ophthalmology. 2022;12(2):138-46.
97. Gil JQ, Marques JP, Hogg R, Rosina C, Cachulo ML, Santos A, et al. Clinical features and long-term progression of reticular pseudodrusen in age-related macular degeneration: findings from a multicenter cohort. Eye. 2017;31(3):364-71.
98. Marsiglia M, Boddu S, Bearelly S, Xu LN, Breaux BE, Freund KB, et al. Association Between Geographic Atrophy Progression and Reticular Pseudodrusen in Eyes With Dry Age-Related Macular Degeneration. Investigative Ophthalmology & Visual Science. 2013;54(12):7362-9.
99. Shijo T, Sakurada Y, Yoneyama S, Sugiyama A, Kikushima W, Tanabe N, et al. Prevalence and characteristics of pseudodrusen subtypes in advanced age-related macular degeneration. Graefes Archive for Clinical and Experimental Ophthalmology. 2017;255(6):1125-31.
100. Dutheil C, Le Goff M, Cougnard-Grégoire A, Gattoussi S, Korobelnik JF, Rougier MB, et al. Incidence and risk factors of reticular pseudodrusen using multimodal imaging. JAMA Ophthalmology. 2020.
101. Arnold JJ, Sarks SH, Killingsworth MC, Sarks JP. RETICULAR PSEUDODRUSEN - A RISK FACTOR IN AGE-RELATED MACULOPATHY. Retina-the Journal of Retinal and Vitreous Diseases. 1995;15(3):183-91.
102. Querques G, Querques L, Forte R, Massamba N, Coscas F, Souied EH. Choroidal Changes Associated with Reticular Pseudodrusen. Investigative Ophthalmology & Visual Science. 2012;53(3):1258-63.
103. Smailhodzic D, Fleckenstein M, Theelen T, Boon CJF, van Huet RAC, de Ven J, et al. Central Areolar Choroidal Dystrophy (CACD) and Age-Related Macular Degeneration (AMD): Differentiating Characteristics in Multimodal Imaging. Investigative Ophthalmology & Visual Science. 2011;52(12):8908-18.
104. Lindner M, Nadal J, Mauschitz MM, Lüning A, Czauderna J, Pfau M, et al. Combined fundus autofluorescence and near infrared reflectance as prognostic biomarkers for visual acuity in foveal-sparing geographic atrophy. Investigative Ophthalmology and Visual Science. 2017;58(6):BIO61-BIO7.
105. Rodrigo-Diaz E, Tahir HJ, Kelly JM, Parry NRA, Aslam T, Murray IJ. The Light and the Dark of Early and Intermediate AMD: Cone- and Rod-Mediated Changes Are Linked to Fundus Photograph and FAF Abnormalities. Investigative Ophthalmology & Visual Science. 2019;60(15):5070-9.
106. Kim JH, Chang YS, Kim JW, Lee TG, Kim CG. PREVALENCE OF SUBTYPES OF RETICULAR PSEUDODRUSEN IN NEWLY DIAGNOSED EXUDATIVE AGE-RELATED MACULAR DEGENERATION AND POLYPOIDAL CHOROIDAL VASCULOPATHY IN KOREAN PATIENTS. Retina-the Journal of Retinal and Vitreous Diseases. 2015;35(12):2604-12.
107. Joachim N, Mitchell P, Rochtchina E, Tan AG, Wang JJ. Incidence and Progression of Reticular Drusen in Age-related Macular Degeneration Findings from an Older Australian Cohort. Ophthalmology. 2014;121(4):917-25.
108. Klein R, Meuer SM, Knudtson MD, Iyengar SK, Klein BEK. The epidemiology of retinal reticular drusen. American Journal of Ophthalmology. 2008;145(2):317-26.
109. Smith RT, Sohrab MA, Busuioc M, Barile G. Reticular Macular Disease. American Journal of Ophthalmology. 2009;148(5):733-43.
110. Yasukawa T, Mori R, Sawa M, Shinojima A, Hara C, Sekiryu T, et al. Fundus autofluorescence and retinal sensitivity in fellow eyes of age-related macular degeneration in Japan. Plos One. 2019;14(2).
111. Zhang XY, Lai TYY. Baseline Predictors of Visual Acuity Outcome in Patients with Wet Age-Related Macular Degeneration. Biomed Research International. 2018;2018.
112. Yang JY, Yuan MZ, Wang EQ, Xia S, Chen YX. Noninvasive multimodal imaging in diagnosing polypoidal choroidal vasculopathy. Bmc Ophthalmology. 2019;19(1).

113. Wolf S, Arend O, Schulte K, Reim M. SEVERE ANAPHYLACTIC REACTION AFTER INDOCYANINE GREEN FLUORESCENCE ANGIOGRAPHY. American Journal of Ophthalmology. 1992;114(5):638-9.
114. Tan CS, Heussen F, Sadda SR. Peripheral Autofluorescence and Clinical Findings in Neovascular and Non-neovascular Age-related Macular Degeneration. Ophthalmology. 2013;120(6):1271-7.
115. Fragiotta S, Kaden TR, Freund KB. Cuticular drusen associated with aneurysmal type 1 neovascularization (polypoidal choroidal vasculopathy). International Journal of Retina and Vitreous. 2018;4(1):44.
116. Fragiotta S, Fernandez-Avellaneda P, Breazzano MP, Scuderi G. Clinical Manifestations of Cuticular Drusen: Current Perspectives. Clinical Ophthalmology. 2021;15:3877-87.
117. Yannuzzi LA, Negrao S, Iida T, Carvalho C, Rodriguez-Coleman H, Slakter J, et al. Retinal angiomatous proliferation in age-related macular degeneration. Retina-the Journal of Retinal and Vitreous Diseases. 2001;21(5):416-34.
118. Martin DF, Maguire MG, Fine SL, Ying G-s, Jaffe GJ, Grunwald JE, et al. Ranibizumab and Bevacizumab for Treatment of Neovascular Age-related Macular Degeneration: Two-Year Results. Ophthalmology. 2012;119(7):1388-98.
119. Holz FG, Sadda SR, Staurenghi G, Lindner M, Bird AC, Blodi BA, et al. Imaging Protocols in Clinical Studies in Advanced Age-Related Macular Degeneration. Ophthalmology. 2017;124(4):464-78.
120. Saito M, Kano M, Itagaki K, Oguchi Y, Sekiryu T. Retinal pigment epithelium tear after intravitreal aflibercept injection. Clinical ophthalmology (Auckland, NZ). 2013;7:1287-9.
121. Lapierre-Landry M, Carroll J, Skala MC. Imaging retinal melanin: a review of current technologies. Journal of Biological Engineering. 2018;12.
122. Yamamoto M, Kohno T, Ataka S, Mini N, Hishida E, Hirabayashi M, et al. Comparison of Fundus Autofluorescence in Age-Related Macular Degeneration Between Fundus Camera and Confocal Scanning Laser Ophthalmoscope Photographs. IOVS. 2006;47:ARVO E-abstract 5688.
123. Lois N, Forrester V. Fundus Autofluorescence. 1st edn. ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2009.
124. Borrelli EA-O, Lei J, Balasubramanian S, Uji A, Cozzi M, Sarao V, et al. Green emission fluorophores in eyes with atrophic age-related macular degeneration: a colour fundus autofluorescence pilot study. (1468-2079 (Electronic)).
125. Vujosevic S, Toma C, Sarao V, Veritti D, Brambilla M, Muraca A, et al. Color Fundus Autofluorescence to Determine Activity of Macular Neovascularization in Age-Related Macular Degeneration. (2164-2591 (Electronic)).
126. Wustemeyer H, Moessner A, Jahn C, Wolf S. Macular pigment density in healthy subjects quantified with a modified confocal scanning laser ophthalmoscope. Graefes Archive for Clinical and Experimental Ophthalmology. 2003;241(8):647-51.
127. Delori FC, Goger DG, Hammond BR, Snodderly DM, Burns SA. Macular pigment density measured by autofluorescence spectrometry: comparison with reflectometry and heterochromatic flicker photometry. Journal of the Optical Society of America a-Optics Image Science and Vision. 2001;18(6):1212-30.
128. Weigert G, Kaya S, Pemp B, Sacu S, Lasta M, Werkmeister RM, et al. Effects of Lutein Supplementation on Macular Pigment Optical Density and Visual Acuity in Patients with Age-Related Macular Degeneration. Investigative Ophthalmology & Visual Science. 2011;52(11):8174-8.
129. Bone RA, Fau LJ, Cains A. Optical density spectra of the macular pigment in vivo and in vitro. Vision, Res. 1992(0042-6989 (Print)).
130. Patel SN, Shi A, Wibbelsman TD, Klufas MA. Ultra-widefield retinal imaging: an update on recent advances. Therapeutic Advances in Ophthalmology. 2020;12.
131. Madhusudhan S, Beare N. Wide-field fluorescein angiography in wet age-related macular degeneration. ScientificWorldJournal. 2014;2014:536161.

132. Shoughy SS, Arevalo JF, Kozak I. Update on wide- and ultra-widefield retinal imaging. Indian Journal of Ophthalmology. 2015;63(7):575-81.
133. Kucukiba K, Erol N, Bilgin M. Evaluation of Peripheral Retinal Changes on Ultra-Widefield Fundus Autofluorescence Images of Patients with Age-Related Macular Degeneration. Turk Oftalmoloji Dergisi-Turkish Journal of Ophthalmology. 2020;50(1):6-14.
134. Guduru A, Fleischman D, Shin S, Zeng DL, Baldwin JB, Houghton OM, et al. Ultra-widefield fundus autofluorescence in age-related macular degeneration. Plos One. 2017;12(6).
135. Delori F, Greenberg JP, Woods RL, Fischer J, Duncker T, Sparrow J, et al. Quantitative Measurements of Autofluorescence with the Scanning Laser Ophthalmoscope. Investigative Ophthalmology & Visual Science. 2011;52(13):9379-90.
136. Delori FC, Duncker T, Sparrow JR. Fundus Autofluorescence. Lois N, Forrester JV, editors. Philadelphia: Wolters Kluwer; 2016.
137. Sparrow JR, Duncker T, Schuerch K, Paavo M, de Carvalho JRL. Lessons learned from quantitative fundus autofluorescence. Progress in Retinal and Eye Research. 2020;74:100774.
138. van de Kraats J, van Norren D. Optical density of the aging human ocular media in the visible and the UV. Journal of the Optical Society of America a-Optics Image Science and Vision. 2007;24(7):1842-57.
139. Bone RA, Landrum JT, Fernandez L, Tarsis SL. ANALYSIS OF THE MACULAR PIGMENT BY HPLC - RETINAL DISTRIBUTION AND AGE STUDY. Investigative Ophthalmology & Visual Science. 1988;29(6):843-9.
140. Greenberg JP, Duncker T, Woods RL, Smith RT, Sparrow JR, Delori FC. Quantitative Fundus Autofluorescence in Healthy Eyes. Investigative Ophthalmology & Visual Science. 2013;54(8):5684-93.
141. Gliem M, Muller PL, Finger RP, McGuinness MB, Holz FG, Issa PC. Quantitative Fundus Autofluorescence in Early and Intermediate Age-Related Macular Degeneration. Jama Ophthalmology. 2016;134(7):817-24.
142. Ly A, Nivison-Smith L, Zangerl B, Assaad N, Kalloniatis M. Advanced imaging for the diagnosis of age-related macular degeneration: a case vignettes study. Clinical and Experimental Optometry. 2018;101(2):243-54.
143. Forte R, Querques G, Querques L, Leveziel N, Benhamou N, Souied EH. MULTIMODAL EVALUATION OF FOVEAL SPARING IN PATIENTS WITH GEOGRAPHICATROPHY DUE TO AGE-RELATED MACULAR DEGENERATION. Retina-the Journal of Retinal and Vitreous Diseases. 2013;33(3):482-9.
144. Khanifar AA, Lederer DE, Ghodasra JH, Stinnett SS, Lee JJ, Cousins SW, et al. COMPARISON OF COLOR FUNDUS PHOTOGRAPHS AND FUNDUS AUTOFLUORESCENCE IMAGES IN MEASURING GEOGRAPHIC ATROPHY AREA. Retina-the Journal of Retinal and Vitreous Diseases. 2012;32(9):1884-91.
145. Sayegh RG, Simader C, Scheschy U, Montuoro A, Kiss C, Sacu S, et al. A Systematic Comparison of Spectral-Domain Optical Coherence Tomography and Fundus Autofluorescence in Patients with Geographic Atrophy. Ophthalmology. 2011;118(9):1844-51.
146. Schmitz-Valckenberg S, Fleckenstein M, Gobel AP, Hohman TC, Holz FG. Optical Coherence Tomography and Autofluorescence Findings in Areas with Geographic Atrophy Due to Age-Related Macular Degeneration. Investigative Ophthalmology & Visual Science. 2011;52(1):1-6.
147. Shinojima A, Sawa M, Mori R, Sekiryu T, Oshima Y, Kato A, et al. Five-year follow-up of fundus autofluorescence and retinal sensitivity in the fellow eye in exudative age-related macular degeneration in Japan. PLoS ONE. 2020;15(3).
148. Sasamoto Y, Gomi F, Sawa M, Tsujikawa M, Nishida K. Effect of 1-year lutein supplementation on macular pigment optical density and visual function. Graefe's Archive for Clinical and Experimental Ophthalmology. 2011;249(12):1847-54.

149. Chew EY, Clemons TE, SanGiovanni JP, Danis R, Ferris FL, Elman M, et al. Lutein plus Zeaxanthin and Omega-3 Fatty Acids for Age-Related Macular Degeneration The Age-Related Eye Disease Study 2 (AREDS2) Randomized Clinical Trial. Jama-Journal of the American Medical Association. 2013;309(19):2005-15.
150. De Luca LM. Beta-carotene increases lung cancer incidence in cigarette smokers. Nutrition Reviews. 1996;54(6):178-80.
151. Tong Y, Ami TB, Hong S, Heintzmann R, Gerig G, Ablonczy Z, et al. Hyperspectral autofluorescence imaging of drusen and retinal pigment epithelium in donor eyes with age-related macular degeneration. Retina. 2016;36:S127-S36.
152. Lin LY, Zhou Q, Hagstrom S, Maguire MG, Daniel E, Grunwald JE, et al. Association of Single-Nucleotide Polymorphisms in Age-Related Macular Degeneration With Pseudodrusen Secondary Analysis of Data From the Comparison of AM D Treatments Trials. Jama Ophthalmology. 2018;136(6):682-8.
153. Lou LX, Hu KM, Jin K, Zhang SZ, Ye J. The Relationship Between Hepatic Lipase Gene Variant and Advanced Age-Related Macular Degeneration A Meta-analysis. Jama Ophthalmology. 2014;132(10):1226-31.
154. Merle BMJ, Maubaret C, Korobelnik JF, Delyfer MN, Rougier MB, Lambert JC, et al. Association of HDL-Related Loci with Age-Related Macular Degeneration and Plasma Lutein and Zeaxanthin: the Alienor Study. Plos One. 2013;8(11).
155. Barbosa DTQ, Mendes TS, Cintron-Colon HR, Wang SY, Bhisitkul RB, Singh K, et al. Age-related macular degeneration and protective effect of HMG Co-A reductase inhibitors (statins): results from the National Health and Nutrition Examination Survey 2005-2008. Eye. 2014;28(4):472-80.
156. Tan JSL, Mitchell P, Rochtchina E, Wang JJ. Statins and the long-term risk of incident age-related macular degeneration: The Blue Mountains Eye Study. American Journal of Ophthalmology. 2007;143(4):685-7.
157. Tsao SW, Fong DS. Do Statins Have a Role in the Prevention of Age-Related Macular Degeneration? Drugs & Aging. 2013;30(4):205-13.
158. Mukkamala SK, Costa RA, Fung A, Sarraf D, Gallego-Pinazo R, Freund KB. Optical coherence tomographic imaging of sub-retinal pigment epithelium lipid. Archives of Ophthalmology. 2012;130(12):1547-53.
159. Deckert A, Schmitz-Valckenberg S, Jorzik J, Bindewald A, Holz FG, Mansmann U. Automated analysis of digital fundus autofluorescence images of geographic atrophy in advanced age-related macular degeneration using confocal scanning laser ophthalmoscopy (cSLO). BMC Ophthalmology. 2005;5(1):8.
160. Schmitz-Valckenberg S, Jorzik J, Roider J, Weinberger A, Wolf S, Holz FG. Analysis of digital fundus autofluorescence photos using scanning laser ophthalmoscopy of geographical atrophies (GA) in AMD (FAM-study). Ophthalmologe. 2001;98(Suppl 1):S157-8.
161. Wikipedia. 2023 [Available from: https://en.wikipedia.org/wiki/East_Riding_of_Yorkshire#Demographics.
162. Buitendijk GHS, Rochtchina E, Myers C, van Duijn CM, Lee KE, Klein BEK, et al. Prediction of Age-related Macular Degeneration in the General Population The Three Continent AMD Consortium. Ophthalmology. 2013;120(12):2644-55.
163. Foster PJ, Ng WS, Nolan WP, Tanner L, Gazzard G, Day AC, et al. Prevention of angle-closure glaucoma: balancing risk and benefit. Eye. 2022;36(12):2229-31.
164. Srivastava AK, Gupta V, Misra RK. Significance of van herick test and its comparison with gonioscopy grading. Indian Journal of Clinical and Experimental Ophthalmology. 2017;3:11-4.
165. Bindewald A, Bird AC, Dandekar SS, Dolar-Szczasny J, Dreyhaupt J, Fitzke FW. Classification of fundus autofluorescence patterns in early age-related macular disease. Invest Ophthalmol Vis Sci. 2005;46:3309-14.

166. Nowroozzadeh MH, Moshksar S, Azimi A, Rasti A, Sedaghat A. Comparison of retinal vascular arcade trajectory between eyes with an idiopathic macular hole and the healthy fellow eye. International Ophthalmology. 2022;42(7):2219-25.
167. Drasdo N Fau - Fowler CW, Fowler CW. Non-linear projection of the retinal image in a wide-angle schematic eye. (0007-1161 (Print)).
168. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. (2008-6164 (Print)).
169. Godec P, Pancur M, Ilenic N, Copar A, Strazar M, Erjavec A, et al. Democratized image analytics by visual programming through integration of deep models and small-scale machine learning. Nature Communications. 2019;10.
170. Park SP, Siringo FS, Pensec N, Hong IH, Sparrow J, Barile G, et al. Comparison of Fundus Autofluorescence Between Fundus Camera and Confocal Scanning Laser Ophthalmoscope-based Systems. Ophthalmic Surgery Lasers & Imaging Retina. 2013;44(6):536-43.
171. Sadda SR, Borrelli E, Fan WY, Ebraheem A, Marion KM, Kwon S. Impact of mydriasis in fluorescence lifetime imaging ophthalmoscopy. Plos One. 2018;13(12).
172. Scanlon PH, Malhotra R, Thomas G, Foy C, Kirkpatrick JN, Lewis-Barned N, et al. The effectiveness of screening for diabetic retinopathy by digital imaging photography and technician ophthalmoscopy. Diabetic Medicine. 2003;20(6):467-74.
173. Bartling H, Wanger P, Martin L. Automated quality evaluation of digital fundus photographs. Acta Ophthalmologica. 2009;87(6):643-7.
174. Lin TC, Chiang YH, Hsu CL, Liao LS, Chen YY, Chen SJ. Image quality and diagnostic accuracy of a handheld nonmydriatic fundus camera: Feasibility of a telemedical approach in screening retinal diseases. Journal of the Chinese Medical Association. 2020;83(10):962-6.
175. Zheng J, Guo LY, Peng LH, Yang JJ, Li JQ, Liang QF, et al., editors. Fundus Image Based Cataract Classification. IEEE International Conference on Imaging Systems and Techniques (IST) / IEEE International School on Imaging; 2014 Oct 14-17; Santorini, GREECE2014.
176. Flaxman SR, Bourne RRA, Resnikoff S, Ackland P, Braithwaite T, Cicinelli MV, et al. Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis. Lancet Global Health. 2017;5(12):E1221-E34.
177. Goh JHL, Lim ZW, Fang XL, Anees A, Nusinovici S, Rim TH, et al. Artificial Intelligence for Cataract Detection and Management. Asia-Pacific Journal of Ophthalmology. 2020;9(2):88-95.
178. Thylefors B, Chylack LT, Konyama K, Sasaki K, Sperduto R, Taylor HR, et al. A simplified cataract grading system. Ophthalmic Epidemiology. 2002;9(2):83-95.
179. Pallant J. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS Program. London: McGraw-Hill Education; 2016.
180. Stremplewski P, Komar K, Szkulmowski M, Motoczynska M, Wojtkowski M, editors. High sensitive Fundus Autofluorescence imaging combined with speckle-free Optical Coherence Tomography. Conference on Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XVII at the SPIE Photonics West Symposia; 2013 Feb 04-06; San Francisco, CA2013.
181. Kolomeyer AM, Baumrind BR, Szirth BC, Shahid K, Khouri AS. Fundus autofluorescence and colour fundus imaging compared during telemedicine screening in patients with diabetes. Journal of Telemedicine and Telecare. 2013;19(4):209-12.
182. Winn B, Whitaker D, Elliott DB, Phillips NJ. FACTORS AFFECTING LIGHT-ADAPTED PUPIL SIZE IN NORMAL HUMAN-SUBJECTS. Investigative Ophthalmology & Visual Science. 1994;35(3):1132-7.
183. Artemenko AR, Filatova E, Vorobyeva YD, Do TP, Ashina M, Danilov AB. Migraine and light: A narrative review. Headache. 2022;62(1):4-10.
184. Wang Y, Tran T, Firl K, Huang N, Yasin O, van Kuijk F, et al. Quantitative fundus autofluorescence in smokers compared to non-smokers. (1096-0007 (Electronic)).

185. Li J, Zhao X, Chen S, Liu B, Li Y, Lian P, et al. Patterns of Fundus Autofluorescence in Eyes with Myopic Atrophy Maculopathy: A Consecutive Case Series Study. (1460-2202 (Electronic)).
186. Shahid H, Khan Jc Fau - Cipriani V, Cipriani V Fau - Sepp T, Sepp T Fau - Matharu BK, Matharu Bk Fau - Bunce C, Bunce C Fau - Harding SP, et al. Age-related macular degeneration: the importance of family history as a risk factor. (1468-2079 (Electronic)).
187. Powers D. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. School of Informatics and Engineering, Finders University. 2007.
188. Ferris FL, Davis MD, Clemons TE, Lee LY, Chew EY, Lindblad AS, et al. A simplified severity scale for age-related macular degeneration - AREDS report no. 18. Archives of Ophthalmology. 2005;123(11):1570-4.
189. Stevens R, Bartlett H, Cooke R. Evaluation of a clinical decision-making aid for nutrition advice in age-related macular degeneration. British Journal of Visual Impairment. 2017;35(3):185-96.
190. Diniz B, Ribeiro R, Heussen FM, Maia M, Sadda S. DRUSEN MEASUREMENTS COMPARISON BY FUNDUS PHOTOGRAPH MANUAL DELINEATION VERSUS OPTICAL COHERENCE TOMOGRAPHY RETINAL PIGMENT EPITHELIAL SEGMENTATION AUTOMATED ANALYSIS. Retina-the Journal of Retinal and Vitreous Diseases. 2014;34(1):55-62.
191. Morgan WH, Cooper RL, Constable IJ, Eikelboom RH. AUTOMATED EXTRACTION AND QUANTIFICATION OF MACULAR DRUSEN FROM FUNDAL PHOTOGRAPHS. Australian and New Zealand Journal of Ophthalmology. 1994;22(1):7-12.
192. Demsar J, Zupan B. Hands-on training about overfitting. Plos Computational Biology. 2021;17(3).
193. Chicco D. Ten quick tips for machine learning in computational biology. Biodata Mining. 2017;10.
194. Domingos P. A Few Useful Things to Know About Machine Learning. Communications of the Acm. 2012;55(10):78-87.
195. Kim DY, Loo J, Farsiu S, Jaffe GJ. COMPARISON OF SINGLE DRUSEN SIZE ON COLOR FUNDUS PHOTOGRAPHY AND SPECTRAL-DOMAIN OPTICAL COHERENCE TOMOGRAPHY. Retina-the Journal of Retinal and Vitreous Diseases. 2021;41(8):1715-22.
196. Lee J, Kang HG, Kim HR, Lee CS, Kim M, Kim SS, et al. Pigmentary abnormality without significant drusen as a risk factor for late age-related macular degeneration. Scientific Reports. 2022;12(1).
197. Davis MD, Gangnon RE, Lee LY, Hubbard LD, Klein BEK, Klein R, et al. The age-related eye disease study severity scale for age-related macular degeneration - AREDS report no. 17. Archives of Ophthalmology. 2005;123(11):1484-98.
198. Sarks SH. NEW VESSEL FORMATION BENEATH RETINAL-PIGMENT EPITHELIUM IN SENILE EYES. British Journal of Ophthalmology. 1973;57(12):951-65.
199. Metz CE. Basic principles of ROC analysis. Seminars in Nuclear Medicine. 1978;8(4):283-98.
200. Le D, Alam M, Yao CK, Lim JI, Hsieh YT, Chan RVP, et al. Transfer Learning for Automated OCTA Detection of Diabetic Retinopathy. Translational Vision Science & Technology. 2020;9(2).
201. Li YH, Feng WB, Zhao XJ, Liu BQ, Zhang Y, Chi W, et al. Development and validation of a deep learning system to screen vision-threatening conditions in high myopia using optical coherence tomography images. British Journal of Ophthalmology. 2022;106(5):633-9.
202. Karri SPK, Chakraborty D, Chatterjee J. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. Biomedical Optics Express. 2017;8(2):579-92.
203. Miere A, Le Meur T, Bitton K, Pallone C, Semoun O, Capuano V, et al. Deep Learning-Based Classification of Inherited Retinal Diseases Using Fundus Autofluorescence. Journal of Clinical Medicine. 2020;9(10).

204. Nguyen JM, Jezequel P, Gillois P, Silva L, Ben Azzouz F, Lambert-Lacroix S, et al. Random forest of perfect trees: concept, performance, applications and perspectives. Bioinformatics. 2021;37(15):2165-74.
205. Dong L, Yang Q, Zhang RH, Wei WB. Artificial intelligence for the detection of age-related macular degeneration in color fundus photographs: A systematic review and meta-analysis. Eclinicalmedicine. 2021;35.
206. Miere A, Capuano V, Kessler A, Zambrowski O, Jung C, Colantuono D, et al. Deep learning-based classification of retinal atrophy using fundus autofluorescence imaging. Computers in Biology and Medicine. 2021;130:104198.
207. Powell S. FDA approves drug for dry AMD in "milestone moment" Online: Optometry Today; 2023 [Available from: https://www.aop.org.uk/ot/professional-support/clinical-and-regulatory/2023/03/01/fda-approves-drug-for-dry-amd-in-milestone-moment.
208. Bindewald A, Schmitz-Valckenberg S, Jorzik JJ, Dolar-Szczasny J, Sieber H, Keilhauer C. Classification of abnormal fundus autofluorescence patterns in the junctional zone of geographic atrophy in patients with age related macular degeneration. Br J Ophthalmol. 2005;89:874-8.
209. Osborne J, Costello AB. Sample size and subject to item ratio in principal components analysis. Prac Assess Res Eval. 2004;9.
210. Velicer WF, Jackson DN. COMPONENT ANALYSIS VERSUS COMMON FACTOR-ANALYSIS - SOME FURTHER OBSERVATIONS. Multivariate Behavioral Research. 1990;25(1):97-114.
211. Suhr, D. Exploratory or Confirmatory Factor Analysis.: Cary: SAS Institute, Inc; 2006. p. 1-17.
212. Hutcheson S. Multivariate Social Scientist: An introduction to generalised linear models. Thousand Oaks, California.: Sage Publications; 1999.
213. Norusis, M. PASW Statistics 18 Statistical Procedures Companion. London: Pearson; 2005.
214. Guadagnoli E, Velicer WF. RELATION OF SAMPLE-SIZE TO THE STABILITY OF COMPONENT PATTERNS. Psychological Bulletin. 1988;103(2):265-75.
215. Pett, M. Making sense of factor analysis. Thousand Oaks: Sage Publications, Inc; 2003.
216. Tabachnick, B, Fidell, L. Using multivariate statistics. Needham Heights: Allyn & Bacon; 2001.
217. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods. 1999;4(3):272-99.
218. Fava JL, Velicer WF. THE EFFECTS OF OVEREXTRACTION ON FACTOR AND COMPONENT ANALYSIS. Multivariate Behavioral Research. 1992;27(3):387-415.
219. Costello AB, Osborne J. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. Practical Assessment, Research & Evaluation. 2005;10:1-9.
220. Henson RK, Roberts JK. Use of exploratory factor analysis in published research - Common errors and some comment on improved practice. Educational and Psychological Measurement. 2006;66(3):393-416.
221. Schonrock-Adema J, Heijne-Penninga M, van Hell EA, Cohen-Schotanus J. Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. Medical Teacher. 2009;31(6):E226-E32.
222. Garson, R. Factor Analysis (Statistical Associates "Blue Book" Series Book 15). Statistical Associates Publisher; 2013th Edition.; 2013.
223. Decoster, J. Overview of factors analysis. University of Alabama: University of Alabama; 1998.
224. Child, D. The essentials of factor analysis. 2nd ed. London: Cassell Educational Ltd; 1990.

225. Beavers, A, Lounsbury, J, Richards, J, et al. Practical considerations for using factor analysis in educational research. Practical Assessment, Research and Evaluation, 18(6). 2013.
226. Guymer RH, Mitchell P, Wong J, Heier J, Singh R, Steinle N, et al. Efficacy and safety of intravitreal pegcetacoplan in geographic atrophy: Results from the phase 3 DERBY and OAKS trials. Clinical and Experimental Ophthalmology. 2022;49(8):833-.
227. McHugh ML. Interrater reliability: the kappa statistic. Biochemia Medica. 2012;22(3)(1330-0962 (Print)):276-82.

END